Real-Time Anomaly Segmentation for Road Scenes

Vincenzo Catalano
Politecnico di Torino

Alessio Gioè
Politecnico di Torino

Riccardo Turco
Politecnico di Torino

Abstract

Anomaly segmentation plays a crucial role in addressing Out-of-Distribution (OoD) objects in deep learning, a significant challenge for applications such as autonomous driving.. Existing models often struggle with real-time performance and memory constraints, particularly on edge devices. This work aims to bridge this gap by developing lightweight, real-time anomaly segmentation models optimized for memory efficiency. Using Cityscapes for training and Fishyscapes for evaluation, the study explores advanced loss functions, including Enhanced Isotropy Maximization and Logit Normalization. Preliminary results show potential for robust anomaly detection while maintaining model efficiency, highlighting the importance of this approach for real-world deployment.

1 Introduction

The deployment of deep learning models in open-world environments, especially in applications like autonomous driving, presents significant challenges in handling Out-of-Distribution (OoD) objects. These models need to perform in real-time while meeting strict memory constraints, especially when deployed on edge devices with limited onboard processing capabilities. This issue is particularly important as real-time anomaly detection is critical for safety and decision-making in autonomous systems.

This research[1] addresses the gap in lightweight anomaly segmentation models that balance real-time performance and memory efficiency. This research specifically investigates the effect of training loss functions tailored for anomaly detection. We analyze the impact of the following losses on model performance: Enhanced Isotropy Maximization Loss, Logit Normalization Loss, and their combination with focal loss and cross-entropy loss. By focusing on these specialized loss functions, we explore how they influence the robustness of anomaly segmentation models. The study uses Cityscapes for training and datasets such as Road Anomaly and Fishyscapes to evaluate the models and provide insights into optimizing anomaly detection performance.

[1] Visita il repository su GitHub

2 Related Works

Recent advancements in lightweight semantic segmentation models aim to balance performance and memory efficiency for edge devices. Models like BiSeNet and BiSeNetV2 use a two-pathway structure to combine spatial and contextual information, ensuring fast inference for real-time applications. ENet, with its encoder-decoder architecture, prioritizes speed and low memory usage for real-time segmentation tasks.

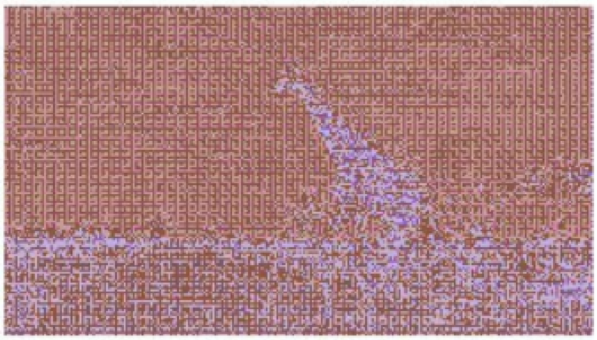
ERFNet, known for its residual factorization strategy, minimizes computational costs while maintaining accuracy, making it ideal for mobile and embedded systems. Although ERFNet excels in general semantic segmentation, its application to anomaly detection remains underexplored. Anomaly segmentation requires identifying out-of-distribution objects.

This study builds on ERFNet by incorporating advanced loss functions for anomaly detection, such as Enhanced Isotropy Maximization Loss (EIML) and Logit Normalization Loss (LN), aimed at improving real-time performance on edge devices. Our work differs from existing approaches by focusing on the effects of specialized loss functions and calibration techniques, such as temperature scaling, for anomaly segmentation.

3 Background

Anomaly detection in deep learning involves identifying data points that deviate from the normal distribution of known categories. In semantic segmentation, this task focuses on detecting Out-of-Distribution (OoD) objects, which do not belong to any predefined class in the training data. The goal is to enable models to identify such anomalies

(a) Base image with OOD object (gyraffe).



(b) OOD image segmentation.

Figure 1: Image segmentation depicted by our model trained with EIML+FL loss.

in real-time, a crucial requirement for applications like autonomous driving.

In the general framework for anomaly detection in semantic segmentation, models like ERFNet, BiSeNetV2, and ENet are commonly employed due to their efficiency and low computational cost, making them suitable for edge devices. During inference, the model classifies each pixel into one of the predefined classes, but may encounter objects outside its training data—potential anomalies.

