Boosting Domain Generalized and Adaptive Detection with Diffusion Models: Fitness, Generalization, and Transferability

Supplementary Material

A. Different Versions of Diff. Detector

Tab. 1 presents performance comparison between our method and GDD across different versions of Stable Diffusion models (SD-3-M means Stable-Diffusion-3-Medium with official weights). Our method achieves both superior accuracy and significantly improved efficiency, reducing inference time by approximately 75% (from 679ms, 698ms, and 976ms to 164ms, 167ms, and 244ms on SD-1.5, SD-2.1, and SD-3-M respectively) while consistently outperforming GDD across all datasets with improvements of up to 8.0% percentage points.

We observe that SD-3-M performs notably worse than SD-1.5 and SD-2.1 on both Real to Artistic and Diverse Weather datasets, with performance gaps of nearly 29.6% and 27.8% mAP on Clipart and Comic datasets. Despite these baseline differences, our method still achieves significant improvements over GDD [15] on SD-3-M, with gains of 12.8% and 13.3% percentage points on Clipart and Comic datasets.

The inferior performance of SD-3-M can be attributed to several factors: First, the UNet [32] denoising structure in SD-1.5 and SD-2.1 outputs multi-level features, which is advantageous for dense prediction tasks, whereas SD-3-M uses MMDiT [12] as its denoising structure, which only outputs single-scale intermediate features, lacking multi-scale feature representation capabilities and making it difficult to capture fine-grained semantic information in images. Additionally, the transformer structure in SD-3-M, while excellent for generation tasks, has inherent limitations for detection tasks that require precise spatial correspondence. Its global attention mechanism may neglect local spatial structure information, resulting in weakened feature alignment capability during cross-domain transfer. This also explains why the performance decline is more pronounced on datasets with larger style differences (such as Clipart and Comic).

However, we believe that SD-3-M with its DiT structure still holds significant exploratory value for domain generalization and adaptation tasks. Future work could explore multi-scale feature utilization, attention mechanism improvements, and domain generalization strategies specifically designed for transformer architectures, offering new possibilities for applying diffusion models to dense prediction tasks.

Table 1. Performance comparison between GDD and our model across different Stable Diffusion versions.

Version	Models	BDD.	Foggy.	Cli.	Com.	Wat.
SD-1.5	GDD [15]	46.6	50.1	58.3	51.9	68.4
	Ours	49.3	50.7	64.1	55.2	69.7
	+Gain	+2.7	+0.6	+5.8	+3.3	+1.3
SD-2.1	GDD [15]	45.8	48.3	51.7	46.6	62.1
	Ours	48.0	50.3	59.7	54.5	68.6
	+Gain	+2.2	+2.0	+8.0	+7.9	+6.5
SD-3-M	GDD [15]	40.4	46.1	28.7	24.1	45.0
	Ours	41.9	50.8	41.5	37.4	54.6
	+Gain	+1.5	+4.7	+12.8	+13.3	+9.6

Version	Models	DF	DR	NR	NS	Inf. (ms)
SD-1.5	GDD [15]	43.3	42.5	27.8	47.0	679
	Ours	48.5	48.4	31.3	51.6	164
	+Gain	+5.2	+5.9	+3.5	+4.6	
SD-2.1	GDD [15]	44.6	41.6	23.2	46.4	698
	Ours	48.7	47.3	29.8	51.8	167
	+Gain	+4.1	+5.7	+6.6	+5.4	
SD-3-M	GDD [15]	36.0	30.5	15.9	32.8	976
	Ours	38.3	32.0	15.9	35.8	244
	+Gain	+2.3	+1.5	+0.0	+3.0	

B. Class-wise Results and Comparisons

B.1. Analysis of Real to Artistic Results

Tab. 2, 3, and 4 present detailed class-wise results on three Real to Artistic datasets. Our methods consistently outperforms existing methods across all three datasets. In DG setting, our diff. detector achieves 64.1%, 55.2%, and 69.7% mAP on Clipart, Comic, and Watercolor respectively, surpassing the previous SOTA results GDD [15] by 5.8%, 3.3%, and 1.3%. Similar improvements are observed in the DA setting, where our approach reaches 58.2%, 50.5%, and 67.5% mAP.

B.2. Analysis of Diverse Weather Results

Tab. 5, 6, 7, and 8 present detailed class-wise results on four Diverse Weather Datasets. In DG setting, our diff. detector achieves 48.5%, 48.4%, 31.3%, and 51.6% mAP on Daytime-Foggy, Dusk-Rainy, Night-Rainy, and Night-Sunny respectively, surpassing the previous best methods GDD [15] by 5.2%, 5.9%, 3.5%, and 4.6%. The improvements are particularly significant in challenging low-light and adverse weather conditions such as Night-Rainy, where our method substantially outperforms traditional approaches. These results demonstrate the robustness of our diffusion-based features in capturing consistent object representations across diverse weather and lighting conditions.

Table 2. Real to Artistic DG and DA Results (%) on Comic.

Methods	Bike	Bird	Car	Cat	Dog	Person	mAP
DG methods *(without target data)*							
Div. [9] *(CVPR'24)*	41.7	12.3	29.0	13.2	20.6	36.5	25.5
DivAlign [9] *(CVPR'24)*	54.1	16.9	30.1	25.0	27.4	45.9	33.2
DDT (SD-1.5) [14] *(MM'24)*	/	/	/	/	/	/	44.4
GDD (SD-1.5) [15] *(CVPR'25)*	63.3	41.7	**58.2**	31.8	40.9	**75.3**	51.9
GDD (R101) [15] *(CVPR'25)*	47.6	21.0	35.3	9.1	21.6	43.5	29.7
***Ours** (Diff. Detector, SD-1.5)*	**64.8**	**50.7**	57.7	**33.0**	**50.1**	75.0	**55.2**
***Ours** (Diff. Guided, R101)*	48.7	16.9	39.0	8.9	20.1	46.6	30.0
DA methods *(with unlabeled target data)*							
DA-Faster [6] *(CVPR'18)*	31.1	10.3	15.5	12.4	19.3	39.0	21.2
SWDA [33] *(CVPR'19)*	36.4	21.8	29.8	15.1	23.5	49.6	29.4
STABR [20] *(ICCV'19)*	50.6	13.6	31.0	7.5	16.4	41.4	26.8
MCRA [42] *(ECCV'20)*	47.9	20.5	37.4	20.6	24.5	50.2	33.5
I3Net [5] *(CVPR'21)*	47.5	19.9	33.2	11.4	19.4	49.1	30.1
DBGL [4] *(ICCV'21)*	35.6	20.3	33.9	16.4	26.6	45.3	29.7
D-ADAPT [18] *(ICLR'22)*	52.4	25.4	42.3	**43.7**	25.7	53.5	40.5
DDT (R101) [14] *(MM'24)*	**63.2**	34.8	**56.6**	31.7	39.0	**75.9**	50.2
***Ours** (Diff. Guided, R101)*	60.1	**37.5**	52.1	30.2	**48.7**	74.3	**50.5**

Table 3. Real to Artistic DG and DA Results (%) on Watercolor.

Methods	Bike	Bird	Car	Cat	Dog	Person	mAP
DG methods *(without target data)*							
Div. [9] *(CVPR'24)*	87.1	51.7	53.6	35.1	23.6	63.6	52.5
DivAlign [9] *(CVPR'24)*	90.4	51.8	51.9	43.9	35.9	70.2	57.4
DDT (SD-1.5) [14] *(MM'24)*	/	/	/	/	/	/	58.7
GDD (SD-1.5) [15] *(CVPR'25)*	**99.8**	70.3	57.5	49.8	51.0	**82.0**	68.4
GDD (R101) [15] *(CVPR'25)*	90.1	51.0	48.5	40.2	28.9	66.7	54.2
***Ours** (Diff. Detector, SD-1.5)*	88.0	**74.9**	**59.0**	**60.5**	**57.2**	78.5	**69.7**
***Ours** (Diff. Guided, R101)*	83.9	53.0	54.0	42.5	36.3	69.6	56.6
DA methods *(with unlabeled target data)*							
SWDA [33] *(CVPR'19)*	82.3	55.9	46.5	32.7	35.5	66.7	53.3
MCRA [42] *(ECCV'20)*	87.9	52.1	51.8	41.6	33.8	68.8	56.0
UMT [10] *(CVPR'21)*	88.2	55.3	51.7	39.8	43.6	69.9	58.1
IIOD [36] *(TPAMI'21)*	95.8	54.3	48.3	42.4	35.1	65.8	56.9
I3Net [5] *(CVPR'21)*	81.1	49.3	46.2	35.0	31.9	65.7	51.5
SADA [7] *(IJCV'21)*	82.9	54.6	52.3	40.5	37.7	68.2	56.0
CDG [24] *(CVPR'19)*	**97.7**	53.1	52.1	47.3	38.7	68.9	59.7
VDD [37] *(ICCV'21)*	90.0	56.6	49.2	39.5	38.8	65.3	56.6
DBGL [4] *(ICCV'21)*	83.1	49.3	50.6	39.8	38.7	61.3	53.8
AT [26] *(CVPR'22)*	93.6	56.1	58.9	37.3	39.6	73.8	59.9
LODS [25] *(CVPR'22)*	95.2	53.1	46.9	37.2	47.6	69.3	58.2
DAVimNet [11] *(ArXiv'24)*	87.2	53.6	51.9	34.9	30.3	70.1	54.8
UMGA [40] *(TPAMI'24)*	67.1	53.4	43.9	46.3	50.5	**79.8**	56.8
DDT (R101) [14] *(MM'24)*	87.1	64.0	55.7	50.6	48.8	75.7	63.7
***Ours** (Diff. Guided, R101)*	93.8	**68.2**	**57.7**	**52.0**	**53.5**	79.7	**67.5**

Table 4. Real to Artistic DG and DA Results (%) on Clipart.

