ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Out-of-Distribution Segmentation via Wasserstein-Based Evidential Uncertainty

Arnold Brosch[1], Abdelrahman Eldesokey[2], Michael Felsberg[3] and Kira Maag[1]

1- Heinrich-Heine-University Düsseldorf, Germany, {arnold.brosch,kira.maag}@hhu.de

2- KAUST, Saudi Arabia, abdelrahman.eldesokey@kaust.edu.sa

3- Linköping University, Sweden, michael.felsberg@liu.se

Abstract. Deep neural networks achieve superior performance in semantic segmentation, but are limited to a predefined set of classes, which leads to failures when they encounter unknown objects in open-world scenarios. Recognizing and segmenting these out-of-distribution (OOD) objects is crucial for safety-critical applications such as automated driving. In this work, we present an evidence segmentation framework using a Wasserstein loss, which captures distributional distances while respecting the probability simplex geometry. Combined with Kullback-Leibler regularization and Dice structural consistency terms, our approach leads to improved OOD segmentation performance compared to uncertainty-based approaches.

1 Introduction

Deep neural networks (DNNs) have achieved remarkable success in computer vision tasks such as semantic segmentation, which is the task of assigning a class label to each pixel of an image from a fixed and predefined set of semantic classes [1]. In safety-critical domains such as automated driving, semantic segmentation enables robust scene understanding. However, DNNs show a significant decrease in performance when used in open-world environments where unseen objects occur that are not present in the training distribution. This challenge of out-of-distribution (OOD) segmentation [2] aims to identify and segment objects that do not belong to known classes to avoid dangerous situations. Many OOD segmentation methods are uncertainty-based techniques that do not require auxiliary models or OOD data. DNNs quantify uncertainty using the softmax output, employing common uncertainty metrics such as the maximum softmax probability or entropy [3], whereby the predicted class probability is interpreted as a measure of confidence. However, these softmax-based measures often tend to produce overly confident predictions, especially for OOD samples. Bayesian approximations such as Monte Carlo (MC) Dropout [4] or deep ensembles [5] attempt to overcome this limitation by estimating prediction uncertainty through model sampling, but they require considerable computational effort during inference due to multiple stochastic forward passes.

In this work, we propose an evidential deep learning (EDL) framework using Wasserstein loss for uncertainty-aware semantic segmentation. EDL encourages the model to express higher uncertainty in unfamiliar or ambiguous regions of the input space. Typically, EDL uses Euclidean objectives like expected mean squared error (MSE)-based loss as it works directly on the expected values of the

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

evidence distribution, but leads to overconfident predictions [6]. Therefore, we propose using a Wasserstein loss as it captures the distance between distributions more accurately in metric terms, i.e., the Wasserstein distance accounts for the underlying geometry of the probability simplex, thereby enabling more smooth and calibrated uncertainty estimates. This novel loss formulation combines a distribution-aware Wasserstein term with Kullback-Leibler regularization, which prevents overconfidence for OOD inputs. In addition, Dice is included to enforce spatial coherence in the segmentation output. Together, these components result in a model that improves OOD segmentation compared to uncertainty-based baselines. The code is publicly available on `https://github.com/A-Brosch/EDL-OOD-Segmentation`.

2 Related Work

Existing approaches for OOD segmentation can be grouped into four main research directions, (i) methods relying on auxiliary models to detect anomalies (e.g., reconstruction networks [7]), (ii) approaches leveraging additional OOD or synthetic data during training [8], (iii) techniques operating in the feature-space to identify outliers [9], and (iv) uncertainty-based methods that quantify the model's confidence. The latter includes approaches such as maximum softmax probability [3] and foreground-background segmentation confidence [10], as well as sampling-based methods such as MC Dropout [4] and deep ensembles [5]. More advanced uncertainty-based approaches go beyond output scores and require access to the model's gradients [11] or intermediate computations [12].

