Supplementary Material: Temporal Resilience in Geo-Localization: Adapting to the Continuous Evolution of Urban and Rural Environments

1. Visualization of the CVUSA Crop

To determine the necessary cropping for our analysis, we evaluated Street View samples from both the CVUSA 1.0 and 2.0 datasets, focusing on images taken before 2015, as shown in Figure 1. This selection highlights a critical aspect of Street View imagery: the infrequent updates. This infrequent update rate presents additional challenges, as the dataset may contain outdated images. When training models on such data, the absence of recent temporal changes could affect the accuracy and relevance of the analysis, underscoring the need for careful consideration of data currency in geospatial studies.

Using these reference images, we determined the necessary cropping parameters to preserve the aspect ratio. This cropping strategy, applied only to the top and bottom of the images, is demonstrated with examples from different years in Figure 2.

2. Visualization of the Orientation

Figure 3 shows an example of aligned Street View imagery. Initially, Street View images obtained from the Google Street View API are not oriented by default. The API provides metadata information about the orientation of each panorama, which we use to align the images so that geographic north is centered in both the Street View and the top corner of the corresponding satellite image. By rolling the panoramas according to this orientation metadata, we restore the visual consistency seen in the CVUSA 1.0 dataset. In addition to providing visual consistency, this alignment process significantly improves the training task, making it easier for the model to understand and learn from the data. The example presented shows the temporal evolution of a newly constructed southwest facing building and illustrates the practical benefits of this alignment approach in capturing dynamic changes within the observed environment.

3. Evaluation on Uncropped CVTemporal

In our study, we compared the pre-trained models CDE [2], TransGeo [3], SAIG-D [4], and Sample4Geo [1] on the CVTemporal dataset. The performance of all models remained consistent when only satellite imagery was introduced, as this did not involve varying aspect ratios. However, the introduction of uncropped Street View imagery, which introduced different aspect ratios, resulted in significant performance drops, especially for CDE and SAIG-D. Both SAIG-D and TransGeo were particularly affected due to their reliance on hard position encodings, which required image resizing and resulted in skewed inputs that degraded performance. CDE faced challenges because its GAN was trained on images without clipping, making a fair comparison with Sample4Geo difficult. Sample4Geo, being a CNN-based model, showed some resistance to aspect ratio variations, as shown in Table 1, but still struggled under the new conditions, although less so than the other models. This highlights the importance of architectural choices in handling aspect ratio changes, where the CNN-based approach of Sample4Geo appeared more robust compared to the Transformer and MLP-Mixer designs of the others.

Dataset	R@1	R@5	R@10	R@1%	Δ in % for R@1
CVUSA 1.0					
Sample4Geo [1]	**98.68**	**99.68**	**99.78**	**99.87**	-
New Sat					
Sample4Geo [1]	**95.21**	**98.51**	**98.96**	**99.61**	-3.64 %
New Pano					
Sample4Geo [1]	**72.30**	**86.25**	**89.42**	**95.42**	-26.73%
New Sat & Pano					
Sample4Geo [1]	**69.55**	**83.65**	**87.47**	**94.50**	-29.51%

Table 1. **Comparison between pre-trained state-of-the-art approaches on the CVTemporal dataset.**

4. Comparison of Our Selection Strategies

In Table 2 we extend our experiments to evaluate different parts of the data. As expected, training performance improves as more data is used. What stands out, however, is that the clustering approach consistently helps select better training samples across all data parts. The only exception is the 1 % setting, where about 350 images are used, leading to rapid overfitting. In this case, false-prediction sampling involves random selection, which can lead to mixed results. A major limitation of the false prediction selection method is that, given the 89.13 % performance of our CVUSA 1.0

(a) CVUSA 1.0 Street View image.

(b) CVTemporal Street View image.

Figure 1. **Example of Street View images from CVUSA 1.0 and 2.0.** Note that both are taken before 2015 and thus can be used to determine the crop size present in CVUSA 1.0. This also highlights the problem of the low update frequency of Street View images compared to satellite images.

pre-trained model, only 10.87 % of the data are actually false predictions. Consequently, in the 20 % and 30 % settings, the remaining samples are randomly selected, which can impact performance.

References

[1] 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 (ICCV)*, pages 16847–16856, October 2023. 1

[2] 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*, pages 6488–6497, 2021. 1

[3] 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. 1

[4] Yingying Zhu, Hongji Yang, Yuxin Lu, and Qiang Huang. Simple, effective and general: A new backbone for cross-view image geo-localization. *arXiv preprint arXiv:2302.01572*, 2023. 1

(a) CVUSA 1.0 Street View image.

