Deep Self-taught Graph Embedding Hashing with Pseudo Labels for Image Retrieval

Yu Liu†, Yangtao Wang†, Jingkuan Song§, Chan Guo†, Ke Zhou†,and Zhili Xiao‡

†Huazhong University of Science and Technology, Wuhan, China

§University of Electronic Science and Technology of China, Chengdu, China

‡Tencent. Inc

Yu Liu

1/6/2020, Wuhan, China

Outline

- **Background**
- Motivation
- Design Overview
- Methodology
- Experiment Results
- Conclusion

Background

Hashing is widely used in large-scale retrieval owing to its efficiency for storage and matching

a=[0.328, -0.112,...,0.131]

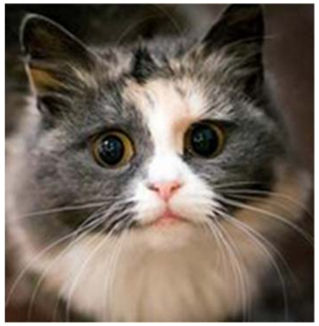
0110000110

|a-b|

$\mathrm{Exp}(a^2-b^2)^{0.5}$

ab/|a||b|

2



b=[0.187, -0.016,...,-0.547]

0110010010

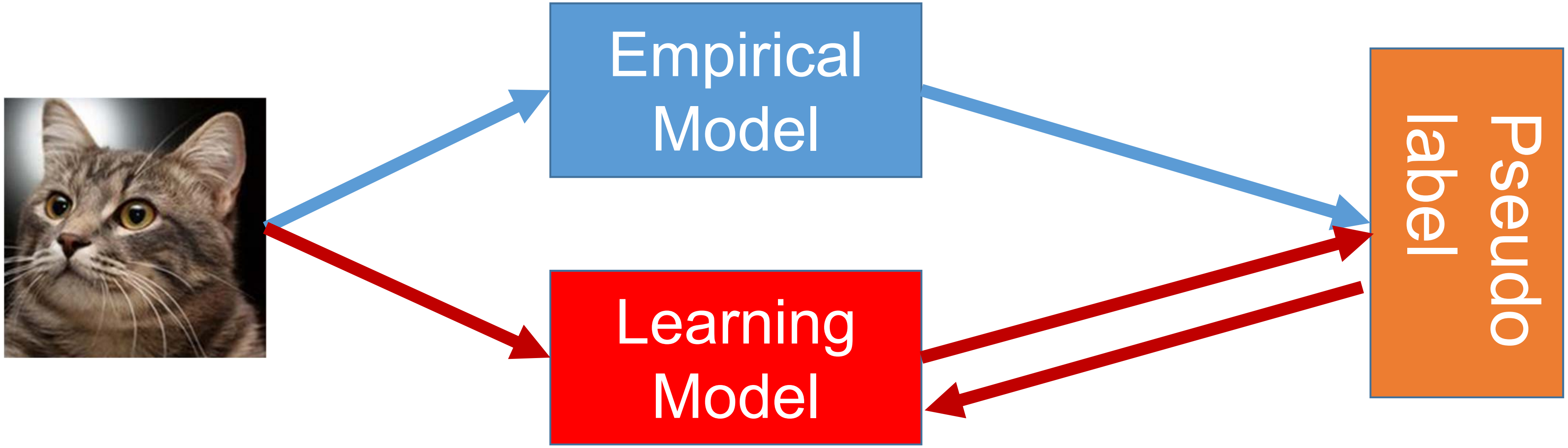
ICME 2020

Background

However, the dependence of label information in deep learning training restricts the practicability of deep hash algorithm

