Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

b-it Bonn-Aachen International Center for Information Technology

German Research Center for Artificial Intelligence

Master Thesis Proposal

Benchmarking Out Of Distribution detection methods in 2D Object Detection

Jaswanth Bandlamudi

Supervised by

Dr. Paul G Plöger
Dr. Nico Hochgeschwender
Dr. Matias Valdenegro Toro
M.Sc. Octavio Arriaga

June 2021

1 Introduction

Science is a way to teach how something gets to be known,
what is not known, to what extent things are known...
— Richard P. Feynman

Deep Neural Networks (DNNs) which are trained discriminatively achieved best performances in tasks like speech recognition [Kłosowski, 2018], object detection [Zhao et al., 2019] and image classification [Voulodimos et al., 2018]. However, DNNs based models are proved to be resulting in overconfident predictions when the model encounters data that is different from the limited domain of expected inputs due to noise, adversarial corruptions, or other changes in distribution possibly due to temporal and spatial effects: a type of data often referred to as being out-of-distribution (OOD) inputs [Goodfellow et al., 2014, Nguyen et al., 2015]. This sort of behavior from DNNs is unsolicited in safety critical application like medical diagnosis and autonomous driving.

One such OOD input scenario is faced in Brazil by a Tesla vehicle operating in Autopilot mode [Ruffo, 2019], where a boy wearing an orange reflective (high-vis) jacket is detected and classified as an orange traffic cone. This erroneous behavior is dangerous since the true class of the input belongs to a movable object while it's classified to be an immovable object. This false detection further results in faulty predictions from motion prediction and planning algorithms which might result in serious injuries or fatalities.

Object detection is one of the tasks that is revolutionized with the application of Convolution Neural Network (CNN) based DNNs [Zhao et al., 2019]. Object detection answers the following questions: what is in an image, where is it, how accurate is the model in detection. But in real-world deployment, object detection systems has suffered from failures when:

1. Input is outside the semantic space formed by the images used for training the perception algorithm.
2. Input consists of objects which are not used in training but has features closer to the object of interest.

To illustrate this failure as shown in Figure 1, let us consider a model trained on Common Object in Context (COCO) object detection dataset [Lin et al., 2014]. We assume that the dataset is sampled from a distribution $p(X)$ at training time, the samples from the distribution $p(X)$ is considered to be in-distribution samples. Training a object detection model refers to modifying the model parameters thereby enabling the model to localize the objects in the image and classifying them into object of interests. When the model is made to perform inference on the input sampled from Indian Driving Dataset (IDD) [Varma et al., 2019] which consists of novel objects like Auto-Rickshaw that are not used to train the model. As illustrated in Figure 1a, the network detects the novel object wrongly with high probability of 81%. Figure 1b represents another OOD sample collected during a rainy day and the model couldn't detect the objects that are present in the image but detected a Dog which is not presented in the scene.

(a) False Positive detection

(b) False Negative detection

Figure 1: Image 1a represents False Positive detection in which an auto-rickshaw marked in blue is detected and classified as a car and Image 1b represents False Negative detection in which cars are present in the scene but are not detected.

Detecting and classifying these OOD samples results in the safe implementation of DNNs based object detection models where intentional or unintentional novel inputs may results in errors that might to lead to catastrophic failures. Autonomous Vehicle (AV) development is one research area which benefits from the implementation of such OOD methods because of change in weather conditions, illumination levels and also because of the novelty present in the semantics of the objects on the road. Detecting and localizing the OOD object present in the scene would help in better planning or fallback strategy.

1.1 Problem Statement

In this work, we study the problem of out-of-distribution inputs in object detection task trained in supervised setting. In particular, our task is to train a model using a dataset collected on urban environment in a clear weather and evaluate different OOD detection methods to detect objects in images sampled from out of distribution dataset.

The majority of OOD detection methods proposed until date are designed entirely for object classification tasks. These methods proposed for object classification are not entirely adaptable to object detection tasks. Except for, ODIN [Liang et al., 2017b] and uncertainty based methods [Malinin and Gales, 2018] because of the design philosophy of object detection problems includes foreground and background class. Background class is not generally included in training but should not be explicitly classified as a OOD data. The task of classifying between background class and OOD class is arduously challenging.

