SUPPLEMENTARY MATERIAL

In this supplementary material, we provide the following sections for better understanding the paper:

A. Examples of All Corruption Types.

B. More Fine-grained Robustness Evaluation Benchmark Experimental Results.

C. Comprehensive Robustness Evaluation Benchmark Experiment Results.

D. Histogram Equalization for Robustness Enhancement.

E. Limitation and Discussion.

A EXAMPLES OF ALL CORRUPTION TYPES

In the main paper, we present examples of various corruption types. Figure 1 in this section illustrates all corruption types, each with a corruption severity level of 2.

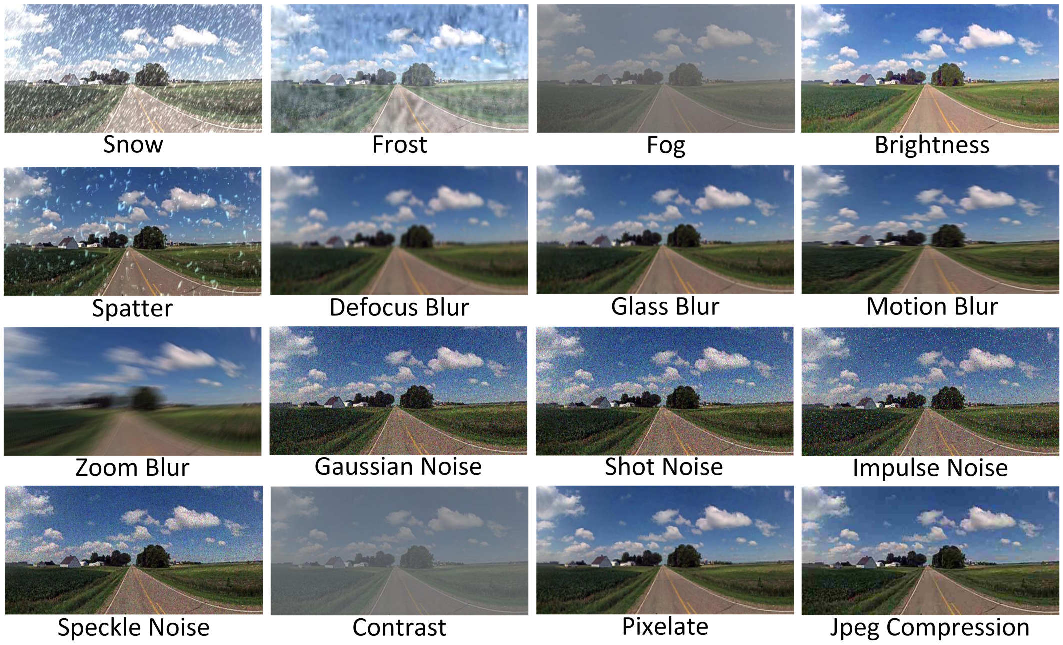


Figure 1: Our generated dataset encompasses 16 distinct corruption types, derived from Weather (Snow, Frost, Fog, Brightness, and Spatter), Blur (Defocus Blur, Glass Blur, Motion Blur, and Zoom Blur), Noise (Gaussian Noise, Shot Noise, Impulse Noise, and Speckle Noise), and Digital (Contrast, Pixelate, and Jpeg Compression) corruption categories (the corrupted images in the figure are sourced from CVUSA-C).

B MORE FINE-GRAINED ROBUSTNESS EVALUATION BENCHMARK EXPERIMENTAL RESULTS

In the main paper, we present the performance of 8 cross-view geo-localization models, including CVM-Net (Hu et al., 2018), OriCNN (Liu & Li, 2019), SAFA (Shi et al., 2019), CVFT (Shi et al., 2020b), DSM (Shi et al., 2020a), L2LTR (Yang et al., 2021), TransGeo (Zhu et al., 2022), and GeoDTR (Zhang et al., 2022), on the fine-grained robustness evaluation benchmarks, CVUSA-C, and CVACT_val-C, specifically focusing on R@1 performance. Within this section, we show the experimental results for R@5, R@10, and R@1% on CVUSA-C in Tables 1, 2, and 3, and for CVACT_val-C in Tables 4, 5, and 6.

