

OBJECT DETECTION BASED ON SPECTRAL ANALYSIS USING SOBEL AND ROBERTS EDGE DETECTION ALGORITHM

Deva Hema D*, Sk Shahid Ali, Arnab Rooj, TanayYeole
SRM Institute of Science and Technology, Ramapuram, Chennai, India

*Email: devahemd@srmist.edu.in

Received: 10th September 2022, Accepted: 3rd October 2022 and Published: 3rd October 2022

ABSTRACT

Background: Object detection in crowded scenes is difficult because of the dense distribution and partial overlap of objects, and existing detectors do not consistently outperform one another across all scenarios.

Objective: This paper proposes a novel object detection approach based on a thorough examination of an image's details and its approximate density chart, using the Sobel and Roberts edge detection algorithm.

Methods: The approach is divided into two phases: spatial distribution knowledge from a density map is used to compute initial object positions, and a saliency map inspired by human attention to detail is then used to calculate precise bounding boxes, with a sparse defocus dictionary learned on a newly constructed dataset used to estimate a focus-prior map.

Results: After one epoch of training, the model predicts object density and bounding boxes with a mean square root error of 1533.91 and a mean absolute error of 1337.67.

Conclusion: The algorithm's output captures spatial features and allows the threshold type to be chosen in a variety of ways.

Keywords: Object Detection, Saliency Map, Sparse Defocus, Spatial Distribution, Sobel and Roberts Edge Detection Algorithm

HIGHLIGHTS:

1. Object detection based on spectral analysis using the Sobel and Roberts edge detection algorithm proved effective compared with existing methodologies.
2. A two-phase pipeline combines a density map for initial object positions with a saliency map for precise bounding-box estimation.
3. The proposed focus-prior map, based on sparse defocus dictionary learning, addresses scale variance caused by uncertain perspective.

INTRODUCTION

Object detection (OD) in crowded scenes is difficult because of the dense distribution and partial overlap of objects. This paper proposes a novel OD approach based on a thorough examination of an image's details and its approximate density chart. The proposed approach is divided into two phases: knowledge of the spatial distribution of objects, obtained from a density map, is used to compute initial object positions, and a saliency map that provides entity boundaries - inspired by human attention to detail - is then used to calculate the bounding boxes with precision.

Figure 1 shows the system architecture. The first step is to estimate the density of the input image using example-based density estimation. The image then passes through an object pre-localization phase, consisting of a sliding Region of Interest (ROI) window, computing the local maximum, and gathering localization results, which aids in understanding the image's region of interest. In the bounding-box estimation phase, the image goes through SLIC (Simple Linear Iterative Clustering) segmentation [1], foreground seeds, and K-means clustering, before the bounding box is estimated and the results are generated. SLIC is a superpixel-generating algorithm that clusters pixels in the image plane by colour similarity and proximity: given a target number of superpixels K , K cluster centres $C_k = [l_k, a_k, b_k, x_k, y_k]$ are chosen at regular grid intervals S , for $k = [1, K]$, and this is used to calculate the size of the bounding box. Sudden changes in pixel density are used to detect edges, using the Sobel and Roberts edge detection algorithms. The technologies used include Python, NumPy, Scikit-learn, Eclipse IDE, web technologies, and Bootstrap.

