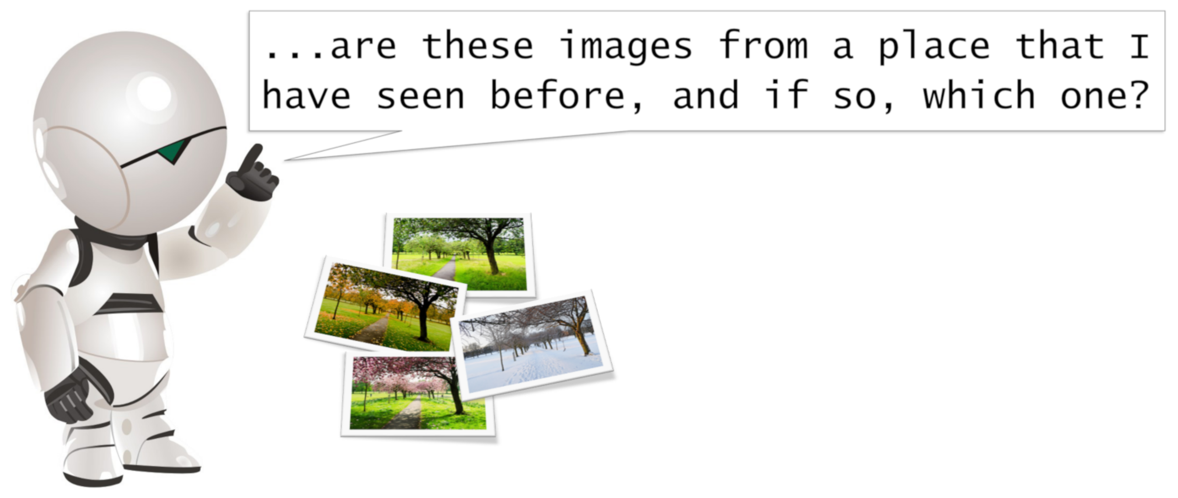
Visual Localization in the Presence of Appearance Changes Using the Partial Order Kernel

Maryam Abdollahyan[1], Silvia Cascianelli[2], Enrico Bellocchio[2], Gabriele Costante[2], Thomas A. Ciarfuglia[2], Francesco Bianconi[2], Fabrizio Smeraldi[1] and Mario L. Fravolini[2]

[1] *School of Electronic Engineering and Computer Science, Queen Mary University of London, UK* ✉ *m.abdollahyan@qmul.ac.uk*
[2] *Department of Engineering, Università degli Studi di Perugia, Italy* ✉ *silvia.cascianelli@studenti.unipg.it*

Localization in Changing Environments

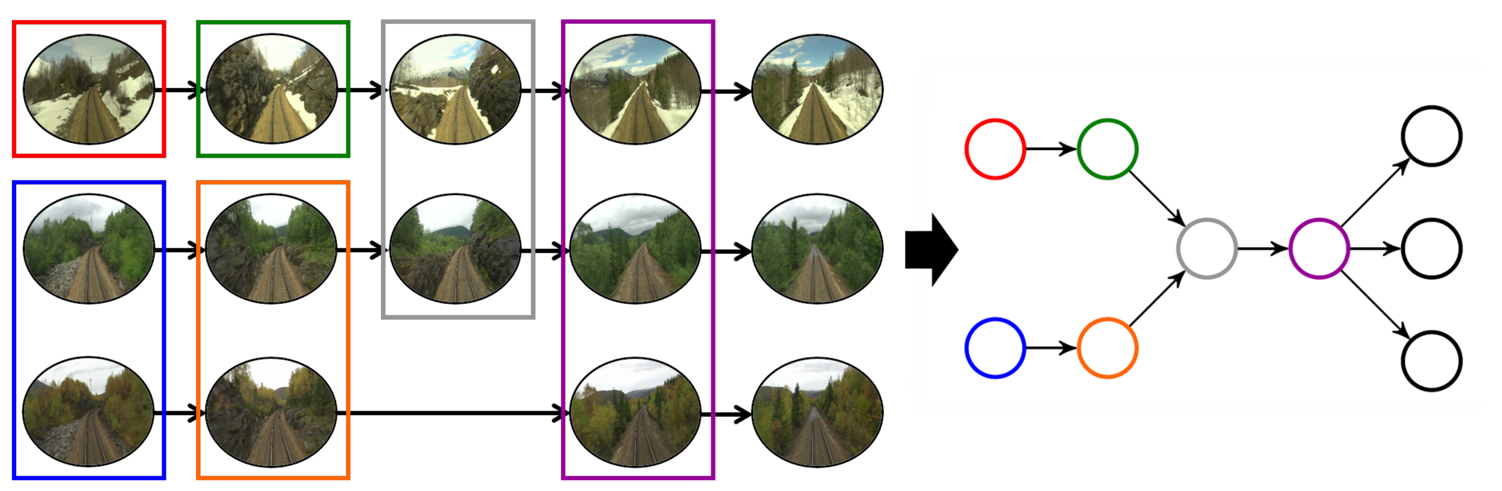
- Accurate and efficient **visual localization** is a fundamental problem in robotics.



