Instance-Aware Observer Network for Out-of-Distribution Object Segmentation

Victor Besnier[1,3,4] Andrei Bursuc[2] Alexandre Briot[1] David Picard[3]
1. Valeo, Créteil, France 2. Valeo.ai, Paris, France
3. LIGM, Ecole des Ponts, Univ Gustave Eiffel, CNRS, Marne-la-Vallée, France
4. ETIS UMR8051, CY Université, ENSEA, CNRS, Cergy France

Abstract

Recent works on predictive uncertainty estimation have shown promising results on Out-Of-Distribution (OOD) detection for semantic segmentation. However, these methods struggle to precisely locate the point of interest in the image, i.e, the anomaly. This limitation is due to the difficulty of fine-grained prediction at the pixel level. To address this issue, we build upon the recent ObsNet approach by providing object instance knowledge to the observer. We extend ObsNet by harnessing an instance-wise mask prediction. We use an additional, class agnostic, object detector to filter and aggregate observer predictions. Finally, we predict an unique anomaly score for each instance in the image. We show that our proposed method accurately disentangles in-distribution objects from OOD objects on three datasets.

1 Introduction

Lately, an ever increasing number of safety-critical systems, such as autonomous driving, are looking into leveraging Deep Neural Networks (DNNs) for the perception of the environment and as well as for subsequent decisions. Despite some success, DNNs still remain unreliable for real world deployment and often over-confident even they are incorrect on both In-Distribution [23] and Out-Of-Distribution (OOD) data [48, 28]. In this work, we aim at detecting OOD objects for 2D object segmentation. In this context, we consider as OOD the objects belonging to a class that is unknown by the perception system, i.e., a class that is not defined nor present in the training data.

Most methods dealing with *unknown-unknown* are based on ensembles[34, 41, 19, 43], pseudo-ensembles [33, 57, 18, 16], or deterministic approaches for computing uncertainty [30, 50, 56, 39]. However, most of them cannot simultaneously satisfy the real world requirements of high performance and real-time inference, typically trading one for the other. Recent works inspired by practices from system validation and monitoring, advance two-stage strategies to detect anomalies in semantic segmentation [5, 4]. An Observer Network is trained to analyze and predict the confidence of a main perception network. Observer-based approaches have been shown to find a good balance between accuracy and computational efficiency [5, 4]. In this work we build on top on observer-based approaches to leverage their properties.

We argue that pixel-wise error map (as shown in [5, 4, 33, 11]) by itself is sub-optimal for anomaly detection in segmentation because these maps lack clarity. Due to the difficulty of fine-grained prediction, most boundaries between two classes as well as small or distant objects are considered uncertain. Therefore, the focus of interest in the image, i.e. the OOD object, is drowned into this noise. The resulting error map does not provide precisely delimited spatial information: we know there is an error on the image but we struggle to accurately locate the corresponding OOD object. In other words, while we can get uncertainty estimates and predictions at pixel-level, extending them to objects is not obvious and we cannot find automatically objects far from the training distribution. As

ML Safety Workshop, 36th Conference on Neural Information Processing Systems (NeurIPS 2022).

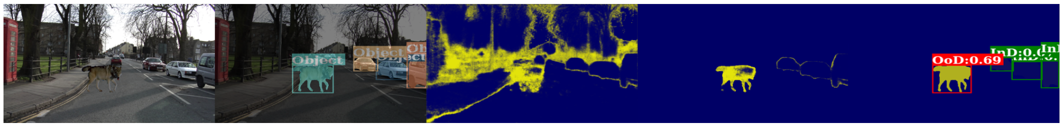

Figure 1: **Flows of the image processing.** From the input image (left), we compute the pixels-wise uncertainty and the object detection masks. Then we filter the uncertainty in the area of object only, and finally aggregate the score in an instance aware manner. We can see that the OOD object is well detected while in-distribution objects with low anomaly score and background errors are erased.

an example, an image depicting a crowd of pedestrian with lot of boundaries has, on average, higher anomaly score than an image with only one OOD object in the middle of a road.

