ETH zürich

Magic Leap

CVG Computer Vision and Geometry Lab

Visual Localization against Universal 3D Models

Eric Tüschenbönner, Patrik Schmuck, Lukas Bernreiter

Master’s Thesis Final Presentation
11.11.2024

Motivation

Visual Localization is essential for **AR** applications

Problem

- State-of-the-art solutions require **mapping**
- Time-consuming for larger scenes, such as **buildings**

Solution

- Existing 3D models for direct localization

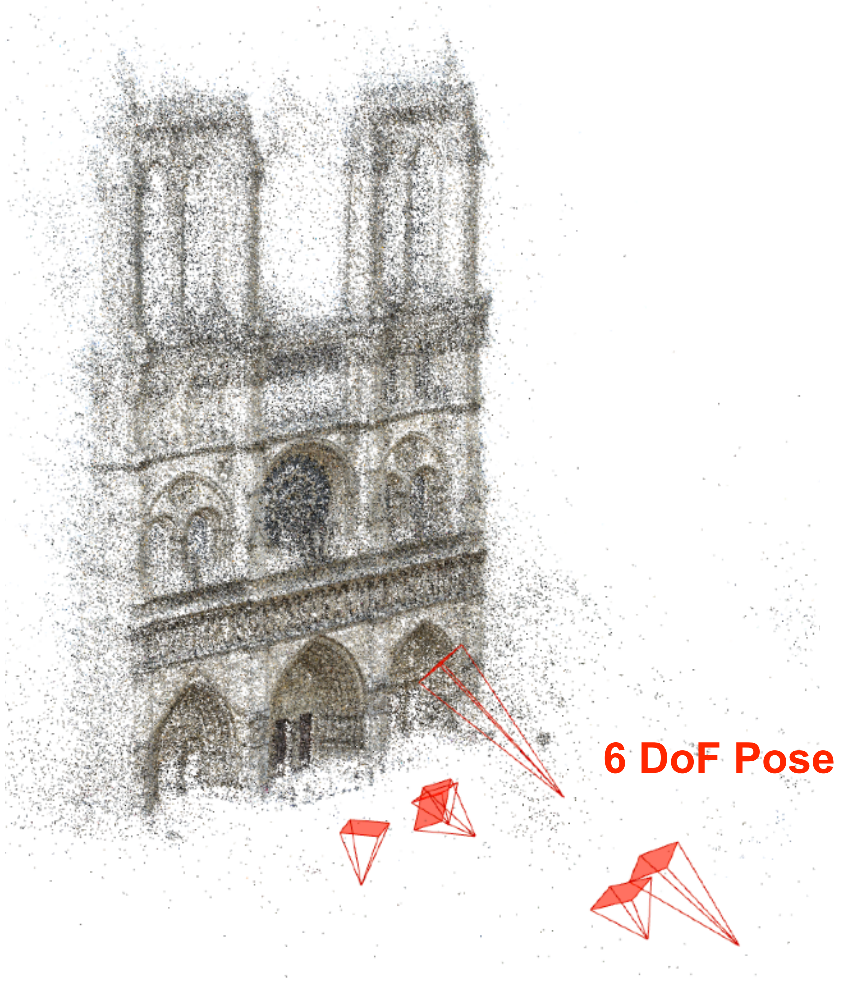
Challenge

- Domain gap between real and synthetic data

Problem Description

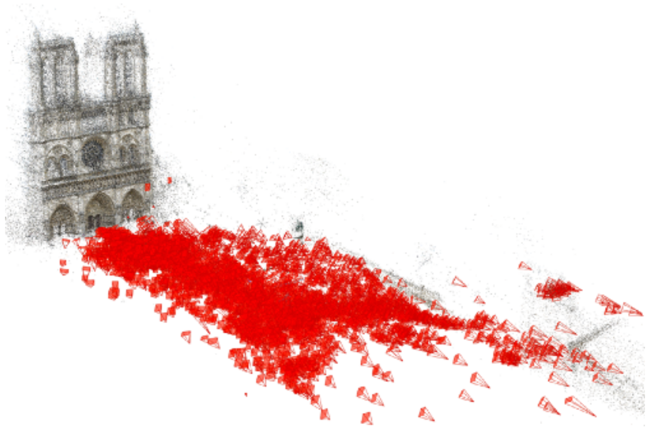
Query: Real Image

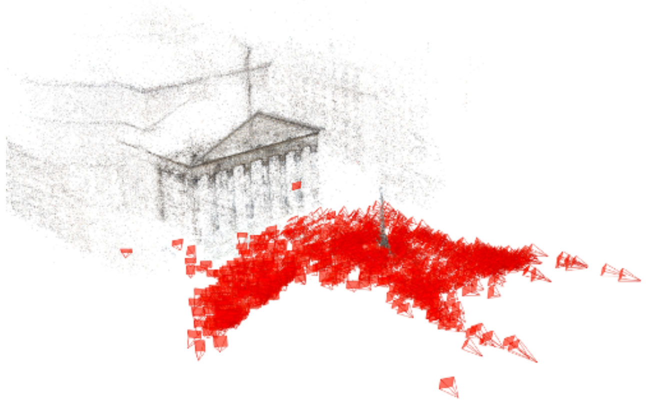
3D Model

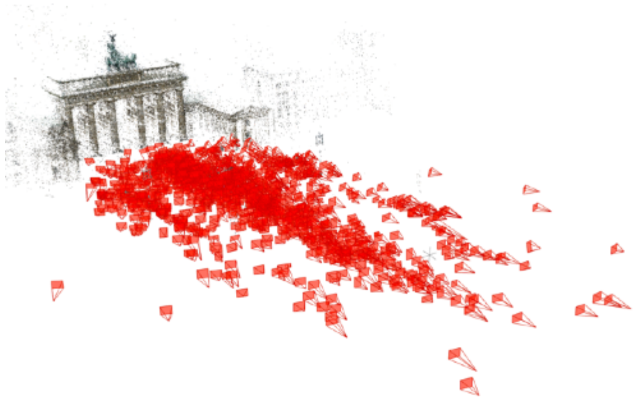


Datasets

Reconstructed Point Clouds [1]



