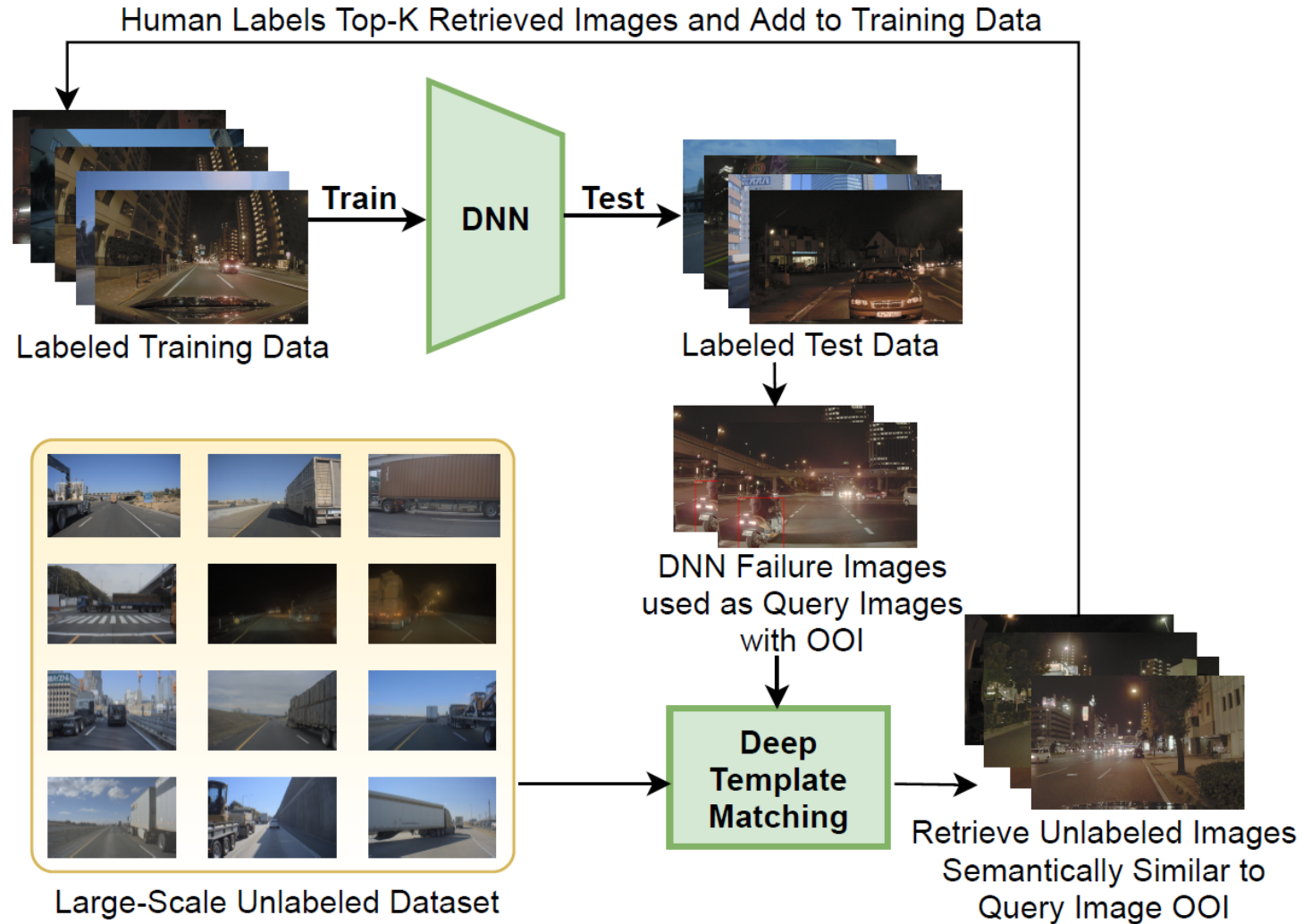# Object-Level Targeted Selection via Deep Template Matching

