

7 Supplemental Material

7.1 Predefined Scene Categories

We list below all categories used to identify our scenes of interest (Section 3.1).

- amphitheatre
- architectural structure
- architecture
- basilica
- building
- castle
- cathedral
- chapel
- château
- church
- destroyed building or structure
- fortification
- hospital
- house
- hotel
- library
- mausoleum
- mosque
- museum
- pagoda
- palace
- theatre
- synagogue
- temple

7.2 Hyperparameters

Prediction Head	DPT
Optimizer	AdamW
Base learning rate	1e-4
Minimum learning rate	1e-6
Weight decay	0.05
Adam β	(0.9, 0.95)
Batch size	48
Epochs	10
Warmup epochs	3
Learning rate scheduler	Cosine decay
Input resolution	512×512
Image augmentations	For each plan-photo input, we only perform augmentations on the plan: color jitter, random cropping, and random rotation.
Initialization	DUST3R

7.3 Custom User Interface to Align Point Clouds to Floor Plans

To facilitate the process of aligning point clouds with floor plans, we create a custom user interface (UI). The UI visualizes the point clouds and floor plans of each scene and enables users to interactively apply transformations. For each scene, our UI has a dropdown menu (Figure 5-right) listing all corresponding reconstructions and floor plans. After making a selection, a floor plan and a top-down view of a point cloud are displayed (Figure 5-top). Users can modify the 2D translations, rotation, and scale of the floor plan and point cloud until they are satisfied with the alignment (Figure 5-bottom). The transformation parameters are saved in a database and will be used in Section 3.3.2. However, aligning the point cloud to the floor plan is often not as straightforward as the example shown in Figure 5. We implement three additional features in the interface to help with alignment. We provide a Google Earth link for each scene at the top of the menu bar, which offers a holistic 3D mesh view of the scene. To help users form a better mental representation of the point cloud geometry, the bottom-left panel allows them to orbit, zoom, and pan a camera around the point cloud. The bottom-right corner displays a set of images used to reconstruct the point cloud, which users can click through to obtain useful information, such as the viewpoint or whether the scene is indoors or outdoors.

