Supplementary file : Deep Hashing with Hash Center Update for Efficient Image Retrieval

Ablation study: The loss function used for training our model is $p*L_{hash}+ q*L_{class}$. When p is low, L_{class} is emphasized in training and the retrieval results are also lower. In a similar way when only L_{hash} is emphasized, i.e q = 1e^5, the results are lower but better than the case when L_{class} is emphasized. This indicates that the gain in performance comes from the L_{hash} and hash center update. In the paper we used q= B-1 /C-1 , and p = 1 which gives equal contribution for both the losses making the theoretical lower bound same. An emphasis on L_{hash} could even increase the performance especially for the single-labeled dataset as with the current setting of q = B-1 / C-1 , L_{class} is slightly emphasized. For the multi-labeled datasets, if the bit length B <= C, L_{hash} will be emphasized in training. Some additional results for a different setting of q is given in Table 1. Surprisingly, higher value of q = 10, deteriorates the performance for larger bit lengths. Also, when both hyperparameters are 1 the retrieval results are better. When q = 0.5, the results are still better indicating the importance of L_{hash}. Also, note that an ablation study making either of the loss function to zero is not meaningful as the hash center update is dependent on the L_{class} and using L_{class} alone will be again a basic classifier.

Bits	q = 1e^5	q = 0.5	q = 1	q = 10	P = 1e^5
16	0.7321	0.7786	0.7945	0.8033	0.6991
32	0.7924	0.8362	0.8399	0.8166	0.7596
48	0.8108	0.8397	0.8457	0.8340	0.8004
64	0.8170	0.8549	0.8480	0.8402	0.8080

Table 1: MAP for ablation study in MS-COCO.

Network architecture: The network architecture of the proposed approach is given in Table 2. As explained in the paper, we used ResNet-50 as the feature extractor followed by hashing layer, intermediate layer and classification layer which are fully connected layers. The hashing layer contains a sigmoid activation function which squashes the output of neural network into the range of 0 and 1. The intermediate layer contains 4096 nodes and is needed since the loss function L_{DCSH} in Eq. 2 is a dimensionality reduction method and is followed by the classification layer.

Layer	Details
ResNet-50	Convolutional layers
Hashing layer	FC (2058 * bits), Sigmoid activation
Intermediate layer	FC (bits * 4096) , ReLU activation
Classification layer	FC (4096 * classes), Sigmoid activation

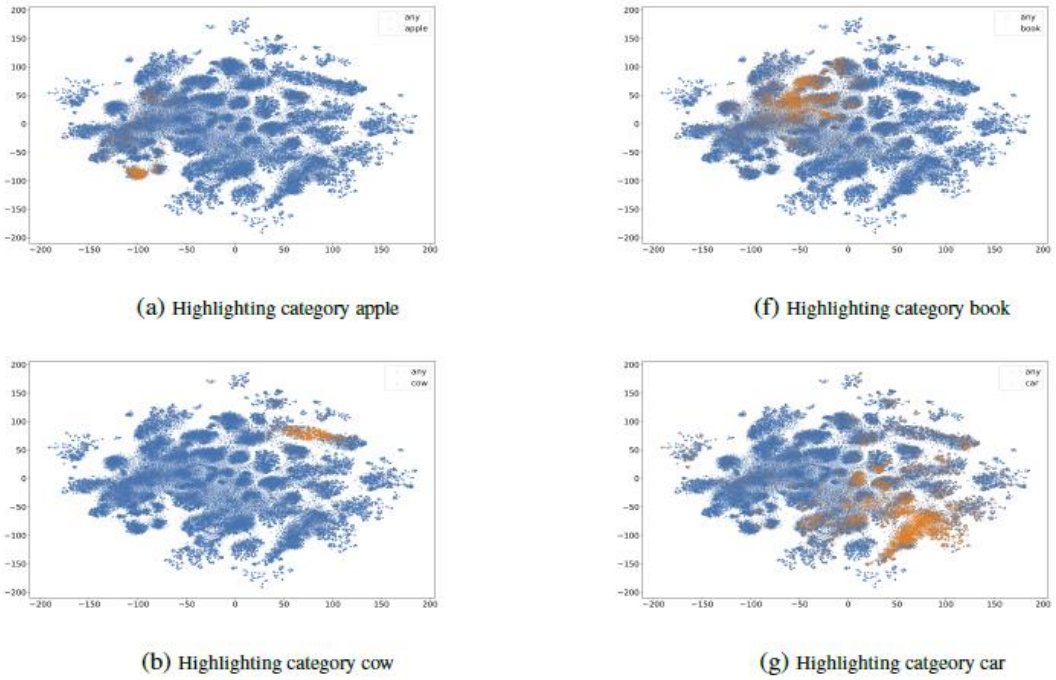
Table 2. Network architecture of the proposed approach.

