Landmarks Image Matching

Team 8: Hongyu Dai, Zihan Fang, Jingyuan Wu

01 INTRO

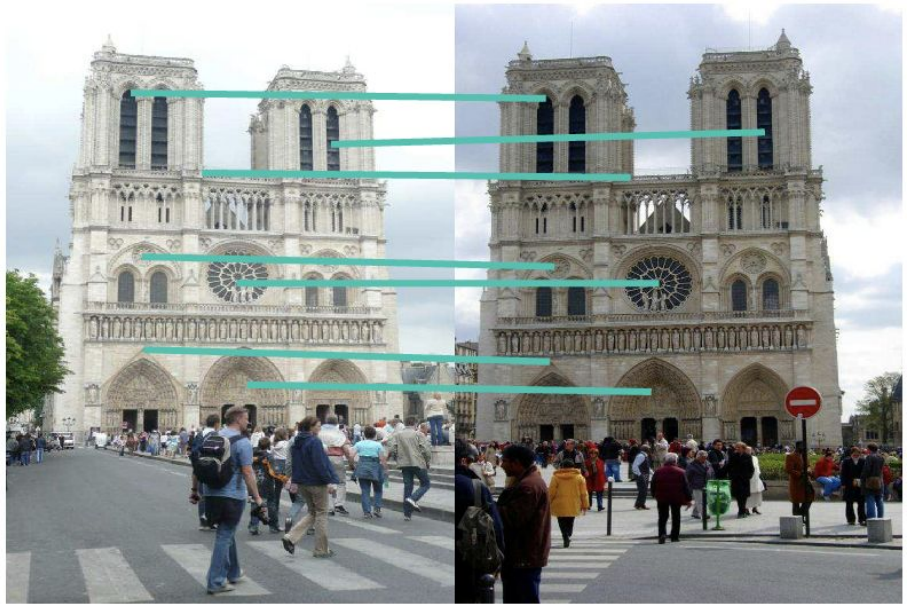
How to infer 3D models

from photos

OVERVIEW

Background

- The process to reconstruct 3D objects from images. The key point of the problem is how to identify the pair co-visibility in two images by local features. The success of this project will make it possible to map the world using unstructured collections.



Method

- LoFTR, SuperGlue, DKM

02 DATA OVERVIEW

Data and EDA

DATA OVERVIEW - INPUT

- There are 16 scenes with 5720 photos in total.
- For each scene:
  - **images folder** - images of that scene from different angles
  - **calibration.csv** - image_id, camera_intrinsics, rotation_matrix, translation_vector
  - **pair_covisibility.csv** - pair, covisibility, fundamental_matrix

