ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Improved Deep Classwise Hashing With Centers Similarity Learning for Image Retrieval

Ming Zhang (Presenter) and Hong Yan

Department of Electrical Engineering, City University of Hong Kong

CityU
香港城市大學
City University of Hong Kong

Outline

Background

Related works

Our method

Experiments and results

Background

- Hashing for image retrieval:

Map high dimensional data into lower dimensional binary codes $\{-1,1\}^l$ in Hamming space while preserving the similarity relation in original space

  - Extremely fast query speed
  - Low storage cost

- Deep supervised hashing:

Apply DNNs to integrate features extraction and hashing learning end-to-end

  - Pairwise labels-based: DPSH [W.-J. Li et al. AAAI17], DSDH [Q. Li et al. NIPS17] …
  - Triplet labels-based: DTSH [X. Wang et al. ACCV16] …
  - Class labels-based: DCWH [X. Zhe et al. TNNLS20], CSQ [L. Yuan et al. CVPR20] …

Outline

Background

Related works

Our method

Experiments and results

Related Works

- Pairwise/triplet labels-based methods:
  - Tremendous computation cost
    - Time complexity $O(n^2)$
    - Compromise: Sample a portion of data for training
  - Only use the local data structure
  - Hard to capture the global similarity relation

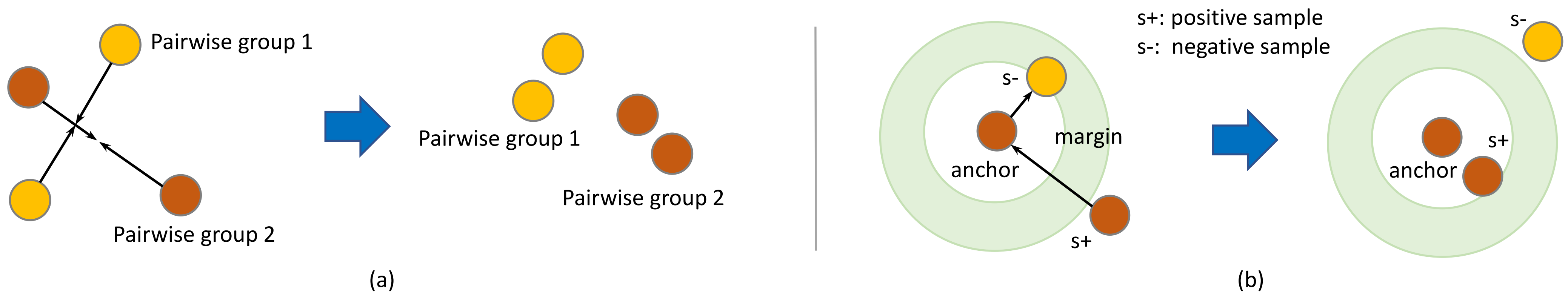

Fig. 1. (a) Pairwise labels-based learning metric (b) Triplet labels-based learning metric

Related Works

- Deep Classwise Hashing (DCWH)
  - Cross-entropy based normalized Gaussian
  - Minimize the error assigning intra-class samples to the corresponding class center

$$\min_{\Theta,M} L_{clw} = -\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ji} \log \frac{\exp\left\{-\frac{\|h_i-\mu_j\|^2}{2\sigma^2}\right\}}{\sum_{j=1}^{C}\exp\left\{-\frac{\|h_i-\mu_j\|^2}{2\sigma^2}\right\}} \quad s.t.\ h_i = b_i, \quad h_i \in \mathbb{R}^l, \quad b_i = \{-1,1\}^l$$

C	Number of class
N	Number of training samples $\{x_i\}_{i=1}^N$
y_{ji}	Label indicator of j-th class for i-th sample
h_i	Hashing output of x_i
$\mu_j = \frac{1}{n_j}\sum_{i=1}^{N} y_{ji}\, h_i$	j-th class center, updated periodically from the intra-class outputs

ICPR 2020 25th International Conference on Pattern Recognition Milan, Italy 10 \| 15 January 2021

香港城市大學 City University of Hong Kong

Outline

Background

Related works

Our method

Experiments and results

Our method

Motivation

- Problem of DCWH:

Samples belonging to different classes but lying far from their corresponding centers are vulnerable to be closer to each other than their intra-class counterparts

- Solution:

Increase the Hamming distance between pairwise class centers for more separable inter-class distribution

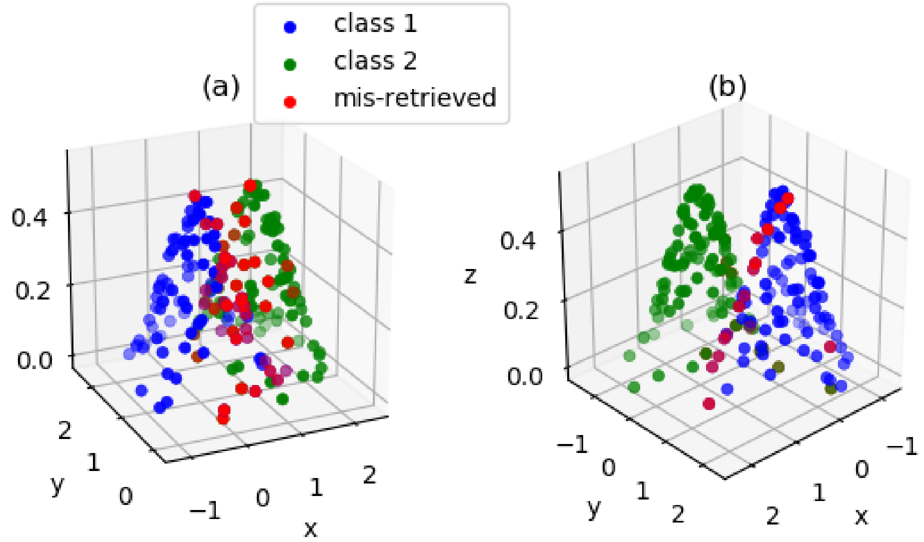

Fig. 2. Draw two-class datapoints from two independent Gaussian distributions with centers/mean: (0,1) and (1,1) in (a), whereas (0,1) and (1,0) in (b).

Our method

Improved Classwise Hashing with Center Similarity Loss (IDCWH)

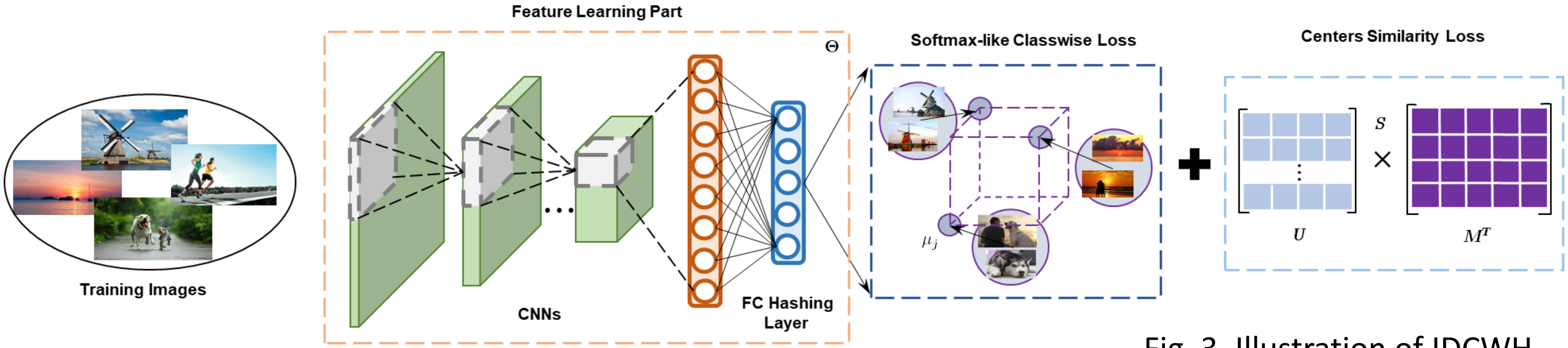

Fig. 3. Illustration of IDCWH.