The task formulation of our problem is to train a DNN model M to perform regression task to extract the bounding box details together with a classification task to classify the type of object present in the image. We also employ another method or model to classify or quantify whether the input is OOD , we denote the model as M_{ood}. The deep learning model is trained using a dataset denoted by D_{train} and validated on a dataset denoted by D_{val} consisting of both in-distribution and OOD samples, Hence $D_{val} = D_{in-val} \bigcup D_{out-val}$. Finally, we test our model for benchmarking purposes on another dataset D_{test}, which is composed of both in-distribution and OOD samples $D_{test} = D_{in-test} \bigcup D_{out-test}$. Hence, an experiment E composes of M, M_{ood}, D_{train}, D_{in-val}, $D_{out-val}$, $D_{in-test}$, $D_{out-test}$.

For OOD dataset, as proposed by [Cao et al., 2020] three distinct out-of-distribution categories are considered

- Dataset from a different domain. For example, data that is collected in an urban environment is considered as in-distribution. While, the dataset collected in remote environment is used as OOD dataset.

- Dataset with poor quality features present. For example, data collected in an environment with poor illumination, rain and snowing conditions.

- Dataset with inputs that are not used or not prominent in the training data. For example, using KITTI dataset [Geiger et al., 2012] as an in-distribution dataset and using Indian Driving Dataset [Varma et al., 2019] which consists of novel unseen classes as OOD dataset.

The research question that are to be answered in this thesis are as follows:

RQ1: How to segregate in-distribution and out-of-distribution data in the context of object detection?

RQ2: How to perform out-of-distribution detection in the context of object detection?

RQ3: Which metrics can be used to compare various out-of-distrbution detection methods?

RQ4: How useful is uncertainty in classifying whether the input is an in-distribution or out-of-distribution?

2 Related Work

In this section, we introduce the relevant prior works in the field of object detection and out-of-distribution fields. We also contextualize the advantages and disadvantages of the prior works and their relevance to the out-of-distribution detection in object detection.

2.1 Object Detection

Object detection is a most sought-after application solved using computer vision. It involves solving multiple tasks like detecting, labelling and regressing the position of multiple objects present in the image.

2.1.1 Two-stage object detectors

Models in the two-stage detection family are all region-based. The detection happens in two stages:

1. First, the model proposes a set of regions of interests by select search or regional proposal network. The proposed regions are sparse as the potential bounding box candidates can be infinite.

2. A classifier only processes the region candidates to regress the bounding box coordinate and the object class.

The seminal work in two-stage object detectors is Regions with Convolution Neural Network (RCNN) proposed in 2013 by Girshick et al. [2014] which extracts proposals for the bounding box using a selective search algorithm. The extracted proposals are used to extract the features using CNN and the extracted features are classified using a Support Vector Machine (SVM). Although this learning based approach has out-performed the classical approaches, its inference times proved to be very high. Also, because of the complexities involved in training the RCNN this method is not preferred for real-time detection.

To address the disadvantages faced by RCNN, another method Fast RCNN by Girshick [2015] in which the entire image as opposed to processing only object proposals is processed using convolution layers. Then for each of the object proposals generated a feature vector is generated from the CNN feature map by using Region Of Interest (ROI) layer. For detection purposes, a Fully-Connected (FC) layer is employed using the feature vectors generated. Fast-RCNN outperformed previously proposed RCNN in terms of detection quality. Since, the trianing pipeline is also modified by including multi-task loss function and the usage of CNN for feature extraction led to reduction in memory usage. But there is no reduction in inference times due to the usage of selective search and edge boxes for generated proposals refinement.

In order to improve the inference speeds, a Region Proposal Network (RPN) is used to generate region proposals from convolution feature map instead of generating from image directly in Faster-RCNN by Ren et al. [2017]. RPN generates proposals in different aspect ratios, scales and provides the information about the location of pooling features to the bounding box regression model. Faster-RCNN model led to improved training and inference speeds due to the usage of RPN.

2.1.2 Single-stage object detectors

The single-stage models proposed till date skips the region proposal stage and runs detection directly over a dense sampling of possible locations. These methods generally has very low inference times but their performance in detection quality is deteriorated.