In this paper, we propose to reduce the granularity of the task in order to improve the relevance of the error map. To this end, we use a class agnostic, instance segmentation network. With this additional prediction, we first filter background errors, and then aggregate uncertainty in an instance-aware manner Figure 2. Ultimately, we only want to highlight object instances with high errors. With this pragmatic and practical solution we can sort objects by anomaly score and then discard all objects close to the training distribution and keep those that are far from the training distribution.

2 Proposed Method

2.1 Observer Networks

Our work builds upon ObsNet [5, 4] that we briefly describe here. Observer networks are a two-stage method to detect pixels-wise errors and OOD. [4] designed two principles to train efficiently an auxiliary network. They improve the architecture by decoupling the OOD detector from the segmentation branch and by observing the whole network via residual connections. Secondly, they generate blind spots in the segmentation network with *local adversarial attacks* (LAA) at a random location of the image, mimicking an OOD object. ObsNet (Obs) outputs a pixels-wise error map corresponding to the probability that the semantic segmentation network (Seg) fails to predict the correct class y:

$$Obs(\boldsymbol{x}, Seg_r(\boldsymbol{x})) \approx Pr[Seg(\boldsymbol{x}) \neq y], \tag{1}$$

where x is the input image and Seg_r the skip connections from intermediate feature maps of segmentation network Seg.

2.2 Instance Anomaly Detection

To this end, we upgrade the semantic segmentation framework with instance hints. We use an additional class agnostic instance segmentation prediction. This detector (Det) produces a binary mask by mapping each object in the image.

Then, the idea is to separate the observer's prediction map into two categories. The background (classes of *stuff*) and the instance (classes of *things*) in the same way as the panoptic segmentation. Background errors correspond to global ambiguities in the scene at different scales: error at the decision boundary between two classes, prediction error between the road and the sidewalk or complexity of the leaves of a tree. In contrast, an instance error corresponds to an object far from the train distribution.

2.3 Uncertainty Aggregation and Filtering

In order to obtain a unique error score for each instance (similar to the well-known objectness score in object detection), we aggregate the per-pixel uncertainty within the predicted object mask to a unique value. In practice, given an image $\boldsymbol{x} \in \mathbb{R}^{3 \times H \times W}$, we predict for each detected object o_i an anomaly score $a_i \in \mathbb{R}$:

$$a_i = \frac{1}{M} \sum_{h=0}^{H} \sum_{w=0}^{W} u^{(h,w)} \odot m_i^{(h,w)}, \tag{2}$$

where $u = Obs(\boldsymbol{x}, Seg_r(\boldsymbol{x})) \in \mathbb{R}^{H \times W}$ is the pixel-wise error map of ObsNet; $m_i \in \mathbb{R}^{H \times W}$ is the binary mask of an instance o_i in the set of the detector prediction $Det(\boldsymbol{x}) = \{m_i\}$; $M = \sum_{h,w=0}^{H \times W} m_i$ the area of the instance o_i; and $\odot$ is the element-wise product. We also filter predicted instance masks m_i by size, in order to remove very small detected objects ($< 16^2$ pixels) in the image.

This strategy shows several benefits. We can discover instances in the dataset that do not match with the training distribution. We can also localize where the anomaly is in the image, which is a primary requirement for safety-critical applications such as autonomous driving. In Figure 1, we show that our framework is able to detect several instances in the images, and the ObsNet succeeds in discriminating in-distribution objects from out-of-distribution ones.

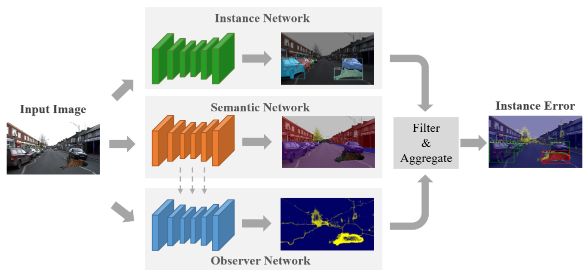

Figure 2: **Architecture.** The image go through the instance, the semantic and the observer network. Next, the ObsNet prediction is filter by the class-agnostic instance prediction and the remaining error is then aggregate object-wise.

