

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



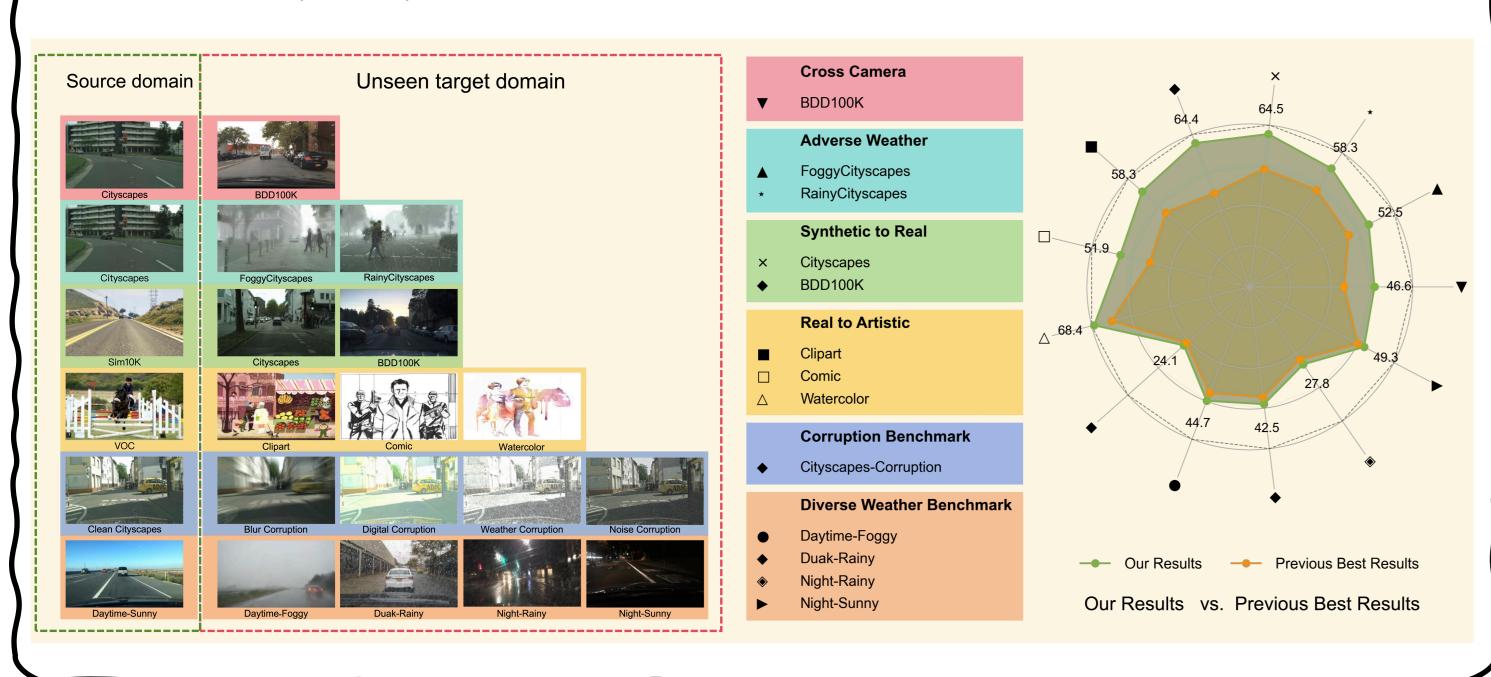
Boyong He, Yuxiang Ji, Qianwen Ye, Zhuoyue Tan and Liaoni Wu*. Xiamen University



