Random Walk on Pixel Manifolds for Anomaly Segmentation of Complex Driving Scenes

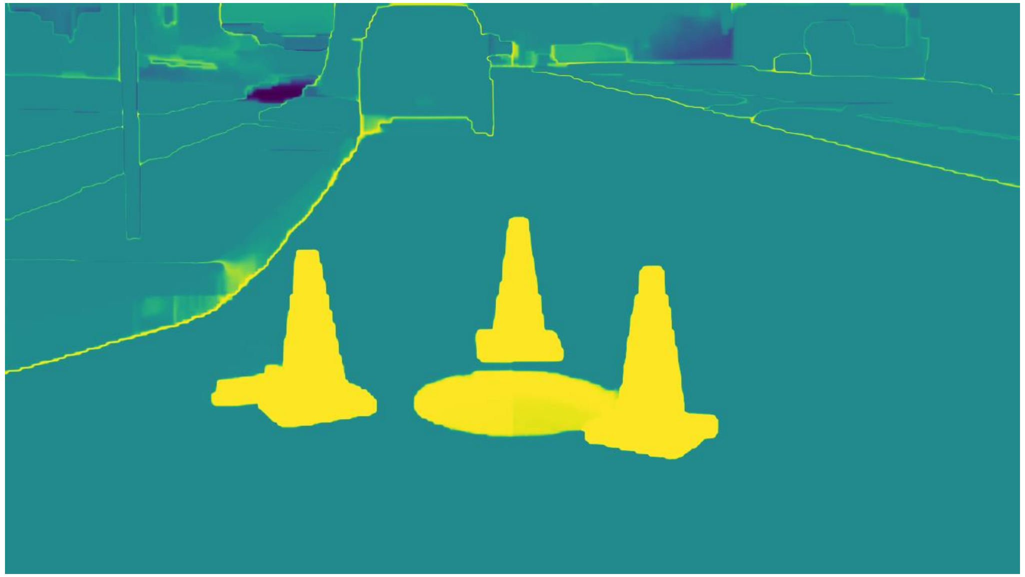
Zelong Zeng, Kaname Tomite
SenseTime Japan, Tokyo, Japan

The 18th European Conference on Computer Vision ECCV 2024

Anomaly Segmentation for Driving Scenes

- Given a driving scene image, anomaly segmentation detects all anomalous objects on the road.

A input image [1]

