Anomaly Detection in Autonomous Driving

GongLei

Outline

Background

Datasets

Methods

Our Works

Background

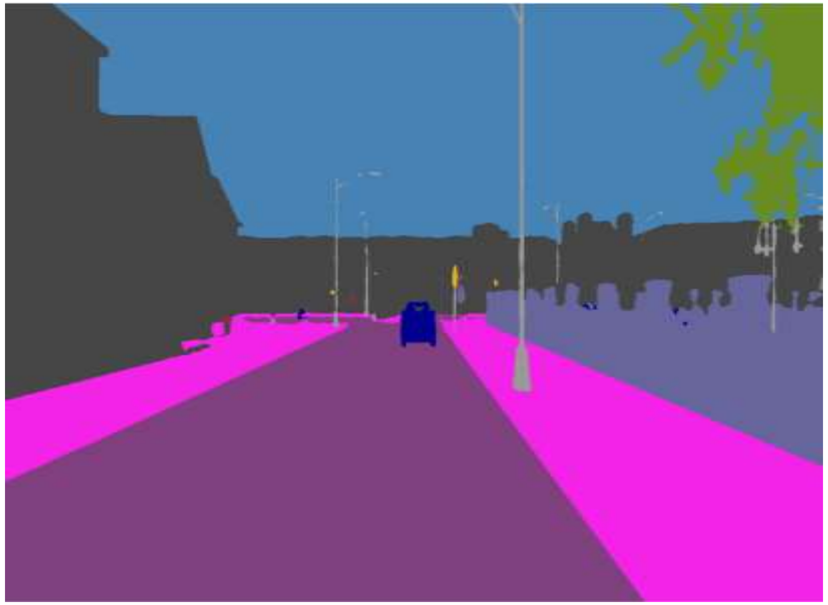
Tesla on Autopilot Crashes into Overturned Truck on Busy Highway

Background

Input

Ground Truth

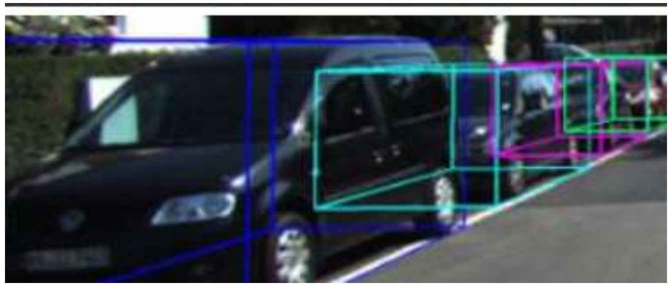
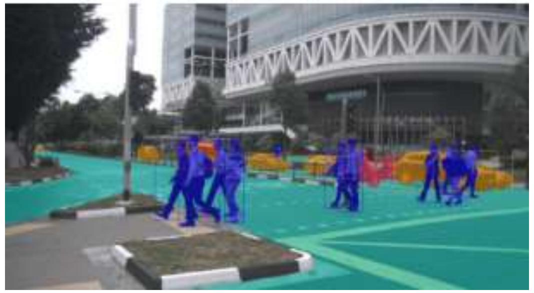
model output



Model trained on data without truck class label incorrectly labels the overturned white truck as a background building.

Background

- **Closed-set VS Open World Environment**

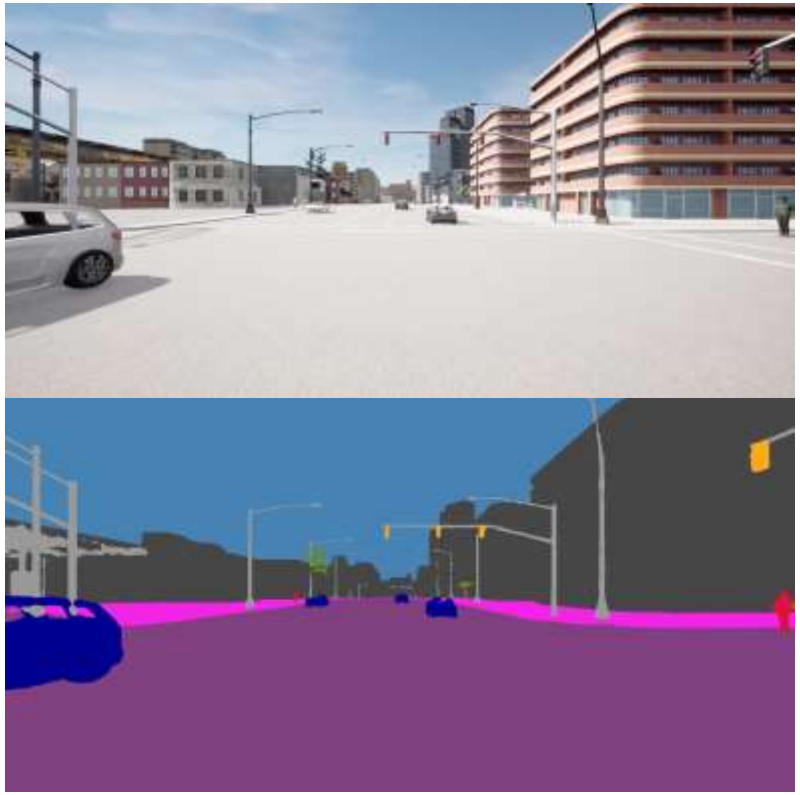
	KITTI	Waymo	nuScenes
Existing DATASETs	**KITTI**	**Waymo**	**nuScenes**
Perception Tasks			

Methods(From CNN to Transformer， From single modality to Fusion) work under the assumption that **all the classes and condition in the testing environment would be available** at training phase

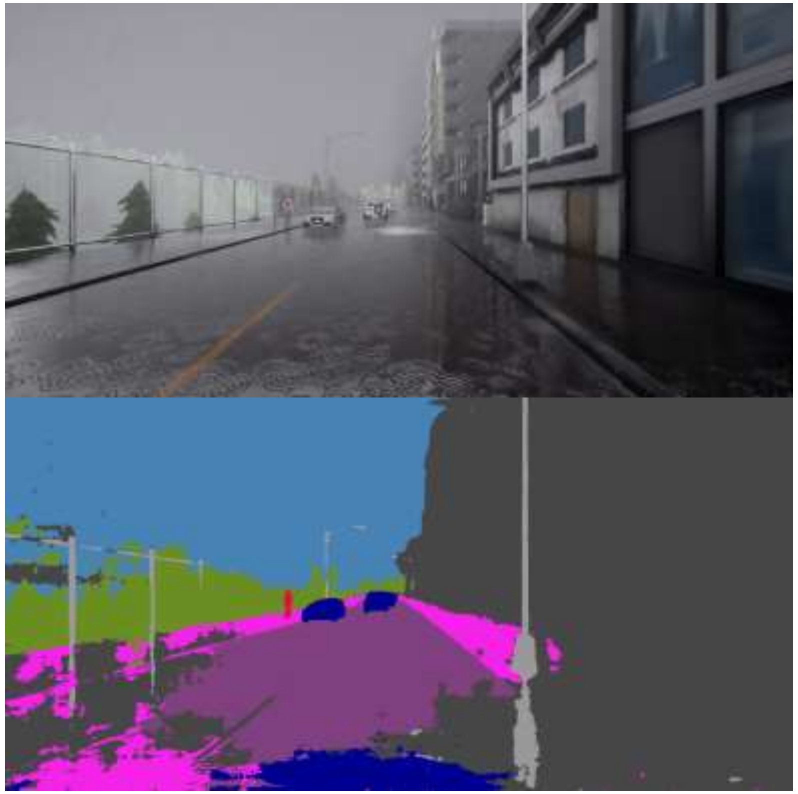
Background

- **Closed-set VS Open World Environment**

Model trained on normal Closed-set datasets suffer significant performance drop in open world condition.



Clear-sunny



Rainy

Background

- Anomaly (Corner Case) : no clear and consistent definition
  - unknown, unseen, unusual

Breitenstein et al. define different corner case levels based on **detection complexity**:

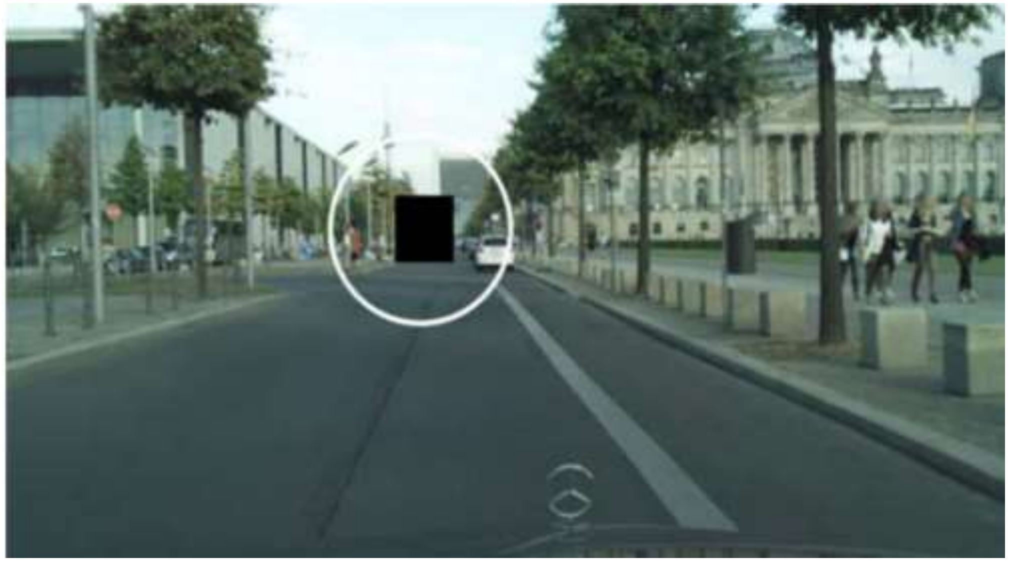
- Pixel Level
- Domain Level
- Object Level
- Scene Level
- Scenario Level

Breitenstein, J.; Term¨ohlen, J.-A.; Lipinski, D.; and Fingscheidt, T. 2020. Systematization of corner cases for visual perception in automated driving. In 2020 IEEE Intelligent Vehicles Symposium (IV), 1257–1264. IEEE.

