Technical summary: Road Obstacle Detection based on Unknown Objectness Scores

Paper by Chihiro Noguchi, Toshiaki Ohgushi, Masao Yamanaka
Toyota Motor Corporation

Summary by J. Edgar Hernandez — May, 2025

Summary

This paper proposes a method for detecting **unknown road obstacles** in driving scenes **using a semantic segmentation network** with a **sigmoid head**.

The method introduces the *Unknown Objectness Score (UOS)*, which **combines pixel-wise anomaly scores with an objectness score to improve accuracy**.

Context

A **Semantic segmentation network** performs pixel-level classification of an image. Assigns a class label to every pixel in input image out of the known classes.

These networks map an image to class scores or probabilities:

$$f_\theta : \mathbb{R}^{3\times H\times W} \rightarrow \mathbb{R}^{K\times H\times W}$$

Where K is the number of semantic classes.

Output interpretation

Each pixel pair (i, j) in the output has a vector of size K, which depending on the chosen strategy, can contain:

- Softmax: A probability distribution over the possible classes:

$$P(y = k|X) \in [0, 1] \qquad \sum_{k=1}^{K} P(y = k|X) = 1$$

- Sigmoid: Allows multi-label assignment. This is the one used for this paper.

Key Contributions

- Introduction of the **Unknown Objectness Score (UOS)**:

  Objectness score p_i^o

  Predefined object classes, such as cars, traffic signals, etc. are merged into a common 'object' class. This will be included in the label of the samples, and the network will be able to predict the probability of pixels of a sample of being part of the 'object' class, regardless of it being one of the known pre-defined classes that originally build the object class.

  Unknown score

  It is the product of the probabilities of a pixel belonging to the known object classes

  $$S_i = \prod_{k=1}^{K}(1 - p_{ik})$$

  where p_{ik} is the predicted probability of class k at pixel i.

  The Unknown Objectness Score (UOS)

  Putting these equations together, we obtain:

  $$S_i = p_i^o \cdot \prod_{k=1}^{K}(1 - p_{ik})$$

  where p_i^o is the **objectness score** of pixel i.

- Use of a **sigmoid head** instead of softmax, allowing multi-label classification per pixel and an extra output for objectness. So we care about two probabilities, the probability of a pixel belonging to the known classes, as well as to a 'joint' object class.

- A **boundary-aware binary cross-entropy loss** function:

  $$\mathcal{L}_n = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k \in C} f(y_{ik}, p_{ik}) - \lambda \cdot \frac{1}{\sum_i \delta_i}\sum_{i=1}^{N}\delta_i \sum_{k \in C} f(y_{ik}, p_{ik})$$

  with $f(y, p) = y \log p + (1 - y)\log(1 - p)$, and δ_i is a boundary indicator.

Model Architecture

- **Backbone:** ResNet-50. This is a pre-trained CNN that allows for the extraction of low-resolution features (edges, textures, shapes, etc.).

- **Segmentation head:** DeepLabv3+ (modified for custom sigmoid activation). This is a **semantic segmentation model** designed to produce pixel-level segmentation capturing local and global details.

Datasets Used

- **LostAndFound**: Real-world road obstacles.
- **Fishyscapes Validation**: Synthetic unknown objects overlaid on Cityscapes.
- **Road Anomaly**: Diverse scenes with rare obstacles like animals or debris.

Results

- The proposed method outperforms state-of-the-art baselines both with and without OoD supervision.
- It excels on metrics such as AUROC, AP, and FPR95, particularly in reducing background false positives.
- Computationally efficient: 15 ms per image on GPU (faster than SynBoost and autoencoder methods).

Limitations

- Detection performance depends on the quality of objectness scores.
- If unknown obstacles do not appear object-like, they may go undetected.

References

1. Noguchi, C., Ohgushi, T., & Yamanaka, M. (2024). Road Obstacle Detection based on Unknown Objectness Scores. arXiv:2403.18207 [cs.CV].