Methods	aero.	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	bike	psn.	plant.	sheep	sofa	train	tv	mAP
DG methods *(without target data)*																					
Div. [9] *(CVPR'24)*	29.3	50.9	23.4	35.3	45.3	49.8	33.4	10.6	43.3	22.3	31.6	4.5	32.9	51.9	40.2	51.1	18.2	29.6	42.3	28.5	33.7
DivAlign [9] *(CVPR'24)*	34.4	64.4	22.7	27.0	45.6	59.2	32.9	7.0	46.8	**55.8**	28.9	14.5	44.4	58.0	55.2	52.1	14.8	38.4	42.5	33.9	38.9
DDT (SD-1.5) [14] *(MM'24)*	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	47.4
GDD (SD-1.5) [15] *(CVPR'25)*	63.7	**86.1**	49.8	56.5	52.9	50.9	67.3	19.7	74.7	34.3	57.7	41.9	63.2	89.4	89.6	59.2	23.5	64.9	65.9	55.2	58.3
GDD (R101) [15] *(CVPR'25)*	19.3	57.8	28.4	37.4	57.8	**81.3**	46.3	3.8	57.8	27.2	28.3	19.6	42.5	50.9	57.8	59.8	15.6	36.0	37.7	50.5	40.8
Ours *(Diff. Detector, SD-1.5)*	**75.7**	72.7	**59.1**	**66.2**	**63.5**	62.9	**76.5**	**19.8**	**78.5**	46.9	**59.1**	**42.7**	**66.9**	**93.7**	**91.2**	**63.4**	**41.0**	**70.4**	**68.3**	**62.5**	**64.1**
Ours *(Diff. Guided, R101)*	31.1	56.2	26.4	33.9	55.9	72.4	42.4	10.2	56.5	11.6	29.0	12.2	36.3	65.4	59.5	57.9	17.9	34.0	39.1	51.6	40.5
DA methods *(with unlabeled target data)*																					
SWDA [33] *(CVPR'19)*	26.2	48.5	32.6	33.7	38.5	54.3	37.1	18.6	34.8	58.3	17.0	12.5	33.8	65.5	61.6	52.0	9.3	24.9	54.1	49.1	38.1
HTCN [3] *(CVPR'20)*	33.6	58.9	34.0	23.4	45.6	57.0	39.8	12.0	39.7	51.3	21.1	20.1	39.1	72.8	63.0	43.1	19.3	30.1	50.2	51.8	40.3
SAPNet [22] *(ECCV'20)*	27.4	70.8	32.0	27.9	42.4	63.5	47.5	14.3	48.2	46.1	31.8	17.9	43.8	68.0	68.1	49.0	18.7	20.4	55.8	51.3	42.2
UMT [10] *(CVPR'21)*	39.6	59.1	32.4	35.0	45.1	61.9	48.4	7.5	46.0	**67.6**	21.4	29.5	48.2	75.9	70.5	56.7	25.9	28.9	39.4	43.6	44.1
IIOD [36] *(TPAMI'21)*	41.5	52.7	34.5	28.1	43.7	58.5	41.8	15.3	40.1	54.4	26.7	28.5	37.7	75.4	63.7	48.7	16.5	30.8	54.5	48.7	42.1
SADA [7] *(IJCV'21)*	29.4	56.8	30.6	34.0	49.5	50.5	47.7	18.7	48.5	64.4	20.3	29.0	42.3	84.1	73.4	37.4	20.5	39.8	41.2	48.0	43.3
UaDAN [13] *(TMM'21)*	35.0	72.7	41.0	24.4	21.3	69.8	53.5	2.3	34.2	61.2	31.0	29.5	47.9	63.6	62.2	61.3	13.9	7.6	48.6	23.9	40.2
DBGL [4] *(ICCV'21)*	28.5	52.3	34.3	32.8	38.6	66.4	38.2	**25.3**	39.9	47.4	23.9	17.9	38.9	78.3	61.2	51.7	26.2	28.9	56.8	44.5	41.6
AT [26] *(CVPR'22)*	33.8	60.9	38.6	49.4	52.4	53.9	56.7	7.5	52.8	63.5	34.0	25.0	62.2	72.1	77.2	57.7	27.2	52.0	55.7	54.1	49.3
D-ADAPT [18] *(ICLR'22)*	56.4	63.2	42.3	40.9	45.3	77.0	48.7	25.4	44.3	58.4	31.4	24.5	47.1	75.3	69.3	43.5	27.9	34.1	60.7	**64.0**	49.0
TIA [41] *(CVPR'22)*	42.2	66.0	36.9	37.3	43.7	71.8	49.7	18.2	44.9	58.9	18.2	29.1	40.7	87.8	67.4	49.7	27.4	27.8	57.1	50.6	46.3
LODS [25] *(CVPR'22)*	43.1	61.4	40.1	36.8	48.2	45.8	48.3	20.4	44.8	53.3	32.5	26.1	40.6	86.3	68.5	48.9	25.4	33.2	44.0	56.5	45.2
CIGAR [27] *(CVPR'23)*	35.2	55.0	39.2	30.7	60.1	58.1	46.9	31.8	47.0	61.0	21.8	26.7	44.6	52.4	68.5	54.4	31.3	38.8	56.5	63.5	46.2
CMT [2] *(CVPR'23)*	39.8	56.3	38.7	39.7	60.4	35.0	56.0	7.1	60.1	60.4	35.8	28.1	**67.8**	84.5	80.1	55.5	20.3	32.8	42.3	38.2	47.0
DAVimNet [11] *(ArXiv'24)*	33.2	**75.5**	33.1	25.5	27.9	69.9	50.1	16.9	40.8	47.0	32.6	24.1	32.5	77.0	69.5	37.6	23.3	41.2	57.0	48.2	43.8
UMGA [40] *(TPAMI'24)*	38.7	77.2	39.0	35.4	53.8	78.1	47.5	17.5	38.2	49.9	20.0	18.0	44.2	83.5	74.6	57.7	26.7	26.0	55.4	58.3	47.0
CAT [19] *(CVPR'24)*	40.5	64.1	38.8	41.0	60.7	55.5	55.6	14.3	54.7	59.6	**46.2**	20.3	58.7	92.9	62.6	57.5	22.4	40.9	49.5	46.0	49.1
DDT (R101) [14] *(MM'24)*	**60.9**	65.7	40.9	52.7	55.2	**82.8**	63.3	14.0	59.2	56.2	39.7	**39.6**	52.7	**97.7**	83.1	60.1	29.0	**44.7**	53.1	61.8	55.6
Ours *(Diff. Guided, R101)*	54.2	53.5	**52.0**	**57.4**	**62.5**	80.6	**65.5**	22.0	**64.1**	60.3	44.3	36.5	60.6	92.5	**85.2**	**63.2**	**42.9**	44.0	**59.8**	61.9	**58.2**

Table 5. Generalization detection Results (%) on Daytime-Foggy.

Methods	Bus	Bike	Car	Motor	Person	Rider	Truck	mAP
IBN-Net [29] *(CVPR'18)*	29.9	26.1	44.5	24.4	26.2	33.5	22.4	29.6
SW [30] *(ICCV'19)*	30.6	26.2	44.6	25.1	30.7	34.6	23.6	30.8
IterNorm [17] *(CVPR'19)*	29.7	21.8	42.4	24.4	26.0	33.3	21.6	28.5
ISW [8] *(CVPR'21)*	29.5	26.4	49.2	27.9	30.7	34.8	24.0	31.8
CDSD [35] *(CVPR'22)*	32.9	28.0	48.8	29.8	32.5	38.2	24.1	33.5
CLIPGap [34] *(CVPR'23)*	36.1	34.3	58.0	33.1	39.0	43.9	25.1	38.5
SRCD [31] *(TNNLS'24)*	36.4	30.1	52.4	31.3	33.4	40.1	27.7	35.9
G-NAS [38] *(AAAI'24)*	32.4	31.2	57.7	31.9	38.6	38.5	24.5	36.4
OA-DG [21] *(AAAI'24)*	-	-	-	-	-	-	-	38.3
DivAlign [9] *(CVPR'24)*	-	-	-	-	-	-	-	37.2
UFR [28] *(CVPR'24)*	36.9	35.8	61.7	33.7	39.5	42.2	27.5	39.6
Prompt-D [23] *(CVPR'24)*	36.1	34.5	58.4	33.3	40.5	44.2	26.2	39.1
DIDM [16] *(ArXiv'25)*	38.5	31.6	62.1	35.8	36.8	42.7	27.3	39.3
PhysAug [39] *(ArXiv'24)*	-	-	-	-	-	-	-	40.8
GDD (SD-1.5) [15] *(CVPR'25)*	37.5	32.4	67.9	35.6	48.3	44.6	**37.1**	43.3
GDD (R101) [15] *(CVPR'25)*	39.3	35.8	69.4	37.7	48.8	49.7	32.3	44.7
Ours (Diff. Detector, SD-1.5)	**39.6**	**41.7**	**72.8**	**43.9**	**53.7**	**53.3**	34.8	**48.5**
Ours (Diff. Guided, R101)	37.6	41.1	70.8	41.0	50.7	51.4	34.6	46.7

Table 6. Generalization detection Results (%) on Dusk-Rainy.