Background

- Pixel Level

A **global outlier** on pixel level:
Overexposure image



A **local outlier** on pixel level:
Pixel errors (dead pixels)

Background

- Domain Level

Domain shifts: a large, constant shift happens in the appearance.
For example, different weather conditions such as **snow, rain or fog**, different lighting conditions such as **daytime, dusk, night**

Example domain shifts: snow weather

Background

- Object Level

Unknown objects : instances that have not been seen during training

A carriage on the road

Background

- Scene Level

non-conformant with the expected patterns, require model not to only detect objects, but also **understand** the entire scene.

Collective Anomaly : Multiple known objects, but in an unseen quantity,e.g. Traffic jam

Contextual Anomaly : A known object, but in an unusual location，e.g. Person lying on the road

Background

- Scenario Level

  Scenario-level corner cases denote the observation of patterns with **temporal context**.
  Risky or Novel scenarios that contains potential for collision

A person is walking unexpectedly onto the street from behind a car.

Datasets

Fishyscapes

CAOS

SegmentMeIfYouCan

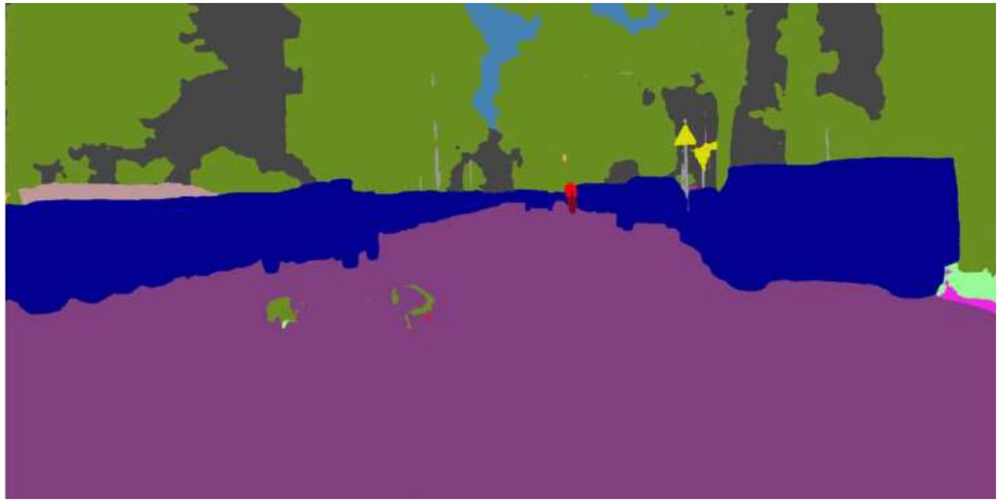
CODA

Datasets

- **Fishyscapes**

Fishyscapes is proposed for **anomaly segmentation** for urban driving, which evaluates pixel-wise uncertainty estimates towards the **detection of anomalous objects** in front of the vehicle.

Input



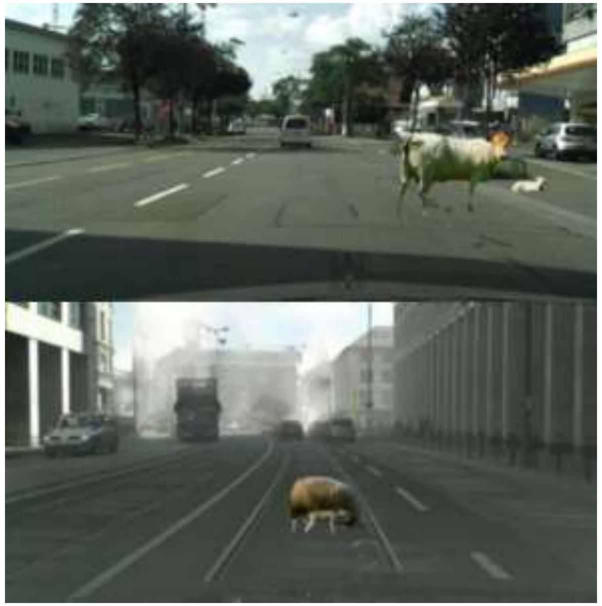
DeepLabv3+ Prediction

H. Blum et al., "Fishyscapes: A Benchmark for Safe Semantic Segmentation in Autonomous Driving," in International Conference on Computer Vision (ICCV) Workshop, 2019.

Datasets

- **Fishyscapes**

FS Static： Images from **Cityscapes** (Cordts et al. 2016) are overlayed with objects from **Pascal VOC** (Everingham et al. 2010) dataset;



overlay objets from classes that cannot be found in Cityscapes: *aeroplane, bird, boat, bottle, cat, chair, cow, dog, horse, sheep, sofa, tvmonitor.*

Datasets

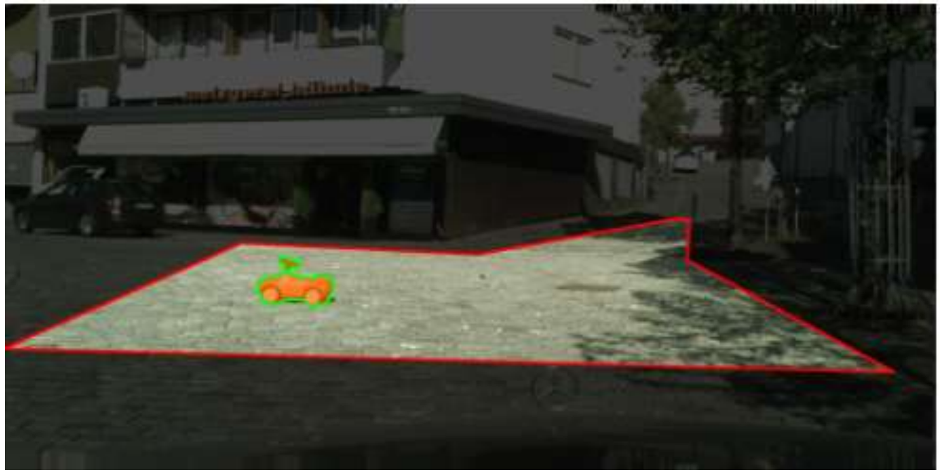
- **Fishyscapes**

FS Web : similarly to FS Static with overlay objects crawled from the **internet**

Datasets

- **Fishyscapes**

FS Lost & Found： based on the original **Lost & Found** dataset, original dataset only included annotations for the **anomalous objects** and a **coarse annotation of the road**.



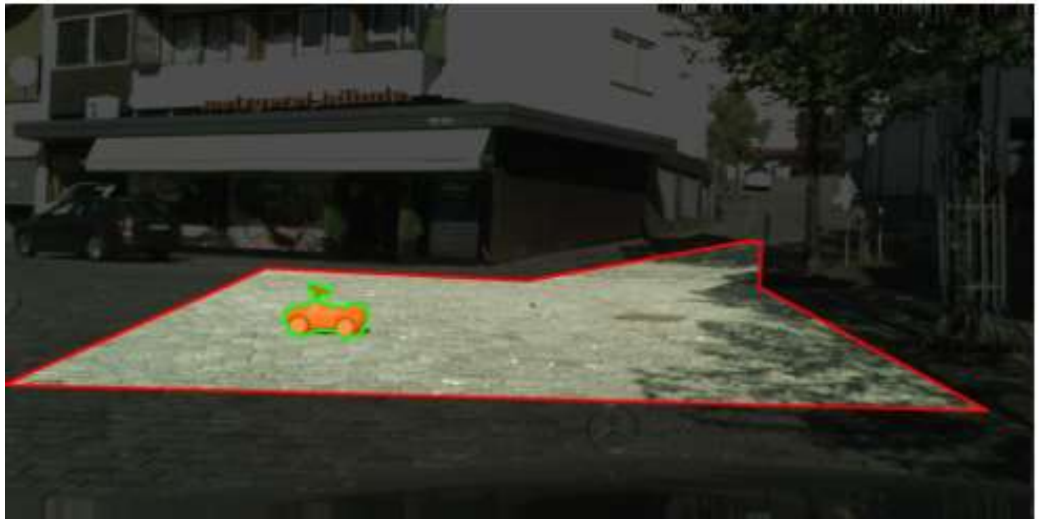
Original L&F

P. Pinggera, S. Ramos, S. Gehrig, U. Franke, C. Rother, and R. Mester, "Lost and found: Detecting small road hazards for self-driving vehicles," in 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2016.

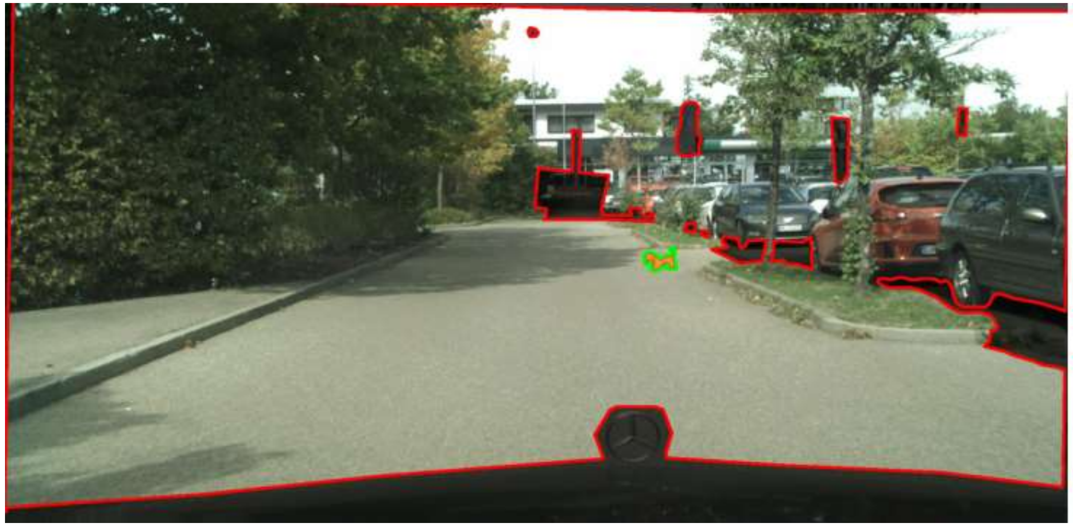
Datasets

- **Fishyscapes**

FS Lost & Found: add pixel-wise annotations that distinguish between **objects** (the anomalies), **background** (classes contained in Cityscapes) and **void** (anything not contained in Cityscapes classes).



