C3Po: Cross-View Cross-Modality Correspondence by Pointmap Prediction

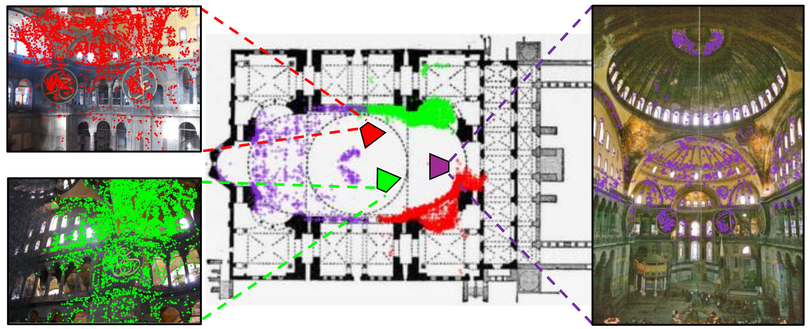
Kuan Wei Huang, Brandon Li, Bharath Hariharan, Noah Snavely

Cornell University

NEURAL INFORMATION PROCESSING SYSTEMS

Project Page

Overview



Task: Draw **correspondences** between **floor plans** (abstract drawings from bird's-eye view) and ground-view **photos**.

Current Limitations: **No existing dataset** provides ground truth correspondences between floor maps and photographs.

Dataset	Scene Photos	Scene Layouts	Correspondences
2D-3D-S	✓		
Matterport3D	✓		
Gibson Env	✓		
Structured3D	✓	✓	
ZInD	✓	✓	
Swiss Dwellings		✓	
MSD		✓	
WAFFLE		✓	
C3Po (ours)	✓	✓	✓

Contributions

- Create the **first cross-view, cross-modality correspondence dataset.**
- Show that **current state-of-the-art correspondence methods fail** to draw accurate correspondences between photos and floor plans.
- **Adapt DUSt3R's pointmap prediction** to estimate correspondences, **outperforming** the best baseline by **34%**.
- **Identify systematic sources of error** due to the **natural ambiguity** in data for future work to explore.

Dataset Statistics

90K Plan-Photo Pairs

153M Correspondences

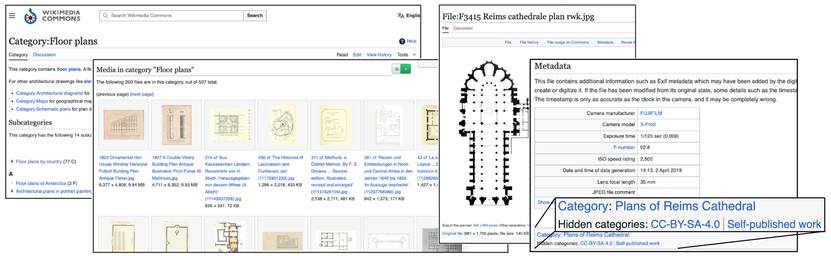
597 Scenes

85K Camera Poses

Dataset Creation

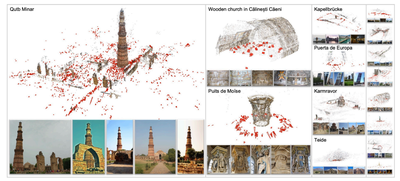
Source Floor Plans by Scenes

Wikimedia Commons → Category: Floor plans → [Images in all Subcategories]



Collect Photos Corresponding to Scenes

MegaScenes



YFCC100M

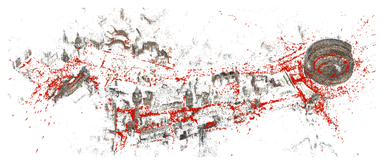
(a) *IMG_9793: Streetcar (Toronto Transit)* by Andy Nystrom CC BY NC SA https://flic.kr/p/jcjMdz.

(b) *Celebrating our 6th wedding anniversary in Villa Mary* by Rita & Tomek CC BY NC SA https://flic.kr/p/fCXEJi.

