Doppelgangers++

Improved Visual Disambiguation with Geometric 3D Features

Yuanbo Xiangli[1], Ruojin Cai[1], Hanyu Chen[1], Jeffrey Byrne[2], Noah Snavely[1]

[1]Cornell University, [2]Visym Labs

Problem Revisit: Doppelgangers in 3D computer vision

Doppelgangers are <u>visually similar</u> but <u>distinct surfaces</u> that confuse 3D reconstruction systems

(a) Internet photos

# **Problem Revisit**: Doppelgangers in 3D computer vision

Pervasive **local feature matching** methods <u>struggle with ambiguity</u>,
lacking <u>global evidence</u> to reject false matches

LightGlue
#matches 579 vs 420

Doppelganger
Stop after 6 layers

True-Match
Stop after 7 layers
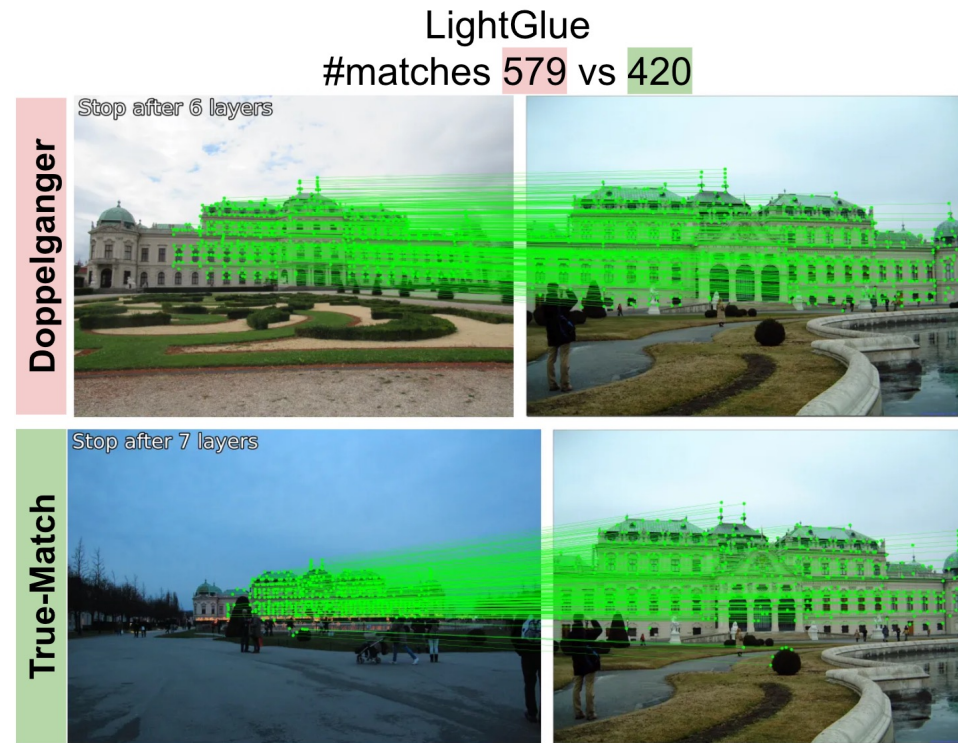
MegaScenes (Tung et.al)

Problem Revisit: Doppelgangers in 3D computer vision

Pervasive **local feature matching** methods <u>struggle with ambiguity</u>,
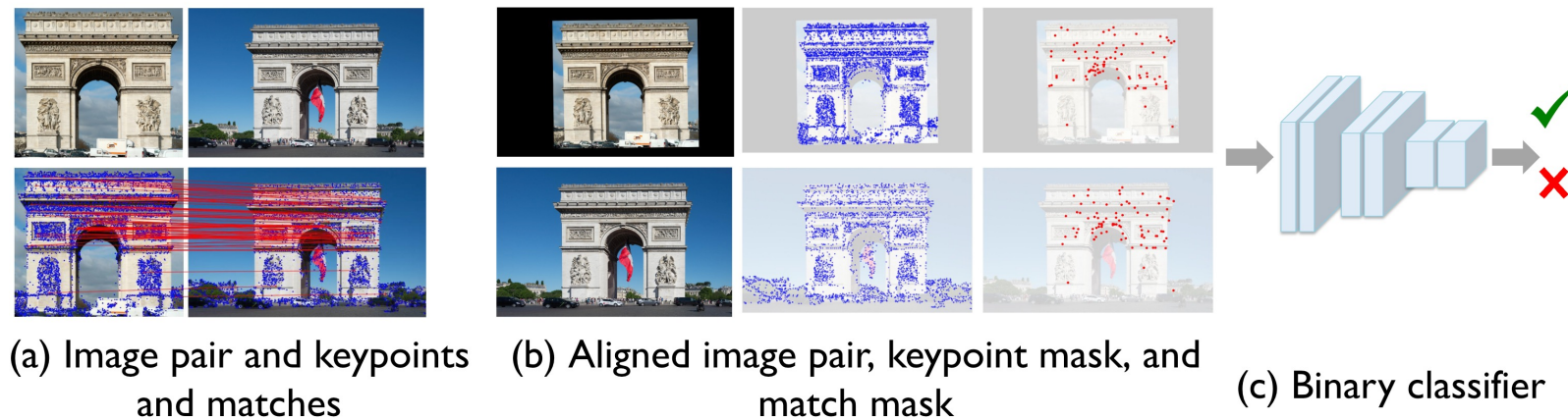lacking <u>global evidence</u> to reject false matches