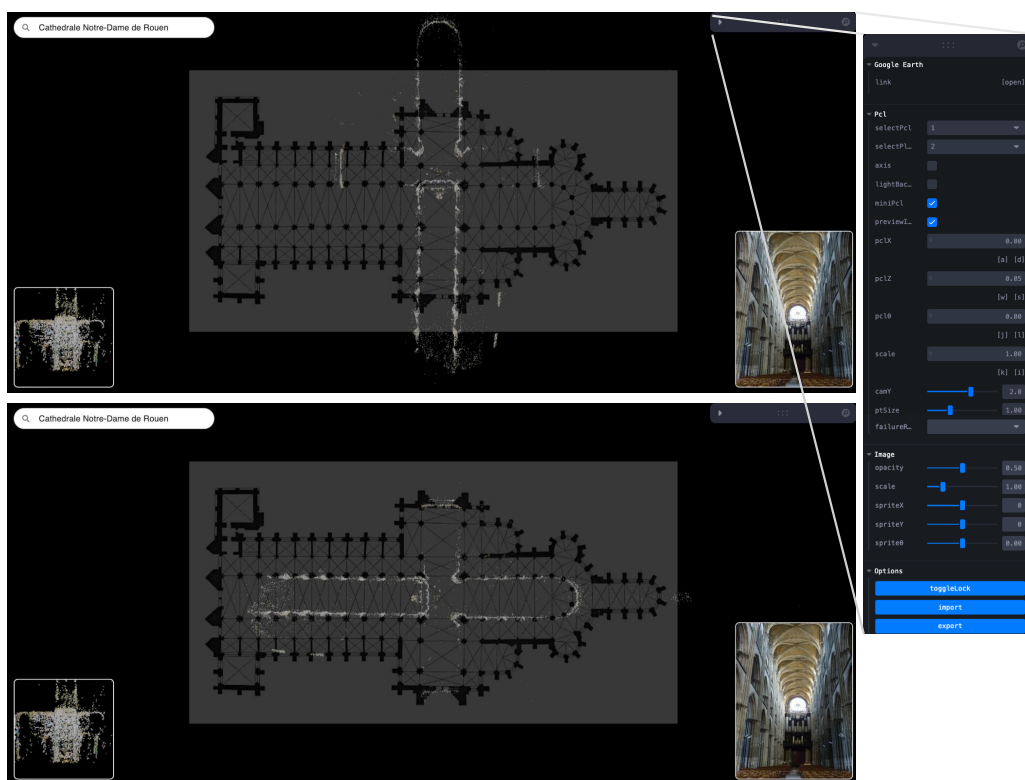


Figure 5: User interface for manual alignment. Top: A sample point cloud and floor plan of a scene that has yet to be aligned. Bottom: A sample point cloud and floor plan of a scene that is aligned. Right: A menu bar with functionalities including Google Earth link for the scene, dropdowns to select floor plans and COLMAP reconstructions, and controls for point cloud and floor plan transformations.

7.4 Qualitative Results with Confidence Scores

We have shown several qualitative results in Figure 3, and in this section, we share additional insights by analyzing the confidence scores of our model’s correspondence predictions. For reference, the DUST3R representation produces a pointmap and confidence score as output for each pixel, and we discuss in Section 4.2 how we turn pointmaps to correspondences. We find that correct correspondence predictions by C3Po are generally accompanied by high confidence scores (Figure 6), while incorrect predictions have low confidence scores (Figure 7). Photos from lower confidence results usually exhibit ambiguity, as in the cases identified in Section 4, whereas the camera pose is more easily identifiable in higher confidence results. This is further corroborated by the PR curves in Figure 3-right, which show that C3Po significantly outperforms state-of-the-art models because more confident correspondence predictions are more likely to be correct.