Methods	Bus	Bike	Car	Motor	Person	Rider	Truck	mAP
IBN-Net [29] *(CVPR'18)*	37.0	14.8	50.3	11.4	17.3	13.3	38.4	26.1
SW [30] *(ICCV'19)*	35.2	16.7	50.1	10.4	20.1	13.0	38.8	26.3
IterNorm [17] *(CVPR'19)*	32.9	14.1	38.9	11.0	15.5	11.6	35.7	22.8
ISW [8] *(CVPR'21)*	34.7	16.0	50.0	11.1	17.8	12.6	38.8	25.9
CDSD [35] *(CVPR'22)*	37.1	19.6	50.9	13.4	19.7	16.3	40.7	28.2
CLIPGap [34] *(CVPR'23)*	37.8	22.8	60.7	16.8	26.8	18.7	42.4	32.3
SRCD [31] *(TNNLS'24)*	39.5	21.4	50.6	11.9	20.1	17.6	40.5	28.8
G-NAS [38] *(AAAI'24)*	44.6	22.3	66.4	14.7	32.1	19.6	45.8	35.1
OA-DG [21] *(AAAI'24)*	-	-	-	-	-	-	-	33.9
DivAlign [9] *(CVPR'24)*	-	-	-	-	-	-	-	38.1
UFR [28] *(CVPR'24)*	37.1	21.8	67.9	16.4	27.4	17.9	43.9	33.2
Prompt-D [23] *(CVPR'24)*	39.4	25.2	60.9	20.4	29.9	16.5	43.9	33.7
DIDM [16] *(ArXiv'25)*	41.6	26.3	66.6	16.6	30.9	21.9	44.1	35.4
PhysAug [39] *(ArXiv'24)*	-	-	-	-	-	-	-	41.2
GDD (SD-1.5) [15] *(CVPR'25)*	49.7	27.9	74.9	18.2	45.5	24.5	56.8	42.5
GDD (R101) [15] *(CVPR'25)*	43.1	23.9	73.6	13.4	33.2	22.1	52.3	37.4
Ours (Diff. Detector, SD-1.5)	**54.8**	**36.1**	**77.6**	**27.6**	**49.3**	**32.4**	**61.1**	**48.4**
Ours (Diff. Guided, R101)	46.7	25.5	74.9	16.4	37.0	21.4	51.9	39.1

Table 7. Generalization detection Results (%) on Night-Rainy.

Methods	Bus	Bike	Car	Motor	Person	Rider	Truck	mAP
IBN-Net [29] *(CVPR'18)*	24.6	10.0	28.4	0.9	8.3	9.8	18.1	14.3
SW [30] *(ICCV'19)*	22.3	7.8	27.6	0.2	10.3	10.0	17.7	13.7
IterNorm [17] *(CVPR'19)*	21.4	6.7	22.0	0.9	9.1	10.6	17.6	12.6
ISW [8] *(CVPR'21)*	22.5	11.4	26.9	0.4	9.9	9.8	17.5	14.1
CDSD [35] *(CVPR'22)*	24.4	11.6	29.5	0.4	10.5	11.4	19.2	15.3
CLIPGap [34] *(CVPR'23)*	28.6	12.1	36.1	9.2	12.3	9.6	22.9	18.7
SRCD [31] *(TNNLS'24)*	26.5	12.9	32.4	0.8	10.2	12.5	24.0	17.0
G-NAS [38] *(AAAI'24)*	28.6	9.8	38.4	0.1	13.8	9.8	21.4	17.4
OA-DG [21] *(AAAI'24)*	-	-	-	-	-	-	-	16.8
DivAlign [9] *(CVPR'24)*	-	-	-	-	-	-	-	24.1
UFR [28] *(CVPR'24)*	29.9	11.8	36.1	9.4	13.1	10.5	23.3	19.2
Prompt-D [23] *(CVPR'24)*	25.6	12.1	35.8	**10.1**	14.2	12.9	22.9	19.2
DIDM [16] *(ArXiv'25)*	31.6	12.1	38.3	3.8	12.8	10.6	25.0	19.2
PhysAug [39] *(ArXiv'24)*	-	-	-	-	-	-	-	23.1
GDD (SD-1.5) [15] *(CVPR'25)*	42.0	15.0	53.6	6.5	26.2	13.8	37.5	27.8
GDD (R101) [15] *(CVPR'25)*	35.4	12.7	46.2	3.2	13.8	10.7	29.7	21.7
Ours (Diff. Detector, SD-1.5)	**42.6**	**21.3**	**57.3**	7.2	**29.7**	**17.7**	**43.5**	**31.3**
Ours (Diff. Guided, R101)	31.6	11.2	47.9	7.0	16.2	12.9	29.8	22.4

Table 8. Generalization detection Results (%) on Night-Sunny.

Methods	Bus	Bike	Car	Motor	Person	Rider	Truck	mAP
IBN-Net [29] *(CVPR'18)*	37.8	27.3	49.6	15.1	29.2	27.1	38.9	32.1
SW [30] *(ICCV'19)*	38.7	29.2	49.8	16.6	31.5	28.0	40.2	33.4
IterNorm [17] *(CVPR'19)*	38.5	23.5	38.9	15.8	26.6	25.9	38.1	29.6
ISW [8] *(CVPR'21)*	38.5	28.5	49.6	15.4	31.9	27.5	41.3	33.2
CDSD [35] *(CVPR'22)*	40.6	35.1	50.7	19.7	34.7	32.1	43.4	36.6
CLIPGap [34] *(CVPR'23)*	37.7	34.3	58.0	19.2	37.6	28.5	42.9	36.9
SRCD [31] *(TNNLS'24)*	43.1	32.5	52.3	20.1	34.8	31.5	42.9	36.7
G-NAS [38] *(AAAI'24)*	46.9	40.5	67.5	26.5	50.7	35.4	47.8	45.0
OA-DG [21] *(AAAI'24)*	-	-	-	-	-	-	-	38.0
DivAlign [9] *(CVPR'24)*	-	-	-	-	-	-	-	42.5
UFR [28] *(CVPR'24)*	43.6	38.1	66.1	14.7	49.1	26.4	47.5	40.8
Prompt-D [23] *(CVPR'24)*	40.9	35.0	59.0	21.3	40.4	29.9	42.9	38.5
DIDM [16] *(ArXiv'25)*	43.5	40.1	65.1	22.4	45.2	32.5	45.3	42.0
PhysAug [39] *(ArXiv'24)*	-	-	-	-	-	-	-	44.9
GDD (SD-1.5) [15] *(CVPR'25)*	49.6	42.1	70.5	21.4	54.5	38.2	52.6	47.0
GDD (R101) [15] *(CVPR'25)*	51.0	42.8	72.2	27.5	55.9	39.5	52.0	48.6
Ours (Diff. Detector, SD-1.5)	**53.1**	**47.5**	72.7	**29.4**	**58.3**	**44.4**	**55.4**	**51.6**
Ours (Diff. Guided, R101)	51.7	45.6	**73.1**	29.3	57.8	42.3	54.0	50.5

C. Error Analysis

C.1. Error Analysis with TIDE

TIDE [1] provides a comprehensive framework for analyzing object detection errors. The main error categories include **Cls** (classification errors with correct location), **Loc** (localization errors with correct class), **Both** (combined class and location errors), **Dupe** (duplicate detections), **Bkg** (false background detections), **Miss** (missed objects), **FP** (false positives not matching any ground truth), and **FN** (ground truth objects not detected).

The TIDE analysis results in Tab. 9 reveals that the most significant factor limiting standard Faster R-CNN performance across DG benchmarks is missed detections (**Miss** and **FN**). This is evident from the consistently high values in these categories, highlighted in red in the table. Our proposed diff. detector significantly reduces these missed detection errors across all test domains, leading to substantial mAP improvements.

Similarly, when ordinary detectors are guided by the diff. detectors (in both DG and DA settings), they also show reduced miss rates, which contributes to their improved performance in target domains.

Notably, as the diff. detector reduces missed detections, the primary performance limitation shifts to false positives (highlighted in green). The diff. detector tends to generate more false positive detections compared to the baseline. This insight provides a clear direction for future improvements: maintaining the diff. detectors strong recall while reducing its false positive rate could further enhance performance in cross-domain detection scenarios.

C.2. Error Analysis with Confusion Matrix

The confusion matrices (Fig. 1, 2, 3, 4, and 5) across various domains provide visual confirmation of our TIDE analysis findings. In the baseline Faster R-CNN matrices, we observe weak diagonal elements and high values in the rightmost column representing missed detections, confirming that false negatives are the primary limitation. In contrast, our diff. detector shows stronger diagonal elements and significantly reduced missed detection rates across all domains. The confusion matrices for our guided detectors (in both DG and DA settings) similarly demonstrate improved detection capabilities with fewer misses, further supporting our conclusion that addressing missed detections is crucial for effective cross-domain detection.

Table 9. Error Analysis with TIDE [1].

Method	Main Errors						Special Error		mAP
	Cls	Loc	Both	Dupe	Bkg	Miss	FP	FN	
Error Analysis on BDD100K									
Faster RCNN R101	10.8	6.8	1.9	0.1	2.4	11.6	19.2	27.2	25.3
Diff. Detector	10.6	6.1	1.9	0.1	3.1	6.1	23.1	17.8	49.3
Diff. Guided Detector for DG	10.4	6.6	2.0	0.1	2.9	7.5	22.2	19.6	46.7
Diff. Guided Detector for DA	10.6	6.9	1.9	0.1	3.0	7.4	19.1	20.4	51.3
Error Analysis on FoggyCityscapes									
Faster RCNN R101	3.3	4.1	0.6	0.1	0.6	38.8	6.6	50.0	30.7
Diff. Detector	6.5	6.2	1.2	0.3	1.5	15.3	13.0	28.2	50.7
Diff. Guided Detector for DG	5.0	6.4	1.0	0.2	1.0	19.1	8.9	31.2	54.1
Diff. Guided Detector for DA	6.3	6.0	1.0	0.3	1.3	14.9	9.9	27.9	56.6
Error Analysis on Clipart									
Faster RCNN R101	9.7	5.2	1.5	0.1	1.8	19.6	13.6	38.9	27.2
Diff. Detector	11.1	4.0	1.4	0.4	4.4	4.1	18.1	12.8	64.1
Diff. Guided Detector for DG	13.1	4.7	2.1	0.1	3.0	10.5	17.1	28.2	40.5
Diff. Guided Detector for DA	10.9	5.4	1.4	0.5	4.1	7.2	17.9	17.1	58.2
Error Analysis on DAytime-Foggy									
Faster RCNN R101	4.4	5.0	1.1	0.1	1.6	26.1	12.1	37.2	35.5
Diff. Detector	5.4	5.8	1.6	0.2	3.4	12.8	17.8	23.7	48.5
Diff. Guided Detector for DG	5.2	5.8	1.6	0.2	2.9	16.2	15.5	27.2	46.7
Error Analysis on Dusk-Rainy									
Faster RCNN R101	8.5	6.1	1.7	0.1	1.2	14.1	17.0	27.7	34.5
Diff. Detector	8.2	6.2	1.9	0.1	2.7	7.3	21.6	18.6	48.4
Diff. Guided Detector for DG	10.9	5.9	1.8	0.1	2.2	10.2	17.5	26.3	39.1
Error Analysis on Night-Rainy									
Faster RCNN R101	7.5	5.3	0.8	0.1	1.1	12.7	12.0	37.9	15.1
Diff. Detector	7.6	5.6	2.4	0.1	3.2	6.7	26.9	19.3	31.3
Diff. Guided Detector for DG	9.0	5.3	1.3	0.1	2.5	10.4	16.9	30.5	22.4
Error Analysis on Night-Sunny									
Faster RCNN R101	9.9	6.4	1.9	0.1	3.1	9.4	21.1	21.7	44.3
Diff. Detector	8.4	6.4	1.7	0.2	3.4	6.8	22.7	16.8	51.6
Diff. Guided Detector for DG	8.5	6.6	1.8	0.2	3.5	7.4	21.6	18.3	50.5

