

CVGL: Causal Learning and Geometric Topology

Appendix

A Extra Dataset

CVUSA-C and CVACT_val-C. To evaluate the robustness of our model under real-world corruptions, we conduct experiments on the Weather and Blur settings from the CVUSA-C and CVACT_val-C dataset. Following the protocol in Zhang, we report the average Recall@1 across five corruption severity levels, denoted as $R@1_{cor}$. As shown in Table 1 and Table 2, our method consistently outperforms the strongest baseline, EP-BEV, across all corruption levels. It achieves state-of-the-art performance under both weather and blur corruptions.

The performance gain comes from our causal enhancement mechanism. By perturbing non-causal components—mainly in the ultra-low and ultra-high frequency bands—the model learns to focus on stable and causally relevant features. This helps suppress spurious correlations that are sensitive to noise and environmental changes, while preserving semantic structures that remain consistent. As a result, the model shows improved robustness against common corruptions.

Table 1: Comparisons with state-of-the-art models under various weather and blur conditions on the CVUSA-C dataset. We report Recall@1 performance for each corruption type. The $R@1_{cor}$ column shows the average Recall@1 across all corruption types. The best performance in each column is highlighted in bold.

Method	Clean	Weather					Blur				$R@1_{cor}$
		Snow	Frost	Fog	Bright	Spatter	Defocus	Glass	Motion	Zoom	
CVM-Net	22.47	0.86	8.42	8.37	13.75	6.11	1.06	4.81	1.47	0.23	5.01
OriCNN	40.79	7.36	6.51	7.57	21.69	22.01	20.46	26.10	19.60	10.32	15.74
SAFA	89.84	19.32	60.42	67.63	81.96	49.86	51.24	80.56	55.49	11.44	53.10
CVFT	61.43	8.00	30.79	47.46	47.54	27.63	24.55	44.93	34.89	8.17	30.44
DSM	91.96	24.24	64.44	84.08	82.44	57.58	62.48	84.52	66.02	25.15	61.22
L2LTR	94.05	67.19	85.00	92.64	91.61	75.24	88.35	93.13	89.33	42.07	80.51
TransGeo	94.08	29.39	69.50	70.89	85.01	64.26	80.97	92.16	85.96	40.97	68.79
GeoDTR	95.43	44.20	84.95	92.80	93.55	73.14	82.64	93.29	76.80	27.19	74.28
Sample4Geo	98.68	79.63	95.46	97.64	96.80	88.73	95.76	97.33	94.85	45.95	88.02
EP-BEV	97.41	55.75	87.62	94.63	95.05	84.71	82.67	95.11	91.39	50.32	81.92
Ours	98.73	79.66	94.90	97.72	96.81	91.48	89.92	97.45	95.63	62.33	89.54
Ours ($\gamma = 0.5$)	98.85	82.30	95.81	98.06	98.11	90.91	88.44	98.45	95.98	60.09	89.79

VIGOR. On VIGOR Table 3, our method ranks second among existing approaches. The performance gap is attributed to VIGOR’s unique challenges: the one-to-many retrieval protocol and street images with high density of dynamic objects that create noisy BEV representations, degrading geometric cue effectiveness. We plan to explore VIGOR-specific adaptations in future work.

B Visualization of Corrupted Images

To further illustrate the generalization ability of our method compared to the baseline, we visualize corrupted images under two categories: weather and noise. The weather corruptions include snow,

Table 2: Comparisons with state-of-the-art models under various weather and blur conditions on the CVACT_val-C dataset. We report Recall@1 performance for each corruption type. The $R@1_{cor}$ column shows the average Recall@1 across all corruption types. The best performance in each column is highlighted in bold.

