SinGeo: Unlock Single Model's Potential for Robust Cross-View Geo-Localization

Supplementary Material

In this supplementary material, we provide the following information for the completeness of our paper.

- Ablations of different rotation transformations (Section A).
- Ablations of different scheduling functions (Section B).
- Ablations of different loss weights (Section C).
- Performance on CVACT test set (Section D).
- Supplementary results on VIGOR (Section E).
- Consistency metrics based on cosine similarity and pearson correlation coefficient (Section F).
- Additional qualitative visualization results (Section G).
- Implementation details (Section H).

A. Ablations of Different Rotation Transformations

Table 1. Performance comparison of different rotation transformations for I_s on the CVUSA dataset.

Transformations	FoV= 360°		FoV= 180°		FoV= 90°	
	R@1	R@1%	R@1	R@1%	R@1	R@1%
$T_s^1(\phi)$	89.2	97.3	77.5	94.2	63.7	96.4
$T_s^2(\phi)$	88.9	97.5	81.1	97.4	60.7	95.8
$T_s^3(p)$	**96.8**	**99.5**	**91.8**	**99.5**	**70.1**	**97.5**

We compare the performance of proposed $T_s^1(\phi)$, $T_s^2(\phi)$, and $T_s^3(p)$ on CVUSA dataset. In Tab. 1, the discrete rotation transformation $T_s^3(p)$ consistently outperforms the continuous variants $T_s^1(\phi)$ and $T_s^2(\phi)$ across all FoVs on the CVUSA dataset. We assume that discrete rotation transformations outperform continuous ones because the former can avoid introducing additional padding boundaries and will not cause information loss. Discrete 90° multiples enable exact pixel permutations, preserving discriminative features.

B. Ablations of Different Scheduling Functions

Table 2. Performance comparison of different scheduling functions on the CVUSA dataset.

Scheduling Function	Avg. R@1	FoV= 360°		FoV= 180°		FoV= 90°	
		R@1	R@1%	R@1	R@1%	R@1	R@1%
$f_1(x)$	**86.2**	96.8	99.5	**91.8**	**99.5**	**70.1**	**97.5**
$f_2(x)(\lambda=3)$	79.9	94.6	99.6	83.7	99.2	61.5	96.0
$f_3(x)(\lambda=3)$	82.8	**97.1**	**99.8**	90.9	99.3	60.4	92.3
$f_2(x)(\lambda=5)$	82.0	95.8	99.5	83.1	98.3	67.0	96.2
$f_3(x)(\lambda=5)$	82.3	96.6	99.8	89.8	99.6	60.5	94.5
Random	76.7	90.3	98.7	83.6	98.8	56.2	93.5

We also considered different scheduling functions. As observed in Tab. 2, a simple linear scheduling function $f_1(x)$ achieves superior overall performance with 86.2 of average R@1 compared to alternative exponential variants $f_2(x)$, $f_3(x)$ and random scheduling. Notably, all scheduling functions that progress from easy to hard outperform random scheduling, as the latter lacks a progressive adaptation mechanism. We hypothesize that the linear schedule maintains a constant gradient in training difficulty between epochs, thereby mitigating abrupt shifts that could lead to model instability, such as insufficient adaptation to escalating challenges. This result validates that a straightforward progressive training paradigm is highly beneficial for model to learn robustness.

C. Ablations of Different Loss Weights

Table 3. Performance comparison of different loss weight configurations on the CVUSA dataset.