DATA OVERVIEW

Brandenburg Gate
1075 x 787 px

British Museum
1041 x 686 px

Buckingham Palace
1077 x 583 px

Colosseum Exterior
1046 x 776 px

Grand Place Brussels
1072 x 786 px

Lincoln Memorial Statue
789 x 1080 px

Notre Dame Front Facade
1038 x 723 px

Pantheon Exterior
764 x 1025 px

Piazza San Marco
1045 x 776 px

Sacre Coeur
768 x 1025 px

Sagrada Familia
688 x 1053 px

St Pauls Cathedral
1077 x 696 px

St Peters Square
1039 x 688 px

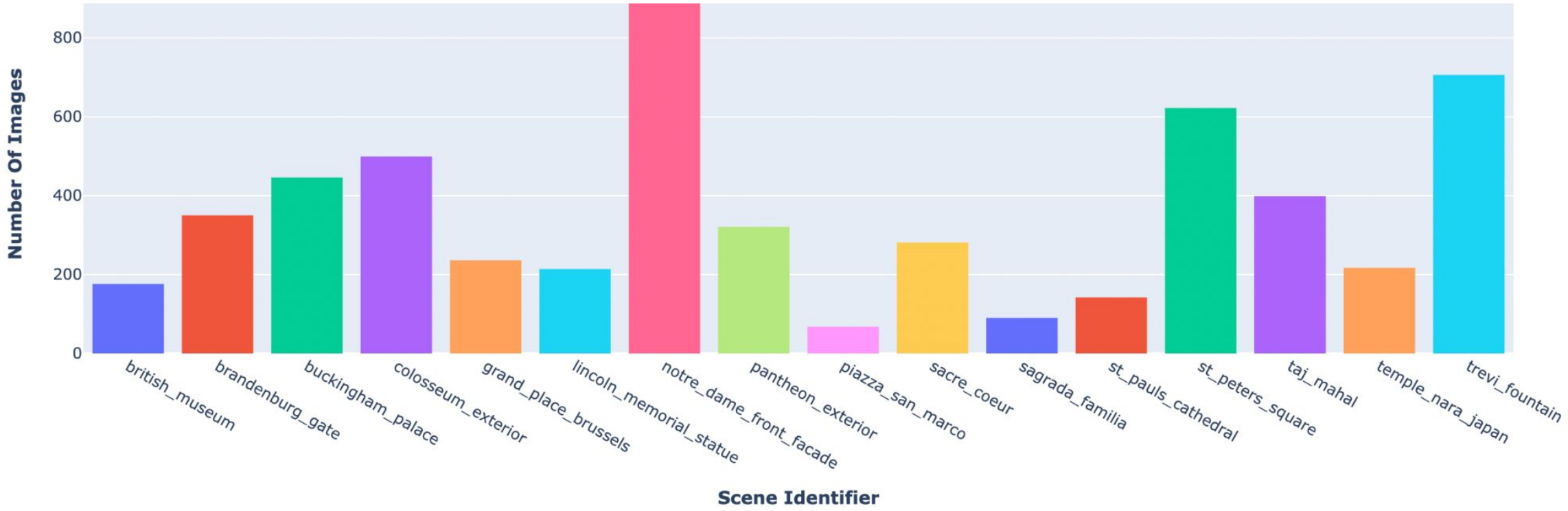
Taj Mahal
1001 x 750 px

Temple Nara Japan
1044 x 686 px

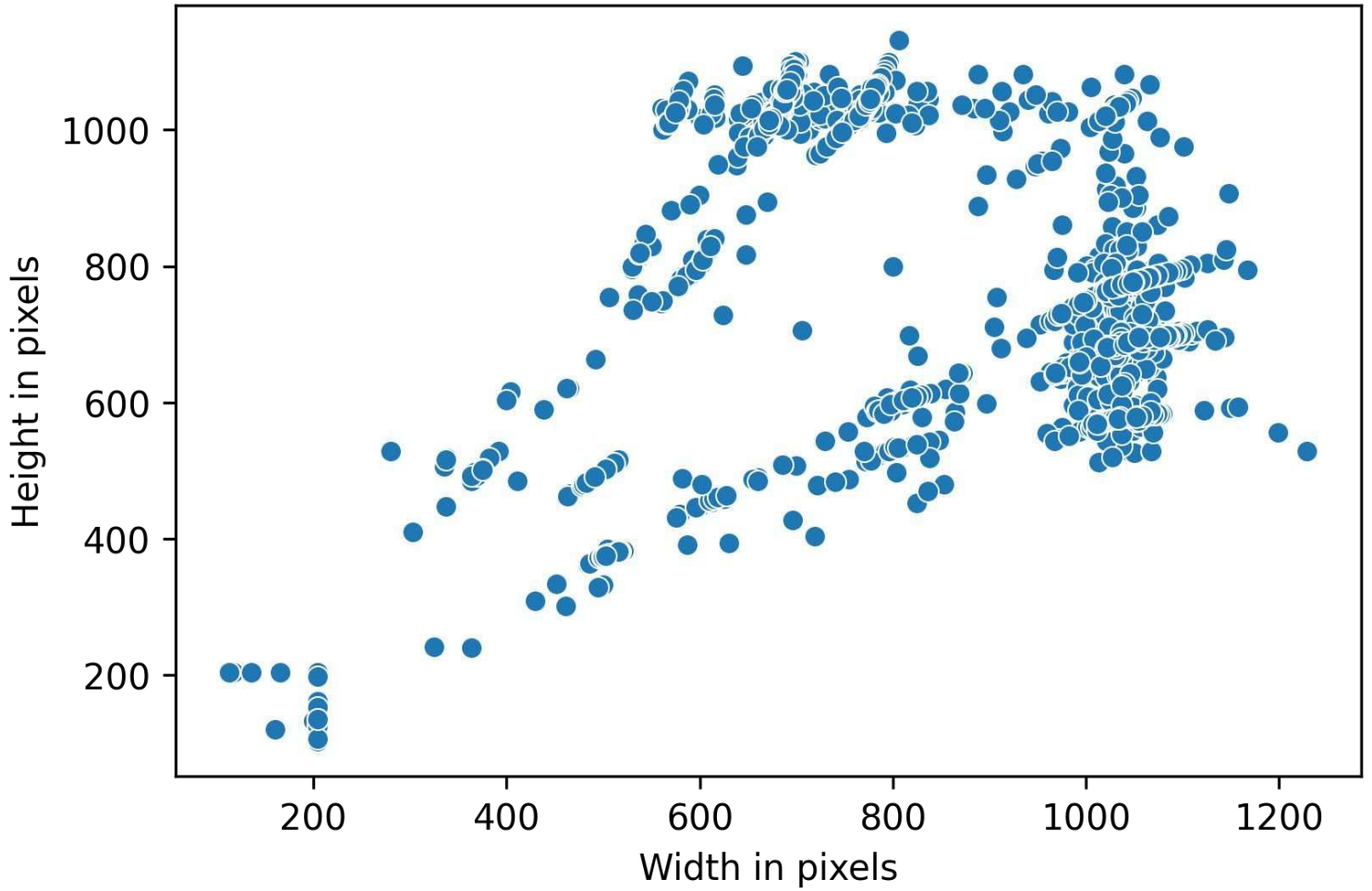
Trevi Fountain
1022 x 684 px

DATA OVERVIEW-EDA

- 5720 images, distributed unevenly over 16 scenes



DATA OVERVIEW - EDA



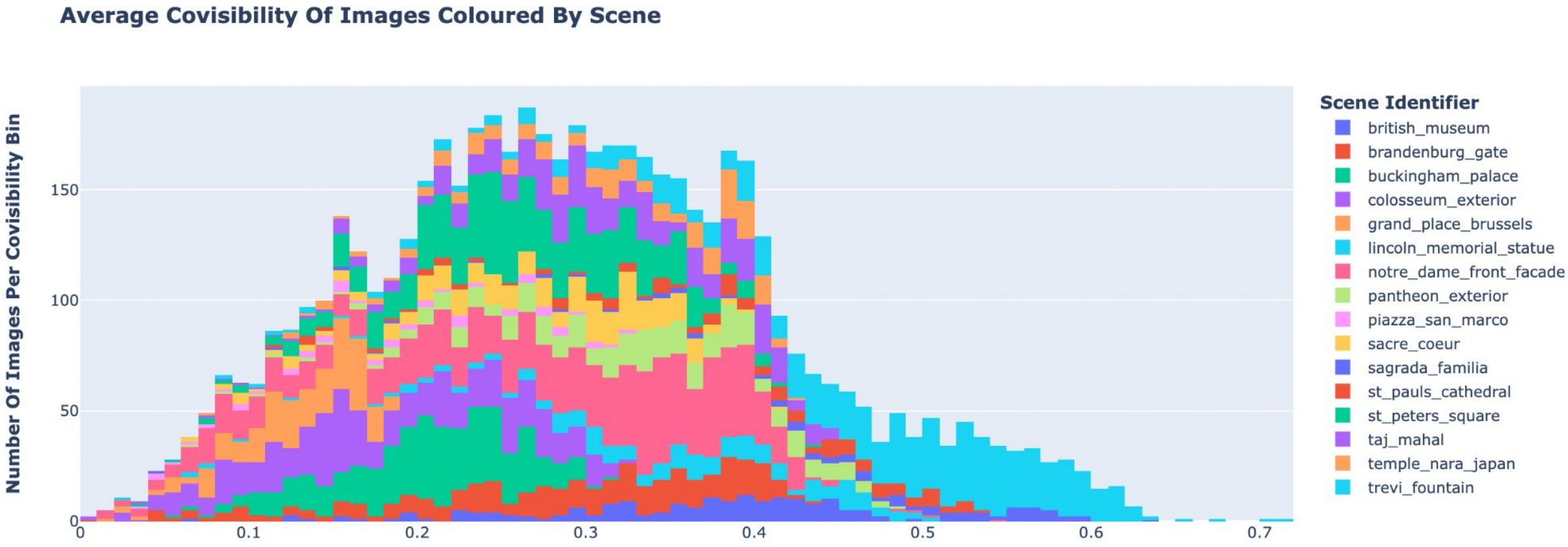
DATA OVERVIEW - IMAGE PAIRS


Bad covisibility (0.007)


Good covisibility (0.713)