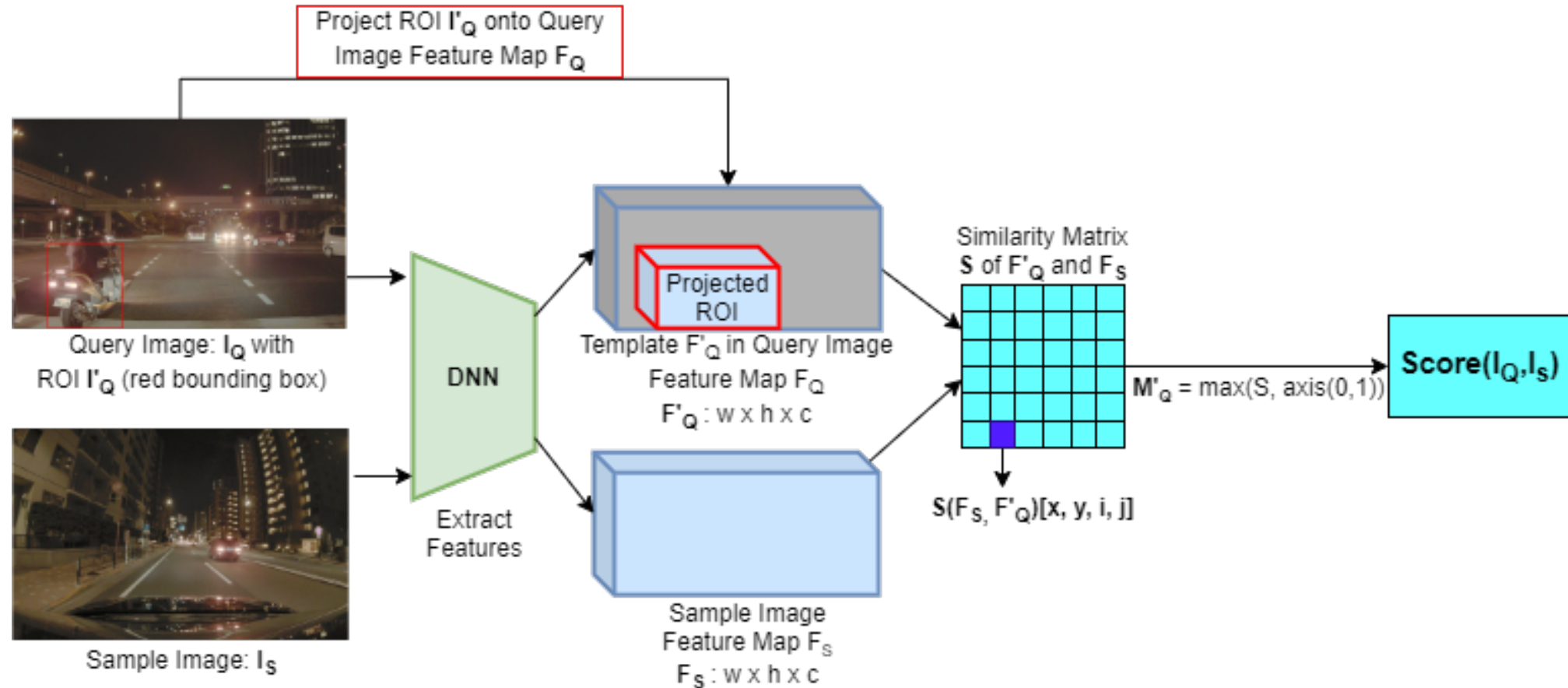Suraj Kothawade*, Donna Roy, Michele Fenzi, Elmar Haussmann, Jose M. Alvarez, Christoph Angerer

