Object-Level Targeted Selection via Deep Template Matching

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

IV22 33rd IEEE INTELLIGENT VEHICLES SYMPOSIUM

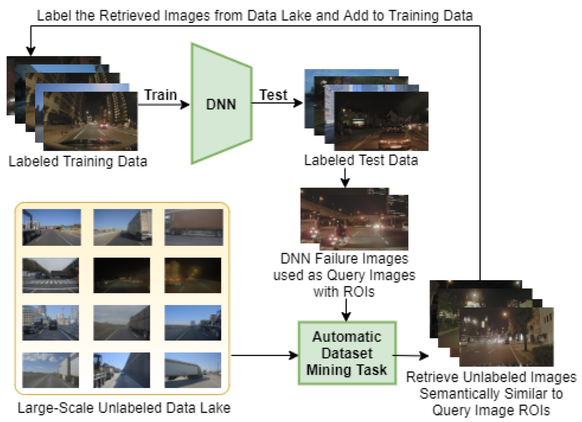
INTRODUCTION

- Retrieving images with objects that are semantically similar to objects of interest (OOI) in a query image has many practical use cases.
- One such important use case is the automatic mining of datasets at an object-level for fixing failures, like false negatives/positives, in deep neural network.
- Existing semantic image retrieval methods often focus on mining for larger sized geographical landmarks, and/or require extra labeled data, such as images/image-pairs with similar objects, for mining images with generic objects.
- We propose a fast and robust template matching algorithm in the DNN feature space, that retrieves semantically similar images at the object-level from a large unlabeled pool of data.

PROBLEM FORMULATION

Goal: Mine for semantically similar images that have some region semantically similar to a template representing an object.

- For instance, the object can be a failure case of an AV based DNN.
- Retrieving semantically similar frames to the failure case and adding them to the training data can fix the failure case.



Deep Template Matching (DTM)