Table 1: The experimental results of 8 cross-view geo-localization methods on the CVUSA-C. We report the R@5 performance of each method under different corruption (obtained by averaging the 5 corruption severities), as well as the average performance R@5_{cor} under all corruption types.

Method	Clean	CVUSA-C																
		Weather					Blur				Noise				Digital			R@5_{cor}
		Snow	Frost	Fog	Bright	Spatter	Defocus	Glass	Motion	Zoom	Gaussian	Shot	Impulse	Speckle	Contrast	Pixel	JPEG	
CVM-Net	49.98	2.95	22.27	22.70	32.74	15.95	3.67	14.21	4.98	1.09	5.42	3.71	4.11	6.76	12.96	18.04	16.68	11.77
OriCNN	66.82	19.21	17.58	20.81	45.67	46.03	44.66	52.91	42.86	27.27	37.68	32.03	42.20	32.09	19.28	55.81	54.34	36.90
SAFA	96.93	31.52	77.78	83.95	93.51	63.79	69.62	93.30	72.39	23.17	45.38	40.24	43.34	51.04	45.32	96.30	92.27	63.93
CVFT	84.69	17.90	53.79	73.88	72.64	46.90	45.79	71.45	58.56	20.59	37.75	33.65	36.49	44.88	61.26	81.76	71.38	51.79
DSM	97.5	34.91	79.34	93.76	92.42	70.51	78.32	94.14	80.04	40.12	61.01	58.10	60.72	74.80	83.33	96.76	93.62	74.49
L2LTR	98.27	83.91	94.19	97.96	97.51	88.50	96.19	98.03	96.89	63.58	91.72	91.69	92.81	95.46	94.51	98.24	97.14	92.40
TransGeo	98.36	44.52	84.42	85.82	94.74	78.90	92.58	97.65	95.27	62.39	84.58	82.85	86.32	93.87	50.68	98.14	96.72	83.09
GeoDTR	98.86	62.54	93.97	98.25	98.51	86.32	93.58	98.39	90.59	48.74	81.67	78.56	82.15	91.03	86.22	98.73	96.78	86.63

Table 2: The experimental results of 8 cross-view geo-localization methods on the CVUSA-C. We report the R@10 performance of each method under different corruption (obtained by averaging the 5 corruption severities), as well as the average performance R@10_{cor} under all corruption types.

Method	Clean	CVUSA-C																
		Weather					Blur				Noise				Digital			R@10_{cor}
		Snow	Frost	Fog	Bright	Spatter	Defocus	Glass	Motion	Zoom	Gaussian	Shot	Impulse	Speckle	Contrast	Pixel	JPEG	
CVM-Net	63.18	4.85	31.64	32.42	43.35	22.38	5.96	20.88	7.89	1.93	8.16	5.80	6.43	10.18	19.16	24.74	23.32	16.82
OriCNN	76.36	26.88	25.12	29.65	57.78	57.93	56.72	64.92	54.91	38.01	48.17	41.91	53.90	42.08	26.00	67.83	66.17	47.37
SAFA	98.14	37.56	83.25	88.53	96.13	69.08	76.25	96.00	78.49	29.44	50.17	44.85	48.61	56.54	51.37	97.82	94.72	68.68
CVFT	90.49	23.81	63.64	82.56	81.02	55.07	55.53	80.41	67.71	28.16	44.88	40.25	43.84	53.10	69.99	88.34	79.05	59.83
DSM	98.54	39.64	83.69	96.01	94.77	74.66	83.06	96.12	84.28	47.21	64.78	62.13	64.62	79.26	86.42	97.87	95.40	78.12
L2LTR	98.99	88.10	95.98	98.70	98.48	91.96	97.60	98.81	98.18	71.26	94.00	94.20	94.92	97.18	96.23	98.96	98.09	94.54
TransGeo	99.04	51.27	88.48	89.41	96.64	83.47	95.09	98.60	97.03	70.56	87.87	86.65	89.52	95.95	52.67	98.85	97.91	86.25
GeoDTR	99.34	69.12	95.71	99.01	99.11	89.75	95.69	99.08	93.70	57.75	85.55	83.00	86.17	93.85	89.27	99.25	97.83	89.62

Table 3: The experimental results of 8 cross-view geo-localization methods on the CVUSA-C. We report the R@1% performance of each method under different corruption (obtained by averaging the 5 corruption severities), as well as the average performance R@$1\%_{\text{cor}}$ under all corruption types.

