Benchmarking the Robustness of Cross-view Geo-localization Models

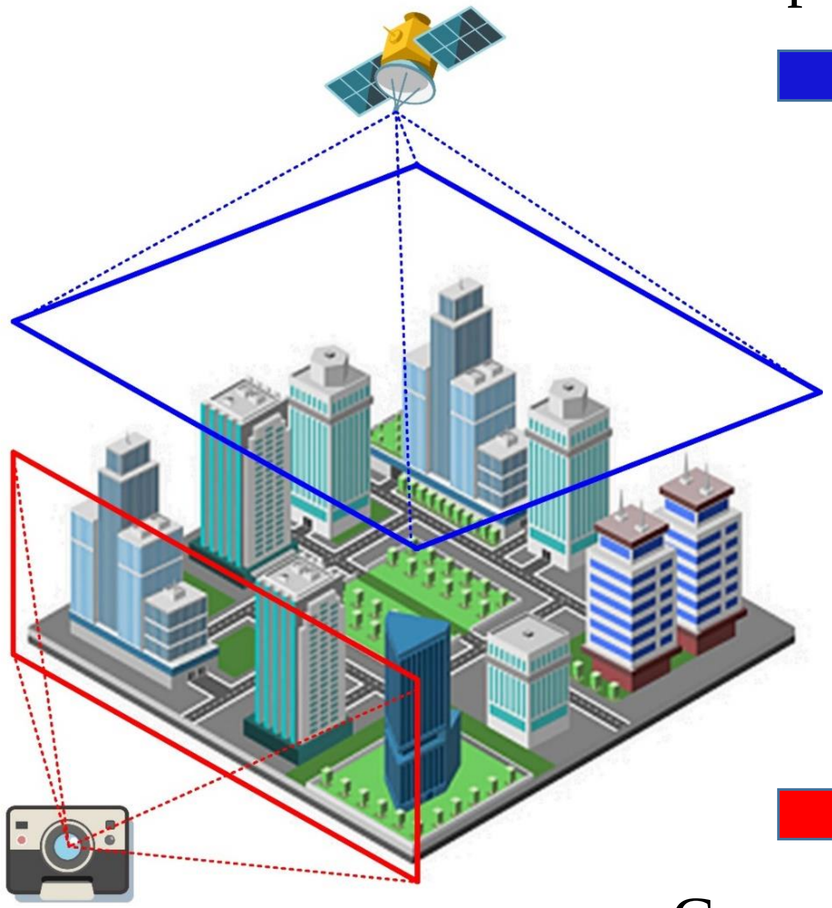
Qingwang Zhang, Yingying Zhu*
College of Computer Science and Software Engineering
Shenzhen University, China

Speaker: Qingwang Zhang

*Corresponding author

Background

What is Cross-view Geo-localization?



