



# Unleashing Unlabeled Data: A Paradigm for Cross-View Geo-Localization

Guopeng Li, Ming Qian, Gui-Song Xia

{guopengli, mingqian, guisong.xia}@whu.edu.cn





#### Motivation

#### Large-Area Cross-View Geo-Localization (CVGL):

Determining the localization of ground images by retrieving the most similar GPS-tagged satellite images.

### Existing CVGL trains with paired ground-satellite images:

- accurately located ground images requires expensive devices.

  matching ground-satellite images brings extra human costs.
- ground and satellite images without GPS cannot be leveraged.
- requiring re-annotation for new or changed scenes.

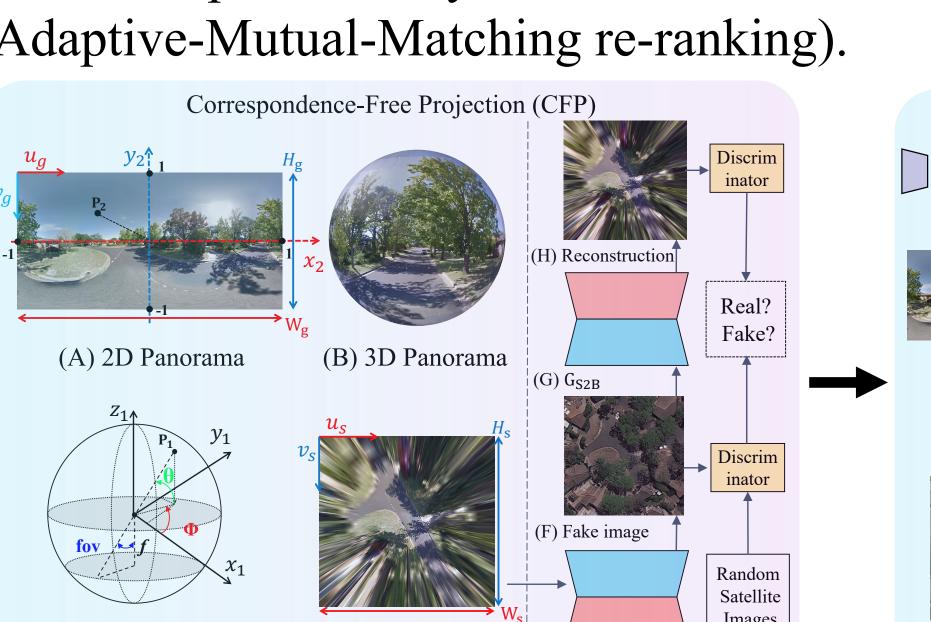
### Contribution

- First work to unleash the unlabeled data in CVGL.
- The first framework to start a unsupervised training without any labeled data and to start a semi-supervised training with few labeled data in CVGL.
- Introducing a unsupervised projection to project the ground images to the satellite view and a fast re-ranking mechanism to refine the noisy labels.
- Competitive performance as the previous supervised SOTA.

