Occlusion-Robustness of Convolutional Neural Networks via Inverted Cutout - Supplementary Material

Matthias Körschens
Computer Vision Group
Friedrich Schiller University Jena
07737 Jena, Germany
Email: matthias.koerschens@uni-jena.de

Paul Bodesheim
Computer Vision Group
Friedrich Schiller University Jena
07737 Jena, Germany
Email: paul.bodesheim@uni-jena.de

Joachim Denzler
Computer Vision Group
Friedrich Schiller University Jena
07737 Jena, Germany
Email: joachim.denzler@uni-jena.de

S-I. Introduction

In this supplementary document we provide several example images of both datasets used to provide a better insight into the different occlusion levels and kinds of occlusion used. Moreover, we also present the numerical results of our experiments again with the addition of the respective standard deviations to give a better indication of the significance of our results.

S-II. Dataset Examples

As mentioned in the main paper, the Occluded-Vehicles dataset, which is based on the Pascal3D+ dataset, has not only four different occlusion levels, but also four different kinds of occlusion, examples of which can be found in Figure S1. The boxes for white box occlusion, texture occlusion and noise occlusion are placed on random locations, however they are consistent over the three mentioned kinds of occlusion. The fourth kind is occlusion by segmented objects whose classes are not contained in the classes to categorize. We can see that, in contrast to the three other occlusion types, the object occlusion comprises objects in strongly differing shapes instead of rectangles. This resembles realistic occlusion scenarios in the wild more closely and gives an indication of the performance of the method in real applications.

In Figure S2 we can see several real occlusion scenarios from the images of the Occluded-COCO-Vehicles dataset. While at level L0 the object is completely visible, with rising occlusion levels the view is being blocked by poles, humans or even bigger objects to the point at which the object in question is only barely visible anymore.

S-III. Experimental Results with Standard Deviation

To provide an indication of the significance of our results, we provide the experimental results from the main paper with inclusion of the standard deviation for each value in Table S-I and Table S-II. On the Occluded-Vehicles dataset, we note that for the experimental setups using A_{GAP} and A_{FC} with BCE and IC in, respectively, 10 and 11 of the 14 different occlusion setups, the state-of-the-art results of CompNet-Multi lie outside the 2σ range our results. In general, we can also see that, regarding the mean over all occlusion setups, the results of CompNet-Multi consistently lie outside the 2σ range of our top results on both datasets. This suggests a significant improvement of our methods over CompNet-Multi of Kortylewski *et al.*[1].

S-IV. Ablation Study with Cutout

To show that, while our Inverted Cutout method is similar to Cutout, it nevertheless yields significantly different results, in this part we present an ablation study with the standard Cutout augmentation [5] in the different setups. All hyperparameters used here are the same as in the main paper and the sampled cutouts also have sizes ranging from 16×16 pixels to 128×128 pixels. The results of the experiments on the Occluded-Images dataset are shown in Table S-III and the respective results for the Occluded-COCO-Vehicles dataset are shown in Table S-IV. We can also see here that, the results with Binary Cross Entropy loss are better than the typical classification Categorical Cross Entropy loss. However, in comparison with the results from our method in Table S-I and Table S-II, using the Cutout augmentation yields significantly worse results than Inverted Cutout. This is especially notable in the higher occlusion levels can be up to around 15% on both datasets when comparing the same base setup, showing a significant advantage of our method. As mentioned in the main paper, this difference is likely caused by the correlations of object parts learned when using Cutout. While the latter only removes a comparably small part from the image, it leaves most of the image intact, resulting in the network learning from co-occurring objects parts. This makes the network unable to identify the object correctly when several of the usually co-occurring object parts are missing due to occlusion. In contrast, with Inverted Cutout the network learns to identify the object from only the small unoccluded image patches, making the network more robust against partial occlusion.

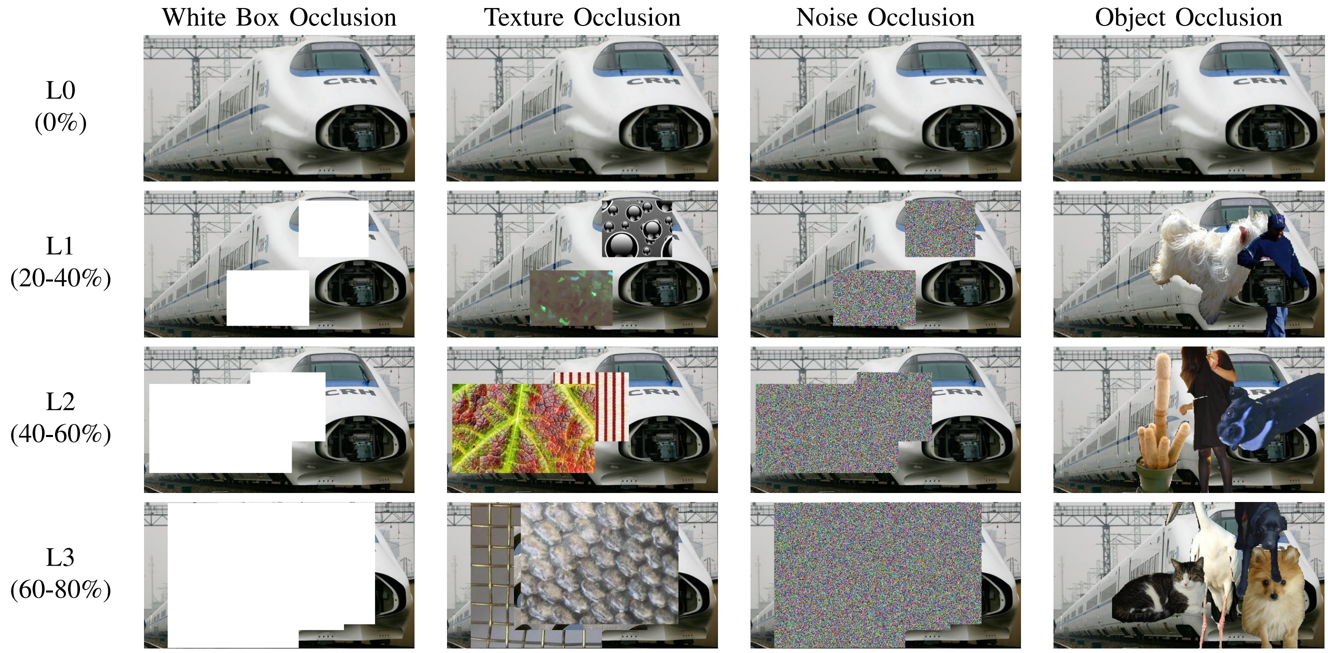

Fig. S1: Example images from the Occluded-Images Pascal3D+ dataset with several different levels and kinds of occlusion.

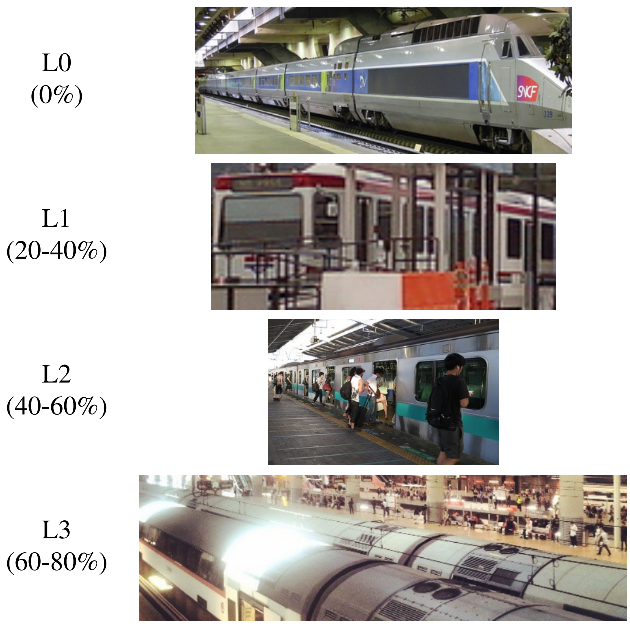

Fig. S2: Example images from the Occluded-COCO-Images dataset with several levels of occlusion.

S-V. Inverted Cutout Mask Size Ablation

To quantify the effect of the mask size when utilizing Inverted Cutout, we also present an ablation study using different mask sizes on both datasets. Similar to the previous ablation, the hyperparameters used are the same as in the main paper. The sizes ablated include 16×16, 32×32, 64×64 and 128×128 pixels. The results for the Occluded-Vehicles dataset are presented in Table S-V and the ones for Occluded-COCO-Vehicles in Table S-VI. Generally, we can see that the larger cutout sizes usually outperform the smaller ones on both datasets. However, we also note that none of the single cutout sizes outperforms the randomly sized cutouts used in the main paper and in Table S-I and Table S-II. While on the Occluded-Vehicles dataset the difference in average occlusion accuracy in comparison to the top results from the main paper for some setups is rather small, the gaps are bigger on the Occluded-COCO-Vehicles dataset. The reason for this is likely that the more regular artificial occlusions in the Occluded-Vehicles are easier to classify despite occlusion in contrast to the much more irregular natural occlusions contained in Occluded-COCO-Vehicles, which, hence, require more sophisticated training to become more robust against more complex occlusions.

In summary, bigger cutouts were found to be more effective, however, the combination of different cutout sizes by randomly sampling the size during the training still outperforms using constant sizes.

S-VI. Feature Extraction Layer Ablation

While in the main paper we used, similar to the top results from [1], the combined features from the two last convolutional blocks of VGG-16, there is also the option to use only the output of each of these blocks without the respective other. In this section, we will discuss this possibility. Experimental results for this setup are shown in Table S-VII and Table S-VIII for the trials on the Occluded-Vehicles and Occluded-COCO-Vehicles dataset, respectively, and the hyperparameters are the same as before.