Original L&F



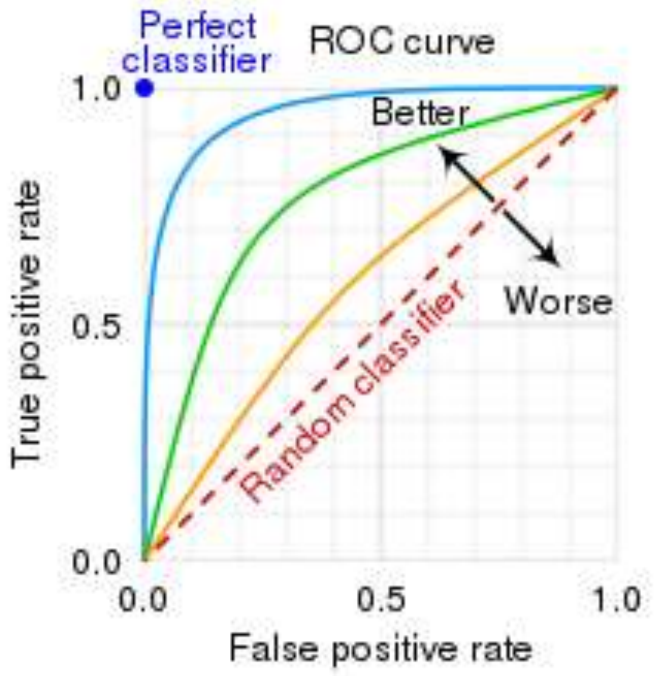
FS Lost & Found

Datasets

- **Fishyscapes**

Metrics : metrics associated with a **binary classification task, anomaly / not anomaly**

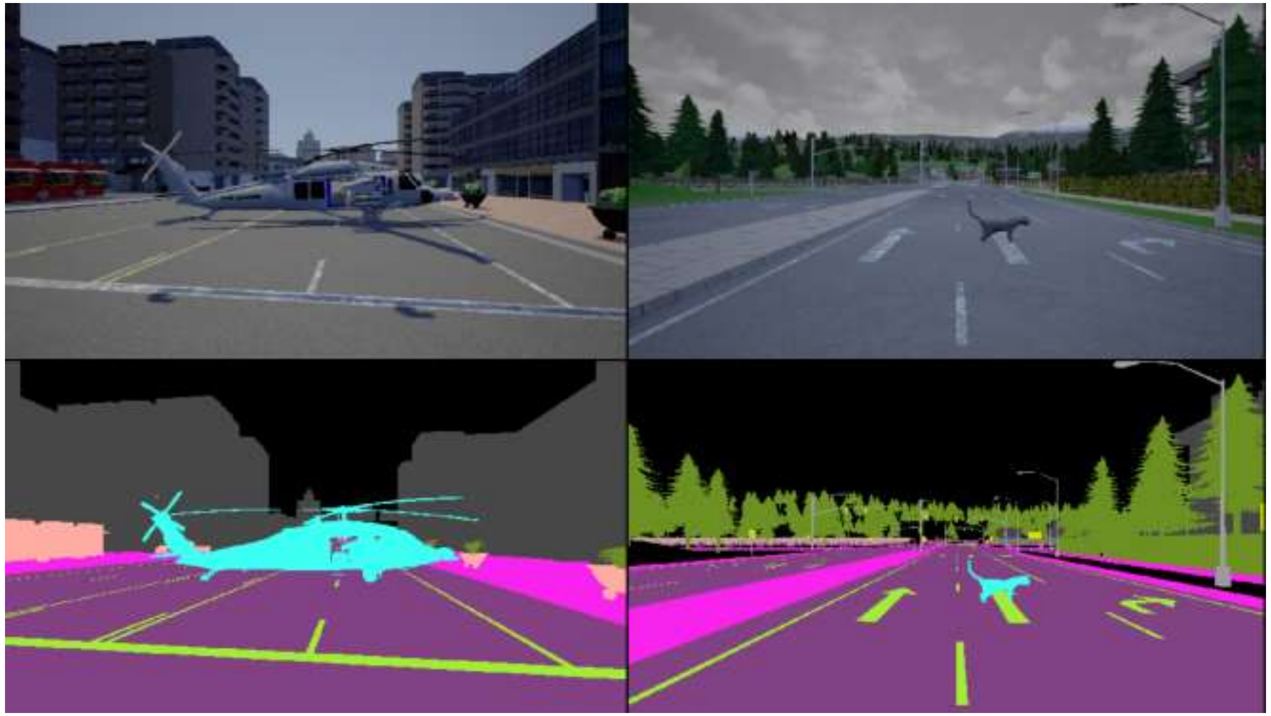
- $AP\uparrow$
- $FPR_{95}\downarrow$ **:false positive rate (FPR) at 95% true positive rate (TPR)**



Datasets

- **CAOS (Combined Anomalous Object Segmentation benchmark)**

StreetHazards : anomaly **segmentation** dataset that leverages simulation to provide anomalous objects. use the Unreal Engine along with the **CARLA simulation environment**



avoids the issues of inconsistent chromatic aberration, inconsistent lighting, edge effects,

Insert in any location, change lighting and weather

Datasets

- **CAOS (Combined Anomalous Object Segmentation benchmark)**

StreetHazards : 12 classes used for training:, thirteenth class is **the anomaly class**

Anomalies are taken from the Digimation Model Bank Library and semantic ShapeNet (ShapeNet-Sem) (Savva et al., 2015)

Datasets

- **CAOS**

BDD-Anomaly : real images, choose *motorcycle, train*, and *bicycle* as the anomalous object classes and remove all images with these objects from the training and validation sets.

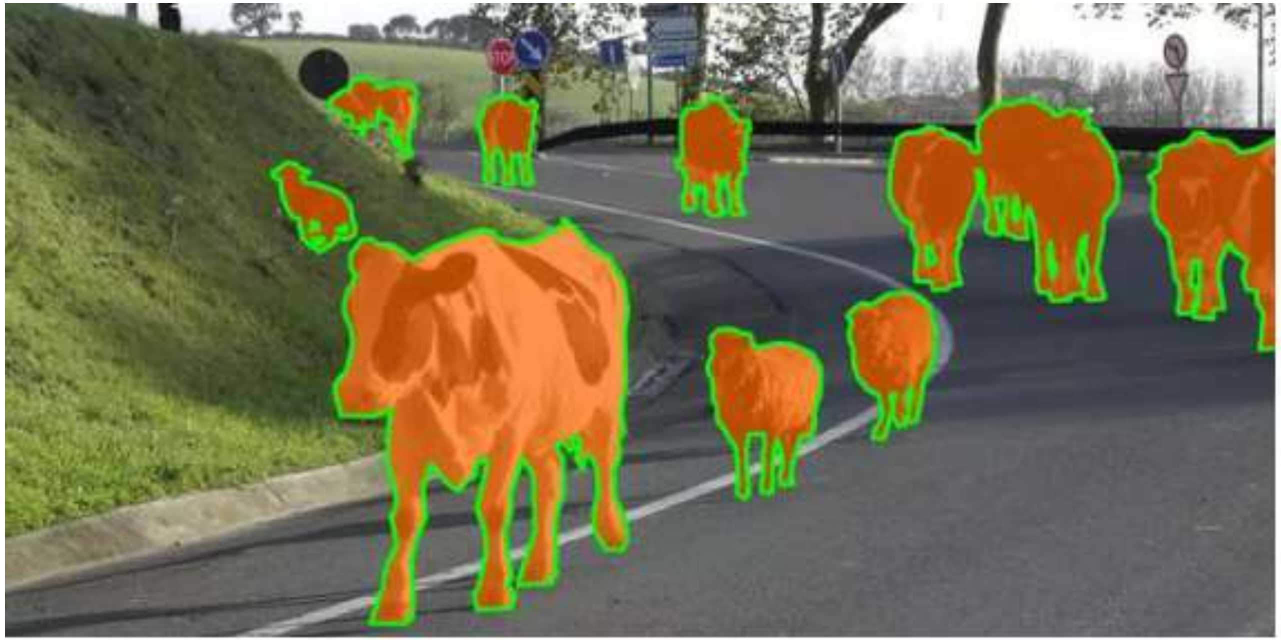
Metric:

- $\boldsymbol{FPR95} \downarrow$
- **AUROC↑**
- **AUPR↑**

Datasets

- **SegmentMeIfYouCan**

RoadAnomaly21: real images, any object **that strictly cannot be seen in the Cityscapes data** as anomalous,



Datasets

- **SegmentMeIfYouCan**

RoadObstacle21: the task of obstacle segmentation, whose goal is to identify all objects on the road, may they be from **known classes or from unknown ones**. focuses only on the road as region of interest.

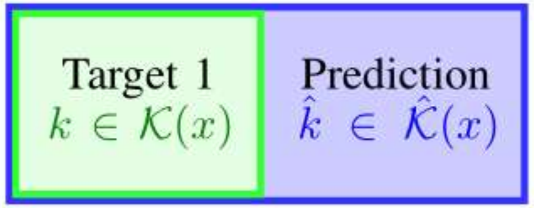
Datasets

- **SegmentMeIfYouCan**

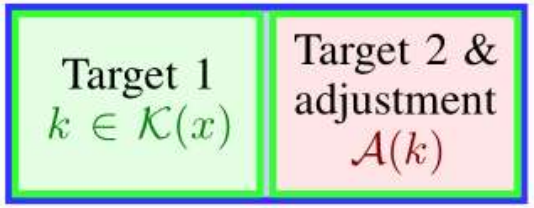
Metric: model with a binary classifier providing anomaly scores, discriminates between the two classes **anomaly and non-anomaly**

- ■ Pixel level **: AUPRC, $\boldsymbol{FPR95}$ ↓**
- ■ Component level **: sIoU**

$$\mathrm{sIoU}(k) := \frac{|k \cap \hat{K}(k)|}{|(k \cup \hat{K}(k)) \setminus \mathcal{A}(k)|} \quad \text{with} \quad \hat{K}(k) = \bigcup_{\hat{k} \in \hat{\mathcal{K}}, \hat{k} \cap k \neq \emptyset} \hat{k} \qquad \mathcal{A}(k) = \{z \in k' : k' \in \mathcal{K} \setminus \{k\}\}$$



