Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

German
Research Center
for Artificial
Intelligence

Bonn-Aachen
International Center for
Information Technology

Benchmarking Out-of-Distribution Detection in 2D Object Detection

Thesis Defense

March 6, 2022

Jaswanth Bandlamudi

Supervisors

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

Introduction

- Deep Neural Networks, current State-Of-The-Art (SOTA) performers in
  - Classification
  - Object Detection
  - Segmentation
- Trained with *closed world assumption*, test data $\sim$ train data
- Deployed in open world $\Longrightarrow$ Out-of-Distribution(OOD) examples
- Applications
  - Product recommendations, recoverable
  - Time series prediction, partially reversible
  - Autonomous driving / Medical diagnosis, irreversiable and catastrophic

Out-of-Distribution (OOD) detection (1/3)

- What is OOD data ?
  - Data that is outside the semantic space formed by the images used for training
  - Input with objects which are not used in training but have features closer to the object of interest.

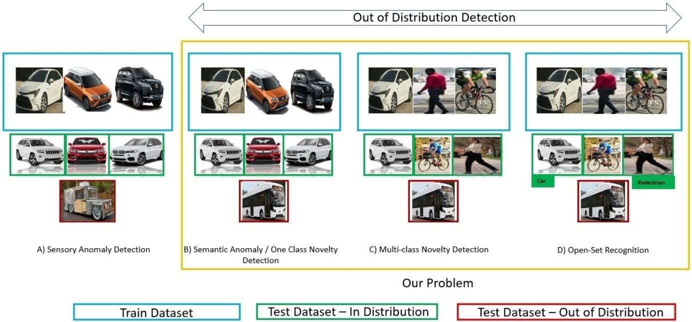


Figure: Class differentiation in generalized OOD detection framework

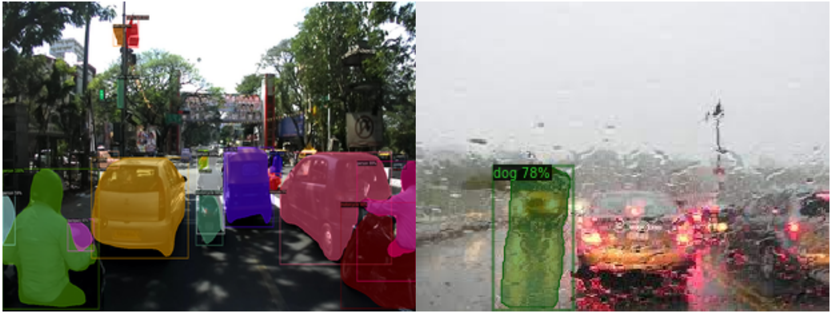
Out-of-Distribution (OOD) detection(2/3)

Different types of OOD data

- Data from a different domain
- Data with poor quality of features
- Data with inputs that are neither used nor prominent in the training data

Out-of-Distribution (OOD) detection(3/3)

Current Object Detection model performance on OOD data



(a) (b)

Figure: Examples of failures in object dedtection

OOD detector - Expectations

- Produce a ***Novelty Score (NS)***.
- NS can be a distance metric, a class-dependent probabilistic value, an entropy value, or a descriptive statistic value
- OOD detection can be posed as a binary classification problem.

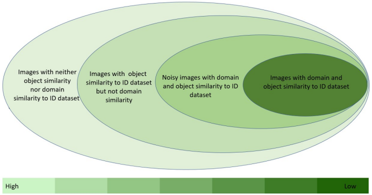


Figure: Expected behavior of OOD detector.

$$X = \begin{cases} \text{ID}, & \text{if } NS \geq \delta \\ \text{OOD}, & \text{otherwise} \end{cases} \tag{1}$$

Previous works

Table: Previous works on OOD detection

Method	Works Proposed
Metric based methods	Devries and Taylor [2018], Oberdiek et al. [2018], Hendrycks et al. [2018] , Lee et al. [2018]
Inconsistency based methods	Liang et al. [2017]
Generative methods	Hendrycks and Gimpel [2017], Ren et al. [2019], Van Den Oord et al. [2016]
Uncertainty based methods	Malinin and Gales [2018], Lakshminarayanan et al. [2017], Van Amersfoort et al. [2020]

- Works only for classification problem
- Not directly adaptable to object detection problem