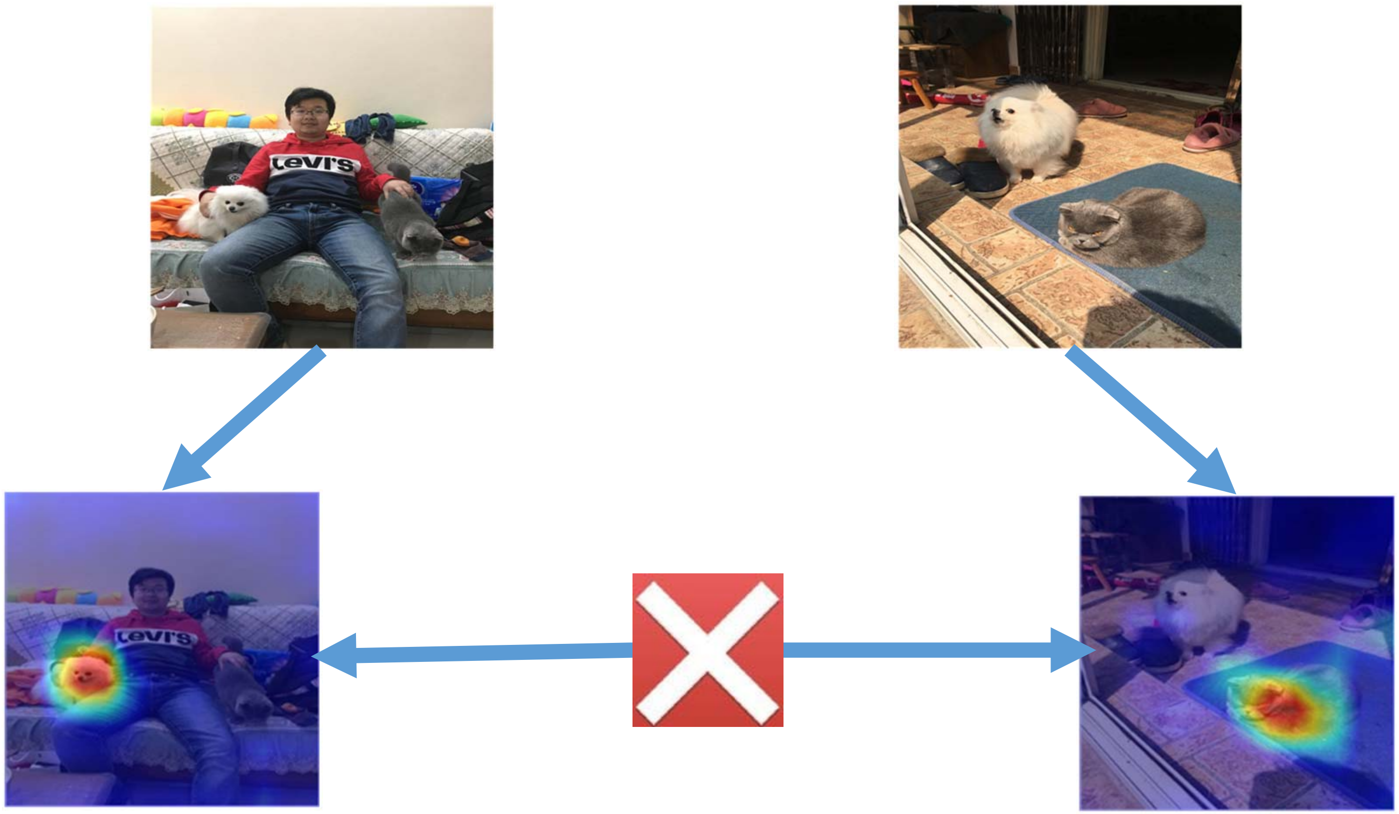
Background

Hash learning with pseudo labels



Background

Problems

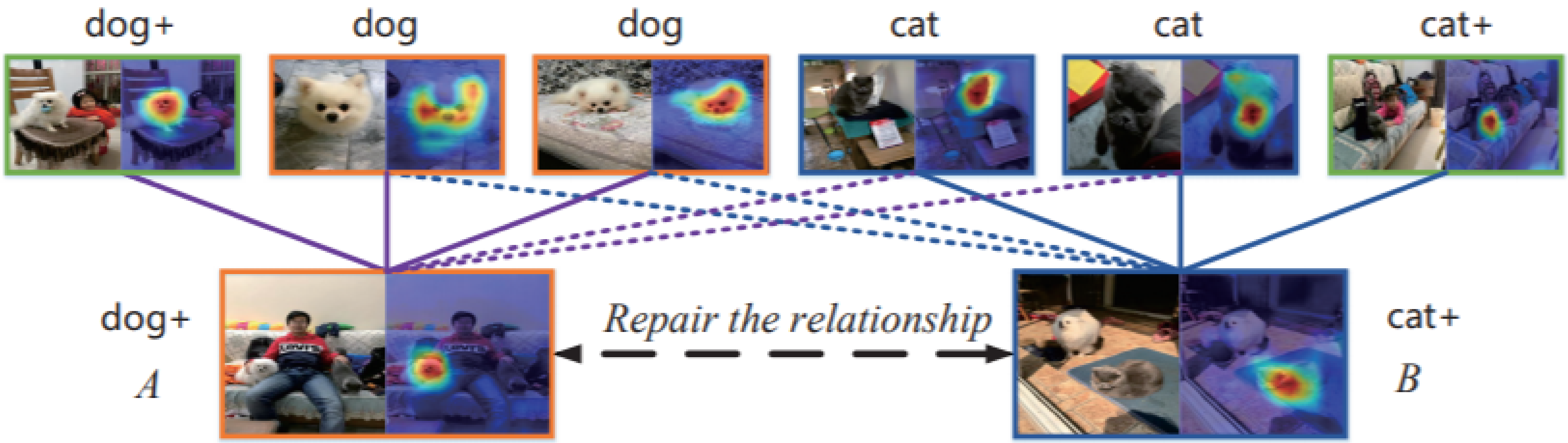


Outline

- Background
- Motivation
- Design Overview
- Methodology
- Experiment Results
- Conclusion

