Deep Residual Hashing Network for Image Retrieval

Jimenez-Lepe Edwin[1], Mendez-Vazquez Andres [1]

[1]Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional

Content-Based Image Retrieval

According to [10] let $X = \{x_1, \dots x_N\}$ be a database with N images and $F = \{f_1 \dots f_N\}$ where f_i is a feature vector associated with x_i and contains the relevant information required for measuring the similarity between images.

Let T represent a mapping from the image space onto the h-dimensional feature space, f_i, i.e.,

$$T : x \to f, \quad (1)$$

where $x \in X$ and $f \in F$.

The similarity between two images x_i and x_j can be measured using similarity function $d(f_i, f_j)$.

The problem of retrieval can be posed as follows: Given a query image q, retrieve a subset of images M from X such that:

$$d(T(q), T(m)) \leq t, \quad m \in M, \quad (2)$$

where t is a user-defined threshold. Instead of this, a user can ask the system to output, say, the top-k images which are most similar to the query image.

Deep Residual Network

DRHN is based on the model proposed by [3] which is formed by residual blocks. A residual block is defined in Eq. 6 and contains convolution (Eq. 3), ReLU (Eq. 4), Batch normalization and Element-wise addition (Eq. 5).

$$y_{ij} = \sum_{c=0}^{C-1} \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} w_{abc} x_{(i+a)(j+b)c} \quad (3)$$

$$y_{ij} = max\{x_{ij}, 0\} \quad (4)$$

$$y_{cij} = x_{cij} + z_{cij} \quad (5)$$

$$y = Add(BN_{\gamma_2,\beta_2}(conv(BN_{\gamma_1,\beta_1}(ReLU(conv(x, w1))), w2)), x) \quad (6)$$

A Residual Group is the join of n Residual Blocks. The architecture of the proposed model is shown in Figure 3 and detailed in Table 1.

Table 1: Definition of Residual Group

Layer	Convolution dimensions	Output dim.
Input		3x32x32
$BN_{\gamma,\beta}(Relu(conv(x, w1)))$	16x3x3x3	16x32x32
Residual Group (n)	16x16x3x3	16x32x32
Residual Group (n) w.i.d[a]	*32x16x3x3, 32x32x3x3	32x16x16
Residual Group (n) w.i.d	*64x32x3x3, 64x64x3x3	64x8x8
Residual Group (n)	64x64x3x3	64x8x8
Residual Group (n)	64x64x3x3	64x8x8
Residual Group (n) w.i.d	*128x64x3x3, 128x128x3x3	128x4x4
Average Pooling Layer		128
Hash Layer		h
FC Layer with Softmax		10

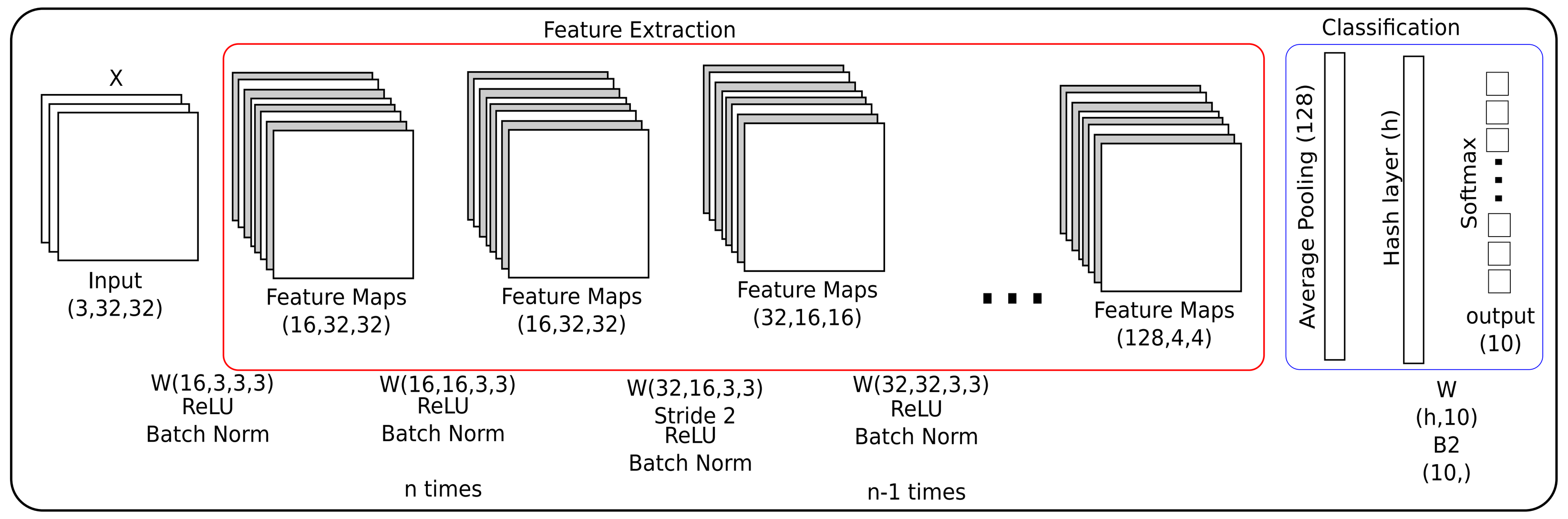

Figure 1: Model of Deep Residual Hashing Network

Generating the binary code

The Hash Layer (H) is a fully conected layer with a sigmoid activation function s:

