Generalize or Detect? Towards Robust Semantic Segmentation Under Multiple Distribution Shifts

Zhitong Gao[1,2], Bingnan Li[1], Mathieu Salzmann[2], Xuming He[1,3]

[1] ShanghaiTech University [2] EPFL

[3] Shanghai Engineering Research Center of Intelligent Vision and Imaging

Background **Semantic Segmentation Under Distribution Shifts**.

Domain Generalization (DG) Techniques focus on generalizing to covariate shifts.

- e.g., different weather or object attributes.

Out-of-distribution (OOD) Detection Techniques focus on detecting semantic shifts.

- e.g., anomalies or novel objects.

Training set (Eg. Cityscapes)

Test image with covariate shifts (Eg. ACDC)

Test image with semantic shifts (Eg. SMIYC)

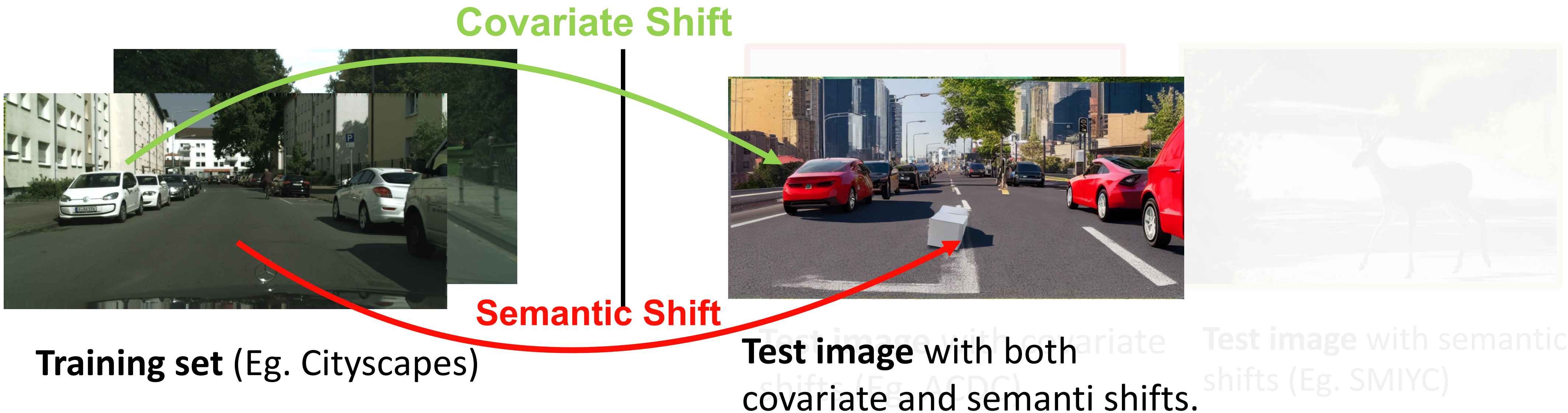
Background **Semantic Segmentation Under Distribution Shifts.**

Domain Generalization (DG) Techniques focus on generalizing to covariate shifts.

- e.g., different weather or object attributes

Out-of-distribution (OOD) Detection Techniques focus on detecting semantic shifts.

Can a model jointly handle both kinds of distribution shift?



Training set (Eg. Cityscapes)

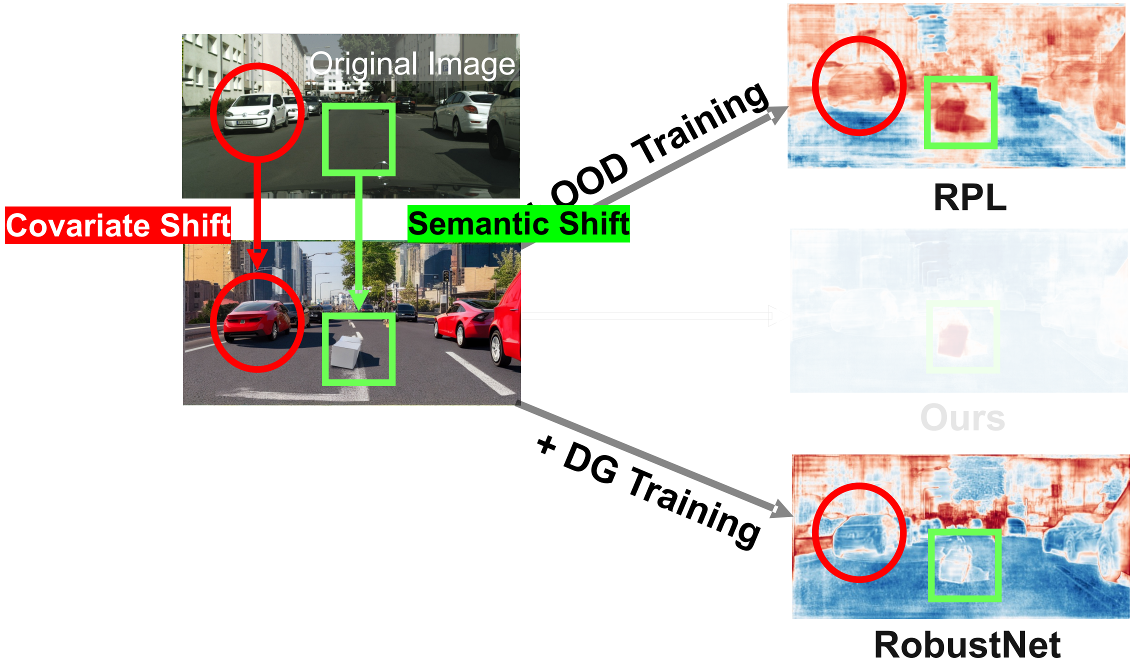
Test image with both covariate and semanti shifts.

