ESTIMATING UNCERTAINTY USING GENERATIVE MODELS IN OBJECT DETECTION

H.J.M., van Genuchten
ID: 1297333
Artificial Intelligence and Data Engineering Lab
Eindhoven University of Technology
`h.j.m.v.genuchten@student.tue.nl`

December 18, 2023

ABSTRACT

Write abstract

1 Overall aim and goals

The main goal of this thesis will be the implementation of a (generative) Object Detection model that is able to provide reliable uncertainty estimates on its own predictions. Secondary to that,

- Create a (generative) model which can predict bounding boxes with Uncertainty Quantification.
  i.e. Detecting weird combinations of objects within a scene
- Validate uncertainty of the model.
  Does certainty and precision correlate. (Statistical analyses)
- Test uncertainty as measure for active learning.
  To make more efficient use of human labelers. One possibility is to take a subset of e.g. 10% of a labeled dataset, and then iteratively select more data. Comparing it to a random subset of equal size.
- Verify uncertainty is a good predictor for OoD data.
- Design a system which can make efficient use of uncertainty of previous stages.

1.1 Possible RQs

Current Ideas for possible RQs. (For the seminar we need to have (at least) 5 RQs, they do not need to all be answered)

- Can generative models be used for uncertainty prediction in Object Detection?
- Does increasing the iterations of a generative model decrease the uncertainty?
- Is uncertainty a good indicator for active learning?
- Is uncertainty a good indicator for relabeling?
- Can temporal data be used to improve Uncertainty Quantification?

1.2 Motivation and Challenges

Artificial Intelligence solutions are increasingly put in new and challenging scenarios. They are known to work well when the training and inference set are from the same distribution [15]. However, it has been shown that models will predict with high scores for inputs that are not relevant [19, 24]. It has also been shown that this can be used to attack these networks[13, 6]. Uncertainty Quantification (UQ) can enable a system to detect when a prediction might be of lesser quality [5], and allow it to preemptively react to that[22]. Either by stopping or requesting human intervention. Furthermore, understanding when models are uncertain, allows for more effective data sampling and labeling by making use of active learning schemes [21, 1, 26]. The latter is especially useful in industries where labeling is expensive or time-consuming.

Object Detection is an especially difficult field as the model has many outputs of different types. It requires both regression (for localization) and classification [10]. Moreover, there are multiple uncertainties present:

- Objectness: Is a bounding box an object (of interests)?
- Classification: Is a bounding box of a certain class?
- Size: Is the size of the bounding box correct?
- Location: Is the location of the bounding box correct?

1.3 Broad Literature Analysis

This project covers broader research areas, each will be covered separately in subsections 1.3.1 and 1.3.2. Related work in the combination will be described in subsection 1.3.3

1.3.1 Uncertainty Quantification

The ability to distinguish certain and uncertain outputs from machine learning models is useful for various reasons. It can be used for active learning [26, 21, 14, 1], which makes better use of limited labeling capacity.

[8] distinguishes two kinds of uncertainty. Aleatoric uncertainty, the uncertainty that is caused by imprecise input data, and epistemic uncertainty, the uncertainty that is caused by the model. Aleatoric uncertainty is often caused by imprecise measurements and can be reduced by improving the quality of our dataset and inputs. The latter, epistemic uncertainty, can be reduced by improving either the training procedure, increasing the amount of data or improving the model architecture.

Uncertainty Quantification (UQ) is an important factor to increase the trust in automated processes based on machine learning. Current methods often sample from existing networks [9, 17, 18] or predict parameters of a distribution [4, 23].

1.3.2 Object Detection

There are two main paradigms within Object Detection, One-Stage detectors[28, 2, 25, 16, 7], which directly predict both the bounding box and class in a single forward pass, and Multi-Stage detectors [12, 11], which first detect regions of interests to then subsequently classify these regions. A more recent development is DiffusionDet [3], which iteratively 'de-noises' a set of random bounding boxes towards the ground truth bounding boxes. Furthermore, there are some general tricks that have shown to provide improved performance at the cost of an increased training time for most object detection algorithms. These have been dubbed "Bag of Freebies" by [27], as they do not increase inference time.

One-Stage Detectors The Single Shot Detector proposed by [16] makes use of many anchor boxes. For each anchor box an offset and class is predicted, which are subsequently merged using Non-Max-Suppression (NMS). During training all bounding boxes with an Intersection over Union (IoU) greater than a threshold are matched and should be predicted with that class. All unmatched boxes should be classified as a 'background' class. This leads to a huge class imbalance. To tackle this, hard negative mining is used. For every positive match, at most n (in the paper 3) negative boxes are included during the loss calculation.

1.3.3 In combination

The combination of Uncertainty Quantification and Object Detection is especially difficult as Object Detection is both a regression and a classification task.

[10]

1.4 Formulation of the problem and objectives

Model formulation Let $x \in \mathbb{R}^{HxWxC}$ be the input image which contains objects O_i, which are a combination of a bounding box: $B_i \in \mathbb{R}^{N4}$ and a class $c \in C$.