Top-view

Geotagged aerial images as Reference Image Database

The images from CVUSA dataset

Ground-view

A ground image as Query

Datasets

CVUSA

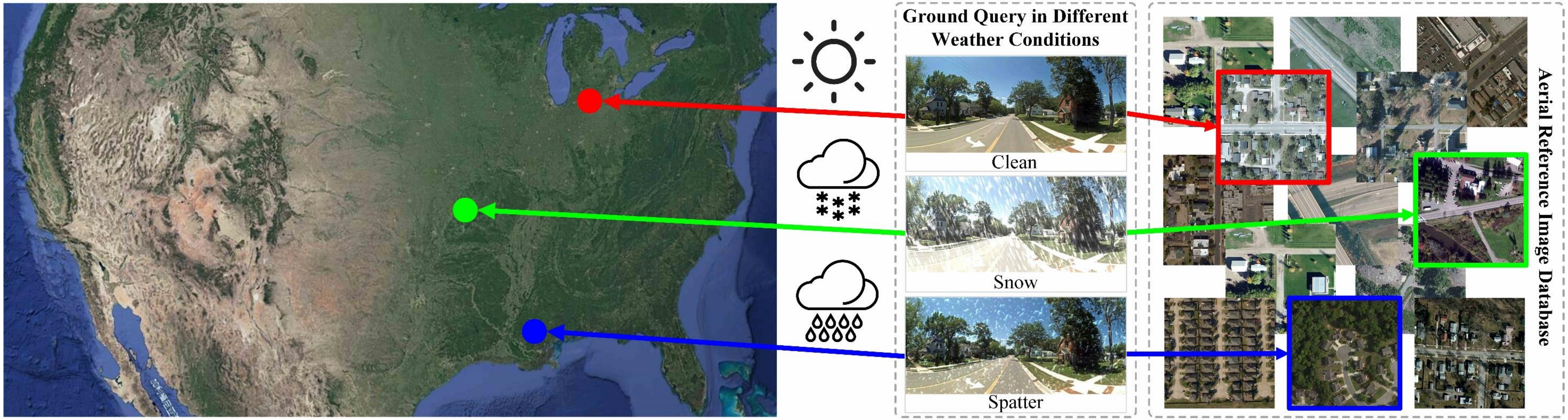
44,416 Pairs of Ground-Aerial Images
Acquired in suburban America

CVACT

137,218 pairs of Ground-Aerial Images
Acquired in suburban Australia

Workman, Scott, Richard Souvenir, and Nathan Jacobs. "Wide-area image geolocalization with aerial reference imagery." Proceedings of the IEEE International Conference on Computer Vision. 2015.
Liu, Liu, and Hongdong Li. "Lending orientation to neural networks for cross-view geo-localization." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.

Problem Statement

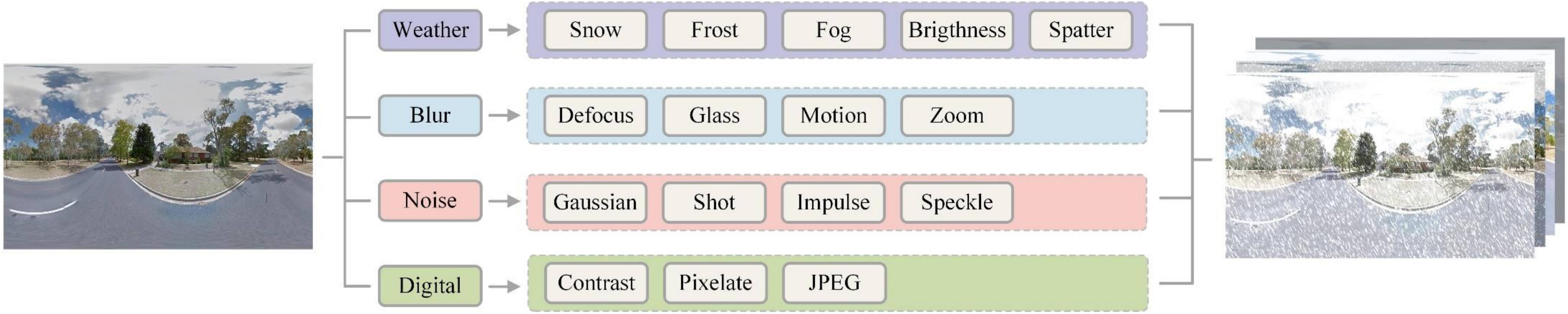


Existing cross-view geo-localization models fail when the ground query image is corrupted

Contributions

- To the best of our knowledge, we have benchmarked the robustness of state-of-the-art cross-view geo-localization models against real-world data corruption challenges for the first time. We have generated about 1.5 million corrupted images based on the CVUSA and CVACT datasets to establish benchmarks for assessing the robustness of cross-view geo-localization models.
- Based on the benchmark study, we have several new insights: (1) in most cases, the clean performance ($R@K_{clean}$) of a model is positively, but not absolutely, correlated with its robustness; (2) snow, spatter, and zoom blur more significantly affect the robustness of various models compared to other corruptions; (3) models trained on more intricate scenarios (e.g., CVACT) exhibit better robustness.
- Introducing stylization and histogram equalization as data augmentation techniques, along with our proposed training strategy, significantly enhances the robustness of various cross-view geo-localization models.