Θ	Network parameters
$M = \{\mu_j\}_{j=1}^{C}$	Learnable class centers
$U = \{u_k\}_{k=1}^{Z}$	Estimated binary center of each unique label in a mini-batch
$s_{ij} = \begin{cases} 1, & y_{u_i}^T y_{\mu_j} = 1 \\ 0, & y_{u_i}^T y_{\mu_j} = 0 \end{cases}$	Similarity matrix describing relations between U and M

Our method

Two-step Centers Similarity Learning:

a. Intra-class samples clustering:

Estimate each center with binary constraint from intra-class binary codes

$$\min_{u_j} \sum_{i=1}^{N} \sum_{j=1}^{C} y_{ji} \parallel u_j - b_i \parallel_2^2 \quad s.t. \;\; u_j \in \{-1,1\}^l$$

$$u_{j,v} = \mathbf{sgn}(m_j) = \begin{cases} 1, & m_{j,v} \geq 0 \\ -1, & m_{j,v} < 0 \end{cases} \text{ where } m_j = \sum_{k=1}^{n_j} b_{jk}$$

- Dynamically attract each class center to concentrate on corresponding intra-class samples

Our method

Two-step Centers Similarity Learning:

b. Center-sample concentration and repelling:

$$p(s_{ij}|u_i;\mu_j) = \begin{cases} \sigma(\theta_{ij}), & s_{ij}=1 \\ 1-\sigma(\theta_{ij}), & s_{ij}=0 \end{cases}$$

where $\theta_{ij} = 0.5l\cos(u_i,\mu_j)$, $\sigma(\cdot)$ is sigmoid function

$$\min_M L_{csl} = -\log \prod_{\substack{i=1,\dots,Z \\ j=1,\dots,C}} p(s_{ij}|u_i;\mu_j) = -\sum_{s_{ij}\in S} (s_{ij}\theta_{ij} - \log(1+e^{\theta_{ij}}))$$

- Minimize the distance between estimated center and learnable class center while maximizing the distance between intra-class samples and centers belonging to other classes

Our method

- Relax h_i to be continuous and introduce $L_{quant.}$ between $b_i = \mathbf{sgn}(h_i)$ and h_i

- Finalized loss function:

$$\min_{\Theta,M} L_{clw} + \gamma L_{csl} + \beta L_{quant.} =$$

$$-\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ji} \log \frac{exp\{-\frac{\| h_i - \mu_j \|^2}{2\sigma^2}\}}{\sum_{j=1}^{C} exp\{-\frac{\| h_i - \mu_j \|^2}{2\sigma^2}\}} - \gamma \sum_{s_{ij}\in S} s_{ij}\theta_{ij} - \log\left(1 + e^{\theta_{ij}}\right) + \beta \sum_{i=1}^{N} \|b_i - h_i\|_2^2$$

Outline

25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Background

Related works

Our method

Experiments and results

香港城市大學
City University of Hong Kong

Experiments and Results

Datasets and Experiments Settings:

Dataset	Training	Query	Remarks
CIFAR-10[1]	10 classes; randomly sample 500/class; total 500	Randomly sample 100/class; total 1,000	*Mini* setting; the rest images are all used as database; Follow the protocol in DPSH
	Official training split: 5,000/class; total 50,000	Official test split: 1,000/class; total 10,000	*Full* setting, the training images are used as database
CIFAR-100[2]	100 classes; Official training split: 500/class; total 50,000	Official test split: 100/class; total 10,000	Follow the protocol in DCWH
MS-COCO[3]	Combine 'train2014' and 'val2014'; Randomly sample 10,000	Randomly sample 5,000	Labeled with 80 semantic concepts; The remainder are all used as database; Follow the protocol in HashNet

[1]https://www.cs.toronto.edu/~kriz/cifar.html
[2]https://www.cs.toronto.edu/~kriz/cifar.html
[3]https://cocodataset.org/#home

Experiments and Results

Results on Mean Average Precision (MAP)

- Hyper parameters: $\sigma^2 = 4, \gamma = 1$ and $\beta = 0.01$
- Optimized by SGD with momentum 0.9, weight decay 5e-4
- Learning rate: 1e-2 and 5e-3 for feature learning and centers learning, respectively. Decay by 0.1 every 50 epochs
- Train for 150 epochs with batch size fixed to 128