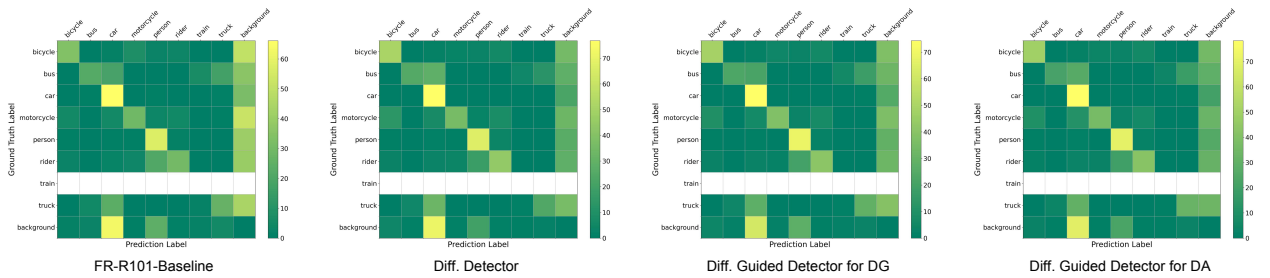

Figure 1. Confusion matrix on BDD100K.

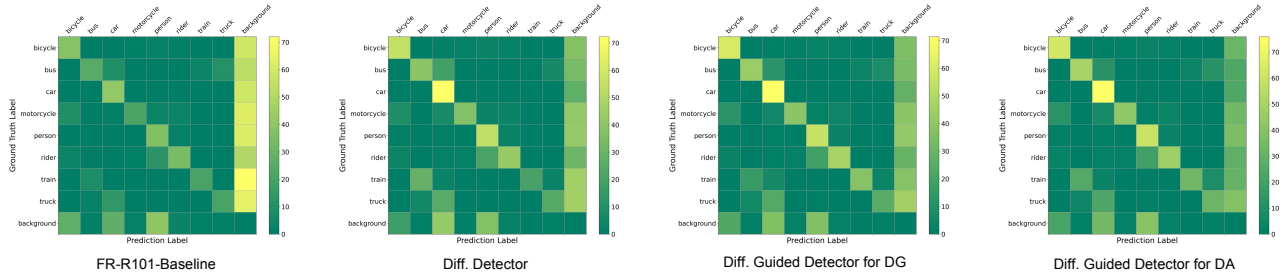

Figure 2. Confusion matrix on FoggyCityscapes.

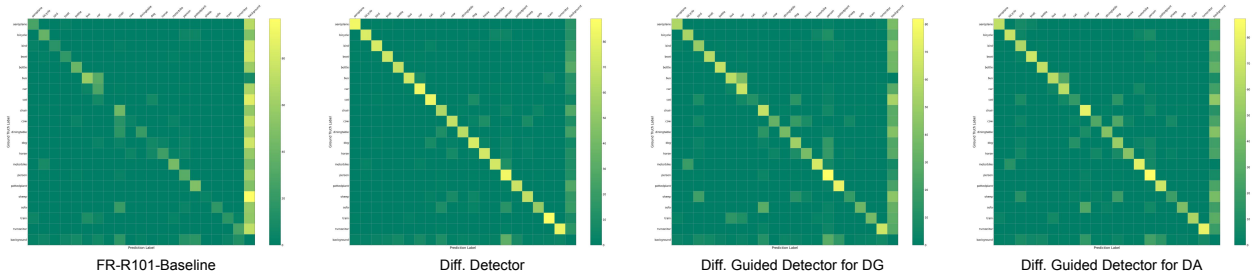

Figure 3. Confusion matrix on Clipart.

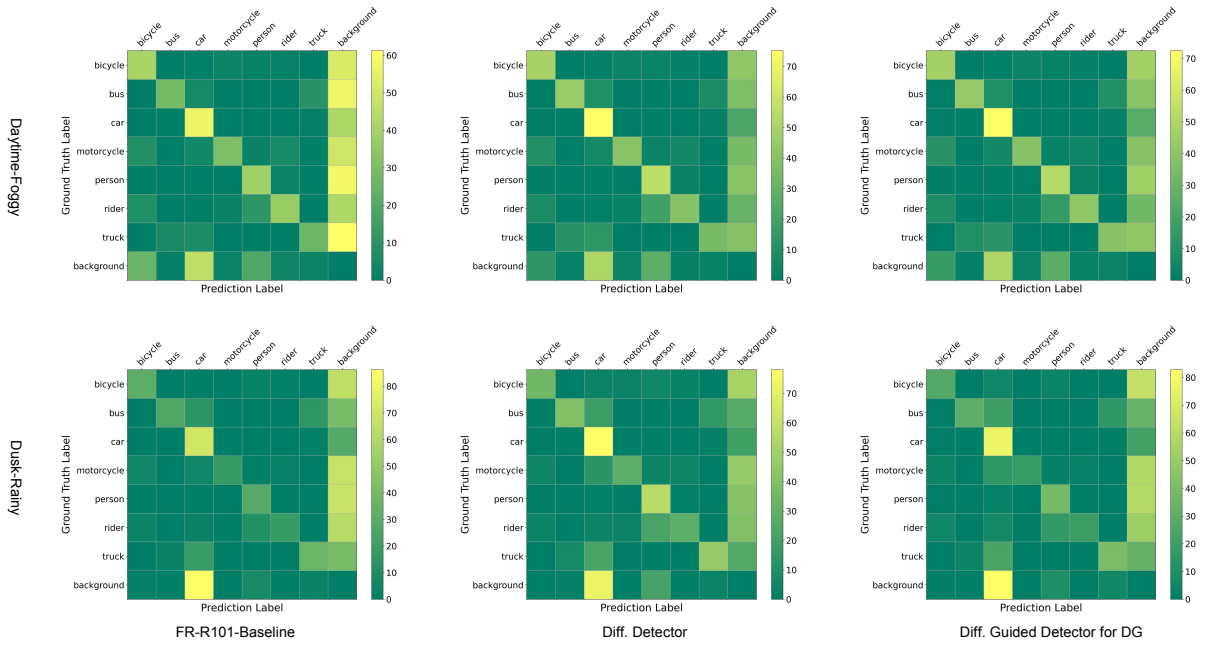

Figure 4. Confusion matrix on Daytime-Foggy and Dusk-Rainy.

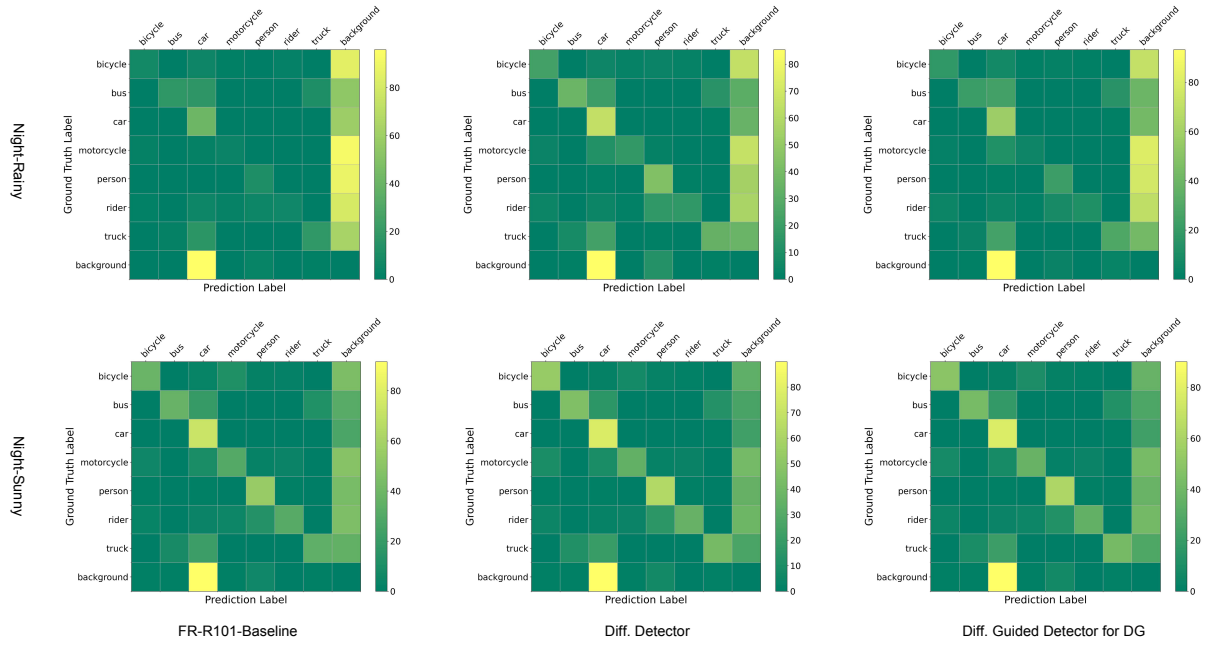

Figure 5. Confusion matrix on Night-Rainy and Night-Sunny.