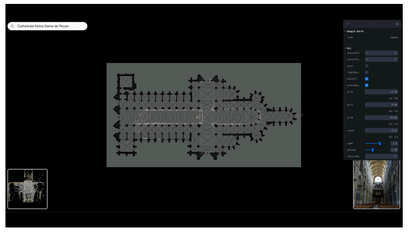
Figure 3: Two photos of real world scenes from photographers in the YFCC100M dataset.

Determine Correspondences

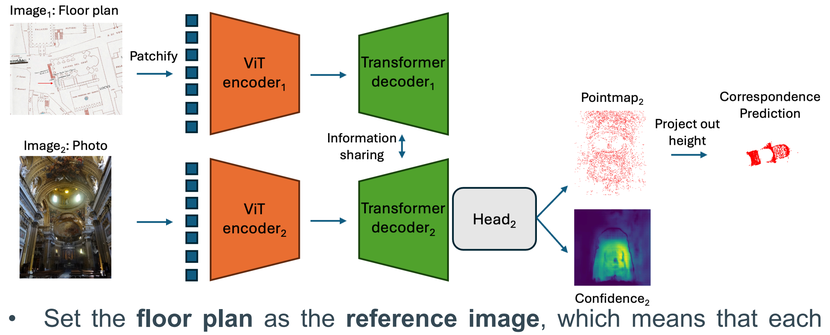
Run **COLMAP** on Photo Collections → Camera Poses + Point Clouds



Custom UI for **Manual Alignment**



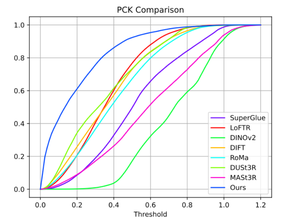
Correspondence by Pointmap Prediction

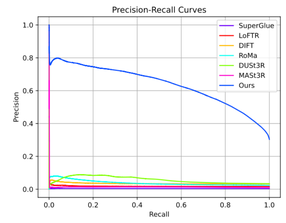


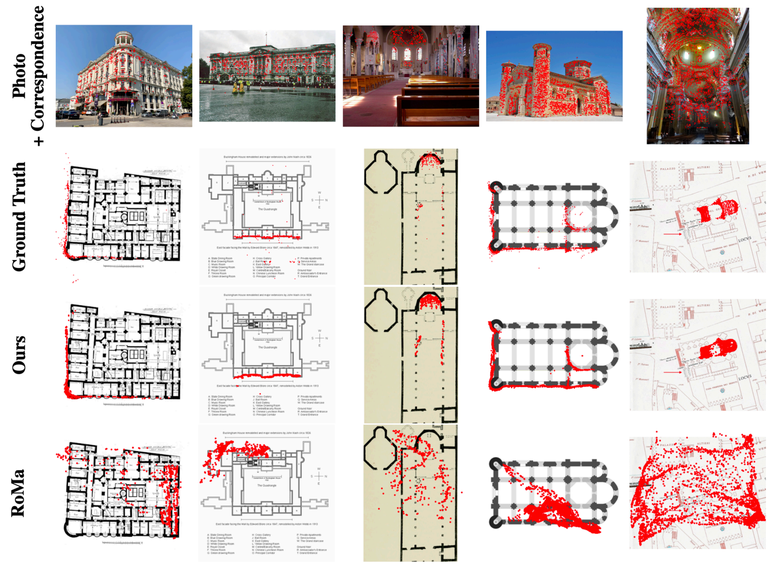
- Set the **floor plan** as the **reference image**, which means that each image pixel can be mapped to a 3D point in the floor plan coordinate frame.
- To obtain a correspondence, **query a photo pixel to retrieve its 3D point** and project it onto the floor plan by **dropping the Y-coordinate**.

Results

	RMSE (↓)
SuperGlue	0.4050
LoFTR	0.2901
DINOv2	0.5338
DIFT	0.3036
RoMa	0.3308
DUSt3R	0.2925
MASt3R	0.4616
Ours	**0.1919**







Open Challenges