LEARNING VISUAL SIMILARITY FOR IMAGE RETRIEVAL

Garcia N., Vogiatzis G. - Aston Univeristy
{garciadn, g.vogiatzis}@aston.ac.uk

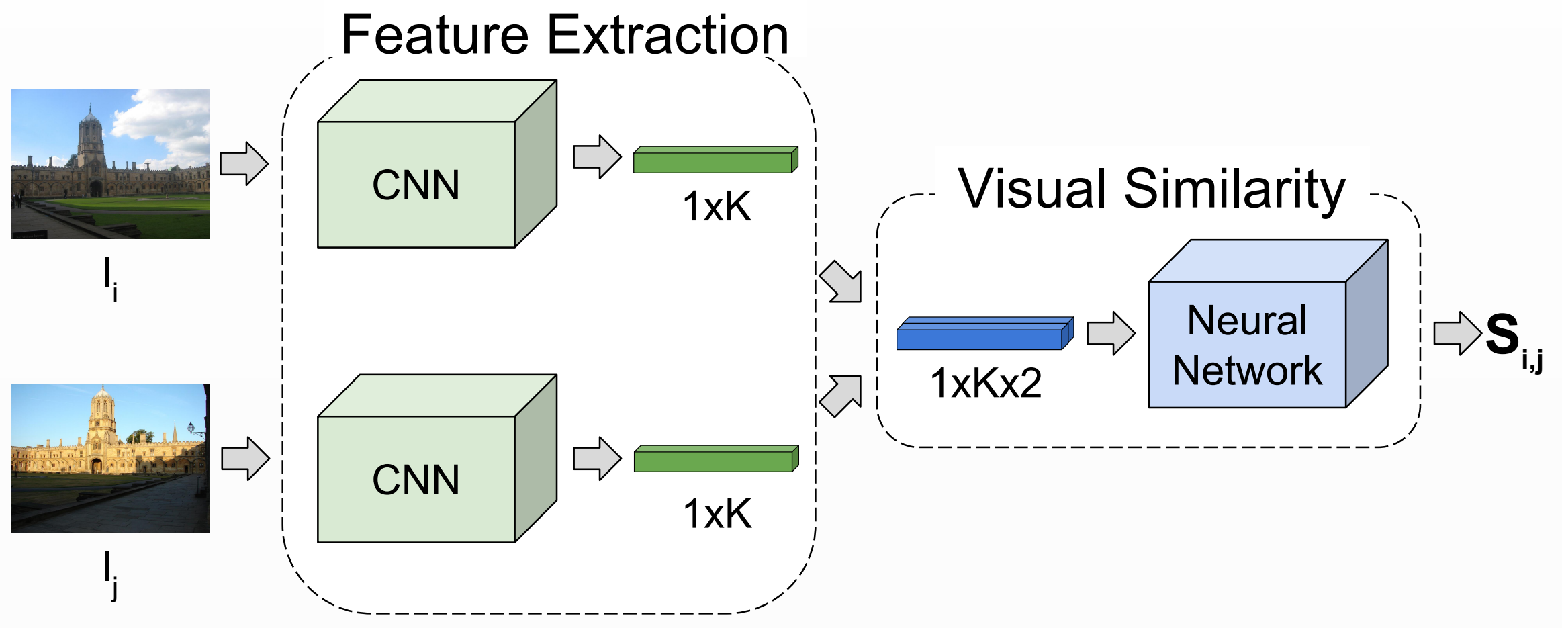
Aston University
Birmingham

Abstract

Can a computer vision system learn the concept of visual similarity? Traditionally, visual similarity in image retrieval is measured by using standard metrics, such as Euclidean distance or cosine similarity. However, these metrics are independent from data and might be missing the nonlinear structure of visual representations. In this work, we propose to learn a nonlinear visual similarity function directly from image representations by optimizing a neural network model.

System Overview

Image Retrieval can be broken into two fundamental tasks: **Feature Extraction** and **Visual Similarity**.