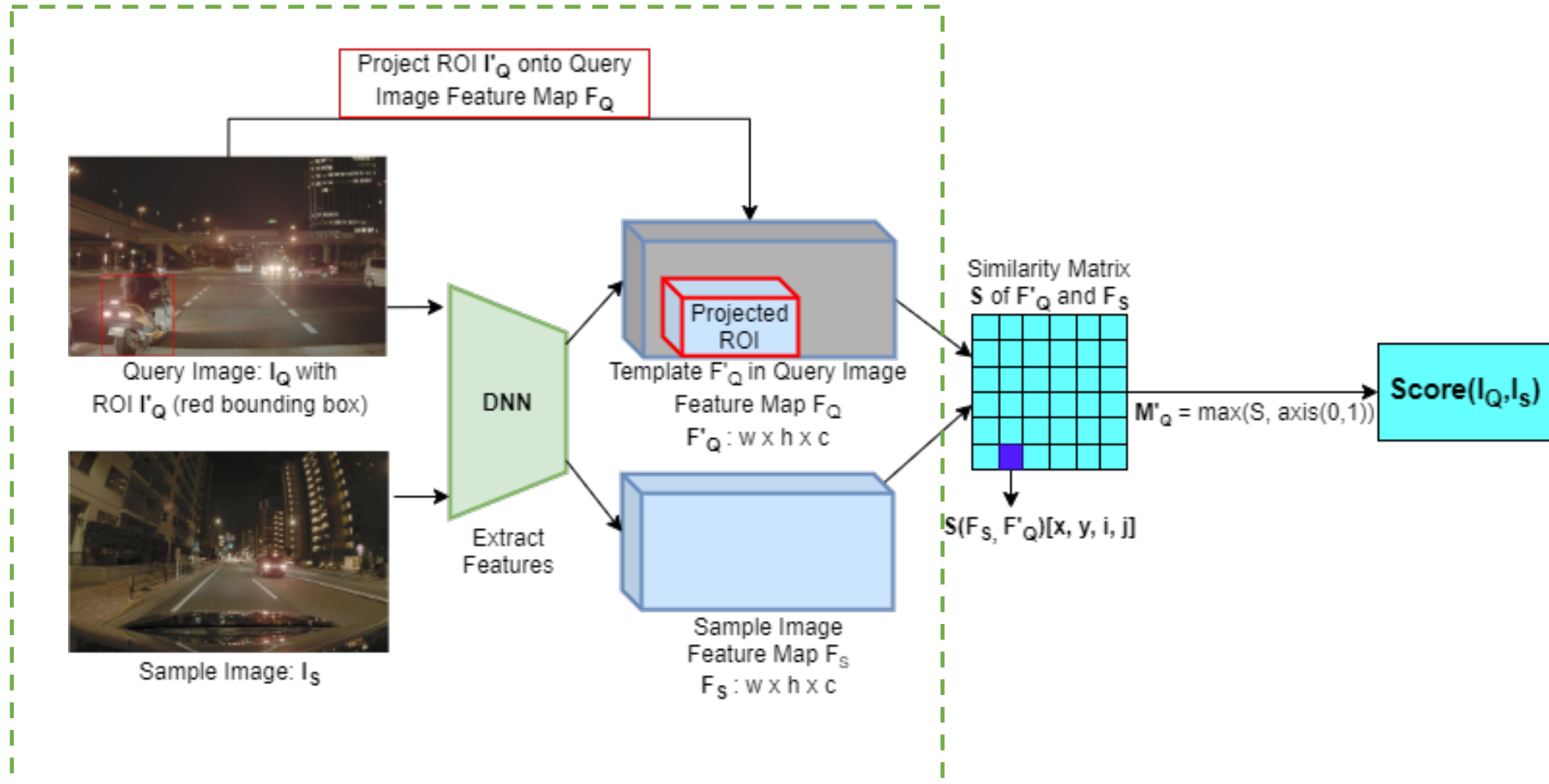*suraj.kothawade@utdallas.edu

NeurIPS 2021

Object Level Targeted Selection

Human Labels Top-K Retrieved Images and Add to Training Data

Labeled Training Data

DNN

Train

Test

Labeled Test Data

DNN Failure Images used as Query Images with OOI

Large-Scale Unlabeled Dataset

Deep Template Matching

Retrieve Unlabeled Images Semantically Similar to Query Image OOI

Deep Template Matching (DTM)

Project ROI I'_Q onto Query Image Feature Map F_Q

Query Image: I_Q with ROI I'_Q (red bounding box)