γ	$\omega_1\omega_2\omega_3$	Avg. R@1	FoV= 360°		FoV= 180°		FoV= 90°	
			R@1	R@1%	R@1	R@1%	R@1	R@1%
0.5	0.25	**86.1**	**96.8**	99.5	**91.5**	99.4	70.1	97.5
0.5	0.5	84.7	96.2	99.6	90.0	99.4	68.0	**97.7**
0.25	0.5	85.6	96.4	**99.7**	90.7	**99.6**	69.6	97.6
0.25	0.25	85.8	96.7	**99.7**	90.2	99.2	**70.6**	97.6

In Tab. 3, we look into the effect of loss weights γ and $\omega_1, \omega_2, \omega_3$ on performance. Results show that different weight configurations cause minor performance fluctuations across FoV settings. The default setup $\gamma = 0.5$, $\omega = 0.25$ achieves the best average R@1 of 86.1 and outperforms other configurations at FoV= 360° and FoV= 180°.

D. Performance on CVACT Test Set

In the main paper, we reported the results on CVACT validation set. In this section, we also evaluate SinGeo on the CVACT test set under unknown orientation and limited FoV settings as shown in Tab. 4, comparing it with Sample4Geo [2] and two FoV-specialized ConGeo variants [3]. ConGeo[360] and ConGeo[180] are trained specifically at 360° and 180° FoV respectively, showing strong performance at their trained angles but significant degradation at unseen narrow FoVs.

In contrast, SinGeo uses a single model without FoV-specific training and achieves the best average R@1 of 35.8 across all FoVs. It outperforms both ConGeo variants at every FoV except 180° where ConGeo[180] holds a marginal

Table 4. Comparison with single-model performance of selected methods on CVACT Test dataset under unknown orientation and limited FoV settings.

Methods	Avg.	FoV 360°				FoV 180°				FoV 90°				FoV 70°			
	R@1	R@1	R@5	R@10	R@1%	R@1	R@5	R@10	R@1%	R@1	R@5	R@10	R@1%	R@1	R@5	R@10	R@1%
Sample4Geo[2]	2.4	8.0	12.5	14.0	31.0	1.0	2.7	3.8	25.1	0.4	1.3	2.2	24.1	0.2	0.7	1.1	16.9
ConGeo[360][3]	25.0	64.3	81.5	84.7	95.6	28.6	47.8	54.9	88.4	5.2	13.0	17.9	67.8	1.8	5.6	8.4	53.3
ConGeo[180][3]	25.4	37.4	55.3	61.1	89.6	45.6	**67.1**	**73.0**	**94.3**	13.7	28.3	35.6	81.7	5.0	13.1	18.0	68.6
SinGeo	**35.8**	**66.7**	**84.2**	**86.8**	**96.0**	**46.0**	66.8	72.6	93.2	**19.6**	**38.8**	**47.1**	**89.3**	**11.0**	**25.5**	**33.3**	**84.2**

Table 5. Comparison of methods on cross-area and same-area splits of VIGOR under unknown orientation and limited FoV settings. "–" denotes metrics not provided in the original paper.

Set	Methods	FoV 360°				FoV 180°				FoV 90°				FoV 70°			
		R@1	R@5	R@10	R@1%	R@1	R@5	R@10	R@1%	R@1	R@5	R@10	R@1%	R@1	R@5	R@10	R@1%
Cross-Area	VIGOR [5]	1.4	-	-	44.6	-	-	-	-	-	-	-	-	-	-	-	-
	TransGeo [6]	5.5	-	-	66.9	-	-	-	-	-	-	-	-	-	-	-	-
	Sample4Geo [2]	9.0	-	-	43.7	-	-	-	-	0.5	-	-	21.6	-	-	-	-
	ConGeo [3]	16.2	-	-	72.9	-	-	-	-	3.9	-	-	54.3	-	-	-	-
	SinGeo	**24.7**	**43.5**	**51.4**	**85.7**	**11.0**	**24.0**	**30.7**	**72.6**	**4.9**	**12.9**	**18.2**	**63.7**	**3.3**	**9.2**	**13.2**	**56.0**
Same-Area	VIGOR [5]	19.1	-	-	95.1	-	-	-	-	-	-	-	-	-	-	-	-
	TransGeo [6]	47.7	-	-	**99.3**	-	-	-	-	-	-	-	-	-	-	-	-
	Sample4Geo [2]	14.2	-	-	54.9	-	-	-	-	1.1	-	-	30.6	-	-	-	-
	ConGeo [3]	61.9	-	-	98.4	-	-	-	-	8.5	-	-	68.7	-	-	-	-
	SinGeo	**62.9**	**88.5**	**91.5**	98.0	**43.0**	**69.0**	**75.2**	**94.0**	**24.0**	**46.3**	**54.6**	**91.5**	**16.1**	**34.2**	**42.3**	**85.5**