Challenges Semantic Segmentation Under <u>Multiple</u> Distribution Shifts.

☹ **Domain Generalization (DG) Techniques** fail to identify unknown objects.

☹ **Out-of-distribution (OOD) Detection Techniques** fail to generalize to unknown domains.

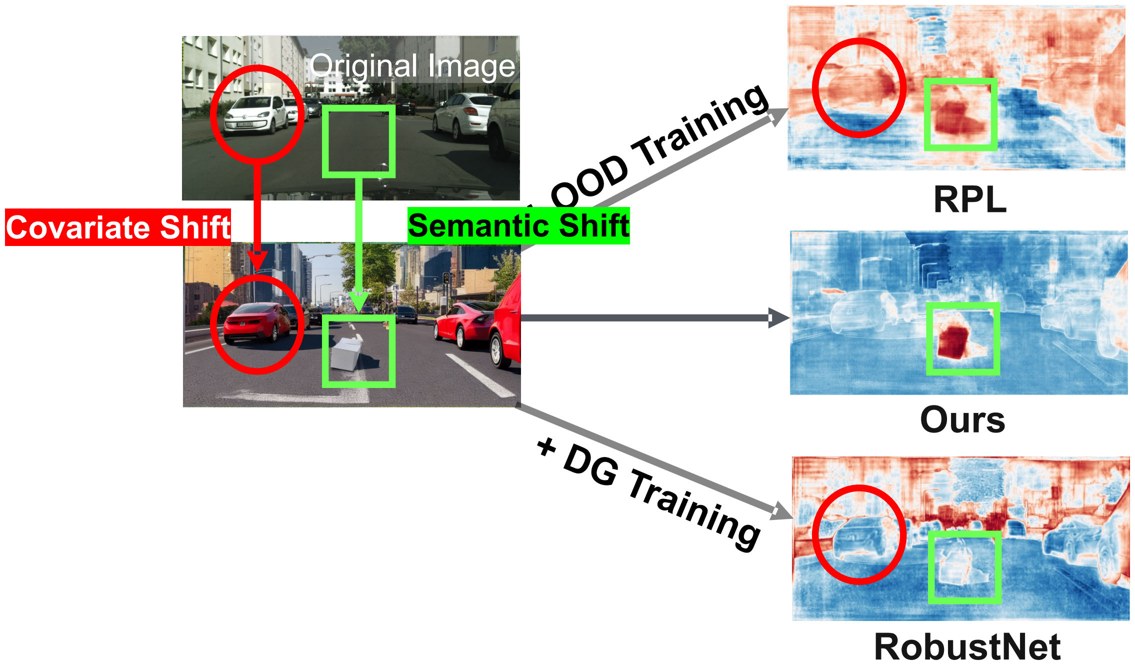
☹ **Simple Combination**: fail to distinguish two distribution shifts in object level.

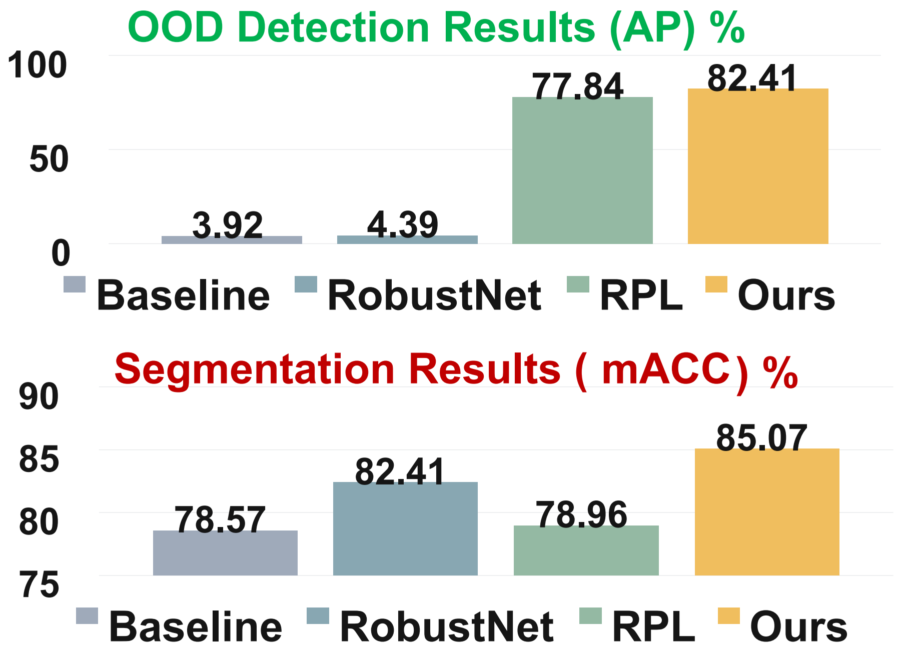


Our Goal **Semantic Segmentation Under Multiple Distribution Shifts.**

We jointly study both semantic and covariate shifts, so that models can:

- generalize effectively to covariate-shift regions, and
- precisely detect semantic-shift regions.