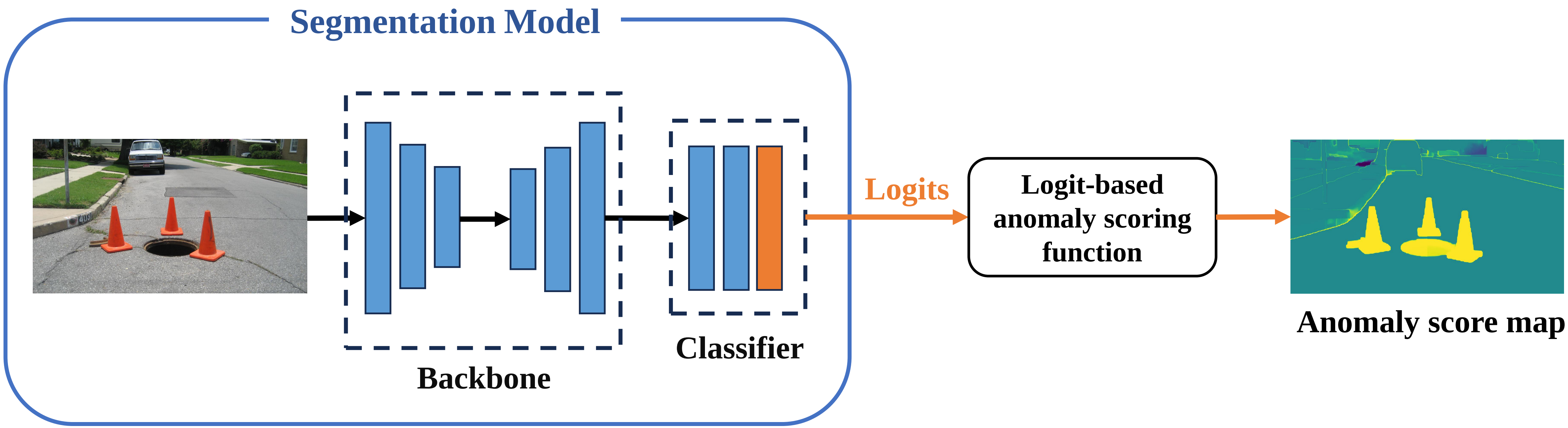


The result of anomaly segmentation

[1] Chan, Robin, et al. "Segmentmeifyoucan: A benchmark for anomaly segmentation." *NeurIPS 2021*.

Anomaly Segmentation for Driving Scenes

- Most existing state-of-the-art methods design and utilize anomaly scoring functions to calculate anomaly scores for each pixel.



Accurately predicting the logits is crucial for precisely inferring the anomaly score.

Motivation

- The diversity of driving scenes.

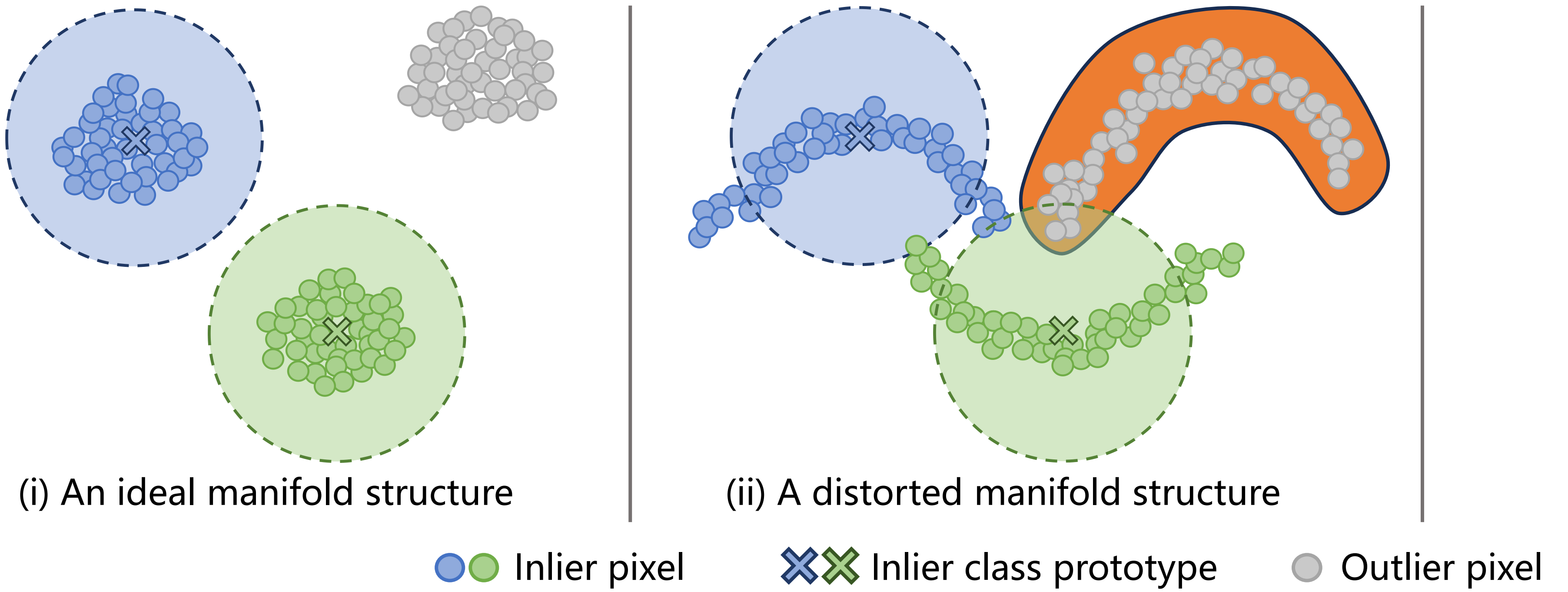
Training data [2]

Test data [3] in various environmental factors.