In both tables, one can see that, when using the layers separately, the network also performs comparably well against

TABLE S-I: The accuracies (in %) and their respective standard deviations (denoted by $\pm$) of our method in comparison with previously introduced methods on the Pascal3D+ Occluded-Vehicles dataset. The values of methods used for comparison have been taken from [1]. * marks the results received after fine-tuning the whole network. The occlusion types are: w - white box occlusion, n - noise box occlusion, t - texture occlusion, o - occlusion by segmented objects. CCE and BCE denote the Categorical (Softmax) and Binary Cross Entropy loss, respectively. A_{GAP} and A_{FC} denote the aggregation module with global average pooling and the one utilizing a convolutional layer, respectively.

Occ. Area	**L0: 0%**	**L1: 20-40%**				**L2: 40-60%**				**L3: 60-80%**				Mean
Occ. Type	-	w	n	t	o	w	n	t	o	w	n	t	o	
VGG [2]	99.2	96.9	97.0	96.5	93.8	92.0	90.3	89.9	79.6	67.9	62.1	59.5	62.2	83.6
CoD [3]	92.1	92.7	92.3	91.7	92.3	87.4	89.5	88.7	90.6	70.2	80.3	76.9	87.1	87.1
VGG+CoD [3]	98.3	96.8	95.9	96.2	94.4	91.2	91.8	91.3	91.4	71.6	80.7	77.3	87.2	89.5
TDAPNet [4]	99.3	98.4	98.6	98.5	97.4	96.1	97.5	96.6	91.6	82.1	88.1	82.7	79.8	92.8
CompNet-Multi [1]	99.3	98.6	98.6	98.8	97.9	98.4	98.4	97.8	94.6	91.7	90.7	**86.7**	88.4	95.4
A_{GAP}+CCE	99.5 ± 0.0	97.5 ± 0.5	97.4 ± 0.2	97.1 ± 0.4	92.9 ± 1.2	90.3 ± 3.0	88.2 ± 1.7	88.3 ± 1.9	68.7 ± 4.1	58.2 ± 7.4	47.2 ± 4.3	46.6 ± 3.9	47.1 ± 4.5	78.4 ± 2.0
A_{GAP}+CCE*	99.5 ± 0.1	97.6 ± 0.5	97.5 ± 0.2	97.3 ± 0.3	93.1 ± 1.0	90.5 ± 2.8	88.9 ± 1.5	88.9 ± 1.7	69.2 ± 4.0	59.3 ± 7.6	48.5 ± 4.3	47.3 ± 3.9	47.4 ± 4.4	78.9 ± 2.0
A_{FC}+CCE	99.7 ± 0.0	97.9 ± 0.3	98.0 ± 0.4	98.0 ± 0.4	95.2 ± 0.4	93.2 ± 1.2	92.7 ± 1.1	91.4 ± 0.9	80.4 ± 1.7	61.8 ± 2.7	56.2 ± 2.8	51.9 ± 1.8	60.8 ± 2.8	82.9 ± 1.1
A_{FC}+CCE*	99.7 ± 0.1	98.1 ± 0.3	98.2 ± 0.3	98.1 ± 0.4	95.3 ± 0.5	93.3 ± 1.0	93.1 ± 1.2	91.6 ± 0.8	80.4 ± 2.0	62.0 ± 2.4	56.8 ± 2.9	52.4 ± 1.4	61.1 ± 2.9	83.1 ± 1.0
A_{GAP}+BCE	98.5 ± 0.3	93.7 ± 0.9	92.7 ± 1.7	92.6 ± 1.9	84.9 ± 4.6	83.2 ± 2.2	80.4 ± 3.2	79.0 ± 4.4	59.9 ± 7.6	49.3 ± 5.0	43.4 ± 6.2	41.2 ± 4.5	41.8 ± 7.1	72.4 ± 3.3
A_{GAP}+BCE*	99.2 ± 0.3	96.8 ± 0.9	97.1 ± 0.8	97.0 ± 0.9	94.5 ± 0.6	90.0 ± 1.6	91.0 ± 2.4	90.6 ± 1.9	83.2 ± 3.5	62.9 ± 3.0	64.6 ± 6.9	61.4 ± 6.6	68.4 ± 5.6	84.3 ± 2.3
A_{FC}+BCE	**99.8** ± 0.1	98.7 ± 0.1	98.5 ± 0.1	98.7 ± 0.2	97.0 ± 0.3	95.4 ± 0.3	94.7 ± 0.5	94.1 ± 0.5	83.6 ± 2.3	66.4 ± 0.7	60.9 ± 1.9	59.5 ± 1.3	61.4 ± 2.9	85.3 ± 0.6
A_{FC}+BCE*	99.7 ± 0.1	98.7 ± 0.1	98.6 ± 0.1	98.7 ± 0.2	97.0 ± 0.4	95.3 ± 0.2	94.8 ± 0.4	94.1 ± 0.5	83.8 ± 2.3	66.3 ± 0.6	61.1 ± 2.0	59.3 ± 1.6	61.8 ± 3.2	85.3 ± 0.6
A_{GAP}+CCE+IC	99.5 ± 0.1	99.2 ± 0.1	99.1 ± 0.1	98.9 ± 0.3	96.8 ± 0.5	99.0 ± 0.1	97.7 ± 0.2	96.3 ± 0.4	90.8 ± 2.5	95.2 ± 0.3	86.6 ± 1.8	74.4 ± 2.7	78.0 ± 5.4	93.2 ± 0.9
A_{GAP}+CCE+IC*	99.7 ± 0.0	99.4 ± 0.1	99.3 ± 0.1	99.0 ± 0.0	97.8 ± 0.2	99.2 ± 0.2	98.3 ± 0.2	96.9 ± 0.4	92.8 ± 1.4	96.4 ± 0.3	87.0 ± 1.9	75.4 ± 2.3	82.3 ± 3.4	94.1 ± 0.6
A_{FC}+CCE+IC	99.5 ± 0.0	99.5 ± 0.0	99.2 ± 0.1	99.3 ± 0.1	97.9 ± 0.1	99.2 ± 0.1	98.2 ± 0.1	97.5 ± 0.2	93.7 ± 0.5	96.8 ± 0.4	87.8 ± 1.9	78.9 ± 1.6	85.1 ± 0.7	94.8 ± 0.3
A_{FC}+CCE+IC*	99.6 ± 0.1	99.6 ± 0.1	99.2 ± 0.2	**99.4** ± 0.1	98.6 ± 0.1	99.3 ± 0.2	97.9 ± 0.5	97.5 ± 0.2	95.7 ± 0.6	95.9 ± 0.8	83.0 ± 4.8	80.4 ± 1.3	88.9 ± 1.1	95.0 ± 0.6
A_{GAP}+BCE+IC	99.1 ± 0.1	98.9 ± 0.1	98.3 ± 0.1	98.3 ± 0.1	97.4 ± 0.2	98.0 ± 0.3	96.4 ± 0.3	96.7 ± 0.1	93.1 ± 0.6	92.8 ± 0.6	83.8 ± 1.7	80.5 ± 1.4	84.8 ± 1.0	93.7 ± 0.3
A_{GAP}+BCE+IC*	99.7 ± 0.1	99.6 ± 0.1	**99.4** ± 0.1	**99.4** ± 0.1	**98.9** ± 0.1	99.2 ± 0.1	98.4 ± 0.2	98.1 ± 0.3	**96.9** ± 0.2	95.8 ± 0.3	87.4 ± 2.4	85.4 ± 1.1	**92.1** ± 1.0	96.2 ± 0.2
A_{FC}+BCE+IC	99.6 ± 0.1	99.6 ± 0.1	99.3 ± 0.1	99.3 ± 0.1	98.7 ± 0.0	99.4 ± 0.1	98.3 ± 0.1	98.1 ± 0.2	95.4 ± 0.6	97.0 ± 0.1	90.5 ± 1.3	82.8 ± 1.2	87.4 ± 1.0	95.8 ± 0.3
A_{FC}+BCE+IC*	99.7 ± 0.1	**99.7** ± 0.1	**99.4** ± 0.1	**99.4** ± 0.1	**98.9** ± 0.1	**99.6** ± 0.1	**98.8** ± 0.2	**98.5** ± 0.2	96.2 ± 0.3	**97.9** ± 0.1	**90.8** ± 1.6	83.9 ± 1.0	89.4 ± 1.0	**96.3** ± 0.2

occlusion. We can also see that using the P5 layer outperforms the P4 layer in most instances, which is expected due to the more sophisticated features generated by the later network layers. While the overall performance is similar to the combined one, we can, however, see some differences. Most notably, the accuracy of the networks with separate blocks is often much worse when training only the last layer instead of fine-tuning the complete network. Hence, the combination of both produces much more valuable features that can be used when training the complete network is not desired. When fine-tuning the complete network, however, the performance using the separate layers comes closer to the combined approach, with the P5 layer even surpassing the latter in one instance. Nevertheless, the combined approach overall outperforms the two separate layers, which was also found in the investigations by Kortylewski *et al.*in [1].

S-VII. ResNet50-Backbone Ablation

While in the main paper we utilized only the VGG-16 network architecture for the comparability to previous works, in this section we will also take a look into the application of Inverted Cutout in conjunction with another popular backbone architecture: ResNet50 [6]. During the training process with the ResNets, we utilized slightly different hyperparameters, which are shown in Table S-IX. Hyperparameters not mentioned are the same as in the main paper.

The experimental results for this ablation study are shown in Table S-X and Table S-XI for the Occluded-Vehicles and Occluded-COCO-Vehicles datasets, respectively.

Generally, it is visible that for ResNet50, Inverted Cutout is also strongly beneficial, especially when utilizing the A_{FC} aggregation module in comparison to the setups without IC. Moreover, we notice that, when training the model without IC,

TABLE S-II: Accuracies (in %) and their respective standard deviations (denoted by $\pm$) of our experiments on the Occluded-COCO-Vehicles dataset. We compare our approaches with previously introduced methods based on classification accuracy. The values of methods used for comparison have been taken from [1]. * marks the results received after fine-tuning the network. CCE and BCE denote the Categorical (Softmax) and Binary Cross Entropy losses, respectively. A_{GAP} and A_{FC} denote the aggregation module with global average pooling and with a convolutional layer, respectively.