Method	fpr95tpr↓	AuPR ↑	AuRoc ↑
Softmax [30]	70.0	11.45	76.7
ObsNet	40.5	22.72	87.9
ObsNet + *in*	31.3	49.4	92.3
ObsNet + *pan*	8.7	70.1	97.3
ObsNet + *gt*	**1.0**	**90.5**	**99.7**

Table 1: **Pixel-wise evaluation on CamVid OOD**. We consider OOD pixels only as the positive class. ObsNet + *gt-detector* has nearly perfect performance as every OOD instances are detected.

3 Experiments

We assess experimentally the effectiveness of our observer network coupled with an class-agnostic instance detector and compare it against several baselines.

3.1 Datasets & Metrics

We conducts experiments on the **CamVid OOD** [5], **StreetHazards** [29] and **BDD Anomaly** [60] datasets of urban streets scenes with anomalies in the test set. Anomalies correspond to OOD objects, not seen during training.

To evaluate each method on these datasets, we select four metrics to detect misclassified and out-of-distribution examples: **fpr95tpr** [36], **Area Under the Receiver Operating Characteristic curve (AuRoc)** [30], **Area under the Precision-Recall Curve (AuPR)** [30] and **Mean Average Prediction** (**mAP_δ**). We compute the latter metric where we discard object smaller than δ^2 pixels.

For each metric, we report the result where an object is considered as well detected if the predicted mask has $IoU > .5$ with the ground truth. We assign to each detected object the anomaly score computed as Equation 2. We use a Bayesian SegNet [2], [32] as the main network for CamVid and a DeepLabv3+ [8] for BDD Anomaly and StreetHazards. The ObsNet follows the same architecture as the corresponding segmentation network.

For our instance segmentation module, we select two Mask R-CNN variants [26]: one trained on CityScapes [12], reported as In-Distribution Detector, and one trained on MS-COCO [37], reported as Pan-Distribution. We do not leverage the class predicted but only the instance mask. Moreover, we use an additional oracle: we take every connected region of the same class in the annotation as one instance of an object, we report this *detector* as GT-detector.

We compare our method against **MCP** [30]: One minus the maximum of the prediction; and **MC Dropout** [21]: The entropy of the mean of 50 softmax prediction with dropout.

Det	Method	CamVid OoD				StreetHazard				Bdd Anomaly			
		$fpr95_{32} \downarrow$	$Roc_{32} \uparrow$	mAP_0	mAP_{32}	$fpr95_{48} \downarrow$	$Roc_{48} \uparrow$	mAP_{16}	mAP_{48}	$fpr95_{32} \downarrow$	$Roc_{32} \uparrow$	mAP_0	mAP_{32}
In	Softmax [30]	*	56.7	7.7	55.2	*	**50.4**	80.0	**81.9**	*	56.7	7.7	55.2
	MC Dropout [21]	*	58.3	8.4	58.5	*	50.3	80.0	81.2	*	58.3	8.4	58.5
	ObsNet	*	**60.5**	**9.9**	**63.7**	*	**50.4**	**80.1**	**81.9**	*	**60.5**	**9.9**	**63.7**
Pan	Softmax [30]	57.2	79.4	4.8	62.1	*	53.7	57.6	77.7	57.2	79.4	4.8	62.1
	MC Dropout [21]	52.8	84.6	6.9	70.8	*	53.6	**57.7**	77.6	52.8	84.6	6.9	70.8
	ObsNet	**46.4**	**89.6**	**11.3**	**81.4**	*	**54.1**	56.9	**77.8**	**46.4**	**89.6**	**11.3**	**81.4**
GT	Softmax [30]	43.3	80.4	10.8	72.1	80.0	85.2	88.9	99.0	43.3	80.4	10.8	72.1
	MC Dropout [21]	32.1	85.5	13.5	79.2	74.5	86.0	86.9	99.0	32.1	85.5	13.5	79.2
	ObsNet	**27.2**	**94.3**	**22.3**	**92.0**	**72.6**	**87.5**	**89.0**	**99.2**	**27.2**	**94.3**	**22.3**	**92.0**

Table 2: Instance-Wise Evaluation on CamVid OoD, StreetHazard and Bdd Anomaly. We consider OOD examples only as the positive class.