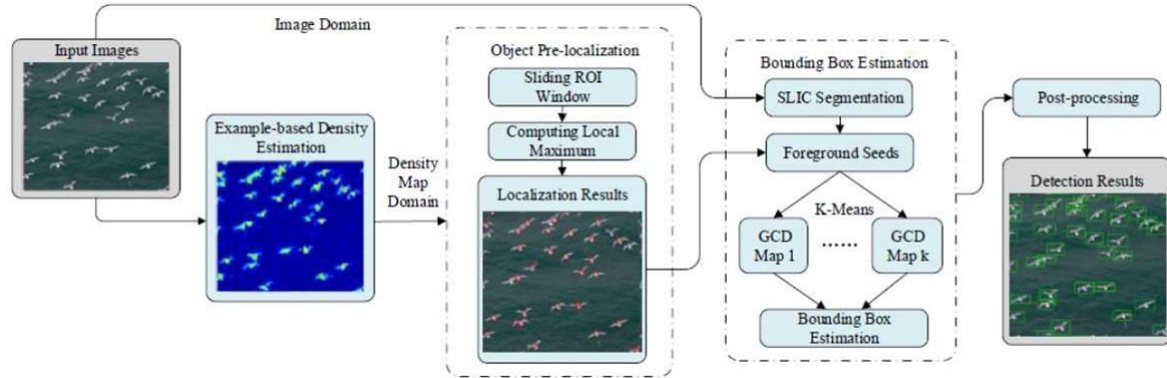


Figure 1: Proposed architecture diagram.

The scale variance of objects induced by uncertain perspective is a common problem in object density map estimation, and irrelevant objects and background context can introduce artefacts into the resulting density maps, adding to the error when density maps are aggregated into heat maps. To address this, a new method for estimating a focus-prior map for any image is proposed: unlike existing edge-density estimation approaches, it is based on sparse defocus dictionary learning on a newly constructed dataset, where the focus power is determined by the number of non-zero coefficients of the dictionary atoms.

MATERIALS AND METHODS

Related Work

Several approaches to object and people counting have been proposed in the literature. Sun et al. presented benchmark data and a method for real-time people counting in cluttered scenes using depth sensors [2], while Basalamah et al. proposed a Scale-Driven Convolutional Neural Network (SD-CNN) on the premise that heads, regardless of density, are the most identifiable feature in crowds [3]. Device-free counting using IoT sensors [4] and an IR-UWB radar sensor [5] have also been explored, alongside participatory smartphone-based crowd-management platforms [6].

Walach and Wolf combined CNNs with layered boosting and selective sampling to improve counting accuracy while reducing processing time [7], and multi-column convolutional neural networks have been used for single-image crowd counting under arbitrary density and perspective [8]. For boundary and saliency detection, Zhu et al. proposed boundary connectivity as a background indicator for robust saliency optimization [9], and Ren et al.'s Faster R-CNN enabled near-cost-free region proposals by sharing full-image convolutional features with the detection network [10]. Despite this range of predictors, no single group of features consistently outperforms the others across all scenarios and applications, motivating the combined density-map and saliency-map approach proposed here.

Sobel and Roberts Edge Detection

For edge detection, the Sobel filter is used. It calculates the image's strength gradient at each pixel, and defines the direction of the most significant transition from light to dark together with the rate of change in that direction: applied to a pixel in a constant-intensity field the output is a zero vector, while applied to a pixel on an edge it produces a vector pointing from darker to lighter values across the edge.

Two 3x3 kernels are used, one for changes along the horizontal axis and one for the vertical axis, and the two kernels are convolved with the original picture to approximate the derivatives. The horizontal kernel used is shown below.

Table 1: Sobel horizontal gradient kernel (G_x).

1	0	-1
2	0	-2
1	0	-1

The kernel (window) is placed over the input image, one pixel's value is computed, and the window is shifted one pixel to the right; this continues until the end of the row is reached, then steps down to the beginning of the next row, until the whole image has been processed.

The proposed system offers several advantages over existing approaches:

4. An image's spatial features are captured.
5. It is one of the more precise learning algorithms available, generating a highly accurate classifier for many datasets.
6. It performs better when the number of dimensions exceeds the number of available samples.

7. The threshold type can be chosen in a variety of ways.

RESULTS AND DISCUSSION

After one epoch of training, the model is able to predict points, as shown in Figure 2, which presents an input image (left) and its corresponding predicted density map (right).