Most works rely on auxiliary models, require additional OOD data, or operate in feature-space, which increases complexity and limits general applicability. In contrast, we present an uncertainty-based method in which the confidence of the model is estimated directly from its predictions, without the need for additional data or auxiliary architectures. Compared to other uncertainty-based techniques, our approach does not rely on sampling or access backward passes, making it most comparable to simpler approaches such as maximum softmax.

3 Method

Evidential Deep Learning. In semantic segmentation, each pixel z of an input image x is assigned a class label y from the set of classes $i \in \{1, \ldots, C\}$. The DNN predicts per-pixel (pre-activation) logits, which are transformed into uncertainty-aware outputs using EDL [6], which has also been applied to semantic segmentation in [13]. Instead of applying a softmax function to obtain class scores (single-point predictions), the network produces *evidence values* e_i for each class using ReLU. These evidences define the parameters of a Dirichlet distribution over class probabilities, $\alpha_i = e_i + 1$, $S = \sum_{i=1}^{C} \alpha_i$, which represents a higher-order distribution capturing both prediction and uncertainty. The estimated probability for class i and the corresponding *uncertainty* are given by $p_i = \frac{\alpha_i}{S}$ and $\mathcal{U} = \frac{C}{S}$. Here, S represents the total collected evidence, reflect-

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

ing the model's overall confidence in its pixel-wise prediction. The idea is that there is high uncertainty for classes not included in the training data, enabling identification of OOD objects.

Geometric Intuition + Loss Construction. Most existing EDL formulations employ Euclidean objectives, like expected MSE. The MSE loss, which incorporates the Dirichlet variance is represented by $\mathcal{L}_{\text{MSE}_i} = (p_i - y_i)^2 + \frac{p_i(1-p_i)}{S+1}$, where y_i denotes the value for the i-th class of one-hot encoded ground truth vector. This formulation enables the reduction of prediction error simultaneous to reducing the variance. The MSE loss is beneficial for obtaining sharp segmentation for in-distribution, pushing predictions near the simplex' vertices (class boundaries), corresponding to overconfident one-hot outputs, see Figure 1 (top).

However, this behavior contradicts the detection of OOD pixels, as we are interested in maintaining a state of uncertainty for OOD objects. To capture the distance between predictive and target distributions in a geometrically meaningful way, we employ a Wasserstein loss formulation. Instead of pushing predictions towards the vertices as MSE, the Wasserstein loss does not shift uncertainties to classes but centers them between the boundaries (see Figure 1 (top)). Using the Dirac metric as the transport cost, the Wasserstein distance between the predicted class probability distribution and the ground truth one-hot distribution reduces to the probability assigned to the true class, $\mathcal{L}_{\mathcal{W}} = \mathbb{E}[\,1 - p_y\,]$, where p_y represents the predicted mean probability for the correct class y.

While the Wasserstein loss enforces distribution alignment and improves uncertainty calibration, it does not explicitly consider the spatial structure of the segmentation results. To ensure that neighboring pixels belonging to the same object have consistent predictions, we add a Dice loss term, $\mathcal{L}_{\text{Dice}_i} = 1 - \frac{2\sum_z p_i(z) y_i(z)}{\sum_z p_i(z) + \sum_z y_i(z)}$ working on image-level, which measures the overlap between predicted and true segmentation masks.

To further prevent the model from becoming overconfident in regions of high uncertainty or OOD inputs, we employ an annealed Kullback-Leibler (KL) regularization term $\mathcal{L}_{\text{KL}} = \text{KL}[\text{Dir}(\alpha) \,||\, \text{Dir}(1)]$. The KL term is gradually introduced during training through an annealing schedule λ_{KL}, allowing the network to focus on learning discriminative features before uncertainty calibration is regulated.