DNN

Extract Features

Projected ROI

Template F'_Q in Query Image Feature Map F_Q

F'_Q : w x h x c

Sample Image: I_S

Sample Image Feature Map F_S

F_S : w x h x c

Similarity Matrix S of F'_Q and F_S

$S(F_S, F'_Q)[x, y, i, j]$

$M'_Q = \max(S, axis(0,1))$

Score(I_Q, I_S)
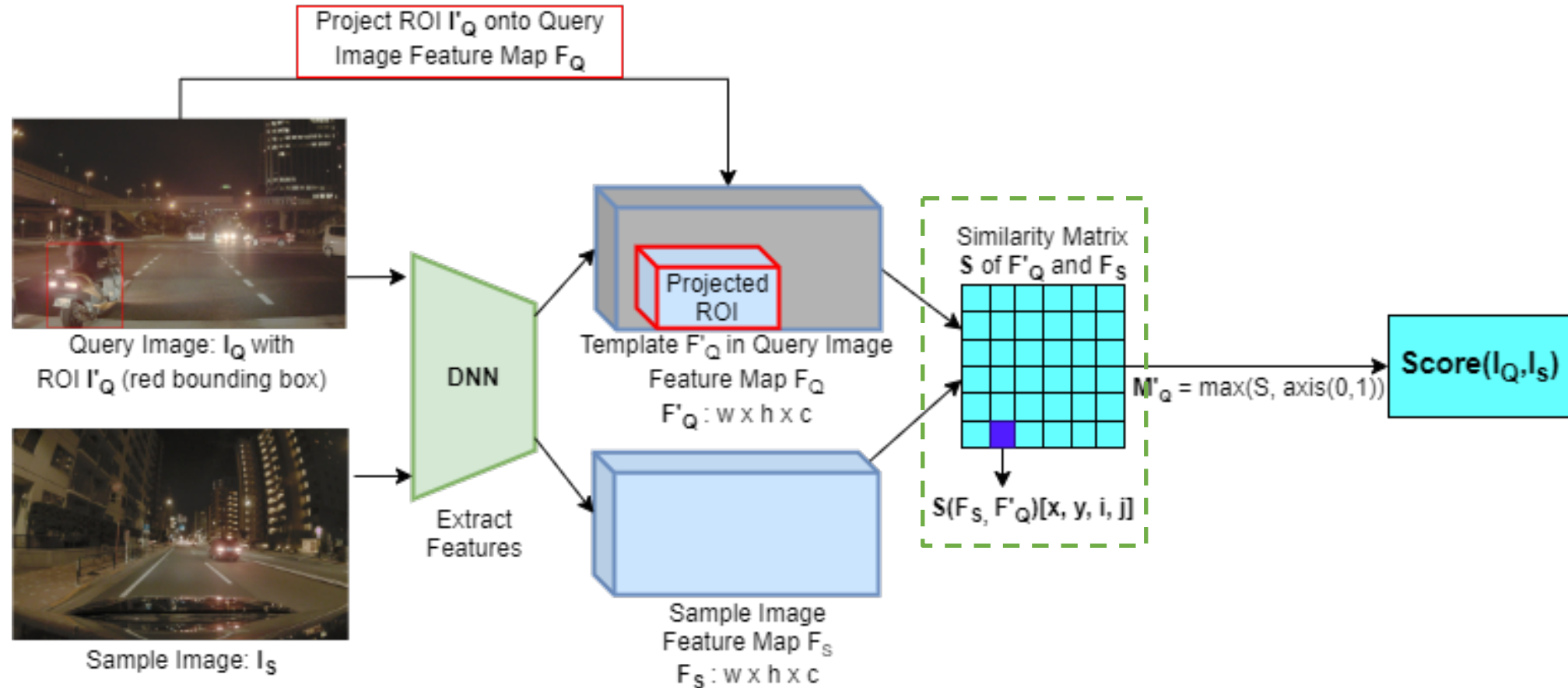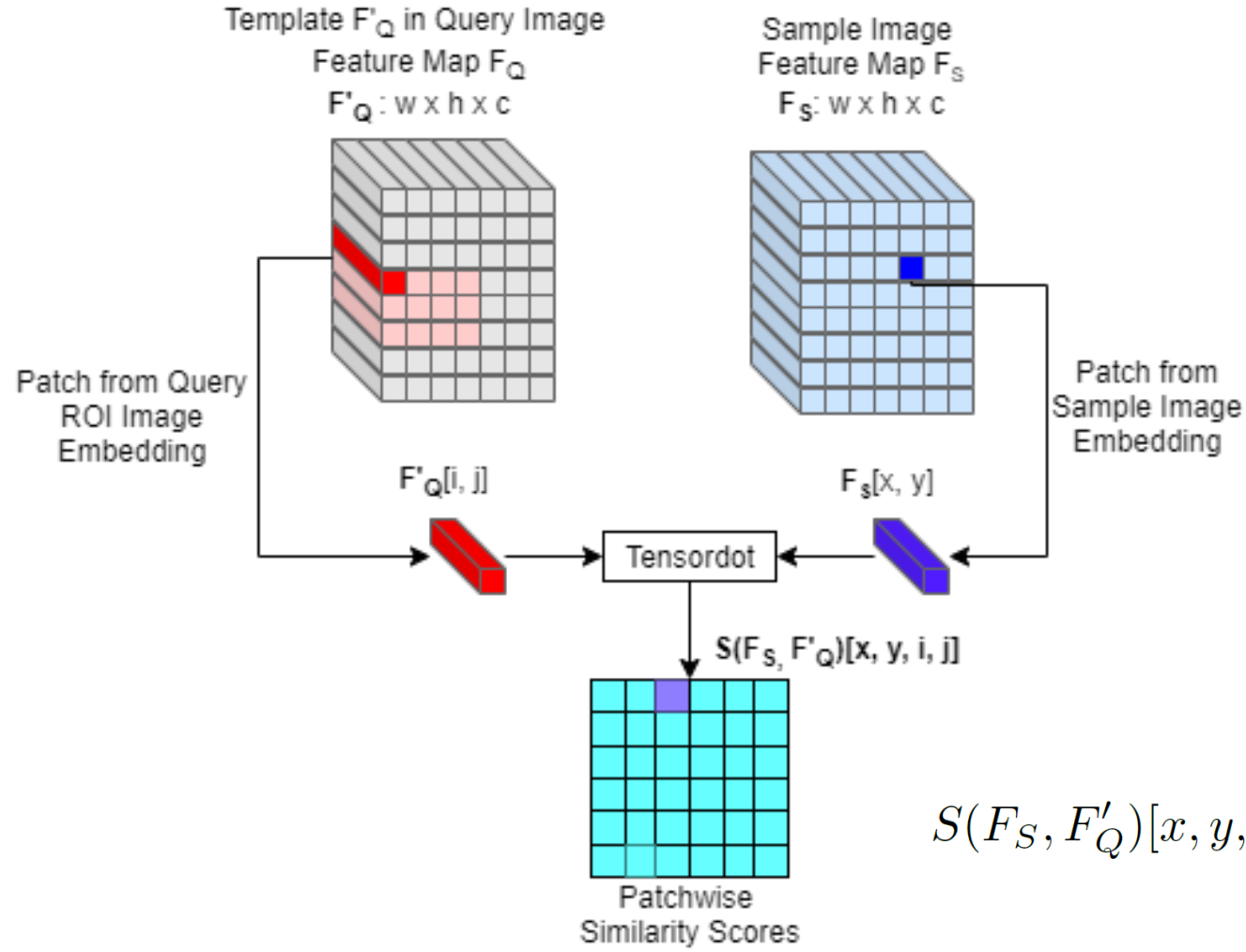
Deep Template Matching (DTM)
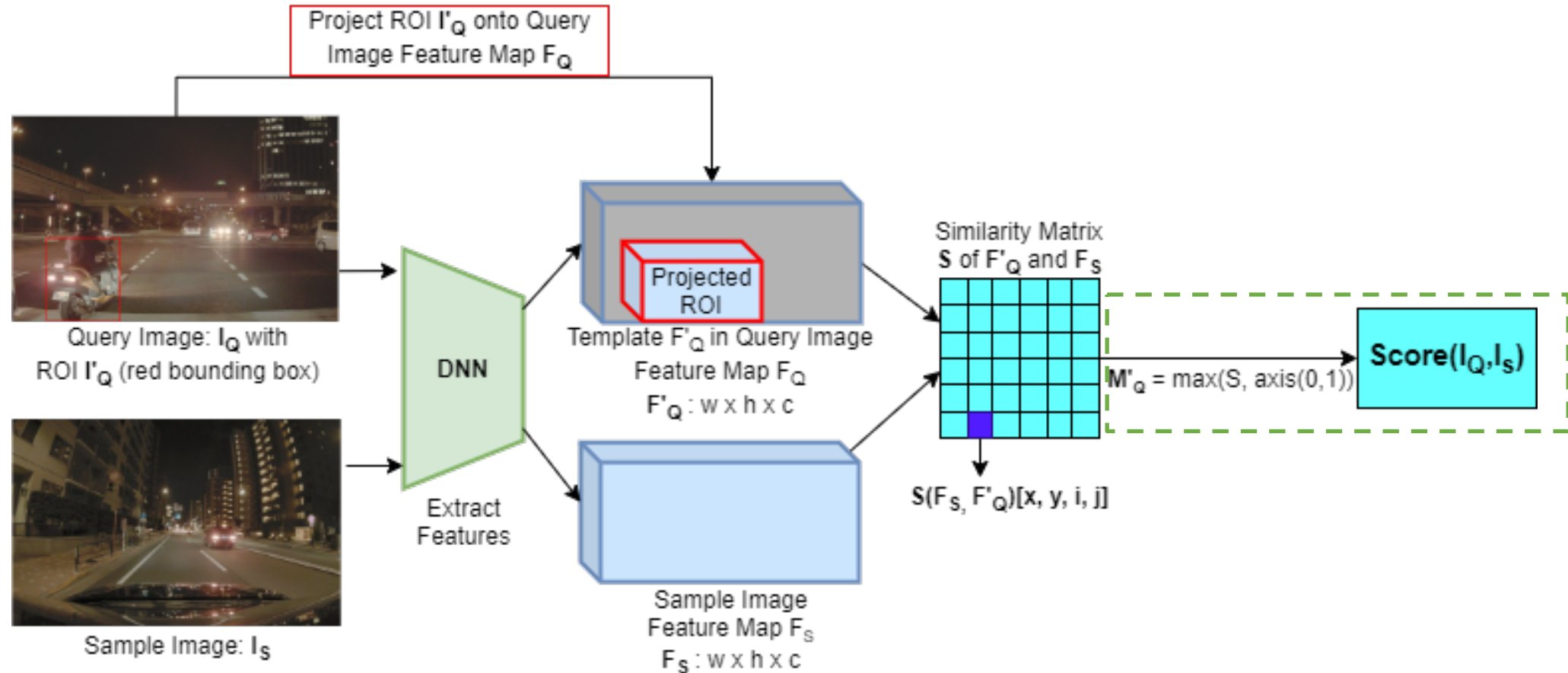
Deep Template Matching (DTM)