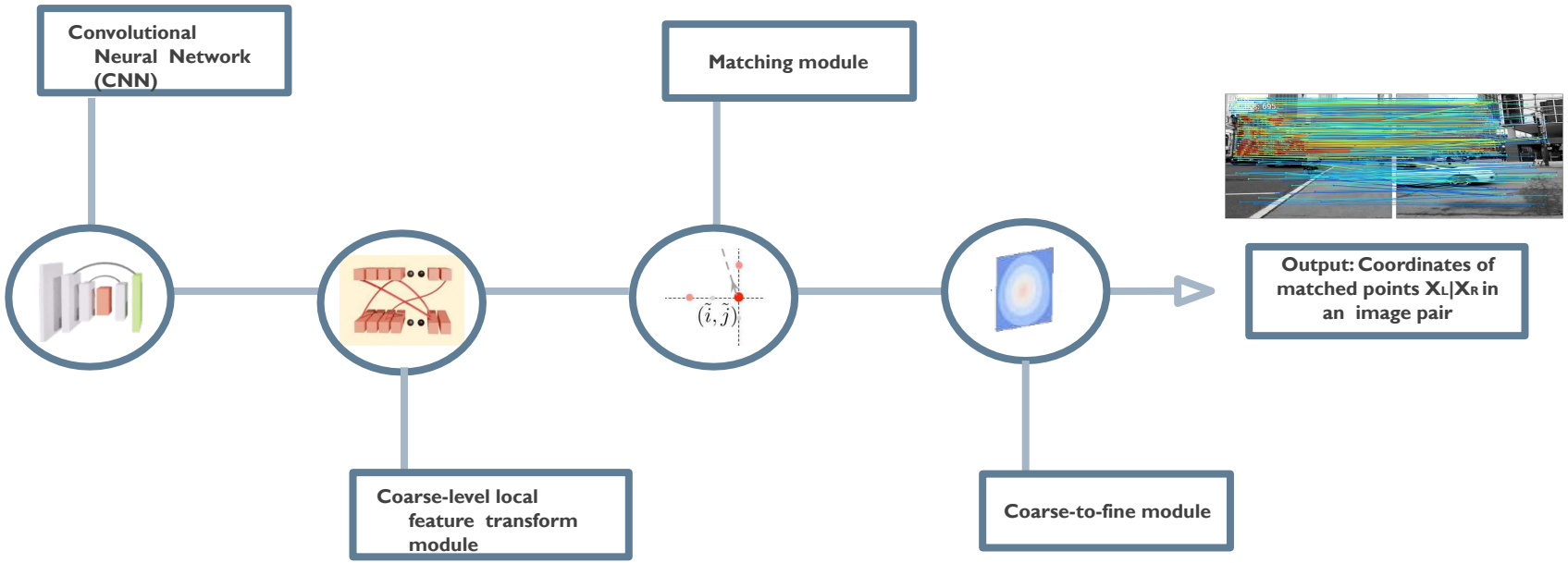
Covisibility matrix is provided by kaggle dataset.