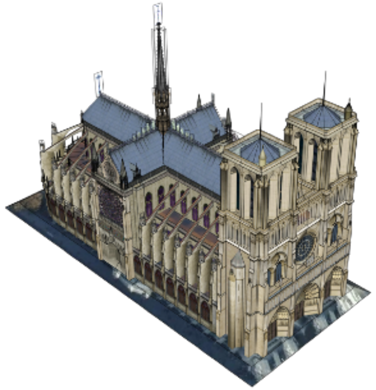




Notre Dame

Pantheon

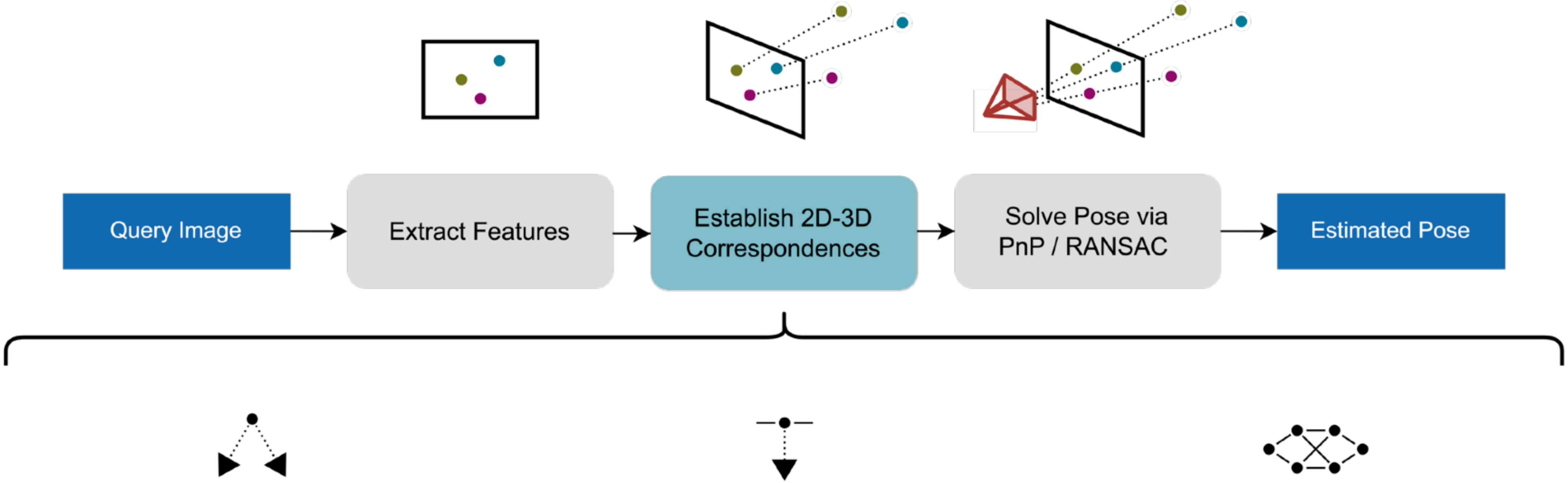
Brandenburg Gate



CAD Models

1. IMC Phototourism Dataset (https://www.cs.ubc.ca/~kmyi/imw2020/data.html)

Structure-based Localization



Reconstruction

Match features with database of 3D coordinates, explicitly calculated via reconstruction.

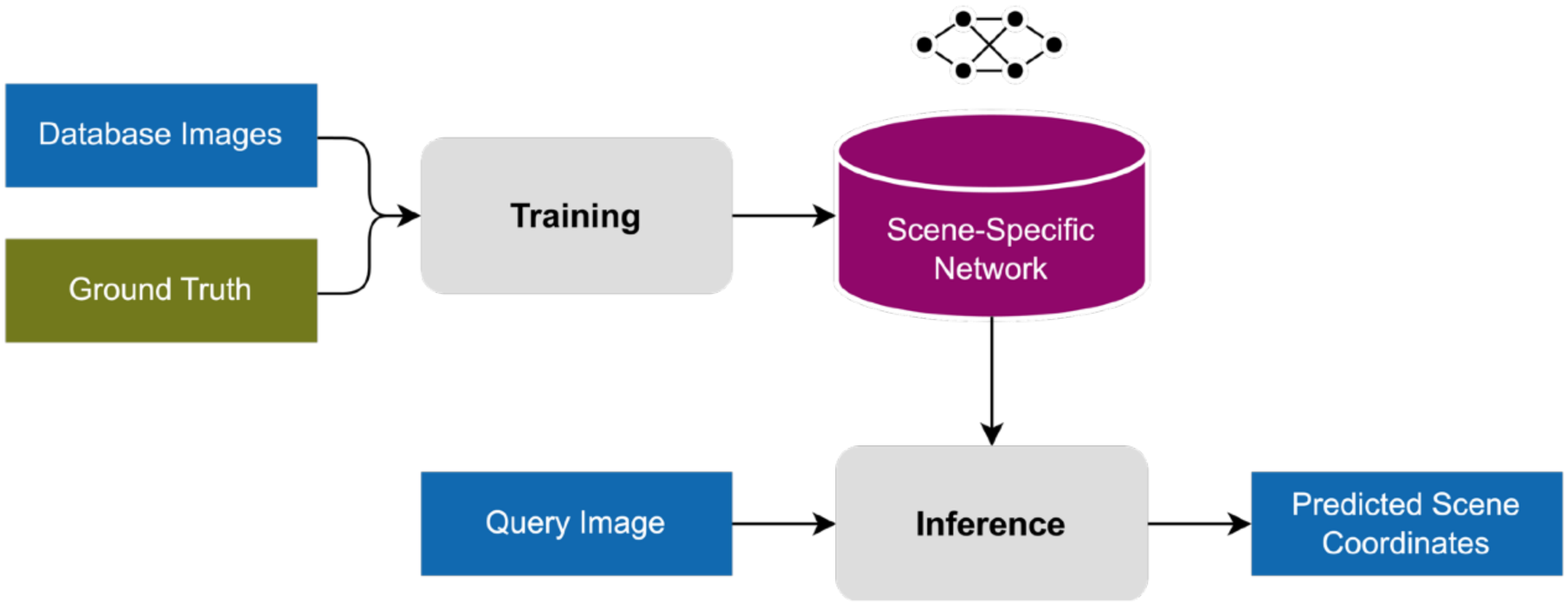
Depth Maps

Match features with database of depth values to directly lift to 3D coordinates.

Scene Coordinate Regression

Regress 3D coordinates via neural network, trained as implicit scene representation.

Scene Coordinate Regression



ACE – Accelerated Coordinate Encoding [1]

- Rapid training (5 minutes), compact model (e.g. 4 MB)

GLACE – Global Local … [2]

- Feature diffícusion, position decoder → improved scalability & generalization

1. Brachmann E. et al. "Accelerated Coordinate Encoding: Learning to Relocalize in Minutes using RGB and Poses" CVPR 2023.
2. Wang F. et al. "GLACE: Global Local Accelerated Coordinate Encoding" CVPR 2024.

Method

Synthetic Data Generation

Create datasets from CAD models that are compatible with visual localization.

Supervised Training against Scene Coordinates

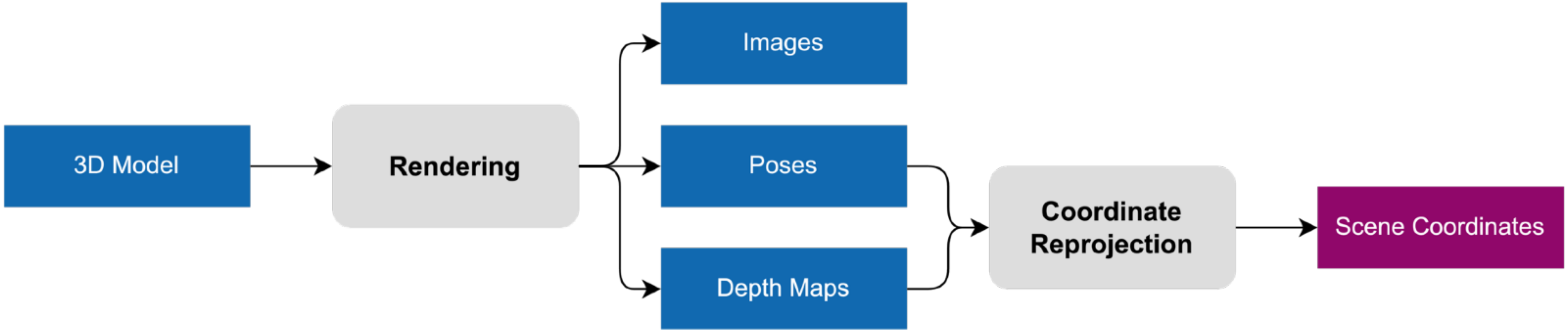
Make use of privileged 3D data to improve training efficiency.