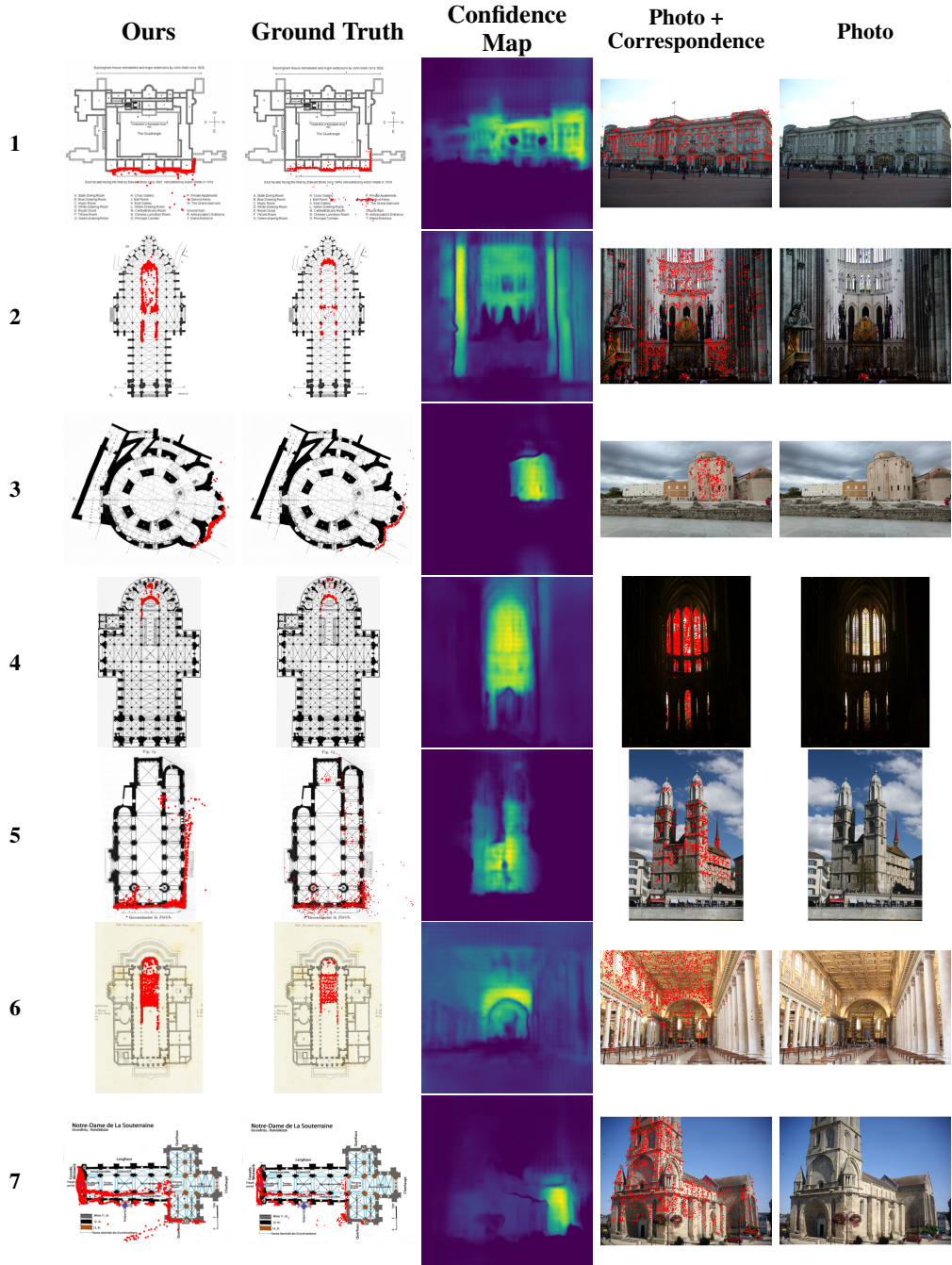


Figure 6: Correct correspondence predictions by C3Po are generally accompanied by high confidence scores.

