

Abstract: Deep Hashing for Large-Scale Medical Image Retrieval

Sailesh Conjeti^{1,2}, Magdalini Paschali², Abhijit Guha Roy^{2,3}, Nassir Navab^{2,4}

¹ Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE), Bonn, Germany

² Computer Aided Medical Procedures, Technische Universität München, Germany.

² AI-med, Ludwig-Maximilian Universität München, Germany.

⁴ Computer Aided Medical Procedures, Johns Hopkins University, USA.

`sailesh.conjeti@dzne.de`

Adoption of content-based image retrieval systems (CBIR) requires efficient indexing of the data contents in order to respond to visual queries without explicitly relying on textual keywords. Searching for similar data is closely related to the fundamental problem of nearest neighbor search. Exhaustive comparison of a query across the database is infeasible in large-scale retrieval as it is computationally expensive [1]. Towards this, we propose scalable image retrieval techniques that employing *hashing* to learn semantics-preserving binary codes for indexing and retrieval. We leverage deep learning for end-to-end learning of hash-codes that are tailored for retrieval in large-scale medical image databases. Developing medical image CBIR systems is particularly challenging due to variability in image acquisition (contrast agents, protocol variations), anatomical variability, high-dimensionality of the data *etc.* We recently introduced *Deep Residual Hashing* (DRH) [2] and *Deep Multiple Instance Hashing* (DMIH) [3] which are deep hashing methods for medical image retrieval. DRH uses residual networks that terminate in dedicated binarization layers that generate hash-codes. It was trained with a multi-class variant of neighborhood component analysis (NCA) with dedicated losses such as quantization and bit-balance to improve hash code quality. DMIH extends DRH to a scenario of *multiple instance* learning by introduction of the multiple-instance pooling layer that aggregates representations from across member instances within a bag of samples. The aforementioned deep hashing methods were exhaustively validated and their fidelity was demonstrated on scenarios of retrieval on highly heterogeneous databases with significant population variability such as cardiopulmonary chest X-rays, breast mammography and breast histology, with highly competitive retrieval time (< 100 ms).

References

1. Conjeti S, Katouzian A, Kazi A, Mesbah S, Beymer D, Syeda-Mahmood TF, et al. Metric hashing forests. *Medical image analysis*. 2016;34:13–29.
2. Conjeti S, Roy AG, Katouzian A, Navab N. Hashing with residual networks for image retrieval. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2017. p. 541–549.
3. Conjeti S, Paschali M, Katouzian A, Navab N. Deep Multiple Instance Hashing for Scalable Medical Image Retrieval. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2017. p. 550–558.