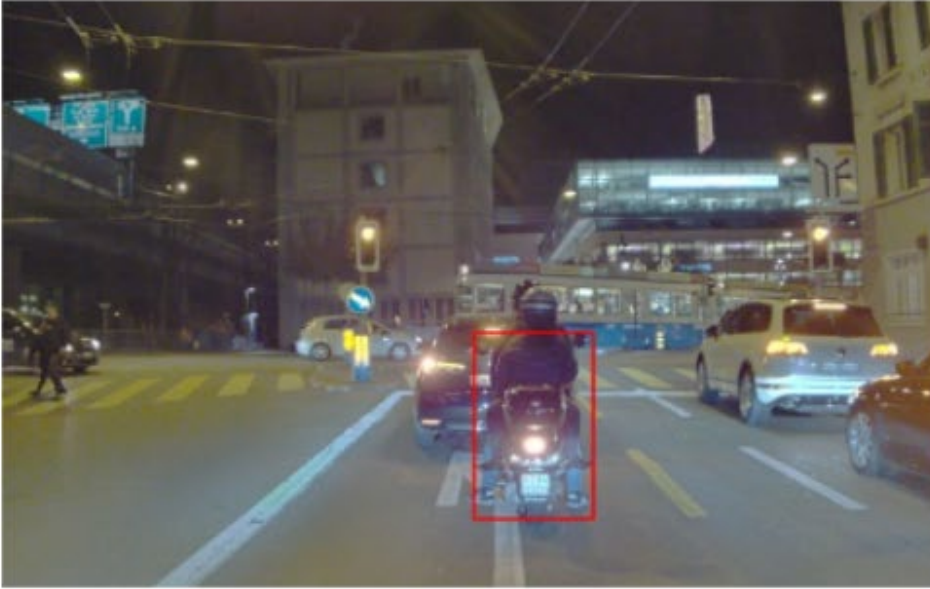# Patchwise Similarity Matching

$$S(F_S, F'_Q)[x, y, i, j] = \frac{F_S[x, y] \cdot F'_Q[i, j]}{|F_S[x, y]|_2 |F'_Q[i, j]|_2}$$

Deep Template Matching (DTM)