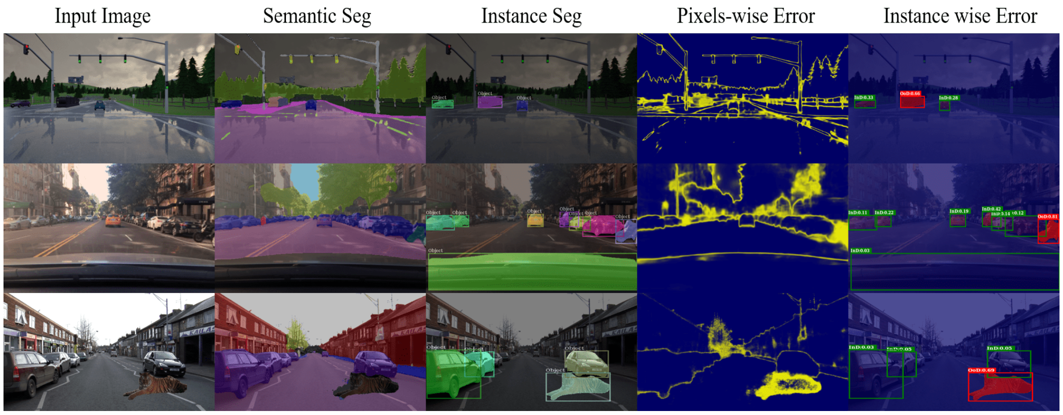

Figure 3: **Qualitative results on the StreetHazards (top row), BDD Anomaly (mid) and CamVid (bot).** From left to right, the input image; the semantic segmentation; the instance segmentation; the pixel-wise error and the instance-wise error. Our method is able to detect numerous objects and disentangle in-distribution objects ($a_i < .5$ in green) from out-of-distribution objects ($a_i > .5$ in red).

3.2 Benefit of the instance module

To validate the benefit of the instance detector, we first check that filtering the pixel-wise error map with the instance detector helps for pixel OOD detection, see Table 1. Using an instance detector improves significantly the performance of ObsNet. Moreover, this experiment shows that keeping raw error map is sub-optimal because many pixels with high anomaly score do not correspond to an OOD object but actually belong to the background of the images, whereas they can easily be filtered out by our instance scheme.

3.3 Instance-Wise Results

We can observe on Table 2 that for each dataset the results are quite different. This is due to the scale of the anomalies and the number of them in each the dataset. For CamVid, all anomalies are above 64^2 pixels, which can explain why the metrics drastically improve as we discard smaller detected objects. For StreetHazard, most of the objects are in fact anomalies, which is why mAP is high, even for the object below 32^2 pixels. Finally, even if on average *pan-detector* outperforms *in-detector*, this is not always the case for Bdd Anomaly. Indeed, *pan-detector* can detect more objects, and among them smaller in-distribution objects, that can hurt performances. Overall, ObsNet outperforms baseline methods, regardless of the detector.

We illustrate a few qualitative results in Figure 3. ObsNet emphasizes OOD objects with higher anomaly scores compared to in-distribution objects. Its predicted error maps are generally clearer and more accurate.

4 Conclusion

In this paper, with propose to use an additional, class-agnostic, object detector to filter and aggregate an anomaly score from ObsNet pixel-wise error map. Our strategy helps to better disentangle in from out-of-distribution objects.

References

[1] Tiago Azevedo, René de Jong, Matthew Mattina, and Partha Maji. Stochastic-yolo: Efficient probabilistic object detection under dataset shifts. In *NeurIPS Workshops*, 2020. 7

[2] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. PAMI*, 2017. 3

[3] Christoph Baur, Benedikt Wiestler, Shadi Albarqouni, and Nassir Navab. Deep autoencoding models for unsupervised anomaly segmentation in brain MR images. In *MICCAI Workshops*, 2018. 7

[4] Victor Besnier, Andrei Bursuc, David Picard, and Alexandre Briot. Triggering failures: Out-of-distribution detection by learning from local adversarial attacks in semantic segmentation. In *ICCV*, 2021. 1, 2