[2] Cordts, Marius, et al. "The cityscapes dataset for semantic urban scene understanding." *CVPR 2016*.
[3] Chan, Robin, et al. "Segmentmeifyoucan: A benchmark for anomaly segmentation." *NeurIPS 2021*.

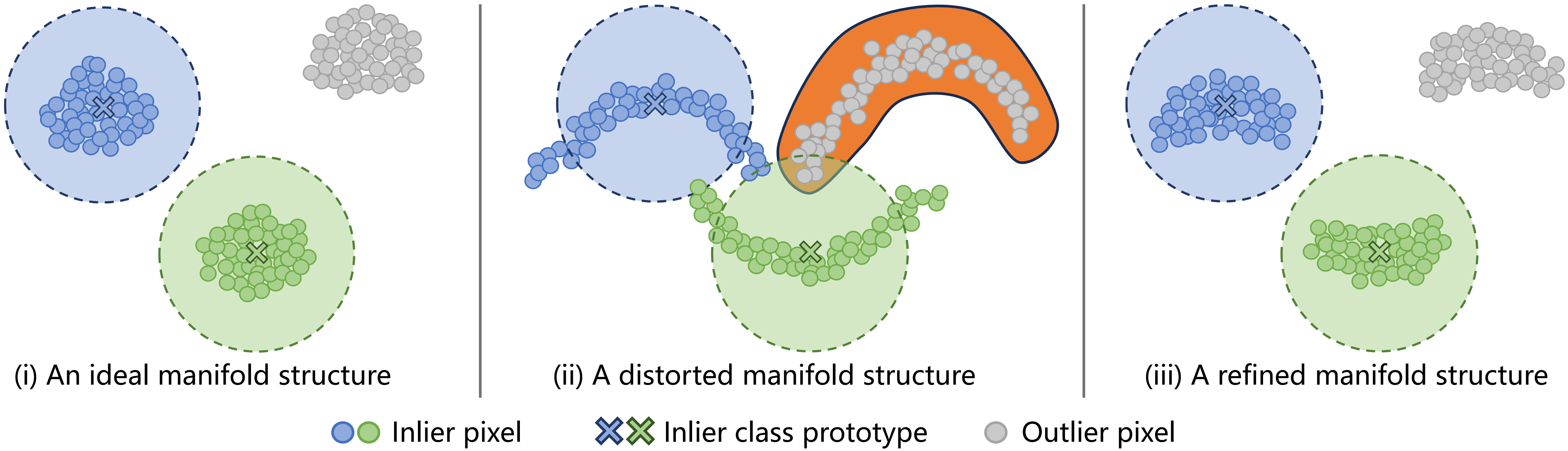
Motivation

- The diversity of driving scenes often result in distorted manifolds.



Motivation

- The structure of data manifolds implicitly contains the intrinsic relationship among pixel points.



