

# Supplementary Material for: MinMaxCAM: Improving object coverage for CAM-based Weakly Supervised Object Localization

Kaili Wang<sup>12</sup>

kaili.wang@esat.kuleuven.be

Jose Oramas<sup>2</sup>

Jose.Oramas@uantwerpen.be

Tinne Tuytelaars<sup>1</sup>

tinne.tuytelaars@esat.kuleuven.be

<sup>1</sup> KU Leuven, ESAT-PSI

Leuven, Belgium

<sup>2</sup> University of Antwerp, imec-IDLab

Antwerp, Belgium

In this supplementary material, we provide: 1) failure case analysis, 2) additional analysis for the proposed two regularizations, 3) additional qualitative results of our method on the ImageNet, CUB-200-2011 and OpenImages-segmentation datasets, and 4) additional implementation details to reproduce our results.

## 1 Failure case

We analyze some failure cases in this section, given the page limitation of the manuscript. The first type is caused by nature, which is unavoidable, like reflections of water. The cause of the second type is related to our first assumption for *CRR*, that different images from a class have very similar background. For example, brown creeper always appear with trunks in the image. Fig. 1 shows some examples.

## 2 Additional ablation study

In this section, we provide one more ablation study conducted on CUB dataset for different backbones. We plot **PxPrec vs. PxRec** in Fig. 2 (using the GT bounding box as proxy for the GT segmentation mask). For VGG (left), CAM-based localization maps are too small. Tuning the hyperparameters  $\lambda_1$  and  $\lambda_2$  in this case makes *FRR* dominate, which increases the object coverage, as evident from an increase in PxRec (curve shifts to the right). For MobileNet (right), we have the opposite situation: CAM-based localization maps are too large. Optimal hyperparameters in this case make *CRR* dominate, increasing the object coverage, as evident from an increase in PxPrec (curve shifts upwards). This experiment further confirms the effectiveness of *CRR* and *FRR* for the over- and under-estimation problems.



Figure 1: Failure cases of the proposed method.

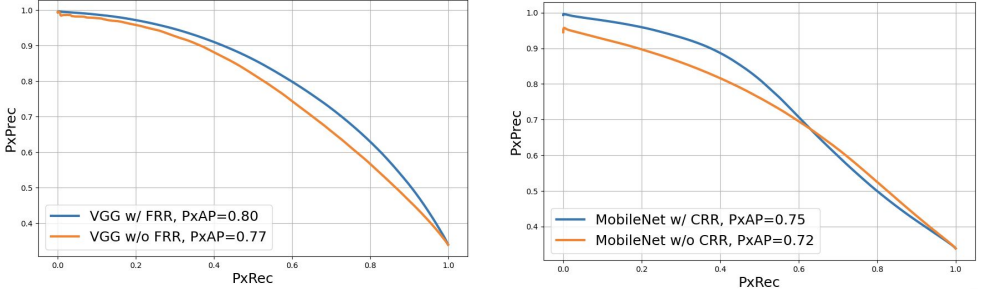


Figure 2: PxPrec vs. PxRec for VGG with and without *FRR*, MobileNet with and without *CRR*.

### 3 Additional qualitative results

Figure 3, Figure 4 and Figure 5 show additional qualitative results on the ImageNet, CUB and OpenImagaes datasets, respectively.

## 4 Implementation Details

Besides the information given in Sec. 4.2, in this section, we given more details. In order to speed up the training process on the ImageNet dataset, we randomly sample 50 images out of around 1000 images per class every epoch.

### 4.1 Learning rate $lr$

For the ImageNet dataset, the base  $lr_{base}$  is set to 0.00002, 0.00004 and 0.00008 for VGG16, Resnet50 and MobilenetV2.  $lr_{new}$ , the learning rate for the newly-added layers, is set to 100 times larger. In addition, the learning rate in stage I is set as twice as the  $lr_{base}$  and  $lr_{new}$  for the ResNet backbone.