LightGlue
#matches 884 vs 913
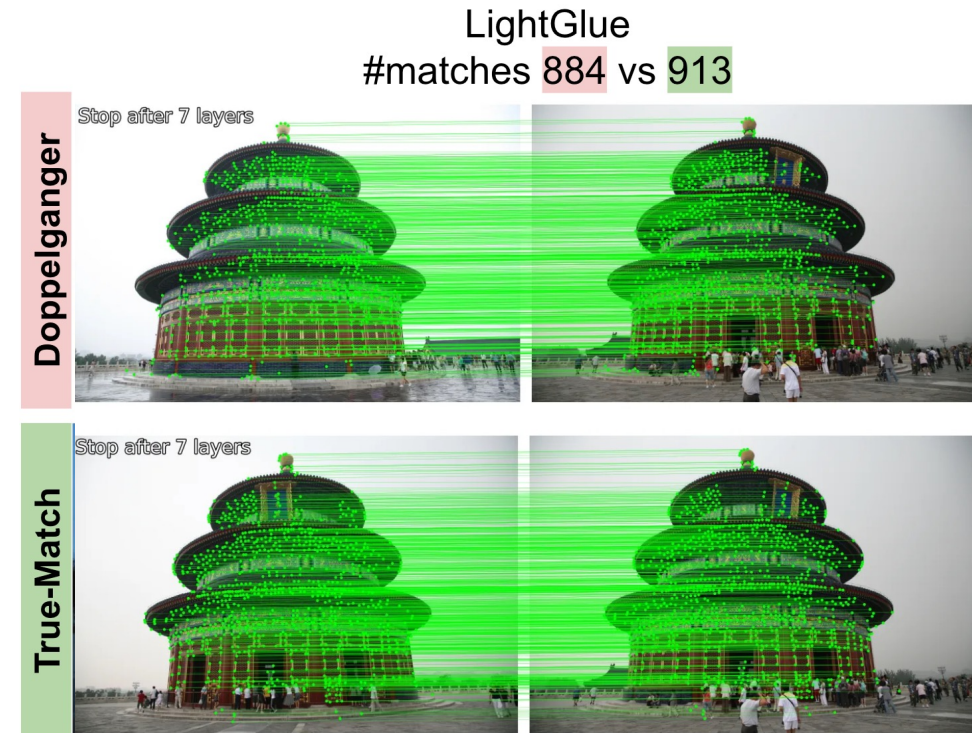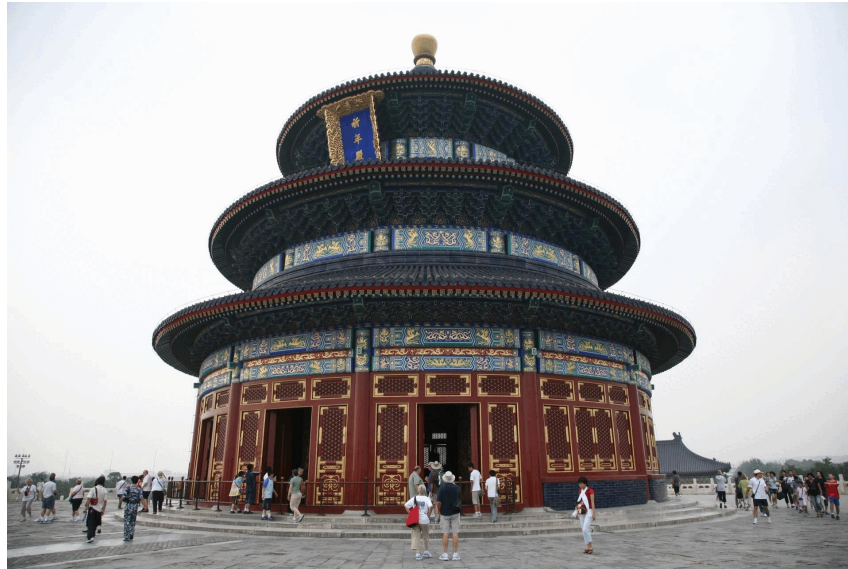
Distinguishing the indistinguishable (Yan et.al)

Prior Work

Construct a Doppelganger dataset from Wikimedia Commons

Train a classifier with keypoint and match locations (LoFTR)