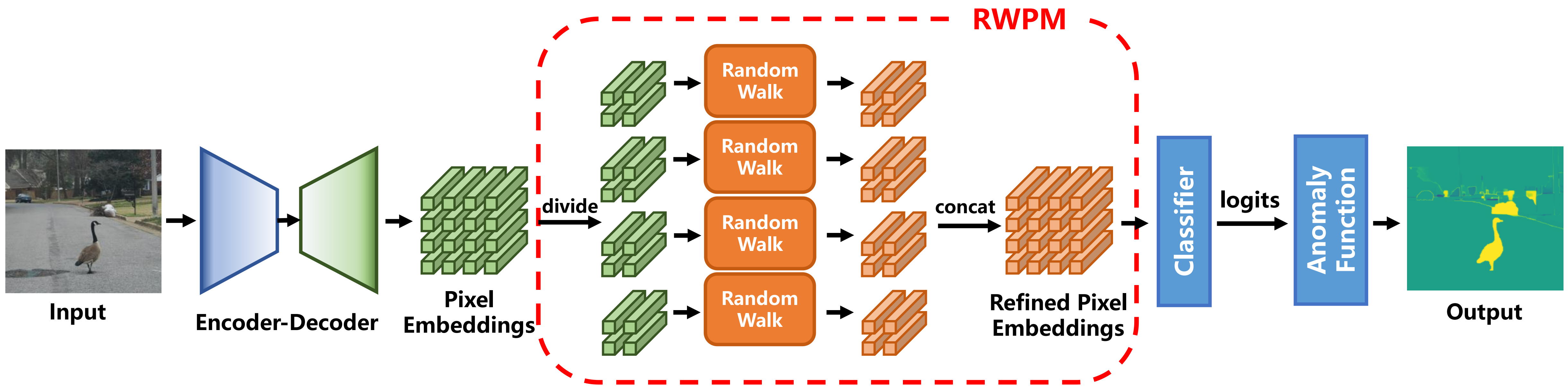
Proposed Method

- Random Walk on Pixel Manifolds (RWPM): RWPM can be directly integrated into existing anomaly segmentation frameworks in the inference phase.

Random Walk on Pixel Manifolds (RWPM)

