Probabilistic Object Detection: Strengths, Weaknesses, Opportunities

Dhaivat Bhatt[*1] **Dishank Bansal**[*1] **Gunshi Gupta**[*1] **Hanju Lee**[2] **Krishna Murthy Jatavallabhula**[1] **Liam Paull**[1]

Abstract

Deep neural networks are the de-facto standard for object detection in autonomous driving applications. However, neural networks cannot be blindly trusted even within the training data distribution, let alone outside it. This has paved way for several *probabilistic* object detection techniques that measure *uncertainty* in the outputs of an object detector. Through this position paper, we serve three main purposes. First, we briefly sketch the landscape of current methods for probabilistic object detecion. Second, we present the main shortcomings of these approaches. Finally, we present promising avenues for future research, and proof-of-concept results where applicable. Through this effort, we hope to bring the community one step closer to performing accurate, reliable, and consistent probabilistic object detection. A project page for this work can be found at montrealrobotics.ca/probod

1. Introduction

Detecting and localizing traffic participants and other objects is of paramount importance in autonomous driving scenarios. While deep neural networks (Krizhevsky et al., 2012) have been the de facto choice for object detection, their predictions are uninterpretable and unreliable outside the operating range (data distribution).

Reliably measuring the predictive uncertainty of blackbox object detection models benefits a range of downstream autonomous driving tasks, ranging from state estimation, to planning, to control. The operative word here is *reliable*: softmax-based classification models have been shown to give highly brittle and overconfident predictions (Guo et al., 2017; Moosavi-Dezfooli et al., 2016; Carlini & Wagner, 2017; Gal, 2016). This has seen *bayesian deep learning* techniques garner plenty of attention (Hall et al., 2018; Harakeh et al., 2019; Malinin & Gales, 2018; Lakshminarayanan et al., 2017). In this position paper, we analyze the strengths and weaknesses of, and opportunities presented by, probabilistic object detection techniques. We highlight open issues that need the attention of the community at large. We succinctly summarize current art and enlist critical open problems to enable widespread adaptation of probabilistic object detectors in figure 1.

Of central importance in the discussion of a *probabilistic* object detection framework is the various types/sources of uncertainties in the system. Drawing from previous studies (Kendall & Gal, 2017; Malinin & Gales, 2018; Hall et al., 2018), the types of uncertainties associated with an object detection framework are the following.

1. **Data uncertainty** (aleatoric uncertainty) is the noise in the data presented to the system. It is "irreducible" (can only be estimated), and corresponds to the underlying entropy in the data distribution.
2. **Model uncertainty** (epistemic uncertainty) is the uncertainty resulting due to the model structure and parameters.
3. **Distributional uncertainty** arises when models are presented with data outside of the training distribution.

The data and model uncertainties can further be decomposed into **spatial** and **semantic** uncertainties, concerning the uncertainties in the location and label of objects respectively.

We begin by providing a background and describing the **strengths** of current probabilistic deep learning methods and how they are employed for object detection. We critique the major **weaknesses** of these methods, and thereafter analyze the research **opportunities** that open up.

We communicate the following key conclusions.

- Uncertainty estimates need to be **calibrated** to down-

[*]Equal contribution Author ordering determined by Krishna holding a cookie behind his back and making Dhaivat guess "left" or "right". [1]Université de Montréal, Montreal, Canada. Mila - Québec AI Institute, Montreal, Canada [2]DENSO CORP.. Correspondence to: Dhaivat Bhatt <bhattdha@mila.quebec>, Dishank Bansal <bansaldi@mila.quebec>.

Workshop on AI for Autonomous Driving, the 37^{th} International Conference on Machine Learning, Vienna, Austria, 2020.

stream tasks; not necessarily to ground-truth data.

- The concept of a **background** class severely impacts the applicability of typical distributional uncertainty estimation methods designed for image classification and segmentation.
- Current metrics evaluate the **deterministic** behaviour (accuracy) of probabilistic object detectors, but fail to characterize their **probabilistic** nature.

2. Strengths

Today, there exist several approaches to estimating the predictive uncertainty of a deep neural network. In this section, we extensively survey the strengths existing approaches to probabilistic object detection techniques, categorized by the types/source of uncertainty measured (*cf*. Fig. 1 (left)).