A common methodology for anomaly detection involves several stages. First, a deep neural network, such as ERFNet, is trained to classify known classes in a labeled dataset. During inference, techniques like Maximum Softmax Probability (MSP) or Maximum Logit are used to measure the model's uncertainty. These methods identify when the model is uncertain, suggesting the presence of an anomaly. Calibration techniques, such as temperature scaling, further refine the model's confidence in its predictions, improving its ability to differentiate between known and unknown objects.

Another approach is the Void Classifier, where the void or background class in datasets like Cityscapes is treated as a potential anomaly. The model's predictions in these regions may indicate uncertainty, suggesting anomalous objects.

To evaluate anomaly detection, common metrics include Area under the Precision-Recall Curve (AU-PRC) and False Positive Rate at 95% recall (FPR95). These metrics help assess the model's ability to correctly identify anomalies (AU-PRC) while minimizing false positives (FPR95), providing insights into its performance in real-world scenarios.

4 Method

In this study, we aim to investigate how various model enhancements, such as advanced loss functions and temperature scaling, can improve anomaly detection in semantic segmentation models, especially in environments with limited computational resources. Our main claim is that, by optimizing model training with specialized loss functions and inference techniques, we can achieve robust anomaly detection without compromising model efficiency. To test this hypothesis, our methodology is structured around the following key objectives:

1. **Benchmarking Lightweight Models**: We apply our methodology to a lightweight semantic segmentation model, ERFNet. Although ERFNet is highly effective in semantic segmentation tasks involving known classes, it may struggle with detecting anomalies. Using a pre-trained version of ERFNet on the benchmark dataset Cityscapes, we evaluate its performance in detecting anomalies when presented with unseen objects or classes. To achieve this, we apply well-established methods for uncertainty estimation, including Maximum Softmax Probability (MSP), MaxLogit, and Maximum Entropy. The objective is to understand how well these methods perform when faced with tasks that require anomaly segmentation.

2. **Improving Model Calibration with Temperature Scaling**: Another critical aspect of our methodology is testing how temperature scaling, a confidence calibration technique, affects anomaly detection. By adjusting the output logits of the model, temperature scal-

ing aims to refine the model's uncertainty estimation. The objective is to identify the optimal temperature that enhances the model's performance in distinguishing known from unknown classes, thereby improving the overall anomaly detection capability.

3. **Void Class Detection as a Proxy for Anomaly Detection**: One approach we explore is using the void class, which typically represents background or unlabeled regions in segmentation datasets, as an indicator of anomaly. By training networks like ENet, ERFNet and BiSeNetV2 to focus specifically on the void class, we examine whether anomalies can be detected by identifying regions of uncertainty (i.e., voids). By using pre-trained versions of these models on benchmark datasets, we evaluate their performance in detecting anomalies when faced with unseen objects.

4. **Role of Advanced Loss Functions**: We focus on the effect of specific loss functions tailored for anomaly detection. The central question here is whether incorporating losses like Enhanced Isotropy Maximization Loss and Logit Normalization Loss improves the model's ability to detect Out-of-Distribution (OoD) objects. These losses are designed to help the model better distinguish between known classes and anomalies. Our goal is to assess how these losses, both individually and combined with focal and cross-entropy losses, can enhance anomaly detection performance.

Through this methodology, we aim to demonstrate that advanced loss functions, combined with effective calibration techniques, can significantly enhance the anomaly detection capability of lightweight models, without compromising their real-time performance or efficiency. The results of this study have the potential to improve the robustness of semantic segmentation models in real-world applications, particularly in autonomous systems, where reliable anomaly detection is crucial for safety and decision-making.

5 Experiments

This section presents the datasets, architectures, and experiments conducted to validate our claims about improving anomaly detection in lightweight semantic segmentation models.

5.1 Datasets

We utilized the Cityscapes dataset for training, which provides high-quality annotations for urban street scenes. It consists of 19 semantic categories and an additional *void* class, which represents unlabeled regions. For evaluation, we employed the following datasets designed for anomaly detection:

- **Fishyscapes Lost and Found (FS Lost & Found)**: A dataset featuring Out-of-Distribution (OoD) objects placed in realistic urban environments, providing a challenging test bed for anomaly segmentation.

- **Fishyscapes Static (FS Static)**: A dataset where anomalies are synthetically embedded into static urban images to evaluate the robustness of anomaly detection methods.