Transfer Learning for Domain Adaptation

Bridge the domain gap between synthetic and real-world data.

Method – Synthetic Data Generation

Automatic Orbit Poses



Method – Synthetic Data Generation

Ground Truth Poses

1. Align reconstruction with CAD model
2. Transform pose between frames
3. Render image

Method

Synthetic Data Generation

Create datasets from CAD models that are compatible with visual localization.

Supervised Training against Scene Coordinates

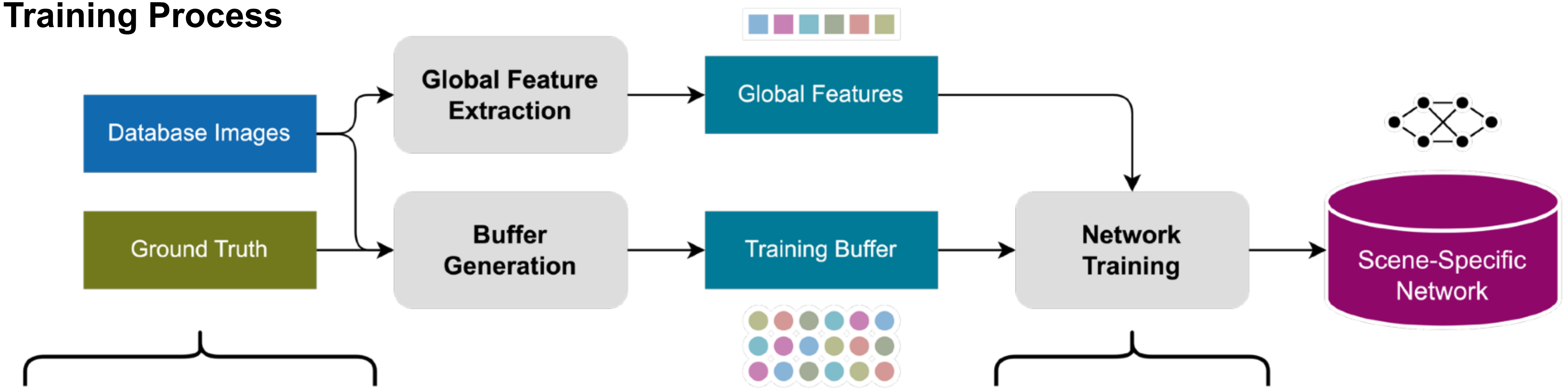
Make use of privileged 3D data to improve training efficiency.

Transfer Learning for Domain Adaptation

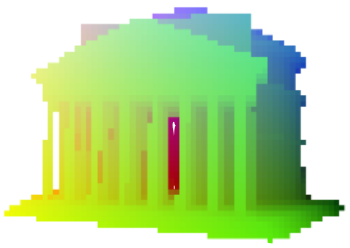
Bridge the domain gap between synthetic and real-world data.

Method – Supervised Training against Scene Coordinates

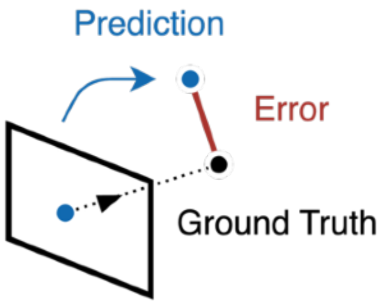
Training Process



Scene Coordinates



3D Distance Loss



Method

Synthetic Data Generation

Create datasets from CAD models that are compatible with visual localization.

Supervised Training against Scene Coordinates

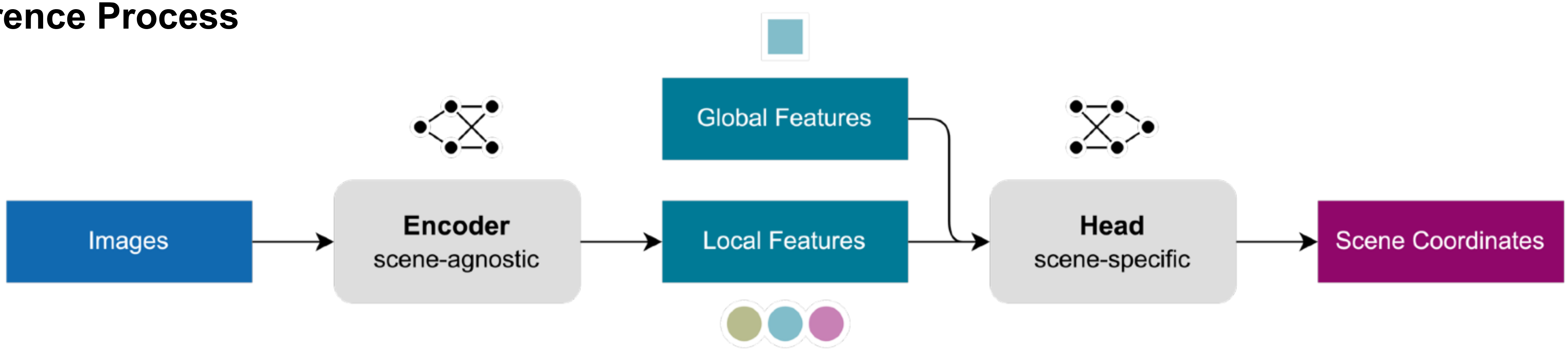
Make use of privileged 3D data to improve training efficiency.

Transfer Learning for Domain Adaptation

Bridge the domain gap between synthetic and real-world data.