Spatial uncertainty indicates the reliability of object localization, which is crucial for downstream tasks such as planning, motion prediction, tracking and collision avoidance. There exist different approaches for estimating uncertainty in object detection such as (He et al., 2019; Harakeh et al., 2019; Choi et al., 2019; Kraus & Dietmayer, 2019; Yoo et al., 2019; Meyer et al., 2019). To estimate the *model uncertainty* in bounding box predictions, (Harakeh et al., 2019) learn a stochastic detection model by employing dropout(Srivastava et al., 2014) at test time, as a Monte-Carlo estimator of variance. As argued in (Gal & Ghahramani, 2015), a neural network with dropout layers can be interpreted as encoding in its parameters a posterior over multiple hypotheses that minimize the training objective. Employing Monte-Carlo dropout at test time is analogous to sampling multiple predictions from this posterior, therefore the variance in these predictions is a measure of model uncertainty. Most approaches (He et al., 2019; Choi et al., 2019; Harakeh et al., 2019) leverage a loss attenuation mechanism to measure the *data uncertainty* in bounding box predictions. The loss attenuation mechanism involves simultaneously regressing to the mean and variance of a random variable, enabling the neural network to trade off accuracy and confidence of predictions. It learns to output variances corresponding to bounding box predictions as a measure of aleatoric uncertainty. It is derived as negative loss likelihood of gaussian:

$$L = \frac{(x-\mu)^2}{\sigma^2} + \log \sigma^2 \tag{1}$$

Semantic uncertainty is particularly useful in object detection, as downstream modules (trajectory forecasting, planning, obstacle avoidance) are often conditioned on a probabilistic labeling of the environment. In probabilistic object detection (Harakeh et al., 2019), semantic uncertainty is exclusively construed as model uncertainty.

[2]The references in red color are not object detection approaches, but methods to tackle OOD inputs in classification setup

Distributional uncertainty, to the best of our knowledge, has not been explored in the context of object detection. Two close sets of approaches exist, however.

1. *Transductive learning* approaches (Rahman et al., 2019b; Bansal et al., 2018; Zhu et al., 2019; Rahman et al., 2018; Rahman et al., 2019a; Gupta et al., 2020) detect unseen objects (zero-shot) by exploiting auxiliary information, such as the relationship between an unseen (test) object to a known (training) class, or pretrained word embeddings.
2. *Few-Shot Object Detection* methods (Wang et al., 2020; Kang et al., 2019; Yan et al., 2019; Karlinsky et al., 2019; Fan et al., 2019) employ meta-learning to localize novel objects, given a limited number of annotated examples.

BayesOD (Harakeh et al., 2019) is—perhaps—the only full-fledged probabilistic object detector (estimates both data and model uncertainty in the spatial and semantic components). It uses a RetiaNet (Lin et al., 2017b) backbone, and is trained with loss attenuation and MC-dropout.

3. Weaknesses

While probabilistic object detection has seen commendable strides over the last few years, current approaches have a long way to go before they are reliable enough to be deployed in real-world autonomous driving systems.

3.1. In-distribution uncertainty estimation

Sampling-based methods incur runtime overhead: To estimate the data uncertainty for in-distribution samples, methods such as MC-dropout (Kendall & Gal, 2017; Harakeh et al., 2019) have been proposed. These methods incur a high computational overhead as they require multiple forward passes at inference time. Non-stochastic methods (Postels et al., 2019; Choi et al., 2017) are yet to be evaluated in an object detection setup.

Lack of "calibration": Another limitation of most current methods is the lack of *calibrated* uncertainty estimates. That is, uncertainty estimates for each sample span an arbitrary, unknown scale, that are not commensurate with each other. While some approaches (Feng et al., 2019; Küppers et al., 2020) have studied this problem in-depth, they either require lidar data (Feng et al., 2019) or only estimate uncertainty in pixel space (Küppers et al., 2020). Pixel-space uncertainty is not directly usable by downstream modules operating in 3D. Feng *et al.* (Feng et al., 2019) show that the predicted variance may under-or-over-estimate the empirical distribution in the absence of calibration.

Figure 1. Left: A taxonomy of probabilistic object detection approaches[2]. *Right*: A summary of our analysis.