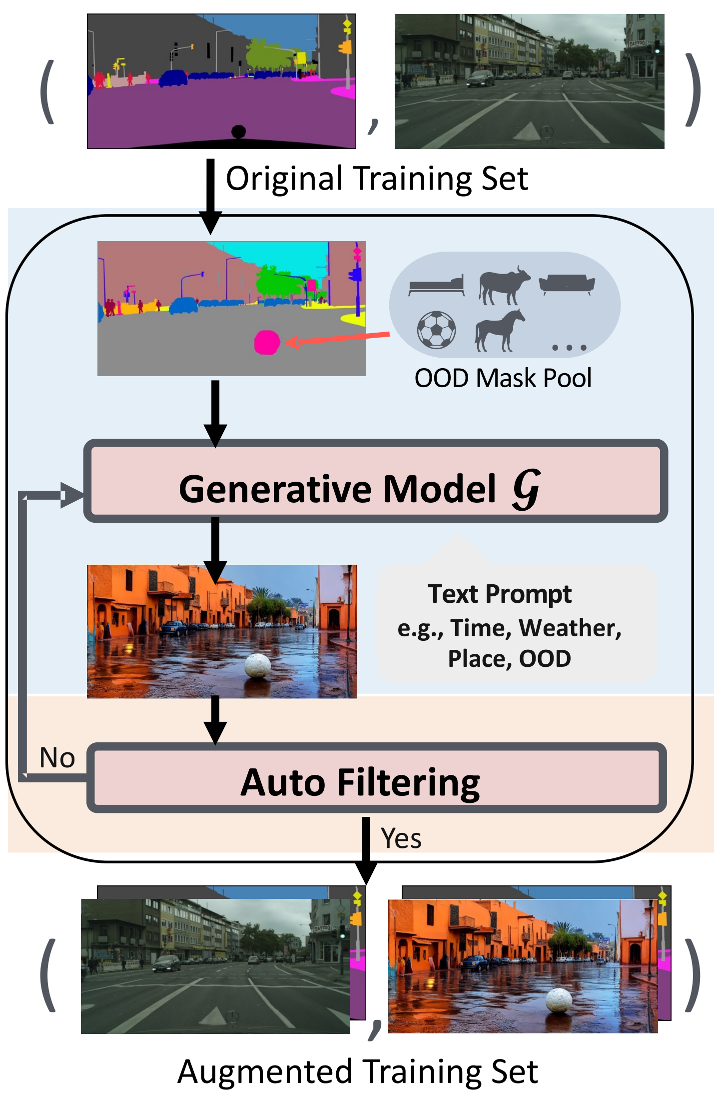


Main Idea

1. Augment training images with various semantic and covariate shifts at both image and object levels in a coherent way.
   - -> Coherent Generative-based Augmentation (CG-Aug)

2. Fully leverage the augmented data, so that the model can **distinguish** between the two types of distribution shifts and **respond appropriately** to each type.
   - -> Two-stage noise-aware training.

Coherent Generative-based Augmentation



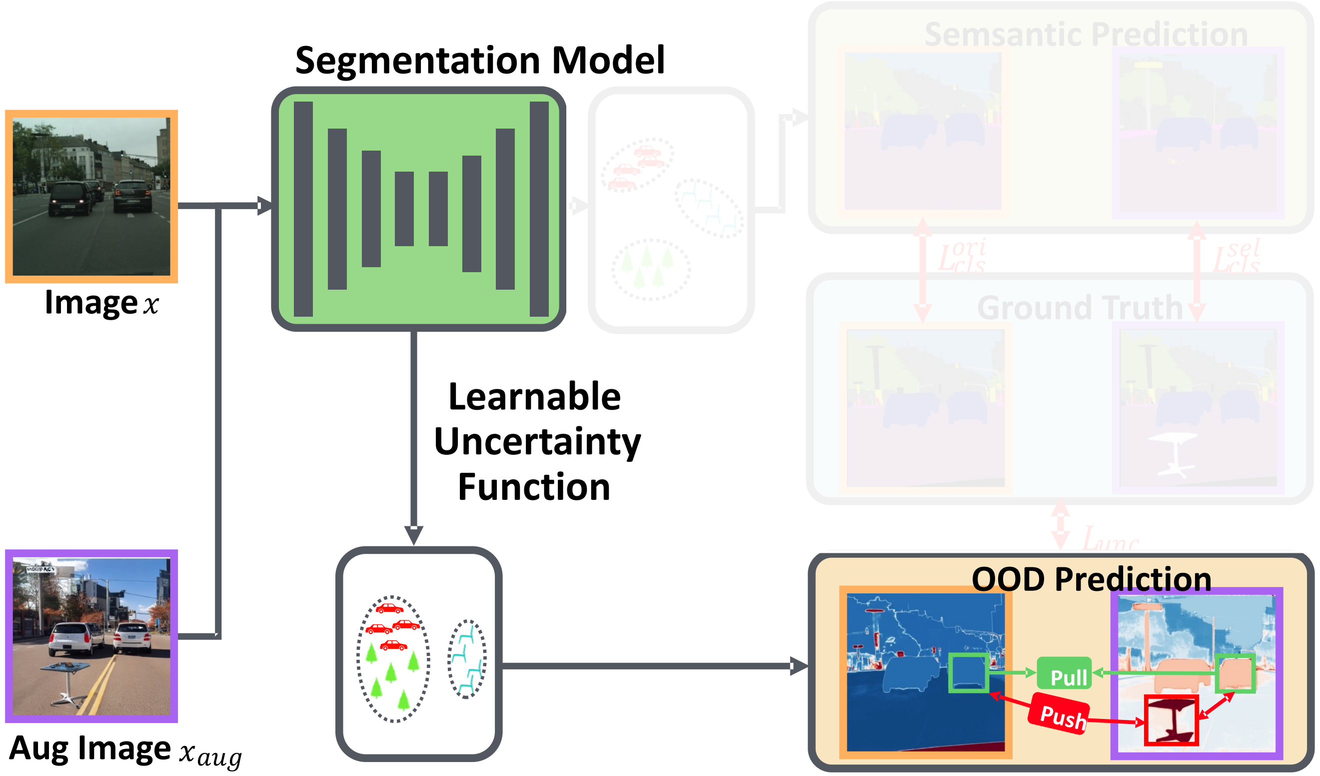
Stage 1: Zero-Shot Semantic-to-Image Generation:

A. **Cut-and-paste** the semantic mask of novel objects to the training labels.

B. **Semantic-to-image** generation via a pretrained generative model (E.g. ControlNet).

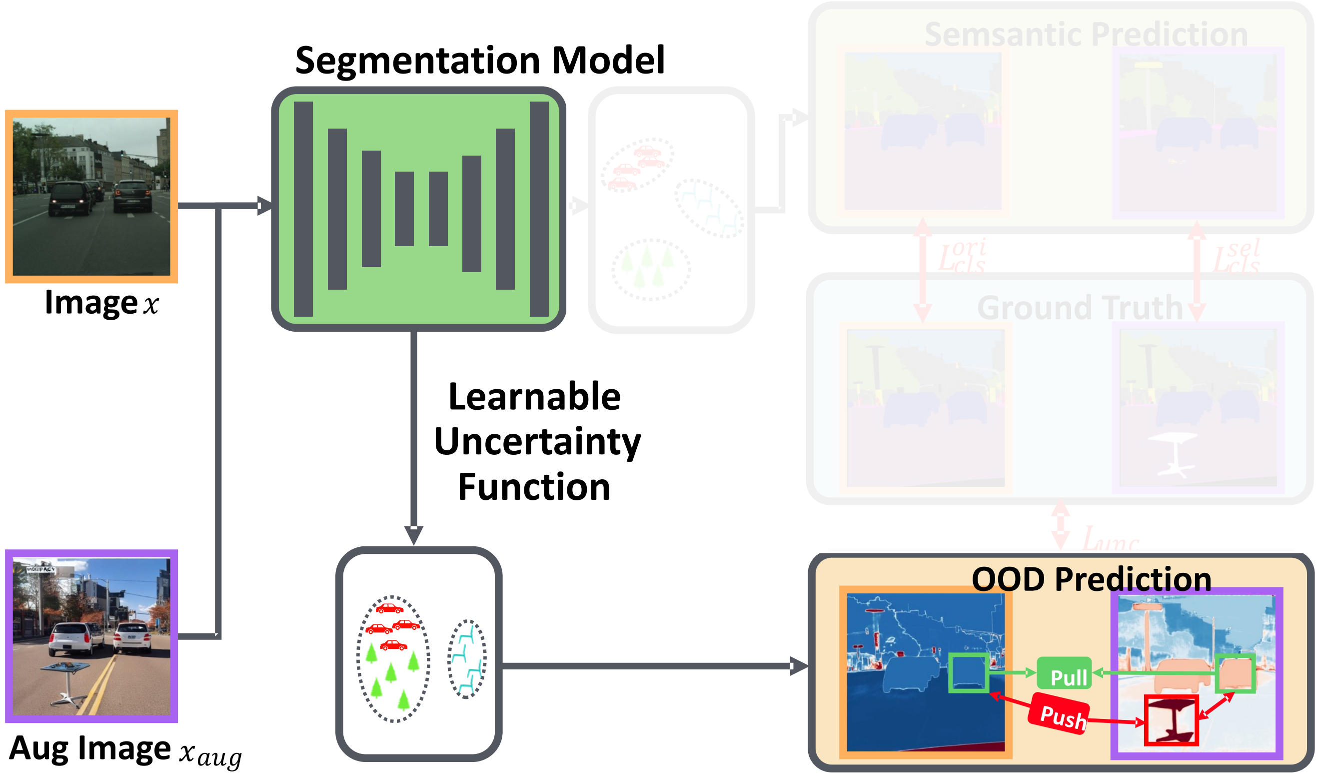
Stage 2: Automatic filter low-quality synthetic data.

Two-Stage Noise-Aware Training



Stage 1: Train a **semantic-exclusive uncertainty function** based on backbone features.

Two-Stage Noise-Aware Training



Stage 1: Train a <u>**semantic-exclusive uncertainty function**</u> based on backbone features.

1. Learnable Uncertianty Function:

$$u(x) = \log \sum_c \exp f(x) \boxed{W_c^o}.$$

Learnable Projection

- Initialize as energy score.

