Generalized Diffusion Detector - Mining Robust Features from Diffusion Models for Domain-Generalized Detection

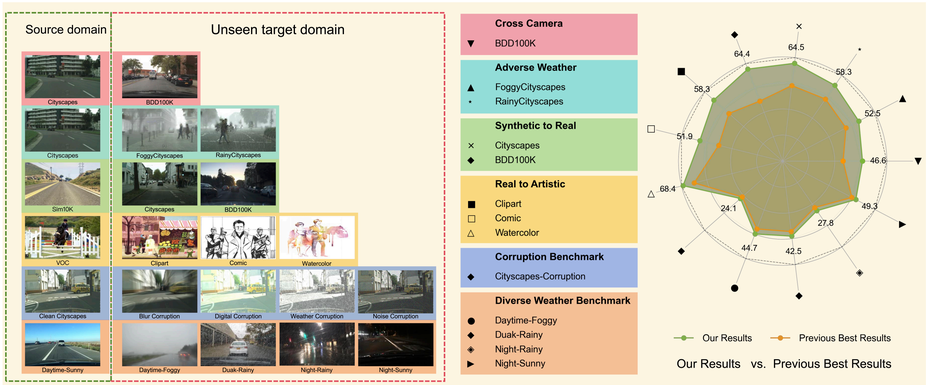
Boyong He, Yuxiang Ji, Qianwen Ye, Zhuoyue Tan and Liaoni Wu*. **Xiamen University**

2026 Ph.D. candidate seeking research internship !!!

CVPR Nashville JUNE 11-15, 2025

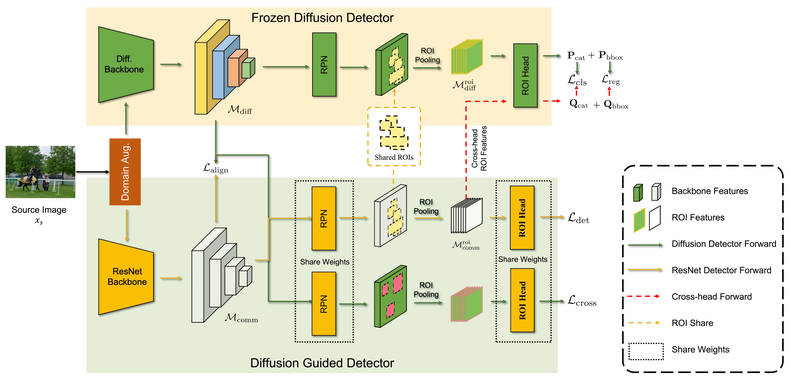
Benchmarks and Comparisons

- We test on 13 datasets across 6 DG detection benchmarks: ***Cross Camera***, ***Adverse Weather***, ***Syn2Real***, ***Real2Artistic***, ***Corruption Benchmark***, and ***Diverse Weather Benchmark***.
- Our method outperforms previous best results across all datasets.



Diff. Guided Framework

- The framework consists of a ***frozen diffusion detector*** and a **trainable ordinary detector**. Knowledge transfer is achieved through **feature- and object-level alignment**.
- Our method enhances cross-domain generalization **without adding inference overhead**.



Objective

Feature-level Alignment

$$\mathcal{L}_{\text{align}} = \sum_{l=1}^{L} \frac{1}{N_l} \|\hat{\mathcal{M}}^l_{\text{comm}} - \hat{\mathcal{M}}^l_{\text{diff}}\|_2^2$$

$$\mathcal{L}_{\text{cross}} = \mathcal{L}^{\text{rpn}}_{\text{comm}}(\mathcal{M}_{\text{diff}}; \theta_{\text{comm}}) + \mathcal{L}^{\text{roi}}_{\text{comm}}(\mathcal{M}_{\text{diff}}; \theta_{\text{comm}})$$

Object-level Alignment

$$\mathcal{L}_{\text{cls}} = \frac{1}{N}\sum_{i=1}^{N} \tau^2 D_{KL}(\mathbf{Q}^i_{\text{cat}} \| \mathbf{P}^i_{\text{cat}})$$

$$\mathcal{L}_{\text{reg}} = \frac{1}{N}\sum_{i=1}^{N} |\mathbf{Q}^i_{\text{bbox}} - \mathbf{P}^i_{\text{bbox}}|_1$$

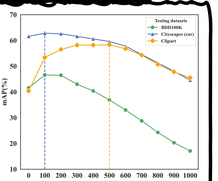
Full Objective

$$\mathcal{L}_{\text{total}} = \underbrace{\mathcal{L}_{\text{det}}(\mathcal{F}_{\text{comm}}(\mathbf{x}_s), y_s)}_{\text{supervised learning on source domain}} + \lambda_{\text{feature}} \underbrace{(\mathcal{L}_{\text{align}} + \mathcal{L}_{\text{cross}})}_{\text{feature-level alignment}} + \lambda_{\text{object}} \underbrace{(\mathcal{L}_{\text{cls}} + \mathcal{L}_{\text{reg}})}_{\text{object-level alignment}}$$