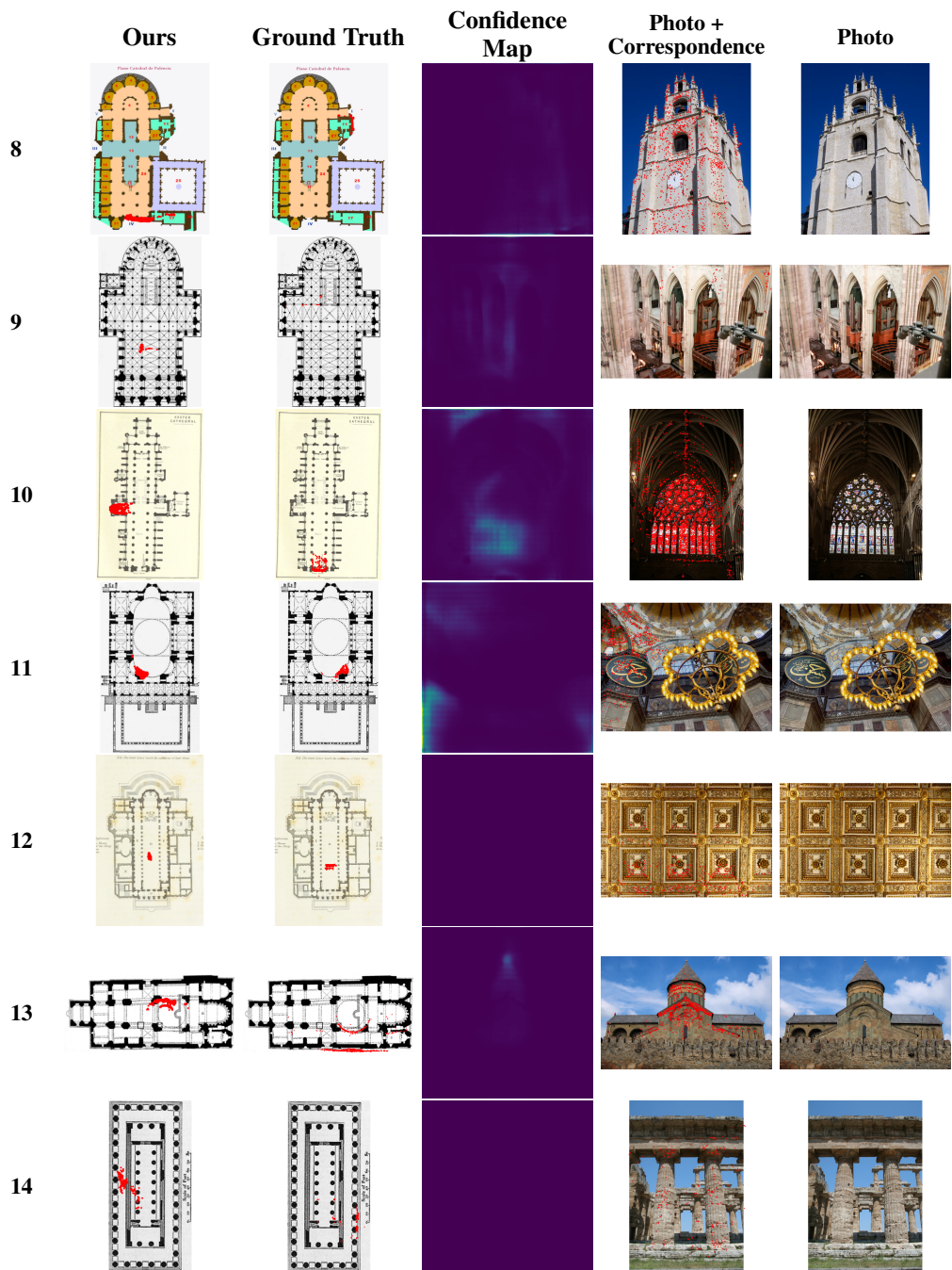


Figure 7: Incorrect predictions often have low confidence scores, and the photos from lower confidence results usually exhibit ambiguity, like the cases identified in Figure 4.

7.5 Validation Set Analysis

We randomly hold out 10% of the train set to use as validation set. We show the RMSE and PR curves of the validation and test sets below. The PR curves are plotted under the same conditions as Section 4.1, where we consider a prediction to be correct if its Euclidean distance from the ground truth is less than 0.05 units in normalized floor plan coordinates. We observe a significant drop in RMSE and an improvement in PR curve on the validation data compared to the test data. More specifically, our model performs better on plan-photo pairs where the plan is seen during training and the photo is unseen.

	RMSE (\downarrow)
Ours (test)	0.1877
Ours (validation)	0.0369

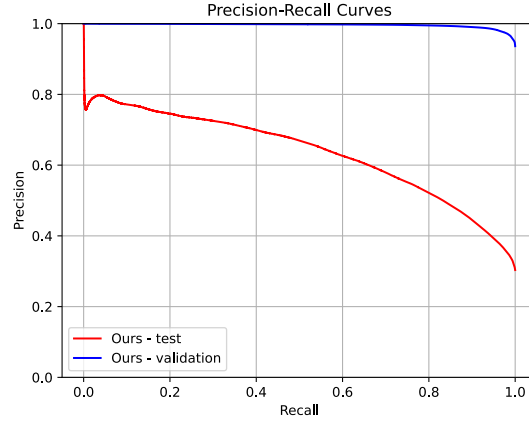


Figure 8: Quantitative results for our model evaluated on the C3 test set and validation set (pairs held out from the train set). Left: table of RMSE errors (lower is better). Right: Precision-Recall curves generated by thresholding on predicted confidence or score for each method.

7.6 Additional Qualitative Results

Below, we show additional qualitative results by C3Po, randomly selected from the test set.