D. Qualitative Prediction Results

Fig. 6, 7, 8, 9, 10, 11, 12, 13, and 14 show qualitative prediction results across all DG and DA benchmarks. These visualizations compare the detection capabilities of our proposed diff. detector against the baseline Faster R-CNN on diverse domain shift scenarios. The visual comparisons consistently demonstrate our method's strong generalization performance across real-world domain shifts, artistic style transfers, challenging weather and lighting conditions, and corrupted images. In all cases, our approach shows better recall with fewer missed objects while maintaining reasonable precision. These qualitative results align with our quantitative findings, confirming that the diff. detector effectively addresses the primary limitation of ordinary detectors missed detections when generalizing to novel domains.

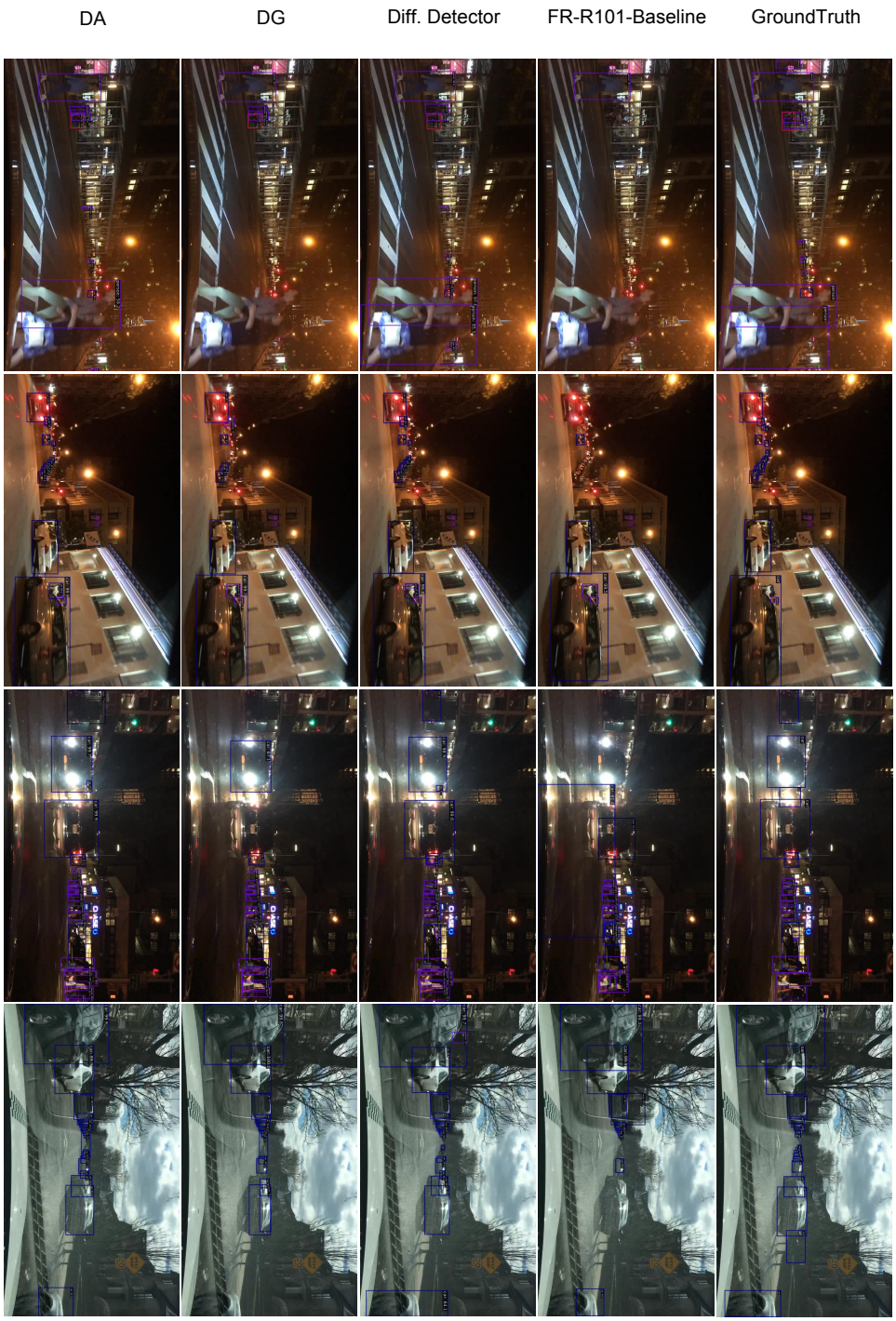

Figure 6. Qualitative prediction results on BDD100K.

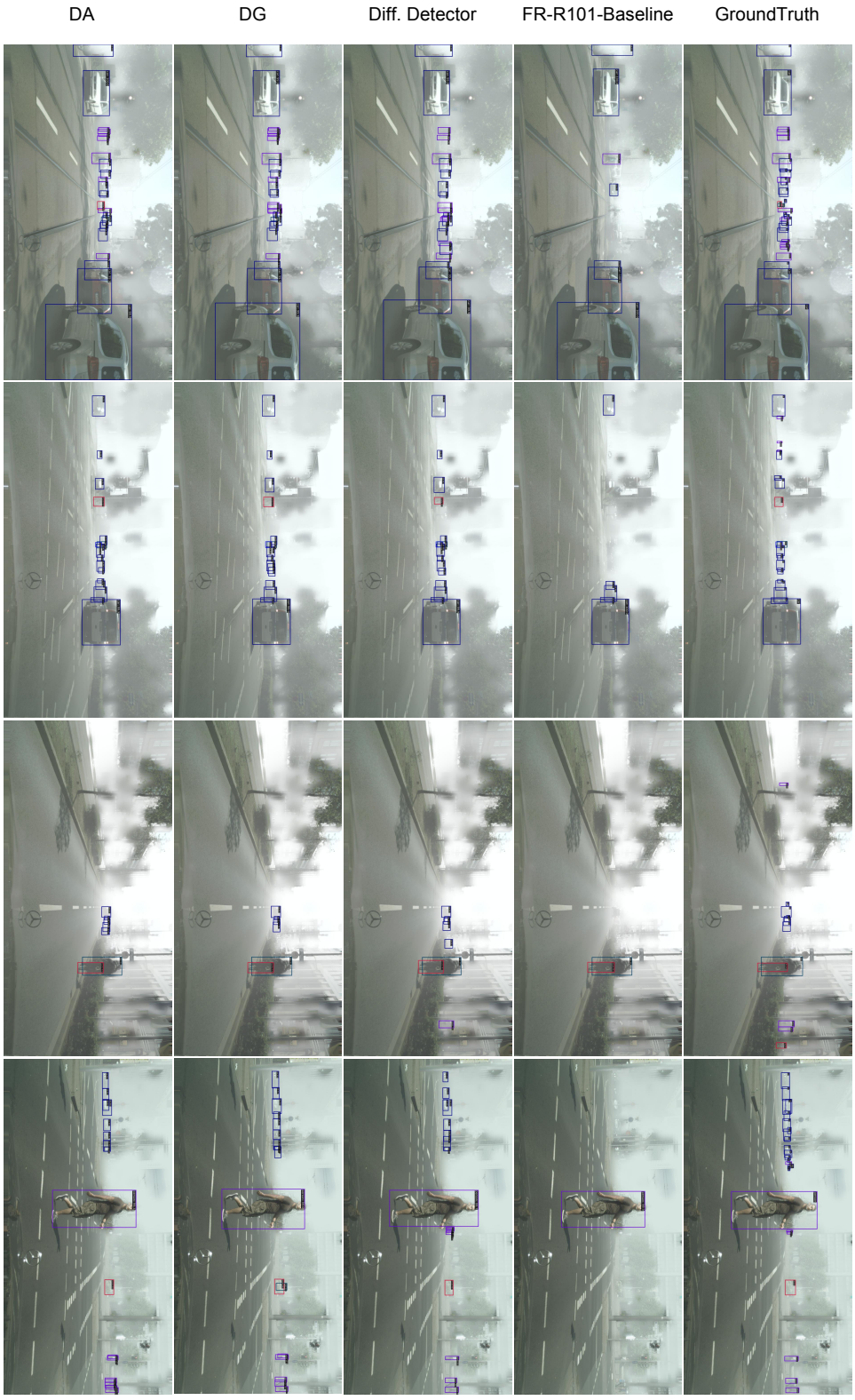

Figure 7. Qualitative prediction results on FoggyCityscapes.