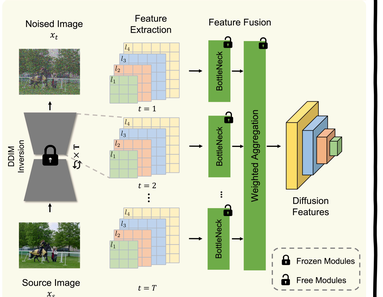
Diff. Backbone

Performance with Different *Timesteps*

T	BDD 100K	Cityscapes (car)	Clipart	Inference Time (ms)
1	28.6	49.8	37.4	270
2	34.9	54.1	48.6	404
5	46.6	**62.8**	58.3	789
10	**47.1**	62.6	**58.9**	1,424
20	45.6	61.4	57.7	2,820



Feature Collection and Fusion Pipeline



Main Results

Table 1. Cross Camera DG and DA Results (%) on BDD100K.

Methods	Bike	Bus	Car	Motor	Psn.	Rider	Truck	mAP
DG methods *(without target data)*								
CDSD [79] *(CVPR'22)*	22.9	20.5	33.8	14.7	18.5	23.6	18.2	21.7
SHADE [91] *(ECCV'22)*	25.1	19.0	36.8	18.4	24.1	24.9	19.8	24.0
SRCD [63] *(TNNLS'24)*	24.8	21.5	38.7	19.0	25.7	28.4	23.1	25.9
MAD [84] *(CVPR'23)*	-	-	-	-	-	-	-	28.0
DA methods *(with unlabeled target data)*								
TDD [28] *(CVPR'22)*	28.8	25.5	53.9	24.5	39.6	38.9	24.1	33.6
PT [10] *(ICML'22)*	28.8	**33.8**	52.7	23.0	40.5	39.9	25.8	34.9
SIGMA [48] *(CVPR'22)*	26.3	23.6	64.1	17.9	46.9	29.6	20.2	32.7
SIGMA++ [49] *(TPAMI'23)*	27.1	26.3	65.6	17.8	47.5	30.4	21.1	33.7
NSA [96] *(ICCV'23)*	-	-	-	-	-	-	-	35.5
HT [18] *(CVPR'23)*	38.0	30.6	63.5	28.2	53.4	40.4	27.4	40.2
Ours (DG settings)								
Diff. Detector (SD-1.5)	**38.9**	31.0	71.5	37.6	61.5	**47.0**	**38.5**	**46.6**
Diff. Detector (SD-2.1)	38.0	33.6	69.9	36.6	**62.1**	46.3	34.2	45.8
Diff. Guided (SD-1.5)	38.4	33.4	**72.0**	**38.3**	60.3	47.0	35.0	46.3+20.9
Diff. Guided (SD-2.1)	38.5	32.6	71.8	37.5	60.2	46.7	35.3	46.1+20.7

Table 2. Adverse Weather DG and DA Results (%) on FoggyCityscapes.

Methods	Bus	Bike	Car	Motor	Psn.	Rider	Train	Truck	mAP
DG methods									
FACT [85] *(CVPR'21)*	27.7	31.3	35.9	23.3	26.2	41.2	3.0	13.6	25.3
FSDR [34] *(CVPR'22)*	36.6	34.1	43.3	27.1	31.2	44.4	11.9	19.3	31.0
MAD [84] *(CVPR'23)*	44.0	40.1	45.0	30.3	34.2	47.4	42.4	25.6	38.6
DA methods									
MGA [95] *(CVPR'22)*	53.2	36.9	61.5	27.9	43.1	47.3	50.3	30.2	43.8
MTTrans [89] *(CVPR'22)*	45.9	46.5	65.2	32.6	47.7	49.9	33.8	25.8	43.4
OADA [87] *(CVPR'22)*	48.5	39.8	62.9	34.3	47.8	46.5	50.9	32.1	45.4
MIC [31] *(CVPR'23)*	52.4	47.5	67.0	40.6	50.9	55.3	33.7	33.9	47.6
SIGMA++ [49] *(TPAMI'23)*	52.2	39.9	61.0	34.8	46.4	45.1	44.6	32.1	44.5
CIGAR [52] *(CVPR'23)*	56.6	41.3	62.1	33.7	46.1	47.3	44.3	27.8	44.9
CMT [3] *(CVPR'23)*	**66.0**	51.2	63.7	41.4	45.9	55.7	38.8	**39.6**	50.3
HT [18] *(CVPR'23)*	55.9	50.3	**67.5**	40.1	**52.1**	55.8	49.1	32.7	50.4
Ours (DG settings)									
Diff. Detector (SD-1.5)	56.2	50.4	66.7	39.9	50.2	59.5	39.9	38.0	50.1
Diff. Detector (SD-2.1)	55.5	49.6	67.0	40.4	50.4	58.2	29.2	36.4	48.3
Diff. Guided (SD-1.5)	53.8	**54.2**	**67.5**	**45.6**	**52.1**	**60.8**	**53.9**	32.4	**52.5**+21.8
Diff. Guided (SD-2.1)	55.1	53.9	67.0	43.4	51.9	59.5	42.2	34.8	51.0+20.8

Table 3. Adverse Weather DG and DA Results (%) on RainyCityscapes.