Our loss function is composed of $\mathcal{L}_{\text{total}} = \lambda_{\mathcal{W}} \frac{1}{Z} \sum_z \mathcal{L}_{\mathcal{W}} + \lambda_{\text{Dice}} \frac{1}{C} \sum_{i=1}^{C} \mathcal{L}_{\text{Dice}_i} + \lambda_{\text{KL}} \frac{1}{Z} \sum_z \mathcal{L}_{\text{KL}} + \lambda_{\text{MSE}} \frac{1}{Z} \sum_z \sum_{i=1}^{C} \mathcal{L}_{\text{MSE}_i}$ where Z is the number of pixels, and $\lambda_{\mathcal{W}}$, λ_{Dice}, λ_{KL} and λ_{MSE} are the different weights. Although the Wasserstein loss serves as the primary objective, we retain a component of the MSE with a small weighting factor to stabilize the early training phase. In practice, the Wasserstein objective may lead to slower convergence or unstable gradients, especially when the predicted evidence distributions are not yet calibrated. In summary, Wasserstein loss promotes distributional accuracy by aligning predictive and target distributions in a geometrically consistent manner, the Dice loss enforces structural accuracy through spatial and boundary coherence, annealed KL regularization ensures uncertainty calibration by penalizing overconfident evidence in ambiguous regions, and the MSE term contributes training stability by providing a reliable gradient signal during optimization.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Table 1: OOD segmentation results for LostAndFound and RoadObstacle21.

				LostAndFound test-NoKnown					RoadObstacle21				
				AuPRC ↑	FPR_{95} ↓	$\overline{sIoU}$ ↑	$\overline{PPV}$ ↑	$\overline{F_1}$ ↑	AuPRC ↑	FPR_{95} ↓	$\overline{sIoU}$ ↑	$\overline{PPV}$ ↑	$\overline{F_1}$ ↑
Ensemble				2.9	82.0	6.7	7.6	2.7	1.1	77.2	8.6	4.7	1.3
MC Dropout				36.8	35.6	17.4	34.7	13.0	4.9	50.3	5.5	5.8	1.1
Maximum Softmax				30.1	33.2	14.2	**62.2**	10.3	15.7	**16.6**	**19.7**	15.9	6.3
Entropy				**47.1**	**21.6**	**30.7**	42.1	**30.2**	**28.4**	26.7	14.7	**20.5**	**9.7**
$\lambda_{\mathcal{W}}$	λ_{Dice}	λ_{KL}	λ_{MSE}										
0.00	0.00	0.00	1.00	25.7	56.8	25.9	27.7	14.6	1.6	62.1	18.5	5.6	2.0
1.00	0.00	0.00	0.00	47.5	27.4	21.5	39.2	19.7	4.1	66.0	6.9	8.5	2.0
1.00	0.00	0.00	0.40	50.4	25.8	21.2	38.0	19.2	**30.7**	42.3	11.6	26.4	**8.7**
1.00	0.00	0.00	0.45	53.2	23.2	25.3	37.7	20.7	20.2	**31.2**	11.0	18.3	5.8
1.00	0.00	0.00	0.50	44.4	30.5	25.2	**42.8**	23.6	12.2	32.6	3.6	**34.0**	3.4
1.00	0.70	0.00	0.45	39.3	36.1	23.4	41.0	20.9	11.7	58.7	**17.2**	14.9	7.3
1.00	0.75	0.00	0.45	47.4	**17.6**	**26.6**	41.1	**24.2**	8.8	35.3	12.1	14.2	5.3
1.00	0.80	0.00	0.45	27.0	39.8	14.7	25.3	8.7	3.3	63.1	5.0	9.1	1.3
1.00	0.75	0.10	0.45	42.1	25.5	19.8	36.2	16.3	2.8	51.1	10.4	7.4	2.7
1.00	0.75	0.15	0.45	**53.6**	31.6	23.4	38.8	22.1	2.8	51.1	10.4	7.4	2.7
1.00	0.75	0.20	0.45	49.1	38.2	21.9	37.5	20.1	4.9	39.0	16.2	5.4	2.7

4 Experiments

4.1 Experimental Setting

Datasets. For training, we use the Cityscapes dataset [14] for semantic segmentation of urban street scenes consisting of 2,975 training images. For evaluation, we consider the *SegmentMeIfYouCan* benchmark [2] using the LostAndFound (LAF, [15]) dataset (1,203 images with small road obstacles), RoadObstacle21 (412 images similar to LAF with greater diversity in OOD objects and situations), and RoadAnomaly21 (100 images with various unique anomalies).