Method	Clean	CVUSA-C																
		Weather					Blur				Noise				Digital			R@$1\%_{\text{cor}}$
		Snow	Frost	Fog	Bright	Spatter	Defocus	Glass	Motion	Zoom	Gaussian	Shot	Impulse	Speckle	Contrast	Pixel	JPEG	
CVM-Net	93.62	18.36	69.00	71.00	76.83	49.47	21.20	53.95	26.45	10.36	22.79	17.99	20.31	27.49	43.39	51.40	51.86	39.49
OriCNN	96.12	60.91	59.41	67.44	90.21	89.81	89.76	93.66	88.59	78.51	80.93	75.56	87.19	76.90	52.56	94.54	93.85	79.99
SAFA	99.64	61.11	94.28	97.84	99.41	84.83	91.90	99.44	92.84	56.52	65.89	60.95	65.44	75.81	73.17	99.62	98.54	82.35
CVFT	99.02	49.46	87.21	97.43	96.47	78.46	82.31	96.68	89.34	60.85	66.00	60.29	65.93	77.25	89.67	98.52	94.10	80.62
DSM	99.67	56.63	92.40	98.95	98.66	85.74	93.40	99.10	93.49	69.10	75.48	74.05	75.93	90.06	93.34	99.51	98.23	87.13
L2LTR	99.67	95.76	98.73	99.65	99.56	97.66	99.46	99.62	99.55	89.04	97.92	98.30	98.25	99.31	98.85	99.65	99.45	98.17
TransGeo	99.77	73.06	95.96	96.17	99.37	93.91	98.91	99.73	99.41	89.25	95.28	95.03	96.19	99.06	57.25	99.73	99.54	92.99
GeoDTR	99.86	84.17	98.61	99.77	99.79	96.00	98.94	99.79	98.46	80.88	93.41	92.59	94.13	98.41	95.37	99.81	99.45	95.60

Table 4: The experimental results of 7 cross-view geo-localization methods on the CVACT_val-C. We report the R@5 performance of each method under different corruption (obtained by averaging the 5 corruption severities), as well as the average performance R@5_{cor} under all corruption types.

Method	Clean	CVACT_val-C																
		Weather					Blur				Noise				Digital			R@5_{cor}
		Snow	Frost	Fog	Bright	Spatter	Defocus	Glass	Motion	Zoom	Gaussian	Shot	Impulse	Speckle	Contrast	Pixel	JPEG	
OriCNN	68.28	28.61	14.31	9.87	50.20	62.47	53.21	61.89	59.51	45.74	55.43	53.19	60.79	54.42	10.34	65.90	64.61	46.91
SAFA	92.80	31.07	46.81	55.18	83.61	60.30	56.74	88.40	67.85	11.02	63.35	58.46	64.06	68.77	28.28	90.69	89.51	60.26
CVFT	81.33	26.95	37.35	64.33	69.51	56.58	50.63	76.11	57.43	10.99	54.17	47.89	55.43	56.75	47.71	79.44	79.03	54.39
DSM	92.44	45.43	68.34	85.01	83.58	66.03	72.42	91.15	81.69	27.33	71.75	66.41	71.64	77.52	65.17	92.04	91.22	72.30
L2LTR	94.59	87.40	91.61	94.14	92.92	89.01	94.30	94.65	94.24	70.77	93.65	93.23	93.96	93.63	91.73	94.66	94.14	91.50
TransGeo	94.14	65.34	76.39	49.55	87.31	82.27	92.49	94.06	92.74	57.06	92.68	92.35	93.17	93.47	30.30	94.10	93.63	80.43
GeoDTR	95.44	66.69	87.28	93.88	94.74	77.57	92.10	95.11	87.62	18.74	89.12	88.51	90.42	92.80	66.35	95.34	95.11	83.21

Table 5: The experimental results of 7 cross-view geo-localization methods on the CVACT_val-C. We report the R@10 performance of each method under different corruption (obtained by averaging the 5 corruption severities), as well as the average performance R@10_{cor} under all corruption types.