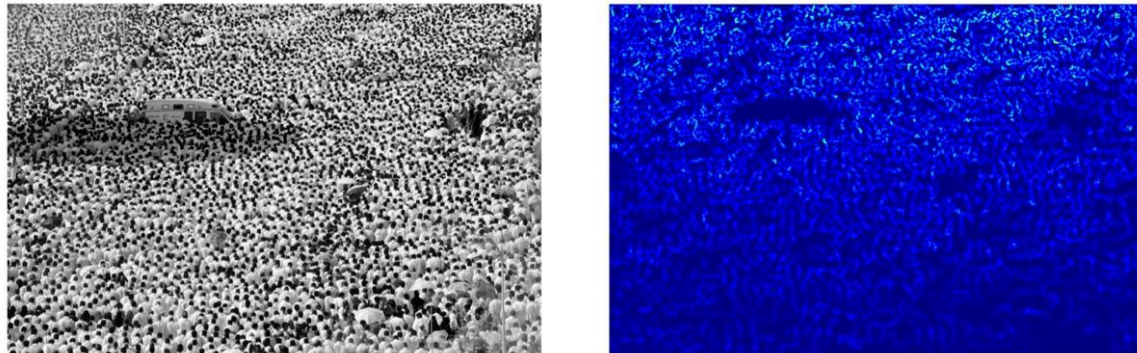


Figure 2: Image detection results — input image (left) and predicted density map (right).

The algorithm is one of the more precise learning algorithms evaluated, producing a highly accurate classifier across a range of datasets, and performs better as the number of dimensions grows relative to the number of available samples; it is also able to capture spatial features and to choose the threshold type in a variety of ways. After training the model and predicting the values, the mean square root error was 1533.91 and the mean absolute error was 1337.67, both obtained after one epoch of training.

CONCLUSION

We propose a method for detecting small objects that takes both the image and the predicted density map into consideration. The method benefits from the density map's knowledge of object spatial distribution while minimising the disadvantages of an obfuscated object boundary: although trained on dotted-annotation datasets, the saliency map's boundary information lets the approximate bounding box match the object accurately.

The proposed focus-prior map can be used as a focus feature, or as a prior, to improve salient object detection accuracy in image analysis. Because of perspective effects, future work will need to make the approach more adaptable to a wide range of individual object sizes, and to reduce the artefacts that complex backgrounds introduce into density-map calculations.

REFERENCES

1. Boemer, F., Ratner, E., & Lendasse, A. (2018). Parameter-free image segmentation with SLIC. *Neurocomputing*, 277, 228-236. <https://doi.org/10.1016/j.neucom.2017.05.096>
2. Sun, S., Akhtar, N., Song, H., Zhang, C., Li, J., & Mian, A. (2019). Benchmark data and method for real-time people counting in cluttered scenes using depth sensors. *IEEE Transactions on Intelligent Transportation Systems*, 20(10), 3599-3612.
3. Basalamah, S., Khan, S. D., & Ullah, H. (2019). Scale driven convolutional neural network model for people counting and localization in crowd scenes. *IEEE Access*, 7, 71576-71584.
4. Sobron, I., & Velez, M. (2018). Device-free people counting in IoT environments: New insights, results and open challenges. *IEEE Internet of Things Journal*, 5(6), 4396-4408.
5. Choi, J. W., & Cho, S. H. (2017). People counting based on an IR-UWB radar sensor. *IEEE Sensors Journal*, 17(17), 5717-5727.
6. Franke, T., Lukowicz, P., & Blanke, U. (2015). Smart crowds in smart cities: Real-life, city-scale deployments of a smartphone based participatory crowd management platform. *Journal of Internet Services and Applications*, 6(1), 27.
7. Walach, E., & Wolf, L. (2016). Learning to count with CNN boosting. *European Conference on Computer Vision (ECCV)*, Springer, 660-676.
8. Zhang, Y., Zhou, D., Chen, S., Gao, S., & Ma, Y. (2016). Single-image crowd counting via multi-column convolutional neural network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 589-597.
9. Zhu, W., Liang, S., Wei, Y., & Sun, J. (2014). Saliency optimization from robust background detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2814-2821.
10. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137-1149.