Method	Clean	Weather					Blur				$R@1_{cor}$
		Snow	Frost	Fog	Bright	Spatter	Defocus	Glass	Motion	Zoom	
OriCNN	46.96	13.94	6.13	3.78	29.45	40.54	31.71	39.99	37.58	24.89	25.33
SAFA	81.03	20.03	31.66	33.19	66.99	45.60	39.83	72.87	49.86	4.62	40.52
CVFT	61.05	15.00	22.32	42.53	47.60	37.25	31.30	53.88	36.91	4.10	32.32
DSM	82.49	31.95	51.70	70.43	69.48	52.35	57.35	80.16	67.38	15.34	55.13
L2LTR	84.89	71.03	77.93	83.50	81.17	73.78	83.98	85.07	84.00	49.79	76.69
TransGeo	84.95	47.65	58.51	32.91	72.67	67.13	81.43	84.83	81.80	36.34	62.59
GeoDTR	86.21	48.24	71.74	83.26	84.60	61.39	79.11	85.51	73.44	8.26	66.17
Sample4Geo	90.81	78.69	84.68	86.97	88.89	84.97	89.17	90.22	89.03	50.00	82.51
DReSS	91.32	81.23	86.50	88.14	89.97	85.48	90.01	90.89	90.12	49.98	83.59
EP-BEV	90.61	76.22	82.90	86.05	88.98	86.94	89.20	90.31	89.58	60.50	83.41
Ours	91.61	85.26	87.81	89.47	90.47	87.50	90.40	91.19	90.62	64.60	86.37
Ours ($\gamma = 0.5$)	91.97	85.08	88.71	90.42	91.24	88.34	90.58	91.73	91.18	68.07	87.26

Table 3: Performance comparison on VIGOR. The top two scores are emphasized, with the highest in bold. (\dagger methods that use polar transformation.)

Model	Same-area					Cross-area				
	R@1	R@5	R@10	R@1%	Hit Rate	R@1	R@5	R@10	R@1%	Hit Rate
SAFA \dagger	33.93	58.42	68.12	98.24	36.87	8.20	19.59	26.36	77.61	8.85
TransGeo	61.48	87.54	91.88	99.56	73.09	18.99	38.24	46.91	88.94	21.21
SAIG-D	65.23	88.08	-	99.68	74.11	33.05	55.94	-	94.64	36.71
Samp4G	77.86	95.66	97.21	99.61	89.82	61.70	83.50	88.00	98.17	69.87
EP-BEV(384×384)	76.08	94.71	96.59	99.64	87.01	60.03	81.63	86.34	97.64	67.88
Ours	76.58	94.18	96.14	99.49	87.80	60.33	81.82	86.54	97.32	68.10
Ours ($\gamma = 0.5$)	77.18	94.57	96.42	99.54	87.95	60.70	82.36	86.92	97.82	68.40

frost, fog, brightness, and spatter, where snow is already presented in the main text, while the noise corruptions consist of Gaussian, impulse, shot, and speckle. We select corrupted images at severity level 5 for heatmap visualization. As shown in Figure 1 and Figure 2, our method demonstrates remarkable generalization ability, maintaining focus on task-relevant discriminative features and regions even under severe perturbations. In contrast, the baseline almost completely fails, as it tends to attend to irrelevant cues such as the sky. These visualizations provide clear evidence of the superior generalization of our approach and further highlight its reliability in practical applications.

C Hyperparameter Analysis of γ

We conducted a hyperparameter study on the parameter γ . The results demonstrate that increasing γ within a reasonable range benefits the model’s generalization ability, while an excessively large value may lead to training collapse. Therefore, in the main text we adopt a conservative choice of $\gamma = 0.1$, while also reporting the best-performing result at $\gamma = 0.5$.

Table 4: Hyperparameter study of γ on the CVACT_val-C-ALL dataset. The best results are highlighted in bold.

γ	R@1	R@5	R@10	R@1%
0.1 (default)	88.68	95.58	96.66	98.49
0.3	88.92	95.78	96.84	98.46
0.5	89.49	95.84	96.92	98.49
0.7	89.25	95.83	96.90	98.63