edge in R@5 and R@10. Notably, SinGeo maintains best performance even at extreme narrow FoVs.

E. Supplementary Results on VIGOR

In the main paper, we reported results of VIGOR dataset at 360° and 90° FoV. We further supplement results at 180° and 70° FoV as shown in Tab. 5. SinGeo achieves excellent performance on VIGOR, a non-center-aligned dataset, outperforming other compared methods across both Cross-Area and Same-Area splits.

F. The Consistency Metrics Based on Cosine Similiarity and PCC

Table 6. Quantitative evaluation of orientation-consistency(OC) and FoV-consistency(FC) on CVUSA dataset based on cosine similarity and pearson correlation coefficient).

Metric	Method	OC_{grd}	OC_{sat}	FC_{grd}	FC_{sat}
Cosine Similarity	Sample4Geo [2]	0.37	0.40	0.25	0.32
	ConGeo [3]	0.15	0.61	0.22	0.46
	SinGeo	**0.77**	**0.91**	**0.54**	**0.63**
PCC	Sample4Geo [2]	0.13	0.18	0.04	0.11
	ConGeo [3]	0.10	0.46	0.17	0.21
	SinGeo	**0.65**	**0.89**	**0.31**	**0.54**

In the main paper, we adopt the Structural Similarity Index (SSIM) [4], a classic 2D similarity measurement method, to calculate the consistency between activation heatmaps. To further validate our conclusions, we flatten the 2D heatmaps into 1D tensors and utilize cosine similarity and Pearson Correlation Coefficient (PCC) [1] to compute the consistency metrics between these tensors. As shown in Tab. 6, SinGeo still achieves the highest scores across all orientation-consistency (OC) and FoV-consistency (FC) metrics when using cosine similarity and PCC. These results confirm that SinGeo preserves stable activation under orientation shifts and FoV variations.

G. Additional Qualitative Visualization Results

We further provide supplementary qualitative visualizations in Fig. 1, Fig. 2 and Fig. 3. When the aligned panorama is applied with a random translation (the second row), SinGeo maintains consistent responses on satellite images. As the FoV further narrows (last three rows), the model consistently responds to key semantic regions of satellite images that correspond to the ground images (the yellow dashed circles). Compared with SinGeo, the results of ConGeo and Sample4Geo are less satisfactory. Even when retrieval is correct, their response regions show certain shifts or errors. As shown in the first and second rows of Fig. 1, after panorama translation, ConGeo's responses on satellite images become completely different despite identical information in panoramas. As shown in the third row of Fig. 1, when FoV=180○,