Train Data	MS-COCO				
Occ. Area	**L0**	**L1**	**L2**	**L3**	**Avg**
VGG [2]	99.1	88.7	78.8	63.0	82.4
VGG [2] + Cutout	99.3	90.9	87.5	75.3	88.3
TDAPNet [4]	99.4	88.8	87.9	69.9	86.5
CompNet-Multi [1]	99.4	95.3	90.9	86.3	93.0
A_{GAP}+CCE	99.2 $\pm$ 0.2	85.3 $\pm$ 2.0	80.0 $\pm$ 1.6	68.8 $\pm$ 4.0	83.3 $\pm$ 1.8
A_{GAP}+CCE*	99.3 $\pm$ 0.1	85.3 $\pm$ 1.6	80.2 $\pm$ 1.2	68.8 $\pm$ 4.0	83.4 $\pm$ 1.7
A_{FC}+CCE	99.6 $\pm$ 0.0	89.9 $\pm$ 0.5	83.7 $\pm$ 0.7	70.2 $\pm$ 2.0	85.9 $\pm$ 0.4
A_{FC}+CCE*	99.6 $\pm$ 0.0	90.2 $\pm$ 0.5	84.2 $\pm$ 0.9	72.3 $\pm$ 2.6	86.6 $\pm$ 0.7
A_{GAP}+BCE	99.3 $\pm$ 0.0	86.3 $\pm$ 1.9	78.3 $\pm$ 8.5	69.5 $\pm$ 11.5	83.4 $\pm$ 5.4
A_{GAP}+BCE*	99.3 $\pm$ 0.0	86.4 $\pm$ 1.6	78.3 $\pm$ 8.8	69.5 $\pm$ 11.6	83.4 $\pm$ 5.4
A_{FC}+BCE	**99.7** $\pm$ 0.1	90.6 $\pm$ 0.5	84.9 $\pm$ 0.7	74.0 $\pm$ 2.2	87.3 $\pm$ 0.7
A_{FC}+BCE*	**99.7** $\pm$ 0.1	90.7 $\pm$ 0.5	85.5 $\pm$ 0.9	74.3 $\pm$ 1.8	87.5 $\pm$ 0.6
A_{GAP}+CCE+IC	99.1 $\pm$ 0.1	90.6 $\pm$ 0.5	87.2 $\pm$ 1.1	87.0 $\pm$ 1.5	91.0 $\pm$ 0.6
A_{GAP}+CCE+IC*	99.5 $\pm$ 0.1	94.5 $\pm$ 0.2	90.8 $\pm$ 1.3	89.4 $\pm$ 3.0	93.5 $\pm$ 1.0
A_{FC}+CCE+IC	99.4 $\pm$ 0.1	94.2 $\pm$ 0.8	90.0 $\pm$ 0.8	82.9 $\pm$ 3.0	91.6 $\pm$ 1.0
A_{FC}+CCE+IC*	99.4 $\pm$ 0.1	94.7 $\pm$ 0.6	90.8 $\pm$ 1.0	84.2 $\pm$ 0.7	92.3 $\pm$ 0.4
A_{GAP}+BCE+IC	99.1 $\pm$ 0.1	89.4 $\pm$ 0.7	88.0 $\pm$ 0.5	86.0 $\pm$ 2.4	90.6 $\pm$ 0.8
A_{GAP}+BCE+IC*	99.4 $\pm$ 0.0	94.6 $\pm$ 0.8	91.6 $\pm$ 0.5	**92.8** $\pm$ 1.8	**94.6** $\pm$ 0.4
A_{FC}+BCE+IC	99.4 $\pm$ 0.1	94.1 $\pm$ 0.6	92.0 $\pm$ 0.4	86.3 $\pm$ 1.7	93.0 $\pm$ 0.3
A_{FC}+BCE+IC*	99.5 $\pm$ 0.0	**95.4** $\pm$ 0.4	**92.1** $\pm$ 0.4	88.7 $\pm$ 2.4	93.9 $\pm$ 0.7

in general training only the aggregation layer often yields better results than fine-tuning the complete network. This effect can likely be attributed to the batch normalization layers [7] contained in the ResNet50 architecture. Batch normalization appears to make the network unable to deal with occlusion if it has not seen any occlusion examples during training. This strong detrimental effect, however, is mitigated when using IC, where the results surpass the ones after training the aggregation layer alone.

When comparing the results on both datasets, we also notice that the network trained without IC performs better on the more natural occlusions seen in the Occluded-COCO-Vehicles dataset in contrast to the artificial occlusions from the Pascal3D+-based Occluded-Vehicles dataset. The reason for this is likely that in the former bigger connected object parts are visible in contrast to more randomly spread small object parts like in the images from the Occluded-Vehicles dataset. Such bigger object parts might make the class easier to identify under occlusion for the ResNet.

Nevertheless, on both datasets, the usage of Inverted Cutout improves the results in comparison to the ones generated after training without IC. Therefore, our novel augmentation method is also beneficial for network architecture beyond VGG-16, yielding results comparable with the latter.

REFERENCES

[1] A. Kortylewski, J. He, Q. Liu, and A. L. Yuille, "Compositional convolutional neural networks: A deep architecture with innate robustness to partial occlusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8940–8949.

[2] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[3] A. Kortylewski, Q. Liu, H. Wang, Z. Zhang, and A. Yuille, "Combining compositional models and deep networks for robust object classification under occlusion," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 1333–1341.

[4] M. Xiao, A. Kortylewski, R. Wu, S. Qiao, W. Shen, and A. Yuille, "Tdapnet: Prototype network with recurrent top-down attention for robust object classification under partial occlusion," *arXiv preprint arXiv:1909.03879*, 2019.

[5] T. DeVries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," *arXiv preprint arXiv:1708.04552*, 2017.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[7] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. PMLR, 2015, pp. 448–456.

TABLE S-III: **Cutout Ablations** - The accuracies (in %) and their respective standard deviations (denoted by ±) of the Cutout ablations on the Pascal3D+ Occluded-Vehicles dataset. * marks the results received after fine-tuning the whole network. The occlusion types are: w - white box occlusion, n - noise box occlusion, t - texture occlusion, o - occlusion by segmented objects. CCE and BCE denote the Categorical (Softmax) and Binary Cross Entropy loss, respectively. A_{GAP} and A_{FC} denote the aggregation module with global average pooling and the one utilizing a convolutional layer, respectively.

Occ. Area	**L0: 0%**	**L1: 20-40%**				**L2: 40-60%**				**L3: 60-80%**				Mean
Occ. Type	-	w	n	t	o	w	n	t	o	w	n	t	o	
A_{GAP}+CCE+Cutout	99.7 ± 0.1	99.2 ± 0.1	99.1 ± 0.0	98.8 ± 0.1	95.8 ± 0.4	98.5 ± 0.2	97.3 ± 0.6	95.3 ± 0.5	79.9 ± 1.4	87.2 ± 2.6	75.5 ± 4.1	65.1 ± 2.6	57.0 ± 2.6	88.3 ± 0.5
A_{GAP}+CCE+Cutout*	99.7 ± 0.1	99.5 ± 0.0	99.3 ± 0.1	99.0 ± 0.1	96.8 ± 0.2	98.8 ± 0.1	97.7 ± 0.2	96.0 ± 0.4	86.1 ± 1.5	92.1 ± 0.9	82.5 ± 2.6	70.7 ± 1.7	67.2 ± 1.5	91.2 ± 0.4
A_{FC}+CCE+Cutout	99.7 ± 0.0	99.5 ± 0.1	99.3 ± 0.1	99.2 ± 0.2	97.4 ± 0.2	98.7 ± 0.3	97.8 ± 0.6	96.6 ± 0.7	85.7 ± 2.1	90.7 ± 1.0	77.6 ± 3.9	71.7 ± 2.4	68.1 ± 1.8	90.9 ± 0.7
A_{FC}+CCE+Cutout*	99.5 ± 0.0	98.9 ± 0.1	97.7 ± 1.0	98.5 ± 0.2	97.7 ± 0.4	97.7 ± 0.5	90.9 ± 6.2	95.9 ± 0.4	92.6 ± 0.7	85.6 ± 2.6	54.2 ± 22.1	75.6 ± 2.6	82.7 ± 1.9	89.8 ± 2.3
A_{GAP}+BCE+Cutout	99.2 ± 0.1	97.5 ± 0.3	97.1 ± 0.4	96.6 ± 0.3	92.2 ± 1.5	94.8 ± 0.9	92.5 ± 0.7	90.5 ± 0.9	76.6 ± 4.8	73.4 ± 4.4	66.3 ± 3.2	59.8 ± 3.3	58.8 ± 6.4	84.2 ± 1.6
A_{GAP}+BCE+Cutout*	99.7 ± 0.1	99.2 ± 0.2	99.0 ± 0.3	98.9 ± 0.1	98.0 ± 0.4	98.2 ± 0.2	97.0 ± 0.3	96.4 ± 0.3	91.5 ± 1.3	88.2 ± 1.8	79.5 ± 3.5	77.6 ± 1.7	80.3 ± 2.7	92.6 ± 0.6
A_{FC}+BCE+Cutout	99.8 ± 0.0	99.5 ± 0.1	99.5 ± 0.1	99.5 ± 0.0	98.5 ± 0.2	99.2 ± 0.1	98.4 ± 0.3	97.7 ± 0.2	90.3 ± 0.9	93.1 ± 0.6	82.5 ± 2.8	75.6 ± 1.8	73.0 ± 1.5	92.8 ± 0.3
A_{FC}+BCE+Cutout*	99.8 ± 0.1	99.6 ± 0.1	99.5 ± 0.1	99.5 ± 0.1	98.6 ± 0.2	99.4 ± 0.1	98.4 ± 0.5	97.9 ± 0.3	91.1 ± 1.0	93.6 ± 0.8	82.6 ± 3.7	76.3 ± 2.1	75.0 ± 2.1	93.2 ± 0.4

TABLE S-IV: **Cutout Ablations** Accuracies (in %) and their respective standard deviations (denoted by ±) of our Cutout ablation study on the Occluded-COCO-Vehicles dataset. * marks the results received after fine-tuning the network. CCE and BCE denote the Categorical (Softmax) and Binary Cross Entropy losses, respectively. A_{GAP} and A_{FC} denote the aggregation module with global average pooling and with a convolutional layer, respectively.