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

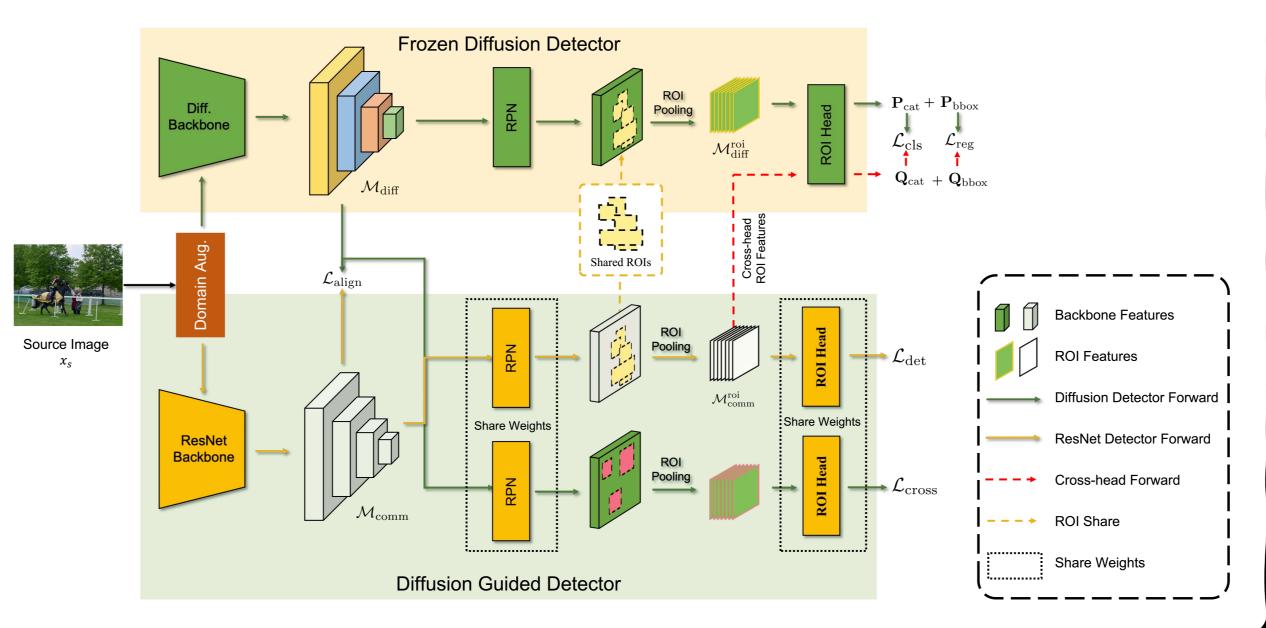
Benchmarks and Comparisons

- 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

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

Object-level Alignment

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

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

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

Div. (CVPR'24)

DivAlign (CVPR'24

SWDA (CVPR'19) MCRA (ECCV'20) I3Net (CVPR'21) DBGL (ICCV'21)

D-ADAPT (ICLR'22)

Diff. Detector (SD-1.5) **Diff. Detector** (SD-2.1)

Diff. Guided (SD-1.5)

Diff. Guided (SD-2.1)

TIA (CVPR'22) LODS (CVPR'22) CIGAR (CVPR'23) CMT (CVPR'23)

Full Objective

object-level alignment

Diff. Backbone

	טוווכ	i Citt 1	iiiics	.cps					
Т	BDD 100K	Cityscapes (car)	Clipart	Inference Time (ms)	50 (%) 40			_	
1	28.6	49.8	37.4	270	mAI 40	Ť			
2	34.9	54.1	48.6	404	30				
<u>5</u>	46.6	62.8	58.3	789	20				
10	47.1	62.6	58.9	1,424					
20	45.6	61.4	57.7	2,820	10	0 1	00 200 300		500 600 700 800 900 1000
							Max	steps o	f noise schedule
		Featu	ıre C	ollecti	on	and	Fusio	on P	ipeline
	Noise	ed Image x_t		Featur Extracti			Feature I	Fusior	1
		↑ ↑	l_1	l_4 l_3 $t = 1$		→	BottleNeck		
DDIM		J ×		l_4 l_3		-	BottleNeck	Weighted Aggregation	→
		†	l_1	! ! ! !3		→	BottleNeck	We	Diffusion Features Frozen Modules
		e Image x_s		t = T					Free Modules
			Multi-s	tep Fea	ture l	Extra	action a	nd F	usion

Main Results

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

Methods

	DG m	ethods	s (witho	out targe	et 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
D A	meth	ods (w	ith unlo	abeled to	arget da	ıta)		
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 s	ettings)				
Diff. Detector (SD-1.5)	38.9	31.0	71.5	<u>37.6</u>	<u>61.5</u>	47.0	38.5	46.6
Diff. Detector (SD-2.1)	38.0	<u>33.6</u>	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