TABLE I
MAP RESULTS BY HAMMING RANKING ON CIFAR-10 UNDER TWO EXPERIMENT PROTOCOLS

Method	CIFAR-10 (mini)				CIFAR-10 (full)			
	12 bits	**24 bits**	**32 bits**	**48 bits**	**12 bits**	**24 bits**	**32 bits**	**48 bits**
SDH+CNN	0.207	0.218	0.223	0.210	0.364	0.433	0.405	0.414
FSDH+CNN	0.196	0.220	0.203	0.212	0.374	0.443	0.410	0.446
DPSH	0.713	0.727	0.744	0.757	0.763	0.781	0.795	0.807
DPSH*	0.797	0.806	0.820	0.802	0.908	0.922	0.925	0.935
DTSH	0.710	0.750	0.765	0.774	0.915	0.923	0.925	0.926
DTSH*	0.790	0.797	0.794	0.775	0.928	0.935	0.940	0.942
DSDH	0.740	0.786	0.801	0.820	0.935	0.940	0.939	0.939
DSDH*	0.800	0.802	0.804	0.808	0.913	0.925	0.943	0.930
DCWH	0.818	0.840	0.848	**0.854**	0.940	0.950	0.954	0.952
IDCWH	**0.828**	**0.865**	**0.868**	0.849	**0.964**	**0.969**	**0.967**	**0.968**

TABLE II
MAP RESULTS BY HAMMING RANKING ON CIFAR-100 DATASET

Method	CIFAR-100			
	12 bits	**24 bits**	**32 bits**	**48 bits**
SDH+CNN	0.0617	0.0624	0.0610	0.0668
FSDH+CNN	0.0596	0.0618	0.0650	0.0665
DPSH	0.0597	0.1008	0.1196	0.1587
DTSH	0.6070	0.7056	0.7122	0.7252
DSDH	0.0784	0.1495	0.1868	0.2272
DCWH	0.7227	0.7441	0.7570	0.7658
IDCWH	**0.7642**	**0.8130**	**0.8236**	**0.8351**

TABLE III
MAP RESULTS BY HAMMING RANKING ON MS-COCO DATASET

Method	MS-COCO			
	16 bits	**32 bits**	**48 bits**	**64 bits**
DPSH	0.3493	0.3545	0.3595	0.3670
DSDH	0.3470	0.3587	0.3661	0.3703
DNNH	0.5932	0.6034	0.6045	0.6099
DHN	0.6774	0.7013	0.6948	0.6944
HashNet	0.6873	0.7184	0.7301	0.7362
DCWH	0.7227	0.7441	0.7570	0.7658
IDCWH	**0.7321**	**0.7597**	**0.7636**	**0.7698**

*Replace the backbone applied in the original works with our employed GoogleNet

Experiments and Results

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Results on P@H=2, R@H=2, P@N and PR curve