For the CUB dataset, the base  $lr_{base}$  is set to 0.0001, 0.0002 and 0.0002 for VGG16, Resnet50 and MobilenetV2.  $lr_{new}$ , the learning rate for the newly-added layers, is set to 10 times larger.

For the OpenImages dataset, the base  $lr$  is set to 0.0002, 0.0004 and 0.0004 for VGG16, Resnet50 and MobilenetV2.  $lr_{new}$ , the learning rate for the newly-added layers, is set to 10 times larger.

## 4.2 Regularization weights $\lambda$

For the ImageNet dataset,  $\{\lambda_1, \lambda_2\}$  are set to  $\{5, 0\}$ ,  $\{0, 30\}$  and  $\{0, 20\}$  for VGG16, Resnet50 and MobilenetV2.

For the CUB dataset,  $\{\lambda_1, \lambda_2\}$  are set to  $\{10, 0\}$ ,  $\{5, 30\}$  and  $\{5, 20\}$  for VGG16, Resnet50 and MobilenetV2.

For the OpenImages dataset,  $\{\lambda_1, \lambda_2\}$  are set to  $\{10, 0\}$ ,  $\{0, 40\}$  and  $\{0, 20\}$  for VGG16, Resnet50 and MobilenetV2.

Please note we do not carefully tune the hyperparameters, the reported performance can be further improved when the optimal hyperparameters are found.