$\mathrm{IoU}(k) = 0.50$
$\mathrm{sIoU}(k) = 0.50$

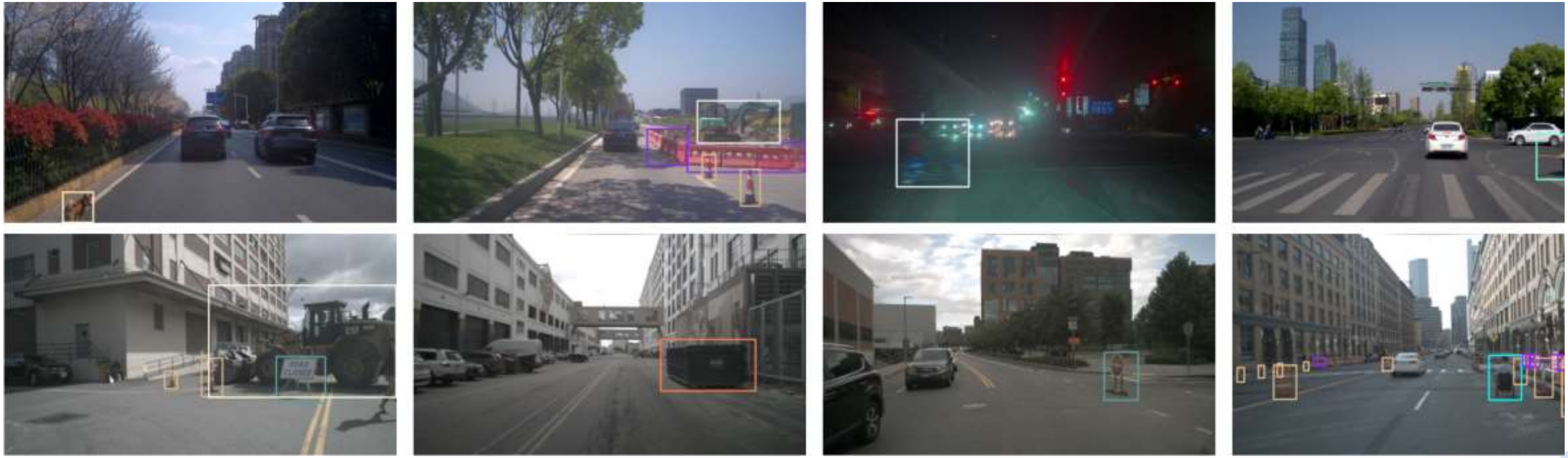


$\mathrm{IoU}(k) = 0.50$
$\mathrm{sIoU}(k) = 0.99$

Datasets

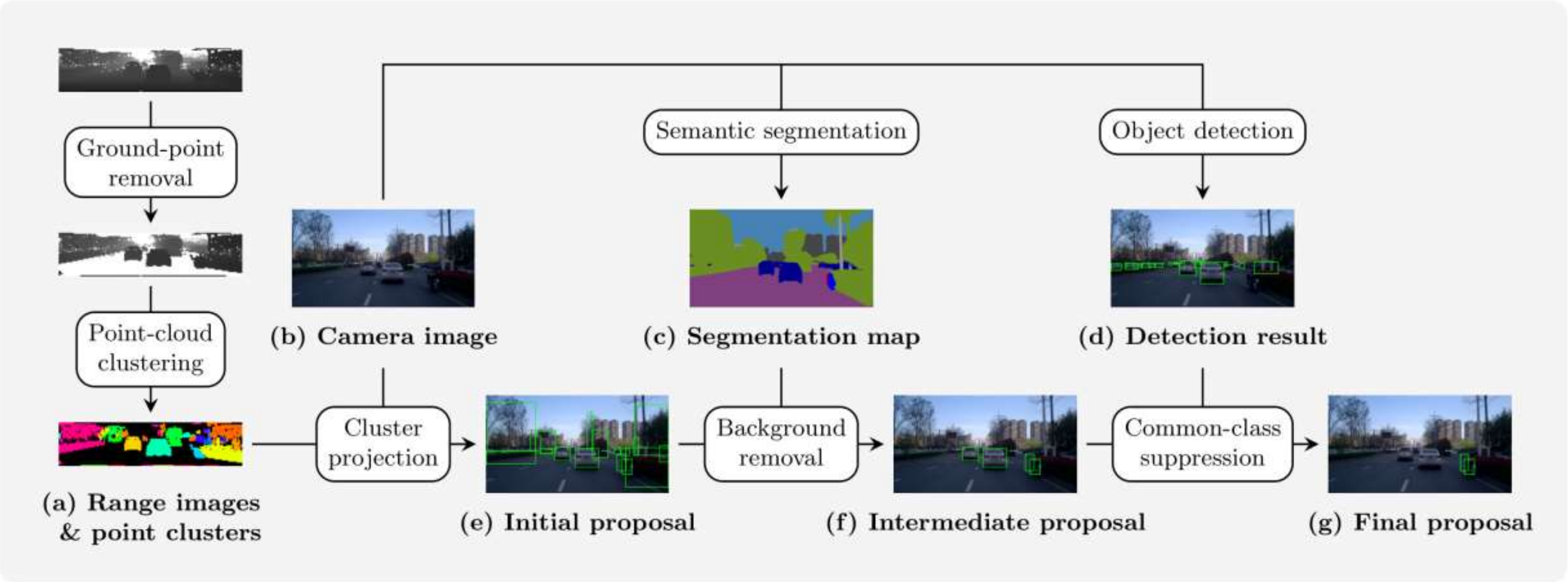
- **CODA**

CODA is a novel dataset of **object-level corner cases** designed for **object detection.** CODA is constructed from 3 major object detection benchmarks for autonomous driving—**KITTI, nuScenes, ONCE**



Datasets

- **CODA**

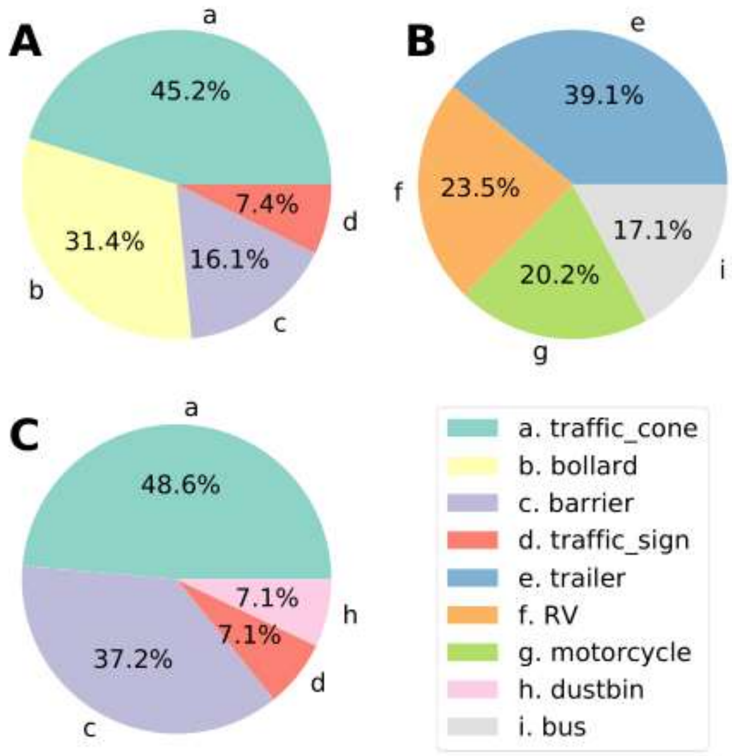


Pipeline for generating proposals of corner case (COPG)

Datasets

- **CODA**

corner cases can be generally grouped into 7 super-classes: *vehicle, pedestrian, cyclist, animal, traffic facility, obstruction, and misc.* these classes can be divided into **novel classes** and **common classes**

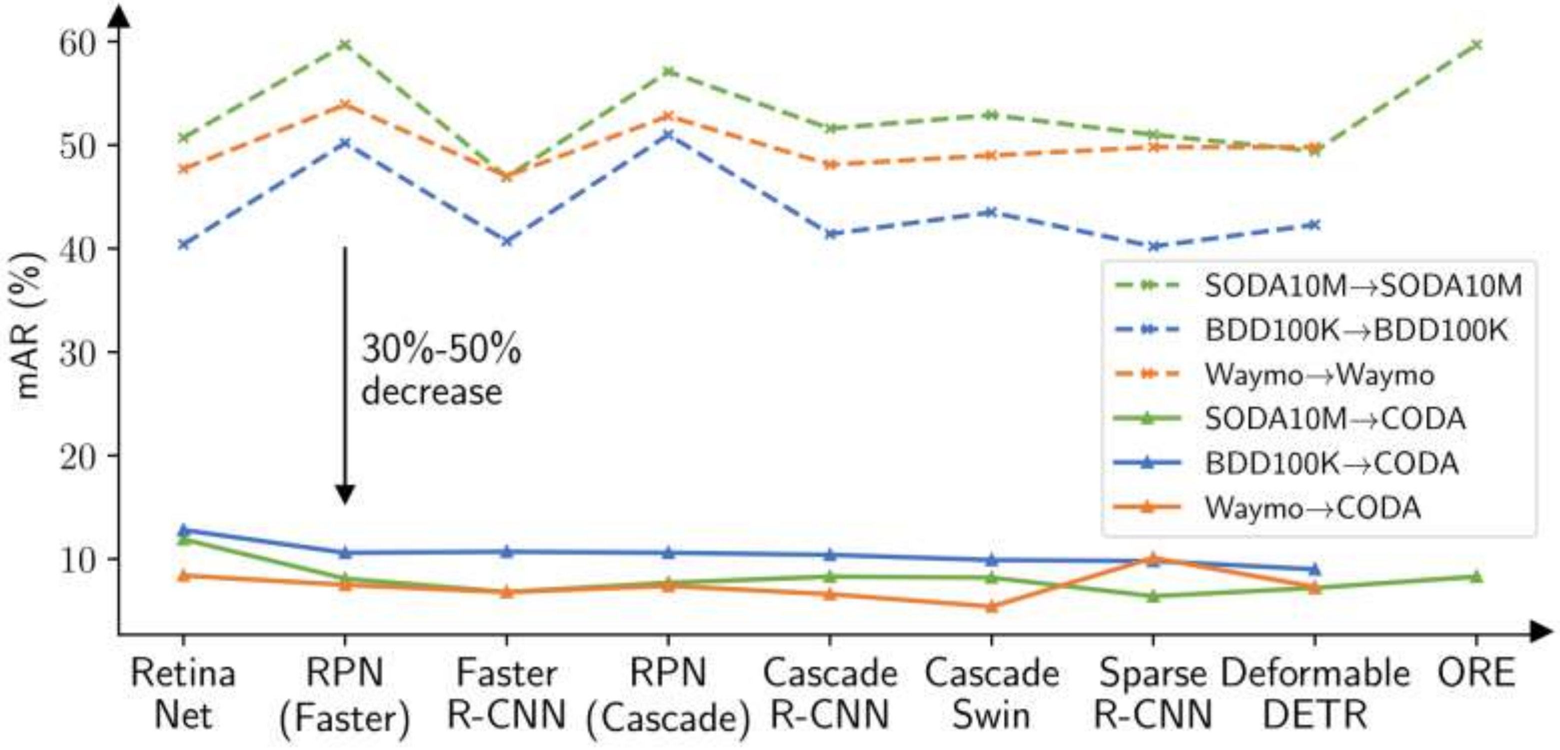


Distribution of the top-4 classes in the three domains of CODA: **A** ONCE, **B** KITTI, and **C** nuScenes.