- Visual localization is a challenging task due to **changes in the appearance of the environment**, which are caused by illumination variation, different weather conditions, seasonal changes, etc.
- We present a **sequence-based** visual localization approach that is robust to appearance changes. Localization requires:
  - I. a **representation** of the environment: we represent different sequences of images of the same place taken under varying appearance conditions as alternative paths in a **directed acyclic graph (DAG)**.
  - II. a **measure of similarity** between the query image sequence and image sequences in the database: we match the sequences by comparing their corresponding DAGs using the **Partial Order Kernel (POKer)**.

I. Graph Representation of Image Sequences

- A set of **database** image sequences of the same place acquired at different times are aligned using the **Partial Order Alignment (POA)** algorithm (Lee et al, 2002) to obtain a DAG **with branches**. To find an optimal alignment:
  - ★ a score is assigned to each aligned pair of nodes. We use the cosine similarity between image descriptors as scores. Nodes that are similar based on this score are merged.
  - ★ gaps are penalized. We use a linear gap penalty.



3 database image sequences of a place in spring, summer and autumn (Nordland dataset)

- A **query** image sequence is represented as a DAG **without branches**.

a query image sequence of the above place in winter (Nordland dataset)

II.a. The Partial Order Kernel

- POKer (Abdollahyan et al, 2017) is a **convolution kernel** for comparison of strings containing alternative substrings and represented by DAGs.
- Let $\Pi_n(G_x)$ and $\Pi_n(G_y)$ be the sets of paths of length n in DAGs G_x and G_y, respectively. POKer is defined as