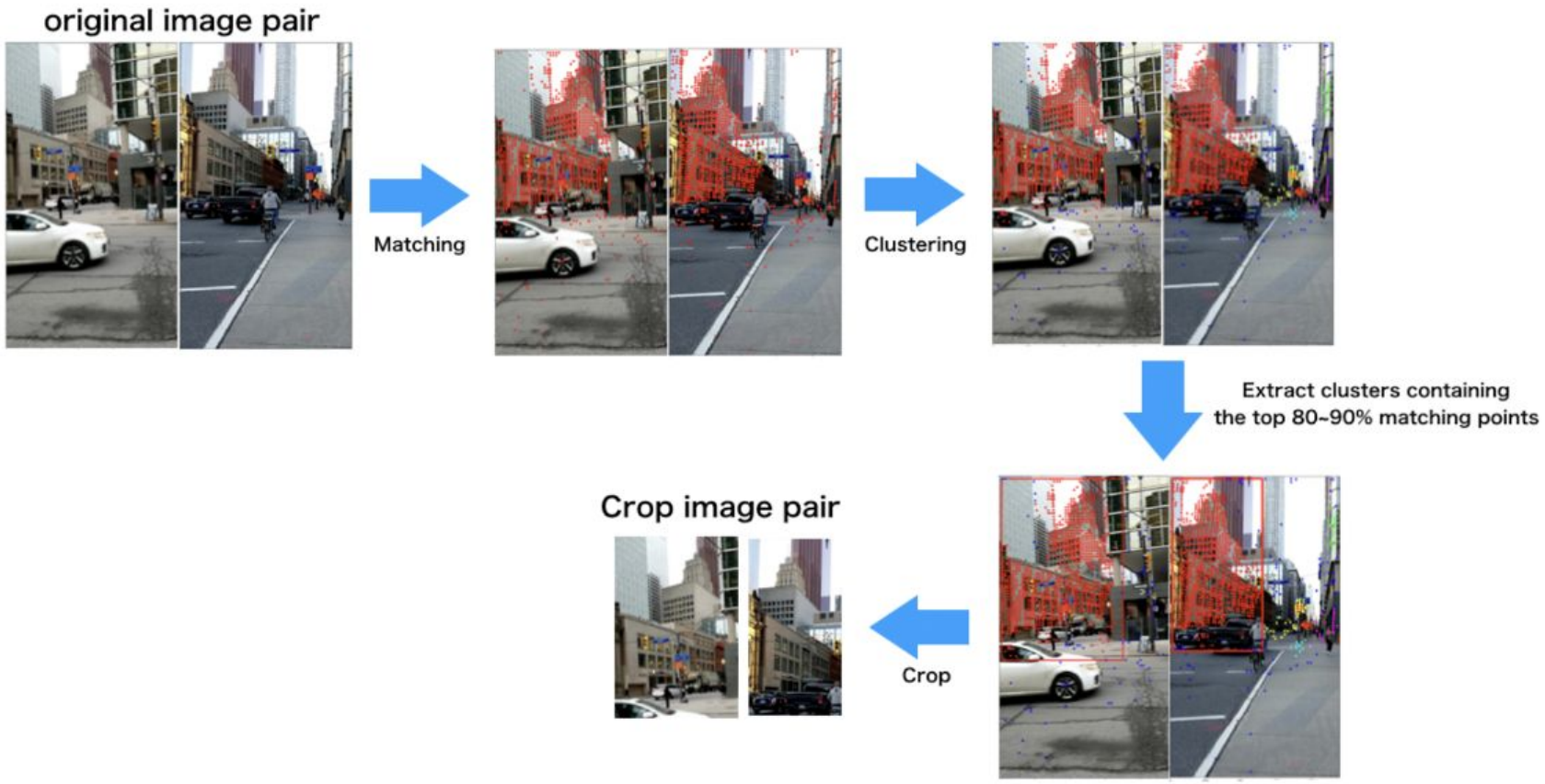
DATA OVERVIEW - EDA



03 MODELING

Baseline Model - LoFTR

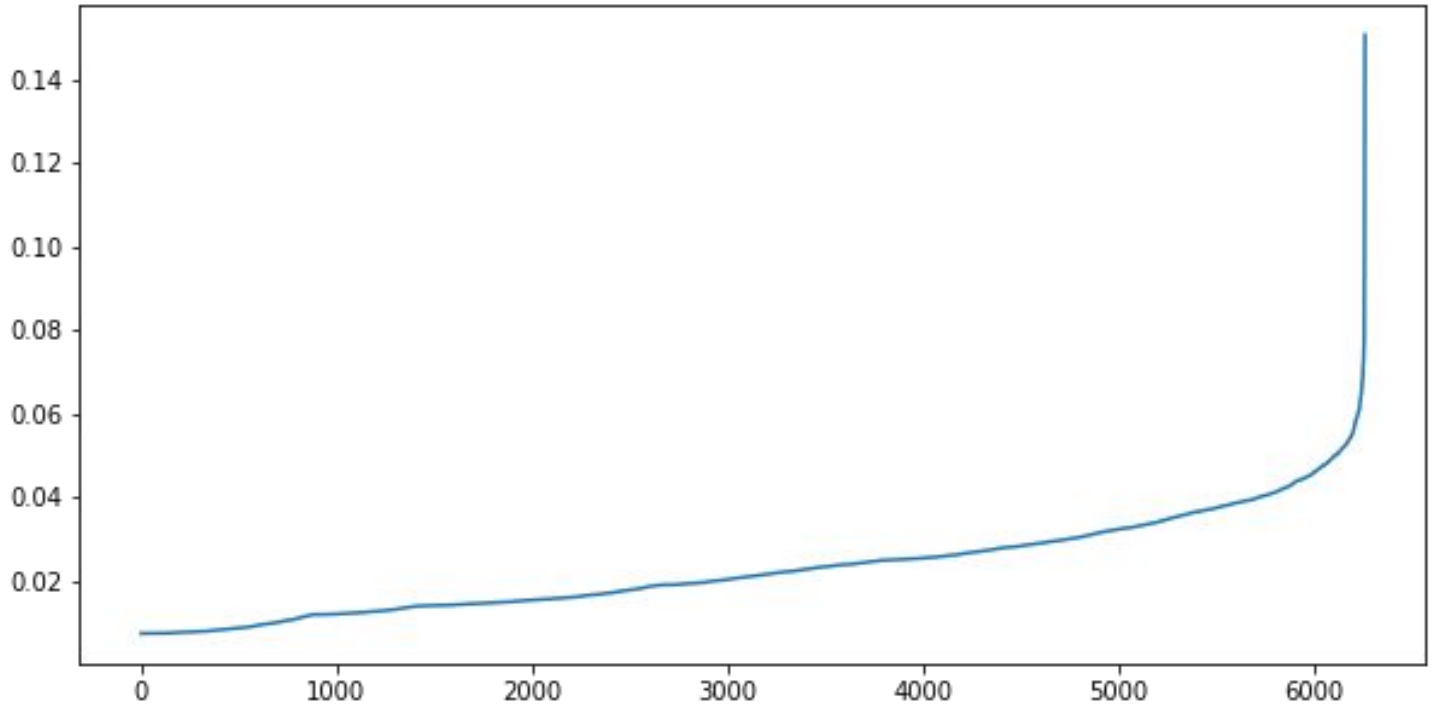




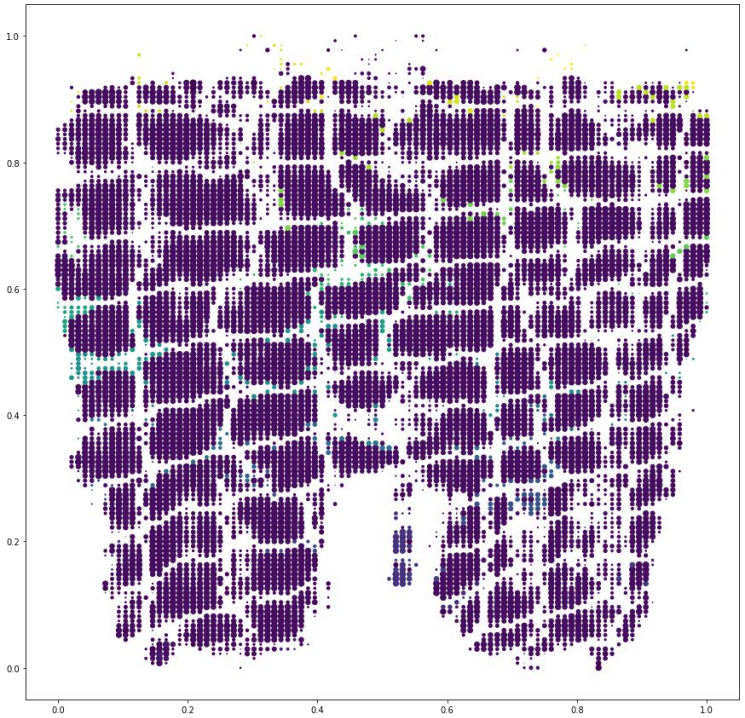