Image Corruption



The proposed fine-grained and comprehensive robustness benchmarks. Each corruption category encompasses 5 severity levels.

Dong, Yinpeng, et al. "Benchmarking robustness of 3d object detection to common corruptions." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.

Image Corruption

(a) Comparison between clean and corrupted images.

(b) Comparison of different severity levels, taking snow as an example.

Robustness Enhancement Methods

Visualization of Stylization / CLAHE applied to the CVUSA training set

Michaelis, Claudio, et al. "Benchmarking robustness in object detection: Autonomous driving when winter is coming." arXiv preprint arXiv:1907.07484 (2019).

Corruption Robustness Benchmarks

Detailed information on the proposed corruption robustness benchmarks

	Fine-grained		Comprehensive		
	CVUSA-C	CVACT_val-C	CVUSA-C-ALL	CVACT_val-C-ALL	CVACT_test-C-ALL
Number of original validation / test ground images	8,884	8,884	8,884	8,884	92,802
Whether or not evaluation subsets are generated for each corruption	✔	✔	✘	✘	✘
Whether all corruptions are included in an independent subset	✘	✘	✔	✔	✔
Number of validation / test ground images for our benchmark	8,884 × 16 × 5	8,884 × 16 × 5	8,884	8,884	92,802
Storage space	∼ 39 GB	∼ 178 GB	∼ 0.5 GB	∼ 2 GB	∼ 21 GB

$$\text{R@}K_{\text{cor}} = \frac{1}{|\mathcal{C}|}\sum_{c\in\mathcal{C}}\frac{1}{5}\sum_{s=1}^{5}\text{R@}K_{c,s} \tag{1}$$

$$\text{RCE}_{c,s} = \frac{\text{R@}K_{\text{clean}} - \text{R@}K_{c,s}}{\text{R@}K_{\text{clean}}};\ \text{RCE} = \frac{\text{R@}K_{\text{clean}} - \text{R@}K_{\text{cor}}}{\text{R@}K_{\text{clean}}} \tag{2}$$

Benchmarking Results

Fine-grained Corruption Robustness Benchmarking Results

Experimental results of 8 cross-view geo-localization methods on the CVUSA-C benchmark.

