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

Zhitong Gao, Bingnan Li, Mathieu Salzmann, Xuming He

Paper Code

Background – Semantic Segmentation Under *Single* Distribution Shift

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

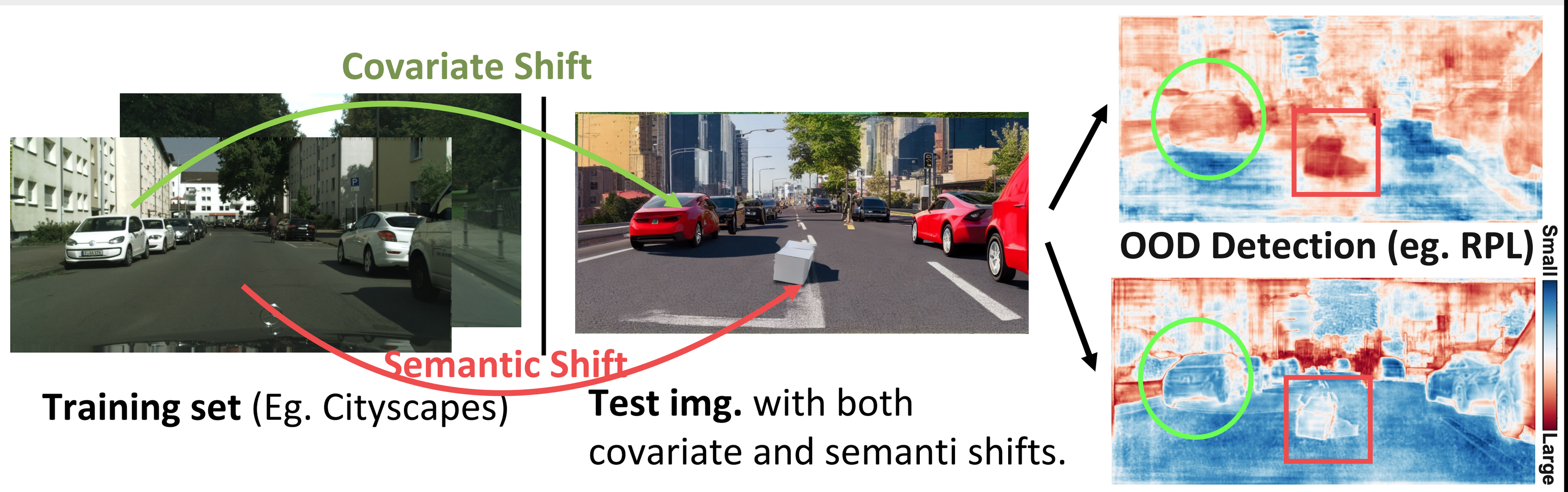
- e.g., different weather or object attributes.

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

- e.g., anomalies or novel objects.

Training set (Eg. Cityscapes) | **Test img.** with covariate shifts (Eg. ACDC) | **Test img.** with semantic shifts (Eg. SMIYC)

Motivation – Can a model jointly handle both kinds of distribution shifts?



☹ **DG Techniques** (eg. RobustNet) fail to identify unknown objects.

☹ **OOD Detection Techniques** (eg. RPL) fail to generalize to unknown domains.

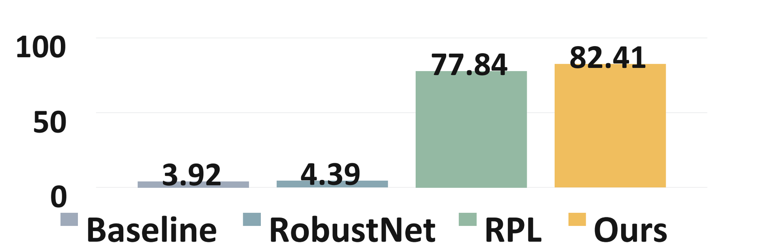
☹ **Simple Combination**: fail to distinguish two distribution shifts of object level.