Method	Clean	CVACT_val-C																
		Weather					Blur				Noise				Digital			R@10_{cor}
		Snow	Frost	Fog	Bright	Spatter	Defocus	Glass	Motion	Zoom	Gaussian	Shot	Impulse	Speckle	Contrast	Pixel	JPEG	
OriCNN	75.48	36.54	19.73	13.86	58.98	70.36	61.95	70.14	68.21	55.61	63.91	62.01	69.03	63.03	13.66	73.76	72.44	54.58
SAFA	94.84	36.34	53.35	64.14	87.82	65.67	63.29	91.89	74.26	15.53	68.50	63.62	69.23	74.44	34.26	93.52	92.57	65.53
CVFT	86.52	32.92	44.14	72.05	76.64	63.68	58.73	82.55	65.49	15.72	60.92	54.46	62.47	64.57	53.57	85.09	84.68	61.11
DSM	93.99	51.08	73.68	88.25	87.07	70.86	76.95	92.99	85.29	33.17	75.62	70.84	75.77	81.65	68.87	93.74	93.08	76.18
L2LTR	95.96	90.89	94.07	95.70	95.01	92.27	95.82	96.08	95.83	77.48	95.41	95.11	95.66	95.40	93.97	96.10	95.71	93.78
TransGeo	95.78	71.46	81.56	56.38	90.66	86.57	94.50	95.59	94.60	64.94	94.83	94.62	95.11	95.31	33.05	95.68	95.31	83.76
GeoDTR	96.72	72.73	90.73	95.65	96.30	82.15	94.39	96.46	90.60	25.05	91.91	91.60	92.96	94.99	69.32	96.64	96.46	86.12

Based on the overall experimental results we obtained, it becomes evident that the performance of models in terms of R@5, R@10, and R@1% aligns with that of R@1. As a result, the relevant analysis regarding R@1 in the main paper remains applicable to R@5, R@10, and R@1%, and is thus not reiterated here.

Table 6: The experimental results of 7 cross-view geo-localization methods on the CVACT_val-C. We report the R@1% performance of each method under different corruption (obtained by averaging the 5 corruption severities), as well as the average performance R@1%$_{\text{cor}}$ under all corruption types.

Method	Clean	CVACT_val-C																
		Weather					Blur				Noise				Digital			R@1%$_{\text{cor}}$
		Snow	Frost	Fog	Bright	Spatter	Defocus	Glass	Motion	Zoom	Gaussian	Shot	Impulse	Speckle	Contrast	Pixel	JPEG	
OriCNN	92.04	66.82	45.45	35.18	84.49	89.97	85.74	89.64	89.01	82.74	85.87	85.27	88.81	86.31	29.14	91.30	90.66	76.65
SAFA	98.17	58.34	76.87	88.53	96.24	83.22	82.66	97.64	90.48	38.27	83.51	80.64	84.90	90.05	55.65	97.94	97.63	81.41
CVFT	95.93	55.81	66.90	89.97	91.96	83.15	80.98	94.40	85.93	40.07	78.94	73.52	81.26	84.72	71.20	95.34	95.14	79.33
DSM	97.32	69.81	87.52	95.25	94.64	84.28	88.58	96.97	93.62	55.66	86.83	83.85	86.94	91.56	79.74	97.22	97.08	86.85
L2LTR	98.37	97.11	97.88	98.31	98.14	97.56	98.30	98.36	98.30	91.48	98.25	98.20	98.27	98.25	97.85	98.36	98.26	97.68
TransGeo	98.37	88.35	92.88	76.97	96.76	95.54	97.90	98.28	97.96	85.92	98.16	98.10	98.21	98.21	41.88	98.33	98.24	91.35
GeoDTR	98.77	87.57	96.88	98.43	98.64	93.15	97.88	98.59	96.34	51.77	97.04	97.24	97.51	98.34	76.79	98.65	98.62	92.71

