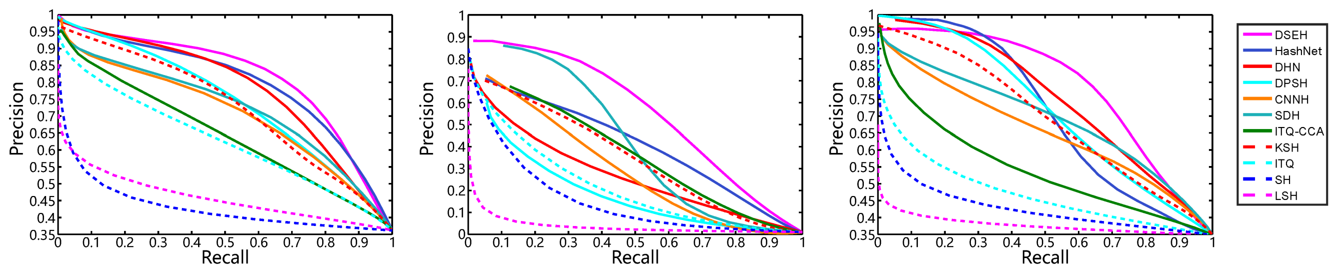
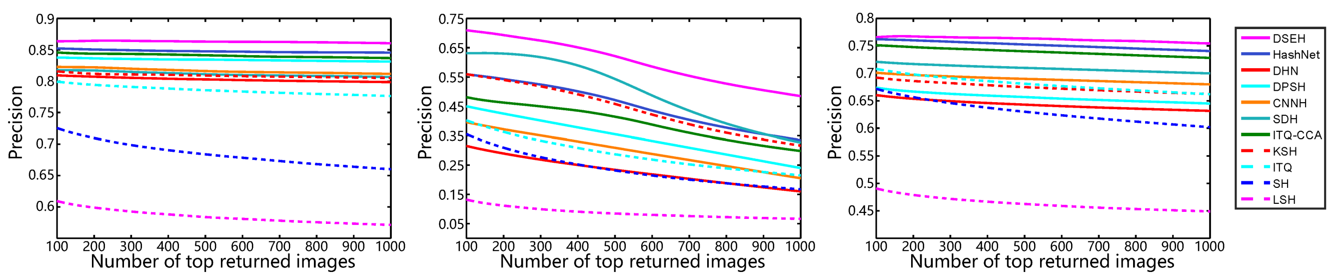
Method	NUS-WIDE				ImageNet				MS-COCO			
	16 bits	32 bits	48 bits	64 bits	16 bits	32 bits	48 bits	64 bits	16 bits	32 bits	48 bits	64 bits
DSEH	**0.7119**	**0.7312**	**0.7372**	**0.7422**	**0.5278**	**0.6137**	**0.6397**	**0.6548**	**0.5897**	**0.6048**	**0.6133**	**0.6188**
HashNet	0.7007	0.7275	0.7301	0.7374	0.3260	0.4563	0.5018	0.5270	0.5600	0.5850	0.5989	0.6056
DHN	0.6512	0.6611	0.6675	0.6741	0.1838	0.2344	0.2375	0.2564	0.5353	0.5456	0.5486	0.5555
DPSH	0.6902	0.7049	0.7130	0.7158	0.2730	0.2841	0.3111	0.3242	0.5618	0.5774	0.5857	0.5901
CNNH	0.6573	0.6601	0.6716	0.6781	0.2488	0.3047	0.3263	0.3387	0.5115	0.5232	0.5283	0.5328
SDH	0.6488	0.6703	0.6811	0.6857	0.3687	0.4292	0.4446	0.4600	0.5312	0.5632	0.5634	0.5741
ITQ-CCA	0.6125	0.6472	0.6655	0.6766	0.2312	0.4061	0.4316	0.4568	0.5418	0.5658	0.5704	0.5715
KSH	0.6404	0.6636	0.6689	0.6731	0.3064	0.3874	0.4006	0.4168	0.5496	0.5574	0.5628	0.5688
ITQ	0.5715	0.5876	0.5910	0.5985	0.1668	0.2452	0.2929	0.3184	0.4834	0.4993	0.5111	0.5153
SH	0.4459	0.4504	0.4342	0.4244	0.1194	0.1776	0.2143	0.2335	0.4494	0.4400	0.4397	0.4316
LSH	0.4624	0.4431	0.4433	0.4816	0.0278	0.0526	0.0720	0.0966	0.3718	0.3807	0.3945	0.4119

Table 1: Mean Average Precision(MAP) of Hamming Ranking on three benchmark datasets.



(a) NUS-WIDE (b) ImageNet (c) MS-COCO

Figure 3: Precision-recall curves @ 32bits of our method and comparison methods on three benchmark datasets.



(a) NUS-WIDE (b) ImageNet (c) MS-COCO

Figure 4: Precision w.r.t. top returned samples curves @ 32bits of our method and comparison methods on three benchmark datasets.

LSH [Gionis *et al.*, 1999], **SH** [Weiss *et al.*, 2009], **ITQ** [Gong *et al.*, 2013], supervised shallow methods **KSH** [Liu *et al.*, 2012], **ITQ-CCA** [Gong *et al.*, 2013], **SDH** [Shen *et al.*, 2015], and deep supervised methods **CNNH** [Xia *et al.*, 2014], **DPSH** [Li *et al.*, 2015], **DHN** [Zhu *et al.*, 2016], **HashNet** [Cao *et al.*, 2017].

For fair comparison, we extract $4,096$-dimensional deep features by CNN-F [Chatfield *et al.*, 2014] model which is retrained on ImageNet dataset. We construct *ImgNet* to reserve first seven layers same with those in CNN-F followed $fc8$ with 512 nodes for semantic layer and K nodes for hash layer, *i.e.*, $(I \rightarrow CNNF \rightarrow 512 \rightarrow K)$. *LabNet* is initialized randomly and constructed as $(L \rightarrow 4096 \rightarrow 512 \rightarrow K \rightarrow c)$, which contains c nodes for total class labels.

Since the semantic layer and hash layer are trained from scratch, we set its learning rate 10 times of the ones for the other layers. The learning rate is chosen from 10^{-2} to 10^{-6} with a validation set. The batch size of *LabNet* and *ImgNet* are set to 32 and 128 respectively. Since the semantic correlation of ImageNet is sparse, we set the values in similarity matrix as $\mathcal{S} \in \{0, 5\}$. For the hyper-parameters in *LabNet*, we conduct cross-validation to search α and γ from 10^{-3} to 10^{2}, and search β from 10^{-6} to 10^{-1}. We find that the optimal result can be obtained when $\alpha = \gamma = 1$, and $\beta = 0.005$. Then we search from 10^{-3} to 10^{2} and discover $\eta = 1$ is the best for *ImgNet*. It is noted that the parameter searching operations are performed with the searching step set to 5. Our model is implemented on **TensorFlow** [Abadi *et al.*, 2016] on a server with two NVIDIA TITAN X GPUs.

4.2 Results and Discussions

Table 1 shows the results of different hashing methods on three benchmark datasets when the code length is 16, 32, 48, and 64 bits respectively. Fig. 3 and Fig. 4 show the Precision-Recall curves and Precision curves respectively for different methods on the code length of 32 bits.

On two multi-label datasets NUS-WIDE and MS-COCO, DSEH substantially outperforms all the compared baseline methods. Besides, almost all deep hashing methods outper-