(a) Image pair and keypoints and matches

(b) Aligned image pair, keypoint mask, and match mask

(c) Binary classifier

Cai, Ruojin et al. "Doppelgangers: Learning to Disambiguate Images of Similar Structures." 2023 IEEE/CVF International Conference on Computer Vision (ICCV) (2023): 34-44.
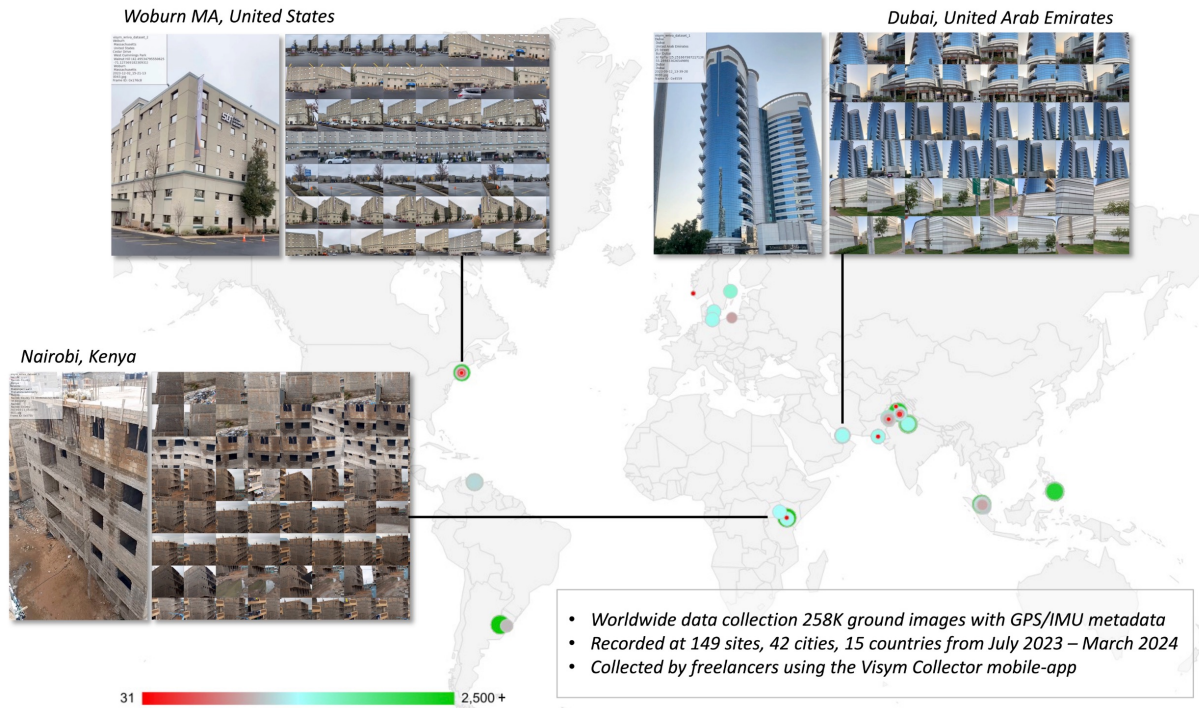
Challenges

- High precision is crucial
  - Even a few incorrect matches can ruin the 3D reconstruction.
- Threshold sensitivity
  - Small changes in classifier thresholds can significantly impact reconstruction;
  - Different scenes require different classifier thresholds.
- Limited generalization
  - Trained on landmark images, the method struggles with everyday scenes.

GT: 1 / DG-OG: 0.473

GT: 1 / DG-OG: 0.349

GT: 0 / DG-OG: 0.921

What's different?

GT: 0 / DG-OG: 0.713

Prior work: **failed, two sides collapsed**
(947 images)