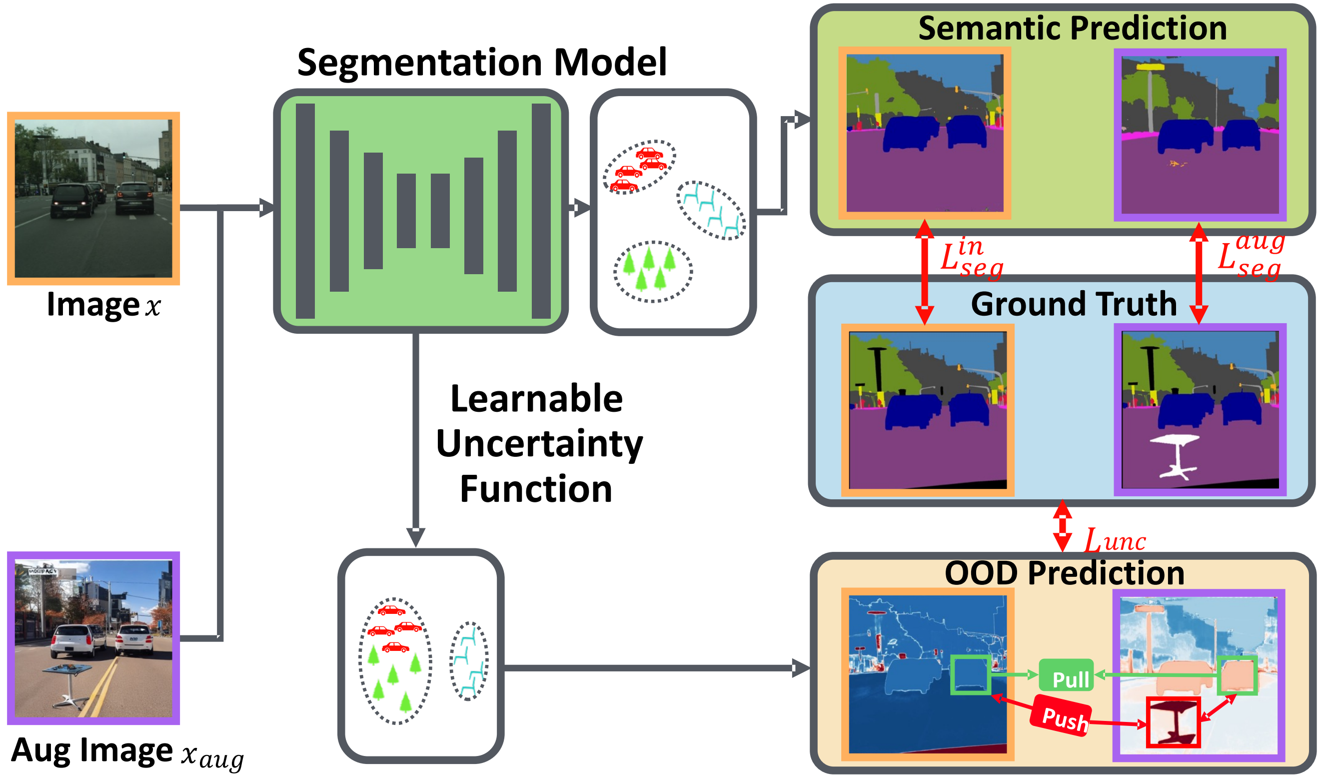
2. Relative Contrastive Loss: $\tau_\lambda(x) = \max(\lambda - x, 0)$

Push uncertinty score farther

Pull uncertinty score closer

$$L_{\text{unc}} = \boxed{\sum_{o\in\Omega^{\text{out}}, i\in\Omega^{\text{in}}} \tau_{\lambda_1}(u_o - u_i) + \sum_{o\in\Omega^{\text{out}}, c\in\Omega^{\text{aug}}} \tau_{\lambda_2}(u_o - u_c)} + \boxed{\sum_{c\in\Omega^{\text{aug}}, i\in\Omega^{\text{in}}} m_{c,i}\cdot\tau_{\lambda_3}(-(u_c - u_i))}$$

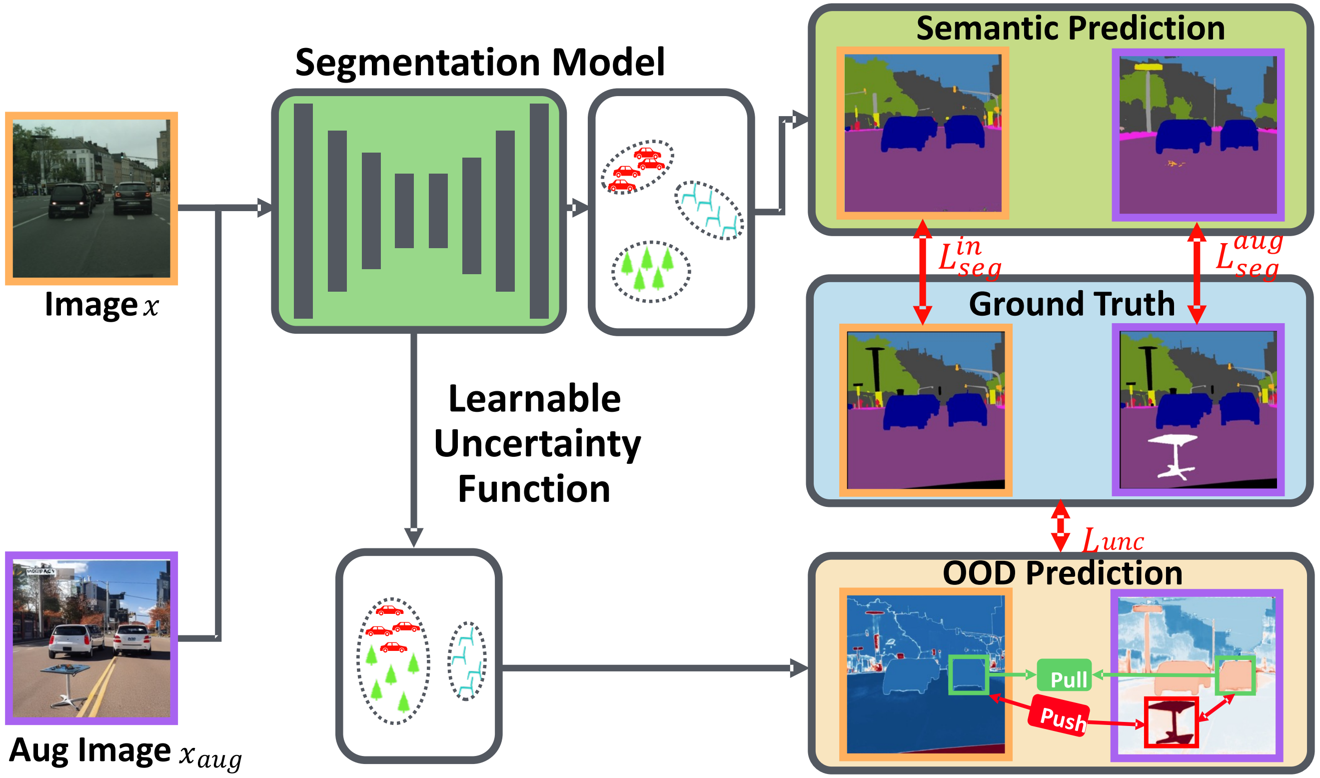
Two-Stage Noise-Aware Training



Stage 1: Train a **semantic-exclusive uncertainty function** based on backbone features.

