A summary on

Deep Learning of Binary Hash Codes for Fast Image Retrieval

Raaghav Radhakrishnan
International Masters in Data Analytics
University of Hildesheim

13th May 2020

Introduction: Deep Neural Networks, especially CNNs, are the state-of-the-art methods in the domains of autonomous driving, robotics and medical-imaging to name but a few. CNNs have shown significant performance and improvement by learning rich image features at different layer levels. The research by Lin *et al.* [3] focuses on fast image retrieval from a large corpus by learning the binary hash codes and image features from the mid-layer of a pre-trained model adapted to a target domain and thereby eliminating the need of pair-wised similarity of the images in the data.

Proposed Framework and Domain Adaption: This paper utilizes the CNN architecture used in ImageNet [2] to learn the binary hash codes. The convolutional (conv) layers of the ImageNet is followed by two fully-connected (fc) layers of 4096 neurons each and a classification layer with 1000 neurons. As the fc layers of an ImageNet represent the image with its rich features [1], the authors Lin *et al.* propose a latent fc layer H with h neurons between the final fc layer and the classification layer. The neurons in the layer H are activated by sigmoid functions approximated to $\{0,1\}$ which are binarized by a threshold (>0.5) to obtain the hash codes. The proposed network uses pre-trained ImageNet weights for the conv and fc layers but randomly initializes weights for the latent layer H and the classification layer. The entire network is fine-tuned to adapt the parameters on the dataset of a target domain and learn the binary hash codes for a rapid image retrieval.

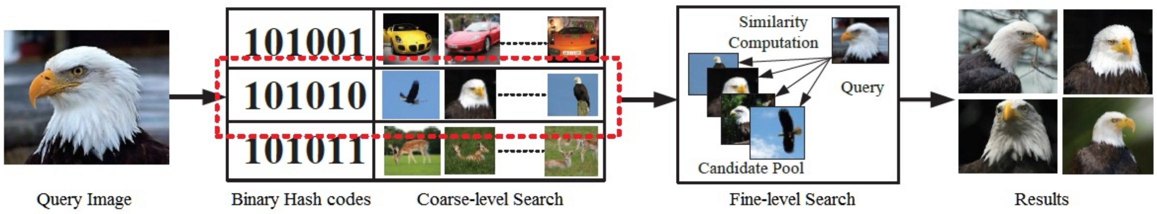

Figure 1: Image matching and retrieval via Hierarchical Deep Search

Hierarchical Deep Search (HDS): Once the network is adapted and binary hash codes are learnt, a coarse-to-fine search is adopted, as shown in Figure 1, for an effective image retrieval. Firstly, a set of images and its corresponding hash codes is considered. Secondly, for a given query image and its binary code, similar images with Hamming distance less than a threshold are pooled to form a set of candidates. Finally, a fine search is done by calculating the Euclidean distance between the final fc layer feature vectors of the query image and the pool of candidates. The candidates are ranked on their similarity with the query image and the top k images are retrieved as required.

Experiments & Results: Experiments were carried out with different settings on 3 datasets. **MNIST & CIFAR-10:** The latent layer H was set to 48 and 128 neurons for comparison. To classify images, the proposed method performed in-line with baselines. However, for image retrieval, the method showed significant improvement in precision ($98.2 \pm 0.3\%$) against other hashing approaches. **Yahoo-1M dataset:** With 116 categories and latent layer H set to 128 neurons, the proposed method showed good classification accuracy (83.75%). The retrieval results obtained by features in different network modes: final fc layer features from ImageNet and proposed method (ES), latent binary code search (BCS) and HDS are compared. The precision of fine-tuned models, ES, BCS & HDS, outperformed ImageNet. Also, BCS performed 971.3x faster than HDS with similar precision.

Conclusion: Although the proposed idea is simple, it is effective and feels more engineering. A future work could utilize the features from semantically labelled data for improved binary hash code.

References

[1] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proc. CVPR*, 2014

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *Proc. NIPS*, 2012

[3] K. Lin, H.-F. Yang, J.-H. Hsiao, and C.-S. Chen. Deep learning of binary hash codes for fast image retrieval. *In Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 27–35, 2015.