The pioneering work in single-stage object detection is proposed by Redmon et al. [2016], titled as You Only Look Once (YOLO). The approach in YOLO involves, dividing an image into a grid of size NxN and predicts the bounding boxes fitting the corresponding objects along with their classification scores from the extracted image features. It out performed the Deformable Part Model (DPM) by Yan et al. [2014] and R-CNN Girshick et al. [2014]. This method struggle from not able to detect objects of smaller sizes. There were revisions done to the initial YOLO models by the same authors and proposed YOLOv2 and YOLOv3, though there are improvements in detection accuracy but it is not considerable. A considerable update is done in YOLOv4 by Bochkovskiy et al. [2020] in which authors used different strategies in data augmentation (Mosaic, CutMix, Label smoothing), loss functions for bounding box (IoU loss, GIoU loss, DIoU loss, CIoU loss) and activation functions (leaky-ReLU, Swish, Mish). The model trained using the above advancements led to an increase in the Average-Precision (AP) values from 38.0% to 42.4%.

Single Shot Multi-box Detector Liu et al. [2016] is another single-stage object detector model in which object detection is done at several different layers of CNN instead of only doing it from the last layer. This technique allows us to detect objects of different sizes and also improved the accuracy predominantly. The current state-of-the-art (SOTA) object detector is EfficientDet by Tan et al. [2020]. The authors proposed Bi-directional Feature Pyramid Network (BiFPN) in which a modified feature pyramid network proposed by Liu et al. [2018] is used as a base network and is modified by varying the connections in Bi-FPN to modify the intermediate features to consider there by increasing the accuracy along with number of parameters that the model constitute. The authors also proposed a compound scaling method, which performs the up-scaling of resolution for regression of bounding box and object classification. There are eight different EfficientDet networks starting from D0 to D7, that is controlled with compound scaling coefficient ϕ.

Recently proposed methods applied probabilistic deep learning methods to the object detection models inorder to extract uncertainty in the bounding box values which acts as a better measure of confidence than the probablistic output from the final softmax layers. Few works which concentrated in this area are from Di Feng et al. [2019a,b], Schubert et al. [2020], Lee et al. [2020], Kraus and Dietmayer [2019]. While some methods concentrated on ensemble methods Lakshminarayanan et al. [2017] for uncertainty estimation other works modelled the parameters of the model as a prior distribution and trained the models with modified training pipeline to achieve better detection quality.

2.2 Out-of-distribution Detection

Out-of-distribution detection methods classifies the inputs that are sampled from another distribution from the inputs used for training the inference model. These methods can be classified into four different groups:Metric, Inconsistency, Generative and Uncertainty based approaches.

2.2.1 Metric based methods

Metric methods determine if the input is OOD based on the behavior of the current input follows the behavior of the samples inside the generalization area of the DNN under investigation.

One of the works done is by Devries and Taylor [2018] in which a classification model is modified to output an additional value c along with a softmax score. The training pipeline is modified to include a two-folded loss while penalizing a misclassification using cross entropy loss along with a penalty for lower confidence score c. Another method proposed by Oberdiek et al. [2018] in which intermediate values like layerwise norm, minimum and maximum value of the layer-wise gradient of the weights with respect to the loss function of the predicted class are fed to a logistic regressor along with entropy of the output class. The logistic regressor produces a score value which helps in deciding whether an input is OOD.

Hendrycks et al. [2018] also adapts the training procedure of the classification model. The authors extract a carefully curated dataset which contains samples outside the generalization area and assign an uniform distribution of values over all the class as the labels and are used while training along with the original training dataset. An input leading to an output with higher entropy is judged to be an OOD sample. Another work proposed by Lee et al. [2018] computes Mahalanobis distance between activation space of the DNN layers to the closest class-conditional Gaussian distribution. All the layerwise distances are fed to a logistic regression network which outputs a score value to decide whether the input is an OOD input

2.2.2 Inconsistency based methods

The basic idea behind inconsistency based methods is to induce a minimal change to the sample from in-distribution dataset. During inference both the image and the modified image are processed and the distance between the outputs is used to classify the in-distribution sample and OOD sample.

ODIN proposed by Liang et al. [2017a] in which the OOD detection is done by scaling the logit layers before the softmax output by a constant value called temperature, as well as perturbs the image input by a small amount. A sample is expected to be in-distribution, if the output value of the modified image for the original predicted class is high. These methods had resulted in a best performing model for detecting OOD data points.

2.2.3 Generative methods

Generative methods are similar to inconsistency based methods except in the way perturbation to the input images being performed. The perturbation is made to shift the image more towards the training distribution and is made using a generative network trained on the in-distribution training dataset. During inference, both the original and the generated image are processed and the higher the distance between both outputs the confident we are that the input sample is OOD.