Scoring method

Input query image with region of interest

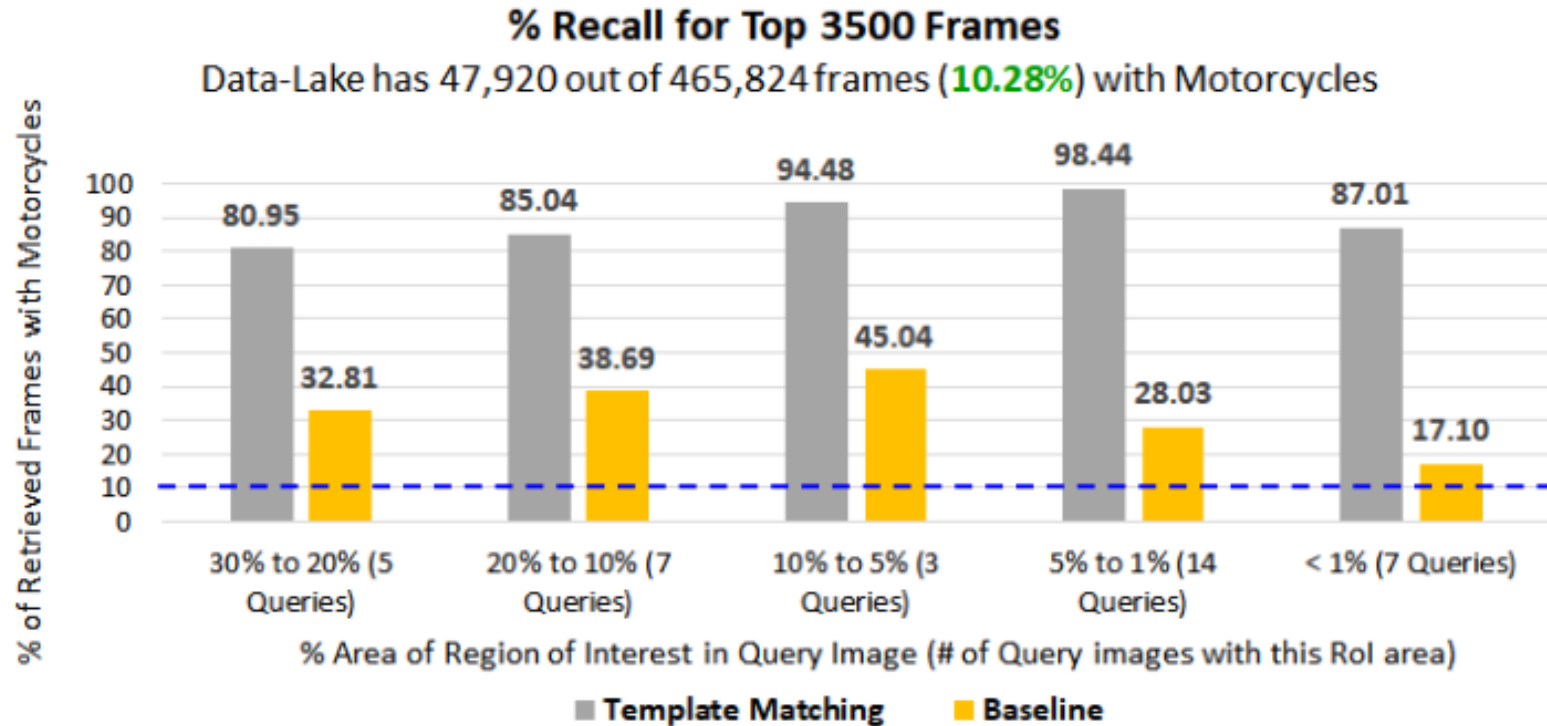Top-1 semantically similar image using DTM with overlaid heat map.

$$M'_Q = \frac{\max\limits_{axis=(0,1)} S(F_S, F'_Q)}{A'_Q}$$

$$Score(I_Q, I_S) = Mean(M'_Q)$$

Recall at Top-3500

Top 3500 retrievals for each of the 36 query images, i.e. 126K retrievals

% Recall for Top 3500 Frames

Data-Lake has 47,920 out of 465,824 frames (**10.28%**) with Motorcycles

% of Retrieved Frames with Motorcycles

Template Matching values	Baseline values
80.95	32.81
85.04	38.69
94.48	45.04
98.44	28.03
87.01	17.10

% Area of Region of Interest in Query Image (# of Query images with this RoI area)

30% to 20% (5 Queries) | 20% to 10% (7 Queries) | 10% to 5% (3 Queries) | 5% to 1% (14 Queries) | < 1% (7 Queries)

■ **Template Matching** ■ **Baseline**

- DTM selection has a recall ranging from 98% to 80%
  - Good recall across entire range of query RoI area
- Baseline approach has a recall ranging from 45% to 17%.
  - As expected, recall degrades for <1% RoI area (small objects)

Relative gain of DTM Recall over Random is at least 8x and over Baseline is at least 2x

