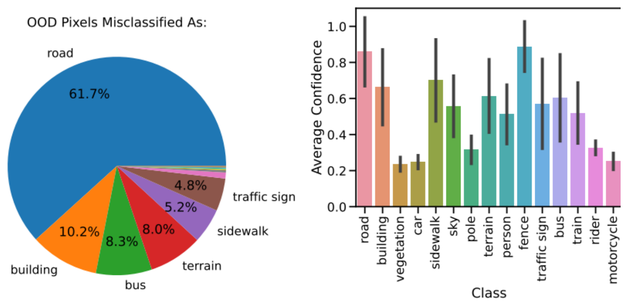
Anomaly-Aware Semantic Segmentation via Style-Aligned OoD Augmentation

Dan Zhang[1,2], Kaspar Sakmann[1], William Beluch[1], Robin Hutmacher[1], and Yumeng Li[1,3]
[1]Bosch Center for Artificial Intelligence,
[2]University of Tübingen, [3]University of Siegen

Motivation

- Within the context of autonomous driving, encountering unknown objects (a.k.a., anomalies) becomes inevitable during deployment in the open world.
- Overlooking objects on the road is a critical error that carries high-level risks. Regrettably, this is a prevailing error pattern observed in semantic segmentation models.
- In this work, we provide a simple finetuning solution to enable their anomaly awareness.



(a) Confusion Rate (b) Confidence of Error Prediction

Experiment Results

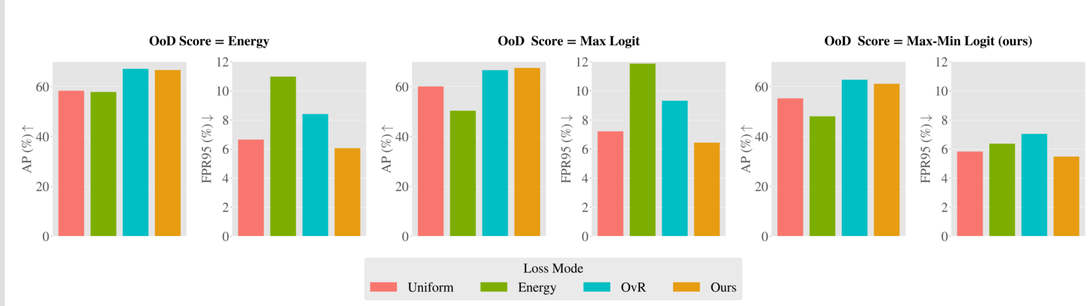
DeepLabv3+ Method	w. WideResNet38 backbone Cityscapes mIoU ↑	Fishyscapes L & F AUC ↑	AP ↑	FPR95 ↓	w. ResNet101 backbone Cityscapes mIoU ↑	Fishyscapes L & F AUC ↑	AP ↑	FPR95 ↓
Max Softmax Pred. [13]		89.29	4.59	40.59		86.99	6.02	45.63
Max Logit (ML)[13]		93.41	14.59	42.21		92.00	18.77	38.13
Entropy [14]	**90.62**	90.82	10.36	40.34	**80.50**	88.32	13.91	44.85
Energy [22]		93.72	16.05	41.78		93.50	25.79	32.26
Standardized ML [16]		94.97	22.74	33.49		96.88	36.55	14.53
Meta-OOD [4]	89.00	93.06	41.31	37.69	-	-	-	-
PEBAL [30]	89.12	**98.96**	58.81	**4.76**	-	**99.09**	59.83	6.49
Ours (Max Logit)		98.71	**71.94**	6.42		98.45	67.35	9.36
Ours (Energy)	90.39	98.79	70.87	5.88	**80.50**	98.58	**69.93**	8.38
Ours (Max-Min Logit)		98.87	70.84	5.52		98.83	66.32	**5.74**

OoD Score	w./o. Style Align. AUC ↑	AP ↑	FPR95 ↓	w. Style Align. AUC ↑	AP ↑	FPR95 ↓
Max Softmax Pred.	94.82	32.32	20.76	+1.55	+18.84	−2.74
Entropy	96.21	47.14	19.76	+1.13	+16.37	−2.85
Max Logit	97.84	51.79	12.61	+0.61	+15.56	−3.25
Energy	98.02	52.32	11.92	+0.56	+17.61	−3.54
Max - Min. Logit	98.24	45.02	9.14	+0.59	+21.30	−3.40

Style alignment greatly improves the performance.