Baseline Model - LoFTR

Remove Outlier Using DBSCAN

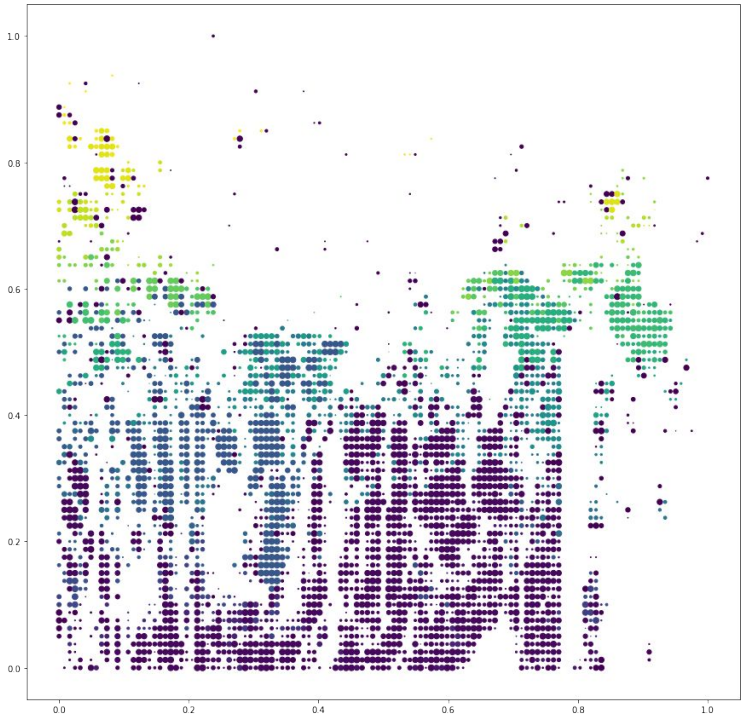
- epsilon: 0.05-0.06



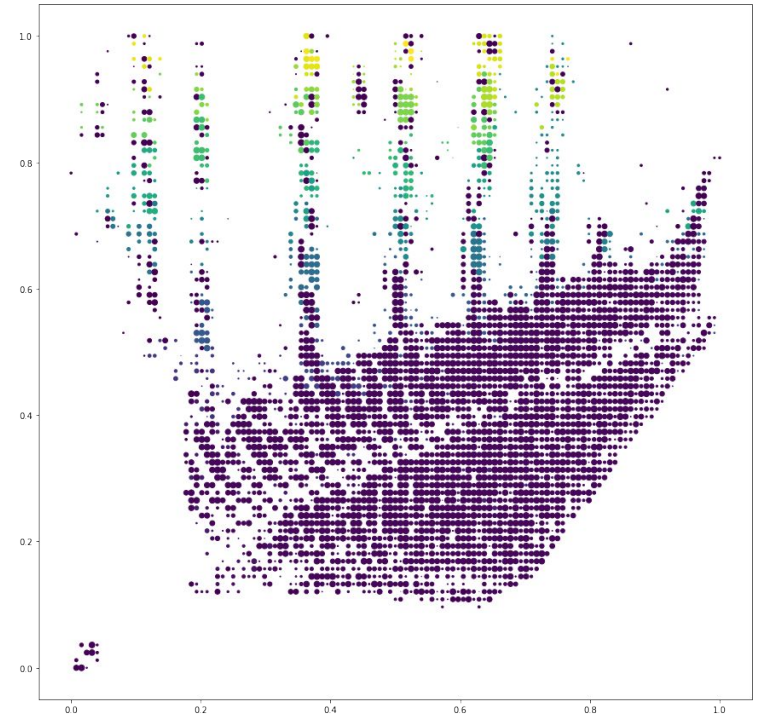
DBSCAN?