Recall at Top-100

Top 100 retrievals for each of the 36 query images, i.e. 3.6K retrievals

% Recall for Top 100 Frames
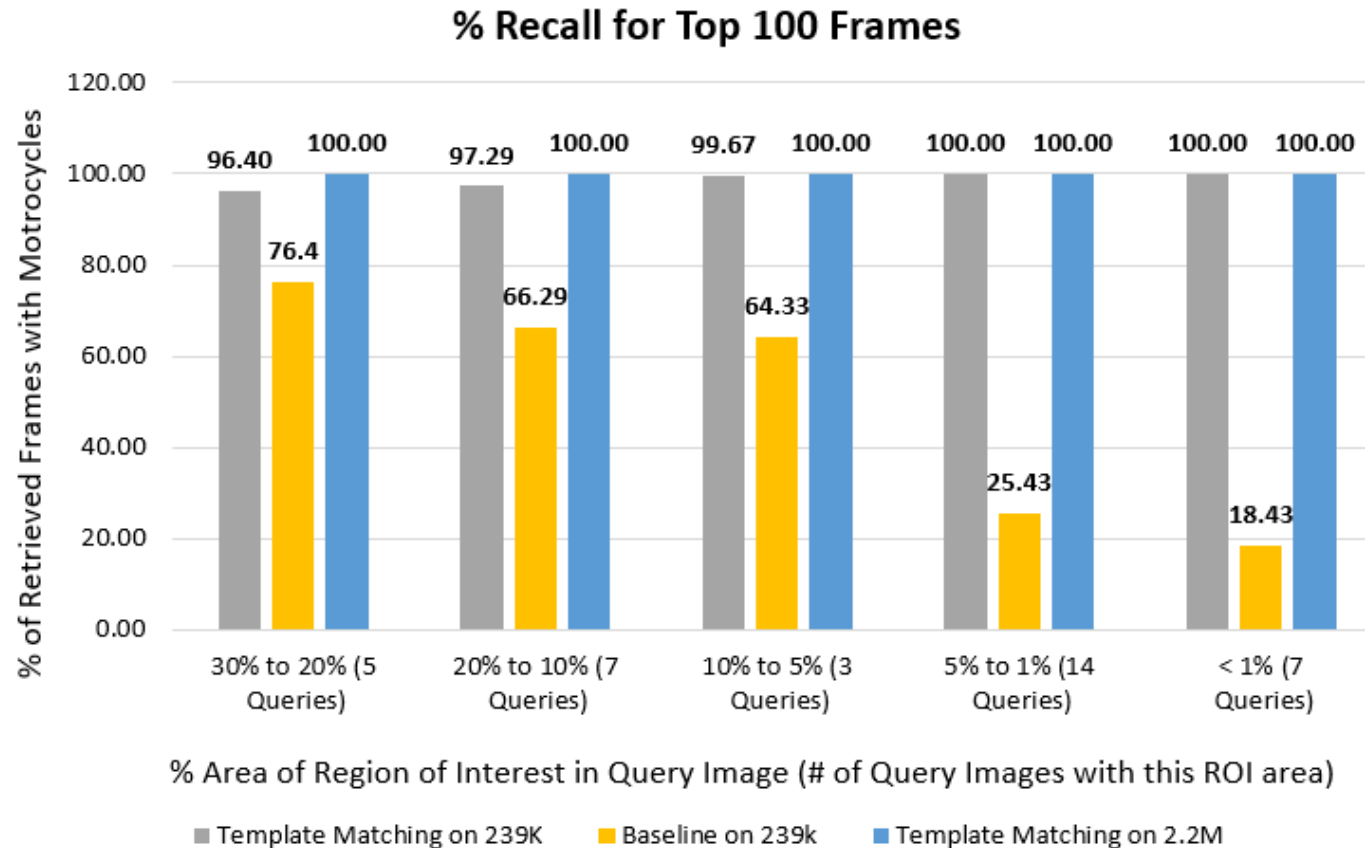
- We observe consistent results on a large unlabeled dataset with 2.2M images.

Relative gain of DTM Recall over Random is at least 32x and over Baseline is at least 1.2x

Conclusion

- We propose a novel approach for object level targeted selection by using deep template matching.

- We study this problem in the context of mining images that are semantically similar to failure cases, like false negatives/positives of object detectors deployed in autonomous vehicles.

- Our method focuses on the semantics of the objects of interest by projecting it onto the feature space and has high recall even when the object is small-sized, amid occlusion and heavy clutter.

- Our method works for multiple co-occurring objects in one or more semantic categories for object-level retrieval.

- Unlike other methods, it does not require extra labeled training data.

Thank You

For more details, do visit our *poster*.