

A Replacement Algorithm of Non-Maximum Suppression Base on Graph Clustering

Xin Zhao¹, Yue-Qing Wang², Yong Dou³

¹National Laboratory for Parallel and Distributed Processing, National University of Defense Technology, Changsha, China

²National Laboratory for Parallel and Distributed Processing, National University of Defense Technology, Changsha, China

³National Laboratory for Parallel and Distributed Processing, National University of Defense Technology, Changsha, China

¹zhaoxin15a@nudt.edu.cn, ²yueqingwang@nudt.edu.cn, ³yongdou@nudt.edu.cn

Abstract: Non-maximum suppression is an important step in many object detection and object counting algorithms. In contrast with the extensive studies of object detection, NMS method has not caused too much attention. Although traditional NMS method has demonstrated promising performance in detection tasks, we observe that it is a hard decision approach, which only uses the confidential scores and Intersection-over-Unions (IoUs) to discard proposals. By this way, NMS method would keep many false proposals whose IoU with the ground truth proposal is smaller than the threshold, which indicates that NMS may not suitable for counting the object number in images. To eliminate the limitation on object counting task, we propose a novel algorithm base on graph clustering to replace the NMS method in this paper. Experiments on faster-rcnn and SSD show that our algorithm achieves better performance than that of NMS on the object counting task.

1. Introduction

Object detection has been extensively studied and applied to a variety of real tasks recently [5], [11]. All of these algorithms provide and score a set of object proposals, and then use the Non-Maximum Suppression (NMS) method to decide which proposals need to be removed, so NMS is considered as an important part of all object detection algorithms. The NMS method greedily merges the higher scoring proposals with lower scoring ones when the IoU between them is large enough (e.g. $\text{IoU} > 0.5$). Classical NMS method contains the following five steps: 1) sort scores in descending order; 2) select the proposal box with the maximum score; 3) calculate the IOUs between the selected proposal box and other ones; 4) remove the proposals whose IoUs are bigger than the IoU threshold; 5) repeat from step 2 to 4 until there is no proposal boxes in the residual set. NMS method is simple, fast, effective and easy to implement, so it is very popular. NMS can also be applied to object counting task. Compared to object detection task, the object counting task not only detects the objects, but also reports the number of objects.

Although the NMS method can efficiently remove some overlapping proposals and achieve promising performance on object detection tasks, we observe that: i) NMS method is a hard decision approach which only uses the confidential scores, IoUs and a threshold value to remove the overlapping proposals; ii) NMS may keep many false proposal boxes which have no real objects in

them. These issues limit the development of NMS method, and further adversely affect the performance of object counting task.

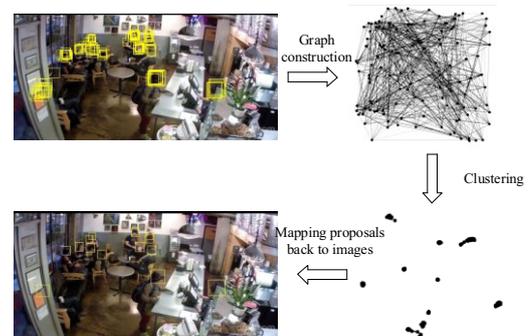


Figure 1. The framework of our approach. We first construct the graph using proposals and IoUs in one image, then cluster all proposals, and finally choose one proposal in each cluster as the final results.

To address these issues, we propose a novel algorithm based on graph clustering to replace NMS method. Compared to the NMS method, our algorithm significantly reduces the false proposals, and is more suitable for object counting task. As is shown in Figure 1, our algorithm first builds a graph which regards the proposals as vertexes, and IOU between proposals as the edges. After that, our algorithm performs a clustering method on all vertexes. Finally, for each clustering center, we choose the proposal with the highest score as the final kept proposals.

In addition, we find that the widely used criterions: MAP (Mean Average Precision) and recall are not suitable for evaluating the performance of object counting task. In fact, when there are many false proposals at last, the MAP and recall might be high, while the number of objects might be considerably different from the number of real objects. For evaluating the performance of object counting task more precisely, we also propose a new criterion.

Extensive experimental results show that our algorithm achieves better performance than NMS on object counting tasks, which indicates our new method is effective.