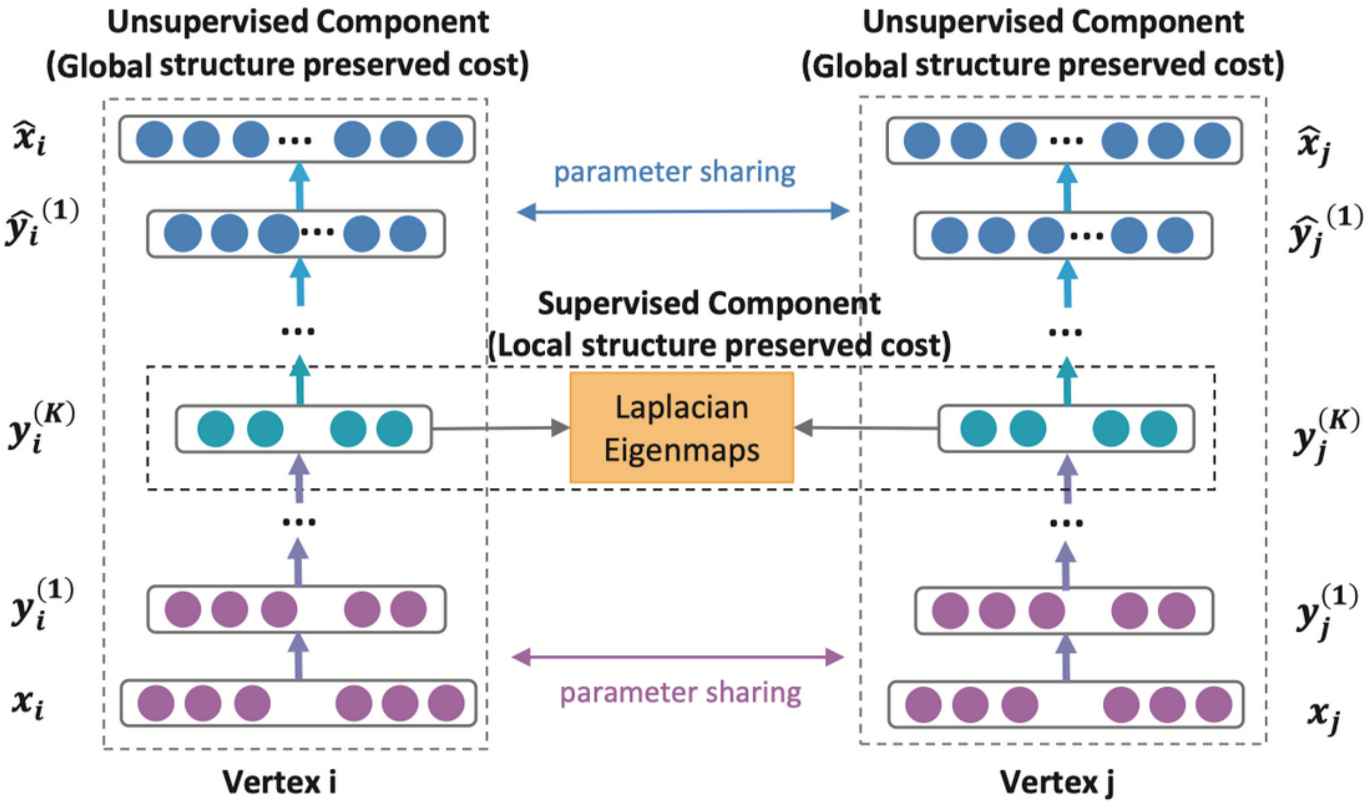
Motivation

Dependency with the indirect correlation



Motivation

Second-order proximity



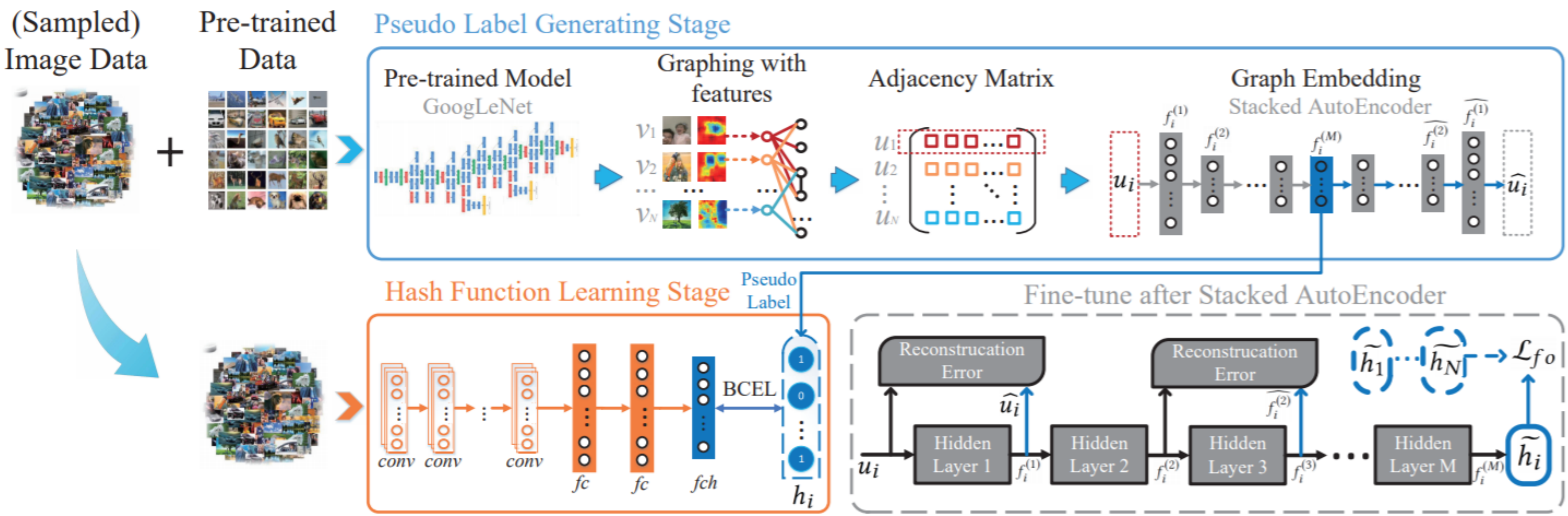
Wang D, Cui P, Zhu W. Structural deep network embedding[C]//Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining. 2016: 1225-1234.

Outline

- Background
- Motivation
- **Design Overview**
- Methodology
- Experiment Results
- Conclusion

