END-TO-END VISUAL PLACE RECOGNITION BASED ON DEEP METRIC LEARNING AND SELF-ADAPTIVELY ENHANCED SIMILARITY METRIC

Chenyang Zhao, Runwei Ding, Hong Liu

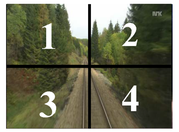
Key Laboratory of Machine Perception, Peking University, Shenzhen Graduate School

INTRODUCTION

- Place recognition: The task aims at recognizing the previously visited places, is the key component of loop closure in most visual simultaneous localization and mapping systems.
- Challenges: The challenges are especially introduced in dynamic and large-scale scenarios as the appearance of an environment can change greatly over time due to scene alterations, variation in illumination or weather phenomenon.
- Major contribution: We proposed an end-to-end network based on deep metric learning with a novel self-adaptively enhanced similarity (SAES) metric layer which strength the discrimination ability in the condition of appearance changing and calculate the similarity between image pairs.

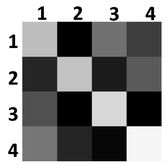
ILLUSTRATION OF RELATIONSHIP BETWEEN ENHANCEMENT FACTOR AND SIMILARITY MATRIX

(a1) Image I

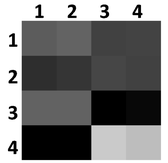


(a2) Image I^+

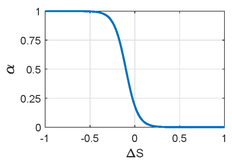
(a3) Image I^-



(b1) Similarity matrix of image I and image I^+



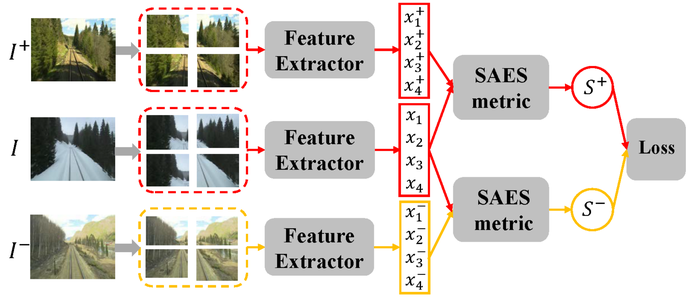
(b2) Similarity matrix of image I and image I^-



(c) Relationship between the enhancement factor α and ΔS.

- If the S_{diag} is distinctly larger than the S_{off}, the two frames have a high probability of coming from the same place, then α approaches to 1. However, if these two values have no distinct difference or even S_{off} is larger than S_{diag}, to a large extent the two images are captured from different places. Then we set α as a small value.

OVERALL FRAMEWORK



SELF-ADAPTIVELY ENHANCED SIMILARITY METRIC

designed to measure the similarity of two frames with the input features.

- Given a frame pair and its corresponding features X and X^p which are obtained from the feature extractor:

$$X = \{x_1, x_2, x_3, x_4\}, X^p = \{x_1^p, x_2^p, x_3^p, x_4^p\}$$

- x_i is the feature generated from the $i-th$ image patch. The similarity of two images can be reflected by the similarities of corresponding patches. The patch similarity between x_i and x_j^p is represented as M_{ij}, as defined as:

$$M_{ij} = \frac{1}{2}(\cos < \overrightarrow{x_i}, \overrightarrow{x_j^p} > +1),\ M_{ij} \in [0,1]$$

- The similarity matrix of the frame pair is:

$$M = \{M_{ij}, 1 \le i \le 4, 1 \le j \le 4\}$$

- The overall similarity S is defined as:

$$S = \alpha \cdot \frac{1}{4}\sum_{i=1}^{4} M_{ii}$$

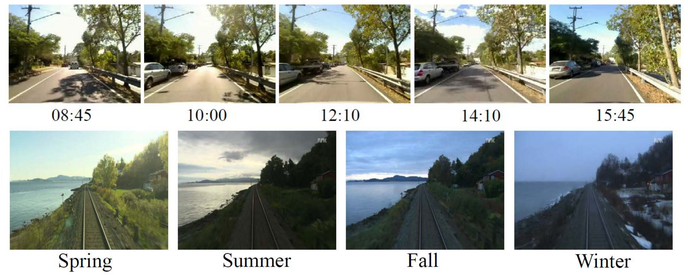
- α is the enhancement factor. It is calculated by comparing the values on diagonal with those on off-diagonal in the similarity matrix:

$$\alpha = \frac{1}{1 + e^{\sigma(\Delta S - \theta)}}$$

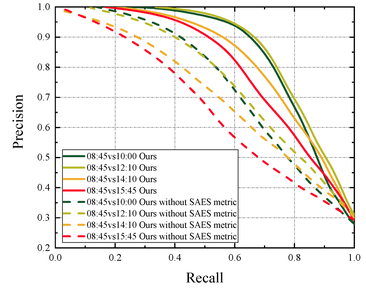
where $\Delta S = S_{off} - S_{diag}$, and σ and θ are constant parameters.

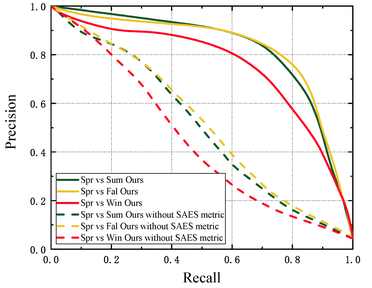
- S_{diag} is the average value of diagonal similarities and S_{off} is that of off-diagonal similarities.

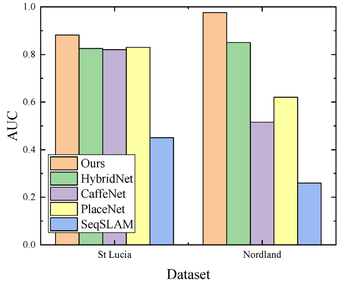
DATASETS



PERFORMANCE







- To evaluate the effect of SAES metric, experiments without SAES metric are conducted.
- Experiments are conducted on the St.Lucia and Nordland datasets to compare our method with some state-of-the-art methods.

CONCLUSION

- In this work an end-to-end deep metric learning network has been attempted on the place recognition problem with appearance changing.
- The effectiveness of the network is significantly improved by using the Self-adaptively Enhanced Similarity (SAES) metric.