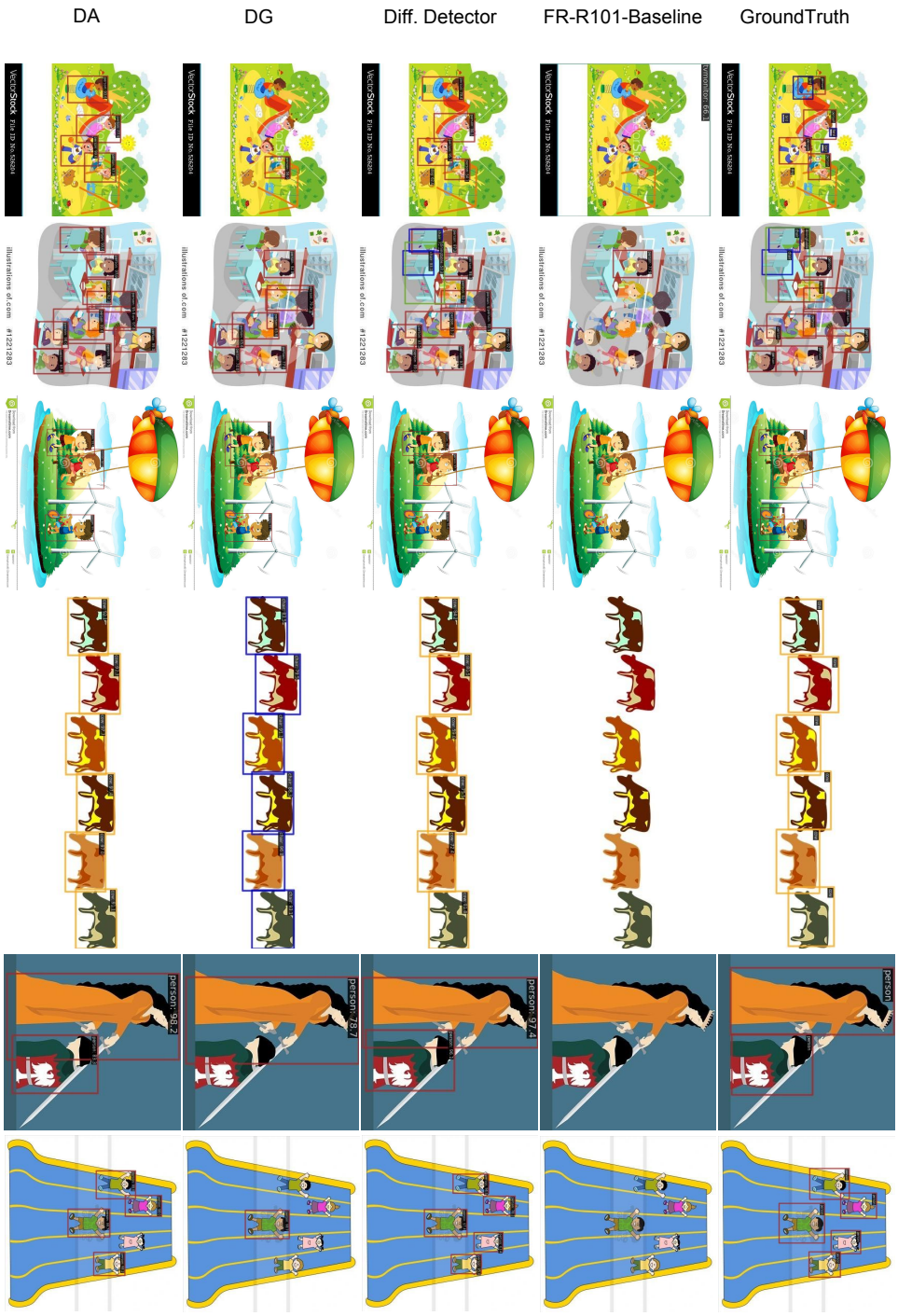

Figure 8. Qualitative prediction results on Clipart.

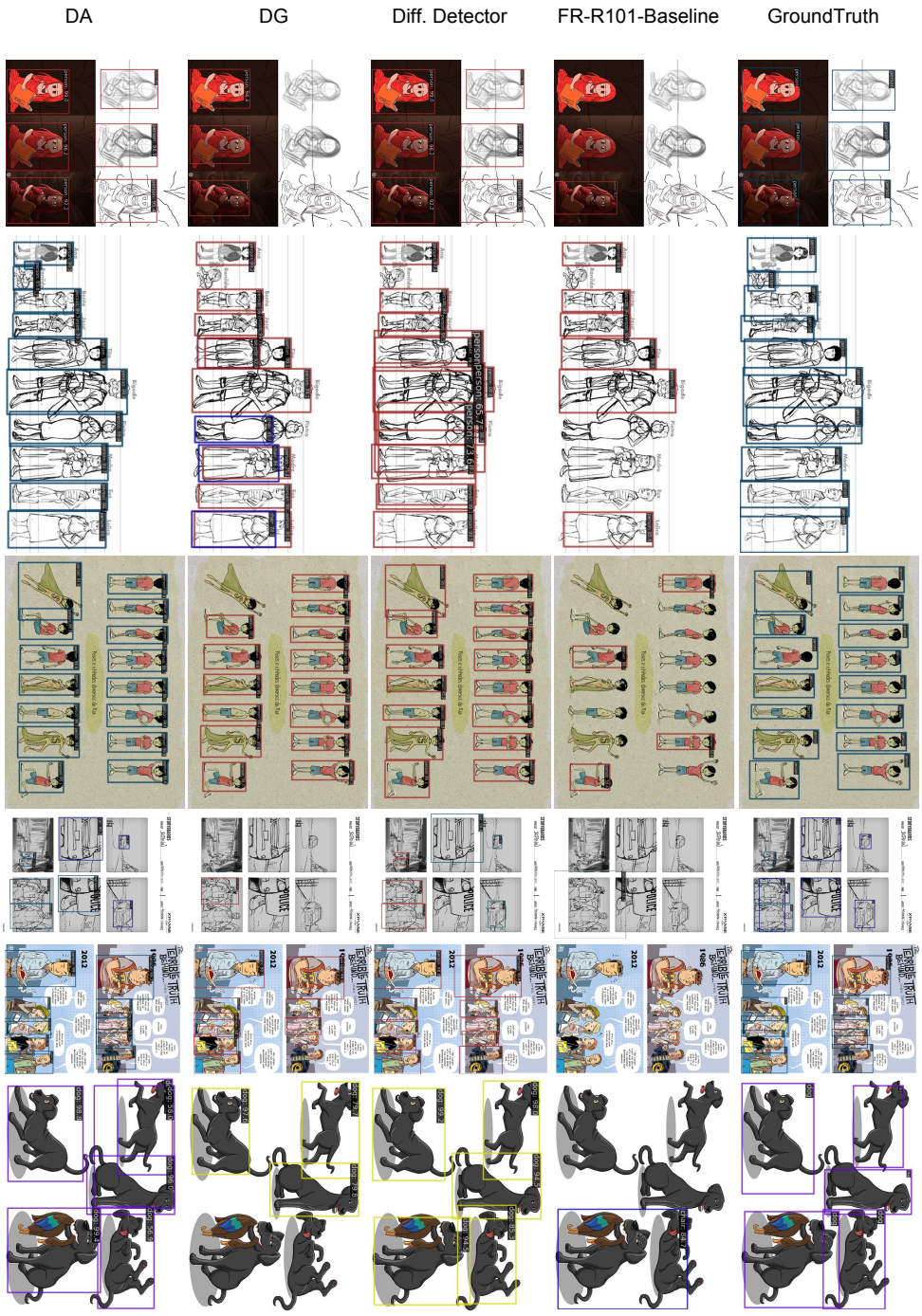

Figure 9. Qualitative prediction results on Comic.

Figure 10. Qualitative prediction results on Watercolor.

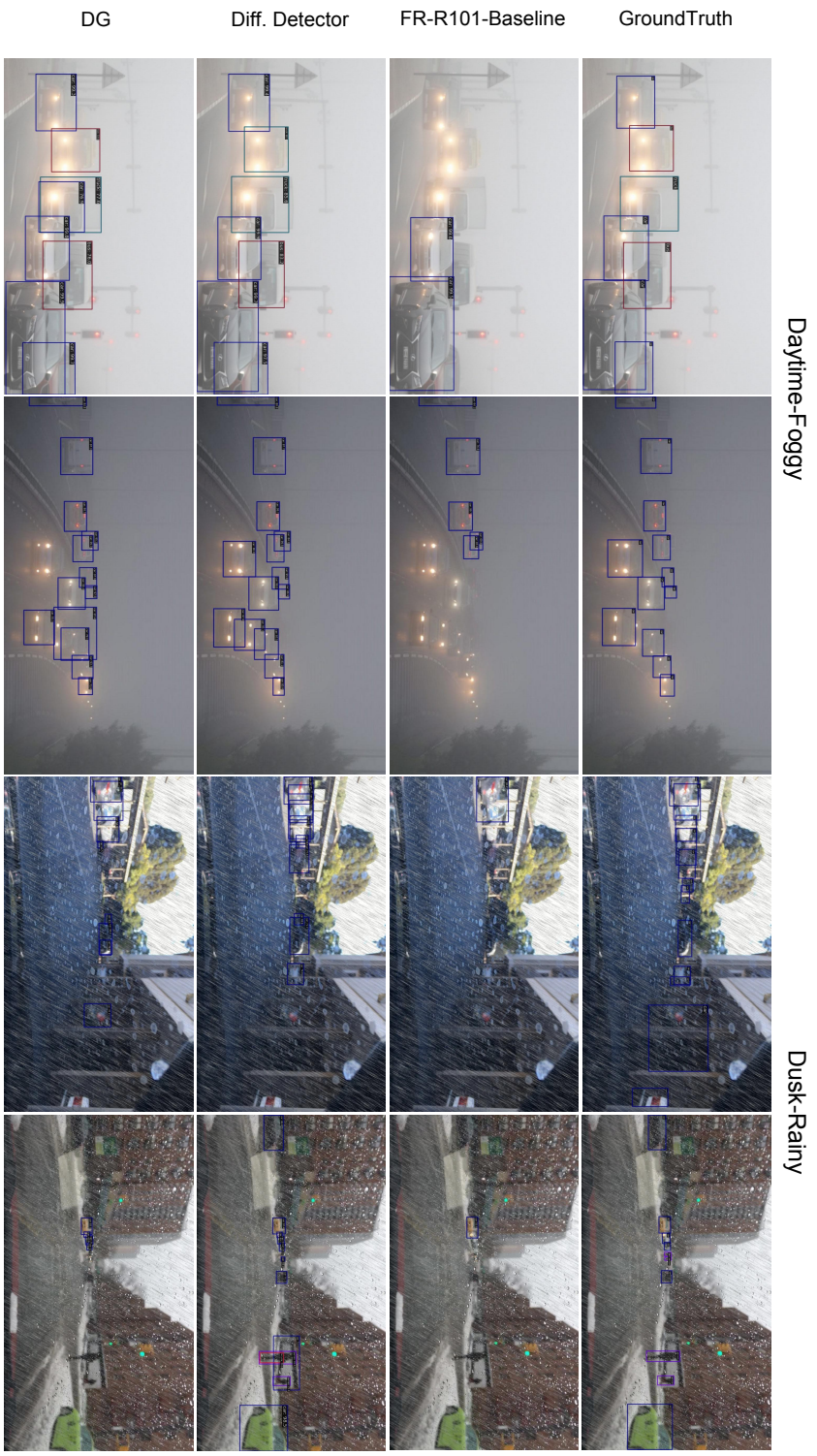

Figure 11. Qualitative prediction results on Daytime-Foggy and Dusk-Rainy.