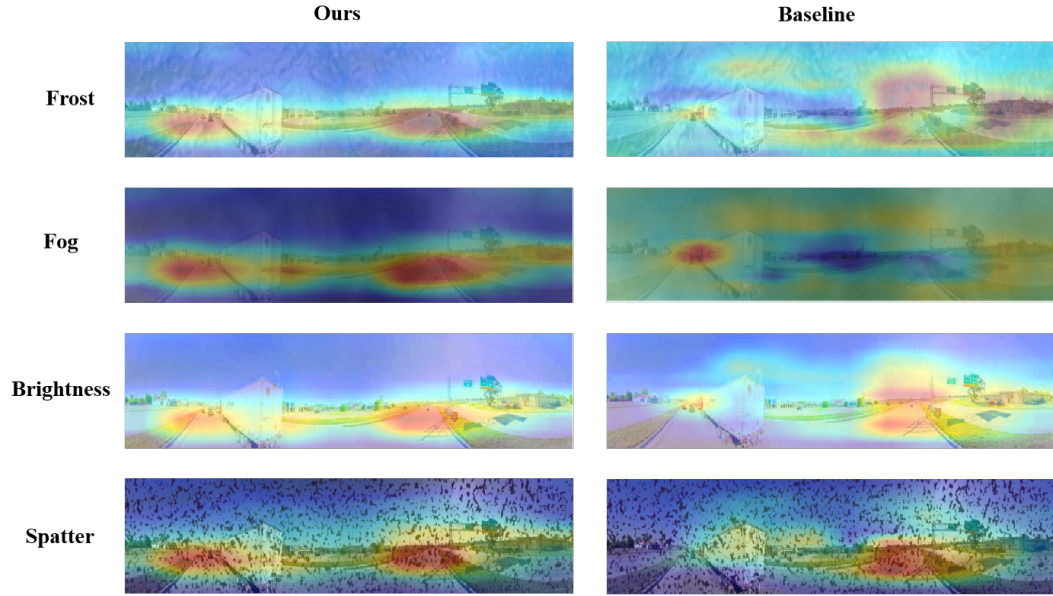


Figure 1: Heatmap visualizations on CVUSA under weather settings. Our method, compared to the baseline, focuses more on task-relevant structural information.

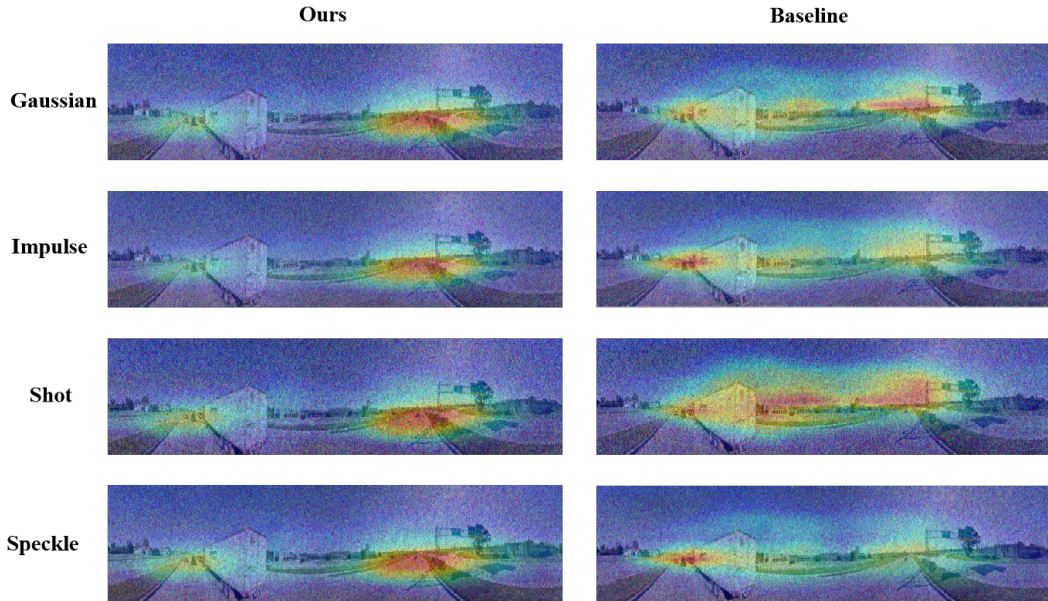


Figure 2: Heatmap visualizations on CVUSA under noise settings. Our method, compared to the baseline, focuses more on task-relevant structural information.