GT:1 / **DG-OG**: 0.190   GT:1 / **DG-OG**: 0.378   GT:0 / **DG-OG**: 0.993   GT:0 / **DG-OG**: 0.994

Incorrect Cameras

Over-split
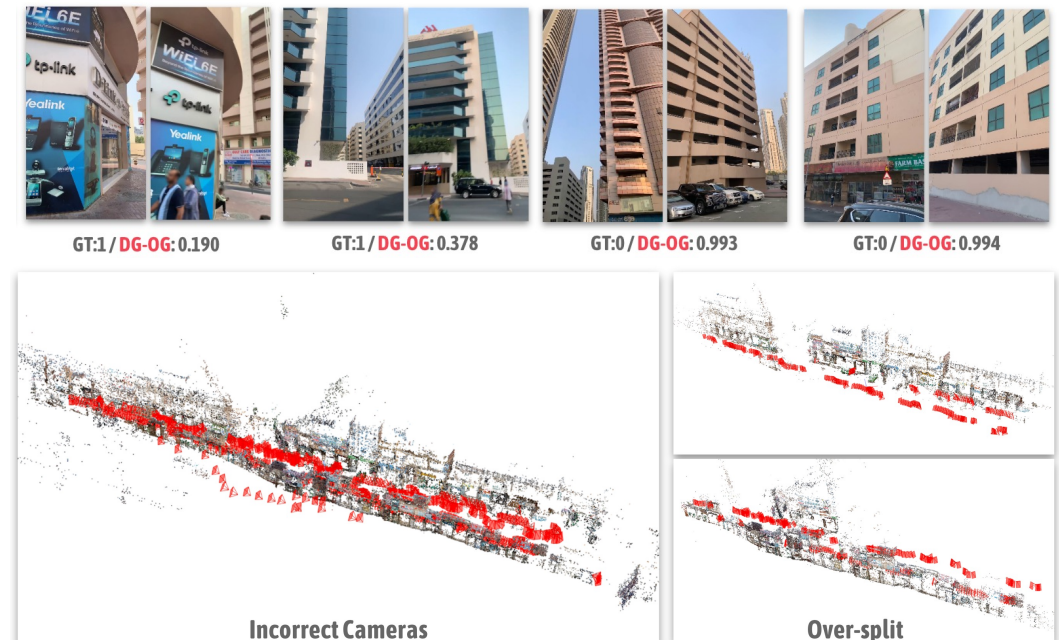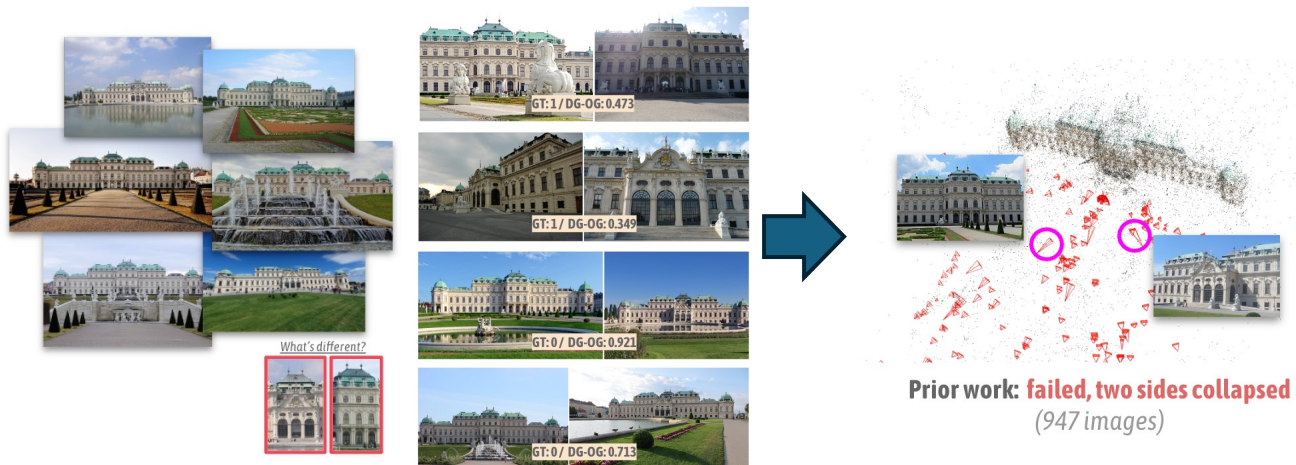
Doppelgangers++: Expand Training Data

VisymScenes Dataset

Woburn MA, United States

Dubai, United Arab Emirates

Nairobi, Kenya

- Worldwide data collection 258K ground images with GPS/IMU metadata
- Recorded at 149 sites, 42 cities, 15 countries from July 2023 – March 2024
- Collected by freelancers using the Visym Collector mobile-app

31   2,500 +

Site01   Images   Doppelganger Example   GPS & IMU
Site02   Images   Doppelganger Example   GPS & IMU
Site03   Images   Doppelganger Example   GPS & IMU
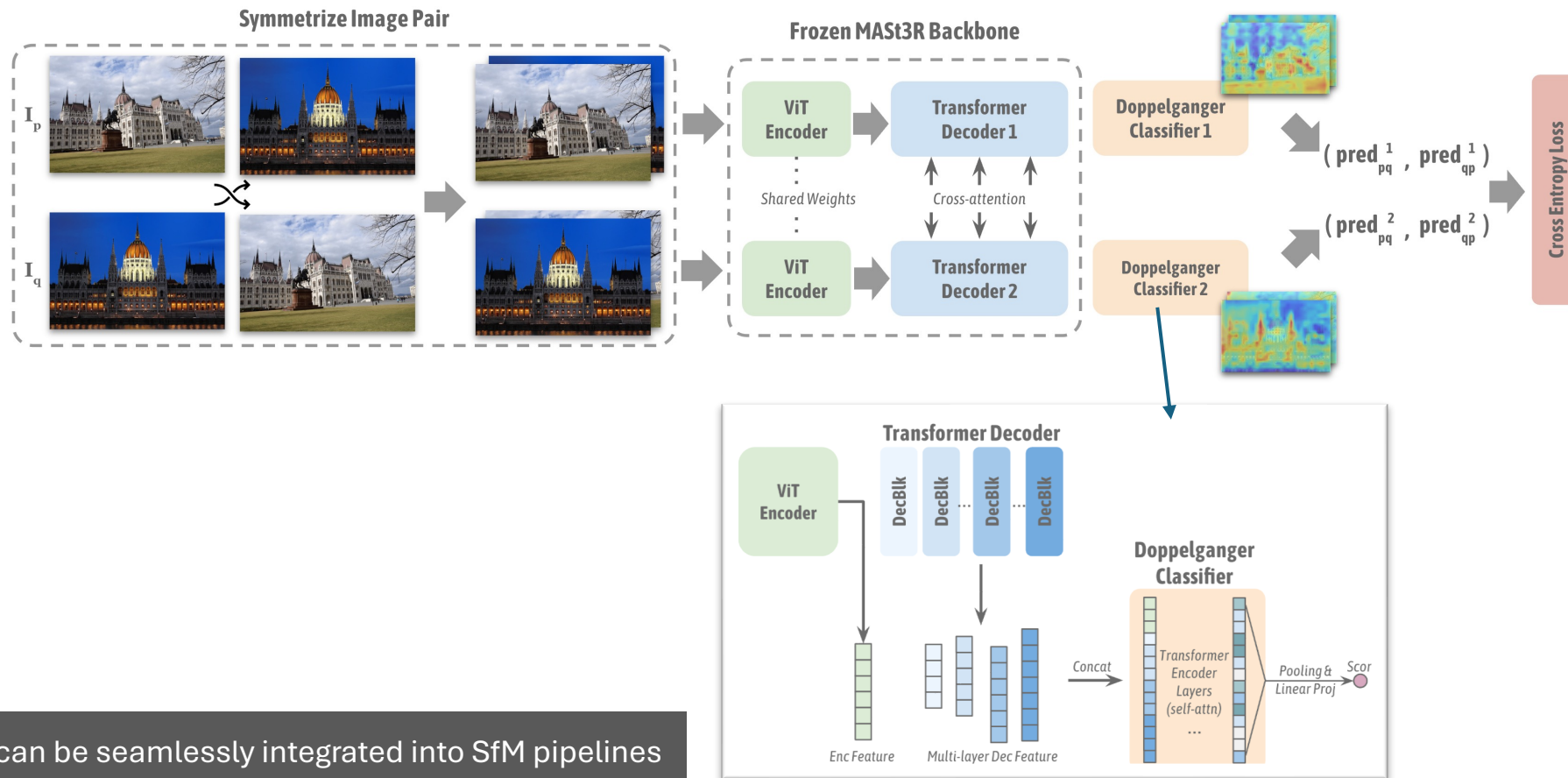Site04   Images   Doppelganger Example   GPS & IMU