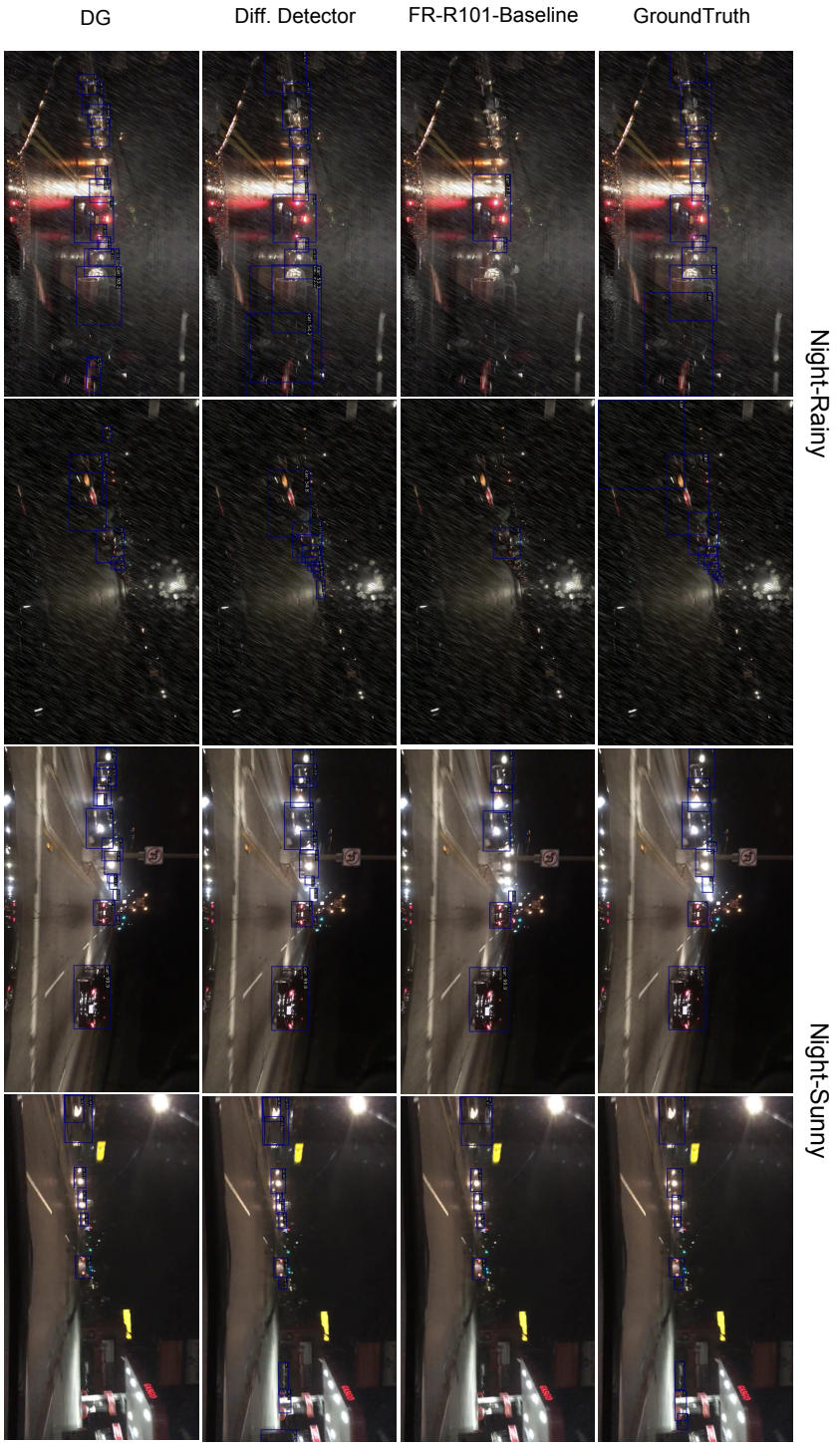

Figure 12. Qualitative prediction results on Night-Rainy and Night-Sunny.

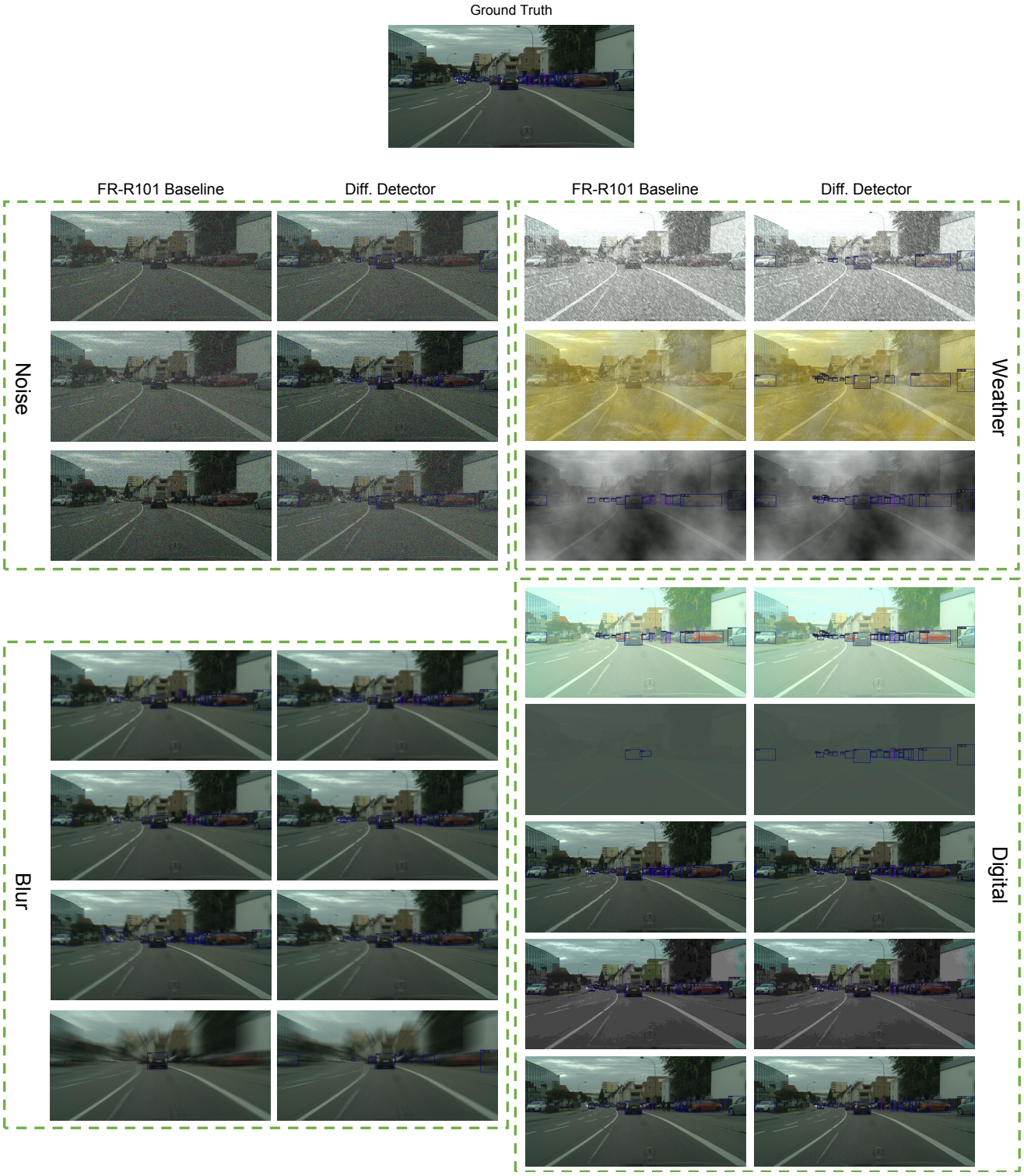

Figure 13. Qualitative prediction results on Cityscapes-Corruption, example 1.

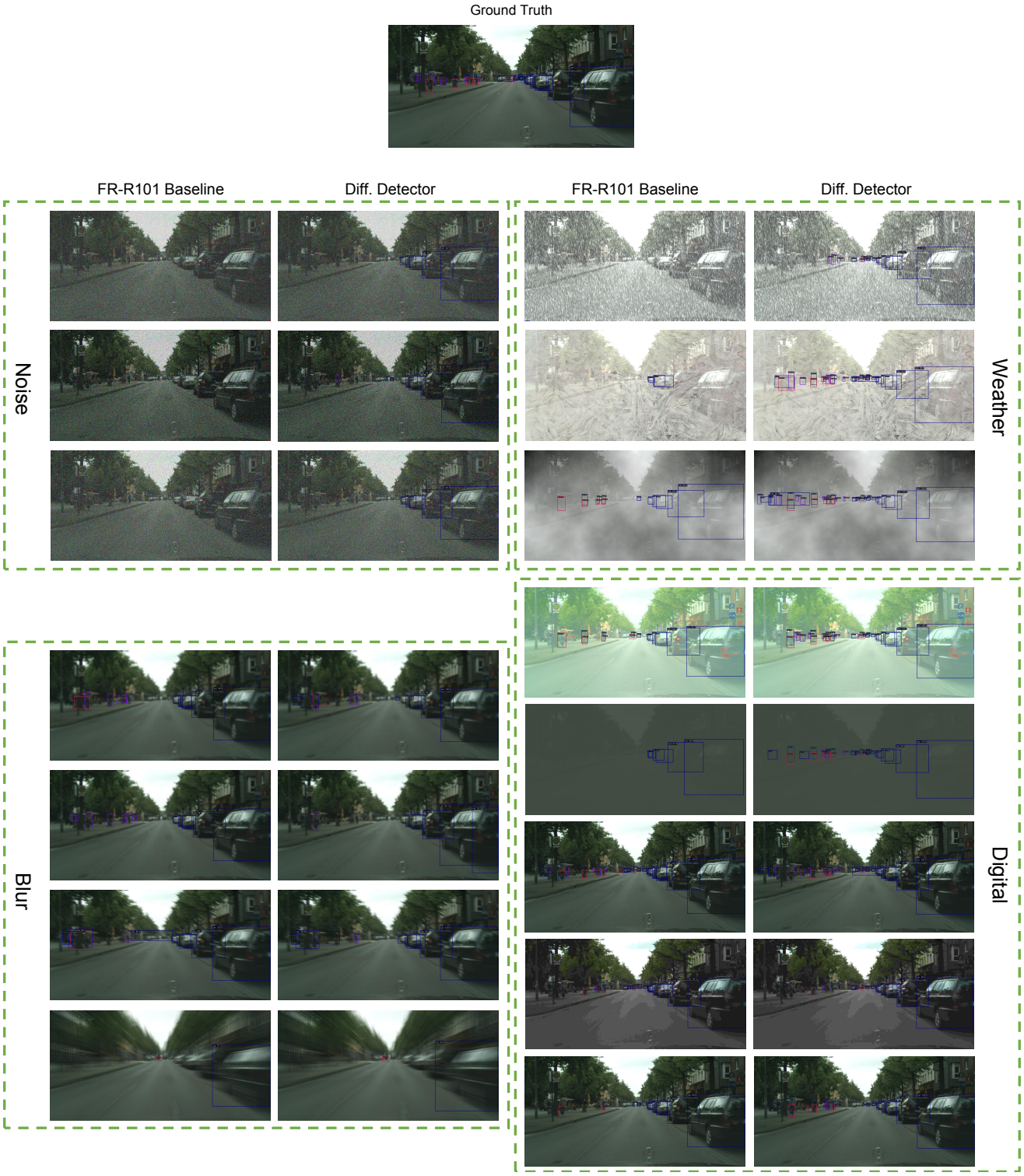

Figure 14. Qualitative prediction results on Cityscapes-Corruption, example 2.