Design Overview

Framework for DSTGeH

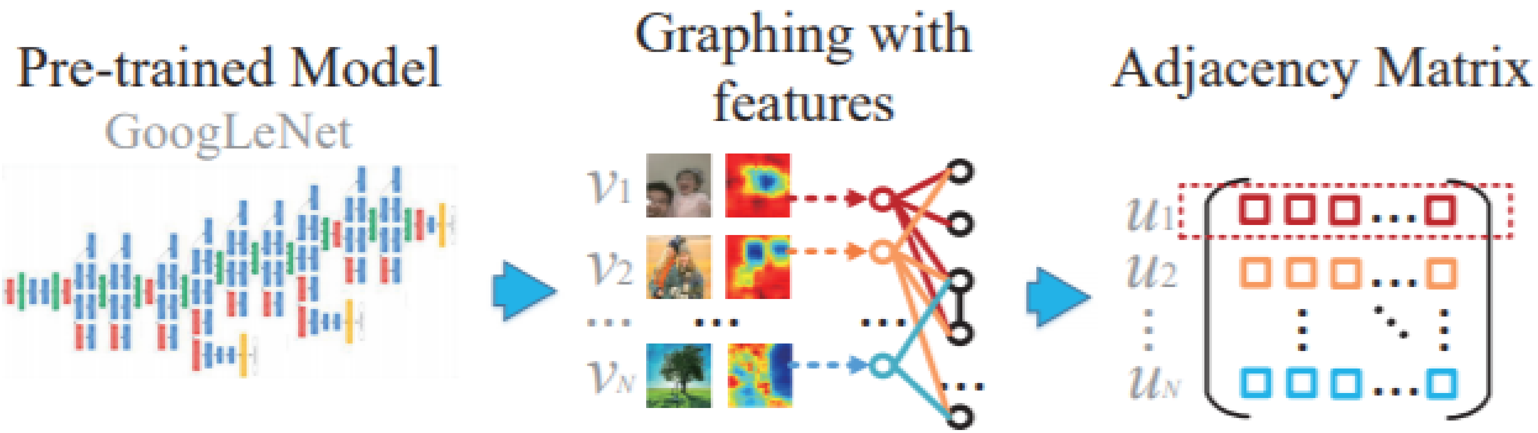


Outline

- Background
- Motivation
- Design Overview
- **Methodology**
- Experiment Results
- Conclusion

Methodology

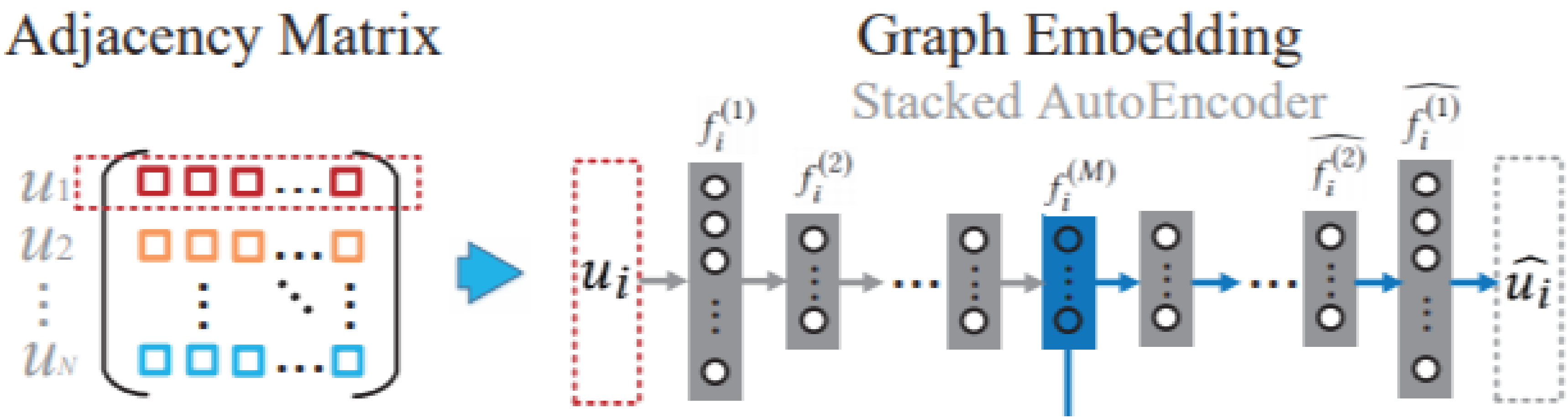
Pseudo label generating stage



cosine distance : $$u_{i,j} = \frac{\vec{v_i} \cdot \vec{v_j}}{\|v_i\| \cdot \|v_j\|}.$$

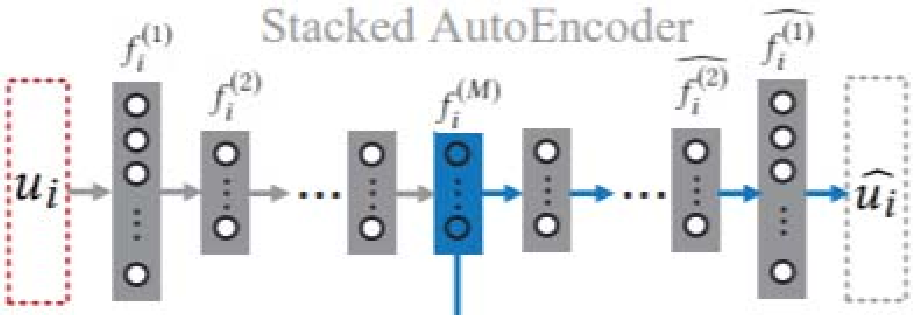
Methodology

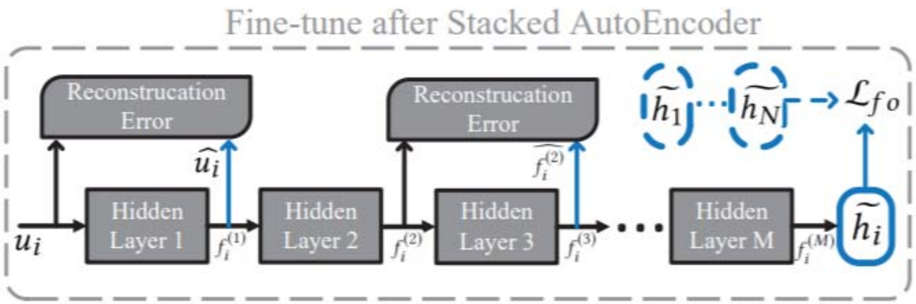
Pseudo label generating stage



Methodology

Pseudo label generating stage





Second-order Proximity of reconstruction error:

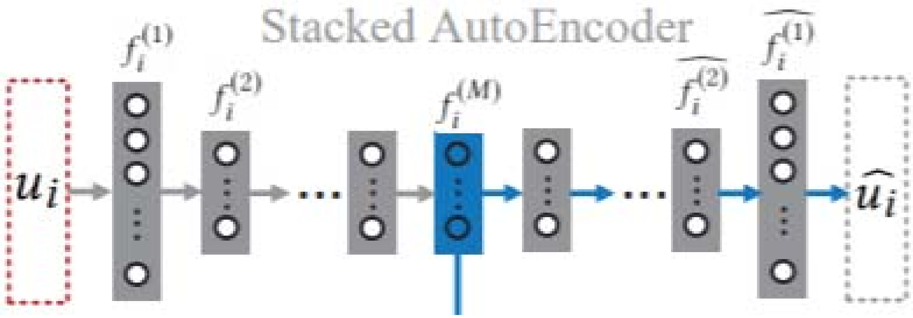
$$\mathcal{L}_{soh} = \sum_{i=1}^{N} \sum_{m=2\times t}^{M} \kappa_t \| f_i^{(m)} - \widehat{f_i^{(m)}} \|_2^2,$$

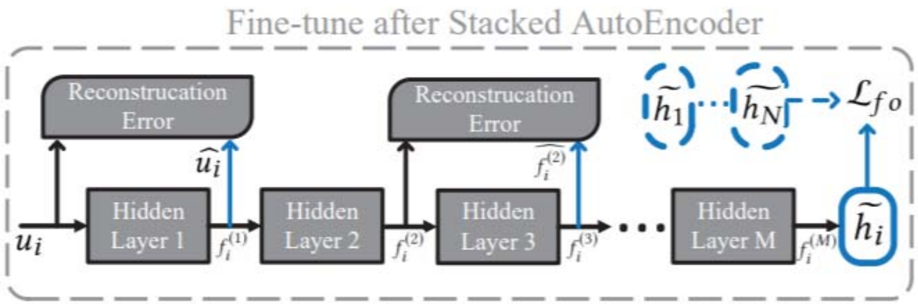
$$s.t. \quad \frac{\kappa_t}{\kappa_{t+1}} = 2, \quad \sum \kappa_t = 1,$$

$$\mathcal{L}_{so} = \mu \sum_{i=1}^{N} \| u_i - \widehat{u_i} \|_2^2 + (1-\mu)\mathcal{L}_{soh}$$

Methodology

Pseudo label generating stage





First-order Proximity of reconstruction error:

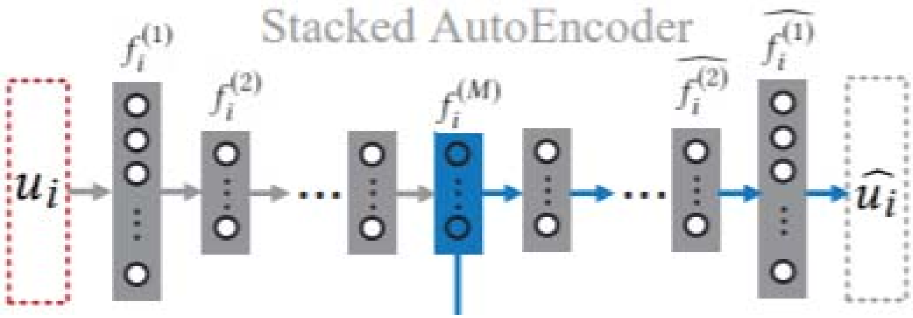
$$\mathcal{L}_{fo} = \sum_{i,j=1}^{N} u_{i,j} \| \widetilde{h_i} - \widetilde{h_j} \|_2^2$$

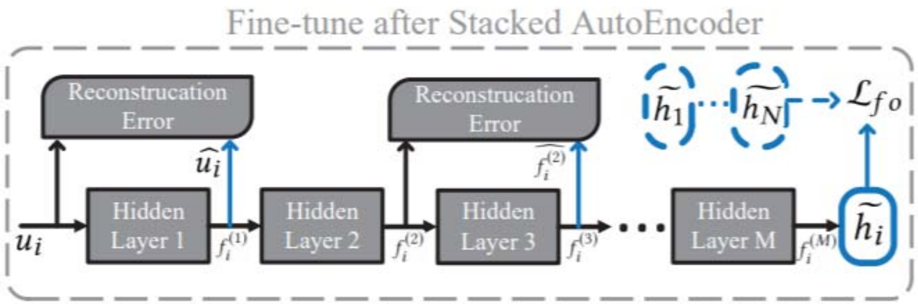
Regularize term:

$$\mathcal{L}_r = \frac{1}{2} \sum_{m=1}^{M} \| W^{(m)} \|_F^2 + \| \widehat{W^{(m)}} \|_F^2$$