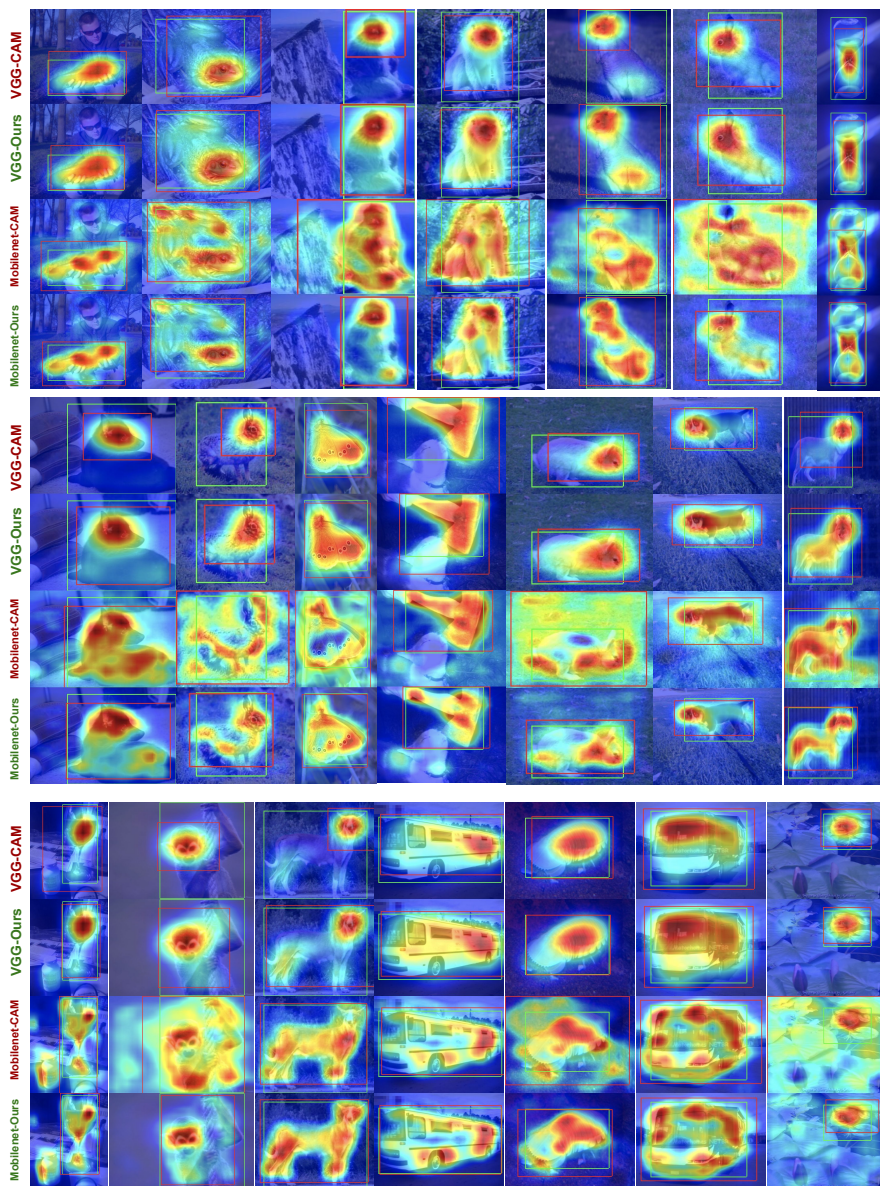


Figure 3: Additional qualitative results on the ImageNet dataset. Ground truth (green) and estimated (red) bounding box.



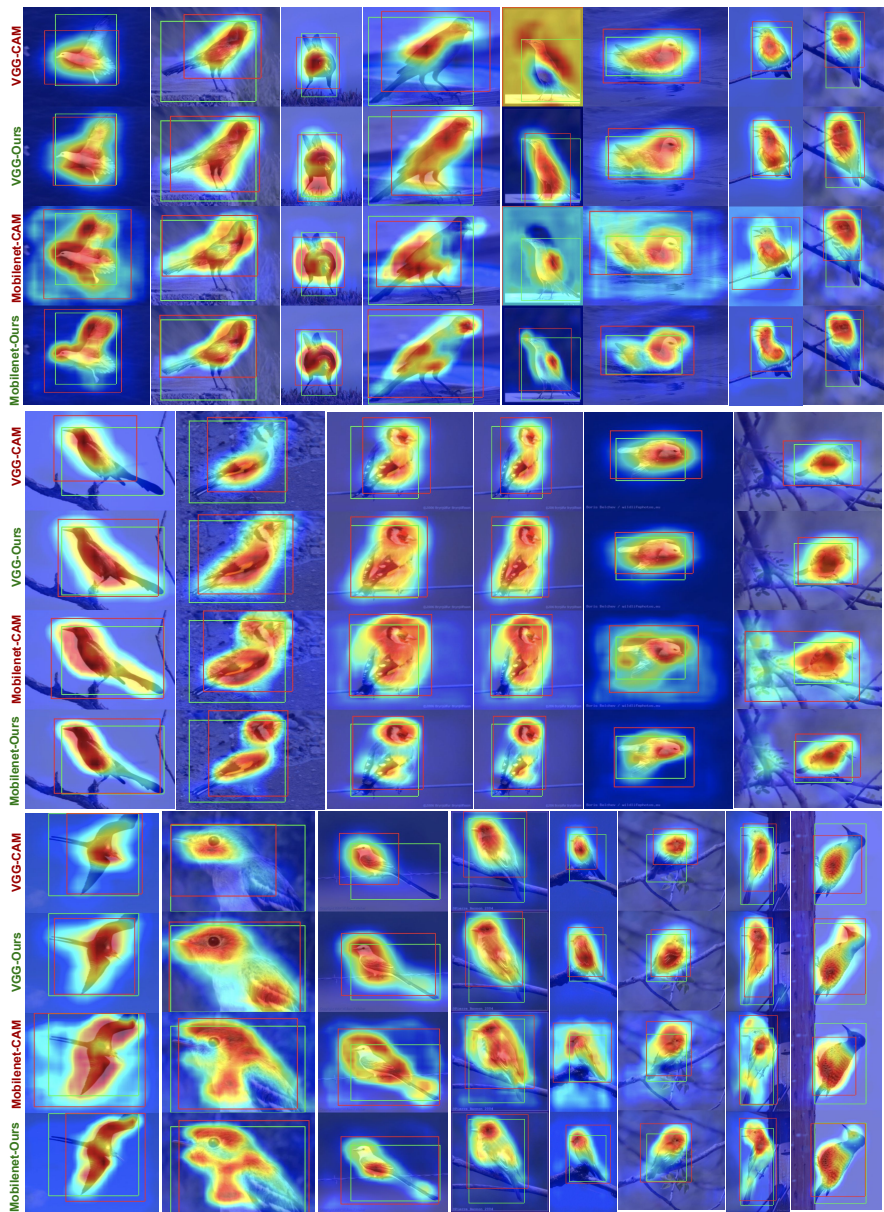


Figure 4: Additional qualitative results on the CUB-200-2011 dataset. Ground truth (green) and estimated (red) bounding box.