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

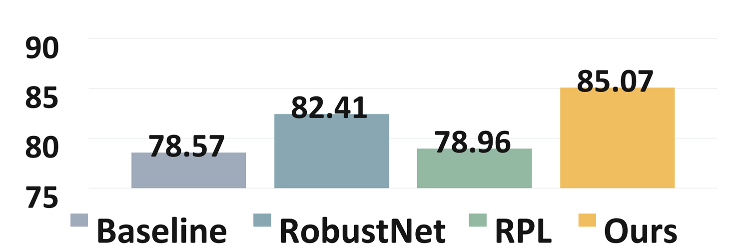
We jointly study both covariate and semantic shifts, so that models can:

☺ Distinguish between the two-types of distribution shifts.

☺ **Generalize** to covariate-shift regions and **detect** semantic-shift regions.

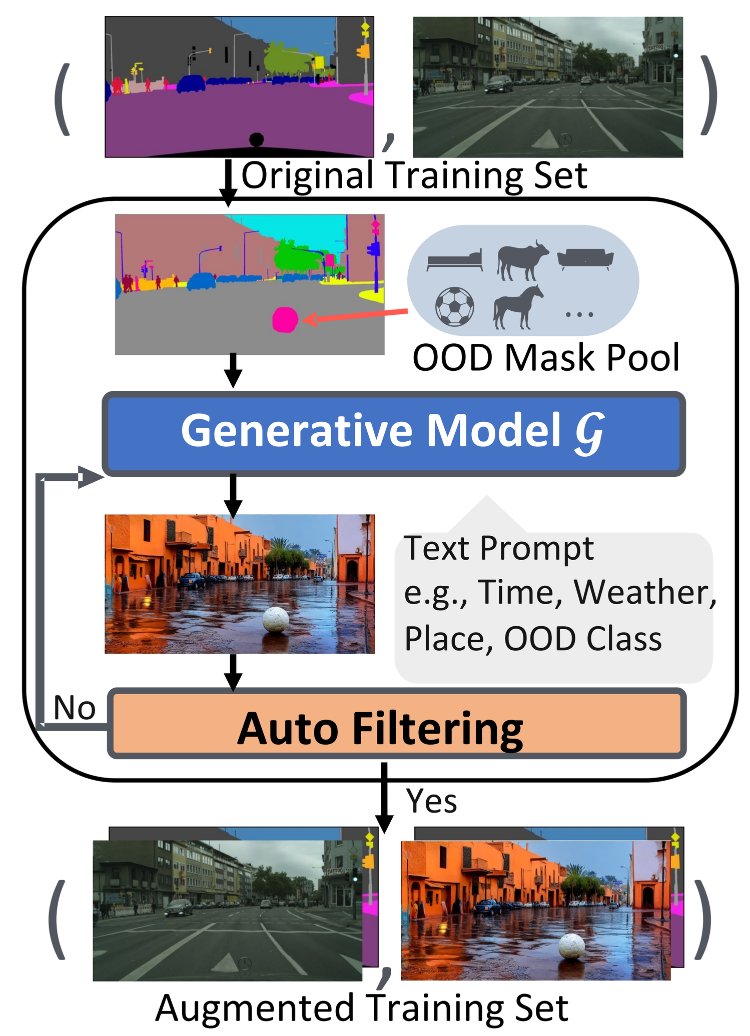


OOD Detection Results (AP) %



Segmentation Results (mAcc) %

Method – I. Coherent Generative-based Augmentation (CG-Aug)



➢ Goal: Augment training images with various semantic and covariate shifts at both image and object levels in a coherent way.