- **Road Anomaly**: A dataset specifically tailored for detecting anomalous objects on roads, including unusual objects that do not belong to known categories.

- **Road Obstacle**: A dataset focused on obstacles and hazards commonly encountered in autonomous driving scenarios, offering a complementary perspective on anomaly detection.

5.2 Architectures

We evaluated three lightweight semantic segmentation architectures:

- **ERFNet**: Known for its efficient use of computational resources and real-time segmentation capabilities, ERFNet was pretrained on Imagenet and trained on Cityscapes and used as a baseline for anomaly detection experiments.

- **ENet**: A compact network designed for efficient semantic segmentation, tested with a focus on its ability to detect anomalies through void classification.

- **BiSeNetV2**: A lightweight model optimized for real-time inference, evaluated for its anomaly segmentation performance when trained on the void class.

5.3 Experimental Setup

Baseline Experiments. We started by conducting anomaly inference using the pretrained ERFNet model. Standard uncertainty estimation methods,

Method		SMIYC RA-21		SMIYC RO-21		FS L&F		FS Static		Road Anomaly	
	mIoU ↑	AuPRC ↑	FPR95 ↓	AuPRC ↑	FPR95 ↓	AuPRC ↑	FPR95 ↓	AuPRC ↑	FPR95 ↓	AuPRC ↑	FPR95 ↓
MSP	72.20	***14.58***	***95.08***	0.72	94.76	22,74	93,14	18.76	93.47	***9.42***	***95.30***
MaxEntropy	72.20	14.31	97.11	0.83	94.08	24.03	96.82	19.31	92.99	9.10	95.31
MaxLogit	72.20	13.20	97,01	***1,15***	***86,81***	***31.04***	***90.11***	***23.20***	***91.90***	8.71	93,76

Table 1: ERFNet: Various anomaly inferences using the pre-trained model and anomaly segmentation test datasets provided.

such as Maximum Softmax Probability (MSP), Max-Entropy, and Max-Logit, were applied to detect anomalies. To refine the calibration of the model, we also experimented with temperature scaling, optimizing the temperature parameter to achieve the best anomaly detection results.

Void Classification. In this experiment, we explored the potential of using the *void* class as a proxy for anomaly detection. The void class in the Cityscapes dataset represents unlabeled or ambiguous regions and is assumed to signify anomalies in this context. We trained ENet, ERFNet, and BiSeNetV2 architectures and performed anomaly inference by selecting only the output corresponding to the void class. The Maximum Softmax Probability (MSP) method was used as the inference technique. Additionally, we evaluated the models based on their speed and memory efficiency to ensure that the proposed approach is suitable for real-time, resource-constrained applications.

Loss Function Analysis. We evaluated the impact of advanced loss functions specifically designed for anomaly detection. Experiments were conducted using Enhanced Isotropy Maximization Loss (EIML) and Logit Normalization Loss (LN), both individually and jointly trained with focal loss and cross-entropy loss. The model was trained for 50 epochs; however, as observed from Figure 2, the loss curve has not yet reached its minimum. This suggests that extending the training duration could potentially lead to further improvements in the model's performance.

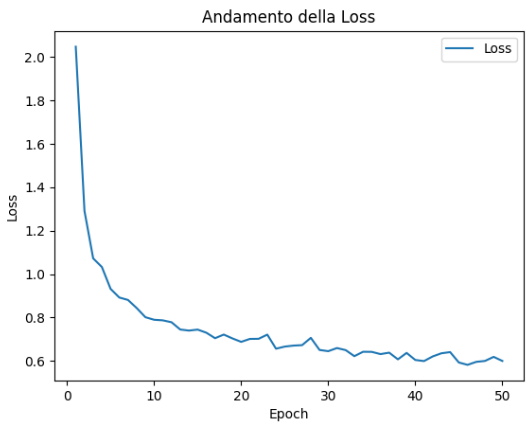


Figure 2: EIML+FL Loss during the training of our model.