T-sne visualization: The t-sne visualizations for additional class categories is given in Fig. 1 for MS-COCO. Two different scenarios are mainly depicted. Images in the left column indicate the categories 'apple', and 'cow' which are clustered together. The categories 'book', and 'car' are some categories shared by few images and hence a strong cluster for these classes are not formed. The t-sne visualizations for

NUS-WIDE is given in Fig. 2. categories such as 'coral' form a strong cluster in the feature space. The categories such as 'beach', 'ocean' and 'mountain' are shared by few images and hence a strong cluster is not formed for these classes.

Training and test losses: The training and test loss for different bit combinations for the two datasets is given in the first two rows in Fig. 3. For multi-labeled case, the theoretical lower bound of $-2(B-1)$ cannot be reached since the correlation coefficients cannot be 1 as the images belong to more than one category. This is because the correlation coefficient can only be 1 when the angle between the feature vectors and labels are 0 which is only possible for single-labeled datasets

Example queries: In this section a few example queries are presented by using the proposed Deep Central Similarity Hashing (DCSH) method. After obtaining the final binary codes of an image dataset, queries can be performed by applying a nearest neighbor search in Hamming space. Therefore, the binary feature representation of a given query image would be compared to each binary code of the retrieval in terms of its Hamming distance. Those images from the gallery set which have the smallest Hamming distance to the query in the binary feature space are retrieved. Final query examples are performed on the same benchmark datasets. Retrieval results on a single labeled dataset CIFAR-10 is given in Fig. 4. Retrieval results on MS-COCO are given in Figs. 5 and 6 and retrieval results on NUS-WIDE in Fig. 7.



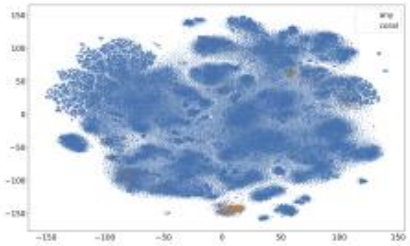
(a) Highlighting category apple

(f) Highlighting category book

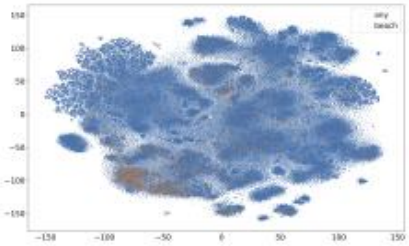
(b) Highlighting category cow

(g) Highlighting catgeory car

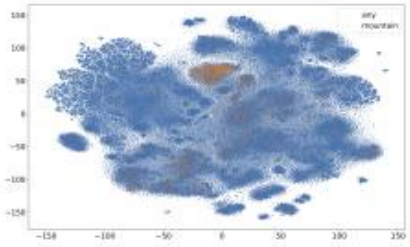
Fig 1. T-SNE visualization for image in MS-COCO dataset for categoreis 'apple', 'cow', 'book', and 'car'.