One method proposed using this philosophy is proposed by Hendrycks and

Gimpel [2017] in which an encoder is employed on top of the penultimate layer of the original DNN, which helps in reconstructing the image. The difference between the original image and the generated image along with the output of the final layer and the penultimate layer to an abnormality module which regresses the confidence score that can be used to differentiate OOD input from an in-distribution input. Another such work is proposed by Ren et al. [2019] in which the authors used likelihood ratio value. if the likelihood ratio determined between the output of the model trained on the original images and the one on the perturbed images is low, the input image is expected to be OOD. The authors had trained two generative PixelCNN models by Van Den Oord et al. [2016], one on the original dataset and another on the slightly perturbed images from training dataset. From the observation they concluded that model trained on original images is sensitive to the true class contents compared to the model trianed on perturbed data. Hence, a low likelihood ratio value points to an OOD input.

2.2.4 Uncertainty based methods

An uncertainty measure could be directly applied to reject OOD samples as we would expect the uncertainty to be high on such inputs. There are multiple uncertainty types coined in various researches proposed by MacKay [1992a,b], Neal [1996]. The uncertainty extracted using Bayesian approaches is called epistemic uncertainty which measures the uncertainty in estimating the model parameters given the training data. In other words epistemic uncertainty measures how well the model is matched to the data in terms of model structure and parameters. The other type of uncertainty is aleatoric uncertainty, which is irreducible uncertainty that arises from the natural complexity of the data, e.g. from class overlap or label noise. Some works like Winkens [2019], Lakshminarayanan et al. [2017], Van Amersfoort et al. [2020] had presented an argument that OOD data is implicitly modelled by epistemic uncertainty.

Exploiting this idea, Lakshminarayanan et al. [2017] proposed the usage of ensemble of multiple models to extract epistemic uncertainty which can be used to judge whether an input is from in-distribution. Similarly, Malinin and Gales [2018] proposed prior networks for uncertainty estimation and dis-entangle the predictive uncertainty into three contributing factors model uncertainty, data uncertainty and distributional uncertainty. The authors used distributional uncertainty, which arises when an input is sampled from a distribution that is very different from the training distribution.

3 Performance Metrics

In this section, we present various possible evaluation metrics that can be used for measuring the performance of the object detectors. They also helps in measuring the prediction quality when faced with OOD samples.

3.1 Intersection over Union

Intersection over Union (IoU) is an evaluation metric that is employed to measure the detection accuracy of an object detector. To apply this metric we need ground-truth encoding of the bounding box and the encoding label of the bounding box.

$$IoU = \frac{Area\ of\ intersection\ of\ predicted\ and\ ground\ truth\ boxes}{Area\ of\ union\ of\ predicted\ and\ ground\ truth\ boxes} \tag{1}$$

3.2 Precision and Recall

Precision refers to the percentage of your results which are relevant and Recall refers to the percentage of total relevant results correctly classified by the model. An object detector is performing well if its precision is high with increase in recall value. Precision and Recall are calculated using True Positive, False Positive and False Negative rates.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

3.3 Average Precision

Average precision summarizes the precision value for 11 equally spaced recall values i,e. [0, 0.1, 0.2,.....,1]. The precision for each recall level is calculated as the maximum precision value measured where the corresponding recall is greater than or equal to corresponding recall value.

$$\mathrm{AP} = \frac{1}{11} \sum_{r \in \{0, 0.1, \ldots, 1\}} p_{\text{interp}}(r) \tag{4}$$

$$p_{\text{interp}}(r) = \max_{\hat{r}: \tilde{r} \geq r} p(\tilde{r}) \tag{5}$$

3.4 FPR @ 95 % TPR

This metric represents the probability of the model to misclassify an OOD sample as in-distribution sample when the in-distribution samples are classified correctly with at least 95% accuracy.