Stage 2: **Fintune the feature extractor** to align features associated with domain shifts.

Two-Stage Noise-Aware Training



Stage 1: Train a **semantic-exclusive uncertainty function** based on backbone features.

Stage 2: **Fintune the feature extractor** to align features associated with domain shifts.

Overall Loss:

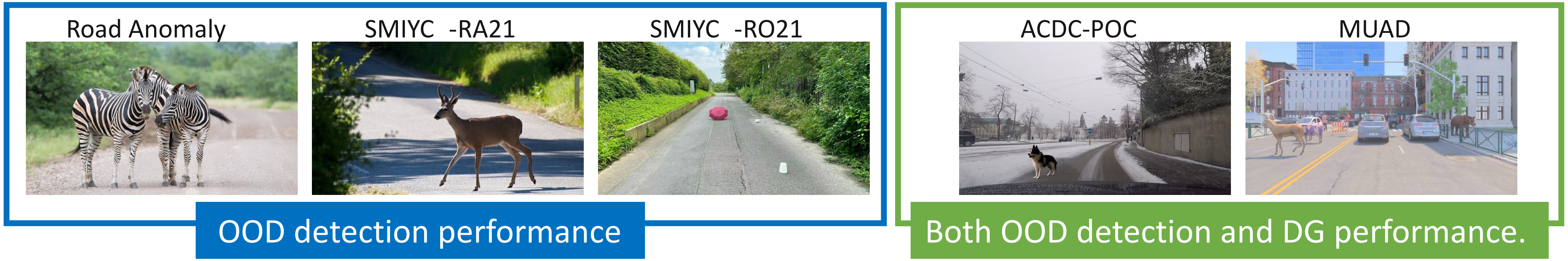
$$L = L_{\text{unc}} + \beta_1 L_{\text{seg}}^{\text{in}} + \beta_2 L_{\text{seg}}^{\text{aug}}.$$

$$L_{\text{seg}}(y, p, \eta) = \sum_i \boxed{\eta_i} \sum_c y_i^c \log p_i^c$$

Indicates whether a pixel i is selected.
Determined via 'small loss' criterion.

Experimental Setup

- **Implementation:** DeepLabv3+ and Mask2Former.
- **Datasets**:
  - Training set: Cityscapes.
  - Test set (below): All contain images with both semantic and domain shifts.



- **Metrics**: AUROC, AP, FPR@95, mAcc, mIoU

Results on Anomaly Segmentation Benchmarks

Table 1: **Results on anomaly segmentation benchmarks:** RoadAnomaly, SMIYC-RA21 and SMIYC-RO21. Our method achieves the best results under both backbones (Best results in Bold).

		RoadAnomaly			SMIYC - RA21		SMIYC - RO21	
Method	Backbone	AUC ↑	AP ↑	FPR_{95} ↓	AP↑	FPR_{95} ↓	AP ↑	FPR_{95} ↓
Maximum softmax [21]	DeepLabv3+	67.53	15.72	71.38	27.97	72.05	15.72	16.60
ODIN [28]	DeepLabv3+	-	-	-	33.06	71.68	22.12	15.28
Mahalanobis [26]	DeepLabv3+	62.85	14.37	81.09	20.04	86.99	20.90	13.08
Image resynthesis [30]	DeepLabv3+	-	-	-	52.28	25.93	37.71	4.70
SynBoost [13]	DeepLabv3+	81.91	38.21	64.75	56.44	61.86	71.34	3.15
Maximized entropy [6]	DeepLabv3+	-	48.85	31.77	85.47	15.00	85.07	0.75
PEBAL [46]	DeepLabv3+	87.63	45.10	44.58	49.14	40.82	4.98	12.68
Dense Hybrid [17]	DeepLabv3+	-	31.39	63.97	77.96	9.81	87.08	**0.24**
RPL+CoroCL [31]	DeepLabv3+	95.72	71.61	17.74	83.49	11.68	85.93	0.58
Ours	DeepLabv3+	**96.40**	**74.60**	**16.08**	**88.06**	**8.21**	**90.71**	0.26
Mask2Anomaly [42]	Mask2Former	-	79.70	13.45	88.7	14.60	93.3	0.20
RbA [36]	Mask2Former	-	85.42	6.92	90.90	11.60	91.80	0.50
M2F-EAM [18]	Mask2Former	-	69.40	7.70	**93.75**	**4.09**	92.87	0.52
Ours	Mask2Former	**97.94**	**90.17**	**7.54**	91.92	7.94	**95.29**	**0.07**

We achieve SOTA anomaly segmentation results with both backbones.

Results on ACDC-POC and MUAD

Table 2: **Results on ACDC-POC and MUAD**. Our model achieves the best performance in both anomaly segmentation (AP↑ , FPR↓) and domain-generalized segmentation (mIoU↑ , mAcc↑). Anomaly segmentation methods typically perform worse than the baseline for known class segmentation, while domain generalization methods fall below the baseline on OOD detection. (Best results are in bold; results below baseline are in blue.)