Notre Dame Front Facade



Trevi Fountain



Pantheon Exterior

DBSCAN?

Accuracy = 0.9
Angle error (degrees) = 1.34
Distance error (meters) = 0.026

Accuracy = 1.0
Angle error (degrees) = 0.44
Distance error (meters) = 0.12

Notre Dame Front Facade

Accuracy = 0.8
Angle error (degrees) = 1.26
Distance error (meters) = 0.31

Accuracy = 0.2
Angle error (degrees) = 8.31
Distance error (meters) = 1.20

Trevi Fountain

Accuracy = 1.0
Angle error (degrees) = 0.81
Distance error (meters) = 0.14

Accuracy = 0.8
Angle error (degrees) = 2.57
Distance error (meters) = 0.31

Pantheon Exterior

Final: Model Ensembling

Model:

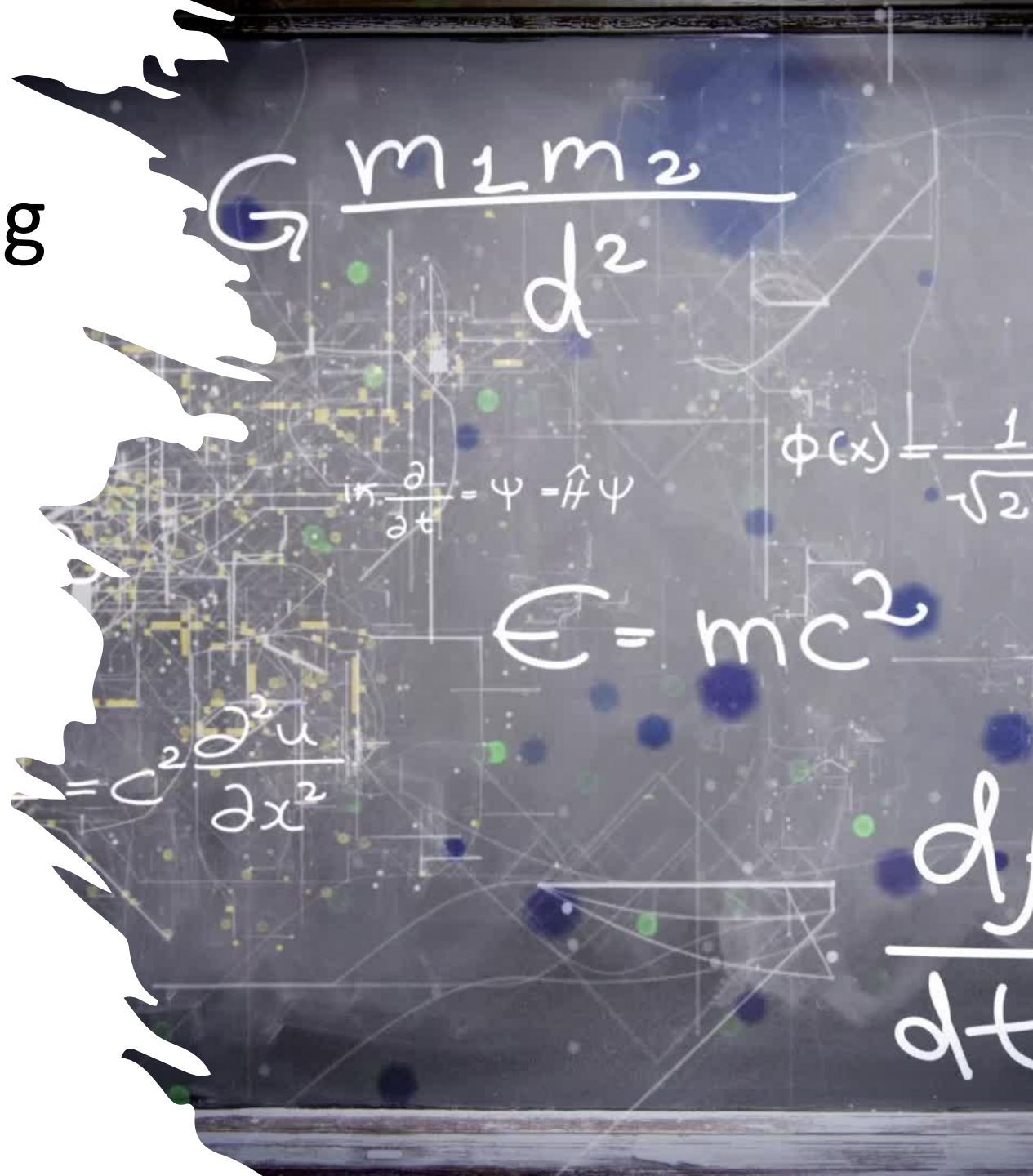
LoFTR, SuperGlue, DKM

Ensembling approach

1. Get the keypoints from all the models.
2. Concatenate them.
3. Get output (fundamental matrix)

