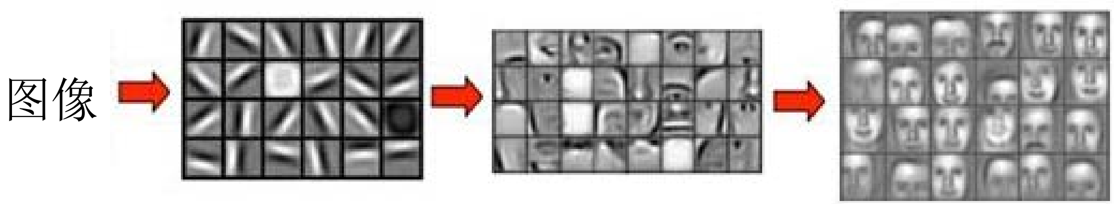
基于弱标注信息的深度学习

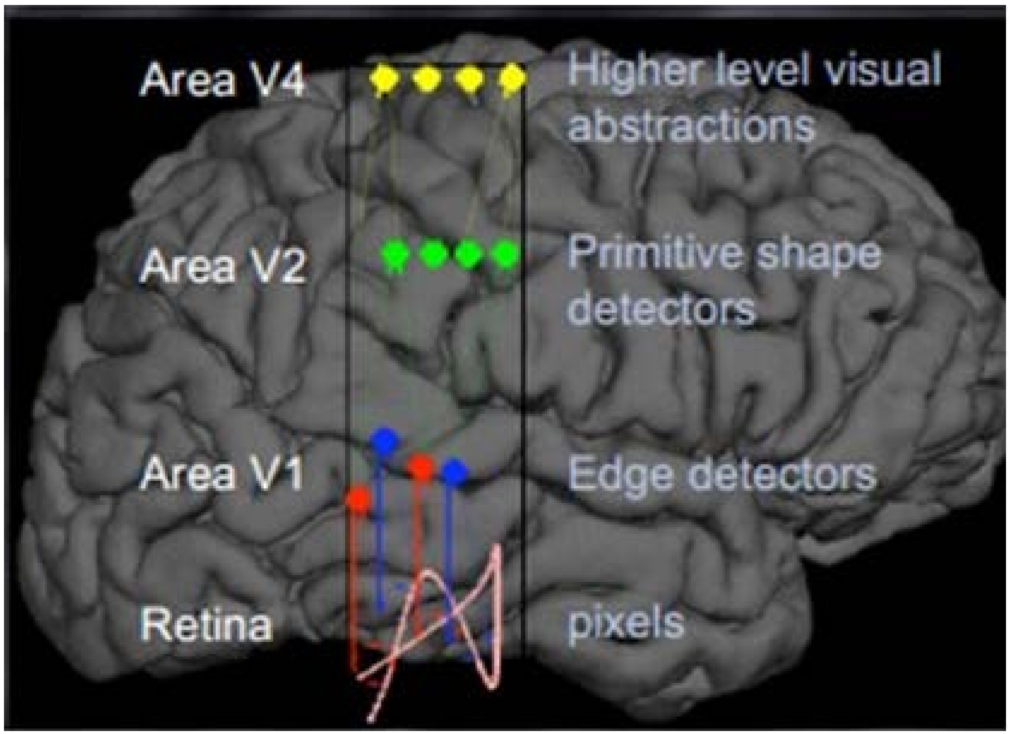
管子玉
西安电子科技大学
2018/5/18

深度学习

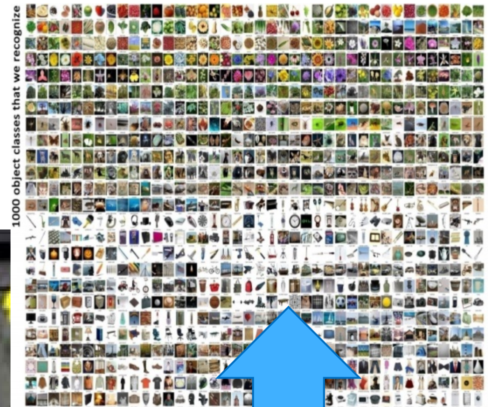
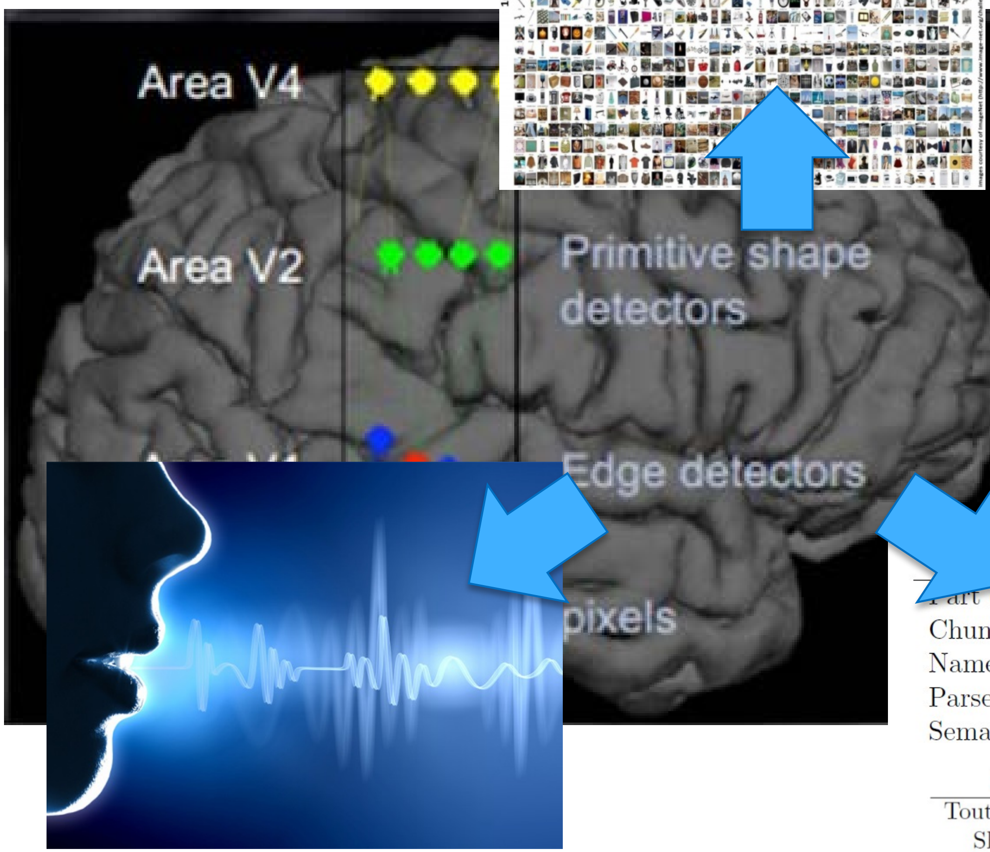
- 从2006年开始重新兴起的一种学习方法，旨在建立类似人脑的神经网络



人脑的神经系统具有深度（多层）结构



深度学习取得的成绩

大规模图像分类

72%, 2010

IN

sNet

- 无监督预学习技术（2006）

✓ RBM

		Benchmark	SENNA
Part of Speech (POS)	(Accuracy)	97.24 %	97.29 %
Chunking (CHUNK)	(F1)	94.29 %	94.32 %
Named Entity Recognition (NER)	(F1)	89.31 %	89.59 %
Parse Tree level 0 (PT0)	(F1)	91.94 %	92.25 %
Semantic Role Labelling (SRL)	(F1)	77.92 %	75.55 %