Method – Transfer Learning for Domain Adaptation

Inference Process

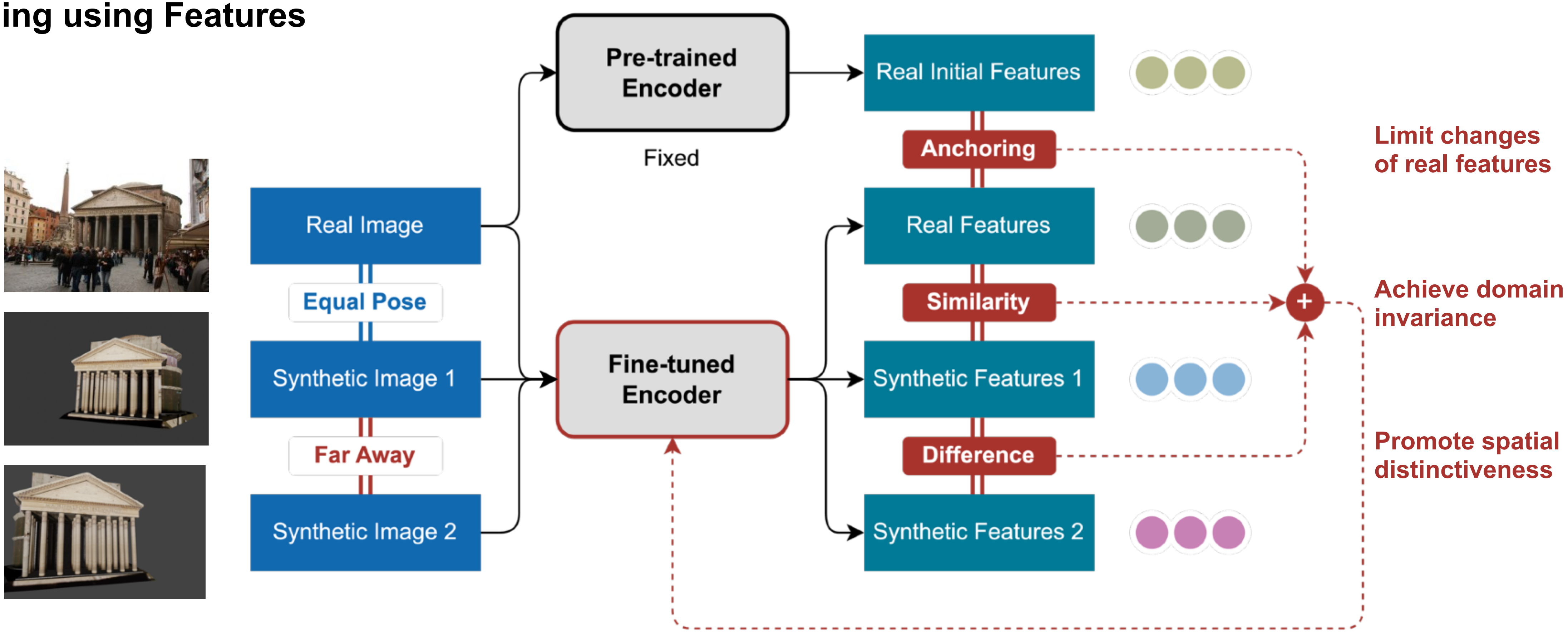


Encoder

- Pre-trained on real images
- Fine-tune for synthetic images
- Goal: invariance between real and synthetic features

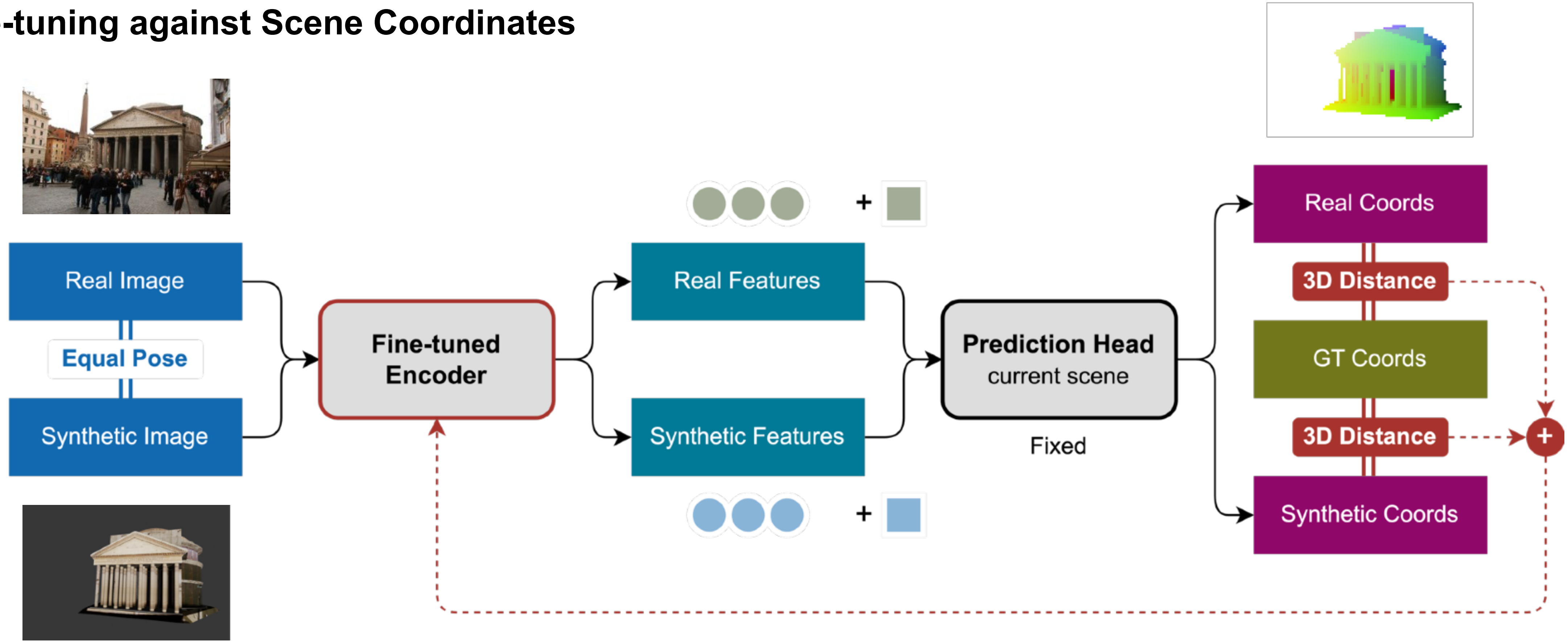
Method – Transfer Learning for Domain Adaptation