																			R
	Bike	Bus	Car	Motor	Psn.	Rider	Truck	mAP											
	DG m	ethods	(with	out targe	et data)				Methods	Bus	Bike	Car	Motor	Psn.	Rider	Train	Truck	mAP	_
	22.9	20.5	33.8	14.7	18.5	23.6	18.2	21.7				DG 1	nethods						N
	25.1	19.0	36.8	18.4	24.1	24.9	19.8	24.0	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	F
	24.8	21.5	38.7	19.0	25.7	28.4	23.1	25.9	MAD [84] (CVPR'23)	44.0	40.1	45.0	30.3	34.2	47.4	42.4	25.6	38.6	F
	-	_	-	-	-	-	-	28.0				DA 1	nethods						S
D A	A meth	ods (w	ith unl	abeled to	arget de	ata)			MGA [95] (CVPR'22)	53.2	36.9	61.5	27.9	43.1	47.3	50.3	30.2	43.8	N
	28.8	25.5	53.9	24.5	39.6	38.9	24.1	33.6	MTTrans [89] (CVPR'22)	45.9	46.5	65.2	32.6	47.7	49.9	33.8	25.8	43.4	_
	28.8	33.8	52.7	23.0	40.5	39.9	25.8	34.9	OADA [87] (CVPR'22)	48.5	39.8	62.9	34.3	47.8	46.5	50.9	32.1	45.4	
	26.3	23.6	64.1	17.9	46.9	29.6	20.2	32.7	MIC [31] (CVPR'23)	52.4	47.5	67.0	40.6	50.9	55.3	33.7	33.9	47.6	N
21		26.3		17.9	47.5	30.4	21.1		SIGMA++ [49] (TPAMI'23)	52.4	39.9	61.0	34.8	46.4	45.1	44.6	32.1	44.5	Γ
3)	27.1	20.3	65.6	17.0	47.3	30.4	21.1	33.7	CIGAR [52] (CVPR'23)	56.6		62.1	33.7	46.1	47.3	44.3	27.8	44.9	(
	-	-	-	-	-	-	-	35.5	CMT [3] (CVPR'23)	$\frac{50.0}{66.0}$	51.2	63.7	41.4	45.9	55.7	38.8	39.6	50.3	S
	38.0	30.6	63.5	28.2	53.4	40.4	27.4	40.2	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)					111 [10] (CVI K 23)	33.7					33.0	77.1	32.1	30.4	Т
	38.9	31.0	71.5	37.6	61.5	47.0	38.5	46.6	D100 D	l 		,	G settin	0 /	- 0 -	• • •	• • •	~~ 1	L
	38.0	33.6	69.9	36.6	62.1	46.3	34.2	45.8	Diff. Detector (SD-1.5)	56.2		66.7	39.9	50.2	<u>59.5</u>	39.9	38.0	50.1	T
	38.4	33.4	72.0	38.3	60.3	47.0	35.0	46.3+20.9	Diff. Detector (SD-2.1)	55.5	49.6	67.0	40.4	50.4	58.2	29.2	36.4	48.3	I
									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	_
	38.5	32.6	<u>71.8</u>	37.5	60.2	46.7	<u>35.3</u>	46.1+20.7	Diff. Guided (SD-2.1)	55.1	<u>53.9</u>	67.0	<u>43.4</u>	<u>51.9</u>	<u>59.5</u>	42.2	34.8	<u>51.0</u> +20.3	

RainyCityscapes.

nAP					
	Methods	mAP	Methods	Cityscapes	BDD1001
25.3	DG methods		DG me	thods	
31.0	FACT [85] (CVPR'21)	39.9	CDSD [79] (CVPR'22)	35.2	27.4
38.6	FSDR [34] (CVPR'22)	42.8	SHADE [91] (CVPR'22)	40.9	30.3
	SCG [84] (CVPR'23)	39.1	SRCD [63] (TNNLS'24)	43.0	31.6
43.8	MAD [84] (CVPR'23)	42.3	DA me	thods	
+3.6 43.4	WIAD [04] (CVFR 23)	12.3	SWDA [70] (CVPR'19)	40.7	42.9
+5.4 45.4	DA methods		MTTrans [89] (CVPR'22)	57.9	-
	MGA [95] (CVPR'22)	43.0	SIGMA [48] (CVPR'22)	53.7	-
47.6	TDD [28] (CVPR'23)	50.3	TDD [28] (CVPR'22)	53.4	-
44.5	CMT [3] (CVPR'23)	52.1	MGA [95] (CVPR'22)	54.1	-
44.9	SIGMA++ [49] (TPAMI'23)	46.9	SIGMA++ [49] (TPAMI'23)	53.7	-
50.3		1	CIGAR [52] (CVPR'23)	58.5	-
50.4	Ours (DG setting	. *	NSA [96] (ICCV'23)	56.3	-
	Diff. Detector (SD-1.5)	<u>58.2</u>	Ours (DG	settings)	
50.1	Diff. Detector (SD-2.1)	56.1	Diff. Detector (SD-1.5)	62.8	64.4
48.3	Diff. Guided (SD-1.5)	57.9+21.5	Diff. Detector (SD-2.1)	64.5	64.1
.5 +21.8	Diff. Guided (SD-2.1)	58.3 +21.9	Diff. Guided (SD-1.5)	59.7+22.3	58.2+30.0
<u>.0</u> +20.3			Diff. Guided (SD-2.1)	57.3+19.9	54.5 <mark>+26.3</mark>

Cityscapes and BDD100K.