$$H = s(Wx + b) \in \mathbb{R}^h \quad (7)$$

where $x \in \mathbb{R}^d$ is the input of the layer, $b \in \mathbb{R}^h$ is a bias and $W \in \mathbb{R}^{h \times d}$ represent the weights that connect the units of x with H.

Given an image $I \in \mathbb{R}^{c \times z \times w}$, where c represents the channels of the image, z the height and w the width, the layers of the model from Input layer to H form a hash function that performs the mapping from $\mathbb{R}^{c \times z \times w}$ to $\mathbb{R}^h$ as defined in Eq. 1.

Consequently, to obtain the binary code related to I as described by [5], we extract the output of H, and binarize the activation by a threshold to obtain the correspondent code. For each element in H we apply the sign function:

$$sign(H^i) = \begin{cases} 1, \text{if } H^i > 0.5 \\ 0, \text{otherwise} . \end{cases} \quad (8)$$

Experiments and Results

We compare DRHN with n=15 against four unsupervised methods LSH [1], SH [8] and ITQ [2], and seven supervised methods CNNH [9], CNNH+ [9], KSH [6], MLH [7], BRE[4], ITQ-CCA [2] and DLBHC [5] in the retrieval task in CIFAR-10 dataset using mean average precision as evaluation metric (Eq. 10).

$$P@K = \frac{\sum_{i=1}^{k} In(i)}{k} \quad (9)$$

$$mAP = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{m_q} \sum_{k=1}^{m_q} P@K \text{ (if } k^{th} \text{ item was relevant)} \quad (10)$$

where $In(i)$ return 1 if the retrieved image and the query image q have the same class label and 0 otherwise, $|Q|$ represent the number of queries and m_q is the number of result images for a given query q.