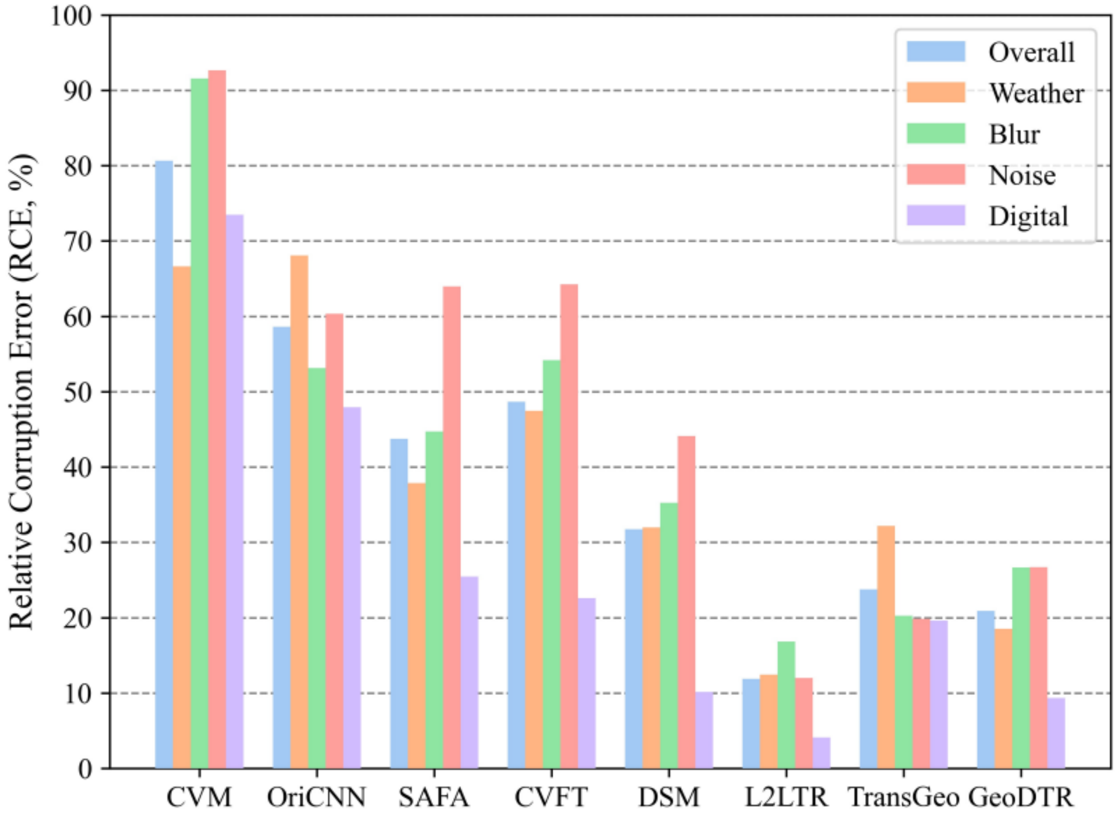
Method	Clean	CVUSA-C																
		Weather					Blur				Noise				Digital			R@1$_{cor}$
		Snow	Frost	Fog	Bright	Spatter	Defocus	Glass	Motion	Zoom	Gaussian	Shot	Impulse	Speckle	Contrast	Pixel	JPEG	
CVM-Net [17]	22.47	0.86	8.42	8.37	13.75	6.11	1.06	4.81	1.47	0.23	1.82	1.18	1.28	2.32	4.75	6.89	6.23	4.35
OriCNN [21]	40.79	7.36	6.51	7.57	21.69	22.01	20.46	26.10	19.60	10.32	17.24	13.95	19.40	14.09	7.94	28.51	27.27	16.88
SAFA [33]	89.84	19.32	60.42	67.63	81.96	49.86	51.24	80.56	55.49	11.44	33.04	28.51	30.37	37.59	31.67	88.05	81.15	50.52
CVFT [35]	61.43	8.00	30.79	47.46	47.54	27.63	24.55	44.93	34.89	8.17	21.83	19.19	20.56	26.25	38.28	57.25	47.11	31.53
DSM [34]	91.96	24.24	64.44	84.08	82.44	57.58	62.48	84.52	66.02	25.15	49.55	46.40	48.84	60.83	72.11	90.20	85.56	62.78
L2LTR [42]	94.05	**67.19**	**85.00**	92.64	91.61	**75.24**	**88.35**	93.13	**89.33**	**42.07**	**81.32**	**80.29**	**82.88**	**86.54**	**86.36**	93.64	**90.56**	**82.88**
TransGeo [45]	94.08	29.39	69.50	70.89	85.01	64.26	80.97	92.16	85.96	40.97	72.95	70.27	74.32	83.99	43.01	93.74	90.13	71.72
GeoDTR [43]	**95.43**	44.20	84.95	**92.80**	**93.55**	73.14	82.64	**93.29**	76.80	27.19	68.40	64.45	68.53	78.28	74.80	**94.45**	90.20	75.48

Experimental results of 7 cross-view geo-localization methods on the CVACT_val-C benchmark.