(b) CVTemporal Street View image.

(c) Cropped CVTemporal Street View image.

Figure 2. **Cropped example on a newer Street view image based on the before determined ratios.**

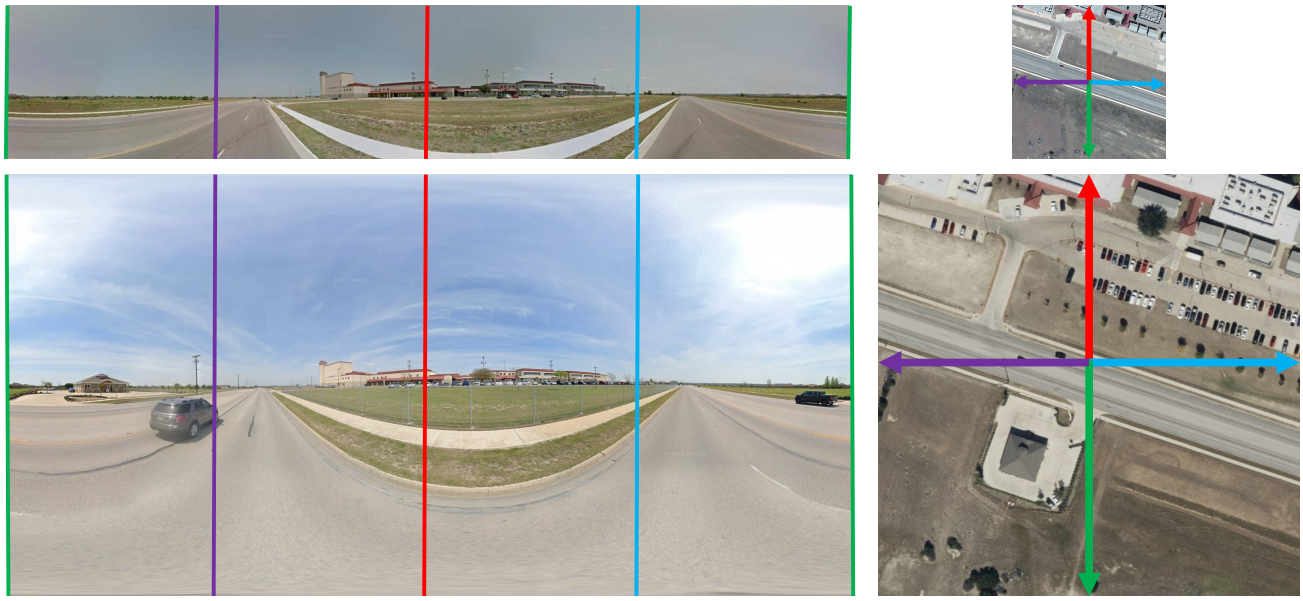

Figure 3. Example of the orientation alignment in CVUSA 1.0 and CVTemporal. Red represents north, blue signifies east, green indicates south, and purple denotes west.

Model Type	R@1	R@5	Trained on		Subset
			CVUSA 1.0	CVTemporal	
Baseline	89.13	94.87	X	-	100%
False preds	90.39	95.67	-	X	1%
Magnitude	90.09	95.48	-	X	1%
Clustering	90.09	95.48	-	X	1%
False preds	90.38	97.29	-	X	5%
Magnitude	89.64	96.77	-	X	5%
Clustering	91.35	96.94	-	X	5%
False preds	91.40	97.97	-	X	10%
Magnitude	91.75	97.55	-	X	10%
Clustering	92.10	97.53	-	X	10%
Clustering + ReRank	**94.65**	**98.51**	-	X	10%
False preds	92.81	98.52	-	X	20%
Magnitude	92.92	97.95	-	X	20%
Clustering	93.11	98.06	-	X	20%
Clustering + ReRank	**94.76**	**98.66**	-	X	20%
False preds	93.70	98.70	-	X	30%
Magnitude	93.71	98.44	-	X	30%
Clustering	93.95	98.44	-	X	30%
Clustering + ReRank	**95.60**	**99.00**	-	X	30%
Full	95.00	98.44	-	X	100%
Full	95.48	98.83	X	X	100%
Full + ReRank	**97.21**	**99.24**	X	X	100%

Table 2. **Comparison of Selection Strategies**