Model Training. We employ the DeepLabV3+ architecture [1] with ResNet-50 backbone, which is mostly used in OOD segmentation approaches. Training is conducted with a batch size of 4 for 80K iterations, a learning rate of $3 \cdot 10^{-4}$, a weight decay of 10^{-4} and ADAM optimizer. We run ablations on the weighting parameters λ of our combined loss function in the following. The KL weight is increased linearly from 0 to λ_{KL} between iterations 40k and 48k and then kept stable, to encourage learning for higher learning rate and regularization when the learning rate has already dropped. We achieved an average mIoU of 72.17%.

Evaluation Metrics. To evaluate OOD segmentation performance, we follow the official benchmark protocol, i.e., use area under the precision-recall curve (AuPRC) and false positive rate at 95% true positive rate (FPR_{95}) on pixel-level, and averaged adjusted mIoU (sIoU), positive predictive value (PPV) and F_1-score on segment-level.

4.2 Numerical Results

The results for LAF and RoadObstacle21 are given in Table 1, and for RoadAnomaly21 in Table 2. Firstly, we compare the use of clean MSE ($\lambda_{\mathcal{W}} = 0$, $\lambda_{\text{MSE}} = 1$) loss vs. Wasserstein ($\lambda_{\mathcal{W}} = 1$, $\lambda_{\text{MSE}} = 0$) and observe that, across all datasets, Wasserstein loss achieves significantly better results than MSE.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Table 2: OOD segmentation results for RoadAnomaly21.

				RoadAnomaly21				
				AuPRC ↑	FPR_{95} ↓	$\overline{\text{sIoU}}$ ↑	$\overline{\text{PPV}}$ ↑	$\overline{F_1}$ ↑
Ensemble				17.7	91.1	16.4	**20.8**	3.4
MC Dropout				28.9	**69.5**	**20.5**	17.3	4.3
Maximum Softmax				28.0	72.1	15.5	15.3	**5.4**
Entropy				**30.0**	73.0	17.8	15.6	5.1
$\lambda_{\mathcal{W}}$	λ_{Dice}	λ_{KL}	λ_{MSE}					
0.00	0.00	0.00	1.00	33.4	68.8	19.0	17.4	3.8
1.00	0.00	0.00	0.00	39.7	61.8	14.5	21.6	4.0
1.00	0.00	0.00	0.40	**54.3**	50.4	26.3	**21.5**	7.8
1.00	0.00	0.00	0.45	50.5	49.6	**28.0**	20.8	**9.0**
1.00	0.00	0.00	0.50	40.0	53.1	19.1	20.3	4.6
1.00	0.70	0.00	0.45	42.8	**43.6**	23.9	19.2	6.3
1.00	0.75	0.00	0.45	41.4	56.3	26.3	16.0	7.1
1.00	0.80	0.00	0.45	37.9	65.5	22.9	15.7	4.8
1.00	0.75	0.10	0.45	35.0	48.2	16.3	6.6	5.0
1.00	0.75	0.15	0.45	43.6	53.6	24.6	15.2	5.9
1.00	0.75	0.20	0.45	35.4	46.9	20.6	17.5	5.6

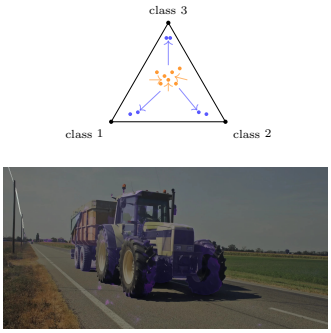

Fig. 1: Comparison of MSE (*blue*) and Wasserstein loss (*orange*) on the three-class probability simplex (*top*) and an uncertainty heatmap out of AnomalyTrack (*bottom*).