Fine-tuning using Features

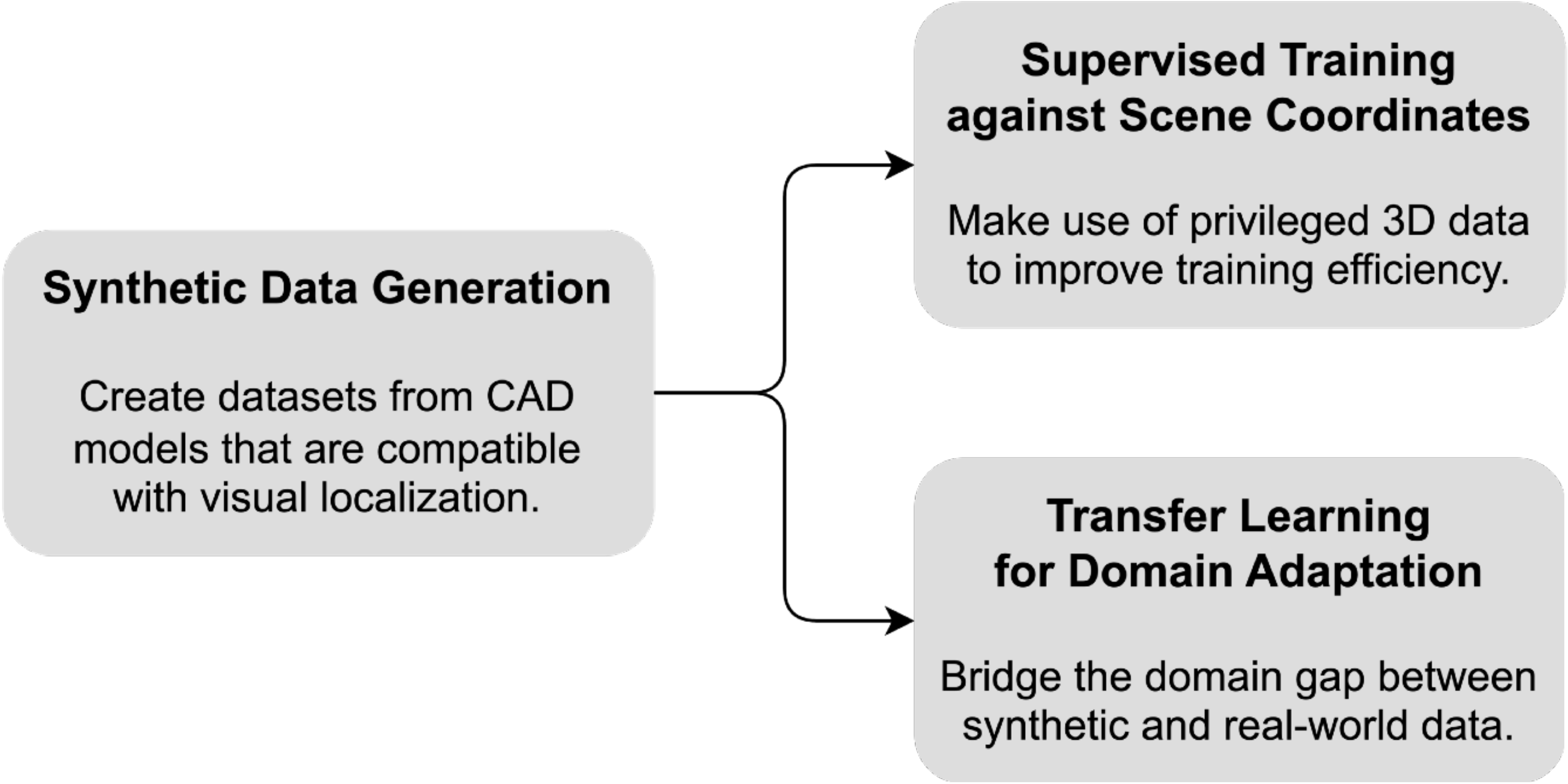


Method – Transfer Learning for Domain Adaptation

Fine-tuning against Scene Coordinates



Method


Synthetic Data Generation
Create datasets from CAD models that are compatible with visual localization.
Supervised Training against Scene Coordinates
Make use of privileged 3D data to improve training efficiency.
Transfer Learning for Domain Adaptation
Bridge the domain gap between synthetic and real-world data.

Results

Supervised Training against Scene Coordinates

Make use of privileged 3D data to improve training efficiency.

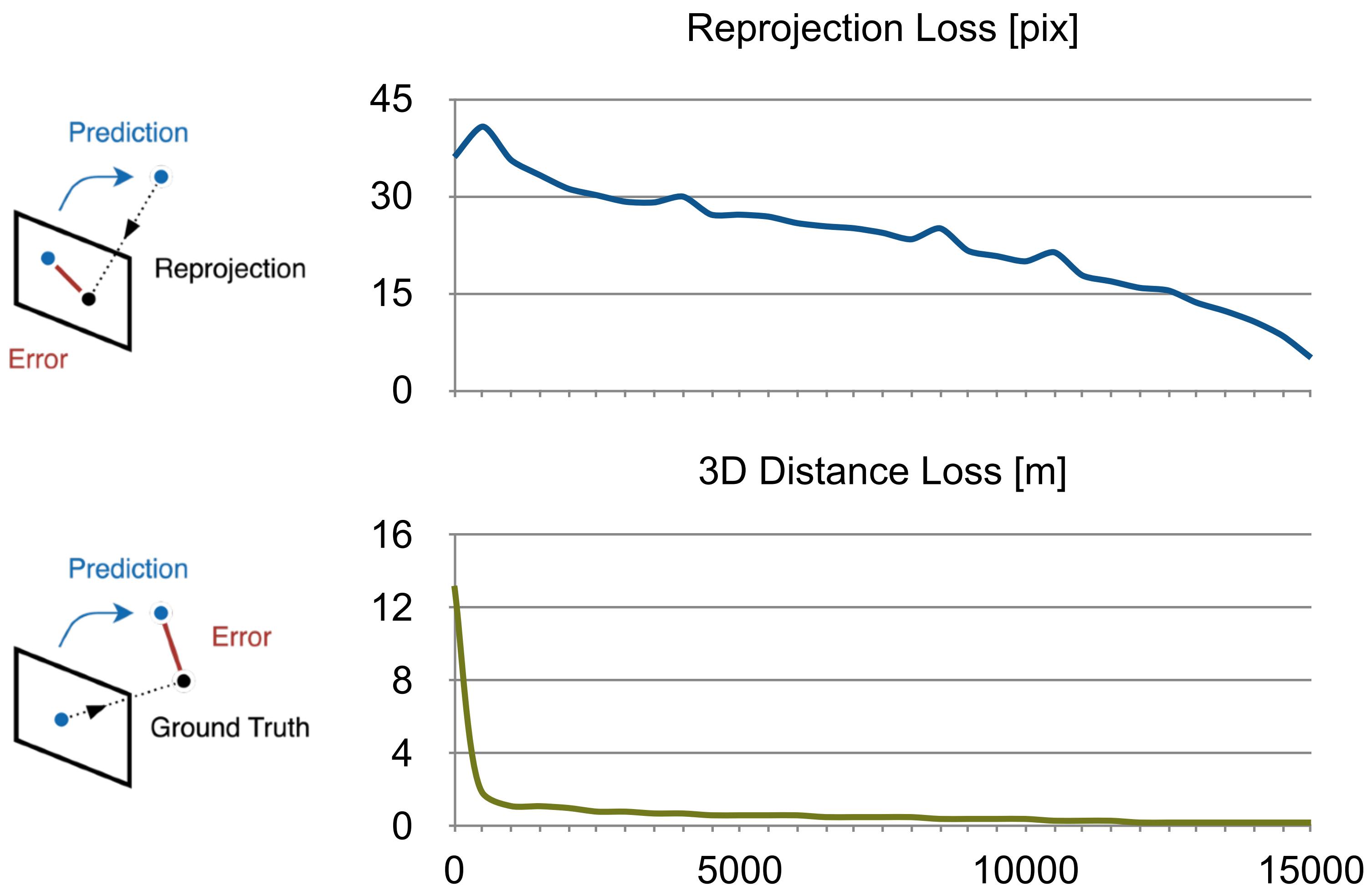
Transfer Learning for Domain Adaptation

Bridge the domain gap between synthetic and real-world data.