2. Related Works

There are only a few studies focus on the NMS method. In this section, we will give a brief introduction about these works. They can roughly be grouped into three categories.

The first category mainly concerns the running time of NMS, so they propose several methods to speed up the NMS method. In specific, Alexander Neubeck et.al. proposed several ways to speed up 1D and 2D NMS [7]. Their work shows that the dependency of the neighborhood can be reduced, and the number of comparisons can be pushed down. The work in [8] speeds up the classical NMS method using general purpose GPU on embedded architecture.

The second category focus on the replacement methods of NMS method. Mean-Shift [6] is widely used in the field of clustering [1], image smoothing, image segmentation, and especially object tracking. Li et.al. [14] speed it up on the GPU platform. Affinity propagation clustering algorithm (APC) [9] also needs not to know the number of clustering centers previously. In [11], Rothe etc. announce a new novel method which is an adapting APC method. The new method reduce the limitation of traditional NMS method on object detection task.

The third category tries to use a neural network to simulate the NMS process. In [3], Wan etc. integrate convolutional network, the deformable part model and NMS method into a single model. In [4], Hosang et.al. propose a convolutional neural network to replace NMS, called Tnet. Different from [3], [4] trained separately.

Although achieving promising performance, the above mentioned algorithms only focus on the object detection task, while ignore the object counting task. In our paper, we propose a novel algorithm based on graph clustering to replace the NMS method, which is suitable for the object counting task.

3. Our Approach based on Graph Clustering

In this section, we present the details of our approach based on graph clustering.

Our approach can be split into three stages. By regarding the proposal boxes and the IoU as the vertexes and the edges respectively, our approach firstly construct a graph for each image. After that, we design a novel

algorithm to cluster the vertexes in each graph, and group these vertexes into several clusters. Finally, we choose one vertex for each cluster. Note that if the score of the chosen proposal is too low, we would regard the chosen proposal as the background, and discard it.

Algorithm 1: Our approach based on graph clustering.

```

Input:  $\{score_i\}_{i=1}^N, E(i, j), i, j \in \{1, \dots, N\}, \epsilon, \eta$ 
Output:  $V_{keep}$ 
1 Add all proposal id into  $V_{unprocess}$ ;
2  $V_{final} \leftarrow []$ ;
3  $k \leftarrow 1$ ;
4 repeat
5   Sort scores of  $V_{unprocess}$  in a decrease order;
6   Choose the  $V_i$  with the highest score;
7   Assign  $k$  to be the label of  $V_i$ ;
8   for  $j = 1; j \leq \text{length}(V_{unprocess}); j++$  do
9     if  $E(i, j) > \epsilon$  then
10      Add  $V_j$  into  $V_{keep}$ ;
11      Set  $k$  to be the label of  $V_j$ ;
12      Delete  $V_j$  from  $V_{unprocess}$ ;
13    $k \leftarrow k + 1$ 
14 until  $V_{unprocess} == []$ ;
15 for  $k = 1; k \leq C; k++$  do
16   for  $j = k; j \leq C; j++$  do
17     if  $d(k, j) < \eta$  then
18      Merge two clusters.

```

We show our clustering method in Algorithm 1. In Algorithm 1, the $V_{unprocess}$ means the vertex set waiting for processing, $E(i, j)$ indicates the IoU between the proposal i and proposal j , and the V_{keep} is the output of our clustering method, which records the vertexes need to be kept in the image. At first, we add all vertexes into $V_{unprocess}$, and initialize V_{keep} to empty. Next, we sort the $V_{unprocess}$ in a decrease order, and choose the V_i who has the highest score. Then, we regard the V_i as the k^{th} clustering center, and find other proposals whose IoU with V_i is bigger than another threshold, and add these vertexes into cluster k . The algorithm would repeat these steps until $V_{unprocess}$ is empty. After that, the algorithm would merge the clustering centers if they are too close to separate. Finally, we would keep one vertex for each cluster, and map these vertexes into proposals. By this way, our method can reduce the false proposals significantly compared with NMS.