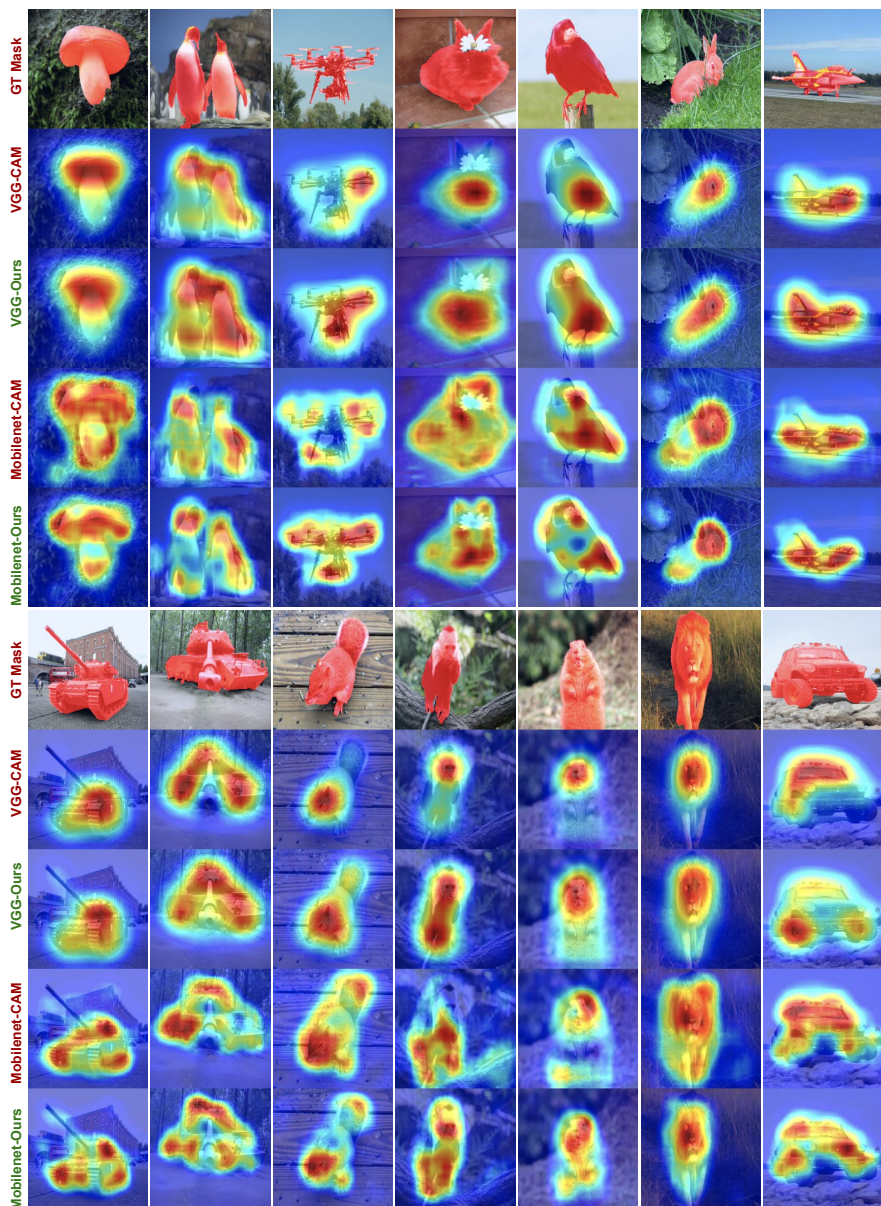


Figure 5: Additional qualitative results on the OpenImages-segmentation dataset.