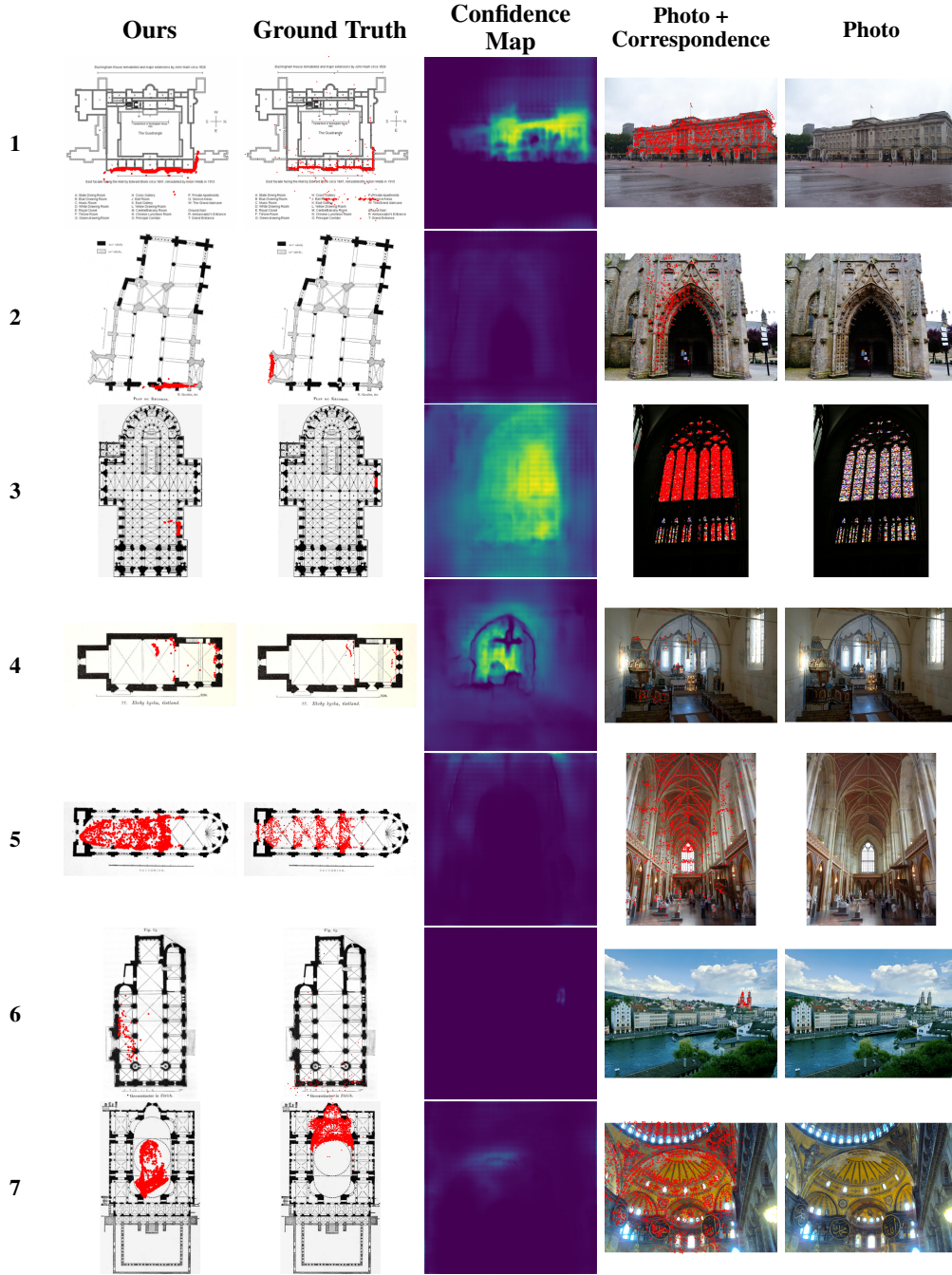


Figure 9: Additional qualitative results