Traffic facilities such as **traffic cone** and **barrier** take up a majority of the corner cases

Datasets

- **CODA**



detectors suffer from a significant **30%-50%** performance drop，A→B represents that the detector is **trained on** dataset A and **evaluated on** dataset B

Methods

Open World Object Detection : ORE

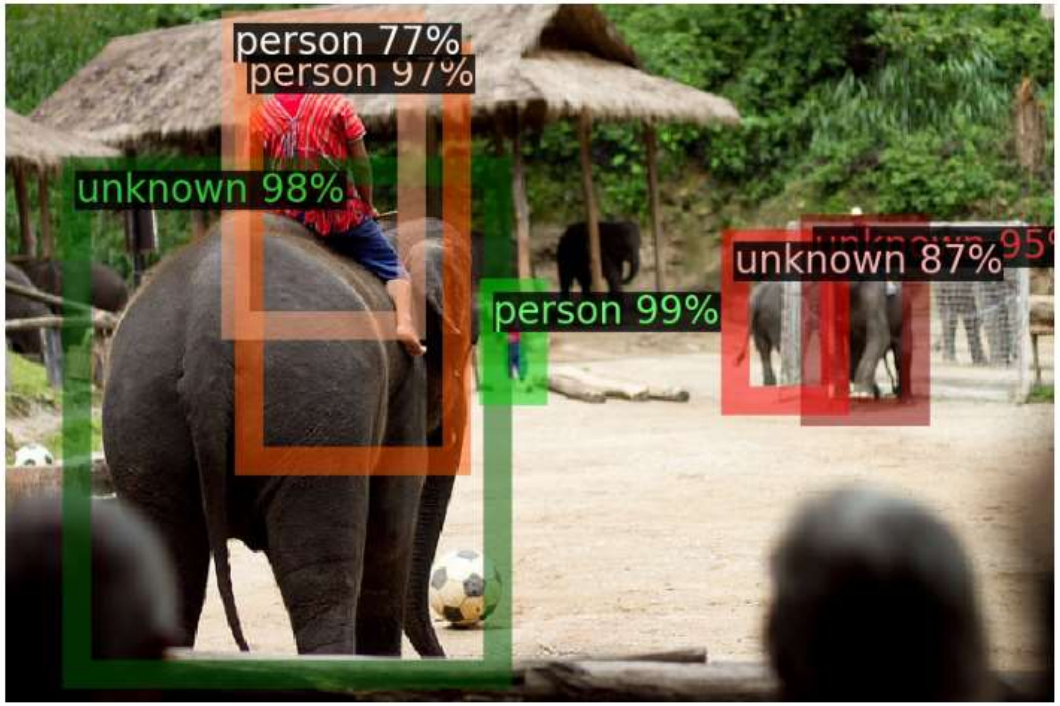
Anomaly Segmentation : SML

Misbehavior detection : SelfOracle

Methods

- ORE (Open World Object Detector)

  Identify objects that have not been introduced to it as **‘unknown’,** without explicit supervision to do so



K J Joseph, Salman Khan, Fahad Shahbaz Khan, and Vineeth N Balasubramanian.Towards Open World Object Detection. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.

Methods

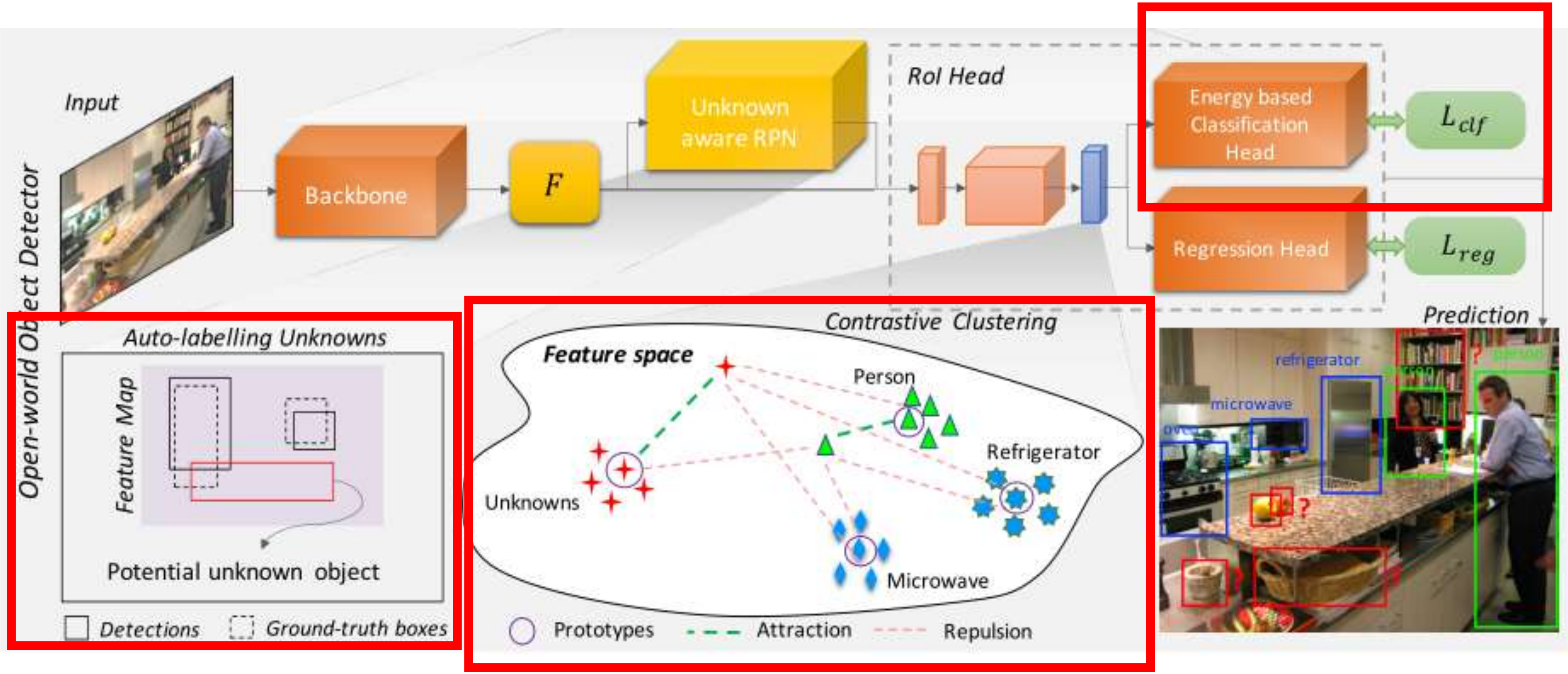
- ORE (Open World Object Detector)

Recognizing an unknown as an unknown requires strong generalization.

Difficulty : The object detector is trained to detect unknown objects as **background.**

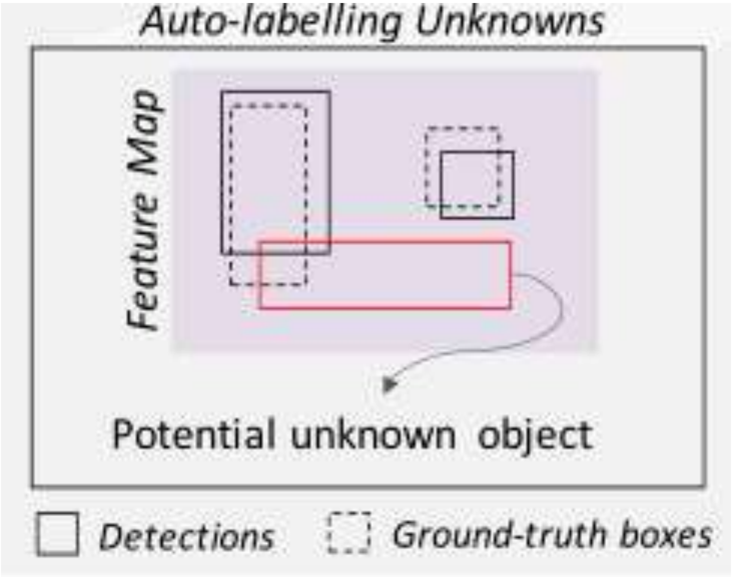
Methods

High-level architectural overview of ORE,choose **Faster-RCNN** as the base detector



Methods

Auto-labelling mechanism based on the Region Proposal Network

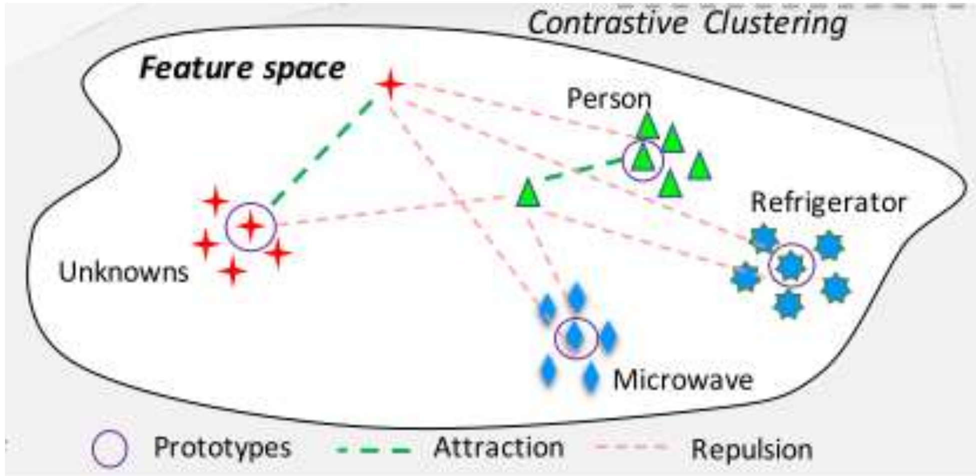


have high **objectness score**, but do **not overlap with a ground-truth object** as a potential unknown object

Simply put, select the **top-k background region proposals**, sorted by its objectness scores, as unknown objects.