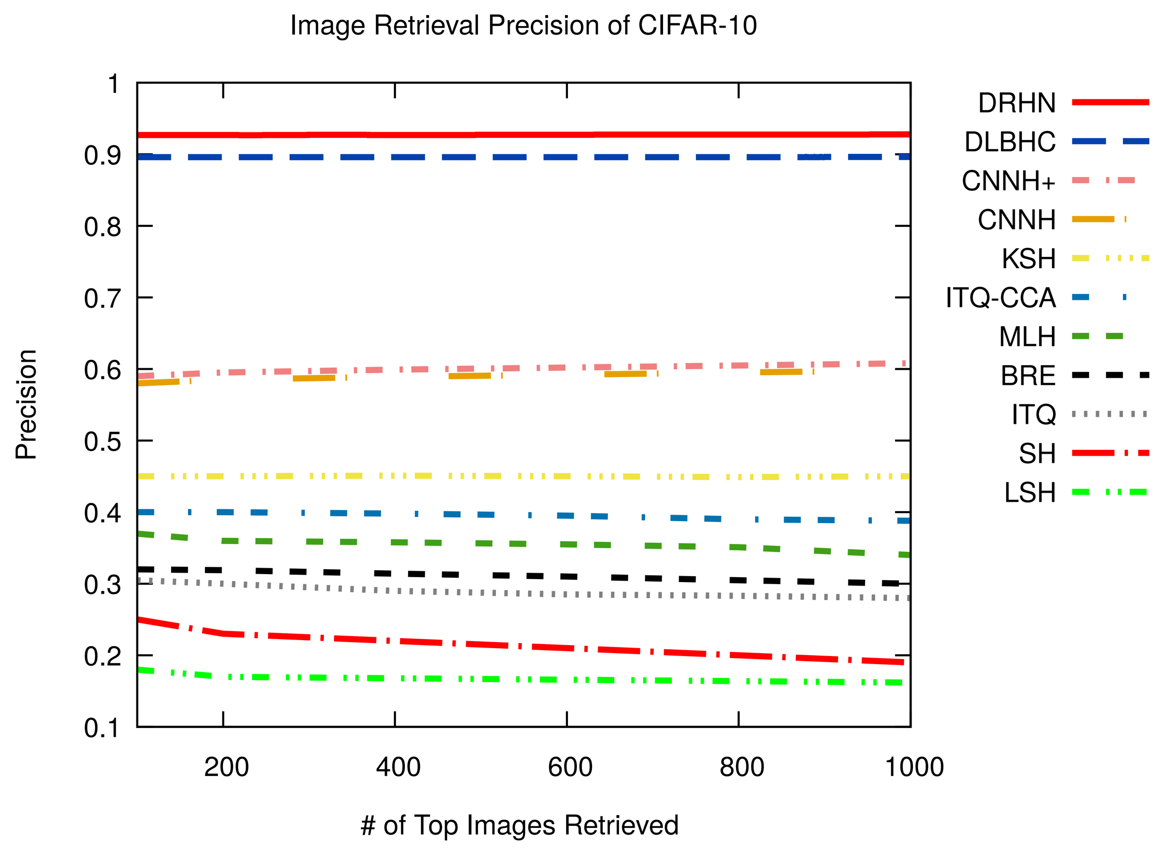

Figure 2: Image retrieval precision with 48 bits on CIFAR-10 dataset. DRHN-15

Method	12 bits	32 bits	48 bits
DRHN-15	**92.65**	**92.23**	**92.91**
DLBHC	89.3	89.72	89.73
CNNH+	46.5	52.1	53.2
CNNH	43.9	50.9	52.2
KSH	30.3	34.6	35.6
ITQ-CCA	26.4	28.8	29.5
LSH	12.1	12.0	12.0

Table 2: mAP comparison of different hashing methods on CIFAR-10 dataset.

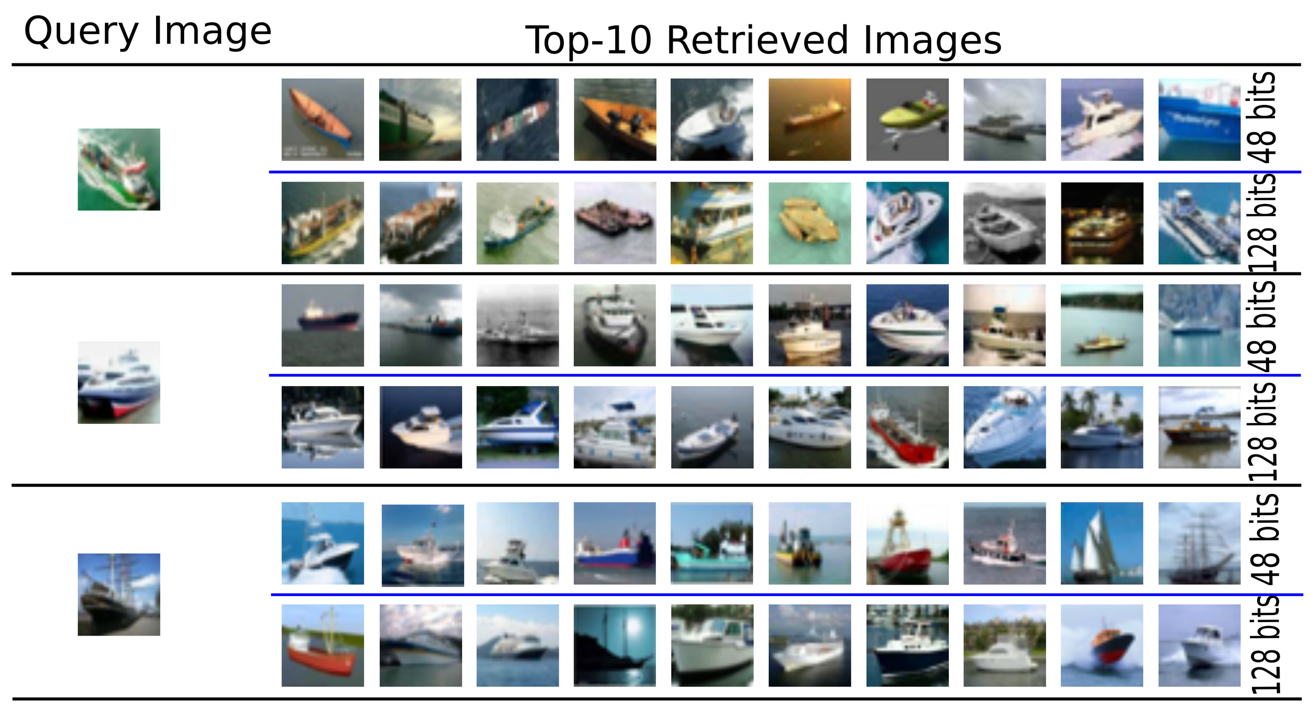

Figure 3: Results for queries with boat images in CIFAR-10 dataset.