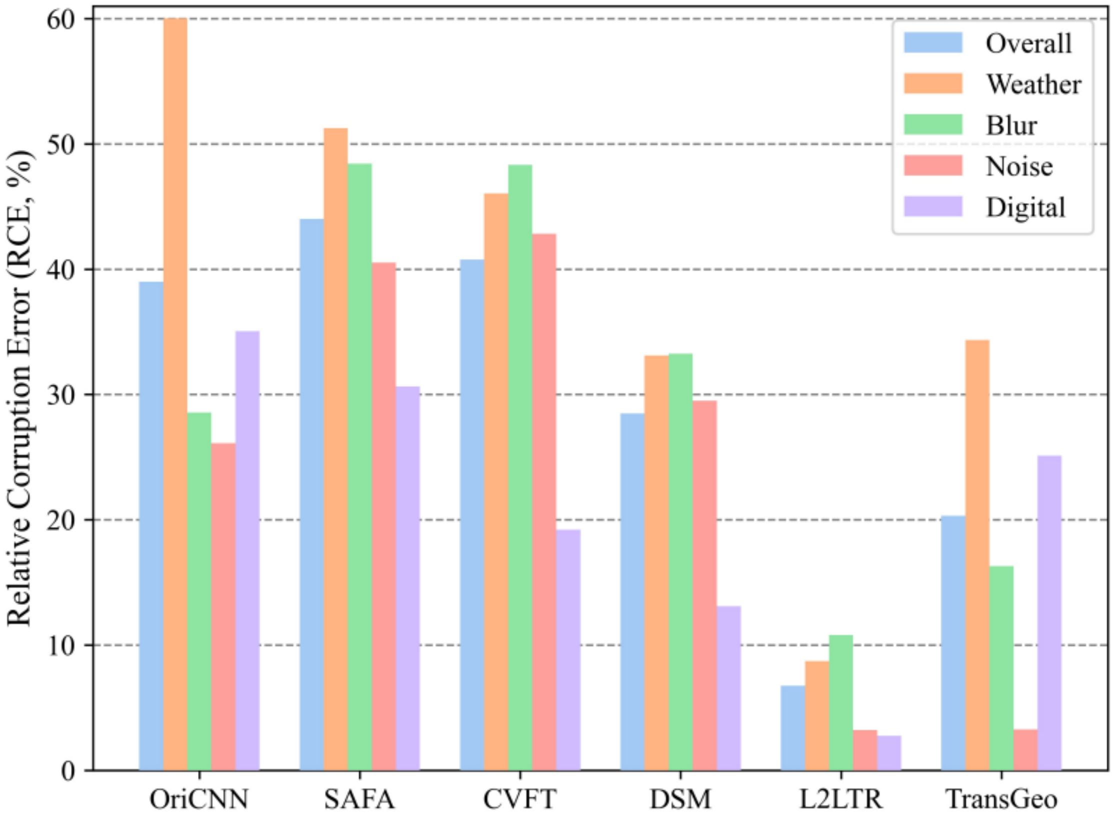
Method	Clean	CVACT_val-C																
		Weather					Blur				Noise				Digital			R@1$_{cor}$
		Snow	Frost	Fog	Bright	Spatter	Defocus	Glass	Motion	Zoom	Gaussian	Shot	Impulse	Speckle	Contrast	Pixel	JPEG	
OriCNN [21]	46.96	13.94	6.13	3.78	29.45	40.54	31.71	39.99	37.58	24.89	34.24	32.27	39.01	33.28	4.56	44.38	42.56	28.65
SAFA [33]	81.03	20.03	31.66	33.19	66.99	45.60	39.83	72.87	49.86	4.62	48.66	43.68	48.82	51.61	15.91	76.90	75.83	45.38
CVFT [35]	61.05	15.00	22.32	42.53	47.60	37.25	31.30	53.88	36.91	4.10	35.68	30.80	36.32	36.84	31.79	58.21	57.97	36.16
DSM [34]	82.49	31.95	51.70	70.43	69.48	52.35	57.35	80.16	67.38	15.34	58.34	53.05	58.18	63.06	52.79	81.72	80.55	58.99
L2LTR [42]	84.89	**71.03**	**77.93**	**83.50**	81.17	**73.78**	**83.98**	85.07	**84.00**	**49.79**	**82.20**	**81.19**	**82.98**	82.23	**79.15**	85.07	83.40	**79.15**
TransGeo [45]	84.95	47.65	58.51	32.91	72.67	67.13	81.43	84.83	81.80	36.34	81.96	80.86	82.84	**83.01**	22.18	84.92	83.74	67.68
GeoDTR [43]	**86.21**	48.24	71.74	83.26	**84.60**	61.39	79.11	**85.51**	73.44	8.26	75.44	73.99	77.06	80.23	55.48	**86.01**	**85.19**	70.56