Methods	Cityscapes	BDD100K
DG me	ethods	
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 me	ethods	
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

Table 2. Adverse Weather DG and DA Results (%) on FoggyCi Table 3. Adverse Weather Table 4. Synthetic to Real DG and Table 5. Generalization detection Results (%) on Diverse Weather Table 6. Real to Artistic DG and DA Results (%) on Clipart, DG and DA Results (%) on DA Results (%) of category car on benchmark. DF: Daytime-Foggy, DR: Dusk-Rainy, NR: Night-Roman Comic, Watercolor. Class-wise results are provided in the sup-Rainy, NS: Night-Sunny, as described in Sec. 4.1. Class-wise results are provided in the supplementary material.

thods	Cityscapes	BDD100K						
DG me SD [79] (CVPR'22)	thods 35.2	27.4	Methods	DF	DR	NR	NS	Average
ADE [91] (CVPR'22)	40.9	30.3		1				
CD [63] (TNNLS'24)	43.0	31.6	CDSD [79] (CVPR'22)	33.5	28.2	16.6	36.6	28.7
	I		SHADE [91] (CVPR'22)	33.4	29.5	16.8	33.9	28.4
DA me DA [70] (CVPR'19)	thods 40.7	42.9	CLIPGap [76] (CVPR'23)	32.0	26.0	12.4	34.4	26.2
Trans [89] (CVPR'22)	57.9	-	SRCD [63] (TNNLS'24)	35.9	28.8	17.0	36.7	29.6
MA [48] (CVPR'22)	53.7	-	G-NAS [81] (AAAI'24)	36.4	35.1	17.4	45.0	33.5
D [28] (CVPR'22)	53.4	-	OA-DG [41] (AAAI'24)	38.3	33.9	16.8	38.0	31.8
A [95] (CVPR'22)	54.1	-	DivAlign [16] (CVPR'24)	37.2	38.1	24.1	42.5	35.5
MA++ [49] (TPAMI'23) GAR [52] (CVPR'23)	53.7 58.5	-	UFR [53] (CVPR'24)	39.6	33.2	19.2	40.8	33.2
A [96] (ICCV'23)	56.3	-		<u> </u>				
Ours (DG	sottings)		Diff. Detector (SD-1.5)	43.3	42.5	27.8	47.0	40.2
Detector (SD-1.5)	62.8	64.4	Diff. Detector (SD-2.1)	44.6	<u>41.6</u>	23.2	46.4	<u>39.0</u>
f. Detector (SD-2.1)	64.5	<u>64.1</u>	Diff. Guided (SD-1.5)	44.7	37.4	21.7	48.7	38.1+13.9
f. Guided (SD-1.5)	59.7+22.3	58.2+30.0						
f. Guided (SD-2.1)	57.3+19.9	54.5+26.3	Diff. Guided (SD-2.1)	44.7	37.1	20.0	49.3	37.8+13.6

plementary material.

Clipart

38.9

40.8 + 13.6

32.7 + 5.5

DG methods

DA methods

Comic

29.7 + 11.6

24.9 + 6.8

Watercolor

57.4

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

 $D-ECE(\%) \downarrow$

Di	ff. Detector	8.5	8.7	5.5
1.0	BDD100K	Citysca	apes (car)	Cliaprt
-	Perfect Calibration Output Gap	Perfect Calibration Output		rfect Calibration utput
.s.o		0.8	0.8	
Accuracy Accuracy		Accuracy	O.6. ACCURACY	
FR-101 Baseline Accuracy Accuracy		₩ 0.4	¥ 0.4	AH
0.2		0.2	0.2	
0.0	0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 Confidence	0.0 0.1 0.2 0.3 0.4 Cont	0.5 0.6 0.7 0.8 0.9 1.0 0.0 0.1 fidence	0.2 0.3 0.4 0.5 0.6 0.7 0.8 Confidence
	Perfect Calibration Output	1.0 Perfect Calibration Output	1.0 Pe	rfect Calibration
0.8	Gap	0.8 - Gap	0.8	
ecto.) 0.6 0.6	9.0	
Detecto Accuracy		Accuracy Accuracy	O.6 ACCURACY	
Diff. Detector				
0.2		0.2	0.2	4
0.0	0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 Confidence	0.0 0.1 0.2 0.3 0.4 Cont	0.5 0.6 0.7 0.8 0.9 1.0 0.0 0.1 fidence	0.2 0.3 0.4 0.5 0.6 0.7 0.8 Confidence