OD^2 Dataset

Table: Table showing various type of images to address the OOD cases

Purpose	Dataset Source	Classes	Novelty Score Behavior	Task
In-Distribution	BDD100K [Yu et al., 2020]	Pedestrian, Rider, Car, Truck, Bus, Motorcycle, Bicycle, Traffic sign	Low Novelty score	Object detector performance
Low light and bad image quality	BDD100K (non-clear weather) and Climate-GAN [Schmidt et al., 2021] generated Smog images	Pedestrian, Rider, Car, Truck, Bus, Motorcycle, Bicycle, Traffic sign	Medium Novelty Score	Detector Robustness
Classes with semantic-variance	IDD [Varma et al., 2019]	Trucks, Motorcycles, Traffic Sign	High Novelty Score	OOD detection
Novel Classes	IDD	Auto-Rickshaws	High Novelty Score	Multi class novelty detection
Out-of-Domain images	Climate-GAN generated Flood and Fire images	Pedestrian, Rider, Car, Truck, Bus, Motorcycle, Bicycle, Traffic sign	High Novelty Score	Out-Of-Domain detection

Single Shot multi-box Detector (SSD) model (1/2)

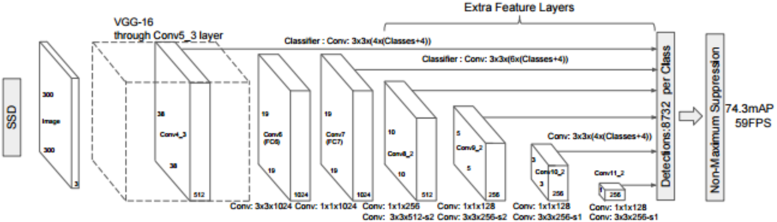


Figure: SSD framework proposed by Liu et al. [2016, p. 24].

- Single network for detection and classification
- No Fully-Connected layers
- Low input resolution

Single Shot multi-box Detector (SSD) model (2/2)

- Default boxes
- Matching strategy is used,
  - $IoU^{groundtruth}_{defaultbox}$ > 0.5
  - overlapped objects and simple learning
- Processing of features from multiple layers
  - Deep feature maps
  - Shallow feature maps

- Loss

$$L(x, c, l, g) = \frac{1}{N}\left(L_{\text{conf}}(x, c) + \alpha L_{\text{loc}}(x, l, g)\right)$$

  - L_{conf} is Softmax Loss
  - L_{loc} is Smooth L_1 Loss
- Filter boxes with low confidence and NMS with 0.45 IOU
- Top 200 detections are considered

OOD methods (1/5)

- Max-Softmax

  Maximum value of softmax scores are used as novelty score

$$s\left(\mathbf{x}^{*}\right)=\max_{c} P\left(y_{c} \mid \mathbf{x}^{*} ; \mathcal{D}\right) \tag{2}$$

- ODIN

$$\tilde{\boldsymbol{x}}=\boldsymbol{x}-\operatorname{esign}\left(-\nabla_{\boldsymbol{x}} \log S_{\hat{y}}(\boldsymbol{x} ; T)\right) \tag{3}$$

$$S_{i}(\boldsymbol{x} ; T)=\frac{\exp \left(f_{i}(\boldsymbol{x}) / T\right)}{\sum_{j=1}^{N} \exp \left(f_{j}(\boldsymbol{x}) / T\right)} \tag{4}$$

  - ϵ is the perturbation magnitude
  - T is the Temperature

OOD methods (2/5)

- Mahalanobis distance based OOD detection
  assuming intermediate layer features follow class-conditional Gaussian distributions with tied covariances

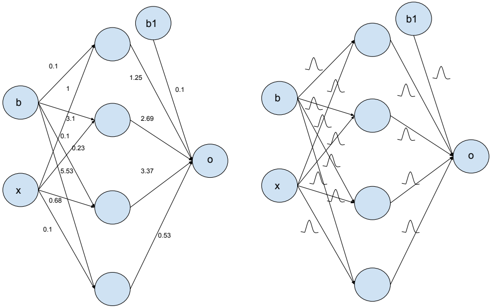
$$M(x) = \max_c -\left(f(x) - \hat{\mu}_c\right)^T \hat{\Sigma}^{-1} \left(f(x) - \hat{\mu}_c\right) \tag{5}$$

$$\hat{\mu}_c = \frac{1}{N_c} \sum_{i:y_c=c} f\left(x_i\right)$$

$$\hat{\Sigma} = \frac{1}{N} \sum_c \sum_{i:y_c=c} \left(f\left(x_i\right) - \hat{\mu}_c\right)\left(f\left(x_i\right) - \hat{\mu}_c\right)^T$$

OOD methods (3/5)