Benchmarking Results

Fine-grained Corruption Robustness Benchmarking Results



(a) CVUSA-C



(b) CVACT_val-C

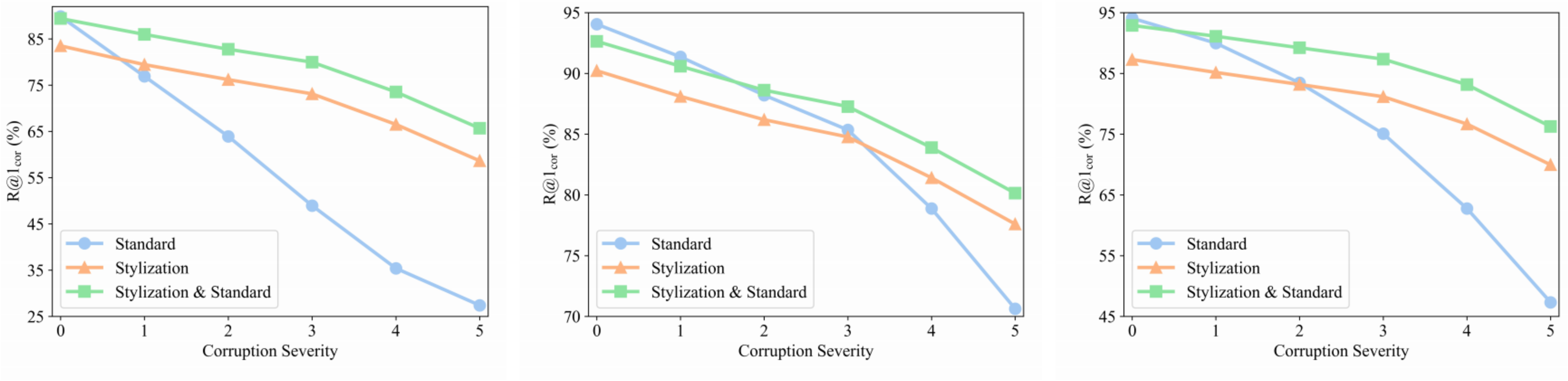
The Relative Corruption Error (RCE) of different cross-view geo-localization models on CVUSA-C and CVACT_val-C benchmarks.

Benchmarking Results

Experimental results of cross-view geo-localization methods on CVUSA-C-ALL, CVACT_val-C-ALL, and CVACT_test-C-ALL benchmarks.

Method	CVUSA-C-ALL				CVACT_val-C-ALL				CVACT_test-C-ALL			
	$R@1_{all}$	$R@5_{all}$	$R@10_{all}$	$R@1\%_{all}$	$R@1_{all}$	$R@5_{all}$	$R@10_{all}$	$R@1\%_{all}$	$R@1_{all}$	$R@5_{all}$	$R@10_{all}$	$R@1\%_{all}$
CVM-Net [17]	6.09	16.05	23.14	52.51	-	-	-	-	-	-	-	-
OriCNN [21]	9.38	22.26	30.04	58.99	15.31	28.31	35.21	58.39	3.69	8.33	11.04	43.93
SAFA [33]	63.68	78.08	82.82	93.91	56.72	73.60	78.59	91.32	31.18	52.06	58.60	90.41
CVFT [35]	41.05	64.01	72.64	91.37	45.69	66.45	72.97	88.38	22.82	43.48	51.07	88.99
DSM [34]	75.27	86.26	89.42	95.07	70.04	82.81	85.86	93.51	47.13	68.41	73.52	93.18
L2LTR [42]	**87.93**	**95.45**	**97.01**	**99.01**	**82.13**	**93.34**	**94.93**	**98.10**	**57.20**	**82.59**	**87.23**	**98.09**
TransGeo [45]	82.72	91.95	94.03	97.92	74.04	86.19	89.10	94.98	52.18	74.35	78.99	95.03
GeoDTR [43]	84.64	93.29	95.01	98.24	77.40	88.95	91.28	95.91	52.87	78.84	83.17	95.84