C Comprehensive Robustness Evaluation Benchmark Experiment Results

The performances of different models on the CVUSA-C-ALL, CVACT_val-C-ALL, and CVACT_test-C-ALL datasets are shown in Table 7 and 8. To facilitate our analysis, we also report their performance on the original validation set. From the experimental results, it is evident that, when evaluate using a comprehensive robustness evaluation benchmark, the performance degradation is closely positively correlated with the original performance, except for L2LTR. The L2LTR exhibits the highest level of robustness, albeit at the expense of increased computational cost and a greater number of trainable parameters. This once again reminds us that, in the pursuit of model lightweighting, we must consider whether there are other associated trade-offs, as there is indeed - no free lunch.

Table 7: The experimental results of 8 cross-view geo-localization methods on the comprehensive robustness evaluation benchmark CVUSA-C-ALL.

Method	CVUSA-C-ALL							
	Clean				Corruption			
	R@1	R@5	R@10	R@1%	R@1$_{\text{cor}}$	R@5$_{\text{cor}}$	R@10$_{\text{cor}}$	R@1%$_{\text{cor}}$
CVM	22.47	49.98	63.18	93.62	6.09	16.05	23.14	52.51
OriCNN	40.79	66.82	76.36	96.12	9.38	22.26	30.04	58.99
SAFA	89.84	96.93	98.14	99.64	63.68	78.08	82.82	93.91
CVFT	61.43	84.69	90.49	99.02	41.05	64.01	72.64	91.37
DSM	91.96	97.50	98.54	99.67	75.27	86.26	89.42	95.07
L2LTR	94.05	98.27	98.99	99.67	87.93	95.45	97.01	99.01
TransGeo	94.08	98.36	99.04	99.77	82.72	91.95	94.03	97.92
GeoDTR	95.43	98.86	99.34	99.86	84.64	93.29	95.01	98.24

Table 8: The experimental results of 7 cross-view geo-localization methods on the comprehensive robustness evaluation benchmarks CVACT_val-C-ALL and CVACT_test-C-ALL.

Method	CVACT_val-C-ALL								CVACT_test-C-ALL							
	Clean				Corruption				Clean				Corruption			
	R@1	R@5	R@10	R@1%	R@1$_{\text{cor}}$	R@5$_{\text{cor}}$	R@10$_{\text{cor}}$	R@1%$_{\text{cor}}$	R@1	R@5	R@10	R@1%	R@1$_{\text{cor}}$	R@5$_{\text{cor}}$	R@10$_{\text{cor}}$	R@1%$_{\text{cor}}$
OriCNN	46.96	68.28	75.48	92.01	15.31	28.31	35.21	58.39	19.21	35.97	43.30	60.69	3.69	8.33	11.04	43.93
SAFA	81.03	92.80	94.84	98.17	56.72	73.60	78.59	91.32	55.50	79.94	85.08	94.49	31.18	52.06	58.60	90.41
CVFT	61.05	81.33	86.52	95.93	45.69	66.45	72.97	88.38	26.12	45.33	53.80	71.69	22.82	43.48	51.07	88.99
DSM	82.49	92.44	93.99	97.32	70.04	82.81	85.86	93.51	59.30	82.27	86.44	97.51	47.13	68.41	73.52	93.18
L2LTR	84.89	94.59	95.96	98.37	82.13	93.34	94.93	98.10	60.72	85.85	89.88	96.12	57.20	82.59	87.23	98.09
TransGeo	84.95	94.14	95.78	98.37	74.04	86.19	89.10	94.98	63.35	86.43	90.10	98.47	52.18	74.35	78.99	95.03
GeoDTR	86.21	95.44	96.72	98.77	77.40	88.95	91.28	95.91	64.52	88.59	91.96	98.74	52.87	78.84	83.17	95.84