- Uncertainty based OOD detection
  - Bayesian Neural Network



» Bayesian Flipout layers [Wen et al., 2018]

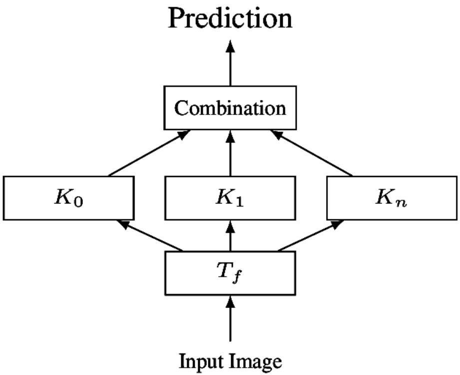
» Reparameterization trick for training [Kingma et al., 2015]

» Prior $P(w) \sim N(0, 1)$

» multiple forward passes for uncertainty quantification

OOD methods (4/5)

– Sub-Ensemble Network



- Model is divided into Trunk and Task layers
- Trunk layers has best performing weights restored and cannot be trained
- Task layers are randomly initialized and re-trained
- Random initialization of layers creates ensemble model.

OOD methods (5/5)

- Novelty Score

Entropy

$$\text{Entropy} = -\sum_{i=1}^{C} P\left(c_i \mid \mathbf{x}^*; \mathcal{D}\right) \ln P\left(c_i \mid \mathbf{x}^*; \mathcal{D}\right) \tag{6}$$

Box deviation is the square root of the trace of the covariance matrix $C(x^*)$.

$$C\left(\mathbf{x}^*\right) = \frac{1}{N} \sum_{i=1}^{N} \hat{\mathbf{v}}_{\mathbf{x}^*}^{i} \hat{\mathbf{v}}_{\mathbf{x}^*}^{i\,T} - \mathbf{I}_{\mathbf{x}^*} \mathbf{I}_{\mathbf{x}^*}^{T} \tag{7}$$

SSD Object Detection Results

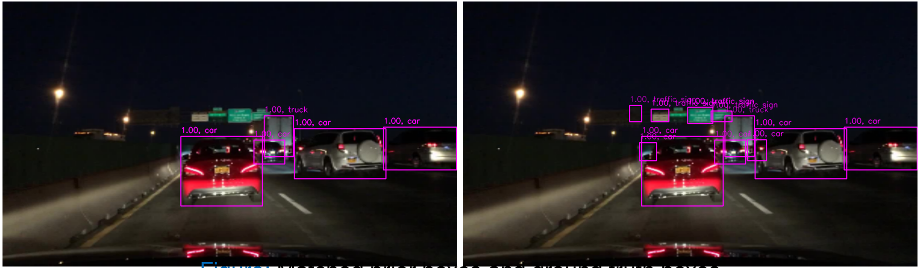


Figure: Matched prior boxes and ground truth boxes

Table: Average Precision values using SSD300 model with and without tuned prior boxes

	Agents								
Models	**Pedestrian**	**Rider**	**Car**	**Truck**	**Bus**	**Motorcycle**	**Bicycle**	**Traffic Sign**	**mean**
SSD300	0.006	0.004	0.095	0.083	0.15	0.045	0.092	0.001	0.059
SSD300 - Tuned	0.165	0.135	0.479	0.339	0.389	0.163	0.213	0.186	0.265

References (1/4)

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

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.

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

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.

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

References (2/4)

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.

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.

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.

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.

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.

References (3/4)

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.

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 **IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**, June 2020.

Victor Schmidt, Alexandra Sasha Luccioni, Mélisande Teng, Tianyu Zhang, Alexia Reynaud, Sunand Raghupathi, Gautier Cosne, Adrien Juraver, Vahe Vardanyan, Alex Hernandez-Garcia, and Yoshua Bengio. Climategan: Raising climate change awareness by generating images of floods, 2021.

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.

Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. Ssd: Single shot multibox detector. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, **Computer Vision – ECCV 2016**, pages 21–37, Cham, 2016. Springer International Pub [illegible] N 97 [illegible] 0.

References (4/4)

Yeming Wen, Paul Vicol, Jimmy Ba, Dustin Tran, and Roger Grosse. Flipout: Efficient pseudo-independent weight perturbations on mini-batches. Technical report, 2018.

Durk P Kingma, Tim Salimans, and Max Welling. Variational dropout and the local reparameterization trick. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, **Advances in Neural Information Processing Systems 28**, pages 2575–2583. Curran Associates, Inc., 2015.