Methods

Contrastive Clustering : Separate Class in latent space.



Instances of same class would be forced to **remain close-by,** while instances of dissimilar class would be **pushed far apart**

contrastive loss:

$$\mathcal{L}_{cont}(\boldsymbol{f}_c) = \sum_{i=0}^{C} \ell(\boldsymbol{f}_c, \boldsymbol{p}_i), \text{ where,} \tag{1}$$

$$\ell(\boldsymbol{f}_c, \boldsymbol{p}_i) = \begin{cases} \mathcal{D}(\boldsymbol{f}_c, \boldsymbol{p}_i) & i = c \\ \max\{0, \Delta - \mathcal{D}(\boldsymbol{f}_c, \boldsymbol{p}_i)\} & \text{otherwise} \end{cases}$$

p_i : prototype vector for each class i

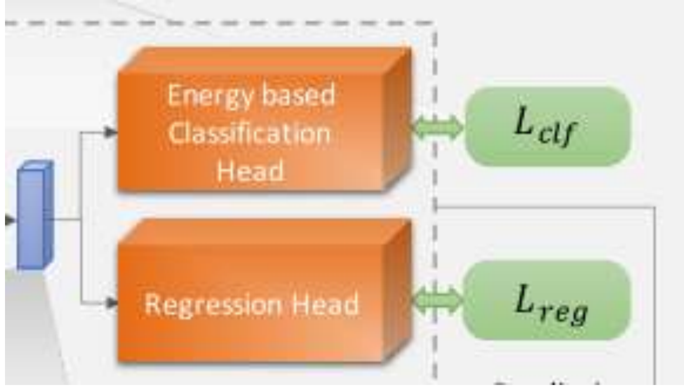
f_c : Feature vector

D : distance function

Δ : defines how close a similar and dissimilar item can be

Methods

Energy Based Unknown Identifier



Free energy in terms of their logits

$$E(\boldsymbol{f}; g) = -T \log \sum_{i=1}^{C} \exp\left(\frac{g_i(\boldsymbol{f})}{T}\right).$$

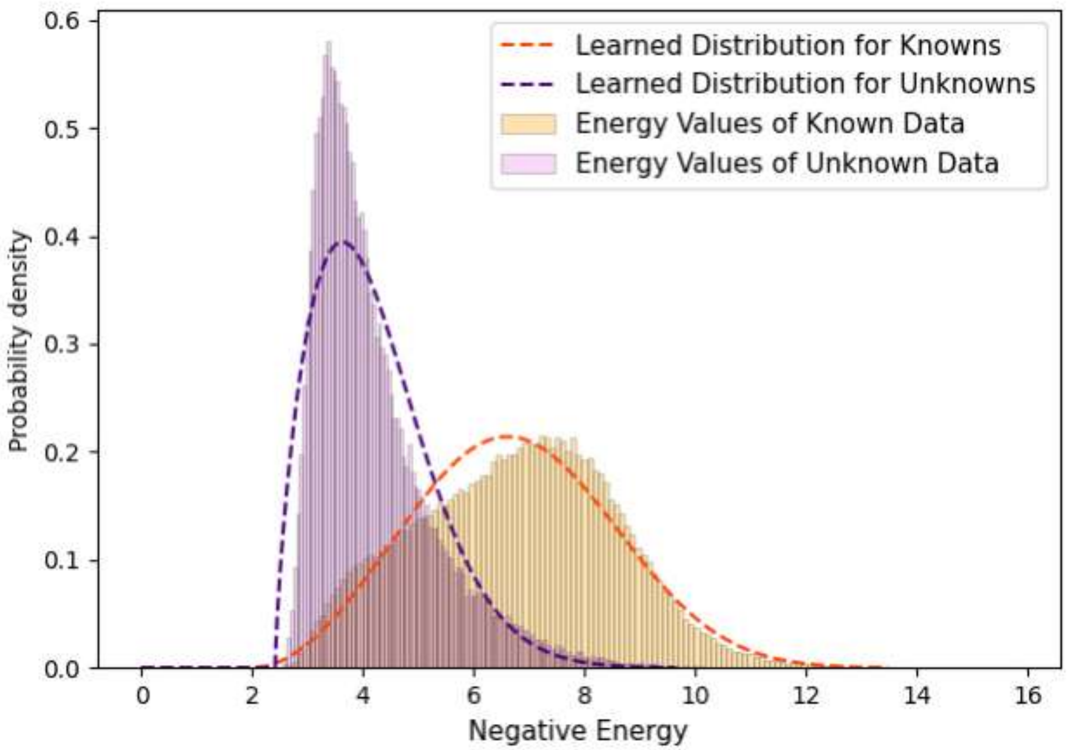
$\boldsymbol{f}$: feature

$\boldsymbol{g_i}(\boldsymbol{f})$: the i^{th} classification logit of classification head g(.)

$\boldsymbol{T}$: temperature parameter

Methods

Energy Based Unknown Identifier



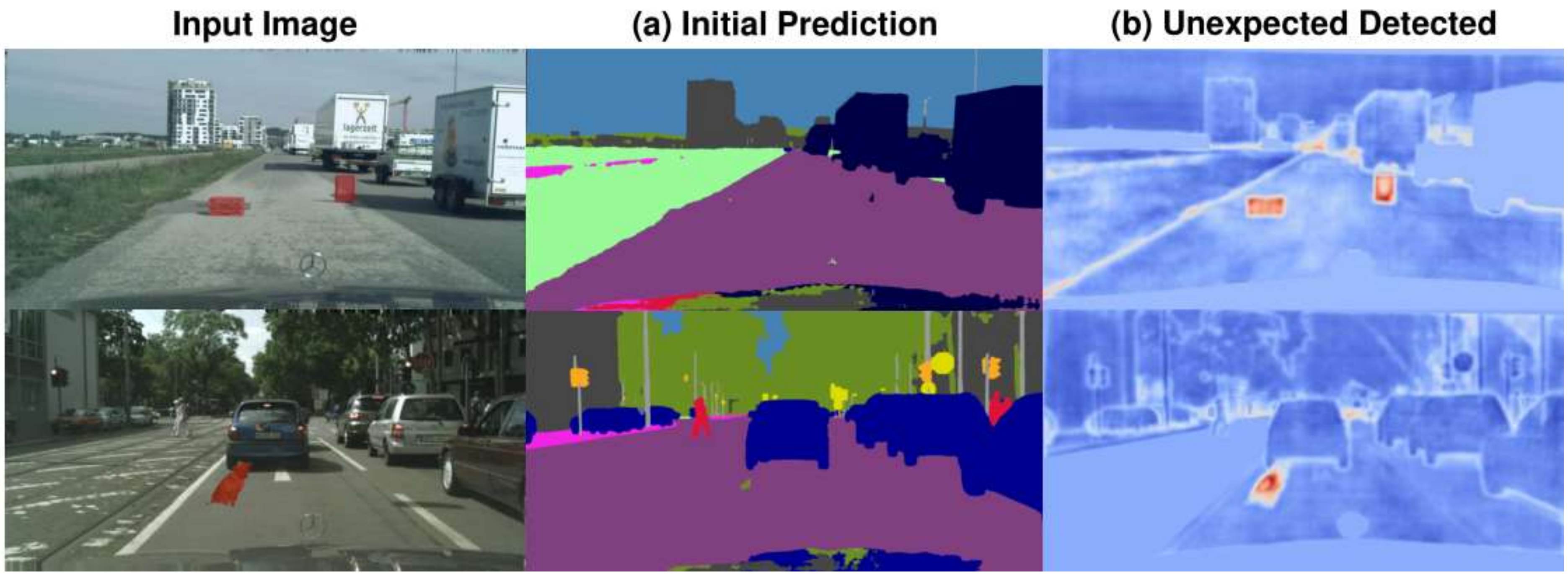
The energy values of the known and unknown datapoints exhibit clear separation

fit a **Weibull distribution** on each of them and use these for identifying unseen known and unknown samples

Methods

- SML (Standardized Max Logits)

  Identify unexpected objects on roads in semantic segmentation

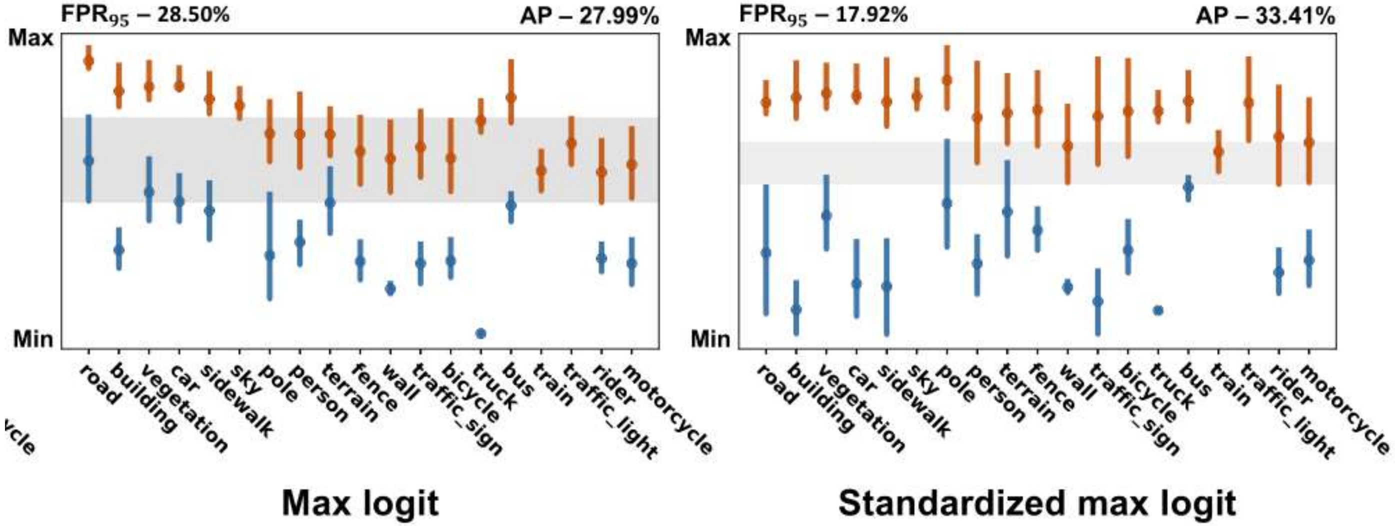


Sanghun Jung, Jungsoo Lee, Daehoon Gwak, Sungha Choi, and Jaegul Choo. Standardized max logits: A simple yet effective approach for identifying unexpected road obstacles in urban-scene segmentation. In IEEE/CVF International Conference on Computer Vision, 2021