Method	Backbone	Technique		ACDC-POC				MUAD			
		OOD	DG	AP↑	FPR_{95} ↓	mIoU↑	mAcc↑	AP↑	FPR_{95} ↓	mIoU↑	mAcc↑
Baseline [7]	DeepLabv3+	-	-	3.92	55.50	46.89	78.57	1.34	72.78	29.47	68.63
RuleAug [45]	DeepLabv3+	-	✓	2.09	72.79	48.60	81.79	0.99	81.08	29.42	69.22
RobustNet [9]	DeepLabv3+	-	✓	4.39	62.65	47.41	82.41	2.27	58.64	**32.18**	72.02
PEBAL [46]	DeepLabv3+	✓	-	20.67	14.35	45.59	81.28	7.81	47.56	29.08	66.41
RPL [31]	DeepLabv3+	✓	✓	77.84	1.20	46.35	78.96	27.70	24.45	29.86	71.60
OOD + RuleAug [45]	DeepLabv3+	✓	✓	80.65	1.30	46.76	73.08	20.97	20.37	27.83	63.02
Ours	DeepLabv3+	✓	✓	**82.41**	**1.01**	**54.12**	**85.07**	**36.08**	**18.74**	31.33	**73.13**
Mask2Anomaly [42]	Mask2Former	✓	-	73.77	3.60	47.32	83.10	39.32	41.24	23.43	61.91
OOD + RuleAug [45]	Mask2Former	✓	✓	82.82	0.79	50.36	82.83	25.43	41.15	26.27	67.51
Ours	Mask2Former	✓	✓	**90.42**	**0.46**	**51.75**	**83.16**	**45.65**	**24.70**	**28.44**	**73.77**

Our method achieves the best results in both anomaly segmentation (OOD detection) and domain-generalized semantic segmentation.

Visualization of Uncertainty Maps

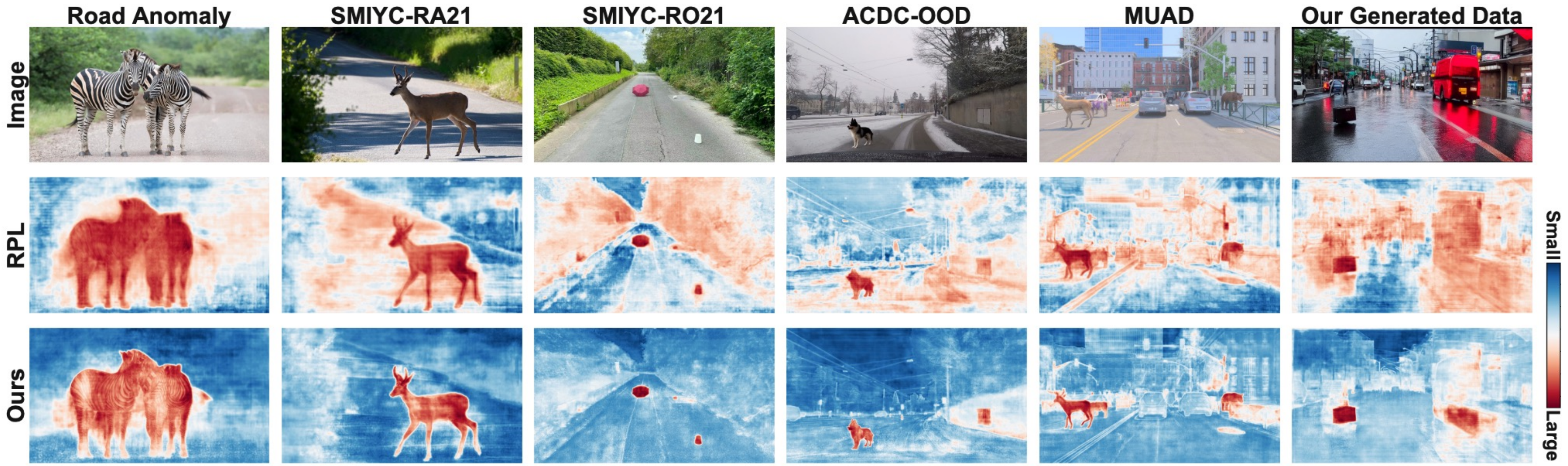


Figure 3: **Comparison of Uncertainty Maps.** Our method robustly detects anomalies under covariate shifts across five datasets (first five columns) and generated data (last column). The previous method RPL [31] failed to distinguish domain from semantic shifts, producing high uncertainty in both cases.

Our method produces semantic-exclusive uncertainty map.

Ablation Study

Table 3: **Impact of CG-Aug and Training Strategy.** The proposed coherent generative-based augmentation consistently enhances the previous OOD method, Mask2Anomaly [42] (M2A for short). Our fine-tuning strategy makes better use of the data and further boosts the performance.

		RoadAnomaly		SMIYC-RA Val		SMIYC-RO Val	
Training	Aug.	AP↑	$\text{FPR}_{95}\downarrow$	AP↑	$\text{FPR}_{95}\downarrow$	AP↑	$\text{FPR}_{95}\downarrow$
M2A [42]	Default	79.70	13.45	94.50	3.30	88.60	0.30
M2A [42]	Ours	85.47	22.38	**97.96**	1.55	89.80	**0.12**
Ours	Ours	**90.17**	**7.54**	97.31	**1.04**	**93.24**	0.14

Ablation Study

Table 3: **Impact of CG-Aug and Training Strategy.** The proposed coherent generative-based augmentation consistently enhances the previous OOD method, Mask2Anomaly [42] (M2A for short). Our fine-tuning strategy makes better use of the data and further boosts the performance.

		RoadAnomaly		SMIYC-RA Val		SMIYC-RO Val	
Training	Aug.	AP↑	FPR_{95} ↓	AP↑	FPR_{95} ↓	AP↑	FPR_{95} ↓
M2A [42]	Default	79.70	13.45	94.50	3.30	88.60	0.30
M2A [42]	Ours	85.47	22.38	**97.96**	1.55	89.80	**0.12**
Ours	Ours	**90.17**	**7.54**	97.31	**1.04**	**93.24**	0.14

Table 4: **Ablation Study of CG-Aug**. Generating data with both Semantic-shift (SS) and Domain-shift (DS) in a coherent manner achieves better results than other variations. The experiments were conducted using the Mask2Former backbone and evaluated on the RoadAnomaly dataset.