Methods	mAP
DG methods	
FACT [85] *(CVPR'21)*	39.9
FSDR [34] *(CVPR'22)*	42.8
SCG [84] *(CVPR'23)*	39.1
MAD [84] *(CVPR'23)*	42.3
DA methods	
MGA [95] *(CVPR'22)*	43.0
TDD [28] *(CVPR'23)*	50.3
CMT [3] *(CVPR'23)*	52.1
SIGMA++ [49] *(TPAMI'23)*	46.9
Ours (DG settings)	
Diff. Detector (SD-1.5)	58.2
Diff. Detector (SD-2.1)	56.1
Diff. Guided (SD-1.5)	57.9+21.5
Diff. Guided (SD-2.1)	**58.3**+21.9

Table 4. Synthetic to Real DG and DA Results (%) of category *car* on Cityscapes and BDD100K.

Methods	Cityscapes	BDD100K
DG methods		
CDSD [79] *(CVPR'22)*	35.2	27.4
SHADE [91] *(CVPR'22)*	40.9	30.3
SRCD [63] *(TNNLS'24)*	43.0	31.6
DA methods		
SWDA [70] *(CVPR'19)*	40.7	42.9
MTTrans [89] *(CVPR'22)*	57.9	-
SIGMA [48] *(CVPR'22)*	53.7	-
TDD [28] *(CVPR'22)*	53.4	-
MGA [95] *(CVPR'22)*	54.1	-
SIGMA++ [49] *(TPAMI'23)*	53.7	-
CIGAR [52] *(CVPR'23)*	58.5	-
NSA [96] *(ICCV'23)*	56.3	-
Ours (DG settings)		
Diff. Detector (SD-1.5)	62.8	**64.4**
Diff. Detector (SD-2.1)	**64.5**	64.1
Diff. Guided (SD-1.5)	59.7+22.3	58.2+30.0
Diff. Guided (SD-2.1)	57.3+19.9	54.5+26.3

Table 5. Generalization detection Results (%) on Diverse Weather benchmark. **DF**: Daytime-Foggy, **DR**: Dusk-Rainy, **NR**: Night-Rainy, **NS**: Night-Sunny, as described in Sec. 4.1. ***Class-wise results are provided in the supplementary material.***

Methods	DF	DR	NR	NS	Average
CDSD [79] *(CVPR'22)*	33.5	28.2	16.6	36.6	28.7
SHADE [91] *(CVPR'22)*	33.4	29.5	16.8	33.9	28.4
CLIPGap [76] *(CVPR'23)*	32.0	26.0	12.4	34.4	26.2
SRCD [63] *(TNNLS'24)*	35.9	28.8	17.0	36.7	29.6
G-NAS [81] *(AAAI'24)*	36.4	35.1	17.4	45.0	33.5
OA-DG [41] *(AAAI'24)*	38.3	33.9	16.8	38.0	31.8
DivAlign [16] *(CVPR'24)*	37.2	38.1	24.1	42.5	35.5
UFR [53] *(CVPR'24)*	39.6	33.2	19.2	40.8	33.2
Diff. Detector (SD-1.5)	43.3	**42.5**	**27.8**	47.0	**40.2**
Diff. Detector (SD-2.1)	44.6	41.6	23.2	46.4	39.0
Diff. Guided (SD-1.5)	**44.7**	37.4	21.7	48.7	38.1+13.9
Diff. Guided (SD-2.1)	44.7	37.1	20.0	**49.3**	37.8+13.6

Table 6. Real to Artistic DG and DA Results (%) on Clipart, Comic, Watercolor. ***Class-wise results are provided in the supplementary material.***

Methods	Clipart	Comic	Watercolor
DG methods			
Div. *(CVPR'24)*	33.7	25.5	52.5
DivAlign *(CVPR'24)*	38.9	33.2	57.4
DA methods			
SWDA *(CVPR'19)*	–	29.4	53.3
MCRA *(ECCV'20)*	–	33.5	56.0
I3Net *(CVPR'21)*	–	30.1	51.5
DBGL *(ICCV'21)*	–	29.7	53.8
AT *(CVPR'22)*	49.3	–	59.9
D-ADAPT *(ICLR'22)*	49.0	40.5	–
TIA *(CVPR'22)*	46.3	–	–
LODS *(CVPR'22)*	45.2	–	58.2
CIGAR *(CVPR'23)*	46.2	–	–
CMT *(CVPR'23)*	47.0	–	–
Ours (DG settings)			
Diff. Detector (SD-1.5)	**58.3**	**51.9**	**68.4**
Diff. Detector (SD-2.1)	51.7	46.6	62.1
Diff. Guided (SD-1.5)	40.8+13.6	29.7+11.6	54.2+12.7
Diff. Guided (SD-2.1)	32.7+5.5	24.9+6.8	50.6+9.1

Table 11. Model calibration performance with D-ECE [40] metric.

Detector	D-ECE(%) ↓ BDD100K	Cityscapes (car)	Clipart
FR-R101 Baseline	10.9	20.2	10.8
Diff. Detector	**8.5**	**8.7**	**5.5**