Methods

- SML (Standardized Max Logits)

The distribution of max logits of each predicted class is significantly different from each other



Red and blue represent the distributions of values in in-distribution pixels and unexpected pixels, respectively

Methods

- SML (Standardized Max Logits)

1. Obtain the mean μ_c and variance σ_c^2 of class c from the training samples

$$\mu_c = \frac{\sum_i \sum_{h,w} \mathbb{1}(\hat{\boldsymbol{Y}}^{(i)}_{h,w} = c) \cdot \boldsymbol{L}^{(i)}_{h,w}}{\sum_i \sum_{h,w} \mathbb{1}(\hat{\boldsymbol{Y}}^{(i)}_{h,w} = c)} \quad (3)$$

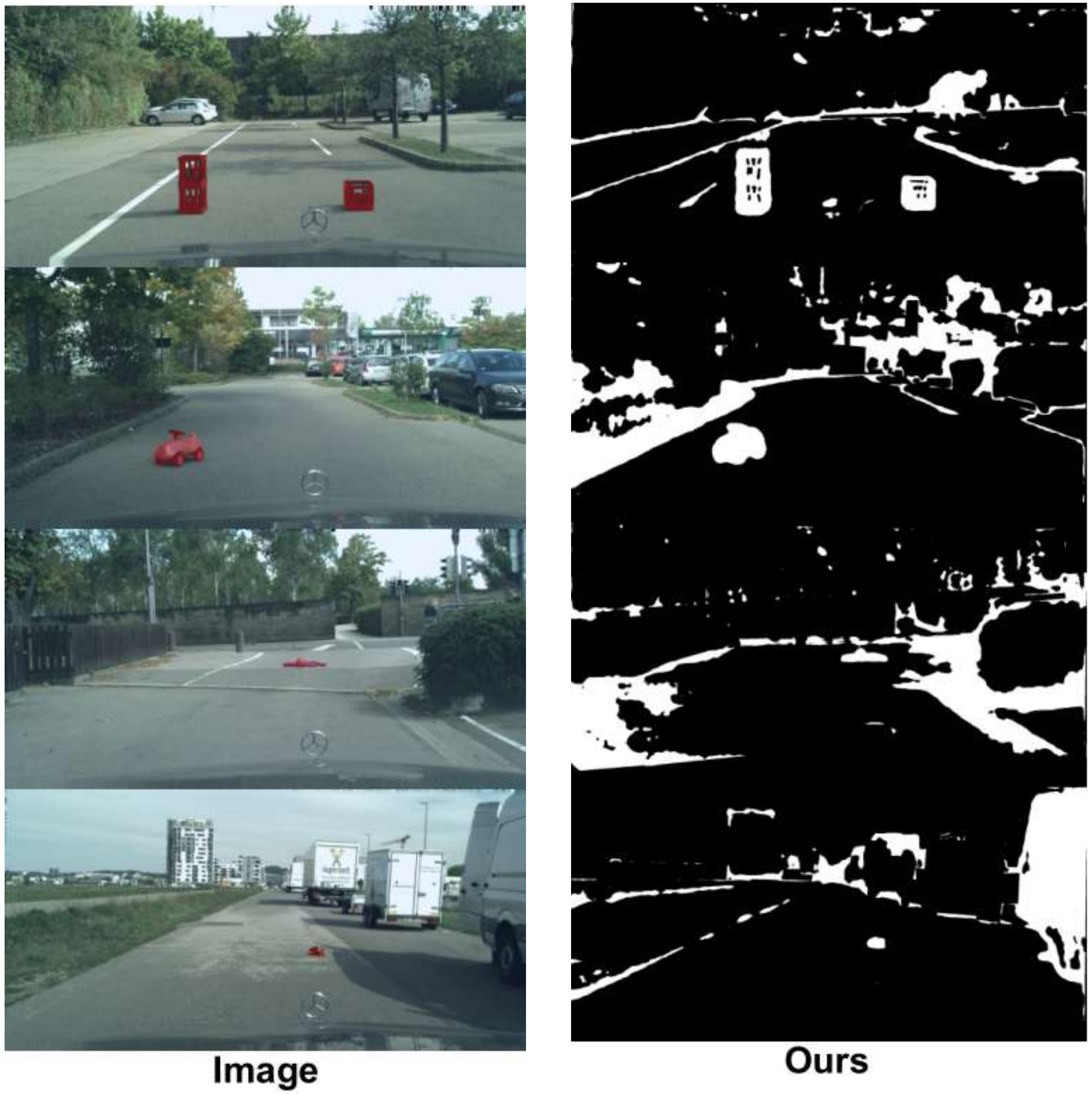
$$\sigma_c^2 = \frac{\sum_i \sum_{h,w} \mathbb{1}(\hat{\boldsymbol{Y}}^{(i)}_{h,w} = c) \cdot (\boldsymbol{L}^{(i)}_{h,w} - \mu_c)^2}{\sum_i \sum_{h,w} \mathbb{1}(\hat{\boldsymbol{Y}}^{(i)}_{h,w} = c)}, \quad (4)$$

i : i-th training example

$\hat{Y}$: prediction label

L : max logit score

Methods

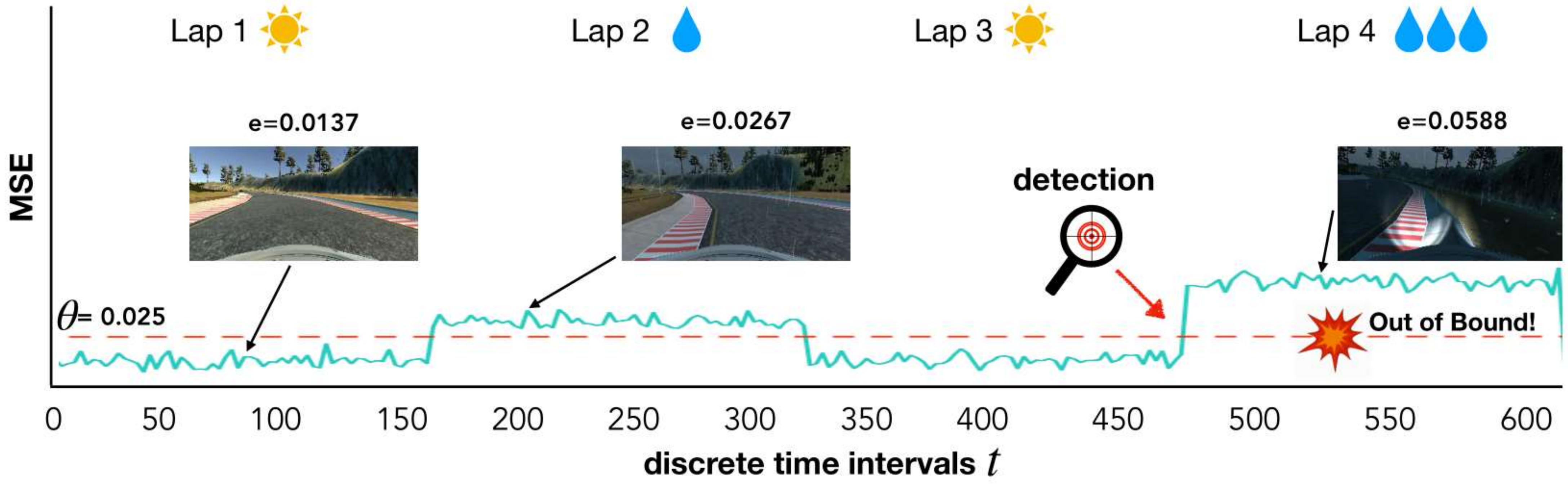


Achieve SOTA performance on
Fishyscapes Lost&Found

Methods

- SelfOracle

  Monitors the DNN confidence at **runtime**, to predict unsupported driving scenarios in advance.