	AUC↑	AP↑	FPR_{95}↓
POC [12] (SS)	95.43	83.66	10.33
DS or SS	95.90	87.64	9.28
DS and SS	96.47	89.08	8.16
CG-Aug (Ours)	**97.94**	**90.17**	**7.54**

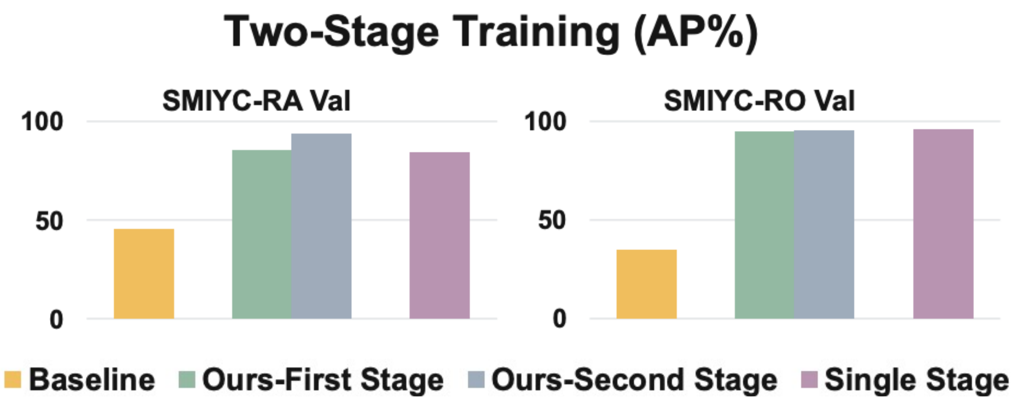
Ablation Study

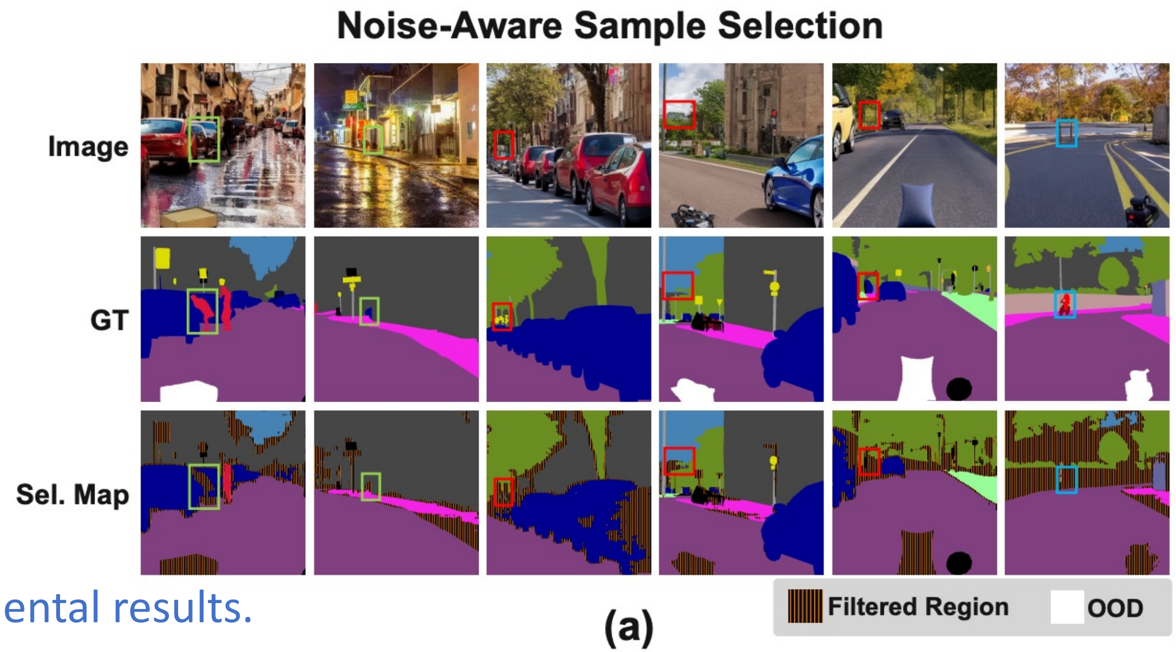
Table 3: **Impact of CG-Aug and Training Strategy.** The proposed coherent generative-based augmentation consistently enhances the previous OOD method, Mask2Anomaly [42] (M2A for short). Our fine-tuning strategy makes better use of the data and further boosts the performance.

		RoadAnomaly		SMIYC-RA Val		SMIYC-RO Val	
Training	Aug.	AP↑	FPR_{95} ↓	AP↑	FPR_{95} ↓	AP↑	FPR_{95} ↓
M2A [42]	Default	79.70	13.45	94.50	3.30	88.60	0.30
M2A [42]	Ours	85.47	22.38	**97.96**	1.55	89.80	**0.12**
Ours	Ours	**90.17**	**7.54**	97.31	**1.04**	**93.24**	0.14

Table 4: **Ablation Study of CG-Aug**. Generating data with both Semantic-shift (SS) and Domain-shift (DS) in a coherent manner achieves better results than other variations. The experiments were conducted using the Mask2Former backbone and evaluated on the RoadAnomaly dataset.

	AUC↑	AP↑	FPR_{95}↓
POC [12] (SS)	95.43	83.66	10.33
DS or SS	95.90	87.64	9.28
DS and SS	96.47	89.08	8.16
CG-Aug (Ours)	**97.94**	**90.17**	**7.54**





(a)

Please refer to our paper for further analysis and experimental results.

Thanks for listening !

For more information please refer to our paper and code.

Paper

Code