Includes residential areas, landmarks, historical sites, business districts, and more;
Recorded with GPS/IMU metadata;
Complement the notable landmarks available on Wiki Commons.

Mined ~53K doppelganger and true matching pairs leveraging their recorded geo-locations and viewing directions.

Doppelgangers++: Improved Doppelganger Classifier

Leverage multi-level geometric features of a SOTA pairwise image matching model (MASt3R)
With a Transformer classification head

The classifier can be seamlessly integrated into SfM pipelines
(E.g. COLMAP, MASt3R-SfM)

Leroy, Vincent et al. "Grounding Image Matching in 3D with MASt3R." European Conference on Computer Vision (2024).

Evaluating Doppelganger Correction in SfM

Prior work uses manual inspection
We introduce an automated, geotag-based method for validating reconstructed models
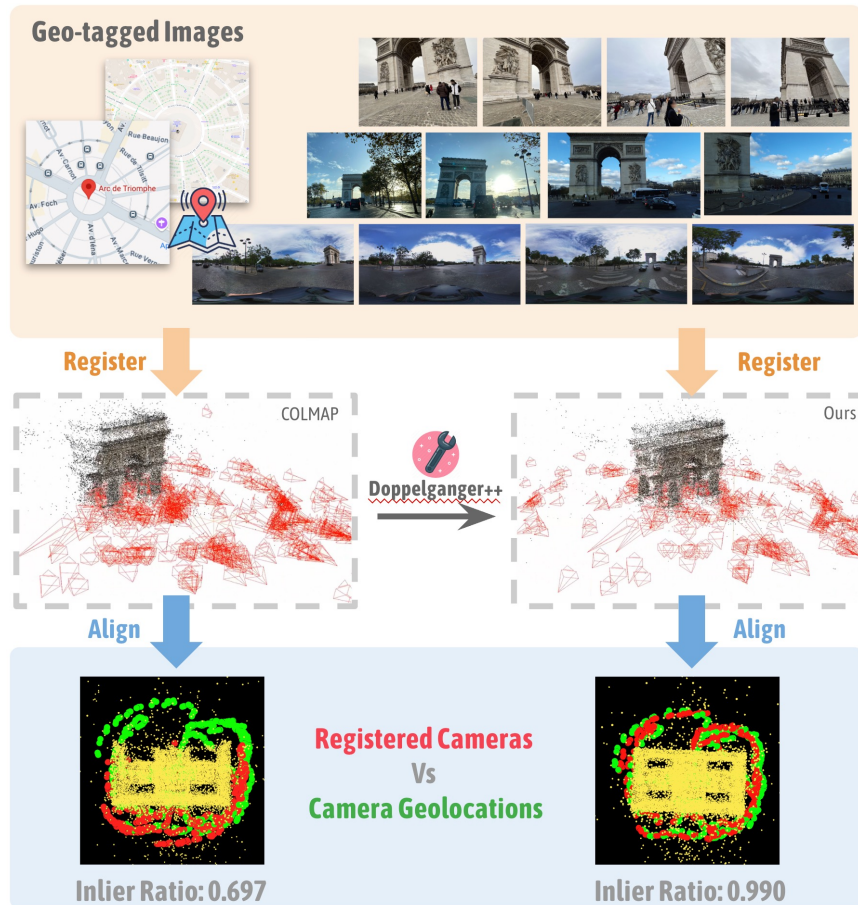
- **Image registration**
  - We collect geo-tagged images from Mapillary near the reconstructed scene;
  - Register geo-tagged Mapilllary images to the SfM model.
- **Alignment cameras**
  - Registered cameras are aligned with their geolocations using RANSAC;
  - We use <u>inlier ratio</u> as an indicator for accuracy.