Results – Supervised Training against Scene Coordinates

Training Progression

- 3D loss converges faster
- Reprojection loss catches up



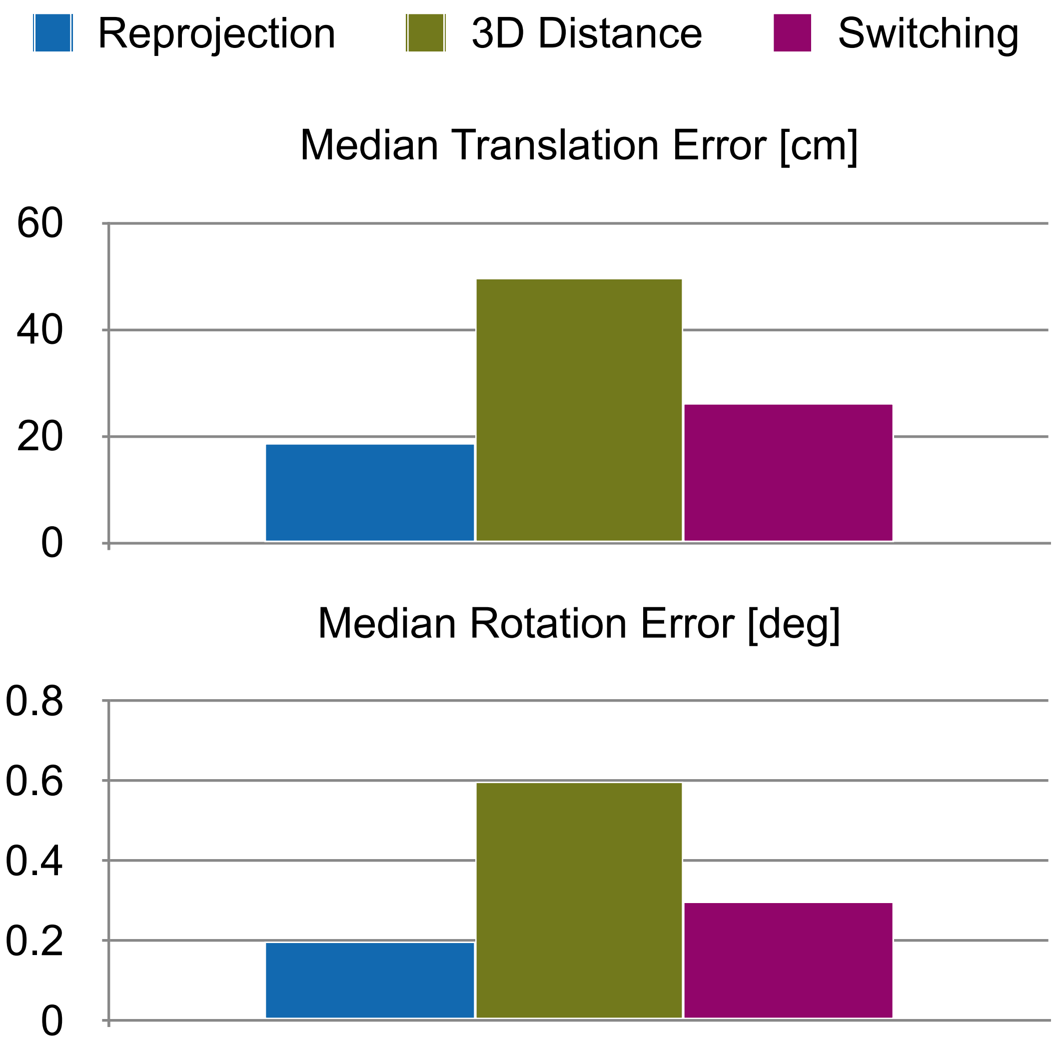
Results – Supervised Training against Scene Coordinates

Localization Accuracy

- **Reprojection** loss performs best
- **3D Distance** loss not as accurate, but still good
- **Switching** from 3D to reprojection loss at 1/3 iterations has limited benefit

Conclusion

- Faster convergence, but no improved accuracy
- Reprojection loss more advanced: ignores outliers



Results

Supervised Training against Scene Coordinates

Make use of privileged 3D data to improve training efficiency.

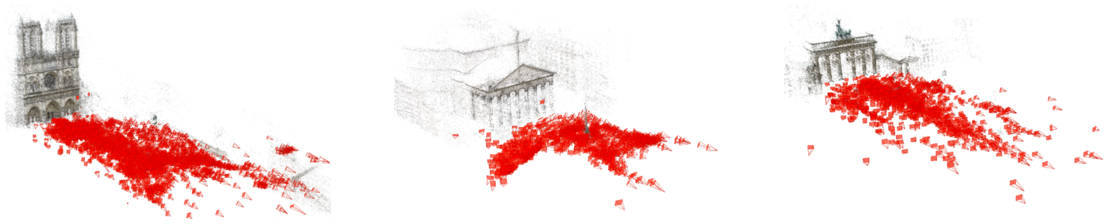
Transfer Learning for Domain Adaptation

Bridge the domain gap between synthetic and real-world data.

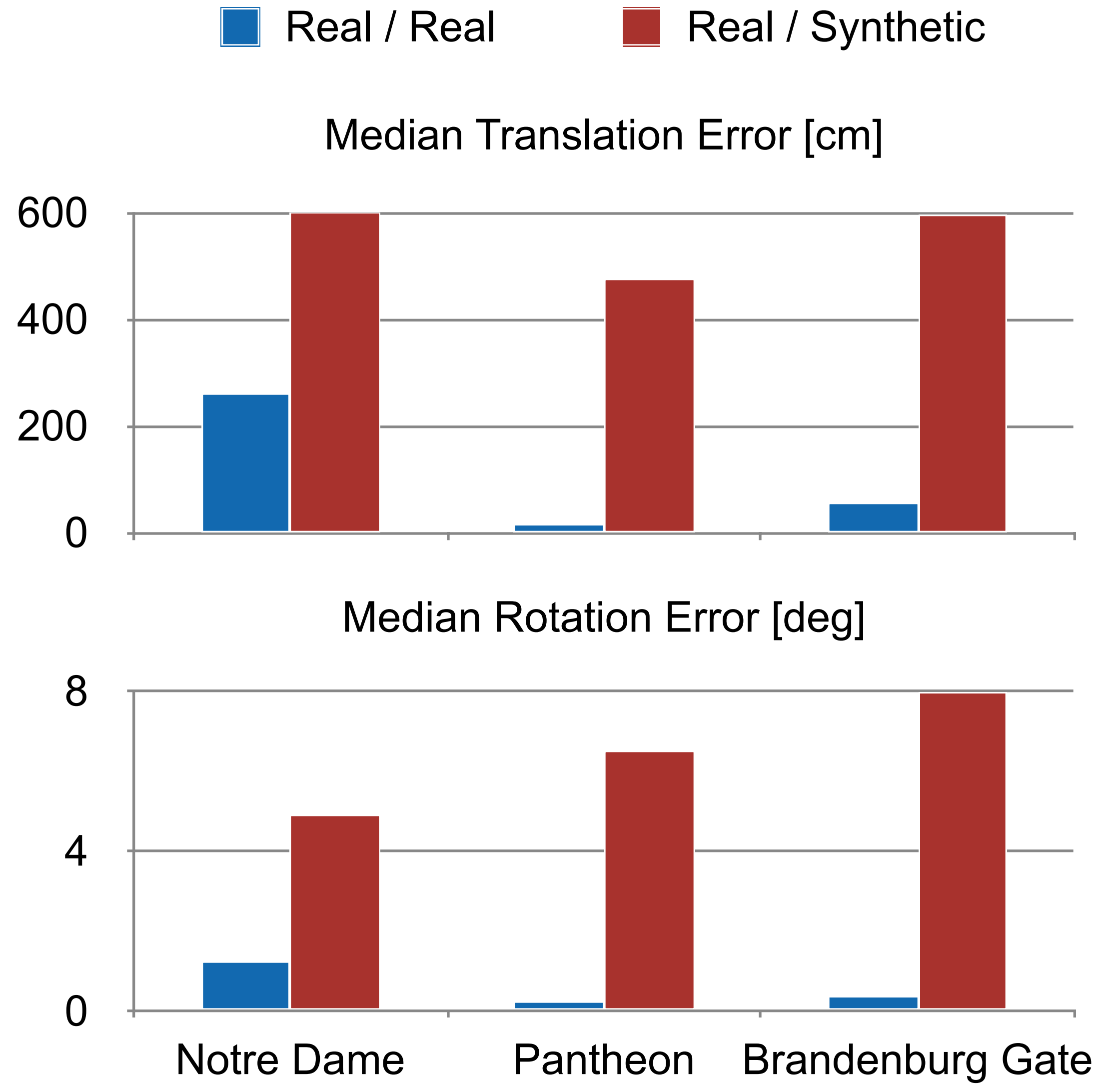
Results – Transfer Learning for Domain Adaptation