Methodology

Pseudo label generating stage



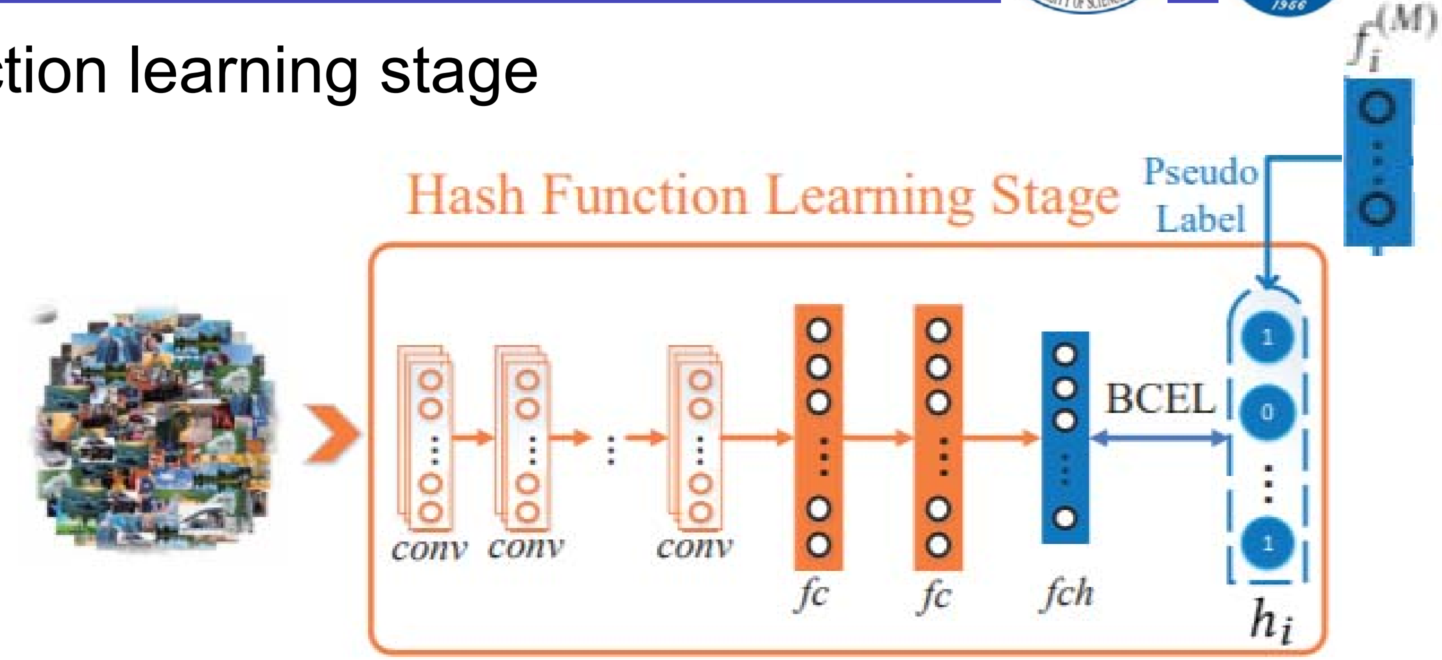


the terminal loss function:

$$\mathcal{L} = \alpha\mathcal{L}_{so} + \beta\mathcal{L}_{fo} + \gamma\mathcal{L}_{r}$$

Methodology

Hash function learning stage



BCEL: $$\mathcal{L}_h(h_i, \widetilde{H_i}) = -\sum_{p=1}^{g} h_i^p log(\widetilde{H_i^p}) + (1 - h_i^p) log(1 - \widetilde{H_i^p})$$

Outline

- Background
- Motivation
- Design Overview
- Methodology
- **Experiment Results**
- Conclusion

Experimental Results

- Datasets: CIFAR-10 dataset

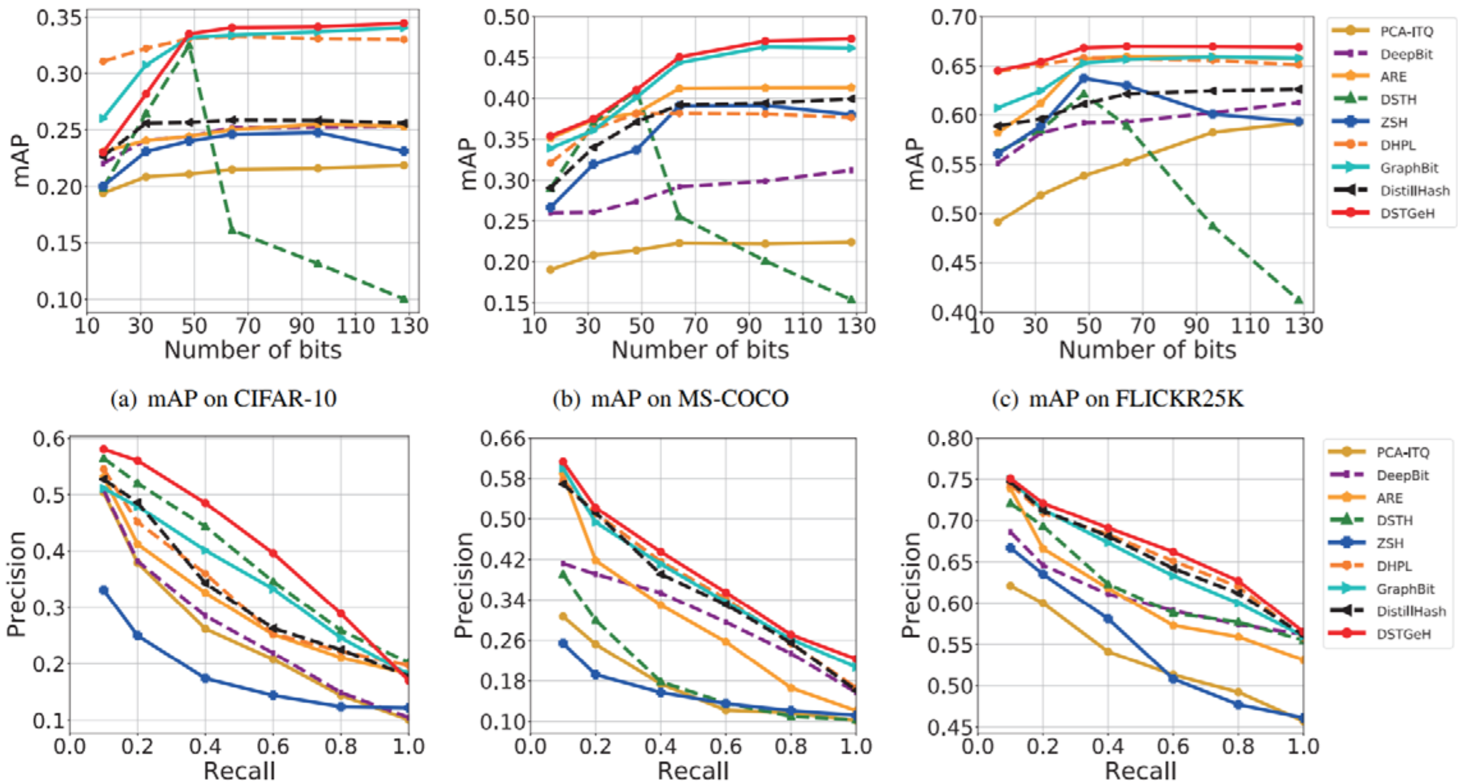
  MS-COCO dataset

  FLICKR25K dataset

- Experimental Procedures:
  - Comparison with the State-of-the-art Methods
  - Ablation Studies