Secondly, we added MSE loss with a small weighting to the Wasserstein loss to increase segmentation performance without losing latent uncertainty. This combination delivers significantly higher performance than if only one factor is considered. From now on, we will set $\lambda_{\text{MSE}} = 0.45$ as it has achieved the best results for LAF and this dataset consists of the most images and therefore serves as a reference value. Thirdly, we are investigating the impact of DICE loss, which has a particularly positive effect on the FPR_{95} metric. We will fix λ_{Dice} to 0.75. Lastly, we include annealed KL regularization to reduce overconfidence. We have found that a value of $\lambda_{\text{KL}} = 0.15$ works best for LAF (corresponding to a small expected calibration error of 0.0347 on in-distribution data). Moreover, we experimented with using lower evidence for the prior distribution (e.g., Dir(0.25)), which weakened its dampening effect on the estimate used in the KL regularization, but this did not lead to measurable improvements. Overall, our experiments show that we achieve the best performance with the combined loss when at least three components have an influence on the prediction.

The baseline methods against which we compare our approach are listed at the top of both tables. We use comparable methods, i.e., uncertainty-based approaches, distinguishing between sampling-based (Ensemble, MC Dropout) and output-based (Maximum Softmax, Entropy) techniques. We obtain improved results compared to these baselines, especially with RoadAnomaly21, where we outperform all others.

5 Conclusion

In this work, we have introduced a Wasserstein-based evidential framework for OOD-aware semantic segmentation that jointly optimizes distribution accuracy, structural consistency, and calibrated uncertainty. Our method better accounts

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

for the geometry of the probability simplex and provides smoother, more reliable uncertainty estimates than MSE. In the SegmentMeIfYouCan benchmark, our approach demonstrates improved performance compared to other uncertainty-based baseline methods, proving that principled uncertainty modeling enables more robust and reliable predictions in open-world environments.

References

[1] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *European Conference on Computer Vision (ECCV)*, 2018.

[2] Robin Chan, Krzysztof Lis, Svenja Uhlemeyer, Hermann Blum, Sina Honari, Roland Siegwart, Pascal Fua, Mathieu Salzmann, and Matthias Rottmann. Segmentmeifyoucan: A benchmark for anomaly segmentation. In *Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2021.

[3] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. 2016.

[4] Jishnu Mukhoti and Yarin Gal. Evaluating bayesian deep learning methods for semantic segmentation. 2018.

[5] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *International Conference on Neural Information Processing Systems*, pages 6405–6416, 2017.

[6] Murat Sensoy, Lance Kaplan, and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 3183–3193. Curran Associates Inc., 2018.

[7] Vojíř, Tomáš and Matas, Jiří. Image-consistent detection of road anomalies as unpredictable patches. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 5480–5489, 2023.

[8] Anja Delić, Matej Grcic, and Siniša Šegvić. Outlier detection by ensembling uncertainty with negative objectness. In *35th British Machine Vision Conference 2024, BMVC 2024, Glasgow, UK, November 25-28, 2024*. BMVA, 2024.

[9] Matteo Sodano, Federico Magistri, Lucas Nunes, Jens Behley, and Cyrill Stachniss. Open-world semantic segmentation including class similarity. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3184–3194, 2024.

[10] Samuel Marschall and Kira Maag. Multi-scale foreground-background confidence for out-of-distribution segmentation. In *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 486–496, 2025.

[11] Kira Maag and Tobias Riedlinger. Pixel-wise gradient uncertainty for convolutional neural networks applied to out-of-distribution segmentation. *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 2024.

[12] Shiyu Liang, Yixuan Li, and Rayadurgam Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks. *International Conference on Learning (ICLR)*, 2018.

[13] Karl Holmquist, Lena Klasén, and Michael Felsberg. Evidential deep learning for class-incremental semantic segmentation. In Rikke Gade, Michael Felsberg, and Joni-Kristian Kämäräinen, editors, *Image Analysis*, pages 32–48. Springer Nature Switzerland, 2023.

[14] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[15] Peter Pinggera, Sebastian Ramos, Stefan Gehrig, Uwe Franke, Carsten Rother, and Rudolf Mester. Lost and found: detecting small road hazards for self-driving vehicles. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016.