4. Experiment

To test the performance of our method, we test our method on four datasets. To evaluate the effectiveness of our approach, we choose two famous object detection frameworks: Faster-RCNN [10] and SSD [5], to produce the proposals and their scores. We only replace the NMS method of these two frameworks using our new method, and test the overall performance. The data sets used in our experiment is shown in Table 1, which includes one scene of the brainwash data set and three scenes in PDL data set. The Brainwash dataset is introduced by Stewart etc. in [12], which is collected from a busy cafe, containing 10461 images: 9977 images for training and 484 image for testing. PDL dataset is collected by ourselves using several different cameras. Our data set would be opened up in the future.

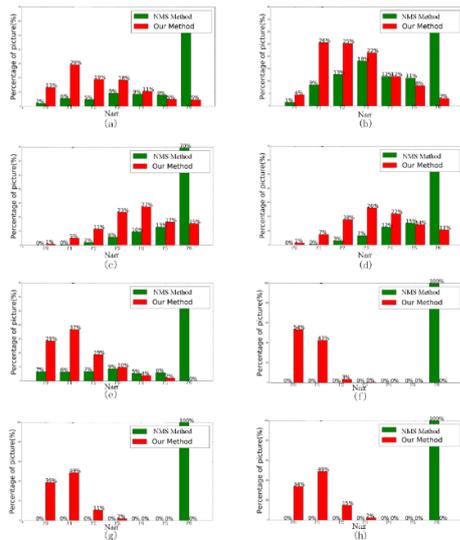


Figure 2. The performance of original method NMS method and our method. These subfigures present the performance of two methods a) in Faster-RCNN framework on brainwash dataset, b) in Faster-RCNN framework on PDL_scene1, c) in Faster-RCNN framework on PDL_scene2, d) in Faster-RCNN framework on PDL_scene3, e) in SSD framework on brainwash, f) in SSD framework on PDL_scene1, g) in SSD framework on PDL_scene2, h) in SSD framework on PDL_scene3.

Table 1. The Data Sets Used In Our Experiments.

| Dataset | Testing samples | Training samples | All samples |
|------------|-----------------|------------------|-------------|
| Brainwash | 484 | 9977 | 10463 |
| Pdl_scene1 | 293 | 0 | 293 |
| Pdl_scene2 | 515 | 0 | 515 |
| Pdl_scene3 | 498 | 0 | 498 |

To initialize our object detection model, we use the pre-trained model on ImageNet [15] as the initial model, and then fine tune the pre-trained model using our data sets. After the training process, we obtain the well trained model, and then we use the model to test on all test images.



Figure 3. Final proposals kept by NMS method and our method. a) The final proposals of NMS on one image of Brainwash data set. b) The final proposals of our method on the same image of Brainwash data set. c) The final proposals of NMS on one image of PDL_scene1. d) The final proposals of our method on the same image of PDL_scene1.

To report the performance of object counting task, we use an new criterion, called N_{diff} . Note that a smaller N_{diff} permits a better performance. For example, when N_{diff} equals to 0, it means the number of objects detected by the method matches the number of the real objects. In specific, we first calculate the difference between the number of real objects and number of detected objects for each image, and compute the absolute value of the difference: $N_{diff} = abs(N_{real} - N_{detected})$. As is described, if the method can correctly count the number of objects in one image, the N_{diff} would equal to 0. After that, we statistic the percentage with different N_{diff} in the test data set. For convenience, we record the percentage with different N_{diff} as P_0, P_1, \dots, P_6 , where P_i means the percentage of images when N_{diff} equals to i . Note that P_7 is the percentage of images whose N_{diff} is larger than 5.

From the definition of these criterions, we can conclude that P_0 means the percentage of images whose real number of objects equals to the detected number. Therefore, the larger the P_0 , the better the performance.

In addition, we also report the typical criterion used in object detection MAP of the NMS and our replacement method.

The MAP of classical NMS method and our method are reported in Table 2. As is shown in Table 2, our method achieves comparative MAP on pdl_scene1 and pdl_scene2 compared to the NMS method. However, on other data sets, NMS achieves better MAP than our method because the calculation of MAP does not consider the influence of false proposals.