5.4 Results and Discussion

The results of our experiments are summarized as follows:

Baseline Anomaly Detection

On Table 1 we see the results for ERFNet model. It was evaluated using multiple uncertainty estimation methods, including Maximum Softmax Probability (MSP), Max-Logit, and Max-Entropy, across five datasets: SMIYC RA-21, SMIYC RO-21, Fishyscapes Lost and Found (FS L&F), Fishyscapes Static (FS Static), and Road Anomaly. The results revealed that MaxLogit achieved the highest performance on several datasets, such as AuPRC values of 1.15 (SMIYC RO-21), 31.04 (FS L&F) and 23.20 (FS Static), along with a lower FPR95 of 86.81 (SMIYC RO-21), 90.11 (FS L&F) and 91.90 (FS Static) compared to other methods. However, MSP demonstrated comparable performance in some scenarios, with an AuPRC of 14.58 and an FPR95 of 95.08 on SMIYC RA-21, outperforming MaxLogit. Conversely, MaxEntropy sometimes outperforms MSP but does not shine on any particular dataset, indicating less reliable anomaly detection under this uncertainty estimation approach. The analysis highlights the utility of MaxLogit as a robust baseline for anomaly segmentation with ERFNet.

Temperature Scaling

Table 2 covers temperature scaling, which is a confidence calibration method applied to improve the anomaly segmentation capabilities of the ERFNet model. We tested multiple temperature values ($t = 0.5$, $t = 0.75$, and $t = 1.1$) to determine the optimal setting for anomaly detection . The results showed that $t = 1.1$ provided the best overall performance. For example, on the FS Lost and Found dataset, it achieved the best AuPRC value of 22.94. Similarly, it achieved the best re-

Method	mIoU ↑	SMIYC RA-21 AuPRC ↑	SMIYC RA-21 FPR95 ↓	SMIYC RO-21 AuPRC ↑	SMIYC RO-21 FPR95 ↓	FS L&F AuPRC ↑	FS L&F FPR95 ↓	FS Static AuPRC ↑	FS Static FPR95 ↓	Road Anomaly AuPRC ↑	Road Anomaly FPR95 ↓
MSP	72.20	14.58	95.08	0.72	94.76	22,74	93,14	18.76	93.47	9.42	95.30
MSP(t=0.5)	72.20	***14.68***	***95.05***	0.69	94.88	21.29	94.17	17.98	94.26	***9.60***	***95.17***
MSP(t=0.75)	72.20	14.63	95.07	0.71	94.82	22.12	93.65	18.45	93.84	9.50	95.24
MSP(t=1.1)	72.20	14.57	95.09	***0.72***	***94.74***	***22.94***	***92.92***	***18.85***	***93.34***	9.39	95.31
MSP(best t)	72.20	14.68	95.05	0.72	94.74	22.94	92.92	18.85	93.34	9.60	95.17

Table 2: Temperature scaling on ERFNet: performance of ERFNet with MSP across various temperature scaling values on five anomaly detection datasets.

sults on the Road Obstacle dataset and FS Static. Across datasets, this temperature setting consistently enhanced model confidence. The temperature of $t = 0.5$ achieved better results on Road Anomaly datasets, but is generally outperformed by $t = 1.1$. Other tested temperatures, such as $t = 0.75$ and $t = 1.1$, demonstrated worse results than the normal MSP model. These results indicate that calibrating the model with a carefully chosen temperature can refine its anomaly segmentation capabilities, making it more effective for real-world deployment.

Void Classifier

On Table 3 we see the results of this experiment, where we assumed the void class in the Cityscapes dataset as an anomaly and performed anomaly inference by exclusively selecting the output of the void class. The Maximum Softmax Probability (MSP) method was applied across three lightweight networks: ENet, ERFNet, and BiSeNetV2. The results revealed that ERFNet achieved the highest mIoU of 58.19, significantly outperforming ENet (40.94) and BiSeNetV2 (41.24). We can deduce from the values that ENet and BiSeNetV2 networks are not well trained, to improve their results we can increase the number of epochs used for training and use a more suitable loss function. For anomaly detection, ERFNet and ENet demonstrated comparable AuPRC values across datasets, such as 14.81 on SMIYC RA-21 and 18.87 on FS Lost and Found. However, BiSeNetV2 exhibited slightly higher AuPRC values on certain datasets, including 14.83 on SMIYC RA-21 and 18.60 on FS Lost and Found, but this came with a trade-off in higher FPR95 values, such as 96.73 on FS Static and 97.25 on Road Anomaly.