Train Data	**MS-COCO**				
Occ. Area	**L0**	**L1**	**L2**	**L3**	**Avg**
A_{GAP}+CCE+Cutout	98.7 ± 0.4	86.1 ± 0.9	80.9 ± 2.0	71.2 ± 5.1	84.2 ± 1.7
A_{GAP}+CCE+Cutout*	99.0 ± 0.2	88.2 ± 2.9	82.3 ± 5.7	66.8 ± 4.3	84.1 ± 3.2
A_{FC}+CCE+Cutout	99.6 ± 0.1	91.6 ± 0.6	84.4 ± 3.2	71.2 ± 3.1	86.7 ± 1.6
A_{FC}+CCE+Cutout*	99.2 ± 0.1	90.4 ± 0.6	79.5 ± 2.4	63.7 ± 5.9	83.2 ± 1.8
A_{GAP}+BCE+Cutout	99.2 ± 0.1	87.9 ± 0.7	83.1 ± 1.5	76.7 ± 1.4	86.7 ± 0.8
A_{GAP}+BCE+Cutout*	99.2 ± 0.2	91.2 ± 1.0	87.2 ± 2.2	75.3 ± 4.8	88.2 ± 1.9
A_{FC}+BCE+Cutout	99.6 ± 0.1	92.1 ± 0.8	86.3 ± 1.2	75.0 ± 2.2	88.2 ± 0.8
A_{FC}+BCE+Cutout*	99.6 ± 0.1	94.1 ± 0.8	87.6 ± 1.2	73.3 ± 3.1	88.6 ± 1.0

TABLE S-V: **Inverted Cutout Mask Size Ablation** The accuracies (in %) and their respective standard deviations (denoted by ±) of the cutout size ablation study of our Inverted Cutout method. * marks the results received after fine-tuning the whole network. The occlusion types are: w - white box occlusion, n - noise box occlusion, t - texture occlusion, o - occlusion by segmented objects. CCE and BCE denote the Categorical (Softmax) and Binary Cross Entropy loss, respectively. A_{GAP} and A_{FC} denote the aggregation module with global average pooling and the one utilizing a convolutional layer, respectively.

Occ. Area		**L0: 0%**	**L1: 20-40%**				**L2: 40-60%**				**L3: 60-80%**				Mean
Occ. Type		-	w	n	t	o	w	n	t	o	w	n	t	o	
A_{GAP}+CCE+IC	Mask size 16	56.4 ± 4.1	55.1 ± 3.1	53.4 ± 3.1	53.8 ± 2.9	52.6 ± 2.1	52.7 ± 3.0	51.7 ± 2.1	51.6 ± 1.8	49.8 ± 2.3	46.4 ± 2.2	45.7 ± 3.0	43.7 ± 2.1	48.0 ± 4.1	50.8 ± 1.4
A_{GAP}+CCE+IC*	Mask size 16	62.6 ± 7.2	58.9 ± 8.0	59.8 ± 5.6	59.8 ± 6.3	48.0 ± 6.4	54.5 ± 8.2	54.6 ± 6.2	54.8 ± 4.8	41.5 ± 6.3	49.1 ± 7.9	41.2 ± 10.8	42.3 ± 2.3	35.0 ± 6.2	50.9 ± 5.7
A_{FC}+CCE+IC	Mask size 16	48.5 ± 1.2	50.9 ± 2.0	48.2 ± 1.0	48.4 ± 1.1	47.3 ± 0.3	52.1 ± 2.3	48.4 ± 1.0	48.4 ± 0.9	47.2 ± 0.2	55.8 ± 2.8	47.1 ± 0.6	46.9 ± 0.7	47.1 ± 0.1	48.9 ± 1.1
A_{FC}+CCE+IC*	Mask size 16	89.7 ± 1.3	88.3 ± 0.9	77.0 ± 2.4	85.5 ± 1.5	80.5 ± 2.1	86.6 ± 1.1	56.9 ± 3.8	80.8 ± 1.6	76.6 ± 2.3	80.8 ± 0.8	12.6 ± 3.8	66.0 ± 2.1	72.2 ± 2.5	73.3 ± 1.1
A_{GAP}+BCE+IC	Mask size 16	49.4 ± 9.9	50.8 ± 9.1	49.4 ± 8.2	47.2 ± 11.2	42.0 ± 11.9	49.3 ± 7.3	48.1 ± 8.3	44.2 ± 11.1	37.8 ± 10.6	41.7 ± 6.3	42.2 ± 9.1	34.9 ± 10.7	35.6 ± 9.9	44.1 ± 9.1
A_{GAP}+BCE+IC*	Mask size 16	69.4 ± 1.0	69.8 ± 2.7	65.3 ± 4.0	68.1 ± 1.5	61.9 ± 3.4	68.3 ± 3.7	55.4 ± 10.9	63.3 ± 1.2	58.1 ± 4.1	60.7 ± 6.5	32.3 ± 15.3	45.3 ± 3.8	52.0 ± 6.0	59.2 ± 1.8
A_{FC}+BCE+IC	Mask size 16	55.2 ± 3.0	59.0 ± 3.0	54.5 ± 2.7	54.6 ± 2.6	50.4 ± 1.5	61.3 ± 2.8	54.7 ± 2.0	54.3 ± 1.9	49.6 ± 1.1	65.2 ± 2.6	54.1 ± 1.7	51.6 ± 1.5	49.2 ± 1.2	54.9 ± 2.1
A_{FC}+BCE+IC*	Mask size 16	73.7 ± 4.6	75.5 ± 4.2	70.4 ± 4.0	71.6 ± 4.4	67.5 ± 5.2	75.6 ± 4.0	67.3 ± 3.7	68.6 ± 3.9	64.4 ± 4.9	73.5 ± 3.5	54.6 ± 1.7	59.5 ± 2.8	62.8 ± 4.5	68.1 ± 3.8
A_{GAP}+CCE+IC	Mask size 32	77.0 ± 2.3	80.0 ± 2.1	74.2 ± 1.6	73.9 ± 1.9	69.4 ± 2.4	80.3 ± 1.9	73.5 ± 1.8	69.8 ± 1.8	62.3 ± 2.4	77.5 ± 2.2	66.9 ± 1.1	56.2 ± 1.1	53.1 ± 3.2	70.3 ± 1.4
A_{GAP}+CCE+IC*	Mask size 32	90.9 ± 1.0	92.9 ± 0.7	88.7 ± 1.1	89.4 ± 1.1	81.8 ± 1.6	93.1 ± 0.9	86.5 ± 1.4	85.0 ± 1.3	73.8 ± 2.1	90.2 ± 0.8	75.8 ± 2.4	67.9 ± 1.8	63.5 ± 2.5	83.1 ± 1.1
A_{FC}+CCE+IC	Mask size 32	88.3 ± 1.3	88.4 ± 0.9	84.0 ± 1.1	84.5 ± 0.9	82.3 ± 1.0	87.7 ± 1.2	82.0 ± 1.2	80.4 ± 0.9	78.9 ± 0.9	85.1 ± 0.6	73.6 ± 1.1	69.1 ± 1.2	74.8 ± 1.5	81.5 ± 0.9
A_{FC}+CCE+IC*	Mask size 32	97.9 ± 0.1	97.9 ± 0.1	96.4 ± 0.2	97.2 ± 0.1	95.2 ± 0.4	97.1 ± 0.2	91.6 ± 2.7	94.0 ± 0.3	91.8 ± 0.3	94.0 ± 0.5	59.0 ± 12.8	78.7 ± 1.0	85.6 ± 1.1	90.5 ± 1.4
A_{GAP}+BCE+IC	Mask size 32	75.4 ± 3.3	77.7 ± 3.0	72.3 ± 2.2	74.3 ± 2.4	72.6 ± 1.7	78.6 ± 3.4	71.3 ± 1.5	72.8 ± 1.6	70.1 ± 3.3	74.4 ± 6.3	65.4 ± 1.6	63.9 ± 2.2	64.5 ± 4.6	71.8 ± 2.7
A_{GAP}+BCE+IC*	Mask size 32	96.4 ± 0.6	96.3 ± 0.6	95.5 ± 0.8	95.8 ± 0.6	93.1 ± 0.8	95.9 ± 0.7	92.4 ± 2.1	93.6 ± 0.7	90.3 ± 0.9	91.7 ± 1.9	67.9 ± 10.3	80.4 ± 0.7	85.1 ± 1.5	90.3 ± 1.4
A_{FC}+BCE+IC	Mask size 32	91.2 ± 0.7	90.7 ± 0.9	87.9 ± 1.1	88.4 ± 1.1	88.5 ± 1.1	90.1 ± 0.7	86.6 ± 0.8	85.9 ± 1.3	85.4 ± 1.1	87.8 ± 1.3	79.6 ± 1.0	75.5 ± 1.3	81.2 ± 1.3	86.1 ± 1.0
A_{FC}+BCE+IC*	Mask size 32	97.0 ± 0.4	97.0 ± 0.6	95.6 ± 1.1	96.0 ± 0.8	94.1 ± 1.2	96.4 ± 0.7	93.7 ± 1.3	93.5 ± 1.4	90.1 ± 1.4	93.8 ± 1.0	84.2 ± 1.9	82.9 ± 1.3	85.7 ± 1.3	92.3 ± 1.0
A_{GAP}+CCE+IC	Mask size 64	98.3 ± 0.1	98.3 ± 0.1	97.4 ± 0.1	97.4 ± 0.2	94.9 ± 0.5	98.0 ± 0.1	96.0 ± 0.2	94.6 ± 0.2	89.6 ± 1.5	95.4 ± 0.2	87.9 ± 0.7	77.5 ± 1.2	80.4 ± 2.8	92.7 ± 0.4
A_{GAP}+CCE+IC*	Mask size 64	98.9 ± 0.1	99.1 ± 0.1	98.8 ± 0.1	98.5 ± 0.2	97.1 ± 0.3	98.8 ± 0.2	97.7 ± 0.2	96.6 ± 0.5	92.7 ± 1.2	97.2 ± 0.2	89.4 ± 2.2	80.8 ± 0.9	84.5 ± 3.1	94.6 ± 0.5
A_{FC}+CCE+IC	Mask size 64	99.1 ± 0.1	99.1 ± 0.1	98.6 ± 0.1	98.5 ± 0.2	96.8 ± 0.1	98.5 ± 0.1	97.4 ± 0.1	96.2 ± 0.3	91.2 ± 0.3	95.6 ± 0.4	87.3 ± 1.1	80.9 ± 0.6	83.6 ± 0.8	94.1 ± 0.1
A_{FC}+CCE+IC*	Mask size 64	99.5 ± 0.1	99.4 ± 0.1	98.7 ± 0.4	99.1 ± 0.1	98.4 ± 0.1	99.3 ± 0.1	95.9 ± 2.4	97.6 ± 0.2	96.4 ± 0.4	96.3 ± 0.2	68.9 ± 13.2	80.8 ± 0.9	90.9 ± 0.9	93.9 ± 1.4
A_{GAP}+BCE+IC	Mask size 64	97.8 ± 0.4	97.7 ± 0.2	96.7 ± 0.3	97.1 ± 0.4	94.6 ± 0.7	96.9 ± 0.3	94.9 ± 0.5	95.4 ± 0.5	89.9 ± 1.2	93.2 ± 0.5	83.2 ± 3.7	81.5 ± 3.2	81.7 ± 3.5	92.4 ± 1.0
A_{GAP}+BCE+IC*	Mask size 64	99.4 ± 0.1	99.5 ± 0.0	99.0 ± 0.1	99.4 ± 0.2	98.6 ± 0.1	99.2 ± 0.0	97.7 ± 0.4	98.3 ± 0.2	96.5 ± 0.2	96.4 ± 0.4	77.7 ± 6.4	85.1 ± 2.3	91.7 ± 0.3	95.3 ± 0.6
A_{FC}+BCE+IC	Mask size 64	99.3 ± 0.1	99.3 ± 0.1	99.0 ± 0.1	98.9 ± 0.2	97.9 ± 0.3	98.9 ± 0.1	98.1 ± 0.3	97.6 ± 0.2	94.9 ± 0.4	96.3 ± 0.2	91.5 ± 0.7	85.0 ± 0.7	87.9 ± 0.7	95.7 ± 0.1
A_{FC}+BCE+IC*	Mask size 64	99.6 ± 0.1	99.7 ± 0.1	99.4 ± 0.1	99.3 ± 0.1	98.5 ± 0.2	99.5 ± 0.1	98.4 ± 0.1	98.5 ± 0.1	96.6 ± 0.4	97.6 ± 0.2	87.0 ± 2.1	84.9 ± 0.3	91.0 ± 0.6	96.2 ± 0.2
A_{GAP}+CCE+IC	Mask size 128	99.7 ± 0.0	99.4 ± 0.0	99.2 ± 0.1	99.0 ± 0.1	97.7 ± 0.2	99.2 ± 0.1	97.8 ± 0.4	96.8 ± 0.3	91.2 ± 1.1	95.9 ± 0.4	84.7 ± 3.0	74.5 ± 1.9	78.2 ± 2.8	93.3 ± 0.7
A_{GAP}+CCE+IC*	Mask size 128	99.7 ± 0.0	99.6 ± 0.1	99.4 ± 0.1	99.1 ± 0.0	97.6 ± 0.3	99.3 ± 0.1	98.0 ± 0.3	96.9 ± 0.3	91.2 ± 1.1	96.4 ± 0.3	81.8 ± 3.3	73.2 ± 1.3	79.3 ± 2.1	93.2 ± 0.5
A_{FC}+CCE+IC	Mask size 128	99.7 ± 0.0	99.5 ± 0.0	99.3 ± 0.1	99.3 ± 0.1	98.4 ± 0.1	99.4 ± 0.1	98.2 ± 0.3	97.6 ± 0.1	93.8 ± 1.0	95.3 ± 0.3	80.9 ± 2.9	75.6 ± 1.2	83.8 ± 2.5	93.9 ± 0.5
A_{FC}+CCE+IC*	Mask size 128	99.7 ± 0.1	99.5 ± 0.0	99.3 ± 0.1	99.3 ± 0.1	98.8 ± 0.2	99.1 ± 0.1	97.7 ± 0.5	97.4 ± 0.4	95.6 ± 0.4	93.1 ± 0.4	78.4 ± 6.8	77.1 ± 1.9	88.6 ± 0.6	94.1 ± 0.5
A_{GAP}+BCE+IC	Mask size 128	99.3 ± 0.0	98.9 ± 0.1	98.6 ± 0.1	98.8 ± 0.2	97.5 ± 0.4	98.1 ± 0.3	96.4 ± 0.6	96.9 ± 0.3	92.2 ± 1.3	91.2 ± 0.8	78.4 ± 2.7	77.5 ± 1.8	81.3 ± 2.9	92.7 ± 0.8
A_{GAP}+BCE+IC*	Mask size 128	99.8 ± 0.0	99.7 ± 0.0	99.5 ± 0.1	99.5 ± 0.1	99.1 ± 0.1	99.4 ± 0.1	98.6 ± 0.3	98.6 ± 0.3	96.7 ± 0.3	94.2 ± 1.5	80.7 ± 3.5	82.5 ± 1.6	89.7 ± 1.0	95.2 ± 0.3
A_{FC}+BCE+IC	Mask size 128	99.8 ± 0.0	99.6 ± 0.1	99.4 ± 0.1	99.4 ± 0.1	98.9 ± 0.1	99.5 ± 0.1	98.5 ± 0.1	98.4 ± 0.2	95.4 ± 0.4	96.2 ± 0.4	84.1 ± 0.3	79.7 ± 0.4	85.2 ± 0.7	94.9 ± 0.1
A_{FC}+BCE+IC*	Mask size 128	99.7 ± 0.1	99.6 ± 0.0	99.4 ± 0.1	99.4 ± 0.0	98.8 ± 0.1	99.6 ± 0.0	98.4 ± 0.2	98.2 ± 0.1	95.6 ± 0.6	97.0 ± 0.2	83.2 ± 1.0	78.5 ± 0.3	85.4 ± 0.9	94.8 ± 0.1