3.2. Out-of-distribution (OOD) uncertainty estimation

Predominant architectures for object detection include: single-stage methods (Szegedy et al., 2013; Sermanet et al., 2013; Redmon et al., 2016; Redmon & Farhadi, 2017; Liu et al., 2016) and two-stage methods (Girshick et al., 2015; He et al., 2015; Girshick, 2015; Ren et al., 2015; Dai et al., 2016; He et al., 2017; Li et al., 2017; Lin et al., 2017c;a). Most of these use "anchor boxes" to handle variation in the number, size, and position of objects in an image. Each anchor box is classified as either an in-distribution class or the background class.

So far, OOD detection methods (Hendrycks & Gimpel, 2016; Lakshminarayanan et al., 2017; Hein et al., 2019; Liang et al., 2017; Guo et al., 2017; Hendrycks et al., 2018; Mohseni et al.; Malinin & Gales, 2018; Sehwag et al., 2019) have been designed exclusively for image classification. Object detection methods, on the other hand, bear a different design philosophy. The notion of a *background* class in object detection obfuscates the distinction between novel objects and the background class. While some zero- and few-shot learning methods recognize novel objects in an image, zero-shot methods can only recognize OOD objects that are close to the training distribution (Rahman et al., 2019b; Bansal et al., 2018; Zhu et al., 2019; Rahman et al., 2018; Rahman et al., 2019a; Gupta et al., 2020), and few-shot methods (Wang et al., 2020; Kang et al., 2019; Yan et al., 2019; Karlinsky et al., 2019; Fan et al., 2019) require a handful of annotated samples for the OOD category at test time. Both these assumptions are impractical in a safety-critical application like autonomous driving.

3.3. Evaluation metrics

There is a lack of consensus on metrics to evaluate uncertainty estimates. The community needs metrics that evaluate "consistency" and "calibration", in addition to "accuracy". Current metrics (Hall et al., 2018) also lack the ability to evaluate OOD detections.

4. Opportunities

The above weaknesses open up several *opportunities* for further research. We discuss two key avenues here.

4.1. Task-calibrated uncertainty estimates

Uncertainty estimates from a probabilistic object detector need to be calibrated with the downstream task they are employed in. Existing approaches (Feng et al., 2019) propose multiple techniques to calibrate uncertainty estimates in image (pixel) space or world (3D) space. However, such uncertainties often need to be transformed suitable to be employed in a downstream task. Such a transformation is often nonlinear, lossy, and can lead to mischaracterization if the estimated uncertainty is improperly handled. A meaningful space to represent uncertainty, therefore, is in the action space of the immediate downstream task of interest. For example, in a driving system, if the role of object detection uncertainty is to inform state estimation or trajectory forecasting modules, the uncertainty estimates must be grounded in the input space of such modules. However, little research has been carried out to this end. While loss-attenuation based schemes demonstrate some correlation between the predicted uncertainties and empirical errors, whether they are consistent and meaningfully inform downstream tasks is an open question. Also, for temporally correlated input, the output maximum likelihood estimate and uncertainty should be correlated and consistent.

As a proof of concept, we show that using the predicted uncertainty in a loop within a Kalman Filter (KF) tracking system improves the calibration error of these uncertainties. We treat the predicted uncertainty as the measurement error within the KF. This supports the hypothesis that grounding the uncertainty with a task leads to more meaningful metrics. Calibration error is defined as $||\sigma - e||_1$ where e is prediction error $x - \mu$.

Method	Calibration Error
Without KF	0.625 px
With KF	0.551 px

4.2. Opportunities in Out of distribution object detection

In section 3, we observed that it is arduously challenging to distinguish between the background class and a novel object, owing to design decisions baked into modern object detectors. We now formalize the definition of OOD for object detection. There are two types of OOD objects in this setup: *seen OOD objects*, *unseen OOD objects*. Seen OOD objects are unannotated objects that are present in training images; these are implicitly learned as background. Unseen OOD objects are object that have never appeared in the train dataset, and only encountered at test time.