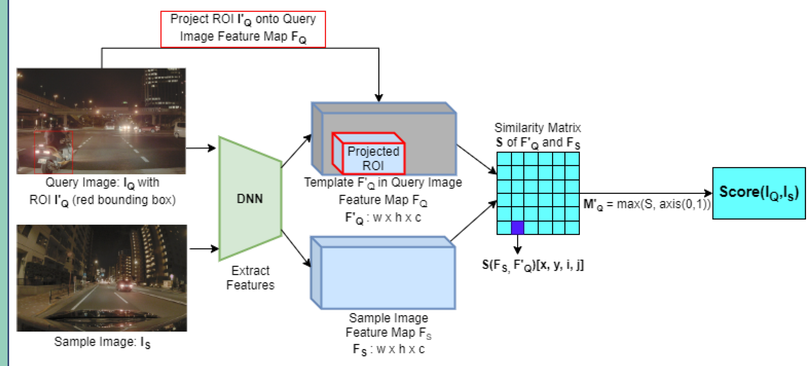


Fig. Architecture of Deep Template Matching

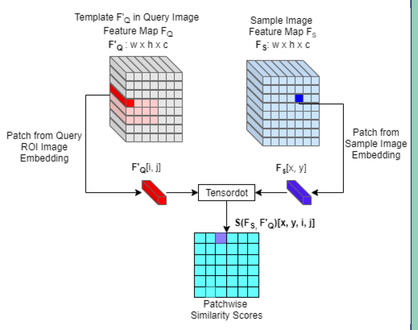


Fig. Patchwise Cosine Similarity

QUALITATIVE RESULTS

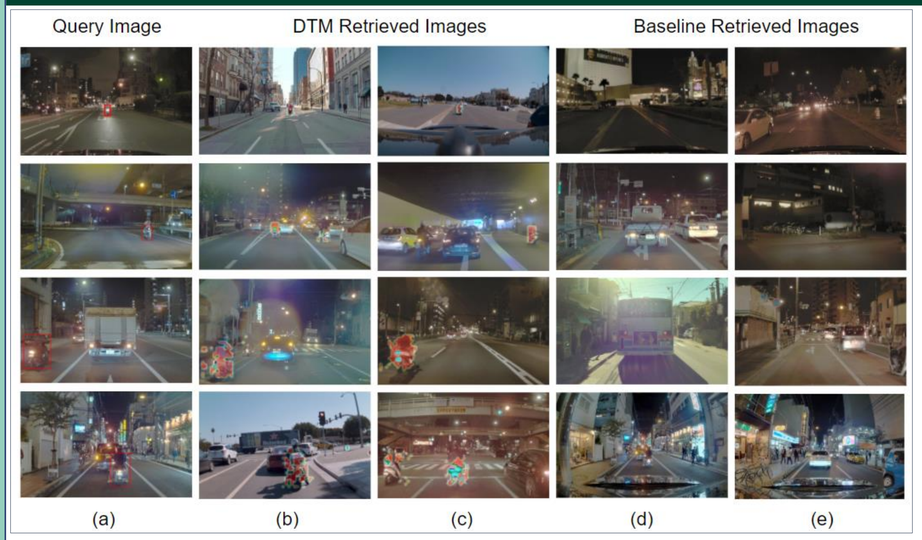


Fig. Results on real AV data: (a) Query image with region of interest (motorcycle). (b-c) DTM retrieved images containing motorcycles with similar semantics. (d-e) Baseline results that uses global features retrieved images that use semantics of the whole query image. Note how accurate the DTM heap maps are even for small objects with $\leq 1\%$ ROI area, e.g., the query image ROI in the top row.

Fig. DTM retrieved images for out-of-distribution queries. (a) Stroller (b) Luggage (c) Stop sign (d) Traffic cone. In (a-b), DTM is able to retrieve semantically similar images from the same scene as the query as well as other scenes. For (c-d), DTM only retrieves images from the same scene since features for those objects are not modeled in the underlying DNN feature representation.

QUANTITATIVE RESULTS

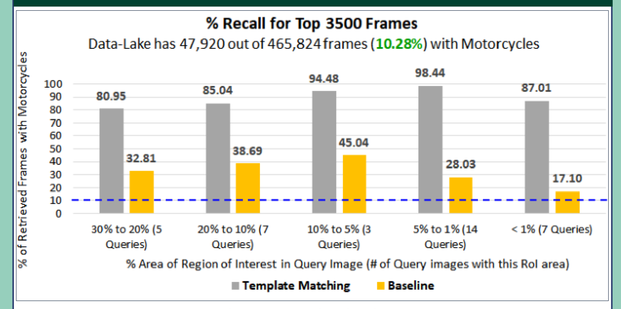


Fig. Results on Labeled Dataset

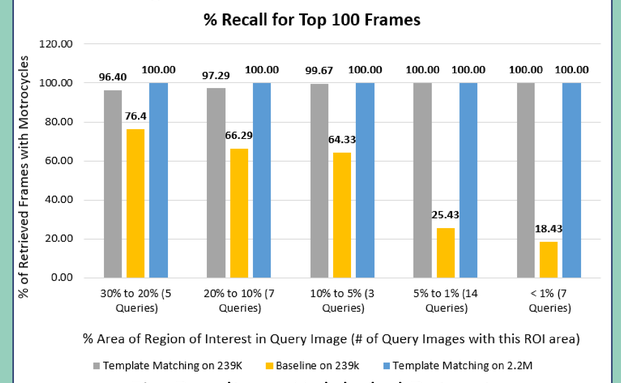


Fig. Results on Unlabeled Dataset

CONCLUSIONS

- We propose a new approach for semantic object retrieval in the wild by using deep features. We study this problem in the context of automatic dataset mining at an object-level for fixing failures, like false negatives/positives, in DNN based object detector.
- Our method seamlessly works for multiple co-occurring objects in one or more semantic categories for object-level retrieval.

PAPER

SCAN ME

Get the paper for more technical details and results:

https://tinyurl.com/3eztcjjz