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

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

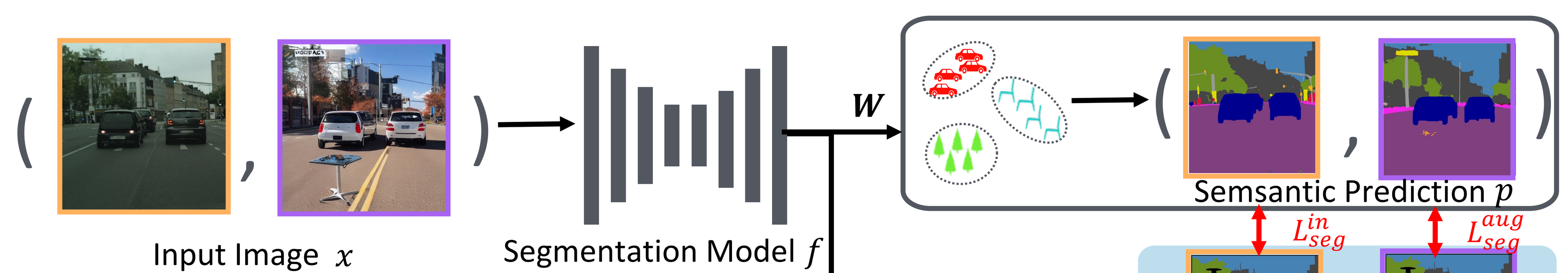
B. **Semantic-to-image** generation via a pretrained generative model (E.g. ControlNet).

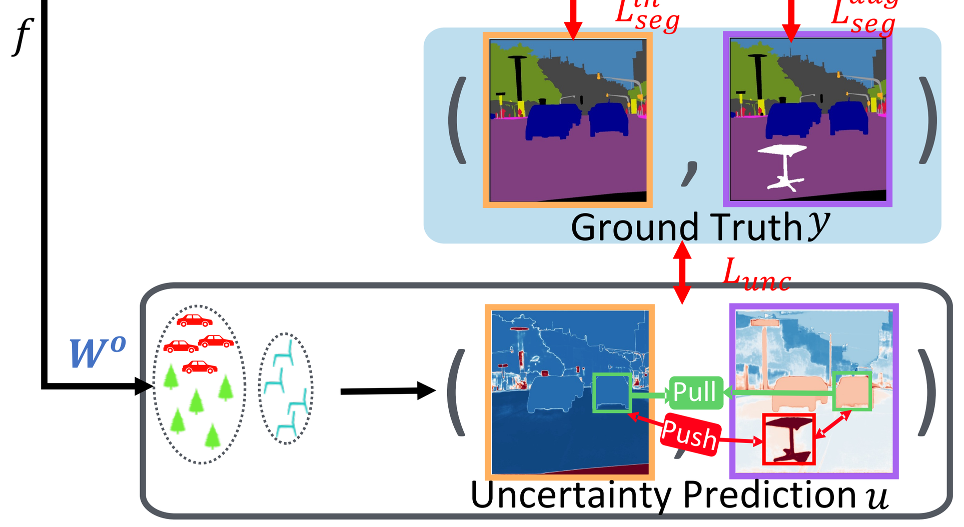
Stage 2: Automatically filtering low-quality synthetic data:

- Identify generation failures, such as missing objects or incorrectly generated known objects.

II. Uncertainty Recalibration & Model Training

➢ Goal: Fully leverage the augmented data, so that the model can **distinguish** between the two types of distribution shifts and **address** each type appropriately.





1. Semantic-Exclusive Uncertainty Recalibration

A. Learnable Uncertainty Function:

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

Learnable Projection

- Initialize as energy score.

B. Relative Contrastive Loss:

Push uncertainty with semantic shifts farther | Pull uncertainty with covariate shifts closer

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

Ω^{out}: Outlier pixel indices; Ω^{in} , Ω^{aug}: Inlier pixel indices from the original and augmented images, respectively; $\tau_\lambda(x) = \max(\lambda - x, 0)$

2. Two-Stage Noise-Aware Training

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

Noise-aware segmentation loss

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

η_i Indicates whether a pixel i is selected, and is determined via 'small loss' criterion..