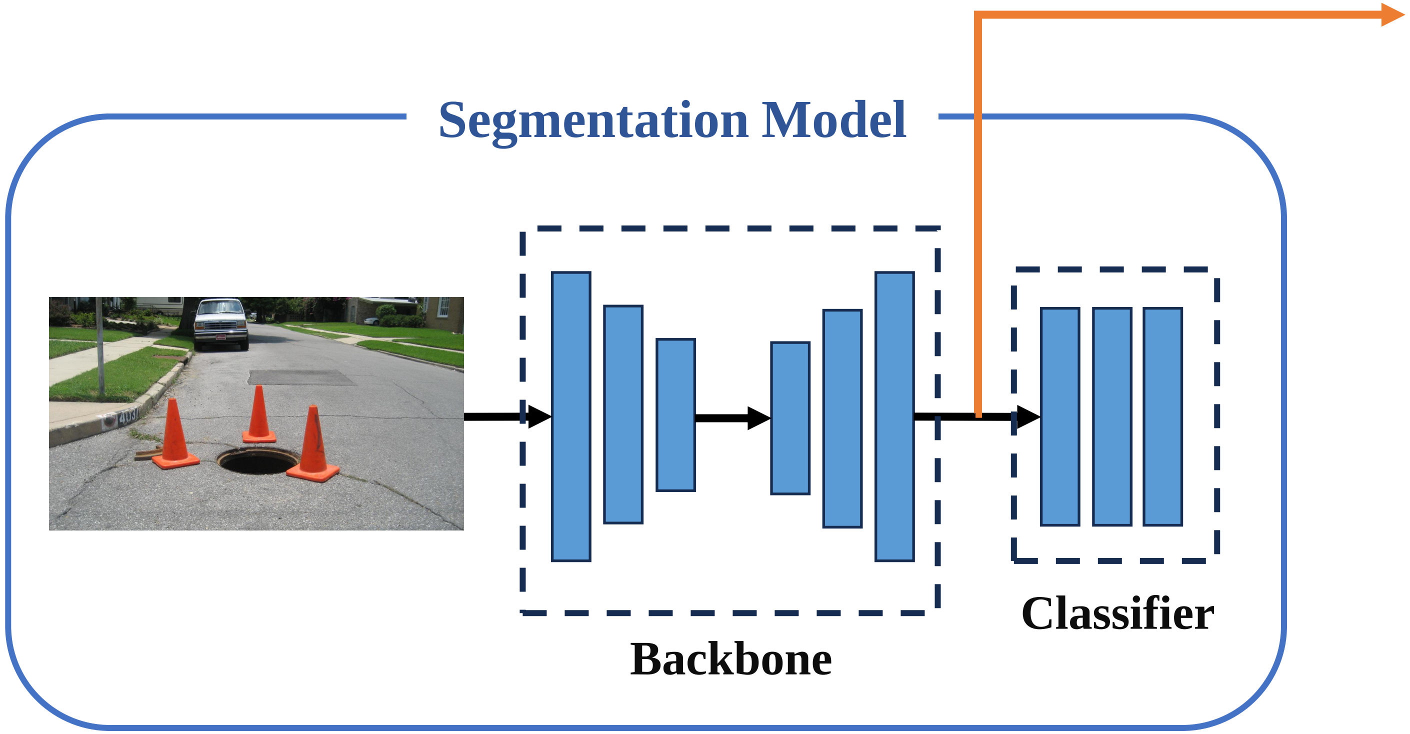
- **Overview**

RWPM consists of 3 parts: *Graph Construction*, *Random Walk Process* and *Partial Random Walk*



Random Walk on Pixel Manifolds (RWPM)

- Graph Construction



Pixel embeddings: $\boldsymbol{p} \in \mathbb{R}^{d\times H\times W}$

$\xrightarrow{\text{Reshape}}$ $\boldsymbol{p}^r \in \mathbb{R}^{HW\times d}$

$\xrightarrow{\text{Normalize}}$ $\widehat{\boldsymbol{p}}_i^r = \boldsymbol{p}_i^r / \|\boldsymbol{p}_i^r\|.$

Affinity matrix:

$$\mathbf{W}_{ij} = \begin{cases} \langle \widehat{\boldsymbol{p}}_i^r, \widehat{\boldsymbol{p}}_j^r \rangle & i \neq j \\ 0 & i = j \end{cases},$$

Locally constrained graph:

$$\mathbf{S}_{ij} = \begin{cases} \frac{\exp(\mathbf{W}_{ij}/\tau)}{\sum_{i=1, i\neq j}^{HW} \exp(\mathbf{W}_{ij}/\tau)} & i \neq j \\ 0 & i = j \end{cases},$$

Random Walk on Pixel Manifolds (RWPM)

- Random Walk Process

Initialization:

$$\mathbf{m}^0 \xrightarrow{\text{Initialize as}} \boldsymbol{p}^r$$

Iteration:

$$\mathbf{m}^{t+1} = \alpha \mathbf{S} \mathbf{m}^t + (1-\alpha)\,\mathbf{m}^0,\ \alpha \in (0,1),$$

$$\xrightarrow{\text{Closed-form}} \boxed{\mathbf{m}^{\infty}} = (1-\alpha)\,(\mathbf{I}-\alpha\mathbf{S})^{-1}\,\mathbf{m}^0,$$

Refined pixel embeddings

Random Walk on Pixel Manifolds (RWPM)