Original Localization Accuracy

- Training on real images, testing real vs. synthetic
- Large domain gap

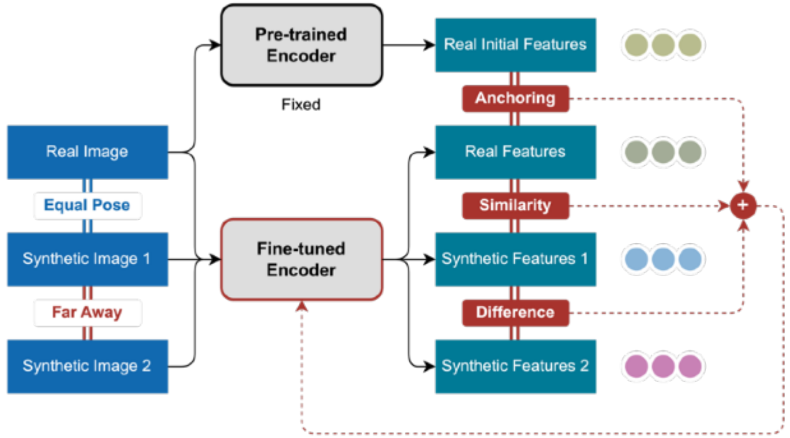


Notre Dame Pantheon Brandenburg Gate

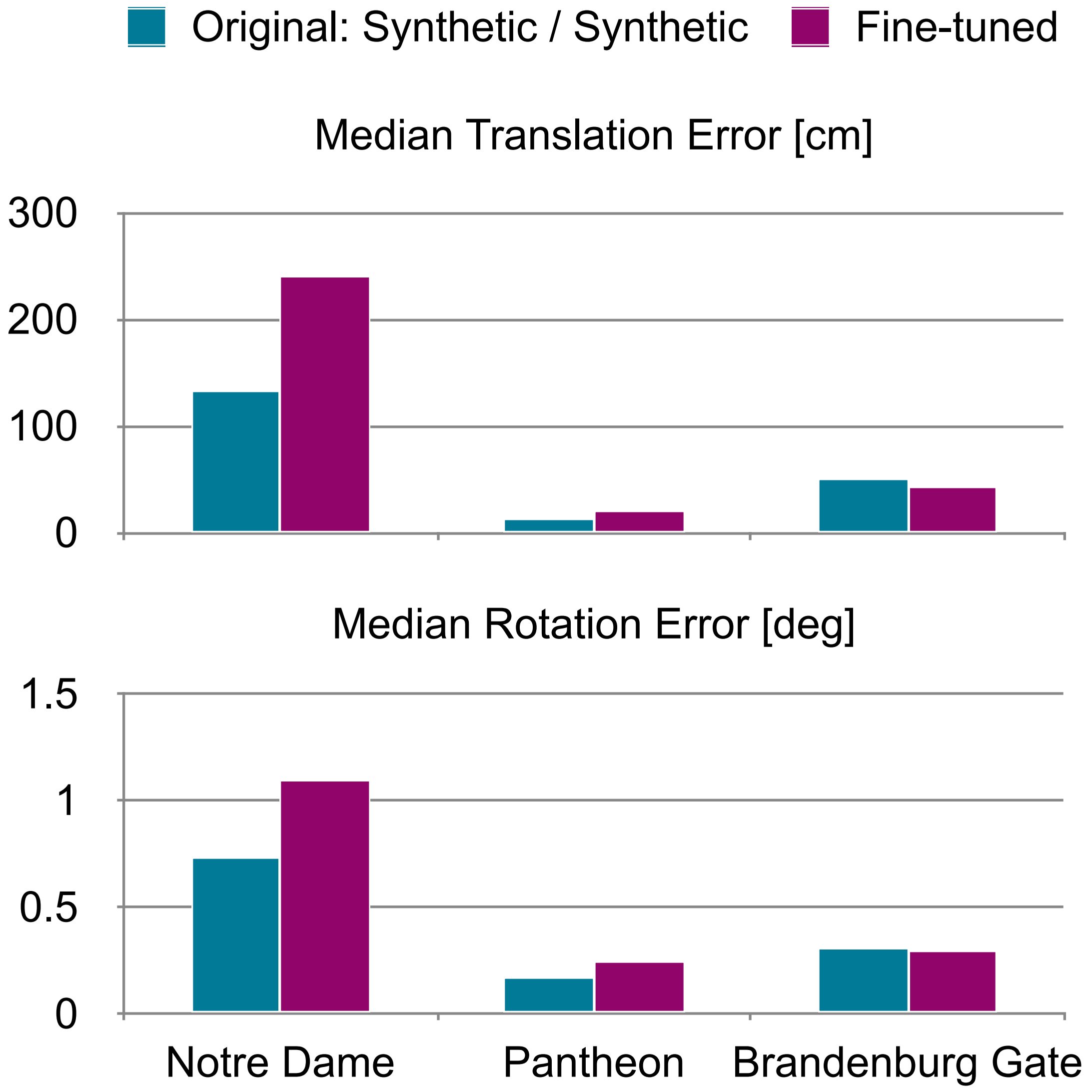


Results – Transfer Learning for Domain Adaptation

Fine-tuning using Features



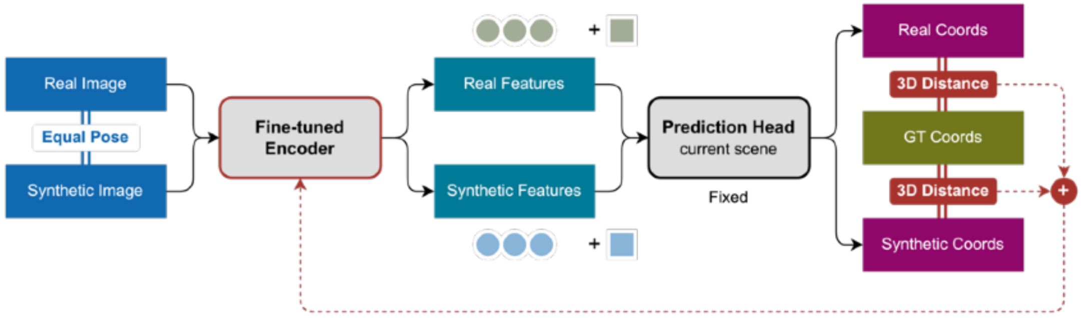
- Improvements in training and validation loss
- Not reflected in localization accuracy
- Features seem to lose usefulness



Results – Transfer Learning for Domain Adaptation

Fine-tuning against Scene Coordinates

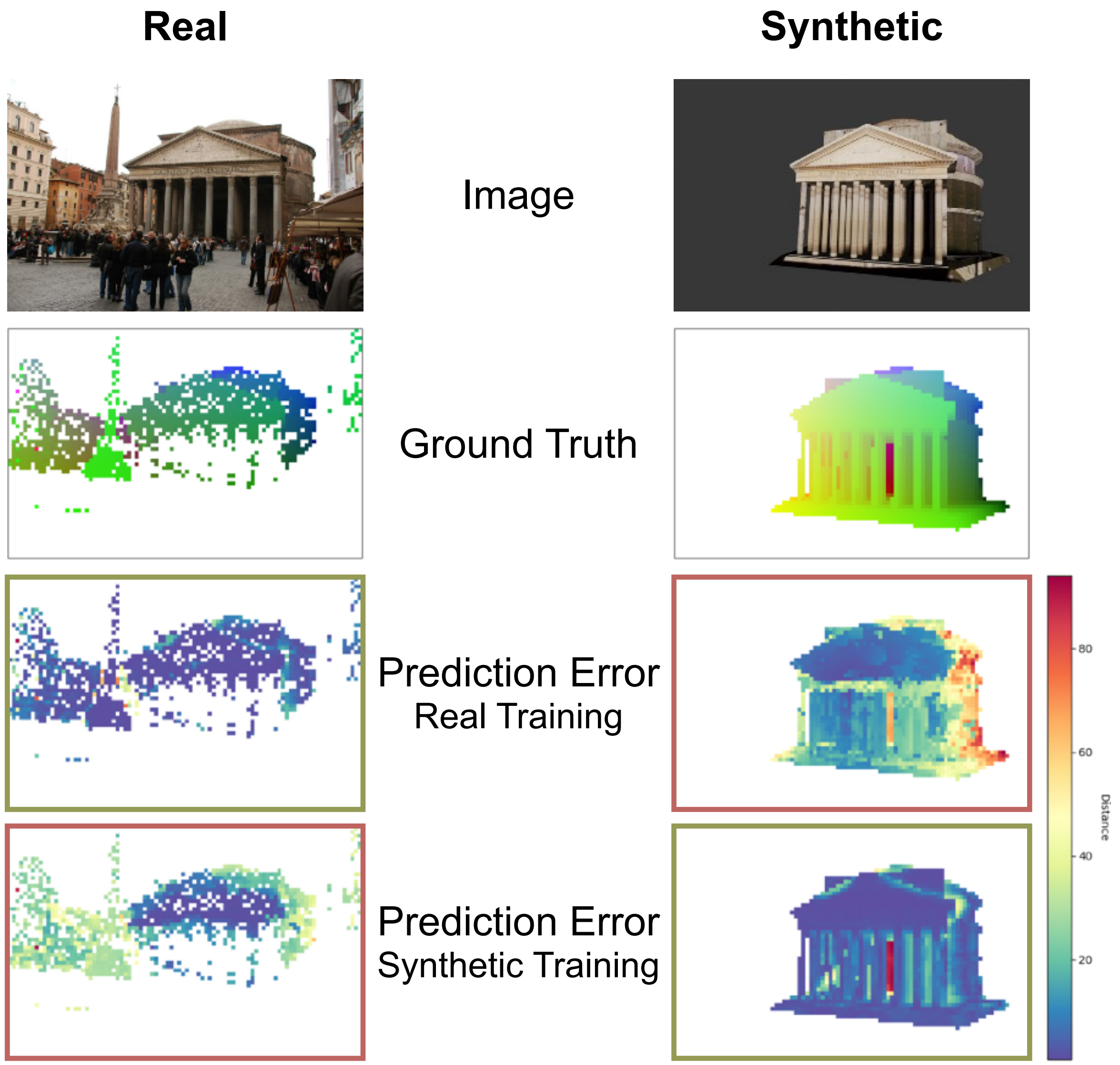
- Direct indicator of localization accuracy



- No consistent improvements
- Lack of useful information passed onto encoder

Conclusion

- Fine-tuning not enough?



Results

Supervised Training against Scene Coordinates

Make use of privileged 3D data to improve training efficiency.

Promising but does not outperform reprojection loss with current implementation.

Transfer Learning for Domain Adaptation

Bridge the domain gap between synthetic and real-world data.

