Method	NUS-WIDE			ImageNet			MS-COCO		
	n_h	map_l	map_i	n_h	map_l	map_i	n_h	map_l	map_i
DSEH	2008	**0.9664**	**0.7312**	100	**1.0000**	**0.6137**	1903	**0.8276**	**0.6048**
DSEH-S	1959	0.9633	0.7208	100	1.0000	0.5988	1933	0.8260	0.5907
DSEH-SS	1164	0.9322	0.7013	98	0.9825	0.5681	1226	0.7452	0.5237
DSEH-L	1036	0.9607	0.7251	100	1.0000	0.6070	802	0.8199	0.5915
DSEH-A	1684	0.9558	0.7234	100	1.0000	0.5576	1574	0.8134	0.5850

Table 2: The results of ablation study @ 32bits of our DSEH.

form the traditional hashing baselines, which highlights the benefit of feature learning by deep networks that more discriminative representation can be obtained. Compared with other deep methods which utilize similarity pairs, DSEH achieves a substantial increase in average MAP at different code lengths. All the results shown in Table 1, Fig. 3 and Fig. 4 illustrate the superiority of our method. One reason may be that instead of utilizing similarity pairs information roughly, DESH exploring label information to generate semantic feature is very effective to generate more sufficient semantic information and thus produce more discriminative hash codes. Another reason is that sufficient semantic information obtained from *LabNet* can be retained completely and thus supervise *ImgNet* effectively when training *ImgNet* with the supervised information on the semantic level and hash codes level.

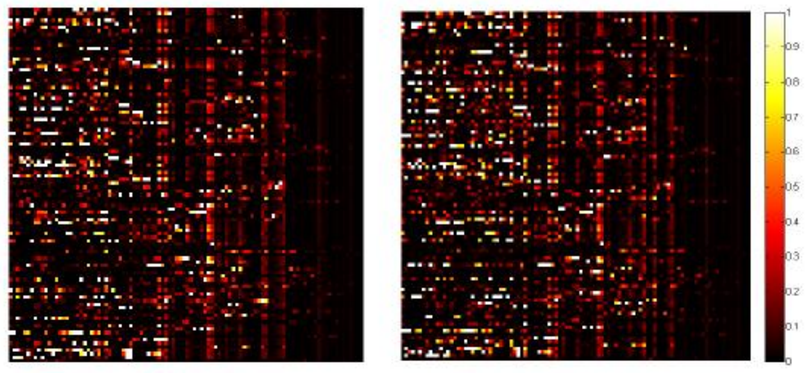
On ImageNet dataset which is annotated with single label. DHN, DPSH, and CNNH achieve under-performing results compared with the shallow baseline SDH, which demonstrates that network learning capacity can be dropped on single-label dataset because of the imbalance of pairs similarity. CNNH generates undiscriminating hash codes only under the supervision of pairwise similarity matrix. By adjusting the weight of similarity correlation, HashNet outperforms other baselines, which shows that adjusting weight can only alleviate influence of the data imbalance. The proposed DSEH significantly outperforms all other baselines. Compared with the state-of-the-art HashNet, we achieve about 34.50% increase in average MAP at different code lengths on this imbalanced dataset. It means that the proposed semantic feature learning and supervision to hashing learning can solve the issue of data imbalance in single-label dataset and thus hash codes can be generated more discriminative.

4.3 Empirical Analysis

Two different experiment settings are designed additionally to analyse the proposed method.

Visualization of Semantic Features: We visualize the semantic features generated by *LabNet* and *ImgNet* on NUS-WIDE at 32 bits in Fig. 5 (for convenience, 100 points are sampled and encapsulated by PCA [Wold *et al.*, 1987]). We observe that the semantic features of *LabNet* are abundant, indicating that the semantic information of labels is effectively exploited. Furthermore, the semantic features of *ImgNet* are similar to those in *LabNet*, inferring that *ImgNet* is well supervised in the common semantic space.

Ablation Study: We investigate the variants of DSEH on the three datasets. **DSEH-S** denotes that *ImgNet* without supervision on semantic layer from *LabNet*. **DSEH-SS** refers to that both *LabNet* and *ImgNet* without semantic supervision.



(a) *LabNet* (b) *ImgNet*

Figure 5: The visualization of semantic features.

DSEH-L denotes that *LabNet* drops direct label supervision. **DSEH-A** refers to that *LabNet* and *ImgNet* are trained only once without alternating manner.

Tabel 2 shows the average results of 10 runs of DSEH variants, where n_h is the total number of hash codes generated from *LabNet*, map_l is the MAP of retrieving labels with hash codes generated by *LabNet*, and map_i is the MAP of retrieving images with the hash codes generated by *ImgNet*. DSEH outperforms all of its variants, which shows the effectiveness of each module. DSEH-SS achieves the worst performance, the main reason of which is that semantic supervision plays a very important role in the proposed framework. It is noted that the higher n_h is, the more diverse hash codes can be generated. DSEH-L reduces n_h dramatically, illustrating that more semantic information can be maintained by adding label supervision to the proposed method.

5 Conclusion

In this paper, we proposed a novel deep hashing method, namely DSEH, for image retrieval, which consists of *LabNet* and *ImgNet*. The *LabNet* is explored to discover abundant semantic correlation and generate accurate hash codes. Meanwhile, the *ImgNet* is jointly constrained with the supervision information from common semantic space and common Hamming space for generating similarity-preserving yet discriminative hash codes. Extensive experiments conducted on three widely-used datasets demonstrate that our proposed method significantly outperforms many state-of-the-art hashing approaches, including both traditional and deep learning-based ones.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61572388 and Grant 61703327, and in part by the Key R&D Program-The Key Industry Innovation Chain of Shaanxi under Grant 2017ZDCXL-GY-05-04-02 and Grant 2017ZDCXL-GY-05-04-02.