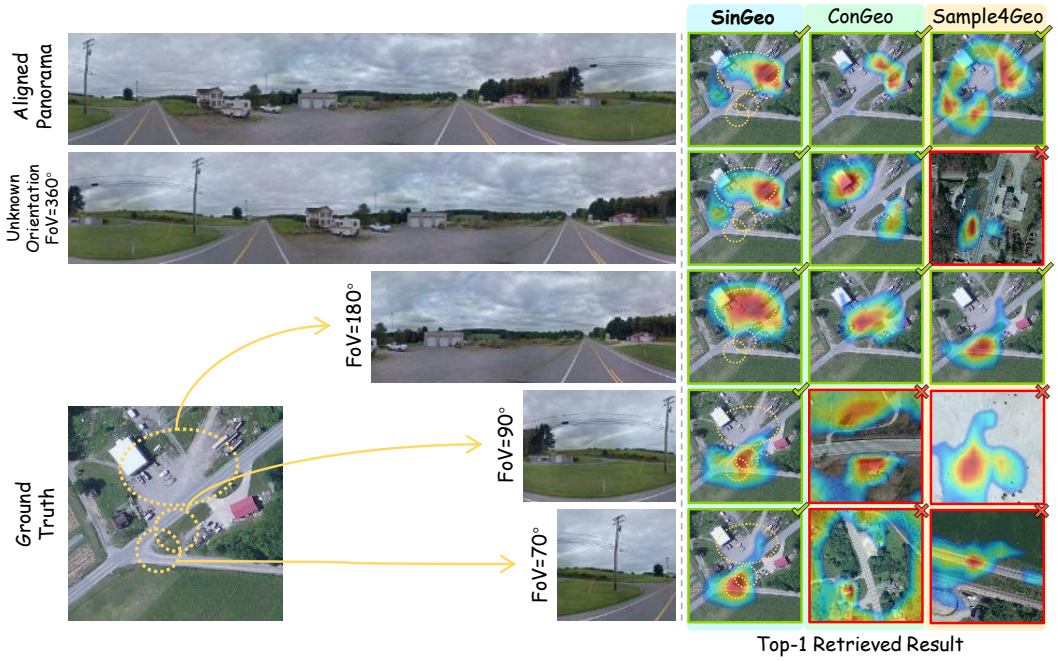

Figure 1. Qualitative visualization result on CVUSA. Yellow circles denote the regions on the satellite image that correspond to the limited-FoV ground images.

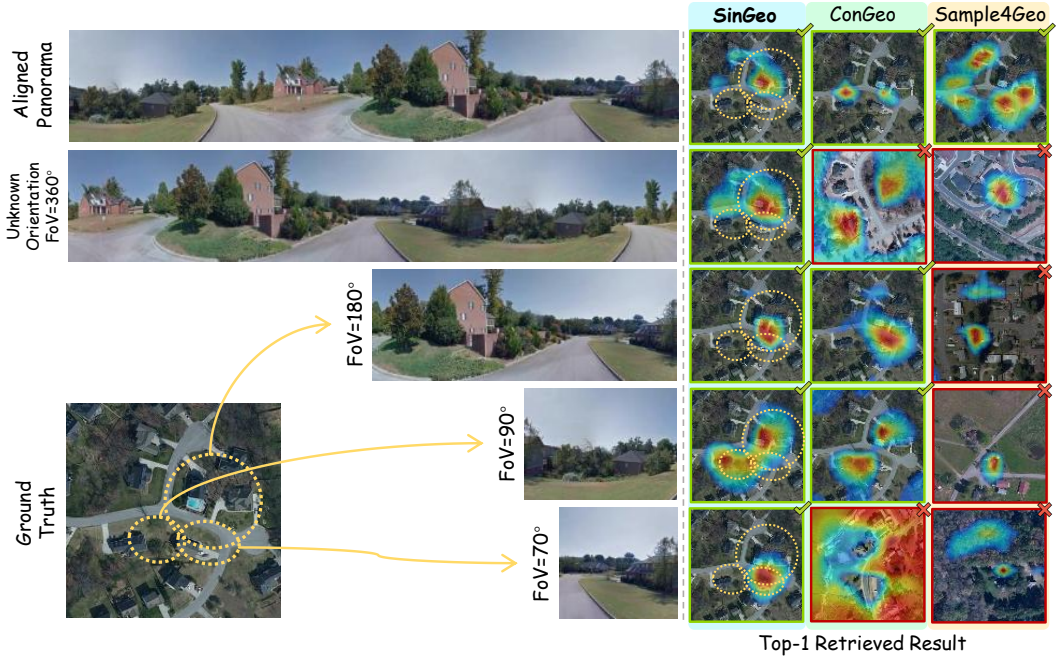

Figure 2. Qualitative visualization result on CVUSA. Yellow circles denote the regions on the satellite image that correspond to the limited-FoV ground images.