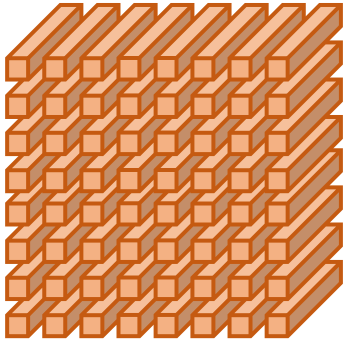
- Partial Random Walk

1024

512

524288

524288

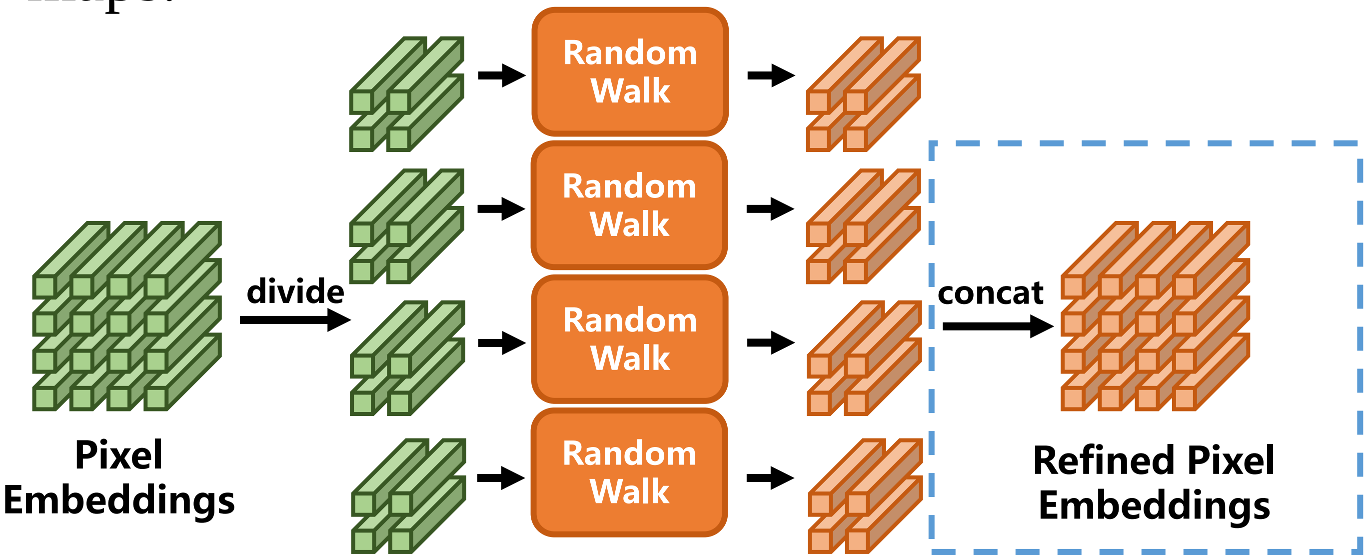


Random Walk on Pixel Manifolds (RWPM)

- Partial Random Walk
  1. Limited iteration: Iterate only T time

$$\mathbf{m}^{t+1} = \alpha \mathbf{S} \mathbf{m}^{t} + (1-\alpha)\,\mathbf{m}^{0},\ \alpha \in (0,1),$$

  2. Embedding map partitioning: Equally divide the pixel embedding map into some sub-maps.



Optional Step: Use calibration to smooth out the difference between patches.

Experiments

- We integrate RWPM with 4 representative anomaly segmentation methods.

Benchmark→		Fishyscapes Lost&Found			Road Anomaly		
Method ↓	Arch	AuROC↑	AP↑	FPR95↓	AuROC↑	AP↑	FPR95↓
PEBAL [38]	Pixel-base	98.96	58.81	4.77	87.63	45.10	44.58
PEBAL + RWPM† (Ours)	Pixel-base	**99.20**	**66.85**	**3.68**	**89.48**	**50.29**	**36.81**
Balanced Energy [12]	Pixel-base	99.03	67.07	2.93	88.31	41.48	41.46
Balanced Energy + RWPM† (Ours)	Pixel-base	**99.29**	**72.95**	**2.38**	**90.51**	**48.26**	**32.10**
Mask2Anomaly[1] [37]	Mask-base	95.41	69.44	9.22	96.54	80.04	13.95
Mask2Anomaly + RWPM (Ours)	Mask-base	**95.92**	**69.54**	**8.13**	**97.44**	**80.09**	**7.45**
RbA [35]	Mask-base	98.62	70.81	6.30	97.99	85.42	6.92
RbA + RWPM (Ours)	Mask-base	**98.82**	**71.16**	**6.12**	**98.04**	**87.34**	**5.27**

Experiments

- Pixel level evaluation: Comparison with state-of-the-art anomaly segmentation methods.