Method	K	AUC ↑	AP ↑	FPR95 ↓
PEBAL [30]	-	**99.09**	59.83	6.49
OvR (Max Logit)	-	97.70	52.24	12.97
OvR (Energy)	-	97.95	59.96	12.09
OvR (Max-Min Logit)	-	98.52	59.19	7.51
Ours (Max Logit)	3	98.34	60.86	10.50
	5	98.45	67.35	9.36
	7	98.12	63.88	11.40
Ours (Energy)	3	98.48	64.07	9.73
	5	98.58	**69.93**	8.38
	7	98.28	68.30	10.47
Ours (Max-Min Logit)	3	98.79	58.59	6.37
	5	98.83	66.32	**5.74**
	7	98.69	66.56	6.43

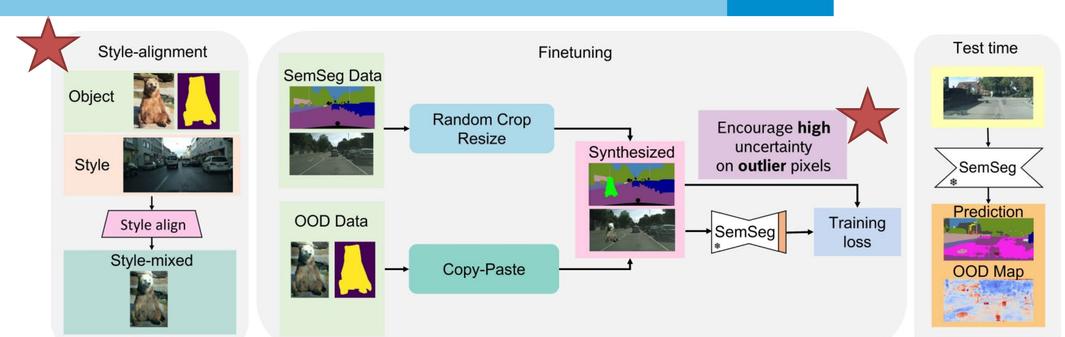
Our finetuning loss consistently outperforms other uncertainty regularization losses across different evaluation metrics.



Conclusion & Outlook

- We developed a simple finetuning method that enables semantic segmentation models to detect unknown objects as well.
- We observed the necessity of considering domain shifts jointly with unknown objects, awaiting for further investigation.

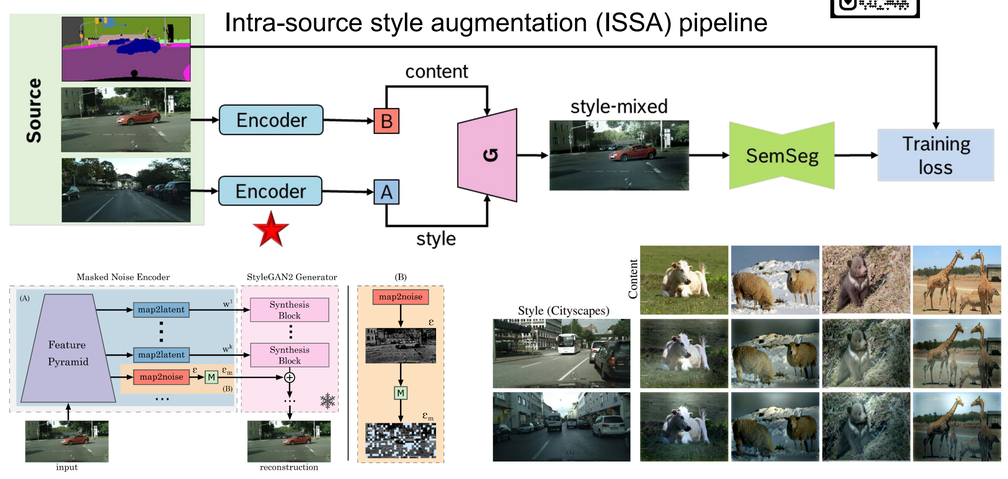
Our Approach



- Only **finetune the final classification head** of semantic segmentation models using **our style-aligned OoD (out-of-distribution) augmentation** & **top-k one-vs-rest (OvR) loss**.
- After finetuning, a **high-quality pixel-wise OoD prediction map can be derived from the output logits** of the model.

Style Alignment

- Synthetic OoD augmentation via Copy & Paste introduces domain gap, i.e., OoD data (e.g., MS COCO objects) has different styles than autonomous driving data (e.g., Cityscapes).
- We advance the synthetic OoD generation process by performing style alignment between the OoD data and driving scene data.
- For style alignment, we exploited the ISSA method.



Top-k OvR Loss

- The One-vs-Rest (OvR) loss induces a pre-trained semantic segmentation model to generate a “none of the given classes” prediction on synthetic OoD pixels.
- We focus on the worst cases by minimizing the top-k terms.

$$\mathcal{L}_{\text{ood}} = \frac{1}{K|\mathcal{N}_{\text{ood}}|} \sum_{i \in \mathcal{N}_{\text{ood}}} \sum_{k \in \mathcal{S}_{\text{topK}}(i)} -\log \sigma(-s\lambda_{i,k})$$

Per-pixel OOD Score

Max. Logit	$\max_k \lambda_{i,k}$
Energy Score	$\sum \log \sum e^{\lambda_{i,k}}$
Max-Min Diff.	$\max_k \lambda_{i,k} - \min_k \lambda_{i,k}$

Visual Examples (Hard Cases)