TABLE S-VI: **Inverted Cutout Mask Size Ablation** Accuracies (in %) and their respective standard deviations (denoted by $\pm$) of our cutout size ablations on the Occluded-COCO-Vehicles dataset. * marks the results received after fine-tuning the network. CCE and BCE denote the Categorical (Softmax) and Binary Cross Entropy losses, respectively. A_{GAP} and A_{FC} denote the aggregation module with global average pooling and with a convolutional layer, respectively.

Train Data		MS-COCO				
Occ. Area		**L0**	**L1**	**L2**	**L3**	**Avg**
A_{GAP}+CCE+IC	Mask size 16	43.2 ± 7.5	19.1 ± 2.9	14.8 ± 3.1	23.3 ± 9.2	25.1 ± 4.9
A_{GAP}+CCE+IC*		54.7 ± 11.0	28.9 ± 7.5	21.4 ± 7.8	20.5 ± 6.5	31.4 ± 7.8
A_{FC}+CCE+IC		72.3 ± 3.4	39.6 ± 3.6	30.0 ± 1.6	29.8 ± 2.6	42.9 ± 2.1
A_{FC}+CCE+IC*		87.9 ± 1.5	69.3 ± 2.2	61.6 ± 2.8	46.6 ± 3.1	66.3 ± 1.6
A_{GAP}+BCE+IC		50.1 ± 2.9	25.7 ± 1.8	21.4 ± 1.3	34.9 ± 2.5	33.0 ± 1.2
A_{GAP}+BCE+IC*		64.3 ± 6.1	34.5 ± 8.7	25.2 ± 8.1	21.6 ± 8.1	36.4 ± 7.6
A_{FC}+BCE+IC		75.5 ± 1.2	41.5 ± 0.9	32.1 ± 1.0	32.5 ± 2.0	45.4 ± 0.9
A_{FC}+BCE+IC*		83.8 ± 3.8	62.8 ± 3.1	53.2 ± 4.0	40.8 ± 3.4	60.1 ± 3.1
A_{GAP}+CCE+IC	Mask size 32	34.1 ± 1.3	21.3 ± 1.0	15.8 ± 1.6	16.8 ± 8.6	22.0 ± 2.6
A_{GAP}+CCE+IC*		74.6 ± 5.4	36.9 ± 7.9	25.5 ± 4.8	30.8 ± 5.2	42.0 ± 3.9
A_{FC}+CCE+IC		95.8 ± 0.5	81.6 ± 2.0	69.3 ± 1.9	52.4 ± 3.7	74.8 ± 1.8
A_{FC}+CCE+IC*		97.9 ± 0.2	88.2 ± 0.8	84.8 ± 1.2	76.0 ± 2.8	86.7 ± 1.1
A_{GAP}+BCE+IC		82.5 ± 6.0	71.4 ± 2.9	58.3 ± 4.2	59.9 ± 8.8	68.0 ± 4.9
A_{GAP}+BCE+IC*		95.0 ± 0.6	82.1 ± 0.7	77.8 ± 1.8	73.6 ± 3.0	82.1 ± 1.3
A_{FC}+BCE+IC		95.5 ± 0.5	81.0 ± 1.8	67.0 ± 1.6	50.7 ± 4.8	73.5 ± 2.0
A_{FC}+BCE+IC*		98.4 ± 0.2	88.2 ± 0.6	85.5 ± 0.7	73.3 ± 2.3	86.3 ± 0.9
A_{GAP}+CCE+IC	Mask size 64	97.7 ± 0.4	86.4 ± 1.3	83.7 ± 1.5	76.0 ± 3.0	86.0 ± 0.9
A_{GAP}+CCE+IC*		99.0 ± 0.1	91.8 ± 1.1	89.6 ± 1.1	84.2 ± 4.4	91.1 ± 1.6
A_{FC}+CCE+IC		99.1 ± 0.1	92.2 ± 0.3	88.0 ± 2.0	76.0 ± 1.5	88.8 ± 0.9
A_{FC}+CCE+IC*		99.3 ± 0.1	93.4 ± 0.7	89.4 ± 0.7	84.2 ± 2.8	91.6 ± 1.0
A_{GAP}+BCE+IC		97.9 ± 0.3	87.3 ± 1.0	85.3 ± 1.3	82.2 ± 1.0	88.2 ± 0.5
A_{GAP}+BCE+IC*		99.3 ± 0.1	92.7 ± 0.4	88.6 ± 0.4	86.3 ± 2.6	91.7 ± 0.8
A_{FC}+BCE+IC		99.1 ± 0.1	92.7 ± 0.4	90.1 ± 0.6	79.1 ± 2.2	90.3 ± 0.6
A_{FC}+BCE+IC*		99.4 ± 0.0	94.3 ± 0.6	91.0 ± 0.6	87.0 ± 2.1	92.9 ± 0.5
A_{GAP}+CCE+IC	Mask size 128	99.0 ± 0.1	90.0 ± 0.7	86.0 ± 1.0	76.0 ± 2.3	87.8 ± 0.9
A_{GAP}+CCE+IC*		99.5 ± 0.1	94.2 ± 0.4	88.4 ± 0.5	86.6 ± 2.6	92.2 ± 0.8
A_{FC}+CCE+IC		99.5 ± 0.1	93.8 ± 0.6	88.9 ± 1.3	79.1 ± 2.6	90.3 ± 0.6
A_{FC}+CCE+IC*		99.2 ± 0.1	93.8 ± 0.5	87.2 ± 0.1	76.7 ± 4.4	89.2 ± 1.3
A_{GAP}+BCE+IC		99.1 ± 0.1	90.0 ± 0.4	86.8 ± 1.2	78.8 ± 2.1	88.7 ± 0.6
A_{GAP}+BCE+IC*		99.5 ± 0.1	93.9 ± 1.0	89.6 ± 0.4	84.9 ± 3.2	92.0 ± 1.1
A_{FC}+BCE+IC		99.6 ± 0.1	94.7 ± 0.3	90.4 ± 0.9	83.2 ± 1.1	92.0 ± 0.2
A_{FC}+BCE+IC*		99.6 ± 0.1	94.8 ± 0.5	90.8 ± 1.1	87.7 ± 1.4	93.2 ± 0.6