References

[1] Daniel Bolya, Sean Foley, James Hays, and Judy Hoffman. Tide: A general toolbox for identifying object detection errors. In *ECCV*, 2020. 5, 6

[2] Shengcao Cao, Dhiraj Joshi, Liang-Yan Gui, and Yu-Xiong Wang. Contrastive mean teacher for domain adaptive object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23839–23848, 2023. 3

[3] Chaoqi Chen, Zebiao Zheng, Xinghao Ding, Yue Huang, and Qi Dou. Harmonizing transferability and discriminability for adapting object detectors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8869–8878, 2020. 3

[4] Chaoqi Chen, Jiongcheng Li, Zebiao Zheng, Yue Huang, Xinghao Ding, and Yizhou Yu. Dual bipartite graph learning: A general approach for domain adaptive object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2703–2712, 2021. 2, 3

[5] Chaoqi Chen, Zebiao Zheng, Yue Huang, Xinghao Ding, and Yizhou Yu. I3net: Implicit instance-invariant network for adapting one-stage object detectors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12576–12585, 2021. 2

[6] Yuhua Chen, Wen Li, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Domain adaptive faster r-cnn for object detection in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3339–3348, 2018. 2

[7] Yuhua Chen, Haoran Wang, Wen Li, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Scale-aware domain adaptive faster r-cnn. *International Journal of Computer Vision*, 129(7):2223–2243, 2021. 2, 3

[8] Sungha Choi, Sanghun Jung, Huiwon Yun, Joanne T Kim, Seungryong Kim, and Jaegul Choo. Robustnet: Improving domain generalization in urban-scene segmentation via instance selective whitening. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11580–11590, 2021. 4

[9] Muhammad Sohail Danish, Muhammad Haris Khan, Muhammad Akhtar Munir, M Saquib Sarfraz, and Mohsen Ali. Improving single domain-generalized object detection: A focus on diversification and alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17732–17742, 2024. 2, 3, 4

[10] Jinhong Deng, Wen Li, Yuhua Chen, and Lixin Duan. Unbiased mean teacher for cross-domain object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4091–4101, 2021. 2, 3

[11] A Enes Doruk and Hasan F Ates. Davimnet: Ssms-based domain adaptive object detection. *arXiv e-prints*, pages arXiv–2502, 2025. 2, 3

[12] Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. Scaling rectified flow transformers for high-resolution image synthesis. In *Forty-first International Conference on Machine Learning*, 2024. 1

[13] Dayan Guan, Jiaxing Huang, Aoran Xiao, Shijian Lu, and Yanpeng Cao. Uncertainty-aware unsupervised domain adaptation in object detection. *IEEE Transactions on Multimedia*, 24:2502–2514, 2021. 3

[14] Boyong He, Yuxiang Ji, Zhuoyue Tan, and Liaoni Wu. Diffusion domain teacher: Diffusion guided domain adaptive object detector. In *ACM Multimedia 2024*, 2024. 2, 3

[15] Boyong He, Yuxiang Ji, Qianwen Ye, Zhuoyue Tan, and Liaoni Wu. Generalized diffusion detector: Mining robust features from diffusion models for domain-generalized detection. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pages 9921–9932, 2025. 1, 2, 3, 4

[16] Zhenwei He and Hongsu Ni. Single-domain generalized object detection by balancing domain diversity and invariance. *arXiv preprint arXiv:2502.03835*, 2025. 4

[17] Lei Huang, Yi Zhou, Fan Zhu, Li Liu, and Ling Shao. Iterative normalization: Beyond standardization towards efficient whitening. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4874–4883, 2019. 4

[18] Junguang Jiang, Baixu Chen, Jianmin Wang, and Mingsheng Long. Decoupled adaptation for cross-domain object detection. In *International Conference on Learning Representations*, 2021. 2, 3

[19] Mikhail Kennerley, Jian-Gang Wang, Bharadwaj Veeravalli, and Robby T Tan. Cat: Exploiting inter-class dynamics for domain adaptive object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16541–16550, 2024. 3

[20] Seunghyeon Kim, Jaehoon Choi, Taekyung Kim, and Changick Kim. Self-training and adversarial background regularization for unsupervised domain adaptive one-stage object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6092–6101, 2019. 2

[21] Wooju Lee, Dasol Hong, Hyungtae Lim, and Hyun Myung. Object-aware domain generalization for object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2947–2955, 2024. 4

[22] Congcong Li, Dawei Du, Libo Zhang, Longyin Wen, Tiejian Luo, Yanjun Wu, and Pengfei Zhu. Spatial attention pyramid network for unsupervised domain adaptation. In *European Conference on Computer Vision*, pages 481–497, 2020. 3

[23] Deng Li, Aming Wu, Yaowei Wang, and Yahong Han. Prompt-driven dynamic object-centric learning for single domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17606–17615, 2024. 4

[24] Shuai Li, Jianqiang Huang, Xian-Sheng Hua, and Lei Zhang. Category dictionary guided unsupervised domain adaptation for object detection. In *Proceedings of the AAAI conference on artificial intelligence*, pages 1949–1957, 2021. 2

[25] Shuaifeng Li, Mao Ye, Xiatian Zhu, Lihua Zhou, and Lin Xiong. Source-free object detection by learning to overlook domain style. In *Proceedings of the IEEE/CVF Conference*

on Computer Vision and Pattern Recognition, pages 8014–8023, 2022. 2, 3

[26] Yu-Jhe Li, Xiaoliang Dai, Chih-Yao Ma, Yen-Cheng Liu, Kan Chen, Bichen Wu, Zijian He, Kris Kitani, and Peter Vajda. Cross-domain adaptive teacher for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7581–7590, 2022. 2, 3

[27] Yabo Liu, Jinghua Wang, Chao Huang, Yaowei Wang, and Yong Xu. Cigar: Cross-modality graph reasoning for domain adaptive object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23776–23786, 2023. 3

[28] Yajing Liu, Shijun Zhou, Xiyao Liu, Chunhui Hao, Baojie Fan, and Jiandong Tian. Unbiased faster r-cnn for single-source domain generalized object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 28838–28847, 2024. 4

[29] Xingang Pan, Ping Luo, Jianping Shi, and Xiaoou Tang. Two at once: Enhancing learning and generalization capacities via ibn-net. In *Proceedings of the european conference on computer vision (ECCV)*, pages 464–479, 2018. 4

[30] Xingang Pan, Xiaohang Zhan, Jianping Shi, Xiaoou Tang, and Ping Luo. Switchable whitening for deep representation learning. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1863–1871, 2019. 4

[31] Zhijie Rao, Jingcai Guo, Luyao Tang, Yue Huang, Xinghao Ding, and Song Guo. Srcd: Semantic reasoning with compound domains for single-domain generalized object detection. *IEEE Transactions on Neural Networks and Learning Systems*, 2024. 4

[32] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015. 1

[33] Kuniaki Saito, Yoshitaka Ushiku, Tatsuya Harada, and Kate Saenko. Strong-weak distribution alignment for adaptive object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6956–6965, 2019. 2, 3

[34] Vidit Vidit, Martin Engilberge, and Mathieu Salzmann. Clip the gap: A single domain generalization approach for object detection. In *CVPR*, pages 3219–3229, 2023. 4

[35] Aming Wu and Cheng Deng. Single-domain generalized object detection in urban scene via cyclic-disentangled self-distillation. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pages 847–856, 2022. 4

[36] Aming Wu, Yahong Han, Linchao Zhu, and Yi Yang. Instance-invariant domain adaptive object detection via progressive disentanglement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8):4178–4193, 2021. 2, 3

[37] Aming Wu, Rui Liu, Yahong Han, Linchao Zhu, and Yi Yang. Vector-decomposed disentanglement for domain-invariant object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9342–9351, 2021. 2

[38] Fan Wu, Jinling Gao, Lanqing Hong, Xinbing Wang, Chenghu Zhou, and Nanyang Ye. G-nas: Generalizable neural architecture search for single domain generalization object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 5958–5966, 2024. 4

[39] Xiaoran Xu, Jiangang Yang, Wenhui Shi, Siyuan Ding, Luqing Luo, and Jian Liu. Physaug: A physical-guided and frequency-based data augmentation for single-domain generalized object detection. *arXiv preprint arXiv:2412.11807*, 2024. 4

[40] Libo Zhang, Wenzhang Zhou, Heng Fan, Tiejian Luo, and Haibin Ling. Robust domain adaptive object detection with unified multi-granularity alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 2, 3

[41] Liang Zhao and Limin Wang. Task-specific inconsistency alignment for domain adaptive object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14217–14226, 2022. 3

[42] Zhen Zhao, Yuhong Guo, Haifeng Shen, and Jieping Ye. Adaptive object detection with dual multi-label prediction. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVIII 16*, pages 54–69. Springer, 2020. 2