Methods

- SelfOracle

Reconstruct & Probability Distribution Fitting

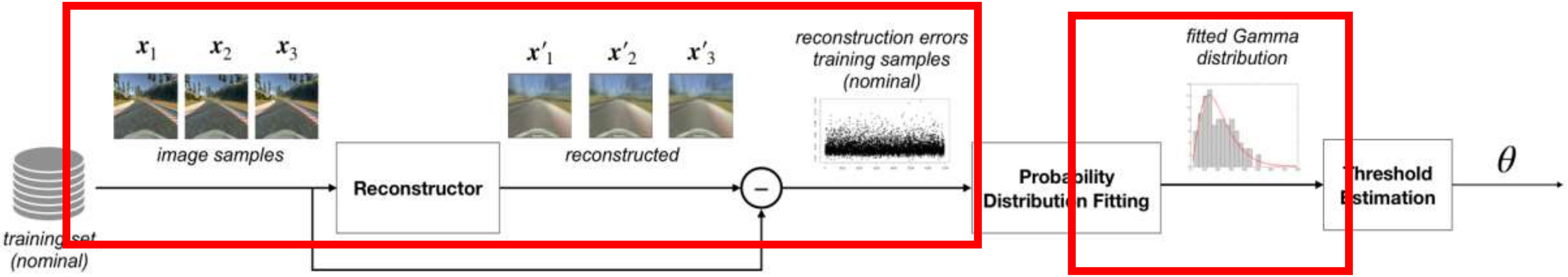
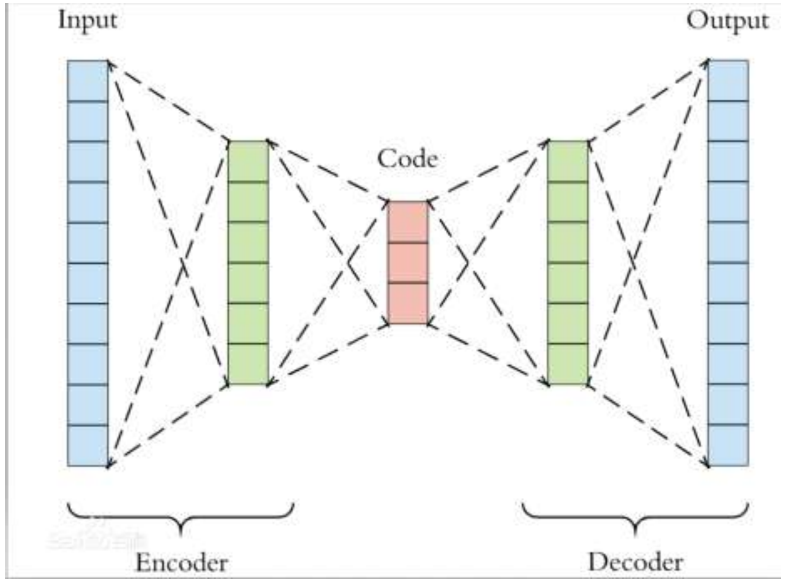


Figure 2: Model Training under Nominal Driving Behaviour.

Methods

- SelfOracle

Reconstructor: Autoencoder



x : input
$f(.)$: encoder
$g(.)$: decoder

Autoencoder minimises a loss function

$$L(x, g(f(x)))$$

which measures the distance between the original data and reconstruction

Methods

- SelfOracle

Reconstruction error

$$d(x, x') = \frac{1}{WHC} \sum_{i=1, j=1, c=1}^{W,H,C} (x[c][i,j] - x'[c][i,j])^2$$

Images have width W , height H and C channels, assume that the pixel-wise error

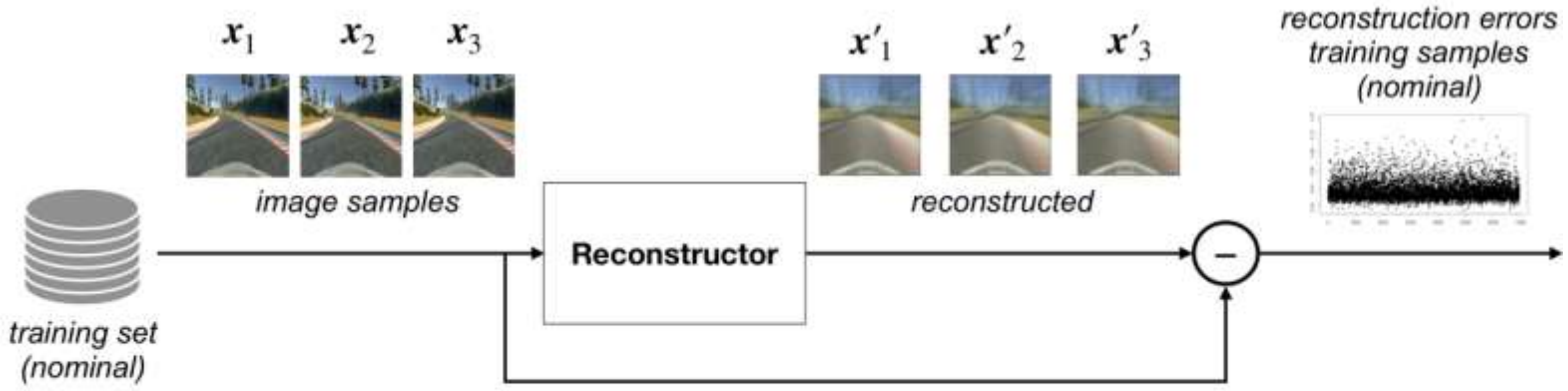
$$e[c][i,j] = x[c][i,j] - x'[c][i,j] \qquad e[c][i,j] \sim \mathcal{N}(0, \sigma_{c,i,j}).$$

Then

$$d(x, x') \sim \Gamma(\alpha, \beta).$$

Methods

- SelfOracle

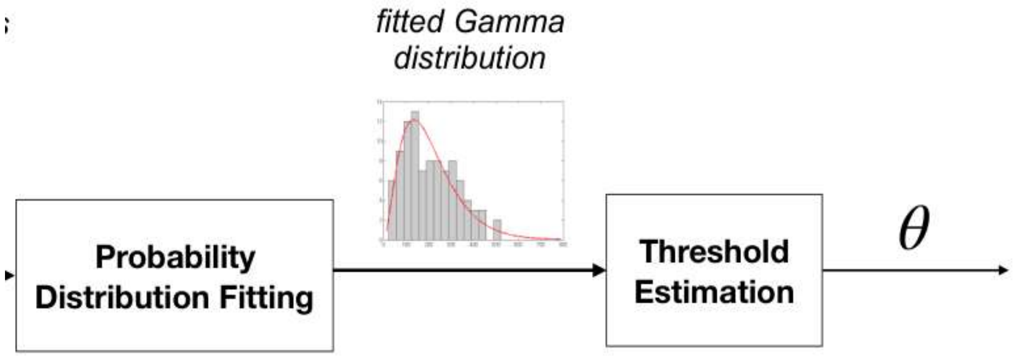


Training Set: user manually controls the car in Udacity simulator while the simulator records the actions

Reconstructor: Autoencoder

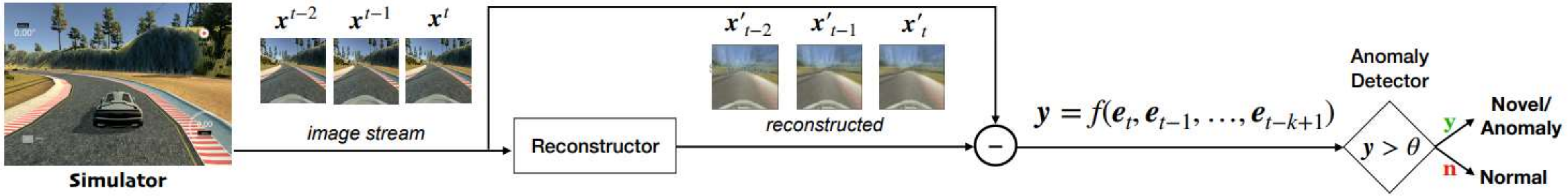
Methods

- SelfOracle



$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \qquad x > 0;\ \alpha, \beta > 0$$

Fit the data (reconstruction errors) by maximum likelihood estimation (MLE).



Ours Works

Sythetic datasets and scenarios for Autonomous Driving testing

Our Works

- Scenario Description language for Road network

Road network data

Property graph Model

Scenario Description

《嵌入路网图模型的自动驾驶场景描述语言》 软件学报，CCF A 中文

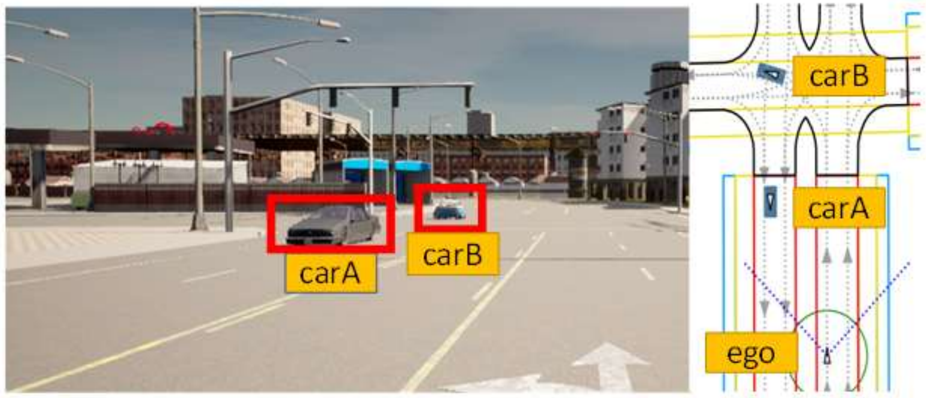
Our Works

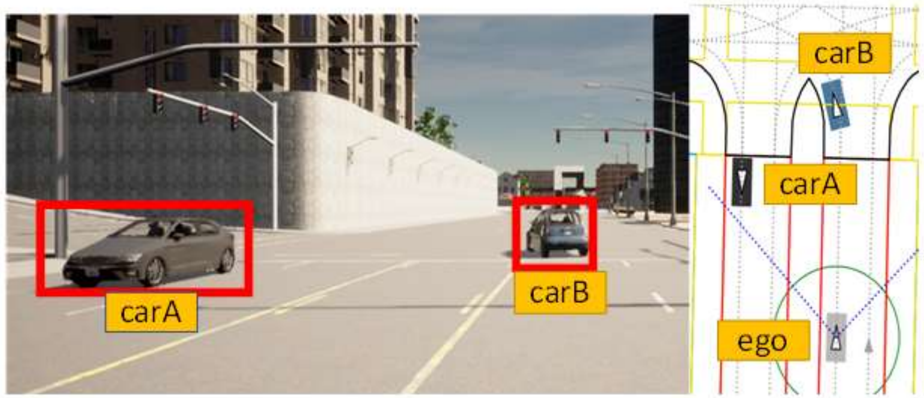
- Scenario Description language for Road network

```
r1 : Road, is2Way = True
g1 : Group , laneNum = 2
g2 : Group , laneNum = 2
g1.road = r1
g2.road  = r1
g1.opposite = g2
```

```
ego_lane : Lane, index = 1
l2 : Lane, index = 2
ego_lane.group = g1
l2.group = g2
```

```
j1 : Junction, is4Way = True
l3 : Lane, turn = LEFT
l3.junction = j1
ego_lane.succ = l3
```





《嵌入路网图模型的自动驾驶场景描述语言》 软件学报，CCF A 中文

Q&A