We conducted an experiment to understand the adverse impact of the background class for the "seen OOD objects" case. In this experiment, we use a clustering based method to classify OOD object. The overall objective in this method is to view an intermediate representation of an object in the trained model (input to softmax layer[3]) as residing on a high-dimensional manifold, and then, assuming that in-distribution samples should be "close" to each other on this manifold (Verma et al., 2018). We employ class conditional Gaussian clustering inspired from (Lee et al., 2018).

We train a model on x classes of the dataset and hold out k classes. We treat these k classes as *seen OOD objects*. We use the KITTI Object dataset (Geiger et al., 2012) with $x = 4$ and $k = 3$. We follow the procedure below,

1. Train the probabilistic detector to convergence
2. Get embeddings of in-distribution objects and background.
3. For each class c calculate a class-conditional Gaussian distribution (μ_c, Σ_c) based on the cluster embeddings
4. For a test datum x, calculate the closest class cluster in the embedding space, where "closeness" is determined by Mahalanobis distance:

$$M(x) = \max_c -(f(x) - \mu_x)\Sigma_c^{-1}(f(x) - \mu_x) \qquad (2)$$

5. If the closest class is above a threshold Mahalanobis distance, then the input is designated as out of distribution since it does not correspond well to any of the known classes.

We treat an acceptable accuracy of the in-distribution samples as the control variable. This control variable automatically presents a threshold on the Mahalanobis distance. Following the procedure above, for each class, we obtain several data points in the validation set and compute a Gaussian mean and covariance using the training data. We then use the validation data to find the Mahalanobis distance threshold that will achieve the acceptable accuracy, across all object classes. Next, using this threshold, we determine whether a new embedding belongs to any of these classes or not. If it doesn't we label it OOD. We perform this experiment for the two problem setups (the "easy" version and the "hard" version), for various acceptable accuracy thresholds, and then we calculate accuracy of OOD classes explicitly. To effectively understand role of the background class, we consider the following two setups. We follow procedure outlined in 4.2 for both the setups (with and without background class). That is, in one setup, we allow OOD embeddings to be classified as background, and in other setup, we do not. Resuts are shown in Fig 2.

This demonstrates that *redesigning object detectors without allowing a catch-all "background" class can boost OOD detection performance*.

[3] Note that getting intermediate representations is possible in 2-stage object detection only, we get intermediate representation from stage-2 of the Faster RCNN

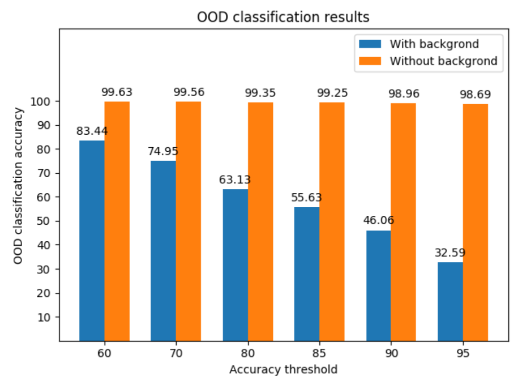

Figure 2. **Background class hampers OOD performance** as the cross-entropy objective forces an OOD object to be very confidently classified as background. Removing the notion of a background class boosts performance. Our results are also corroborated by (Denouden, 2020).

5. Conclusion

The highest level of safety guarantees are a necessity when deploying any system that can endanger human life. Autonomous vehicles are no exception. Probabilistic detectors are substantially more informative compared to their deterministic counterparts, and the measures of uncertainty they offer can be used to develop reliable and safe detection schemes. By summarizing the strengths and weaknesses of current art in probabilistic object detection, and by highlighting the most critical issues for further research, we hope to drive community efforts towards the right directions that will move us one step closer to safer autonomous vehicles.

References

Bansal, A., Sikka, K., Sharma, G., Chellappa, R., and Divakaran, A. Zero-shot object detection. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 384–400, 2018.

Carlini, N. and Wagner, D. Towards evaluating the robustness of neural networks. In *2017 IEEE Symposium on Security and Privacy (SP)*, pp. 39–57, 2017.

Choi, J., Chun, D., Kim, H., and Lee, H. J. Gaussian YOLOv3: An accurate and fast object detector using localization uncertainty for autonomous driving. *Proceedings of the IEEE International Conference on Computer Vision*, 2019-Octob:502–511, 2019. ISSN 15505499. doi: 10.1109/ICCV.2019.00059.