[5] Victor Besnier, David Picard, and Alexandre Briot. Learning uncertainty for safety-oriented semantic segmentation in autonomous driving. In *ICIP*, 2021. 1, 2, 3, 7

[6] Petra Bevandić, Ivan Krešo, Marin Oršić, and Siniša Šegvić. Simultaneous semantic segmentation and outlier detection in presence of domain shift. In *GCPR*, 2019. 7

[7] Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. Weight uncertainty in neural networks. *ICML*, 2015. 7

[8] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *ECCV*, 2018. 3

[9] Jiwoong Choi, Dayoung Chun, Hyun Kim, and Hyuk-Jae Lee. Gaussian YOLOv3: An accurate and fast object detector using localization uncertainty for autonomous driving. In *ICCV*, 2019. 7

[10] Jiwoong Choi, Ismail Elezi, Hyuk-Jae Lee, Clement Farabet, and Jose M Alvarez. Active learning for deep object detection via probabilistic modeling. In *ICCV*, 2021. 7

[11] Charles Corbière, Nicolas Thome, Avner Bar-Hen, Matthieu Cord, and Patrick Pérez. Addressing failure prediction by learning model confidence. In *NeurIPS*, 2019. 1, 7

[12] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The Cityscapes dataset for semantic urban scene understanding. In *CVPR*, 2016. 3

[13] Terrance DeVries and Graham W Taylor. Learning confidence for out-of-distribution detection in neural networks. *arXiv preprint arXiv:1802.04865*, 2018. 7

[14] Xuefeng Du, Xin Wang, Gabriel Gozum, and Yixuan Li. Unknown-aware object detection: Learning what you don't know from videos in the wild. In *CVPR*, 2022. 7

[15] Xuefeng Du, Zhaoning Wang, Mu Cai, and Sharon Li. Towards unknown-aware learning with virtual outlier synthesis. In *ICLR*, 2022. 7

[16] Nikita Durasov, Timur Bagautdinov, Pierre Baque, and Pascal Fua. Masksembles for uncertainty estimation. In *CVPR*, 2021. 1, 7

[17] Michael W. Dusenberry, Ghassen Jerfel, Yeming Wen, Yian Ma, Jasper Snoek, Katherine Heller, Balaji Lakshminarayanan, and Dustin Tran. Efficient and scalable bayesian neural nets with rank-1 factors. In *ICML*, 2020. 7

[18] Gianni Franchi, Andrei Bursuc, Emanuel Aldea, Séverine Dubuisson, and Isabelle Bloch. Encoding the latent posterior of bayesian neural networks for uncertainty quantification. *arXiv preprint arXiv:2012.02818*, 2020. 1, 7

[19] Gianni Franchi, Andrei Bursuc, Emanuel Aldea, Séverine Dubuisson, and Isabelle Bloch. TRADI: Tracking deep neural network weight distributions. In *ECCV*, 2020. 1, 7

[20] Yarin Gal. Uncertainty in Deep Learning. *PhD*, 2016. 7

[21] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *ICML*, 2016. 3, 4, 7

[22] Stefano Gasperini, Jan Haug, Mohammad-Ali Nikouei Mahani, Alvaro Marcos-Ramiro, Nassir Navab, Benjamin Busam, and Federico Tombari. Certainnet: Sampling-free uncertainty estimation for object detection. *RAL*, 2022. 7

[23] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural networks. In *ICML*, 2017. 1

[24] Ali Harakeh and Steven L. Waslander. Estimating and evaluating regression predictive uncertainty in deep object detectors. In *ICLR*, 2021. 7

[25] Elmar Haussmann, Michele Fenzi, Kashyap Chitta, Jan Ivanecky, Hanson Xu, Donna Roy, Akshita Mittel, Nicolas Koumchatzky, Clement Farabet, and Jose M Alvarez. Scalable active learning for object detection. In *IV*, 2020. 7

[26] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask R-CNN. In *ICCV*, 2017. 3

[27] Simon Hecker, Dengxin Dai, and Luc Van Gool. Failure prediction for autonomous driving. In *IV*, 2018. 7