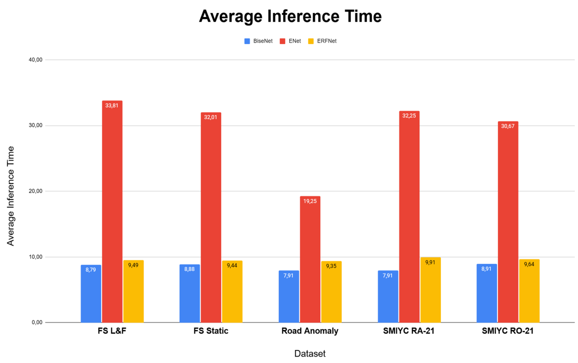


Figure 3: Average inference time of the model using Enhanced Isotropy Maximization Loss combined with Focal Loss during the training of our model.

From Figure 3, we observe that the inference time for ENet is significantly higher compared to ERFNet and BiSeNet, rendering it unsuitable for our task. While BiSeNetV2 is slightly faster than ERFNet, Figure 4 also shows that BiSeNetV2 requires considerably more memory, making ERFNet the more favorable choice overall.

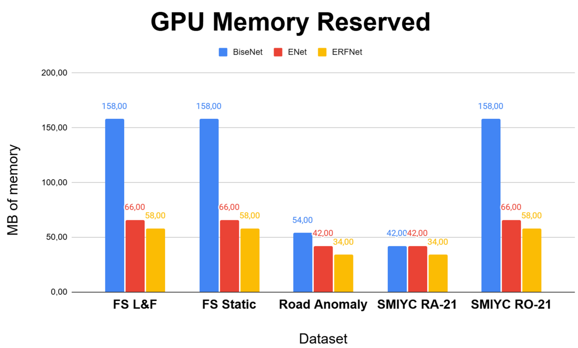


Figure 4: GPU memory reserved of the model using EIML combined with Focal Loss.

Method	mIoU ↑	SMIYC RA-21 AuPRC ↑	SMIYC RA-21 FPR95 ↓	SMIYC RO-21 AuPRC ↑	SMIYC RO-21 FPR95 ↓	FS L&F AuPRC ↑	FS L&F FPR95 ↓	FS Static AuPRC ↑	FS Static FPR95 ↓	Road Anomaly AuPRC ↑	Road Anomaly FPR95 ↓
ENet	40.94	14.81	95.00	0.67	95.00	18.87	95.00	16.36	95.00	9.83	95.24
ERFNet	58.19	14.81	95.00	0.67	95.00	18.87	95.00	16.35	95.08	9.85	94.99
BiSeNetV2	41.24	14.83	94.82	0.67	95.00	18.60	96.73	16.02	97.25	9.82	95.36

Table 3: Void Classifier: we will assume the void class as an anomaly and we will perform the anomaly inference by only selecting the output of the Void class (method = MSP).

Method	Loss		SMIYC RA-21		SMIYC RO-21		FS L&F		FS Static		Road Anomaly	
		mIoU ↑	AuPRC ↑	FPR95 ↓	AuPRC ↑	FPR95 ↓	AuPRC ↑	FPR95 ↓	AuPRC ↑	FPR95 ↓	AuPRC ↑	FPR95 ↓
MSP	LN	67.33	16,13	94,44	0,65	95,30	19,17	94,46	16,18	94,58	10,63	94,58
	LN + CE	66.62	16,41	94,33	0,80	93,97	19,74	94,52	16,61	94,79	11,20	94,16
	LN + FL	67.00	16,11	94,47	0,62	95,79	19.35	94.91	15.26	95.64	11.23	94,20
	EIML	56.33	12.74	96.79	0.52	97.68	21.10	96.22	17.32	96.31	10.04	93.53
	EIML + FL	56.96	17.75	93.68	0.75	94.22	21.81	93.15	17.44	94.28	12.89	92.66
	EIML + CE	50.12	15.90	94.51	1.37	93.52	18.13	95.40	15.62	95.52	12.16	93.69
MaxEntropy	LN	67.33	17.19	94.39	0.61	96.96	19.79	90.79	16.30	97.26	11.12	93.93
	LN + FL	67.00	16.92	95.71	0.57	98.45	20.58	93.10	15.05	92.43	11.96	93.03
	LN + CE	66.62	17.87	95.44	0.90	95.37	20.42	92,68	16.81	92.45	12.39	86.04
	EIML	56.33	12.19	98.28	0.47	97.47	20.97	98.19	17.26	97.24	10.36	90.13
	EIML + FL	56.96	19.13	89.96	0.92	80.49	22.42	93.99	17.69	94.17	13.84	87.81
	EIML + CE	50.12	17.48	95.40	1.99	92.34	17.41	93.10	15.01	93.76	14.05	85.93
MaxLogit	LN	67.33	18.92	98.39	0.83	99.14	17.43	92.36	15.34	94.74	12.17	90.65
	LN + CE	66.62	15.83	96.69	1.19	98.86	14,62	96,27	13.44	94.93	11.70	83.97
	LN + FL	67.00	20.18	95.34	0.51	98.72	23.16	95.07	16.23	94.35	14.32	89.81
	EIML	56.33	12.91	98.26	0.46	97.56	21.78	98.33	17.74	97.44	11.03	90.74
	EIML + FL	56.96	20.01	86.33	1.53	81.38	27.40	92.45	19.73	91.80	13.35	85.42
	EIML + CE	50.12	23.56	95.17	4.11	93.24	16.54	94.47	14.32	94.01	16.21	86.22