TABLE S-VII: **Feature Extraction Layer Ablations** The accuracies (in %) and their respective standard deviations (denoted by ±) of the feature extraction layer ablation study of our Inverted Cutout method. * marks the results received after fine-tuning the whole network. The occlusion types are: w - white box occlusion, n - noise box occlusion, t - texture occlusion, o - occlusion by segmented objects. CCE and BCE denote the Categorical (Softmax) and Binary Cross Entropy loss, respectively. A_{GAP} and A_{FC} denote the aggregation module with global average pooling and the one utilizing a convolutional layer, respectively.

Occ. Area		**L0: 0%**	**L1: 20-40%**				**L2: 40-60%**				**L3: 60-80%**				Mean
Occ. Type		-	w	n	t	o	w	n	t	o	w	n	t	o	
A_{GAP}+CCE+IC	Extraction Block P4	99.4 ± 0.1	99.0 ± 0.2	98.6 ± 0.1	98.4 ± 0.1	96.5 ± 0.2	98.7 ± 0.1	97.6 ± 0.3	95.6 ± 0.2	90.3 ± 1.0	94.1 ± 0.4	85.1 ± 1.5	71.4 ± 0.7	77.6 ± 1.5	92.5 ± 0.3
A_{GAP}+CCE+IC*		99.5 ± 0.1	99.3 ± 0.1	99.1 ± 0.1	98.8 ± 0.1	97.4 ± 0.2	98.9 ± 0.1	98.3 ± 0.1	95.7 ± 0.3	92.3 ± 0.7	95.6 ± 0.2	87.0 ± 1.4	71.8 ± 1.3	81.8 ± 1.7	93.5 ± 0.2
A_{FC}+CCE+IC		99.4 ± 0.1	99.4 ± 0.1	98.9 ± 0.1	99.0 ± 0.1	97.1 ± 0.4	99.0 ± 0.0	97.5 ± 0.2	96.5 ± 0.1	91.0 ± 1.1	95.7 ± 0.2	83.7 ± 0.6	75.2 ± 0.6	80.0 ± 2.2	93.3 ± 0.3
A_{FC}+CCE+IC*		99.6 ± 0.0	99.5 ± 0.0	99.0 ± 0.2	99.4 ± 0.1	98.3 ± 0.0	99.3 ± 0.1	97.4 ± 0.2	97.0 ± 0.2	94.9 ± 0.5	95.1 ± 1.1	79.6 ± 2.5	77.2 ± 0.9	86.7 ± 1.1	94.1 ± 0.3
A_{GAP}+BCE+IC		98.7 ± 0.2	98.2 ± 0.1	98.0 ± 0.2	97.7 ± 0.3	96.2 ± 0.5	97.2 ± 0.5	95.6 ± 0.6	94.5 ± 0.5	88.9 ± 1.1	90.6 ± 1.6	84.2 ± 2.0	73.7 ± 1.2	76.8 ± 1.8	91.6 ± 0.7
A_{GAP}+BCE+IC*		99.5 ± 0.1	99.4 ± 0.1	98.9 ± 0.1	99.2 ± 0.1	98.6 ± 0.1	99.0 ± 0.1	96.7 ± 0.8	97.5 ± 0.2	96.0 ± 0.4	95.5 ± 0.5	78.0 ± 2.6	82.0 ± 1.2	89.7 ± 0.8	94.6 ± 0.2
A_{FC}+BCE+IC		99.5 ± 0.1	99.4 ± 0.1	99.2 ± 0.1	99.1 ± 0.0	98.2 ± 0.2	99.2 ± 0.1	98.2 ± 0.2	97.5 ± 0.1	93.0 ± 0.5	95.8 ± 0.3	90.4 ± 0.1	80.0 ± 0.5	82.8 ± 1.4	94.8 ± 0.2
A_{FC}+BCE+IC*		99.6 ± 0.1	99.6 ± 0.1	99.4 ± 0.1	99.3 ± 0.0	98.5 ± 0.1	99.3 ± 0.1	98.6 ± 0.2	97.6 ± 0.2	94.4 ± 0.2	97.2 ± 0.2	91.8 ± 0.8	80.8 ± 0.6	85.5 ± 0.7	95.5 ± 0.1
A_{GAP}+CCE+IC	Extraction Block P5	99.3 ± 0.1	99.1 ± 0.1	98.5 ± 0.1	98.5 ± 0.2	94.6 ± 0.5	97.9 ± 0.1	95.9 ± 0.5	95.6 ± 0.3	79.6 ± 2.4	90.6 ± 0.7	74.2 ± 3.5	70.5 ± 2.1	58.6 ± 3.7	88.7 ± 0.9
A_{GAP}+CCE+IC*		99.8 ± 0.0	99.6 ± 0.1	99.4 ± 0.1	99.3 ± 0.1	97.7 ± 0.1	99.2 ± 0.1	97.8 ± 0.5	97.9 ± 0.1	91.1 ± 1.3	95.3 ± 0.1	79.3 ± 3.4	78.8 ± 0.6	77.1 ± 3.2	93.2 ± 0.6
A_{FC}+CCE+IC		99.2 ± 0.1	99.2 ± 0.1	99.0 ± 0.2	98.8 ± 0.1	96.8 ± 0.2	98.5 ± 0.1	96.7 ± 0.6	96.4 ± 0.4	86.9 ± 2.1	92.4 ± 0.2	72.9 ± 2.9	70.8 ± 1.2	69.7 ± 3.9	90.6 ± 0.3
A_{FC}+CCE+IC*		99.6 ± 0.1	99.5 ± 0.1	99.4 ± 0.1	99.4 ± 0.1	98.7 ± 0.2	99.2 ± 0.1	98.8 ± 0.2	98.2 ± 0.1	96.3 ± 0.2	96.1 ± 0.8	88.3 ± 1.6	83.4 ± 0.3	89.6 ± 1.3	95.9 ± 0.1
A_{GAP}+BCE+IC		99.0 ± 0.1	98.7 ± 0.2	98.2 ± 0.1	98.3 ± 0.2	95.6 ± 0.5	97.6 ± 0.3	95.8 ± 0.4	95.9 ± 0.2	86.6 ± 1.6	90.1 ± 0.5	78.2 ± 2.5	75.2 ± 1.5	69.3 ± 2.9	90.7 ± 0.6
A_{GAP}+BCE+IC*		99.6 ± 0.0	99.6 ± 0.1	99.4 ± 0.0	99.4 ± 0.0	98.9 ± 0.1	99.2 ± 0.1	98.9 ± 0.2	98.5 ± 0.1	96.8 ± 0.3	96.9 ± 0.4	91.8 ± 1.6	86.2 ± 0.6	91.0 ± 0.7	96.6 ± 0.1
A_{FC}+BCE+IC		99.4 ± 0.1	99.3 ± 0.0	98.9 ± 0.1	99.1 ± 0.1	97.5 ± 0.3	98.6 ± 0.2	97.1 ± 0.2	97.1 ± 0.3	88.9 ± 1.8	93.5 ± 0.4	78.2 ± 1.6	74.2 ± 0.2	71.8 ± 3.9	91.8 ± 0.4
A_{FC}+BCE+IC*		99.7 ± 0.1	99.6 ± 0.1	99.4 ± 0.1	99.4 ± 0.1	98.9 ± 0.1	99.5 ± 0.1	98.6 ± 0.1	98.6 ± 0.1	95.0 ± 1.1	97.0 ± 0.2	86.6 ± 1.7	84.7 ± 1.0	84.3 ± 2.3	95.5 ± 0.2
A_{GAP}+CCE+IC	Extraction Block P4 & P5	99.5 ± 0.1	99.2 ± 0.1	99.1 ± 0.1	98.9 ± 0.3	96.8 ± 0.5	99.0 ± 0.1	97.7 ± 0.2	96.3 ± 0.4	90.8 ± 2.5	95.2 ± 0.3	86.6 ± 1.8	74.4 ± 2.7	78.0 ± 5.4	93.2 ± 0.9
A_{GAP}+CCE+IC*		99.7 ± 0.0	99.4 ± 0.1	99.3 ± 0.1	99.0 ± 0.0	97.8 ± 0.2	99.2 ± 0.2	98.3 ± 0.2	96.9 ± 0.4	92.8 ± 1.4	96.4 ± 0.3	87.0 ± 1.9	75.4 ± 2.3	82.3 ± 3.4	94.1 ± 0.6
A_{FC}+CCE+IC		99.5 ± 0.0	99.5 ± 0.0	99.2 ± 0.1	99.3 ± 0.1	97.9 ± 0.1	99.2 ± 0.1	98.2 ± 0.1	97.5 ± 0.2	93.7 ± 0.5	96.8 ± 0.4	87.8 ± 1.9	78.9 ± 1.6	85.1 ± 0.7	94.8 ± 0.3
A_{FC}+CCE+IC*		99.6 ± 0.1	99.6 ± 0.1	99.2 ± 0.2	99.4 ± 0.1	98.6 ± 0.1	99.3 ± 0.2	97.9 ± 0.5	97.5 ± 0.2	95.7 ± 0.6	95.9 ± 0.8	83.0 ± 4.8	80.4 ± 1.3	88.9 ± 1.1	95.0 ± 0.6
A_{GAP}+BCE+IC		99.1 ± 0.1	98.9 ± 0.1	98.3 ± 0.1	98.3 ± 0.1	97.4 ± 0.2	98.0 ± 0.3	96.4 ± 0.3	96.7 ± 0.1	93.1 ± 0.6	92.8 ± 0.6	83.8 ± 1.7	80.5 ± 1.4	84.8 ± 1.0	93.7 ± 0.3
A_{GAP}+BCE+IC*		99.7 ± 0.1	99.6 ± 0.1	99.4 ± 0.1	99.4 ± 0.1	98.9 ± 0.1	99.2 ± 0.1	98.4 ± 0.2	98.1 ± 0.3	96.9 ± 0.2	95.8 ± 0.3	87.4 ± 2.4	85.4 ± 1.1	92.1 ± 1.0	96.2 ± 0.2
A_{FC}+BCE+IC		99.6 ± 0.1	99.6 ± 0.1	99.3 ± 0.1	99.3 ± 0.1	98.7 ± 0.0	99.4 ± 0.1	98.3 ± 0.1	98.1 ± 0.2	95.4 ± 0.6	97.0 ± 0.1	90.5 ± 1.3	82.8 ± 1.2	87.4 ± 1.0	95.8 ± 0.3
A_{FC}+BCE+IC*		99.7 ± 0.1	99.7 ± 0.1	99.4 ± 0.1	99.4 ± 0.1	98.9 ± 0.1	99.6 ± 0.1	98.8 ± 0.2	98.5 ± 0.2	96.2 ± 0.3	97.9 ± 0.1	90.8 ± 1.6	83.9 ± 1.0	89.4 ± 1.0	96.3 ± 0.2