Experimental Results

Comparison with the State-of-the-art Methods



(a) mAP on CIFAR-10 (b) mAP on MS-COCO (c) mAP on FLICKR25K

(d) PR on CIFAR-10 (e) PR on MS-COCO (f) PR on FLICKR25K

Fig. 5: mAP with different code lengths and 48-bits PR curve on CIFAR-10, MS-COCO and FLICKR25K.

Experimental Results

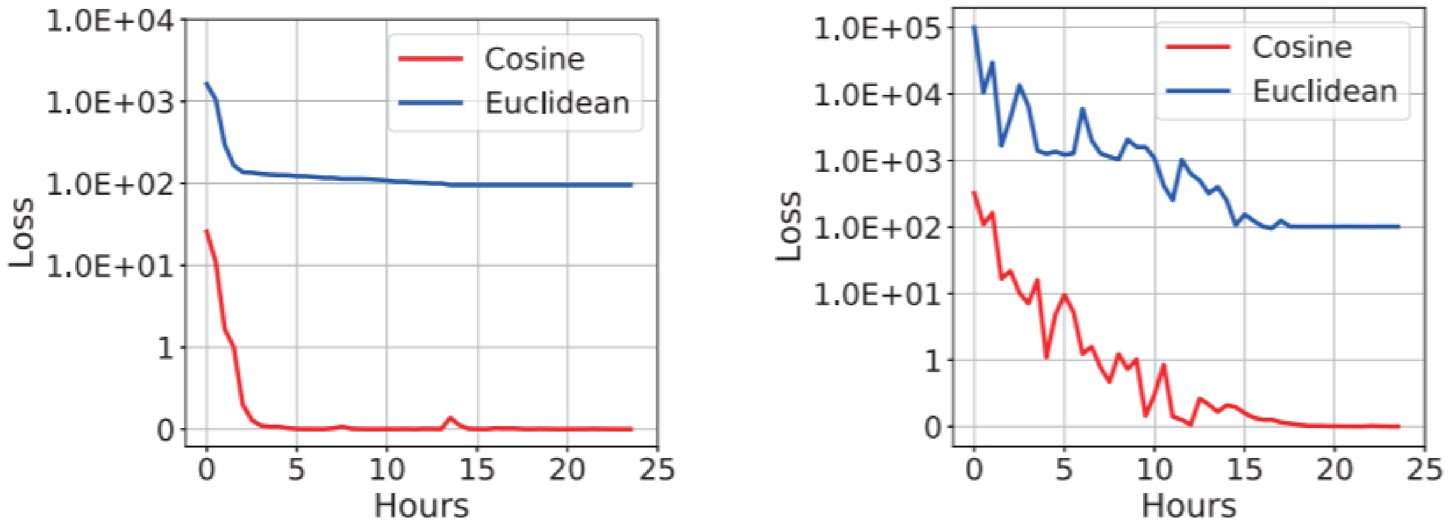
Ablation Studies: Influence of hash length and distance metric

Table 1: Performance (mAP) variances varying code length L and distance metric.

Dataset	Distance Metric	mAP@Hash Code Length L					
		16-bits	32-bits	48-bits	64-bits	96-bits	128-bits
CIFAR-10	Cosine	**0.2303**	**0.2818**	**0.3352**	**0.3408**	**0.3415**	**0.3447**
	Euclidean	0.2223	0.2702	0.3165	0.3252	0.3247	0.3269
MS-COCO	Cosine	0.3540	**0.3750**	**0.4102**	**0.4508**	**0.4703**	**0.4731**
	Euclidean	**0.3543**	0.3552	0.4008	0.4388	0.4607	0.4621

Experimental Results

Ablation Studies: Influence of distance metric



(a) CIFAR-10 (b) MS-COCO

Fig. 3: GE convergence on CIFAR-10 and MS-COCO with different distance metrics in the first stage.

Experimental Results

Ablation Studies: Influence of GE component

Table 2: Performance (PR) variance on 48-bits varying GE method, recall and loss function.

Dataset	GE Method (Loss Function)	Precision@Recall 10%	20%	40%	60%	80%	F_1
CIFAR-10	LE (BCEL)	0.3944	0.3302	0.2835	0.2397	0.2086	0.2828
	Word2Vec (BCEL)	0.5623	0.5037	0.4193	0.3025	0.2343	0.3260
	SDNE (BCEL)	**0.5919**	0.5578	0.4573	0.3604	0.2673	0.3487
	SDNE+SAE (BCEL)	0.5804	**0.5601**	**0.4848**	**0.3962**	**0.2885**	**0.3610**
	SDNE+SAE (MLML)	0.4311	0.3534	0.2942	0.2599	0.2217	0.2933
	SDNE+SAE (MLSML)	0.2319	0.2206	0.2032	0.1623	0.1303	0.1197
MS-COCO	LE (BCEL)	0.3120	0.2402	0.2001	0.1657	0.1304	0.2241
	Word2Vec (BCEL)	0.5509	0.5050	0.4001	0.3075	0.2265	0.3231
	SDNE (BCEL)	0.5932	0.5215	0.4188	0.3070	0.2173	0.3235
	SDNE+SAE (BCEL)	**0.6136**	**0.5223**	**0.4347**	**0.3552**	**0.2709**	**0.3458**
	SDNE+SAE (MLML)	0.3561	0.2437	0.2074	0.1763	0.1410	0.2322
	SDNE+SAE (MLSML)	0.2019	0.1882	0.1724	0.1592	0.1333	0.2098