Benchmark→		Fishyscapes L&F		Road Anomaly		Anomaly Track		Obstacle Track		Average	
Method ↓	Arch	AP↑	FPR95↓	AP↑	FPR95↓	AP↑	FPR95↓	AP↑	FPR95↓	AP↑	FPR95↓
Synboost [15] (CVPR'21)	Pixel-based	60.58	31.02	41.84	59.72	56.44	61.86	71.34	3.15	57.46	38.94
SML [26] (ICCV'21)	Pixel-based	22.74	33.49	25.82	49.47	46.8	39.5	3.4	36.8	24.69	39.82
Meta-OOD [7] (ICCV'21)	Pixel-based	41.31	37.69	48.84	31.77	85.47	15.00	85.07	0.75	65.13	21.30
Learning Embedding [5](IJCV'21)	Pixel-based	4.65	24.36	-	-	37.52	70.76	0.82	46.38	-	-
Void Classifier [5](IJCV'21)	Pixel-based	10.29	22.11	-	-	36.61	63.49	10.44	41.54	-	-
GMMSeg-DL [28] (NeurIPS'22)	Pixel-based	43.47	13.11	34.42	47.90	-	-	-	-	-	-
DenseHybrid [18] (ECCV'22)	Pixel-based	69.79	5.09	31.39	63.97	77.96	**9.81**	87.08	0.24	65.56	19.78
PEBAL [38] (ECCV'22)	Pixel-based	58.81	4.77	45.10	44.58	49.14	40.82	4.98	12.68	39.58	25.71
Balanced Energy [12] (CVPR'23)	Pixel-based	67.07	2.93	41.48	41.46	-	-	-	-	-	-
RPL-CoroCL [33] (ICCV'23)	Pixel-based	70.61	**2.52**	71.61	17.74	83.49	11.68	85.93	0.58	77.91	8.13
Mask2Anomaly [37] (ICCV'23)	Mask-based	69.44	9.22	80.04	13.95	88.62	14.57	93.10	**0.20**	82.87	9.49
M2F-EAM [19] (CVPRW'23)	Mask-based	52.03	20.51	66.67	13.42	76.3	93.9	66.9	17.9	65.48	36.43
RbA [35] (ICCV'23)	Mask-based	70.81	6.30	85.42	6.92	90.9	11.6	91.8	0.5	84.73	6.33
RbA + RWPM (Ours)	Mask-based	**71.16**	6.12	**87.34**	**5.27**	**92.20**	10.15	**93.30**	0.28	**86.00**	**5.46**