Metric formulation

Mathematically formulate the following metrics

- IoU
- mean Average Precision
- Uncertainty Error [18]
- (AU)ROC [18]
- (AU)Precision-Recall [20]

Formulate (uncertainty) Metrics

References

[1] Mélanie Bernhardt et al. “Active label cleaning for improved dataset quality under resource constraints”. In: *Nature Communications* 13.1 (Mar. 2022). ISSN: 2041-1723. DOI: 10.1038/s41467-022-28818-3. URL: http://dx.doi.org/10.1038/s41467-022-28818-3.

[2] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. *YOLOv4: Optimal Speed and Accuracy of Object Detection*. 2020. arXiv: 2004.10934 [cs.CV].

[3] Shoufa Chen et al. *DiffusionDet: Diffusion Model for Object Detection*. 2023. arXiv: 2211.09788 [cs.CV].

[4] Jiwoong Choi et al. “Gaussian yolov3: An accurate and fast object detector using localization uncertainty for autonomous driving”. In: *Proceedings of the IEEE/CVF International conference on computer vision*. 2019, pp. 502–511.

[5] Sina Däubener et al. *Detecting Adversarial Examples for Speech Recognition via Uncertainty Quantification*. 2020. arXiv: 2005.14611 [eess.AS].

[6] Yinpeng Dong et al. “Boosting adversarial attacks with momentum”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 9185–9193.

[7] Kaiwen Duan et al. “Centernet: Keypoint triplets for object detection”. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2019, pp. 6569–6578.

[8] Yarin Gal et al. “Uncertainty in deep learning”. In: (2016).

[9] Yarin Gal and Zoubin Ghahramani. *Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning*. 2016. arXiv: 1506.02142 [stat.ML].

[10] Stefano Gasperini et al. “CertainNet: Sampling-Free Uncertainty Estimation for Object Detection”. In: *IEEE Robotics and Automation Letters* 7.2 (Apr. 2022), pp. 698–705. ISSN: 2377-3774. DOI: 10.1109/lra.2021.3130976. URL: http://dx.doi.org/10.1109/LRA.2021.3130976.

[11] Ross Girshick. “Fast r-cnn”. In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1440–1448.

[12] Ross Girshick et al. “Rich feature hierarchies for accurate object detection and semantic segmentation”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 580–587.

[13] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. “Explaining and harnessing adversarial examples”. In: *arXiv preprint arXiv:1412.6572* (2014).

[14] Neil Houlsby et al. *Bayesian Active Learning for Classification and Preference Learning*. 2011. arXiv: 1112.5745 [stat.ML].

[15] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “Imagenet classification with deep convolutional neural networks”. In: *Advances in neural information processing systems* 25 (2012).

[16] Wei Liu et al. “Ssd: Single shot multibox detector”. In: *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer. 2016, pp. 21–37.

[17] Wesley J Maddox et al. “A Simple Baseline for Bayesian Uncertainty in Deep Learning”. In: *Advances in Neural Information Processing Systems*. Ed. by H. Wallach et al. Vol. 32. Curran Associates, Inc., 2019. URL: https://proceedings.neurips.cc/paper_files/paper/2019/file/118921efba23fc329e6560b27861f0c2-Paper.pdf.

[18] Dimity Miller et al. “Evaluating merging strategies for sampling-based uncertainty techniques in object detection”. In: *2019 international conference on robotics and automation (icra)*. IEEE. 2019, pp. 2348–2354.

[19] Anh Nguyen, Jason Yosinski, and Jeff Clune. “Deep Neural Networks Are Easily Fooled: High Confidence Predictions for Unrecognizable Images”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2015.

[20] Takaya Saito and Marc Rehmsmeier. “The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets”. In: *PloS one* 10 (Mar. 2015), e0118432. DOI: `10.1371/journal.pone.0118432`.

[21] Burr Settles. “Active learning literature survey”. In: (2009).

[22] David John Stracuzzi et al. “Quantifying Uncertainty to Improve Decision Making in Machine Learning.” In: (Oct. 2018). DOI: `10.2172/1481629`. URL: `https://www.osti.gov/biblio/1481629`.

[23] Jakub Swiatkowski et al. *The k-tied Normal Distribution: A Compact Parameterization of Gaussian Mean Field Posteriors in Bayesian Neural Networks*. 2020. arXiv: `2002.02655 [cs.LG]`.

[24] Christian Szegedy et al. “Intriguing properties of neural networks”. In: *arXiv preprint arXiv:1312.6199* (2013).

[25] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors”. In: *arXiv preprint arXiv:2207.02696* (2022).

[26] Bishan Yang et al. “Effective multi-label active learning for text classification”. In: *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2009, pp. 917–926.

[27] Zhi Zhang et al. *Bag of Freebies for Training Object Detection Neural Networks*. 2019. arXiv: `1902.04103 [cs.CV]`.

[28] Xingyi Zhou, Dequan Wang, and Philipp Krähenbühl. “Objects as Points”. In: *arXiv preprint arXiv:1904.07850*. 2019.