D Histogram Equalization for Robustness Enhancement

We further examined the impact of employing histogram equalization on the robustness of cross-view geo-localization models. In our study, we employ Contrast Limited Adaptive Histogram Equalization (CLAHE) (Pizer et al., 1987) to enhance the robustness of existing methods. On the CVUSA dataset, we evaluated the performance of 3 classic cross-view geo-localization models using the training strategy outlined in Section 3.2 of the main paper, as illustrated in the Figure 2.

From the experimental results, it becomes evident that training solely on data subjected to histogram equalization does not significantly enhance the robustness of models. Conversely, combining histogram equalization with clean data in equal proportions can to some extent improve the robustness of models, although the extent of improvement is notably inferior to stylization-based approaches. Furthermore, it is noteworthy that the same training data yields varying effects on different models. This underscores the importance of focusing on model robustness and highlights the challenges in achieving universal enhancements across diverse model architectures.

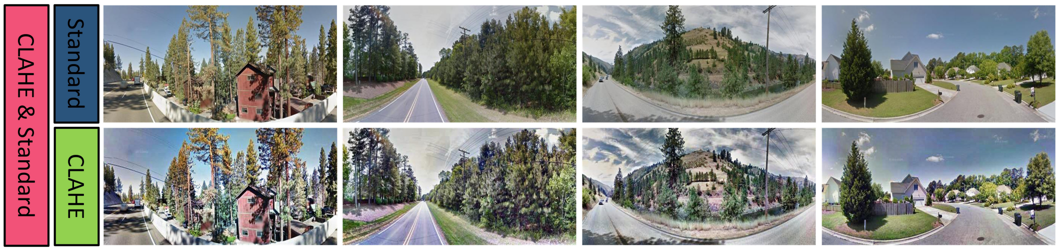
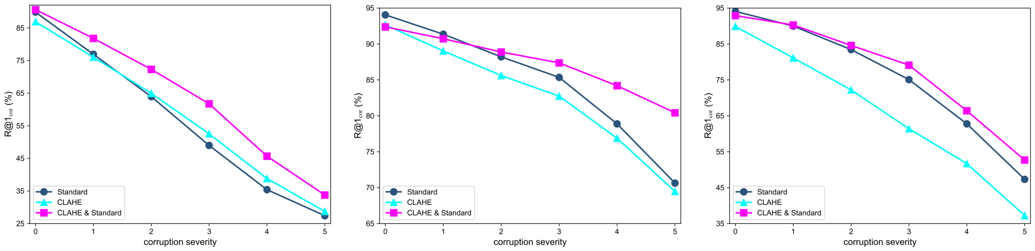


Figure 2: Visualization of CLAHE applied to CVUSA dataset. The illustration depicts standard images (top row), and histogram-equalized images (bottom row). The rectangular sections on the left represent different training strategies.



(a) SAFA

(b) L2LTR

(c) TransGeo

Figure 3: Training on histogram-equalized images enhance the robustness of SAFA, L2LTR, and TransGeo on the CVUSA dataset, with each severity level representing the average across all 16 corruption types. Severity = 0 corresponds to clean images for testing . The `Standard` denotes the original, unaltered training data, while `CLAHE` denotes training exclusively on images subjected to histogram equalization. `CLAHE & Standard` denotes histogram equalization and original images are equally interleaved during the training process. Notably, the 3 different training strategies require identical training complexity, and the experimental configurations and model structures remain consistent throughout.