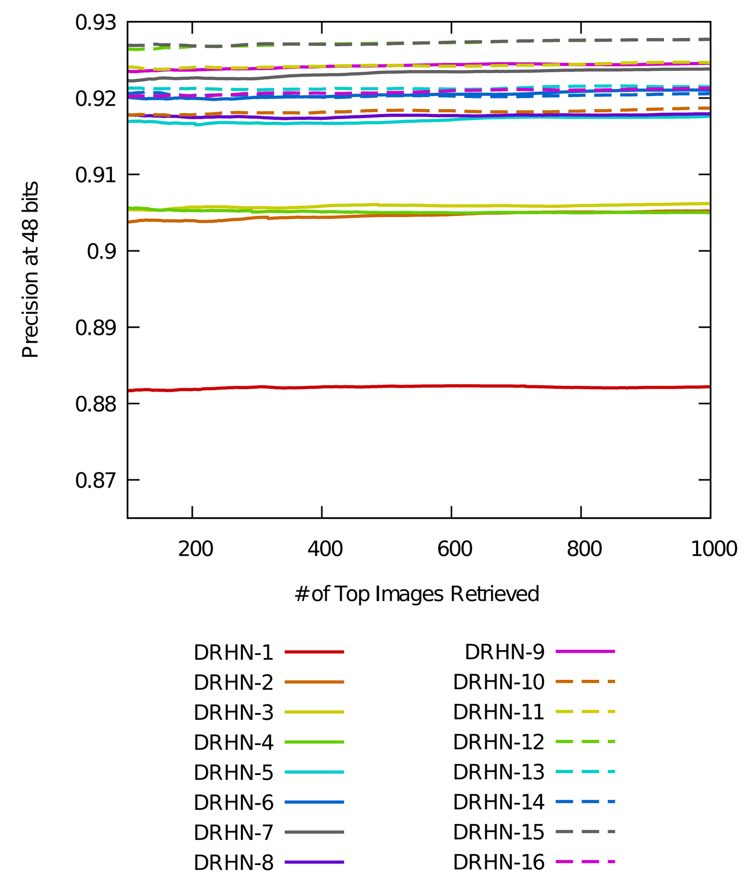

Figure 4: Image retrieval precision with 48 bits on CIFAR-10 dataset with DRHN.

Table 3: mAP comparison of different depth in our method on CIFAR-10 dataset.

Method	12 bits	24 bits	32 bits	48 bits
DRHN-1	87.50	88.01	88.59	88.43
DRHN-2	90.40	90.22	90.62	90.66
DRHN-3	90.81	91.04	90.99	90.78
DRHN-4	91.32	90.83	91.43	90.71
DRHN-5	91.75	91.21	91.21	91.89
DRHN-6	91.85	92.49	91.36	92.23
DRHN-7	91.76	91.74	92.30	92.50
DRHN-8	92.17	92.03	92.49	91.95
DRHN-9	91.97	91.92	92.35	92.58
DRHN-10	91.93	92.55	**92.94**	91.95
DRHN-11	**92.66**	92.35	92.57	92.57
DRHN-12	91.55	92.71	92.78	92.89
DRHN-13	92.52	92.58	92.18	92.30
DRHN-14	91.78	**92.74**	92.93	92.19
DRHN-15	92.65	92.02	92.23	**92.91**
DRHN-16	92.53	92.32	92.15	92.25

References

[1] Gionis, A., Indyk, P., & Motwani, R. (1999, September). Similarity search in high dimensions via hashing. In VLDB (Vol. 99, No. 6, pp. 518-529).

[2] Gong, Y., Lazebnik, S., Gordo, A., & Perronnin, F. (2013). Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(12), 2916-2929.

[3] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[4] Kulis, B., & Darrell, T. (2009). Learning to hash with binary reconstructive embeddings. In Advances in neural information processing systems (pp. 1042-1050).

[5] Lin, K., Yang, H. F., Hsiao, J. H., & Chen, C. S. (2015). Deep learning of binary hash codes for fast image retrieval. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 27-35).

[6] Liu, W., Wang, J., Ji, R., Jiang, Y. G., & Chang, S. F. (2012, June). Supervised hashing with kernels. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on (pp. 2074-2081). IEEE.

[7] Norouzi, M., & Blei, D. M. (2011). Minimal loss hashing for compact binary codes. In Proceedings of the 28th international conference on machine learning (ICML-11) (pp. 353-360).

[8] Weiss, Y., Torralba, A., & Fergus, R. (2009). Spectral hashing. In Advances in neural information processing systems (pp. 1753-1760).

[9] Xia, R., Pan, Y., Lai, H., Liu, C., & Yan, S. (2014, July). Supervised Hashing for Image Retrieval via Image Representation Learning. In AAAI (Vol. 1, pp. 2156-2162).

[10] Yoo, H. W., Jang, D. S., Jung, S. H., Park, J. H., & Song, K. S. (2002). Visual information retrieval system via content-based approach. Pattern Recognition, 35(3), 749-769.

Acknowledgments. We appreciate the financial support given by CONACYT.