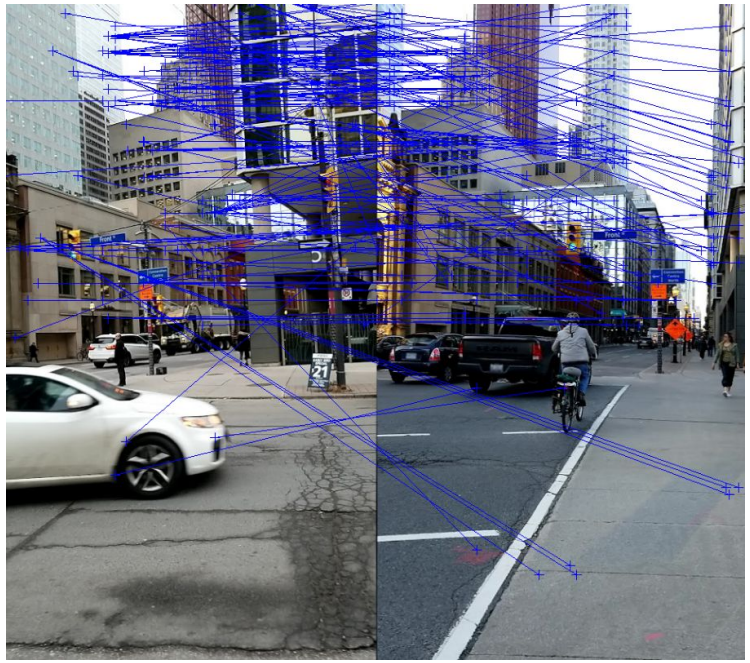
Submit to kaggle for metric calculation (mean average accuracy)

Plot match checkpoints



Final Model - LoFTR + DKM + SuperGlue

Baseline



mean Average Accuracy:

73.38%

Optimized

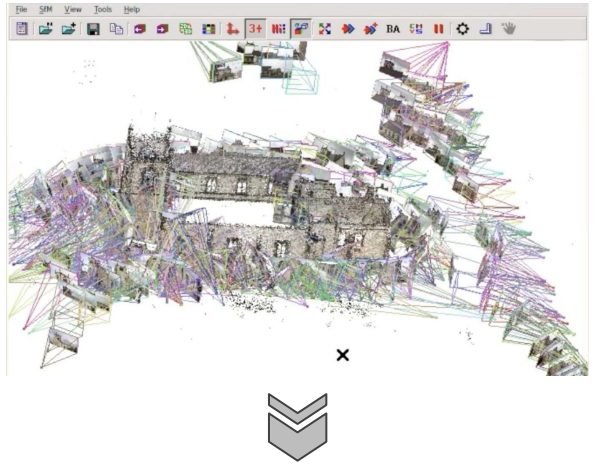
mean Average Accuracy:

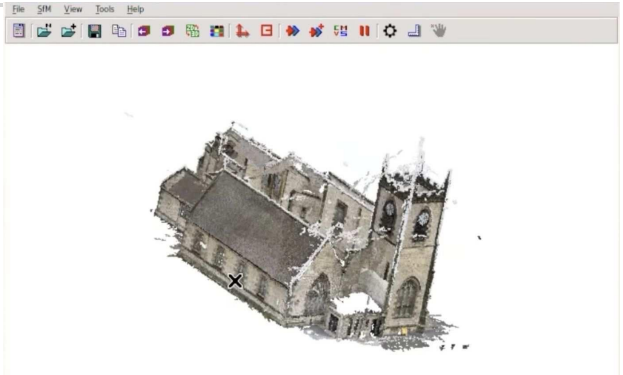
85.35%

Learning & Challenges

- General idea of image processing
- Image sizes / resize
- Keypoints selection

FUTURE WORK





VisualSFM

1. OPTIMIZATION
   - Try different keypoint selection method, like RANSAC

2. DASHBOARD
   - Images can be selected based on position and paires
   - Show image point pairings as basis and predicted camera positions

3. 3D RECONSTRUCTION
   - Existing 3D reconstruction tools, like VisualSFM, could be improved by implementing LoFTR

Thanks!