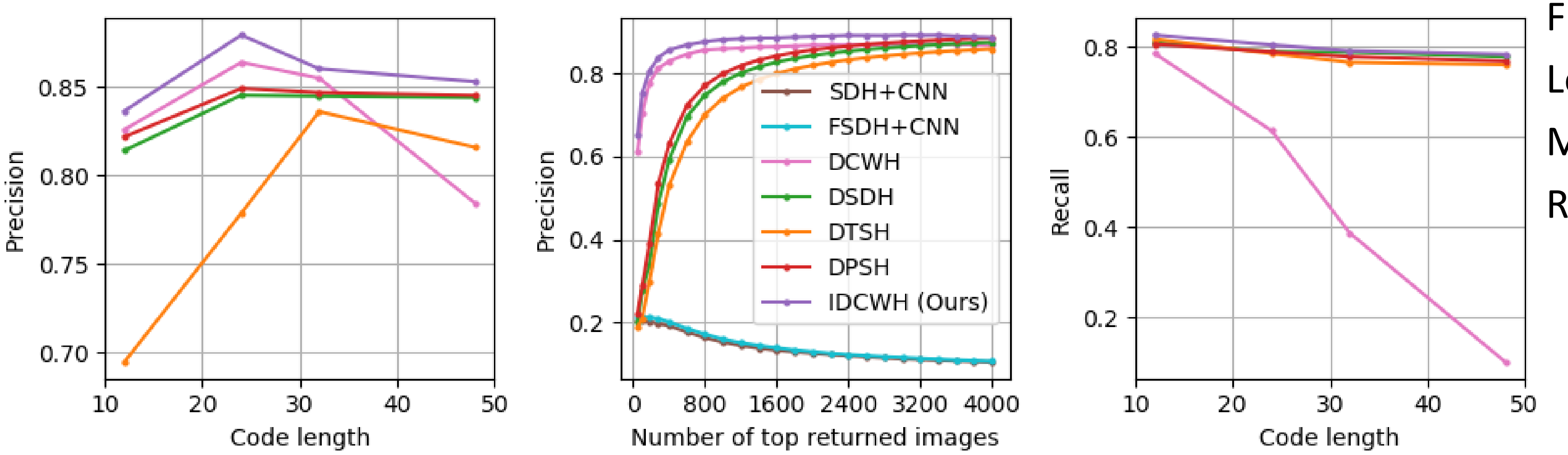

Fig. 4. Results on CIFAR-10 dataset.
Left: P@H=2 w.r.t. different code lengths
Middle: P@N with 32-bit codes
Right: R@H=2 w.r.t. different code lengths

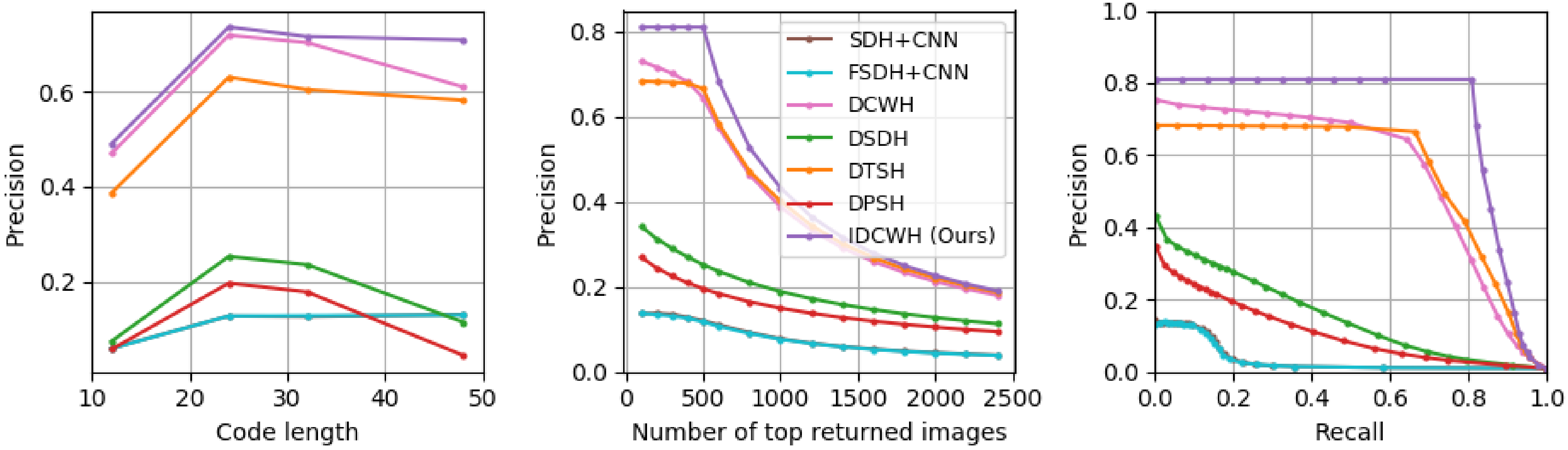

Fig. 5. Results on CIFAR-100 dataset.
Left: P@H=2 w.r.t. different code lengths
Middle: P@N with 48-bit codes
Right: PR-curve with 48-bit codes