Stage 2: Further fintune the feature extractor to improve feature representations of both known and OOD classes

Overall Loss:

$$L = L_{\text{unc}} + \beta_1 \underbrace{L_{\text{seg}}^{\text{in}}}_{L_{seg}(y,p,1)} + \beta_2 \underbrace{L_{\text{seg}}^{\text{aug}}}_{L_{seg}(y,p,\eta)}.$$

Note: our method can be applied to pixel-wise models (e.g. DeepLabv3+) or mask-wise models (e.g. Mask2Former). Please refer to our paper for details.

Experiments

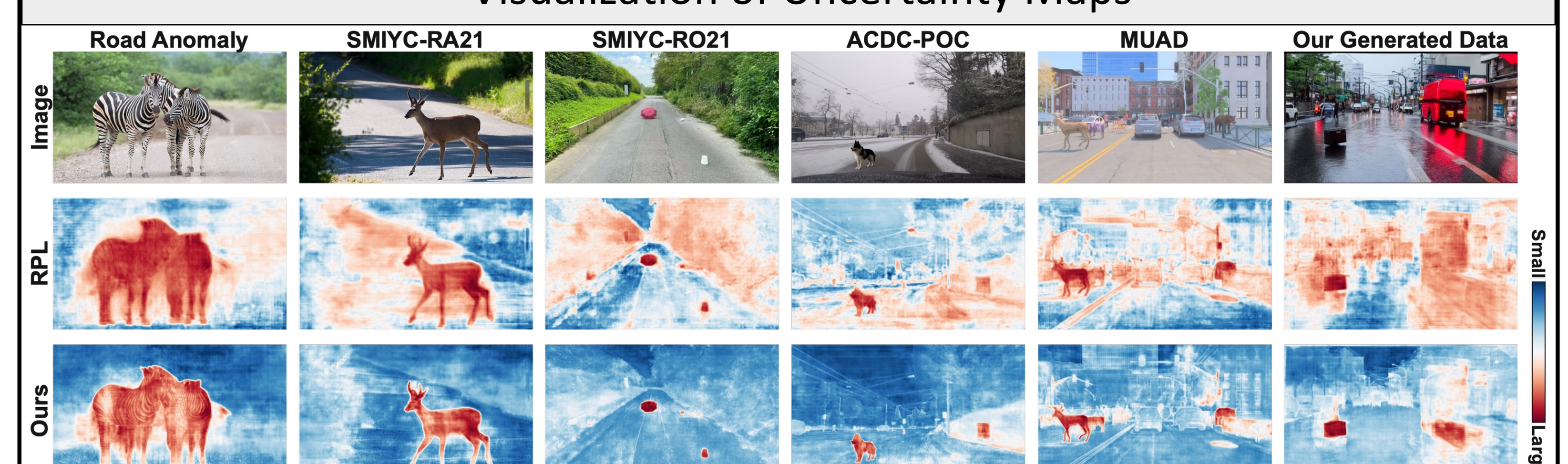
- Results on Anomaly Segmentation Benchmarks (**RoadAnomaly** & **SMIYC**)

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

- Results on **ACDC-POC** and **MUAD**

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

- Visualization of Uncertainty Maps



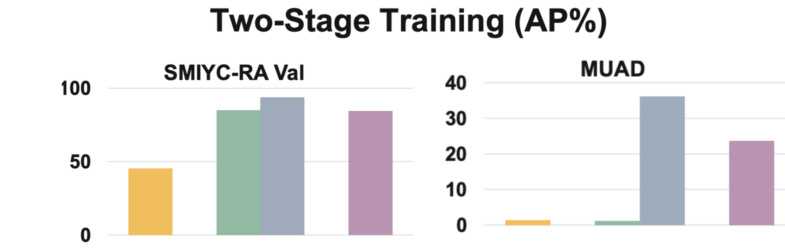
- Visualization of Generated Images & Selection Maps.

- Analysis & Ablation Study

	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**

SS: Semantic-shift; DS: Domain shift

Two-Stage Training (AP%)



Please refer to our paper for more results.