Experimental Results

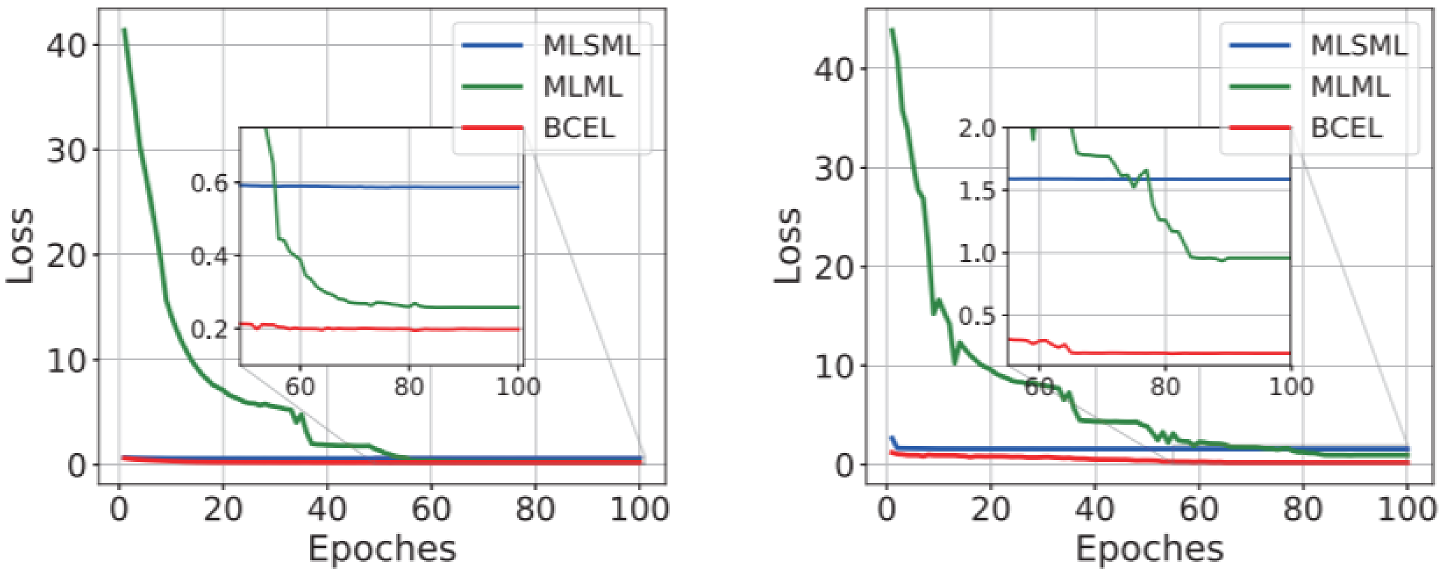
Ablation Studies: Influence of loss function

Table 2: Performance (PR) variance on 48-bits varying GE method, recall and loss function.

Dataset	GE Method (Loss Function)	Precision@Recall					F_1
		10%	20%	40%	60%	80%	
CIFAR-10	LE (BCEL)	0.3944	0.3302	0.2835	0.2397	0.2086	0.2828
	Word2Vec (BCEL)	0.5623	0.5037	0.4193	0.3025	0.2343	0.3260
	SDNE (BCEL)	**0.5919**	0.5578	0.4573	0.3604	0.2673	0.3487
	SDNE+SAE (BCEL)	0.5804	**0.5601**	**0.4848**	**0.3962**	**0.2885**	**0.3610**
	SDNE+SAE (MLML)	0.4311	0.3534	0.2942	0.2599	0.2217	0.2933
	SDNE+SAE (MLSML)	0.2319	0.2206	0.2032	0.1623	0.1303	0.1197
MS-COCO	LE (BCEL)	0.3120	0.2402	0.2001	0.1657	0.1304	0.2241
	Word2Vec (BCEL)	0.5509	0.5050	0.4001	0.3075	0.2265	0.3231
	SDNE (BCEL)	0.5932	0.5215	0.4188	0.3070	0.2173	0.3235
	SDNE+SAE (BCEL)	**0.6136**	**0.5223**	**0.4347**	**0.3552**	**0.2709**	**0.3458**
	SDNE+SAE (MLML)	0.3561	0.2437	0.2074	0.1763	0.1410	0.2322
	SDNE+SAE (MLSML)	0.2019	0.1882	0.1724	0.1592	0.1333	0.2098

Experimental Results

Ablation Studies: Influence of loss function



(a) CIFAR-10 (b) MS-COCO

Fig. 4: Hash function convergence on CIFAR-10 and MS-COCO with different loss functions in the second stage.

Outline

- Background
- Motivation
- Design Overview
- Methodology
- Experiment Results
- **Conclusion**

Conclusion

- We propose a deep self-taught graph embedding hashing (DSTGeH) framework for image retrieval.
- We introduce the second-order proximity into the hash learning, which improves the precision of hash code in the self-taught framework with pseudo labels.
- DSTGeH achieves the best performances on most evaluation metrics and produce an overwhelming advantage on multi-label datasets.

ICME 2020

Thank You!

Yu Liu

Yangtao Wang

Chan Guo

Ke Zhou

Jingkuan Song

Zhili Xiao