Sample4Geo's response region does not match the corresponding area of the ground query image even with correct retrieval.

H. Implementation Details

We follow the data preprocessing method of our baseline Sample4Geo [2]. Our code of this paper uses Python 3.8.20, PyTorch 2.1.1, timm 0.9.0, and scikit-learn 1.3.2. All train-

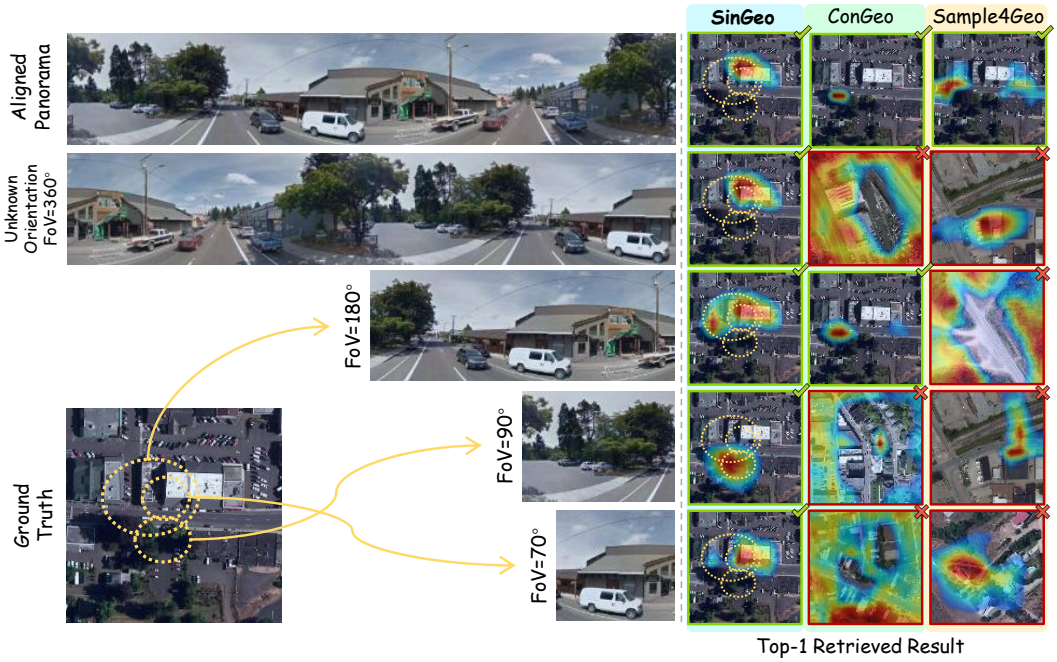

Figure 3. Qualitative visualization result on CVUSA. Yellow circles denote the regions on the satellite image that correspond to the limited-FoV ground images.

ing and testing are conducted on a single NVIDIA A100 GPU.

References

[1] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. Pearson correlation coefficient. In *Noise reduction in speech processing*, pages 1–4. Springer, 2009. 2

[2] Fabian Deuser, Konrad Habel, and Norbert Oswald. Sample4geo: Hard negative sampling for cross-view geo-localisation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16847–16856, 2023. 1, 2, 3

[3] Li Mi, Chang Xu, Javiera Castillo-Navarro, Syrielle Montariol, Wen Yang, Antoine Bosselut, and Devis Tuia. Congeo: Robust cross-view geo-localization across ground view variations. In *European Conference on Computer Vision*, 2024. 1, 2

[4] Zhou Wang. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 2

[5] Sijie Zhu, Taojiannan Yang, and Chen Chen. Vigor: Cross-view image geo-localization beyond one-to-one retrieval. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3640–3649, 2021. 2

[6] 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*, pages 1162–1171, 2022. 2