E Limitation and Discussion

Limitation. This paper primarily discusses the robustness exhibited by classic cross-view geo-localization models when ground query images are subjected to various forms of corruption. To conduct our research, we applied existing image corruption algorithms to publicly available CVUSA (Workman et al., 2015) and CVACT (Liu & Li, 2019) datasets, forming the foundation for our robustness evaluation benchmarks. Nevertheless, it is important to acknowledge the limitations imposed by whether these image corruption algorithms faithfully replicate real-world scenarios, which is evidently challenging. Consequently, in the future, we aspire to develop more advanced image corruption algorithms to generate corruption scenarios that better align with real-world conditions. Additionally, we explored two robustness enhancement techniques, namely stylization and histogram equalization, aimed at enhancing the robustness of existing models. However, it is worth noting that the principal constraint in utilizing these techniques is the necessity for preprocessing and retraining models using the training data.

Discussion. While our primary focus lies in the evaluation of the robustness of cross-view geo-localization models when subjected to input image corruption, we propose that the evaluation benchmarks we introduce can have broader applicability. These benchmarks not only to cross-view geo-localization tasks but can also be leveraged for cross-view camera pose estimation (Shi & Li, 2022; Xia et al., 2022; Lentsch et al., 2023) and cross-view image synthesis (Regmi & Borji, 2018; Tang et al., 2020; Toker et al., 2021; Shi et al., 2022).

REFERENCES

Sixing Hu, Mengdan Feng, Rang MH Nguyen, and Gim Hee Lee. Cvm-net: Cross-view matching network for image-based ground-to-aerial geo-localization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7258–7267, 2018.

Ted Lentsch, Zimin Xia, Holger Caesar, and Julian FP Kooij. Slicematch: Geometry-guided aggregation for cross-view pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17225–17234, 2023.

Liu Liu and Hongdong Li. Lending orientation to neural networks for cross-view geo-localization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5624–5633, 2019.

Stephen M Pizer, E Philip Amburn, John D Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B Zimmerman, and Karel Zuiderveld. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3):355–368, 1987.

Krishna Regmi and Ali Borji. Cross-view image synthesis using conditional gans. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 3501–3510, 2018.

Yujiao Shi and Hongdong Li. Beyond cross-view image retrieval: Highly accurate vehicle localization using satellite image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17010–17020, 2022.

Yujiao Shi, Liu Liu, Xin Yu, and Hongdong Li. Spatial-aware feature aggregation for image based cross-view geo-localization. *Advances in Neural Information Processing Systems*, 32, 2019.

Yujiao Shi, Xin Yu, Dylan Campbell, and Hongdong Li. Where am i looking at? joint location and orientation estimation by cross-view matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4064–4072, 2020a.

Yujiao Shi, Xin Yu, Liu Liu, Tong Zhang, and Hongdong Li. Optimal feature transport for cross-view image geo-localization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 11990–11997, 2020b.

Yujiao Shi, Dylan Campbell, Xin Yu, and Hongdong Li. Geometry-guided street-view panorama synthesis from satellite imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):10009–10022, 2022.

Hao Tang, Dan Xu, Yan Yan, Philip HS Torr, and Nicu Sebe. Local class-specific and global image-level generative adversarial networks for semantic-guided scene generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 7870–7879, 2020.

Aysim Toker, Qunjie Zhou, Maxim Maximov, and Laura Leal-Taixé. Coming down to earth: Satellite-to-street view synthesis for geo-localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6488–6497, 2021.

Scott Workman, Richard Souvenir, and Nathan Jacobs. Wide-area image geolocalization with aerial reference imagery. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3961–3969, 2015.

Zimin Xia, Olaf Booij, Marco Manfredi, and Julian FP Kooij. Visual cross-view metric localization with dense uncertainty estimates. In *European Conference on Computer Vision*, pp. 90–106. Springer, 2022.

Hongji Yang, Xiufan Lu, and Yingying Zhu. Cross-view geo-localization with layer-to-layer transformer. *Advances in Neural Information Processing Systems*, 34:29009–29020, 2021.

Xiaohan Zhang, Xingyu Li, Waqas Sultani, Yi Zhou, and Safwan Wshah. Cross-view geo-localization via learning disentangled geometric layout correspondence. *arXiv preprint arXiv:2212.04074*, 2022.

Sijie Zhu, Mubarak Shah, and Chen Chen. Transgeo: Transformer is all you need for cross-view image geo-localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1162–1171, 2022.