We compute a similarity score by extracting image representations from a pre-trained CNN and comparing them in a visual similarity network.

References

[1] A. Babenko, A. Slesarev, A. Chigorin, and V. Lempitsky. Neural codes for image retrieval. In *ECCV*, 2014.

[2] G. Chechik, V. Sharma, U. Shalit, and S. Bengio. Large scale online learning of image similarity through ranking. *Journal of Machine Learning Research*, 2010.

[3] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *CVPR*, 2007.

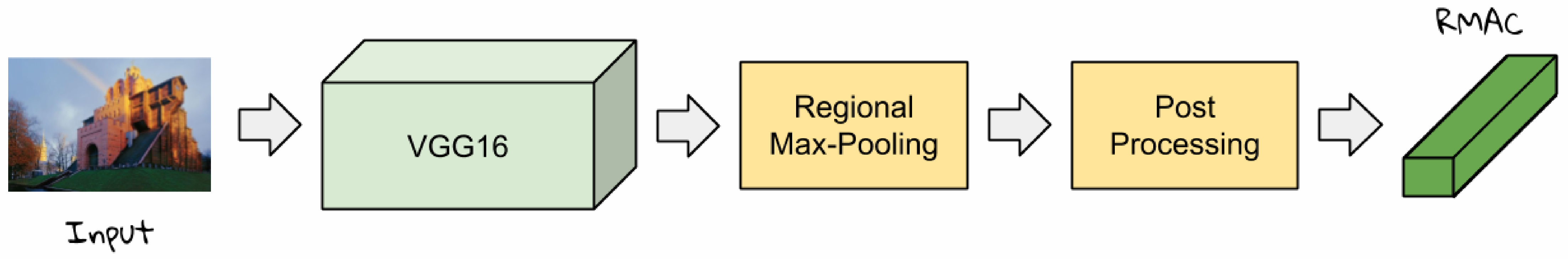
[4] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Lost in quantization: Improving particular object retrieval in large scale image databases. In *CVPR*, 2008.

[5] G. Tolias, R. Sicre, and H. Jégou. Particular object retrieval with integral max-pooling of cnn activations. In *ICLR*, 2016.

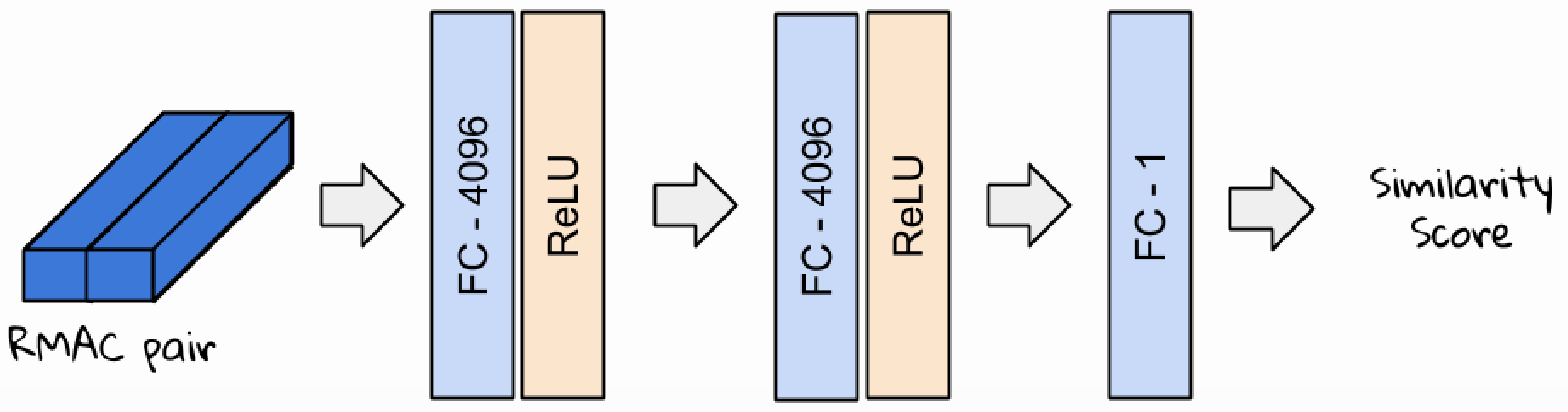
Methodology

The proposed visual similarity neural network learns the mapping from a pair of images to a similarity score.