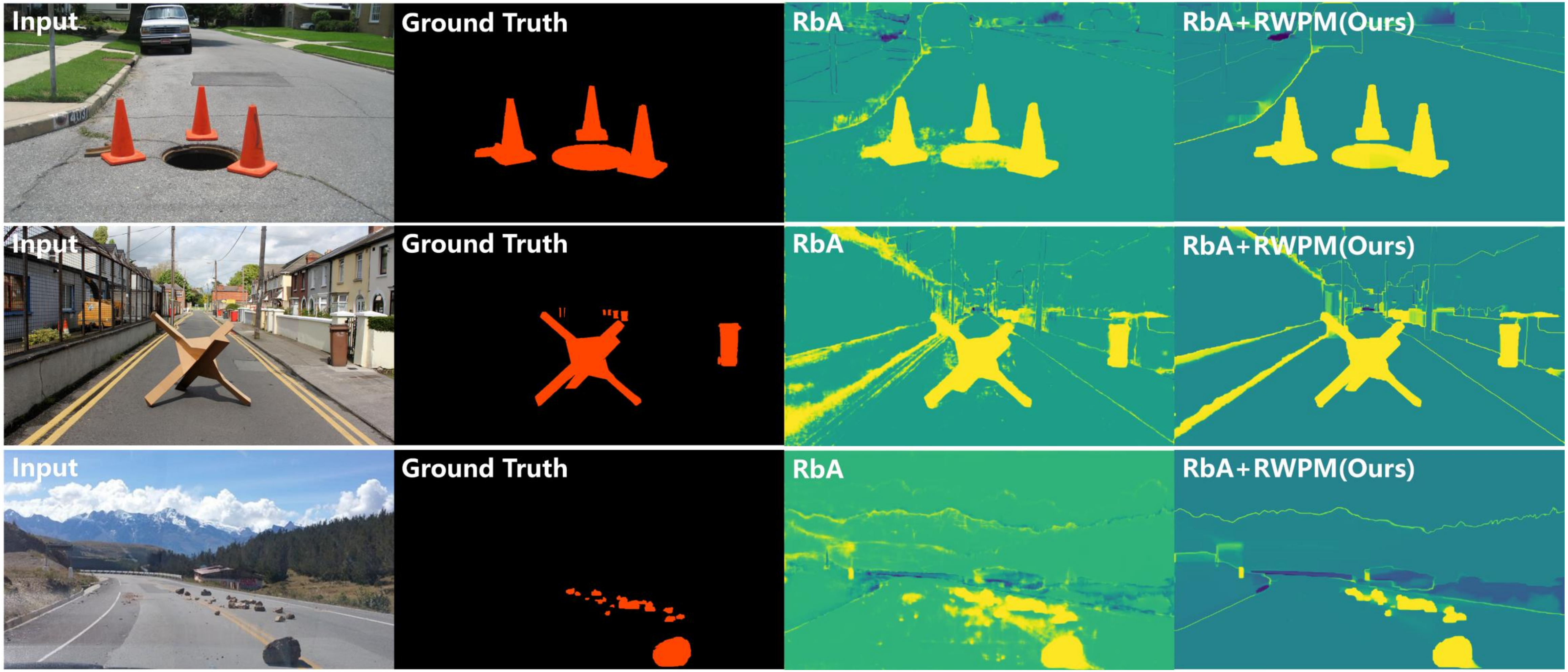
Experiments

- Component level evaluation: Comparison with state-of-the-art anomaly segmentation methods.

Benchmark→		Anomaly Track			Obstacle Track		
Method ↓	Arch	sIoU ↑	PPV↑	mean F1↑	sIoU ↑	PPV↑	mean F1↑
Synboost [15] (CVPR'21)	Pixel-based	34.68	17.81	9.99	44.28	41.75	37.57
SML [26] (ICCV'21)	Pixel-based	26.00	24.70	12.20	5.10	13.30	3.00
Meta-OOD [7] (ICCV'21)	Pixel-based	49.31	39.51	28.72	47.87	63.64	48.51
Void Classifier [5](IJCV'21)	Pixel-based	21.14	22.13	6.49	6.34	20.27	5.41
Learning Embedding [5](IJCV'21)	Pixel-based	33.86	20.54	7.90	35.64	2.87	2.31
JSRNet [39] (ICCV'21)	Pixel-based	20.20	29.27	13.66	18.55	24.46	11.02
DenseHybrid [18] (ECCV'22)	Pixel-based	54.17	24.13	31.08	45.74	50.10	50.72
PEBAL [38] (ECCV'22)	Pixel-based	38.88	27.20	14.48	29.91	7.55	5.54
RPL-CoroCL [33] (ICCV'23)	Pixel-based	49.76	29.96	30.16	52.61	56.65	56.69
Mask2Fromer [37] (ICCV'23)	Mask-based	25.20	18.20	15.30	5.00	21.90	4.80
Maskomaly [2] (BMVC'23)	Mask-based	55.4	51.6	50.0	-	-	-
RbA [35] (ICCV'23)	Mask-based	55.69	52.14	46.80	58.36	58.78	60.85
RbA + RWPM (Ours)	Mask-based	**57.00**	**61.25**	**58.44**	**59.47**	**72.51**	**69.85**

Experiments

- Qualitative Result.



Thank you !

The Code Link

The 18th European Conference on Computer Vision ECCV 2024