Table 2 the map of nms and our methods in two object detection frameworks on four data sets.

| MAP | Faster-RCNN | | SSD | |
|------------|-------------|-------------|------------|------|
| | NMS Method | Ours | NMS Method | Ours |
| brainwash | 0.76 | 0.69 | 0.71 | 0.65 |
| pdl_scene1 | 0.90 | 0.89 | 0.77 | 0.70 |
| pdl_scene2 | 0.83 | 0.83 | 0.72 | 0.64 |
| pdl_scene3 | 0.85 | 0.85 | 0.75 | 0.66 |

The P_i ($i = 1, 2, 6$) of the two method are shown in Figure 2. As can be seen from Figure 2, the P_0 of our method equals to 13% for faster-rcnn and 26% for SSD, which is significantly larger than that of traditional NMS method on Brainwash data set. This means our method could count the objects in one image more accurate. Therefore, our method is more suitable for object counting task.

To compare the performance of the NMS method and our method more vividly, we draw the final proposals on the original images and choose several images to display the results. Figure 3 draws the final proposals of our method and NMS method. It can be seen from Figure 3 that: i) our method can count the number of objects more precisely, for example, in the PDL scene, the true number of people is 4, the number counted by our method is correct, while that of NMS is larger than 30; ii) the locations of the final kept proposals are more accurate than that of NMS method. We conclude from these results that NMS always keeps many false proposals, so it cannot

count the objects correctly in the image. By contrast, our clustering method can count the objects more precisely.

5. Conclusion

In this paper, we propose a novel method based on graph clustering to replace traditional NMS method, and apply this method on the object counting task. Our approach can be split into three stages. By regarding the proposal boxes and the IoU as the vertexes and the edges respectively, our approach firstly construct a graph for each image. After that, we design a novel algorithm to cluster the vertexes in each graph, and group these vertexes into several clusters. Finally, we choose one vertex for each cluster. Experiments on faster-rcnn and SSD show that our algorithm achieves better performance than that of NMS on object counting tasks.

References

- [1] Mean Shift, Mode Seeking, and Clustering. page 10, 1995.
- [2] Adrien Descamps and Cyril Carincotte. Counting people in the crowd using a generic head detector 1 Introduction 2 Previous Work 3 Overview of our approach.
- [3] David Eigen. “End-to-End Integration of a Convolutional Network , Deformable Parts Model and Non-Maximum Suppression,” (September 2010):5540096, 2011.
- [4] Jan Hosang, Rodrigo Benenson, and Bernt Schiele. “A convnet for non- maximum suppression,” pages 1–14, 2015.
- [5] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, and Scott Reed. “SSD: Single Shot MultiBox Detector,” Arxiv, pages 1–15, 2015.
- [6] Peter Meer. Mean Shift Analysis and Applications.
- [7] Alexander Neubeck and Luc Van Gool. “Efficient non-maximum suppression. Proceedings,” International Conference on Pattern Recognition, 3:850– 855, 2006.
- [8] David Oro, Xavier Martorell, and Javier Hernando. “WORK-EFFICIENT PARALLEL NON-MAXIMUM SUPPRESSION FOR EMBEDDED GPU”.
- [9] ARCHITECTURES Carles Fern Herta Security , Barcelona , Spain Uni- versitat Polit ‘ ecnica de Catalunya , Barcelona , Spain Barcelona Super- computing Center Email : david.oro@hertasecurity.com. Icassp 2016, pages 1026–1030, 2016.
- [10] Michael Pfeiffer. Clustering by Passing Messages between Data Points.2007.
- [11] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. “Faster R- CNN: Towards Real-Time Object Detection with Region Proposal Networks,” Nips, pages 1–10, 2015.
- [12] Rasmus Rothe, Matthieu Guillaumin, and Luc van Gool. “Non-maximum suppression for object detection by passing messages between windows,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9003:290–306, 2015.
- [13] Russell Stewart and Mykhaylo Andriluka. “End-to-end people detection in crowded scenes,” Arxiv, page 9, 2015.
- [14] Li P, Xiao L. “Mean Shift Parallel Tracking on GPU[C],” Iberian Conference on Pattern Recognition and Image Analysis. Springer-Verlag, 2009:120-12
- [15] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. “ImageNet: A Large-Scale Hierarchical Image Database,” In CVPR09, 2009