Table 4: Effect of Training Loss function on ERFNet: impact of different training loss functions on ERFNet's performance, evaluated using mIoU, AuPRC, and FPR95 across five anomaly detection datasets.

Effect of Training Loss Functions on ERFNet
Table 4 shows the performance of ERFNet evaluated with different combinations of anomaly detection-specific loss functions, such as Enhanced Isotropy Maximization Loss (EIML), Logit Normalization Loss (LN), and Focal Loss (FL), alongside standard Cross-Entropy (CE) loss. The results revealed that the usage of loss function that are specific to anomaly segmentation tasks showed notable improvement over cross-entropy loss. The loss that generally achieved the best results is the combination of EIML and Focal Loss, outperforming other variants on most datasets, especially with the MSP method. On Max Entropy and Max Logit the combination of EIML and Cross Entropy Loss achieved great results, even better than EIML and Focal Loss on some instances . EIML and its combinations proved more effective than Logit Normalization, with the latter obtaining on very few cases. The combination of LN with other loss functions seems to have decreased its performance. Interestingly, the usage of these loss functions lead to a decrease in mIoU perfomance, but dramatically increased their performance on anomaly detection. Overall, our findings show that the joint use of advanced loss functions, especially EIML with Focal Loss, significantly enhances ERFNet's anomaly segmentation capabilities compared to standard Cross-Entropy loss. These results highlight the importance of carefully selecting and combining loss functions for improving anomaly detection in semantic segmentation tasks. These experiments demonstrate that advanced loss functions and calibration techniques can significantly enhance the anomaly detection capabilities of lightweight models. The findings underscore the importance of combining architectural efficiency with specialized training methodologies to achieve robust and reliable anomaly segmentation.

6 Conclusion

This study demonstrated that advanced loss functions, particularly Enhanced Isotropy Maximization Loss (EIML) combined with Focal Loss, significantly improve ERFNet's performance in anomaly segmentation tasks. While the use of these loss functions led to a decrease in mIoU, they greatly enhanced anomaly detection metrics, such as AuPRC and FPR95. Our findings highlight the importance of selecting appropriate loss functions to balance anomaly detection and segmentation accuracy. Future work could explore further optimizations and hybrid loss functions by analyzing how assigning different weights to the combination of loss components impacts performance. Additionally, it would be valuable to investigate whether incorporating temperature scaling, which enhances model performance with standard loss functions, could yield improvements when applied in this context. These efforts aim to address trade-offs and refine real-time performance, particularly for deployment on edge devices.

7 References

I. Git repository

II. Google Sheets of the results

III. BiSeNet: Bilateral Segmentation Network for Real-time Semantic Segmentation

IV. BiSeNet V2: Bilateral Network with Guided Aggregation for Real-time Semantic Segmentation

V. ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation

VI. The Fishyscapes Benchmark Anomaly Detection for Semantic Segmentation

VII. ERFNet: Efficient Residual Factorized ConvNet for Real-Time Semantic Segmentation

VIII. Starting Code Repository

IX. Scaling Out-of-Distribution Detection for Real-World Settings

X. Max Entropy

XI. Temperature scaling

XII. Enhanced Isotropy Maximization Loss

XIII. Mitigating Neural Network Overconfidence with Logit Normalization