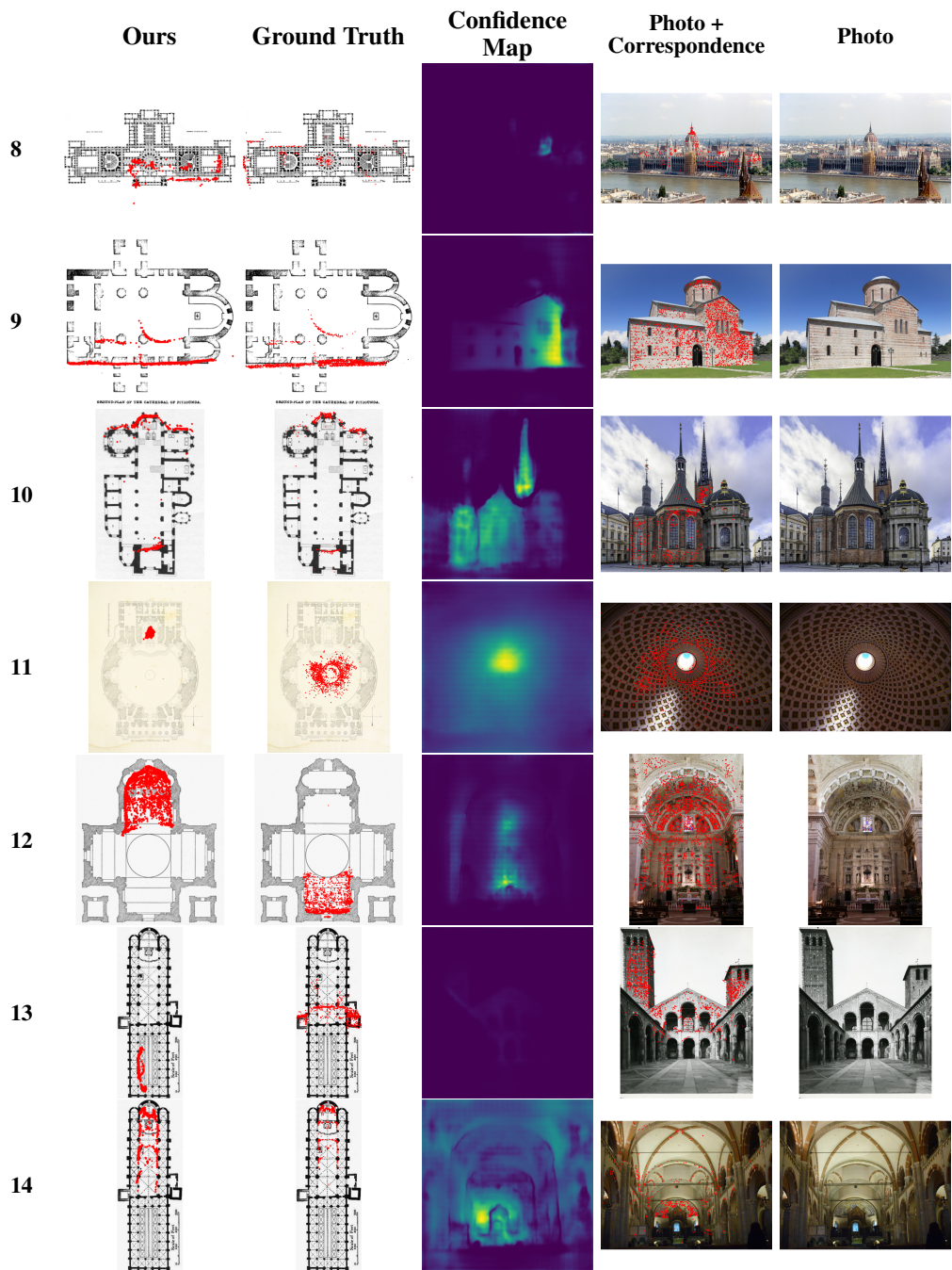


Figure 10: Additional qualitative results.