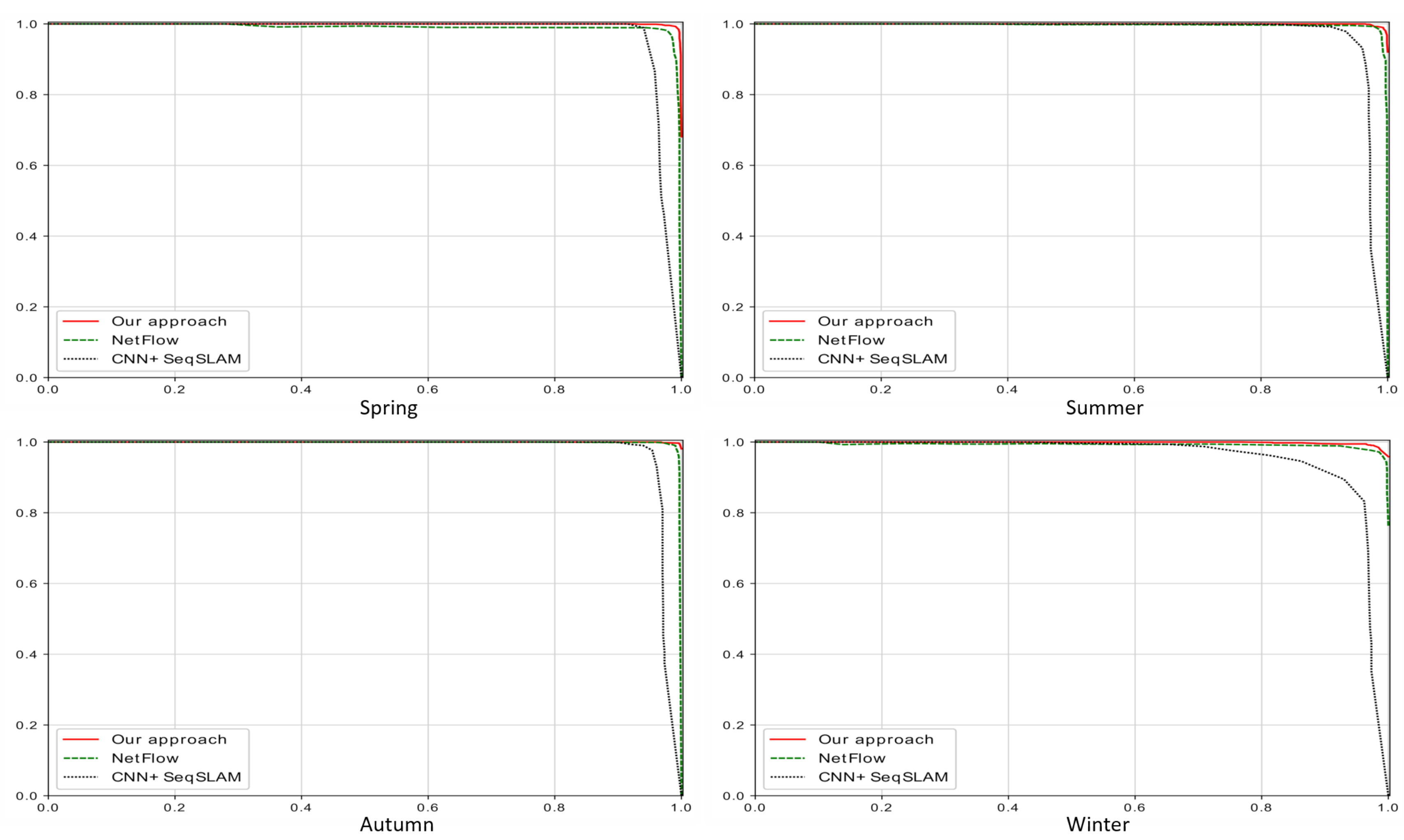
$$K(G_x, G_y) = \sum_{n \geq 0} K_n(G_x, G_y) = \sum_{n \geq 0} \sum_{\substack{\pi_x \in \Pi_n(G_x) \\ \pi_y \in \Pi_n(G_y)}} \exp(\beta S(\pi_x, \pi_y))$$

  - ★ $S(\pi_x, \pi_y)$ is the score of the **local alignment** of n nodes along path π_x with the same number of nodes along path π_y. We use the same scores and gap penalty as in the POA.
  - ★ $\beta \geq 0$ is a parameter. Valid values for it are those for which POKer is **positive semi-definite**.