3.5 Probabilistic Detection Quality

Probabilistic Detection Quality (PDQ) proposed by Hall et al. [2020] not only captures the detection quality but also the spatial and label uncertainties regressed by the probabilistic object detection network. The calculation of PDQ is as follows

1. Calculate Foreground Loss L_{FG}

$$L_{FG}\left(\mathcal{G}_i^f, \mathcal{D}_j^f\right) = -\frac{1}{\left|\hat{\mathcal{S}}_i^f\right|} \sum_{\mathbf{x} \in \hat{\mathcal{S}}_i^f} \log\left(P\left(\mathbf{x} \in \mathcal{S}_j^f\right)\right) \tag{6}$$

2. Calculate Background Loss L_{BG}

$$L_{BG}\left(\mathcal{G}_i^f, \mathcal{D}_j^f\right) = -\frac{1}{\left|\hat{\mathcal{S}}_i^f\right|} \sum_{\mathbf{x} \in \mathcal{V}_{ij}^f} \log\left(\left(1 - P\left(\mathbf{x} \in \mathcal{S}_j^f\right)\right)\right) \tag{7}$$

3. Calculate Spatial Quality Q_S, e spatial quality measures how well the detection

describes where the object is within the image,

$$Q_S\left(\mathcal{G}_i^f,\mathcal{D}_j^f\right)=\exp\left(-\left(L_{FG}\left(\mathcal{G}_i^f,\mathcal{D}_j^f\right)+L_{BG}\left(\mathcal{G}_i^f,\mathcal{D}_j^f\right)\right)\right) \tag{8}$$

4. Calculate Label Quality Q_L,it measures how effectively a detection identifies what the object is, in other words it is defined as the probability estimated by the detector for the object's ground-truth class.

$$Q_L\left(\mathcal{G}_i^f,\mathcal{D}_j^f\right)=\mathbf{I}_j^f\left(\hat{c}_i^f\right) \tag{9}$$

5. Calculate pairwise PDQ (pPDQ) between a detection D_i^f and ground truth object G_i^f, it is the geometric mean of spatial quality Q_S and label quality Q_L.

$$\text{pPDQ}\left(\mathcal{G}_i^f,\mathcal{D}_j^f\right)=\sqrt{Q_S\left(\mathcal{G}_i^f,\mathcal{D}_j^f\right)\cdot Q_L\left(\mathcal{G}_i^f,\mathcal{D}_j^f\right)} \tag{10}$$

6. Finally, PDQ is calculated as

$$PDQ(\mathcal{G},\mathcal{D})=\frac{1}{\sum_{f=1}^{N_F}N_{TP}^f+N_{FN}^f+N_{FP}^f}\sum_{f=1}^{N_F}\sum_{i=1}^{N_{TP}^f}\mathbf{q}^f(i) \tag{11}$$

where, $q^f(i)$ indicates the pairwise PDQ score of the i^{th} detection-groundtruth pair in the f^{th} frame, $\sum_{f=1}^{N_F}N_{TP}^f+N_{FN}^f+N_{FP}^f$ is the total number of true positive, false negative and false positive assignments across all frames and N_F is the total number of frames.

4 Project Plan

In this section different tasks planned to solve the problem are listed and are assigned to respective work packages.

4.1 Work packages

WP1: Literature search

- Literature search on different image based object detection models
- Literature search for different out-of-distribution detection methods which can be adapted to the object detection task
- Literature search for probabilistic object detectors
- Documentation of literature

WP2: Propose a benchmark dataset with in-distribution and out-of-distribution samples

- Collect datasets proposed for solving 2D object detection task
- Classify dataset into in-distribution and out-of-distribution datasets
- Documentation of dataset chose

WP3: Implementation of object detection pipelines with OOD detection methods

- Implement a state-of-the-art two-stage object detector
- Implement a state-of-the-art one-stage object detector
- Implement a out-of-distribution detection method and integrate it with the implemented models
- Implement probabilistic object detection model
- Implement a method to model out-of-distribution decision making based on epistemic uncertainty from the probabilistic model
- Documentation of the models implemented

WP4: Experimentation and Comparative analysis

- Running experiments and collecting the metrics chose for benchmarking
- Documentation of experiments and results

WP5: Documentation

- Documentation of literature review (revision)
- Documentation of chosen datasets (revision)
- Documentation of implemented models and modifications (revision)
- Documentation of experimental results and comparisions (revision)
- Report completion and submission

4.2 Deliverables

Minimum Viable

- A survey on 2D object detection methods and out-of-distribution detection methods
- Propose a dataset for benchmarking out-of-distribution detection in 2D object detection
- Train and evaluate at least one state-of-the-art 2D object detection model
- Integrate one of the current out-of-distribution detection methods into detection pipeline

