Learning Encodings using Deep Neural Networks with Generalized Max Pooling Layers

State-of-the-art systems for writer recognition use the activation features of deep convolutional neural networks (CNN) to compute local image descriptors [1, 2]. Subsequently, a global image descriptor, i.e. a single-vector image-level representation, is computed by encoding the obtained CNN activation features. The standard approach to computing a global descriptor of an image given a set of local descriptors consists of two steps: In the *embedding step* an embedding function maps each local descriptor into a high-dimensional space. During the *aggregation step* an aggregation function, e.g. sum pooling or max pooling, computes a single vector from the embedded feature vectors. The global image descriptors are used to compute the similarity score between two images.

However, the aggregated descriptors can suffer from interference of unrelated descriptors that influence the similarity, even if they have low individual similarity [3]. The issue of interference is exacerbated in the presence of bursty descriptors, that means very similar frequent descriptors that together form a mode in descriptor space and therefore dominate the similarity metric. Murray and Perronnin [4] propose a generalized form of max pooling to address this issue by reducing the influence of bursty descriptors on the global descriptor.

In this work, generalized max pooling is incorporated as a layer into a deep neural network architecture, allowing for end-to-end training of the network for writer recognition.

The thesis consists of the following milestones:

- Incorporating generalized max pooling [3, 4] into a neural network.

- Evaluating performance on the ICDAR17 competition dataset on historical document writer identification [5].

- Comparison with other pooling techniques.

- Further experiments regarding learning procedure and network architecure.

The implementation should be done in Python / C++.

Supervisors:	Dr.-Ing. V. Christlein, Prof. Dr.-Ing. habil. A. Maier, Mathias Seuret, M. Sc.
Student:	Timo Bohnstedt
Start:	November 8th, 2021
End:	April 0th, 2022

References

[1] Vincent Christlein, Martin Gropp, Stefan Fiel, and Andreas K. Maier. Unsupervised feature learning for writer identification and writer retrieval. *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 01:991–997, 2017.

[2] Vincent Christlein and Andreas K. Maier. Encoding CNN activations for writer recognition. *CoRR*, abs/1712.07923, 2017.

[3] Naila Murray, Hervé Jégou, Florent Perronnin, and Andrew Zisserman. Interferences in match kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9):1797–1810, 2017.

[4] Naila Murray and Florent Perronnin. Generalized max pooling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2473–2480, 2014.

[5] S. Fiel, F. Kleber, M. Diem, V. Christlein, G. Louloudis, S. Nikos, and B. Gatos. Icdar2017 competition on historical document writer identification (historical-wi). In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 01, pages 1377–1382, Nov. 2018.