Robustness Enhancement Results

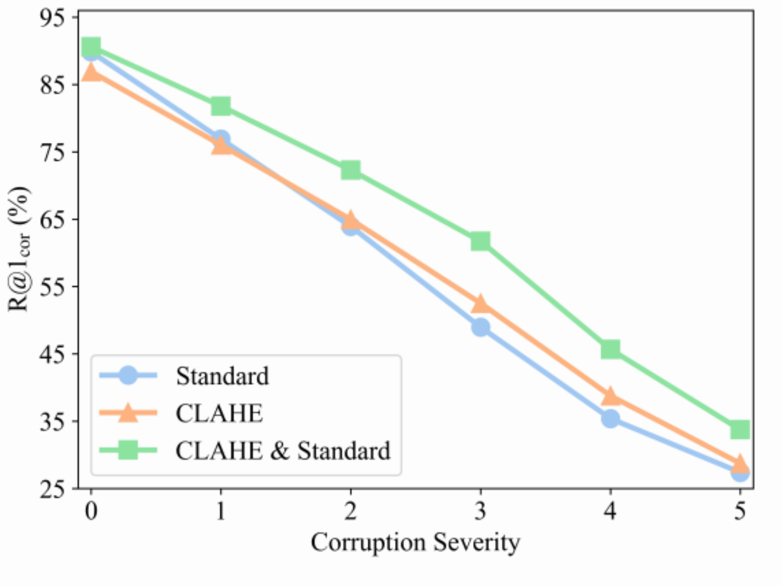


(a) SAFA **(b)** L2LTR **(c)** TransGeo

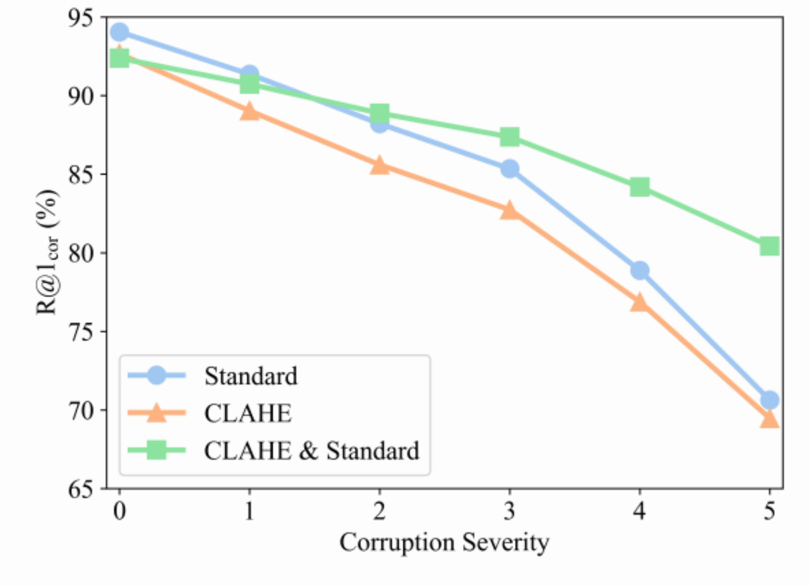
Stylization improves the robustness of SAFA, L2LTR, and TransGeo on the CVUSA-C benchmark

Shi, Yujiao, et al. "Spatial-aware feature aggregation for image based cross-view geo-localization." Advances in Neural Information Processing Systems 32 (2019).
Yang, Hongji, Xiufan Lu, and Yingying Zhu. "Cross-view geo-localization with layer-to-layer transformer." Advances in Neural Information Processing Systems 34 (2021): 29009-29020.
Zhu, Sijie, Mubarak Shah, and Chen Chen. "Transgeo: Transformer is all you need for cross-view image geo-localization." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