Expected

- Train and evaluate a probabilistic 2D object detector and extract epistemic uncertainty in detecting the objects
- Compare various uncertainty quantification methods in the context of out-of-distribution detection

Maximum

- Proposing a modification that can be made to improve the ability of the DNN models to detect out-of-distribution samples

- Evaluate the out-of-distribution methods on samples which consists of human-induced perturbation for confusing the DNN models (adverserial attack samples)

5 Project Schedule

The Thesis is scheduled for a 6 months period and the detailed plan can be found in Figure 2

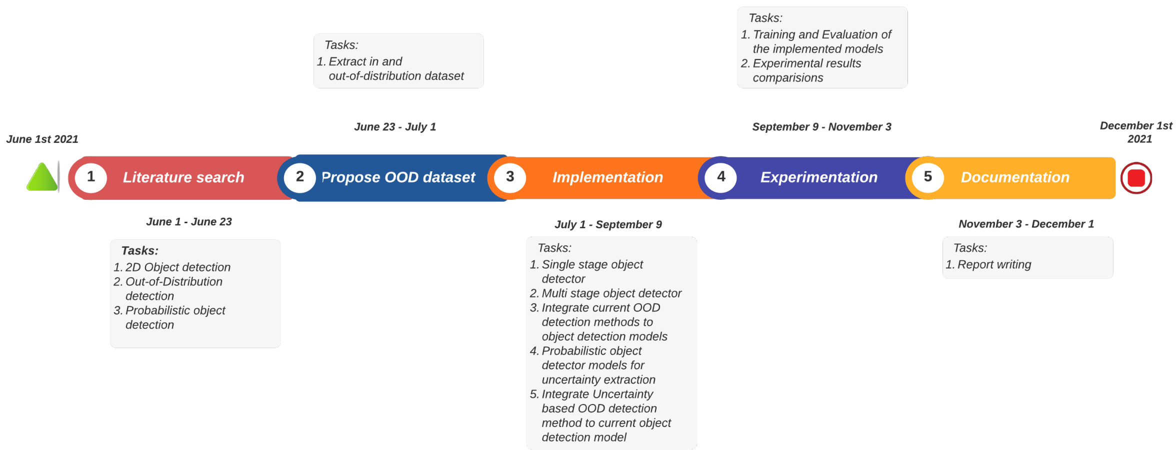

Figure 2: Project Timeline

References

Alexey Bochkovskiy, Chien Yao Wang, and Hong Yuan Mark Liao. YOLOv4: Optimal Speed and Accuracy of Object Detection. In *arXiv*, 2020.

Tianshi Cao, Chin-Wei Huang, David Yu-Tung Hui, and Joseph Paul Cohen. A Benchmark of Medical Out of Distribution Detection. Technical report, 2020.

Terrance Devries and Graham W Taylor. Learning Confidence for Out-of-Distribution Detection in Neural Networks. 2018.

Di Feng, Lars Rosenbaum, Claudius Gläser, Fabian Timm, and Klaus Dietmayer. Can we trust you? On calibration of a probabilistic object detector for autonomous driving. In *arXiv*, 2019a.

Di Feng, Lars Rosenbaum, Claudius Gläser, Fabian Timm, and Klaus Dietmayer. Can we trust you? On calibration of a probabilistic object detector for autonomous driving. In *arXiv*, 2019b.

A Geiger, P Lenz, and R Urtasun. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

Ross Girshick. Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2015 International Conference on Computer Vision, ICCV 2015, pages 1440–1448, 2015. ISBN 9781467383912.

Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 580–587, 2014. ISBN 9781479951178.

Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. 2014. cite arxiv:1412.6572.

David Hall, Feras Dayoub, John Skinner, Haoyang Zhang, Dimity Miller, Peter Corke, Gustavo Carneiro, Anelia Angelova, and Niko Sünderhauf. Probabilistic object detection: Definition and evaluation. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2020.

Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, 2017. ISBN 1610.02136v3.

Dan Hendrycks, Mantas Mazeika, and Thomas Dietterich. Deep anomaly detection with outlier exposure. In *arXiv*, 2018.

Florian Kraus and Klaus Dietmayer. Uncertainty Estimation in One-Stage Object Detection. Technical report, 2019.

P. Kłosowski. Deep learning for natural language processing and language modelling. In *2018 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, pages 223–228, 2018.

Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in Neural Information Processing Systems*, volume 2017-December, pages 6403–6414, 2017.

Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. Technical report, 2018.

Youngwan Lee, Joong Won Hwang, Hyung Il Kim, Kimin Yun, and Jongyoul Park. Localization uncertainty estimation for anchor-free object detection. In *arXiv*, 2020.

Shiyu Liang, Yixuan Li, and Rayadurgam Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks. *arXiv preprint arXiv:1706.02690*, 2017a.

Shiyu Liang, R. Srikant, and Yixuan Li. Enhancing the reliability of out-of-distribution image detection in neural networks. Technical report, 2017b.

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 David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 740–755, Cham, 2014. Springer International Publishing. ISBN 978-3-319-10602-1.

Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, and Jiaya Jia. Path Aggregation Network for Instance Segmentation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 8759–8768, 2018. ISBN 9781538664209.

W. Liu, Dragomir Anguelov, D. Erhan, Christian Szegedy, Scott E. Reed, Cheng-Yang Fu, and A. Berg. Ssd: Single shot multibox detector. In *ECCV*, 2016.

David J. C. MacKay. A practical bayesian framework for backpropagation networks. *Neural Comput.*, 4(3):448–472, May 1992a. ISSN 0899-7667.

David J.C. MacKay. *Bayesian methods for adaptive models*. PhD thesis, California Institute of Technology, 1992b.

Andrey Malinin and Mark Gales. Predictive uncertainty estimation via prior networks. In *Advances in Neural Information Processing Systems*, volume 2018-December, pages 7047–7058, 2018.

Radford M. Neal. *Bayesian Learning for Neural Networks*. Springer-Verlag, Berlin, Heidelberg, 1996. ISBN 0387947248.

Anh Nguyen, Jason Yosinski, and Jeff Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 427–436, 2015.

Philipp Oberdiek, Matthias Rottmann, and Hanno Gottschalk. Classification uncertainty of deep neural networks based on gradient information. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 11081 LNAI, pages 113–125, 2018. ISBN 9783319999777.

Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2016-December, pages 779–788, 2016. ISBN 9781467388504.

Jie Ren, Peter J. Liu, Emily Fertig, Jasper Snoek, Ryan Poplin, Mark A. DePristo, Joshua V. Dillon, and Balaji Lakshminarayanan. Likelihood ratios for out-of-distribution detection. In *arXiv*, 2019.

Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In *IEEE Transac-*

tions on Pattern Analysis and Machine Intelligence, volume 39, pages 1137–1149, 2017.

Gustavo Henrique Ruffo. Tesla autopilot confuses boy in orange shirt for a cone in brazil, Dec 2019. URL `https://insideevs.com/news/388253/tesla-autopilot-confuses-boy-for-cone/`.

Marius Schubert, Karsten Kahl, and Matthias Rottmann. MetaDetect: Uncertainty quantification and prediction quality estimates for object detection. In *arXiv*, 2020.

Mingxing Tan, Ruoming Pang, and Quoc V. Le. EfficientDet: Scalable and efficient object detection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 10778–10787, 2020.

Joost Van Amersfoort, Lewis Smith, Yee Whye Teh, and Yarin Gal. Simple and scalable epistemic uncertainty estimation using a single deep deterministic neural network. Technical report, 2020.

Aäron Van Den Oord, Nal Kalchbrenner, Oriol Vinyals, Lasse Espeholt, Alex Graves, and Koray Kavukcuoglu. Conditional image generation with PixelCNN decoders. In *Advances in Neural Information Processing Systems*, pages 4797–4805, 2016.

G. Varma, A. Subramanian, A. Namboodiri, Manmohan Chandraker, and C. V. Jawahar. Idd: A dataset for exploring problems of autonomous navigation in unconstrained environments. *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1743–1751, 2019.

Athanasios Voulodimos, Nikolaos Doulamis, Anastasios Doulamis, Eftychios Protopapadakis, and Diego Andina. Deep learning for computer vision: A brief review. *Intell. Neuroscience*, 2018, January 2018.

Jim Winkens. Out-of-distribution detection for computational pathology with multi-head ensembles. Master's thesis, University of Amsterdam, 2019.

Junjie Yan, Zhen Lei, Longyin Wen, and Stan Z. Li. The fastest deformable part model for object detection. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2497–2504, 2014.

Zhong-Qiu Zhao, Peng Zheng, Shou-Tao Xu, and Xindong Wu. Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*, 30(11):3212–3232, 2019.