No consistent improvements with different losses, suggesting that fine-tuning is insufficient.

Enhancements

- Dynamic thresholding for large distances
- Camera-based constraints: depth, visibility

- End-to-end learning of encoder + head
- Gradient decorrelation to find patterns
- More datasets, longer training times

Conclusion

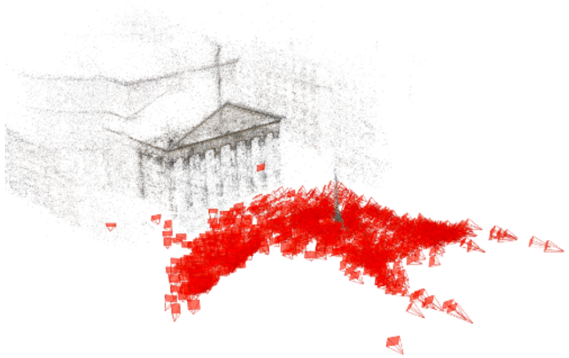
- Investigated adapting SCR to train on CAD models for real-world localization
- Goal: bridge real-synthetic domain gap

1. Developed synthetic data generation pipeline
2. Implemented supervised training
3. Experimented with transfer learning

- Identified several areas for future research
- Highlighted challenges & opportunities

- Complex problem, needs further investigation

Reconstruction



CAD Model

Synthetic Data Generation

Create datasets from CAD models that are compatible with visual localization.

Supervised Training against Scene Coordinates

Make use of privileged 3D data to improve training efficiency.

Transfer Learning for Domain Adaptation

Bridge the domain gap between synthetic and real-world data.