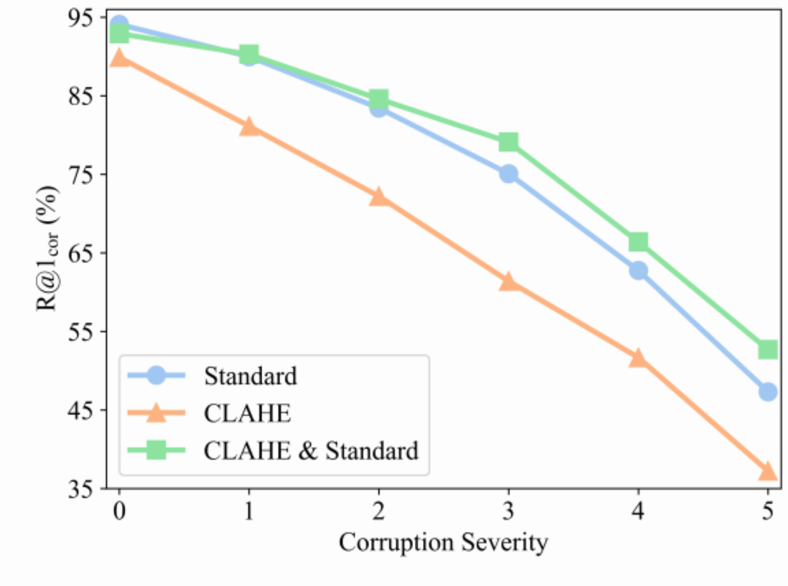
Robustness Enhancement Results



(a) SAFA



(b) L2LTR



(c) TransGeo

CLAHE improves the robustness of SAFA, L2LTR and TransGeo on the CVUSA-C benchmark

Shi, Yujiao, et al. "Spatial-aware feature aggregation for image based cross-view geo-localization." Advances in Neural Information Processing Systems 32 (2019).
Yang, Hongji, Xiufan Lu, and Yingying Zhu. "Cross-view geo-localization with layer-to-layer transformer." Advances in Neural Information Processing Systems 34 (2021): 29009-29020.
Zhu, Sijie, Mubarak Shah, and Chen Chen. "Transgeo: Transformer is all you need for cross-view image geo-localization." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

Further Discussion

- **Real-World Corruption Complexity:** Real-world corruptions are extensive and complex, and not all can be enumerated.
- **Designed Corruptions:** 16 corruption types, each with 5 severity levels, were systematically designed to generate rich data for research.
- **Synthetic Imagery Advantages:** Synthetic images offer controllable and diverse data by adjusting conditions like weather and scene types, providing safer and more ethical research alternatives.
- **Practical Testbed:** The benchmarks created serve as a practical testbed for evaluating the robustness of cross-view geo-localization models.
- **Broad Applicability:** While focused on cross-view geo-localization, these robustness benchmarks are applicable to other research areas like cross-view image synthesis, camera pose estimation, and autonomous driving.

Conclusion

- This paper systematically investigates the impact of corruption data on cross-view geo-localization models, which is a challenge previously overlooked in the context of cross-view geo-localization studies.
- We propose two fine-grained corruption robustness benchmarks (CVUSA-C and CVACT_val-C) and three comprehensive corruption robustness benchmarks (CVUSA-C-ALL, CVACT_val-C-ALL, and CVACT_test-C-ALL) for the cross-view geo-localization.
- Extensive experiments are conducted to evaluate existing classical methods on these corruption robustness benchmarks, revealing new insights.
- Furthermore, we introduce two simple techniques (stylization and histogram equalization) and the training strategy to effectively enhance robustness.

Thanks!

If you have any questions, please contact

zhangqingwang2022@email.szu.edu.cn

Project Homepage