II.b. The Partial Order Kernel

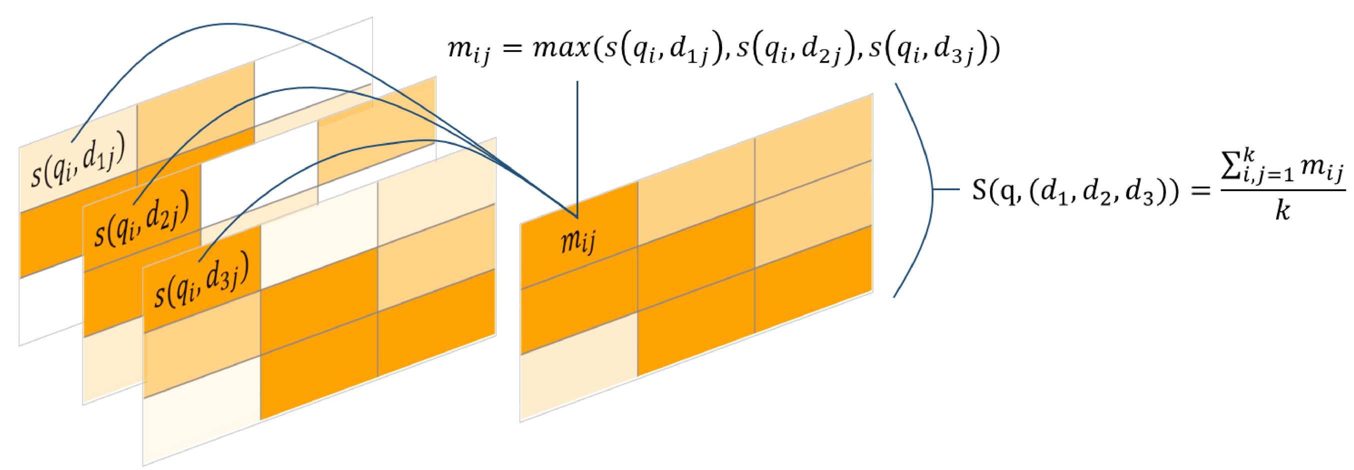
- ★ POKer accounts for the contributions of **all** the possible local alignments between any path in the query DAG and any path in the database DAG. For $\beta \to \infty$, only the best alignments are taken into account.
- ★ POKer is computed using **dynamic programming**, with a **linear** time complexity with respect to the number of nodes in the strong product of input DAGs.

Experiments & Results



precision-recall curves from the experiments

- We evaluated our approach on the **Nordland dataset**, which consists of footage of a 728*km*-long train journey recorded once in every season. We subsampled each video at 0.5*fps* and cut its sequence into subsequences of length 15 to obtain 4 sets of image sequences: **Spring**, **Summer**, **Autumn** and **Winter**. We performed 4 experiments, each time using the sequences from a different season as the query DAGs and the triplets of sequences from other seasons as the database DAGs.
- Image descriptors were extracted from the **Places365-VGG Convolutional Neural Network** (Zhou et al, 2017), pre-trained for place recognition. Gap penalty and β were set to -1 and 1, respectively.
- We used the methods presented in (Naseer et al, 2014) and (Milford et al, 2012), with the same descriptors as those used in our method, as baselines. We refer to them as **NetFlow** and **CNN+SeqSLAM**, respectively.



since the baselines compute similarity s of an image to another, not a sequence to a triplet of sequences, similarity S between query sequence q and triplet of database sequences (d_1, d_2, d_3) is computed as shown above

Precision (%)	Recall (%)		
	Our approach	*NetFlow*	*CNN+SeqSLAM*
100	90.7	37.0	75.5
95	99.9	99.2	92.6
90	99.9	99.6	95.1

average recall values for each method

- Our method either matches or outperforms the baselines at all **precision-recall** values.
- Experiments on the Nordland dataset show that POKer **effectively** computes the similarity between DAGs, in an approach that is robust to appearance changes and significantly outperforms 2 state-of-the-art methods.

References

1. M. Abdollahyan and F. Smeraldi, POKer: a partial order kernel for comparing strings with alternative substrings in *25th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, 2017, pp. 263268.
2. C. Lee, C. Grasso, and M. F. Sharlow, Multiple sequence alignment using partial order graphs *Bioinformatics*, vol. 18, no. 3, pp. 452464, 2002.
3. M. J. Milford and G. F. Wyeth, SeqSLAM: visual route-based navigation for sunny summer days and stormy winter nights *in 2012 IEEE International Conference on Robotics and Automation (ICRA)*, 2012, pp. 16431649.
4. T. Naseer, L. Spinello, W. Burgard, and C. Stachniss, Robust visual robot localization across seasons using network flows in *AAAI Conference on Artificial Intelligence*, 2014, pp. 25642570.
5. B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, Places: a 10 million image database for scene recognition *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.