TABLE S-VIII: **Feature Extraction Layer Ablations** Accuracies (in %) and their respective standard deviations (denoted by ±) of our feature extraction layer ablations on the Occluded-COCO-Vehicles dataset. * marks the results received after fine-tuning the network. CCE and BCE denote the Categorical (Softmax) and Binary Cross Entropy losses, respectively. A_{GAP} and A_{FC} denote the aggregation module with global average pooling and with a convolutional layer, respectively.

Train Data		**MS-COCO**				
Occ. Area		**L0**	**L1**	**L2**	**L3**	**Avg**
A_{GAP}+CCE+IC	Extraction Block P4	98.5 ± 0.1	87.6 ± 1.2	82.9 ± 2.4	71.2 ± 4.1	85.1 ± 1.3
A_{GAP}+CCE+IC*	Extraction Block P4	99.2 ± 0.1	92.9 ± 0.1	89.7 ± 0.5	84.6 ± 2.6	91.6 ± 0.8
A_{FC}+CCE+IC	Extraction Block P4	99.4 ± 0.0	92.7 ± 0.3	88.5 ± 1.4	79.1 ± 3.5	89.9 ± 1.1
A_{FC}+CCE+IC*	Extraction Block P4	99.5 ± 0.1	94.1 ± 0.0	90.4 ± 0.7	79.5 ± 2.6	90.9 ± 0.8
A_{GAP}+BCE+IC	Extraction Block P4	98.8 ± 0.1	88.1 ± 0.6	85.4 ± 1.3	79.8 ± 1.8	88.0 ± 0.7
A_{GAP}+BCE+IC*	Extraction Block P4	99.1 ± 0.2	93.2 ± 0.5	89.8 ± 0.6	89.4 ± 2.8	92.9 ± 0.7
A_{FC}+BCE+IC	Extraction Block P4	99.3 ± 0.0	92.8 ± 0.5	88.7 ± 0.6	80.1 ± 1.5	90.2 ± 0.4
A_{FC}+BCE+IC*	Extraction Block P4	99.5 ± 0.1	94.7 ± 0.3	90.8 ± 1.5	87.0 ± 3.1	93.0 ± 1.2
A_{GAP}+CCE+IC	Extraction Block P5	98.9 ± 0.1	88.2 ± 0.8	84.6 ± 1.2	73.6 ± 0.6	86.3 ± 0.5
A_{GAP}+CCE+IC*	Extraction Block P5	99.6 ± 0.1	94.2 ± 0.6	91.7 ± 1.0	85.3 ± 0.6	92.7 ± 0.5
A_{FC}+CCE+IC	Extraction Block P5	99.4 ± 0.1	92.7 ± 0.3	86.5 ± 1.6	78.1 ± 3.5	89.2 ± 1.3
A_{FC}+CCE+IC*	Extraction Block P5	99.6 ± 0.1	95.1 ± 0.2	92.2 ± 0.4	84.6 ± 0.6	92.9 ± 0.1
A_{GAP}+BCE+IC	Extraction Block P5	99.1 ± 0.1	89.8 ± 0.2	85.7 ± 2.1	76.0 ± 1.5	87.7 ± 0.6
A_{GAP}+BCE+IC*	Extraction Block P5	99.4 ± 0.0	94.9 ± 0.2	91.1 ± 1.2	88.4 ± 3.4	93.4 ± 1.0
A_{FC}+BCE+IC	Extraction Block P5	99.5 ± 0.0	93.1 ± 0.3	88.2 ± 0.9	81.8 ± 2.0	90.6 ± 0.7
A_{FC}+BCE+IC*	Extraction Block P5	99.5 ± 0.1	95.2 ± 0.6	94.0 ± 0.7	90.4 ± 0.0	94.8 ± 0.2
A_{GAP}+CCE+IC	Extraction Block P4 & P5	99.1 ± 0.1	90.6 ± 0.5	87.2 ± 1.1	87.0 ± 1.5	91.0 ± 0.6
A_{GAP}+CCE+IC*	Extraction Block P4 & P5	99.5 ± 0.1	94.5 ± 0.2	90.8 ± 1.3	89.4 ± 3.0	93.5 ± 1.0
A_{FC}+CCE+IC	Extraction Block P4 & P5	99.4 ± 0.1	94.2 ± 0.8	90.0 ± 0.8	82.9 ± 3.0	91.6 ± 1.0
A_{FC}+CCE+IC*	Extraction Block P4 & P5	99.4 ± 0.1	94.7 ± 0.6	90.8 ± 1.0	84.2 ± 0.7	92.3 ± 0.4
A_{GAP}+BCE+IC	Extraction Block P4 & P5	99.1 ± 0.1	89.4 ± 0.7	88.0 ± 0.5	86.0 ± 2.4	90.6 ± 0.8
A_{GAP}+BCE+IC*	Extraction Block P4 & P5	99.4 ± 0.0	94.6 ± 0.8	91.6 ± 0.5	92.8 ± 1.8	94.6 ± 0.4
A_{FC}+BCE+IC	Extraction Block P4 & P5	99.4 ± 0.1	94.1 ± 0.6	92.0 ± 0.4	86.3 ± 1.7	93.0 ± 0.3
A_{FC}+BCE+IC*	Extraction Block P4 & P5	99.5 ± 0.0	95.4 ± 0.4	92.1 ± 0.4	88.7 ± 2.4	93.9 ± 0.7

TABLE S-IX: The hyperparameters used in the training process of the ResNet50-networks. * denotes hyperparameters used during fine-tuning of the network, A_{GAP} denotes the aggregation module with global average pooling, and A_{FC} the module with kernel size equal to the input feature map. CCE and BCE denote the Categorical (Softmax) and Binary Cross Entropy losses, respectively.

Dataset	Agg.	Loss	LR	EP	LR*	EP*
Pascal3D+	A_{GAP}	CCE	1e-3	90	1e-5	90
		BCE	1e-3	90	1e-5	90
	A_{FC}	CCE	1e-4	90	1e-5	90
		BCE	1e-4	90	1e-5	90
MSCOCO	A_{GAP}	CCE	1e-3	180	1e-4	90
		BCE	1e-4	180	1e-4	90
	A_{FC}	CCE	1e-4	180	1e-5	90
		BCE	1e-4	180	1e-4	90

TABLE S-X: **ResNet50 Ablations** The accuracies (in %) and their respective standard deviations (denoted by $\pm$) of the ResNet50 ablation study of our Inverted Cutout method. * marks the results received after fine-tuning the whole network. The occlusion types are: w - white box occlusion, n - noise box occlusion, t - texture occlusion, o - occlusion by segmented objects. CCE and BCE denote the Categorical (Softmax) and Binary Cross Entropy loss, respectively. A_{GAP} and A_{FC} denote the aggregation module with global average pooling and the one utilizing a convolutional layer, respectively.