Feature Extraction Image's visual content is extracted into RMAC vectors [5].

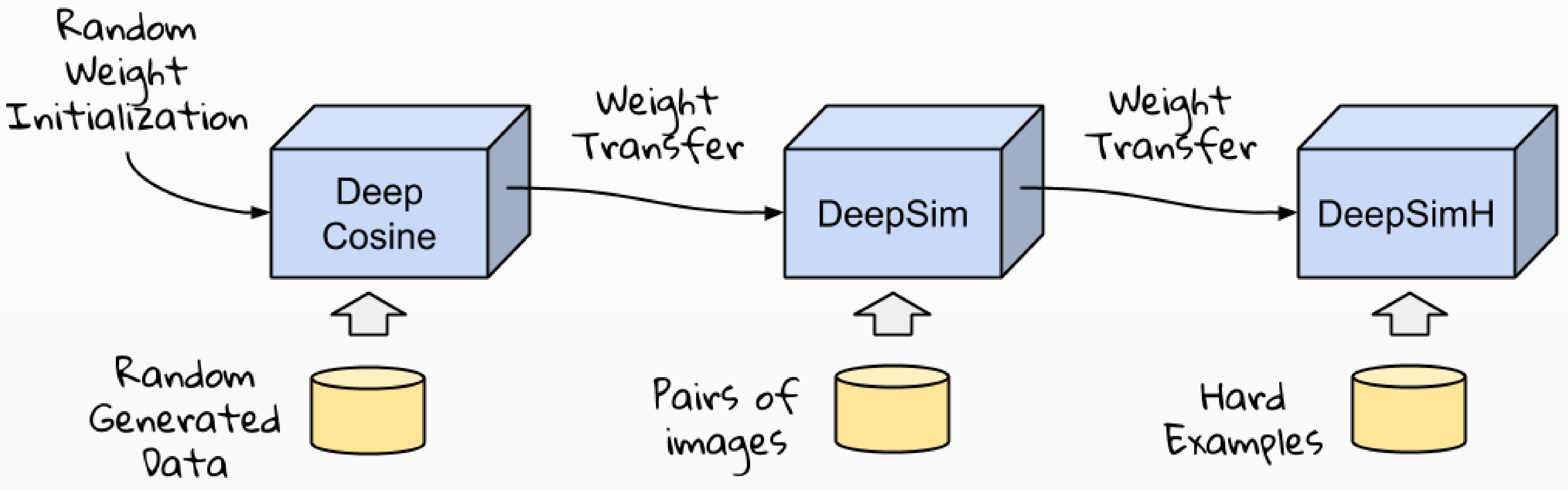


Similarity Network Pairs of RMAC vectors are fed into the similarity network.



Training The similarity network is trained in 3 stages.

1. DeepCosine: learns cosine similarity by using random generated data.
2. DeepSim: learns visual similarity by increasing (decreasing) Δ to the standard cosine similarity score for matching (non-matching) pairs of images.
3. DeepSimH: reinforces specifically the learning of difficult pairs of images.



Results

The proposed methodology is evaluated in terms of mAP on three challenging image retrieval datasets: Oxford5k [3] with 5,062 images, Paris6k [4] with 6,412 images and Landmarks5k [1] with 4,915 images.

	Oxford5k	Paris6k	Landmarks5k
Cosine Similarity (Baseline)	0.665	0.638	0.564
OASIS [2]	0.619	0.853	0.579
DeepCosine	0.638	0.596	0.549
DeepSim	0.718	0.757	**0.668**
DeepSimH	**0.786**	**0.860**	0.662

The network is trained using 35,342 images from the Landmarks [1] collection, including 1,000 images from each of the Oxford and Paris datasets.