[28] Matthias Hein, Maksym Andriushchenko, and Julian Bitterwolf. Why relu networks yield high-confidence predictions far away from the training data and how to mitigate the problem. In *CVPR*, 2019. 1

[29] Dan Hendrycks, Steven Basart, Mantas Mazeika, Mohammadreza Mostajabi, Jacob Steinhardt, and Dawn Song. A benchmark for anomaly segmentation. *arXiv preprint arXiv:1911.11132*, 2019. 3

[30] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *ICLR*, 2017. 1, 3, 4

[31] Dan Hendrycks, Mantas Mazeika, and Thomas Dietterich. Deep anomaly detection with outlier exposure. In *ICLR*, 2018. 7

[32] Alex Kendall, Vijay Badrinarayanan, , and Roberto Cipolla. Bayesian SegNet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. *arXiv preprint arXiv:1511.02680*,

2015. 3
[33] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? In *NeurIPS*, 2017. 1, 7
[34] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *NeurIPS*, 2017. 1, 7
[35] Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. In *NeurIPS*, 2018. 7
[36] Shiyu Liang, R. Srikant, and Yixuan Li. Enhancing the reliability of out-of-distribution image detection in neural networks. In *ICLR*, 2018. 3, 7
[37] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common objects in context. In *ECCV*, 2014. 3
[38] Krzysztof Lis, Krishna Nakka, Pascal Fua, and Mathieu Salzmann. Detecting the unexpected via image resynthesis. In *ICCV*, 2019. 7
[39] Jeremiah Zhe Liu, Zi Lin, Shreyas Padhy, Dustin Tran, Tania Bedrax-Weiss, and Balaji Lakshminarayanan. Simple and principled uncertainty estimation with deterministic deep learning via distance awareness. In *NeurIPS*, 2020. 1
[40] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *ECCV*, 2016. 7
[41] Wesley J Maddox, Pavel Izmailov, Timur Garipov, Dmitry P Vetrov, and Andrew Gordon Wilson. A simple baseline for bayesian uncertainty in deep learning. In *NeurIPS*, 2019. 1, 7
[42] Andrey Malinin and Mark Gales. Predictive uncertainty estimation via prior networks. In *NeurIPS*, 2018. 7
[43] Alireza Mehrtash, Purang Abolmaesumi, Polina Golland, Tina Kapur, Demian Wassermann, and William M Wells III. Pep: Parameter ensembling by perturbation. In *NeurIPS*, 2020. 1, 7
[44] Dimity Miller, Feras Dayoub, Michael Milford, and Niko Sünderhauf. Evaluating merging strategies for sampling-based uncertainty techniques in object detection. In *ICRA*, 2019. 7
[45] Mahdi Pakdaman Naeini, Gregory Cooper, and Milos Hauskrecht. Obtaining well calibrated probabilities using bayesian binning. In *AAAI*, 2015. 7
[46] Mahdi Pakdaman Naeini and Gregory F Cooper. Binary classifier calibration using an ensemble of near isotonic regression models. In *KDD*, 2016. 7
[47] Radford M Neal. Bayesian learning for neural networks. *arXiv*, 2012. 7
[48] A. Nguyen, J. Yosinski, and J. Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *CVPR*, 2015. 1
[49] John Platt et al. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *ALMC*, 1999. 7
[50] Janis Postels, Francesco Ferroni, Huseyin Coskun, Nassir Navab, and Federico Tombari. Sampling-free epistemic uncertainty estimation using approximated variance propagation. In *ICCV*, 2019. 1
[51] Quazi Marufur Rahman, Niko Sünderhauf, and Feras Dayoub. Online monitoring of object detection performance post-deployment. In *IROS*, 2020. 7
[52] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *NeurIPS*, 2015. 7
[53] Laurens Samson, Nanne van Noord, Olaf Booij, Michael Hofmann, Efstratios Gavves, and Mohsen Ghafoorian. I bet you are wrong: Gambling adversarial networks for structured semantic segmentation. In *ICCV Workshops*, 2019. 7
[54] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *IPMI*, 2017. 7
[55] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *JMLR*, 2014. 7
[56] Joost Van Amersfoort, Lewis Smith, Yee Whye Teh, and Yarin Gal. Uncertainty estimation using a single deep deterministic neural network. In *ICML*, 2020. 1
[57] Yeming Wen, Dustin Tran, and Jimmy Ba. Batchensemble: an alternative approach to efficient ensemble and lifelong learning. In *ICLR*, 2020. 1, 7
[58] Yingda Xia, Yi Zhang, Fengze Liu, Wei Shen, and Alan Yuille. Synthesize then compare: Detecting failures and anomalies for semantic segmentation. In *ECCV*, 2020. 7
[59] Donggeun Yoo and In So Kweon. Learning loss for active learning. In *CVPR*, 2019. 7
[60] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. BDD100K: A diverse driving dataset for heterogeneous multitask learning. In *CVPR*, 2020. 3
[61] Bianca Zadrozny and Charles Elkan. Obtaining calibrated probability estimates from decision trees and naive bayesian classifiers. In *ICML*, 2001. 7
[62] Bianca Zadrozny and Charles Elkan. Transforming classifier scores into accurate multiclass probability estimates. In *KDD*, 2002. 7

