Deep Feature Deblurring Diffusion for Detecting Out-of-Distribution Objects: Supplementary Material

Aming Wu[1] Da Chen[2] Cheng Deng[1*]
[1]School of Electronic Engineering, Xidian University, Xi'an, China
[2]University of Bath
amwu@xidian.edu.cn, da.chen@bath.edu, chdeng@mail.xidian.edu.cn

1. Further Discussion of the OOD Map

To reduce the impact of lacking OOD data for training, we focus on synthesizing virtual OOD features that are close to the classification boundary of ID and OOD objects. In this paper, by gradually performing Gaussian Blur on the extracted features, the blurred features are related to the original input features and are difficult to be classified correctly. At this time, the synthesized OOD features are the expected features that are beneficial for improving the ability of detecting OOD objects.

Fig. 5 in the submitted paper shows some visualization examples. Compared with the Enhanced map, we can observe that **the synthesized OOD map contains plentiful input-related content. Meanwhile, it is hard to discriminate the corresponding categories.** Since there is no OOD-related information available, one feasible way is to make the OOD map to be different from the original map while retaining certain input-related information. To this end, by the blur operation, compared with the original feature, the synthesized features contain much less detail information but still involve certain content relevant to the input, which is conducive to strengthening the discrimination.

In the experiments, we separately evaluate our method on OOD-OD, Incremental Object Detection, and Open-Set Object Detection. The significant performance gains over baselines demonstrate the effectiveness of our method.

2. Analysis of the Deblurred Map

For our method, the reverse process aims to continually perform the deblurring operation to recover the lost detail in the forward blurring process. In Fig. 1, we show some deblurred maps. We can see that compared with the Enhanced maps, the deblurred maps contain plentiful object-related information. Meanwhile, we also observe that the deblurred results involve rich object-irrelevant noise information. Taking the deblurred results for training is instrumental in enhancing the discrimination ability of the object classifier, which is beneficial for keeping the superior performance of ID object detection.

*Corresponding author

3. More details of the training process

During training, since we could only access the supervision information of ID objects, for the input images, the features P_{in} of object proposals extracted by the RPN module are used as the ID features to calculate the loss $\mathcal{L}_{\text{in}}$ (Eq. (7)). The synthesized OOD features are utilized to train the classifier of discriminating OOD objects. Finally, by optimizing the objective $\mathcal{L}$ (Eq. (9)), a clear classification boundary for discriminating OOD and ID objects could be learned.

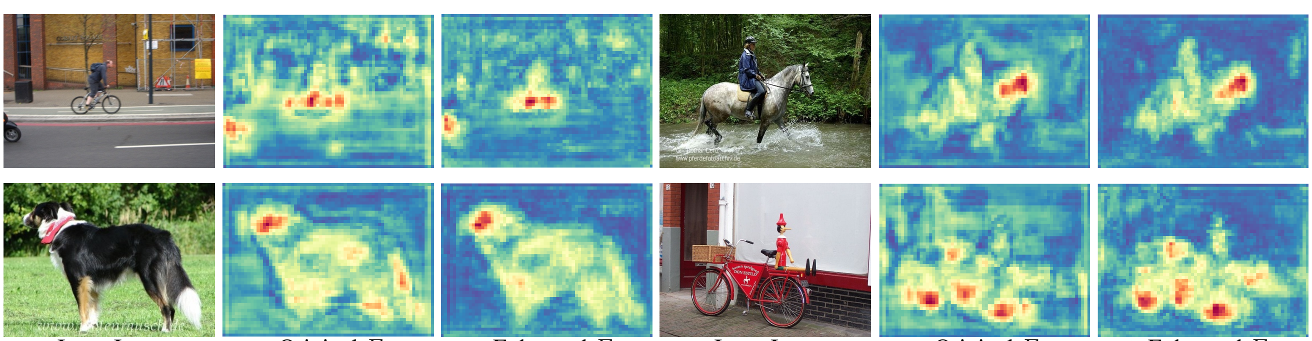
4. More Visualization Analysis

In Fig. 2 and 3, we show more visualization examples to further verify the proposed method. We can see that the enhanced map E involves plentiful object-related information and less object-irrelevant information (e.g., the background information). This indicates that the residual result between the original feature and the blurred feature indeed contains rich content about objects, which is instrumental in keeping the superior detection performance. Meanwhile, we can observe that the blurred features still involve certain input-related information but are hard to correctly classify these features. By taking these features as virtual OOD features, it is beneficial for improving the ability of discriminating OOD objects from ID objects.

In Fig. 4 and 5, we show more detection results. We can see that our method could detect OOD objects and ID objects accurately. Particularly, we observe that the categories of the OOD objects are very diverse. And the background style is different from the training data. At this time, only leveraging the ID data for detecting OOD objects is full of challenges. This further demonstrates that synthesizing specific virtual features is an effective solution for OOD-OD. And our method is conducive to generating expected virtual features to resolve the two challenges of OOD-OD.