Choi, S., Lee, K., Lim, S., and Oh, S. Uncertainty-aware learning from demonstration using mixture density networks with sampling-free variance modeling, 2017.

Dai, J., Li, Y., He, K., and Sun, J. R-fcn: Object detection via region-based fully convolutional networks. In *Advances in neural information processing systems*, pp. 379–387, 2016.

Denouden, T. An application of out-of-distribution detection for two-stage object detection networks. Master's thesis, University of Waterloo, 2020.

Fan, Q., Zhuo, W., and Tai, Y.-W. Few-shot object detection with attention-rpn and multi-relation detector. *ArXiv*, abs/1908.01998, 2019.

Feng, D., Rosenbaum, L., Glaeser, C., Timm, F., and Dietmayer, K. Can We Trust You? On Calibration of a Probabilistic Object Detector for Autonomous Driving. 2, 2019. URL http://arxiv.org/abs/1909.12358.

Gal, Y. Uncertainty in deep learning. *University of Cambridge*, 1(3), 2016.

Gal, Y. and Ghahramani, Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning, 2015.

Geiger, A., Lenz, P., and Urtasun, R. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

Girshick, R. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, 2015.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. Region-based convolutional networks for accurate object detection and segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 38(1):142–158, 2015.

Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q. On calibration of modern neural networks. *arXiv preprint arXiv:1706.04599*, 2017.

Gupta, D., Anantharaman, A., Mamgain, N., S, S. K., Balasubramanian, V. N., and Jawahar, C. V. A multi-space approach to zero-shot object detection. In *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1198–1206, 2020.

Hall, D., Dayoub, F., Skinner, J., Zhang, H., Miller, D., Corke, P., Carneiro, G., Angelova, A., and Sünderhauf, N. Probabilistic object detection: Definition and evaluation, 2018.

Harakeh, A., Smart, M., and Waslander, S. L. BayesOD: A Bayesian Approach for Uncertainty Estimation in Deep Object Detectors. 2019. URL http://arxiv.org/abs/1903.03838.

He, K., Zhang, X., Ren, S., and Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(9):1904–1916, 2015.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.

He, Y., Zhu, C., Wang, J., Savvides, M., and Zhang, X. Bounding box regression with uncertainty for accurate object detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019-June:2883–2892, 2019. ISSN 10636919. doi: 10.1109/CVPR.2019.00300.

Hein, M., Andriushchenko, M., and Bitterwolf, J. Why relu networks yield high-confidence predictions far away from the training data and how to mitigate the problem. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 41–50, 2019.

Hendrycks, D. and Gimpel, K. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *arXiv preprint arXiv:1610.02136*, 2016.

Hendrycks, D., Mazeika, M., and Dietterich, T. Deep anomaly detection with outlier exposure. *arXiv preprint arXiv:1812.04606*, 2018.

Kang, B., Liu, Z., Wang, X., Yu, F., Feng, J., and Darrell, T. Few-shot object detection via feature reweighting. pp. 8419–8428, 10 2019. doi: 10.1109/ICCV.2019.00851.

Karlinsky, L., Shtok, J., Harary, S., Schwartz, E., Aides, A., Feris, R., Giryes, R., and Bronstein, A. Repmet: Representative-based metric learning for classification and few-shot object detection. pp. 5192–5201, 06 2019. doi: 10.1109/CVPR.2019.00534.

Kendall, A. and Gal, Y. What uncertainties do we need in Bayesian deep learning for computer vision? *Advances* *in Neural Information Processing Systems*, 2017-Decem (Nips):5575–5585, 2017. ISSN 10495258.

Kraus, F. and Dietmayer, K. Uncertainty estimation in one-stage object detection. *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, Oct 2019. doi: 10.1109/itsc.2019.8917494. URL http://dx.doi.org/10.1109/ITSC.2019.8917494.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Neural Information Processing Systems*, 2012.

Küppers, F., Kronenberger, J., Shantia, A., and Haselhoff, A. Multivariate confidence calibration for object detection, 2020.

Lakshminarayanan, B., Pritzel, A., and Blundell, C. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in neural information processing systems*, pp. 6402–6413, 2017.