Incorrect models have lower inlier ratio (largely misaligned trajectories);
Correct models have higher inlier ratio (better aligned trajectories)

Results

Evaluation on Pairwise Visual Disambiguation
- In-domain test: DG, VisymScenes
- Out-of-domain test: Mapillary

Test Data	Method	Metrics (trained on DG / trained on DG + VisymScenes)			
		AP↑	ROC AUC↑	Prec@Recall=0.85↑	Recall@Prec=0.99↑
DG	DG-OG	0.954 / 0.956	0.944 / 0.947	0.901 / 0.910	0.611 / 0.614
	Ours	0.980 / 0.981	0.981 / 0.981	0.972 / 0.982	0.690 / 0.642
VisymScenes	DG-OG	0.816 / 0.938	0.726 / 0.921	0.498 / 0.831	0.340 / 0.623
	Ours	0.936 / 0.991	0.909 / 0.990	0.892 / 0.999	0.542 / 0.901
Mapillary	DG-OG	0.566 / 0.692	0.581 / 0.701	0.523 / 0.572	0.003 / 0.000
	Ours	0.950 / 0.968	0.944 / 0.958	0.927 / 0.942	0.310 / 0.736

Dataset expansion improves Doppelganger classifier performance
Ours is more generalized than previous work (DG-OG)

Results

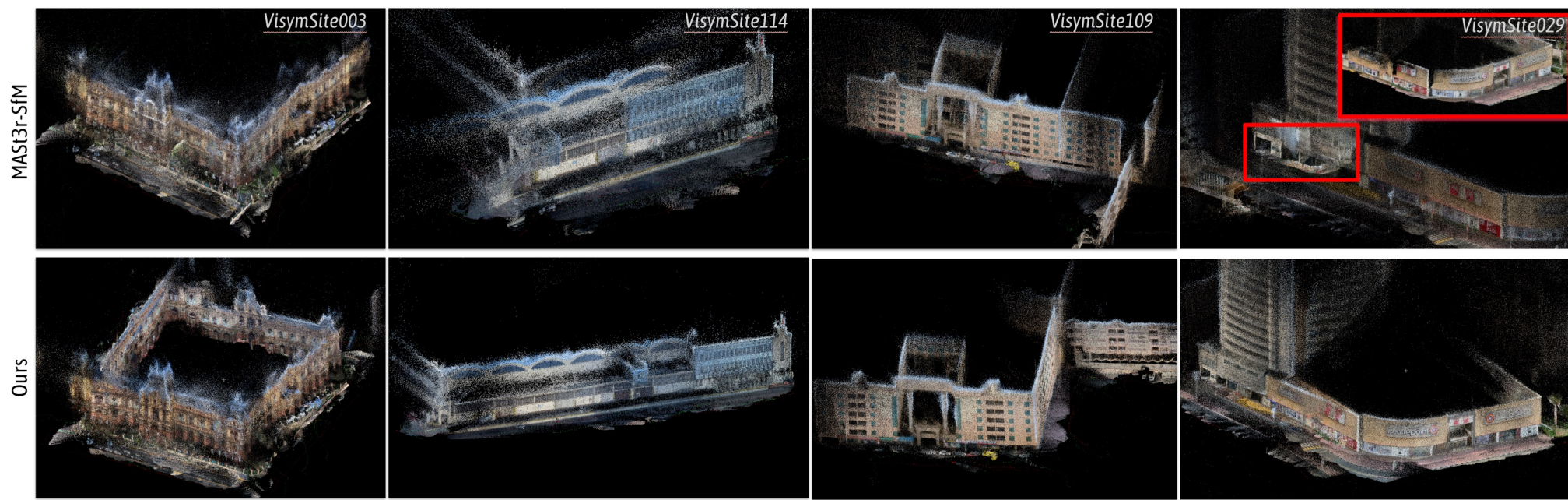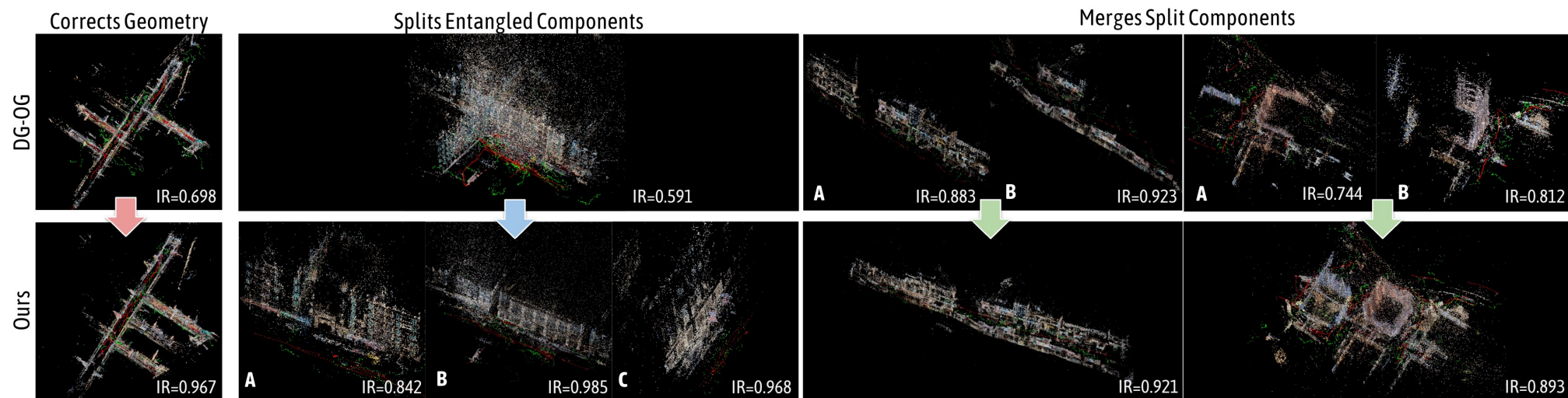Evaluation on SfM Reconstruction Disambiguation
- Classifier threshold τ=0.8 is used unless otherwise stated.
- The '+' symbol indicates split reconstruction components;