(a) Input Image (b) Enhanced Map (c) Deblurred Map (d) Input Image (e) Enhanced Map (f) Deblurred Map

Figure 1. Visualization of the Enhanced map and the Deblurred map based on the OOD data (MS-COCO). For each feature map, the channels corresponding to the maximum value are selected for visualization. We can see that after the deblurring operation, the generated results contain rich object-related information.



(a) Input Image (b) Original F_0 (c) Enhanced E (d) Input Image (e) Original F_0 (f) Enhanced E

Figure 2. Visualization of the Enhanced map E (i.e., $E = \Psi([F_0, F_0 - F_T])$) and original map F_0 based on the ID data (PASCAL VOC). For each feature map, the channels corresponding to the maximum value are selected for visualization.

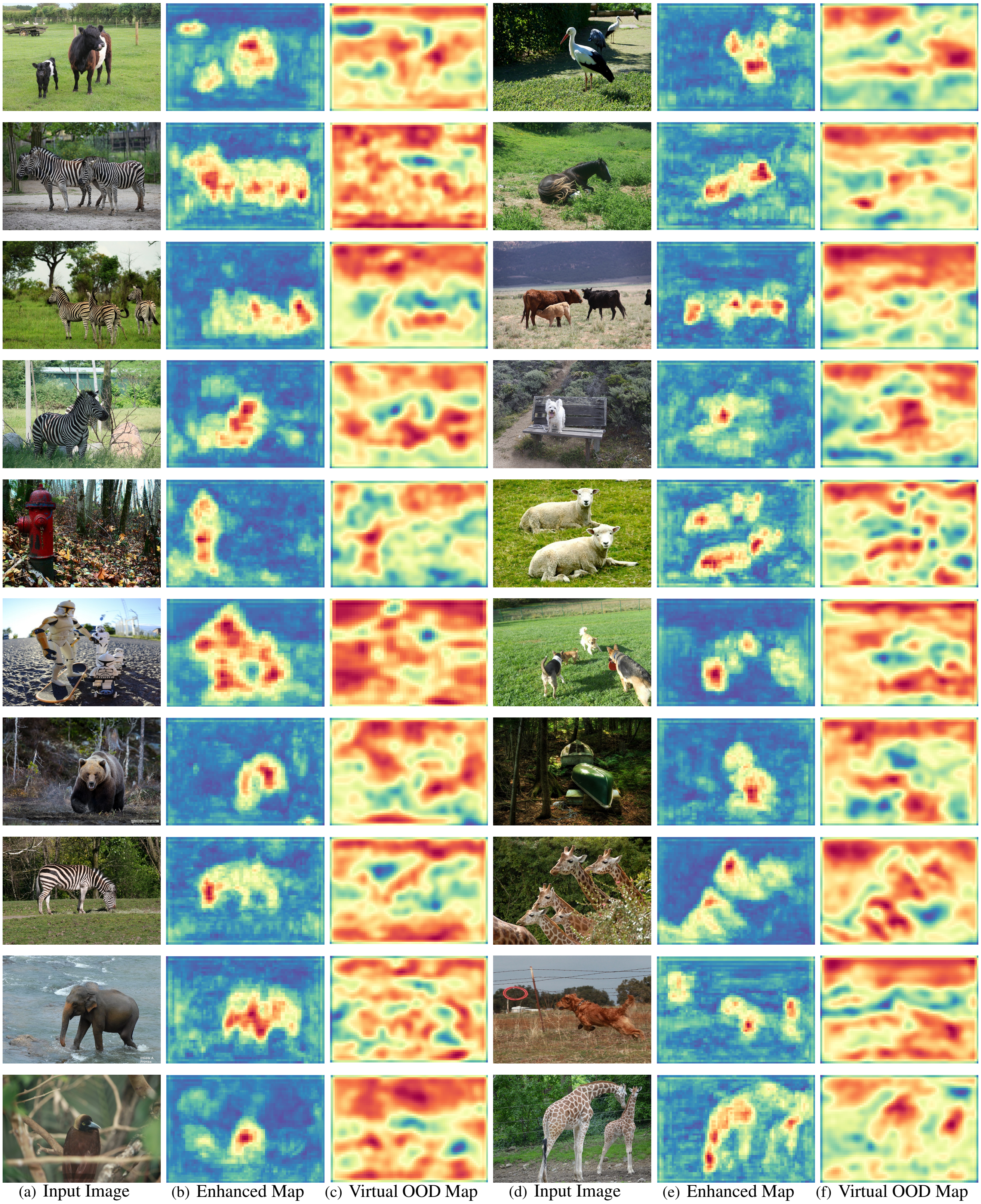


Figure 3. Visualization of the Enhanced map E (i.e., $E = \Psi([F_0, F_0 - F_T])$) and Virtual OOD map F_T (Eq. (3)) based on the OOD data (MS-COCO). For each feature map, the channels corresponding to the maximum value are selected for visualization.

Figure 4. Detection results on the OOD images from MS-COCO. We can see that our method detects OOD objects accurately, which further demonstrates the effectiveness of our method.

Figure 5. Detection results on In-Distribution dataset, i.e., PASCAL VOC. We can see that our method effectively detects objects in these images, which shows the superiorities of our method.