CityU
香港城市大學
City University of Hong Kong

Experiments and Results

Visualization

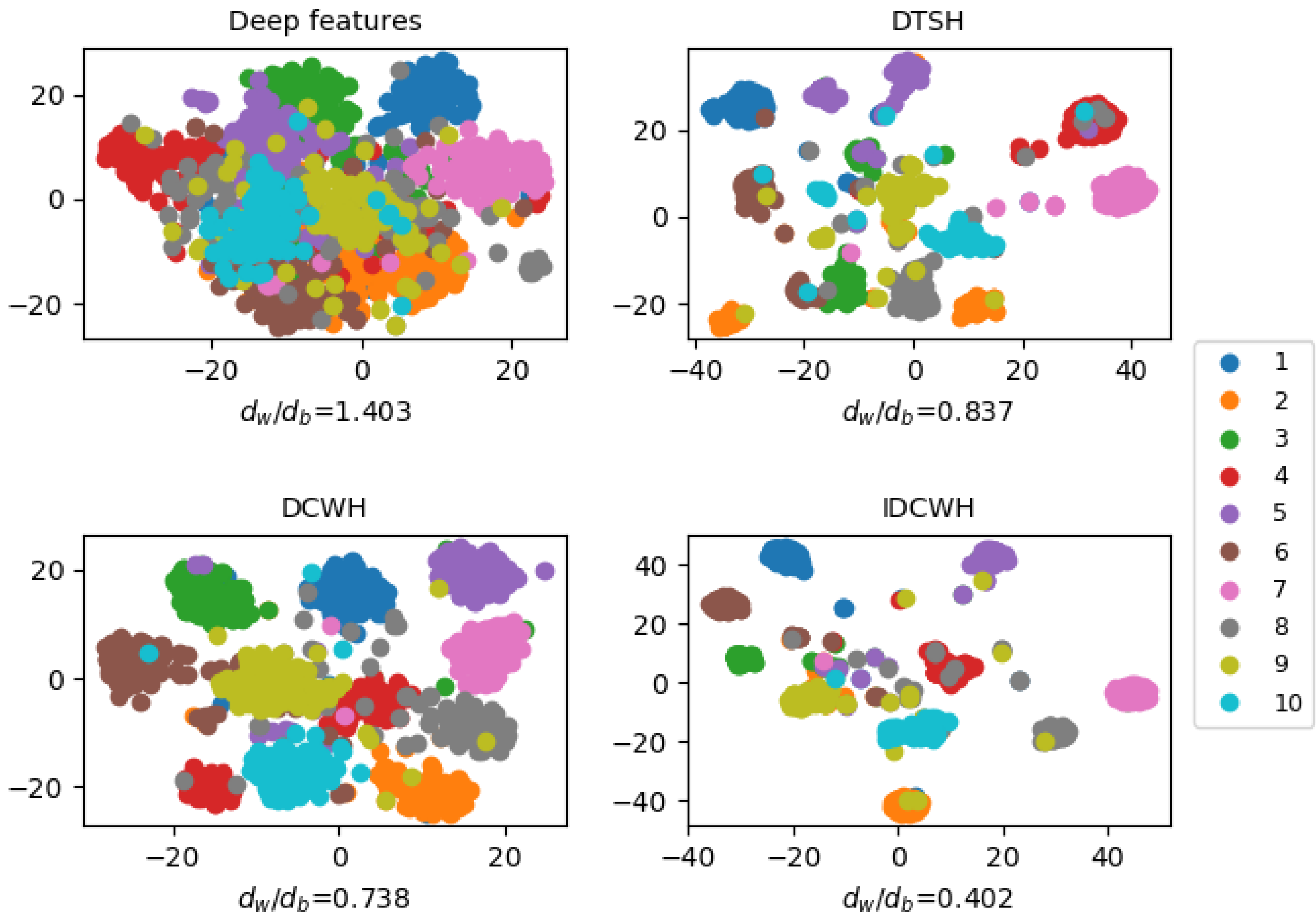

Fig. 6. The t-SNE visualization of hashing codes learned from deep features, DTSH, DCWH and the proposed IDCWH, respectively.

d_w/d_b represents the ratio of within-class distance and the between-class distance.

香港城市大學
City University of Hong Kong

Thanks

Q&A

mzhang367-c@my.cityu.edu.hk