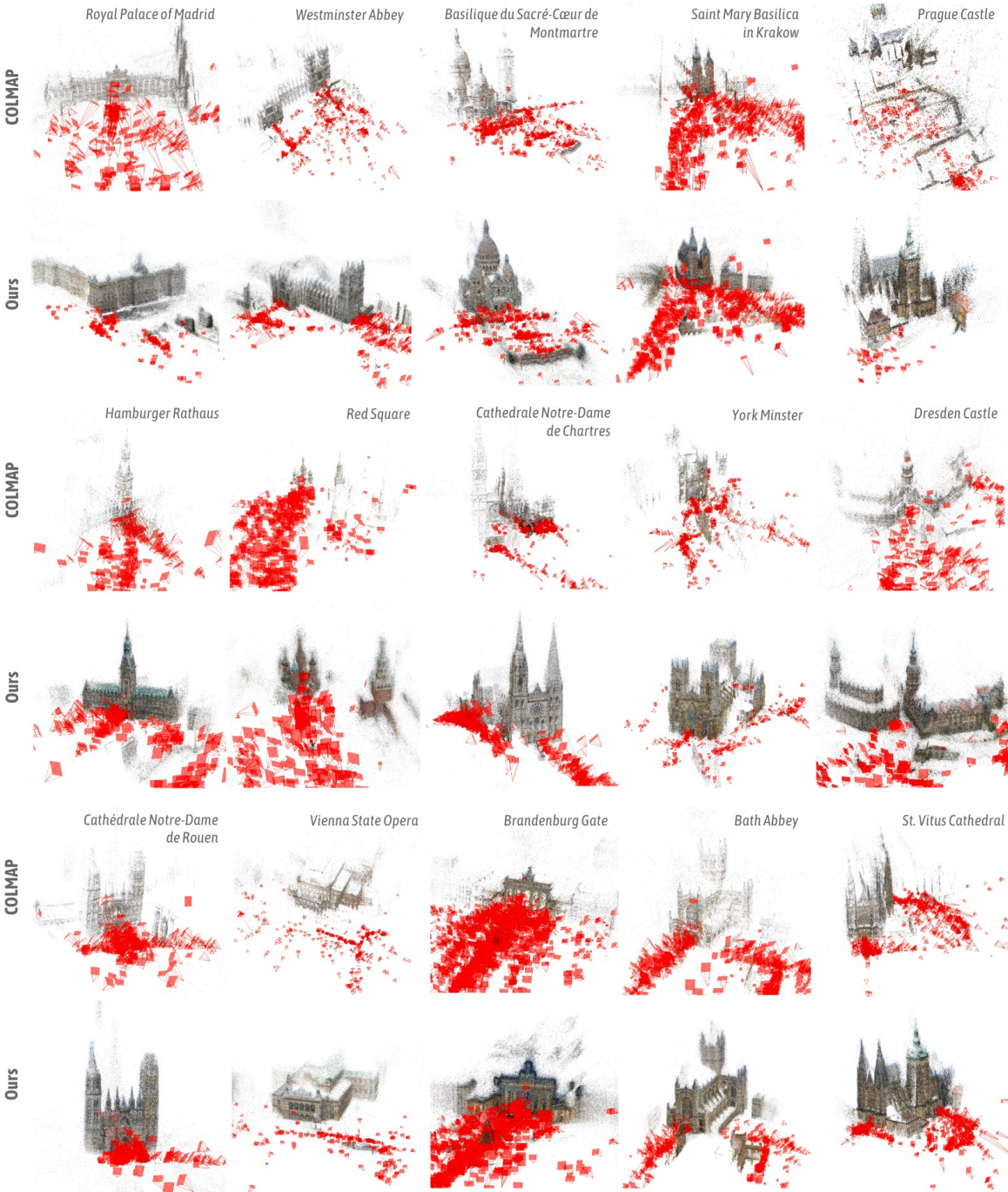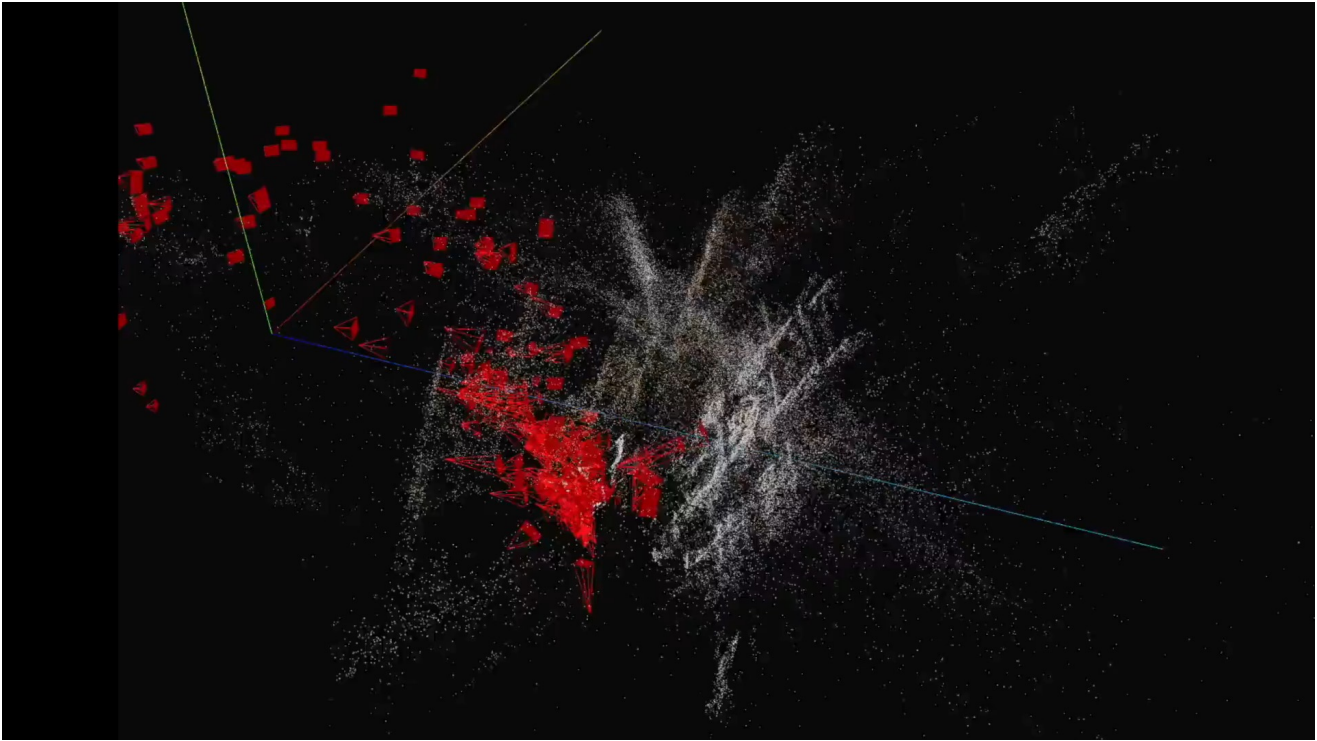| Test Scenes	# SfM-registered Images			Inlier Ratio (Sec. 3.3)		
	COLMAP	DG-OG	Ours	COLMAP	DG-OG	Ours
Alexander Nevsky Cathedral [6]	447	444	447	0.565	1.0	1.0
Arc de Triomphe [6]	424	384	423	0.697	0.966	0.990
Berliner Dom [6]	1603	1588	1606	0.709	0.987	0.992
Big Ben [6]	398	379	394	0.564	0.827	0.831
Church on Spilled Blood [6]	273	84+94 (τ=0.97)	157+106	0.542	0.881	0.962
Radcliffe Camera [6]	281	91+84	94+186	0.495	0.955	0.970
Seville [18]	1498	585+272+515	615+303+552	0.450	0.772	0.854
Brandenburg Gate [14]	2137	1361+570	1398+603	0.440	0.900	0.909
Palacio de Comunicaciones [14]	727	307+80 (τ=0.6)	308+84	0.229	0.823	0.934
Ponte di Rialto [14]	652	538+101 (τ=0.6)	540+107	0.627	0.834	0.844
York Minster [14]	636	200+362	206+284	0.727	0.858	0.901
Town Hall Tower, Kraków [14]	298	255	280	0.609	0.731	0.838
Belvedere (Vienna) [14]	1038	851 (fail)	457+500	0.521	0.451	0.874