A Related Work

The widespread DNN adoption in this field has led to a fresh wave of approaches to improve OOD detection by input reconstruction [54, 3, 38, 58], predictive uncertainty [21, 33, 42], ensembles [34, 19, 16], adversarial attacks [36, 35], using a void or background class [52, 40] or dataset [6, 31, 42], *etc.*., to name just a few.

Bayesian approaches and ensembles. BNNs [47, 7] can capture predictive uncertainty from learned distributions over network weights, but don't scale well [17] and approximate solutions are preferred in practice. Deep Ensembles (DE) [34] is a highly effective, yet costly approach, that trains an ensemble of DNNs with different initialization seeds. Efficient or pseudo-ensemble approaches are a pragmatic alternative to DE that bypass training of multiple networks and generate predictions from different random subsets of neurons [20, 55, 16] or from networks sampled from approximate weight distributions [41, 19, 43, 57, 18]. However they all require multiple forward passes and/or storage of additional networks in memory.

Learning to predict errors. Inspired by early approaches from model calibration literature [49, 61, 62, 45, 46], a number of methods propose endowing the task network with an error prediction branch allowing self-assessment of predictive performance. This branch can be trained jointly with the main network [13, 59], however better learning stability and results are achieved with two-stage sequential training [11, 27, 5, 53].

OOD in object detection. Compared to classification and segmentation, OOD identification for 2D object detection is more challenging (mix of classification and regression predictions) and is less explored. Here, ensembles and pseudo-ensembles are often used for predictive uncertainty [44, 25, 1, 51]. Spatial uncertainty can be computed with slight changes in popular architectures and losses [9, 24, 10]. Other works leverage sampling-free approaches for both spatial and class uncertainty for real-time applications [22]. OOD detection itself has started to be addressed only recently by generating outliers in the feature space [15] or by "seeing" unknown objects from videos in the wild [14].

B Detected object histogram

In Figure 4, we report the histogram of objects detected by our detector, ranked by our framework. We can well disentangle in-distribution objects as cars, bicycles, or pedestrians, from OoD objects.

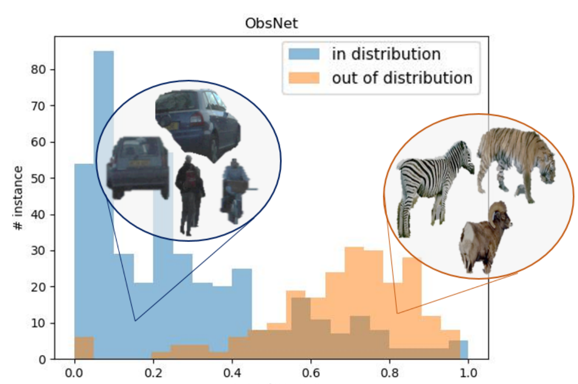

Figure 4: **Histogram on CamVid OOD**. Anomaly score from obsnet and detection from mask RcNN trained on the pan-distribution. We show here some examples of well-detected objects and predicted as in distribution in blue (left). While objects detected with high anomaly score (right) are considered as OOD in orange.