## Method

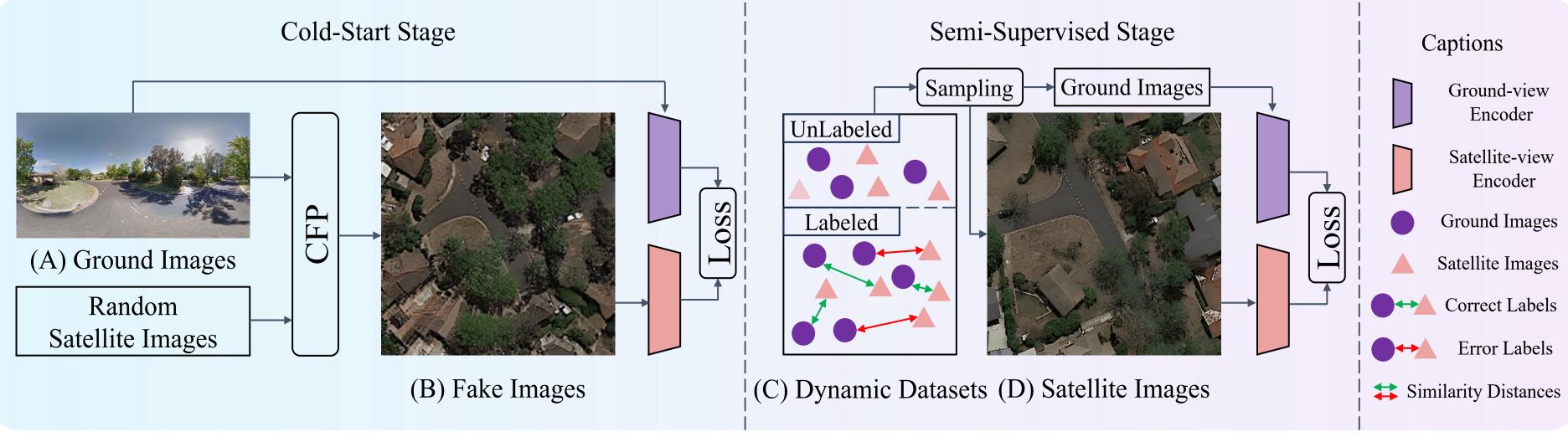
### **Unsupervised CVGL:**

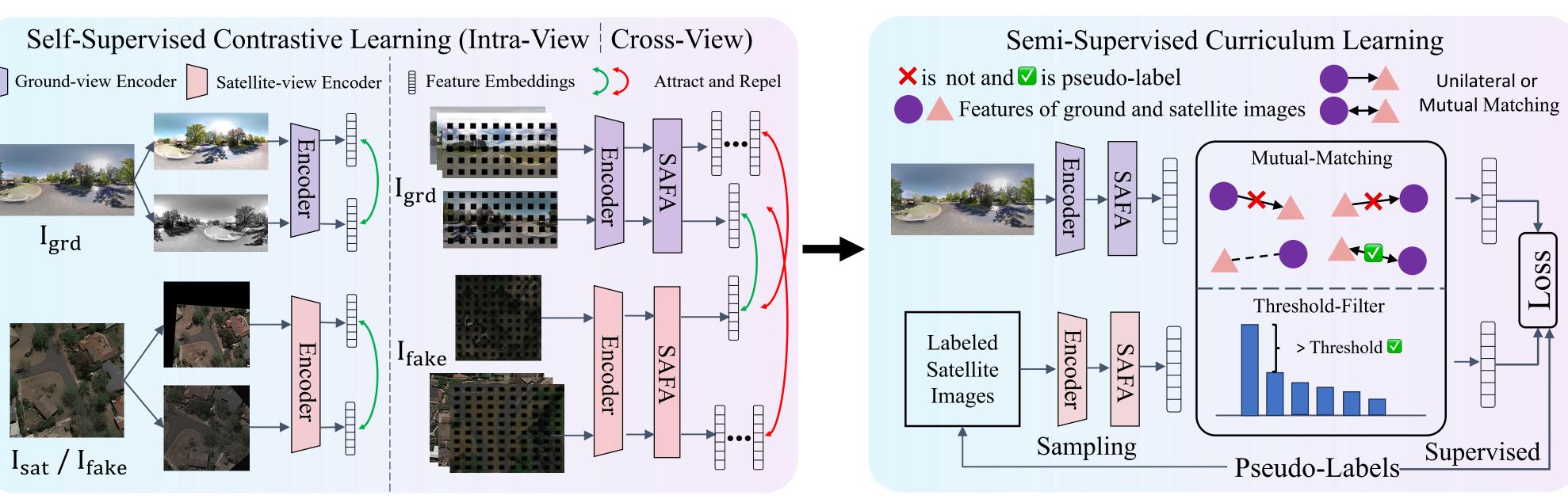
- 1. Cold-Start Stage learns cross-view retrieval by training with ground images and ground images in satellite view (generated by CFP).
- 2. Semi-Supervised Stage trains with ground and pseudo-labeled satellite images. Pseudo labels are produced by the former model and Adaptive-Mutual-Matching re-ranking).



(D) BEV

(C) Spherical Transform





#### Results

# Unsupervised

A namacah	GT Ratio	CVUSA ↑	CVACT Val ↑	CVACT Test ↑	
Approach		R@1	R@1	R@1	
TransGeo (CVPR22)	100%	94.08	84.95	-	
Sample4Geo (ICCV23)	100%	97.83	87.49	60.57	
Ours (ConvNeXt-Small)	100%	93.53	84.44	57.71	
Ours (ConvNeXt-Small)	0%	87.90	82.96	58.85	
Ours (ConvNeXt-Base)	0%	92.56	84.58	60.53	
Ours (ConvNeXt-Base)	10%	94.88	87.89	65.30	

Ablation	CVUSA			CVACT			
(w/o)	R@1 ↑	Labels	Correct ↑	R@1 ↑	Labels	Correct ↑	
(a) BEV	0.0225	14	0	0.0	6	0	
(b) Fake	9.18	119	98	22.78	1349	923	
(c) Intra	15.58	1476	798	42.98	5539	4281	
(d) Cross	0.011	21	0	0.011	48	0	
(e) Ours	17.96	1477	968	44.81	6012	4752	

# Semi-Supervised

GT	CVACT			GT	VIGOR(Chicago)		
Ratio	R@1	R@5	R@10	Ratio	R@1	R@5	R@10
10%	56.10	81.69	88.18	30%	36.82	65.16	74.28
1%	68.29	85.18	88.80	5%	25.82	42.81	49.81
5%	78.10	90.87	93.11	10%	44.17	63.30	69.81
10%	78.88	91.31	93.53	20%	55.90	75.44	81.20
20%	79.60	91.98	93.96	30%	60.42	80.12	84.88
100%	84.44	94.85	98.53	100%	68.40	88.49	92.44

Ablation	CVU	J <b>SA</b> ↑	CVACT ↑	
(w/o)	R@1	R@10	R@1	R@10
(a) Mutual-Matching	14.40	35.43	61.90	82.80
(b) Threshold-Filter	36.52	56.10	82.46	94.04
(c) Curriculum-Learning	32.56	54.35	63.14	82.83
Ours	87.90	97.51	82.96	94.43