Occ. Area	**L0: 0%**	**L1: 20-40%**				**L2: 40-60%**				**L3: 60-80%**				Mean
Occ. Type	-	w	n	t	o	w	n	t	o	w	n	t	o	
A_{GAP}+CCE	99.7 ± 0.0	98.7 ± 0.2	98.5 ± 0.2	98.5 ± 0.2	93.6 ± 0.5	94.4 ± 0.4	92.9 ± 0.3	93.1 ± 0.5	72.5 ± 1.2	69.1 ± 1.7	61.3 ± 0.7	61.8 ± 2.0	52.8 ± 1.8	83.6 ± 0.5
A_{GAP}+CCE*	99.7 ± 0.1	98.3 ± 0.5	98.3 ± 0.3	97.9 ± 0.3	92.2 ± 2.4	91.9 ± 1.3	90.0 ± 1.9	89.9 ± 0.4	70.8 ± 6.5	58.4 ± 4.5	51.3 ± 8.1	51.7 ± 2.5	52.2 ± 5.7	80.2 ± 1.4
A_{FC}+CCE	99.8 ± 0.1	98.9 ± 0.2	98.4 ± 0.3	98.4 ± 0.1	94.2 ± 0.9	94.3 ± 0.5	93.2 ± 1.9	92.6 ± 1.1	76.2 ± 1.6	64.9 ± 3.1	59.8 ± 6.2	57.8 ± 4.9	57.9 ± 2.6	83.6 ± 1.5
A_{FC}+CCE*	99.8 ± 0.1	97.9 ± 0.8	97.3 ± 0.8	97.6 ± 0.6	91.8 ± 2.0	91.2 ± 3.2	88.0 ± 4.0	89.2 ± 3.0	72.3 ± 6.0	53.6 ± 11.7	46.8 ± 9.6	47.5 ± 7.5	53.6 ± 7.5	79.0 ± 4.0
A_{GAP}+BCE	99.6 ± 0.0	98.8 ± 0.1	98.5 ± 0.2	98.6 ± 0.1	94.0 ± 0.5	95.1 ± 0.3	93.1 ± 0.4	93.0 ± 0.1	75.8 ± 1.0	70.6 ± 1.2	60.5 ± 1.2	59.9 ± 1.9	58.3 ± 1.2	84.3 ± 0.2
A_{GAP}+BCE*	99.8 ± 0.1	98.7 ± 0.3	97.9 ± 0.5	97.9 ± 0.5	91.6 ± 2.1	92.6 ± 1.5	87.3 ± 2.6	89.3 ± 1.7	70.2 ± 3.8	59.8 ± 2.8	45.8 ± 3.5	48.1 ± 2.5	51.2 ± 4.0	79.3 ± 1.8
A_{FC}+BCE	99.8 ± 0.1	99.1 ± 0.1	98.7 ± 0.2	98.8 ± 0.0	94.3 ± 0.2	95.1 ± 0.3	94.1 ± 0.4	93.8 ± 0.2	76.6 ± 1.0	65.5 ± 1.1	60.8 ± 1.3	57.7 ± 1.8	57.5 ± 1.2	84.0 ± 0.3
A_{FC}+BCE*	99.8 ± 0.0	98.9 ± 0.2	98.6 ± 0.3	98.5 ± 0.4	92.2 ± 1.7	93.5 ± 0.8	91.0 ± 1.6	90.8 ± 1.7	70.9 ± 2.9	56.3 ± 2.1	48.8 ± 3.7	50.3 ± 4.4	53.5 ± 3.3	80.2 ± 1.7
A_{GAP}+CCE+IC	99.5 ± 0.1	98.9 ± 0.1	96.6 ± 0.2	97.7 ± 0.2	90.7 ± 1.0	96.9 ± 0.5	90.1 ± 0.6	91.5 ± 0.6	71.2 ± 3.3	87.2 ± 0.6	61.2 ± 3.1	57.4 ± 1.6	53.3 ± 3.6	84.0 ± 1.0
A_{GAP}+CCE+IC*	99.7 ± 0.0	99.5 ± 0.0	99.3 ± 0.1	99.3 ± 0.1	98.3 ± 0.2	99.2 ± 0.1	97.9 ± 0.2	97.7 ± 0.3	92.0 ± 1.2	95.7 ± 0.4	79.0 ± 3.8	78.1 ± 1.5	78.7 ± 1.8	93.4 ± 0.6
A_{FC}+CCE+IC	99.5 ± 0.1	99.4 ± 0.1	98.3 ± 0.2	98.8 ± 0.1	96.8 ± 0.3	98.7 ± 0.1	95.2 ± 1.0	96.0 ± 0.3	89.0 ± 0.6	92.9 ± 0.8	72.8 ± 3.9	70.9 ± 2.6	78.2 ± 1.6	91.3 ± 0.6
A_{FC}+CCE+IC*	99.8 ± 0.1	99.5 ± 0.1	99.3 ± 0.1	99.4 ± 0.1	98.7 ± 0.2	99.4 ± 0.1	98.3 ± 0.3	98.2 ± 0.2	95.7 ± 1.0	96.9 ± 0.4	83.1 ± 1.8	79.2 ± 1.1	87.2 ± 1.8	95.0 ± 0.4
A_{GAP}+BCE+IC	99.3 ± 0.1	99.1 ± 0.1	97.7 ± 0.4	98.3 ± 0.2	92.9 ± 0.6	97.7 ± 0.2	93.0 ± 1.1	94.4 ± 0.3	76.5 ± 0.8	89.9 ± 0.8	66.4 ± 1.2	68.8 ± 0.6	60.8 ± 1.8	87.3 ± 0.1
A_{GAP}+BCE+IC*	99.8 ± 0.0	99.6 ± 0.1	99.4 ± 0.1	99.5 ± 0.1	98.7 ± 0.1	99.2 ± 0.1	98.2 ± 0.2	98.5 ± 0.2	95.0 ± 0.5	95.6 ± 0.4	85.1 ± 2.7	84.5 ± 1.5	83.8 ± 0.9	95.1 ± 0.4
A_{FC}+BCE+IC	99.5 ± 0.1	99.4 ± 0.0	98.8 ± 0.1	99.0 ± 0.1	96.4 ± 0.1	98.9 ± 0.1	96.8 ± 0.4	97.1 ± 0.3	88.8 ± 1.0	93.4 ± 0.3	80.1 ± 2.0	77.5 ± 1.6	77.3 ± 1.9	92.6 ± 0.5
A_{FC}+BCE+IC*	99.8 ± 0.1	99.6 ± 0.0	99.4 ± 0.1	99.5 ± 0.1	99.0 ± 0.2	99.5 ± 0.1	98.8 ± 0.2	98.7 ± 0.3	95.8 ± 0.7	96.9 ± 0.4	87.5 ± 3.6	85.5 ± 1.2	86.1 ± 1.3	95.9 ± 0.5

TABLE S-XI: **ResNet50 Ablations** Accuracies (in %) and their respective standard deviations (denoted by ±) of our ResNet50 ablations on the Occluded-COCO-Vehicles dataset. * marks the results received after fine-tuning the network. CCE and BCE denote the Categorical (Softmax) and Binary Cross Entropy losses, respectively. A_{GAP} and A_{FC} denote the aggregation module with global average pooling and with a convolutional layer, respectively.

Train Data	**MS-COCO**				
Occ. Area	**L0**	**L1**	**L2**	**L3**	**Avg**
A_{GAP}+CCE	99.6 ± 0.1	92.4 ± 0.5	89.0 ± 0.7	79.8 ± 1.1	90.2 ± 0.3
A_{GAP}+CCE*	99.5 ± 0.1	89.3 ± 1.4	79.8 ± 2.0	67.5 ± 6.2	84.0 ± 2.3
A_{FC}+CCE	99.6 ± 0.1	93.1 ± 0.9	87.8 ± 1.8	75.0 ± 3.5	88.9 ± 1.5
A_{FC}+CCE*	99.5 ± 0.1	90.0 ± 1.0	82.7 ± 1.6	63.4 ± 3.3	83.9 ± 1.4
A_{GAP}+BCE	99.5 ± 0.0	92.1 ± 0.2	88.5 ± 0.3	72.9 ± 0.6	88.3 ± 0.1
A_{GAP}+BCE*	99.7 ± 0.1	91.1 ± 0.8	84.6 ± 2.9	71.6 ± 4.2	86.8 ± 1.9
A_{FC}+BCE	99.6 ± 0.1	93.5 ± 0.2	88.7 ± 0.4	75.7 ± 1.1	89.4 ± 0.3
A_{FC}+BCE*	99.5 ± 0.1	91.3 ± 0.4	84.1 ± 2.7	74.7 ± 4.1	87.4 ± 0.9
A_{GAP}+CCE+IC	99.2 ± 0.1	91.4 ± 0.5	85.4 ± 2.8	72.6 ± 1.9	87.2 ± 1.2
A_{GAP}+CCE+IC*	99.4 ± 0.1	93.1 ± 0.1	90.0 ± 1.6	85.6 ± 3.1	92.0 ± 1.2
A_{FC}+CCE+IC	99.6 ± 0.0	94.8 ± 0.4	91.3 ± 1.1	87.0 ± 1.2	93.2 ± 0.5
A_{FC}+CCE+IC*	99.4 ± 0.1	94.2 ± 0.8	91.0 ± 1.3	88.4 ± 2.5	93.3 ± 1.1
A_{GAP}+BCE+IC	99.4 ± 0.0	90.6 ± 0.3	84.2 ± 0.6	76.7 ± 1.7	87.7 ± 0.6
A_{GAP}+BCE+IC*	99.5 ± 0.0	93.7 ± 0.4	92.2 ± 1.1	88.4 ± 2.1	93.4 ± 0.7
A_{FC}+BCE+IC	99.5 ± 0.1	95.2 ± 0.3	90.6 ± 1.0	85.3 ± 0.6	92.6 ± 0.1
A_{FC}+BCE+IC*	99.6 ± 0.0	94.7 ± 0.4	92.6 ± 0.5	90.8 ± 2.0	94.4 ± 0.7