Lee, K., Lee, K., Lee, H., and Shin, J. A simple unified framework for detecting out-of-distribution samples and adversarial attacks, 2018.

Li, Z., Peng, C., Yu, G., Zhang, X., Deng, Y., and Sun, J. Light-head r-cnn: In defense of two-stage object detector. *arXiv preprint arXiv:1711.07264*, 2017.

Liang, S., Li, Y., and Srikant, R. Enhancing the reliability of out-of-distribution image detection in neural networks. *arXiv preprint arXiv:1706.02690*, 2017.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2117–2125, 2017a.

Lin, T. Y., Goyal, P., Girshick, R., He, K., and Dollar, P. Focal Loss for Dense Object Detection. *Proceedings of the IEEE International Conference on Computer Vision*, 2017-Octob:2999–3007, 2017b. ISSN 15505499. doi: 10.1109/ICCV.2017.324.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017c.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C. Ssd: Single shot multibox detector. In *European conference on computer vision*, pp. 21–37. Springer, 2016.

Malinin, A. and Gales, M. Predictive uncertainty estimation via prior networks. In *Advances in Neural Information Processing Systems*, pp. 7047–7058, 2018.

Meyer, G. P., Laddha, A., Kee, E., Vallespi-Gonzalez, C., and Wellington, C. K. Lasernet: An efficient probabilistic 3d object detector for autonomous driving, 2019.

Mohseni, S., Pitale, M., Yadawa, J., and Wang, Z. Self-supervised learning for generalizable out-of-distribution detection.

Moosavi-Dezfooli, S.-M., Fawzi, A., and Frossard, P. Deepfool: A simple and accurate method to fool deep neural networks. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2574–2582, 2016.

Postels, J., Ferroni, F., Coskun, H., Navab, N., and Tombari, F. Sampling-free epistemic uncertainty estimation using approximated variance propagation. *Proceedings of the IEEE International Conference on Computer Vision*, 2019-Octob:2931–2940, 2019. ISSN 15505499. doi: 10.1109/ICCV.2019.00302.

Rahman, S., Khan, S., and Barnes, N. Polarity Loss for Zero-shot Object Detection. *arXiv e-prints*, art. arXiv:1811.08982, November 2018.

Rahman, S., Khan, S., and Barnes, N. Transductive learning for zero-shot object detection. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019a.

Rahman, S., Khan, S., and Porikli, F. Zero-shot object detection: Learning to simultaneously recognize and localize novel concepts. In Jawahar, C. V., Li, H., Mori, G., and Schindler, K. (eds.), *Computer Vision – ACCV 2018*, pp. 547–563, Cham, 2019b. Springer International Publishing. ISBN 978-3-030-20887-5.

Redmon, J. and Farhadi, A. Yolo9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7263–7271, 2017.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016.

Ren, S., He, K., Girshick, R., and Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pp. 91–99, 2015.

Sehwag, V., Bhagoji, A. N., Song, L., Sitawarin, C., Cullina, D., Chiang, M., and Mittal, P. Analyzing the robustness of open-world machine learning. In *Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security*, pp. 105–116, 2019.

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., and LeCun, Y. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*, 2013.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014. URL http://jmlr.org/papers/v15/srivastava14a.html.

Szegedy, C., Toshev, A., and Erhan, D. Deep neural networks for object detection. In *Advances in neural information processing systems*, pp. 2553–2561, 2013.

Verma, V., Lamb, A., Beckham, C., Najafi, A., Mitliagkas, I., Courville, A., Lopez-Paz, D., and Bengio, Y. Manifold mixup: Better representations by interpolating hidden states, 2018.

Wang, X., Huang, T. E., Darrell, T., Gonzalez, J. E., and Yu, F. Frustratingly Simple Few-Shot Object Detection. *arXiv e-prints*, art. arXiv:2003.06957, March 2020.

Yan, X., Chen, Z., Xu, A., Wang, X., Liang, X., and Lin, L. Meta r-cnn: Towards general solver for instance-level low-shot learning. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9576–9585, 2019.

Yoo, J., Seo, G., Chung, I., and Kwak, N. Mixture-model-based bounding box density estimation for object detection, 2019.

Zhu, P., Wang, H., and Saligrama, V. Zero shot detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(4):998–1010, 2019.