Reichstag (building) [14]	1504	997+310	1024+356	0.469	0.804	0.862
St. Vitus Cathedral [14]	752	673	692	0.853	0.909	0.933
Royal Liver Building [14]	212	171	180	0.7	1.0	1.0
VisymSite0010	1544	520+290	1446	0.770	0.820	0.913
VisymSite0023	849	471+81	566+82	0.867	0.848	0.942
VisymSite0028	450	238+179	267+120	0.818	0.711	0.909
VisymSite0042	540	206+207	467	0.863	0.924	0.963
VisymSite0109	1245	237+458+78	239+612+127	0.857	0.862	0.927

Previous work (DG-OG) require frequent threshold adjustment;

Ours is more discriminative and generalized; and has better reconstruction correctness and completeness.

SfM Disambiguition on VisymScenes

Corrects Geometry

Splits Entangled Components

Merges Split Components

DG-OG

IR=0.698 IR=0.591 A IR=0.883 B IR=0.923 A IR=0.744 B IR=0.812

Ours

IR=0.967 A IR=0.842 B IR=0.985 C IR=0.968 IR=0.921 IR=0.893

MASt3r-SfM

VisymSite003 VisymSite114 VisymSite109 VisymSite029

Ours

SfM Disambiguition on MegaScenes

Check out project page for more!