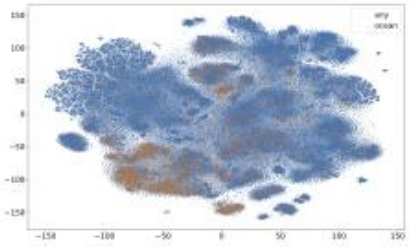
(a) Highlighting category coral



(f) Highlighting category beach

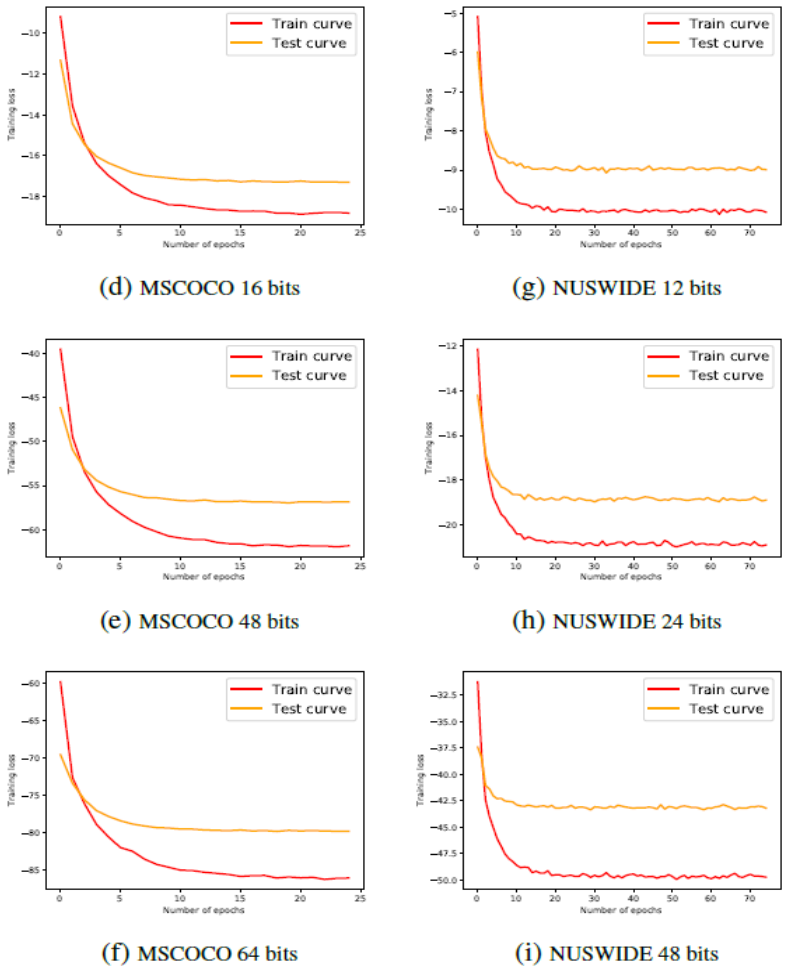


(b) Highlighting category mountain



(g) Highlighting catgeory ocean

Fig 2. T-SNE visualization for image in NUS-WIDE dataset for categoreis ‘coral’, ‘mountain’, ‘beach’, and ‘ocean’.



(d) MSCOCO 16 bits (g) NUSWIDE 12 bits

(e) MSCOCO 48 bits (h) NUSWIDE 24 bits

(f) MSCOCO 64 bits (i) NUSWIDE 48 bits

Fig 3. Training and test curves for different bit combinations for the MS-COCO and NUS-WIDE datasets.

(a) Query image 1 (b) Top 10 retrieval results

Figure 4: Retrieval results on CIFAR-10 with category label 'airplane'.

(a) Query image 1 (b) Top 3 retrieval results

Figure 5: Retrieval results on MS-COCO with category labels 'keyboard', 'mouse', and 'tv'.

(a) Query image 2 (b) Top 3 retrieval results

Figure 6: Retrieval results on MS-COCO with category labels 'bus', 'car, 'person', and 'truck'.

(a) Query image 1 (b) Top 3 retrieval results

Figure 7: Retrieval results on NUS-WIDE with category labels 'buildings', 'clouds', 'sky', and 'water'.