POS System	Time (s.)
Toutanova et al. (2003)	1065
Shen et al. (2007)	833
SENNA	$\boxed{4}$

SRL System	Time (s.)
Koomen et al. (2005)	6253
SENNA	$\boxed{51}$

语音识别：性能**30%**相对提升

自然语言处理：与基准方法可比较的性能；无需特征设计；

深度学习的缺点

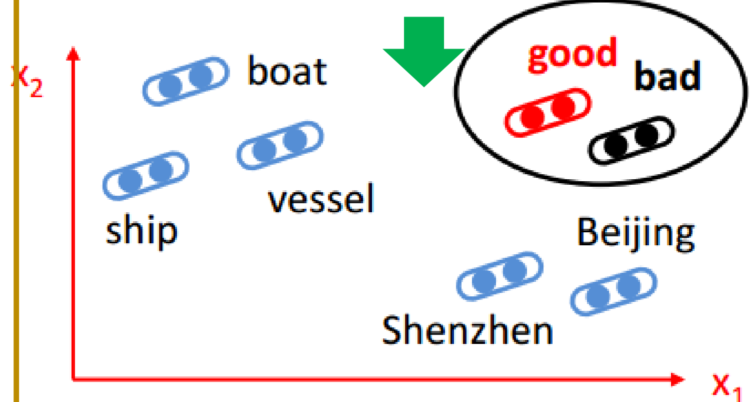
1. **需要大量训练数据来训练深度模型**，而标注训练数据需要耗费高昂的人工成本。

2. **无监督训练方法需要满足数据与语义间具有强相关性的条件**，但数据分布与语义分布很多时候存在不一致的情况，在这种情况下使用无监督训练方法的效果并不理想。

文本数据分布与语义分布不一致情况举例：
在舆情分析领域，计算机无法从无标注的文本数据中区分正负情感。

he formed the ***good*** *habit of …*

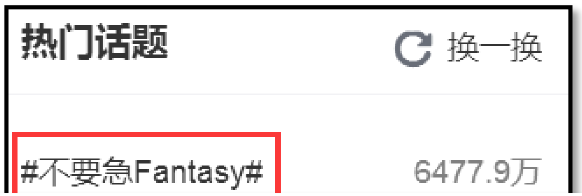
he formed the ***bad*** *habit of …*



弱标注信息

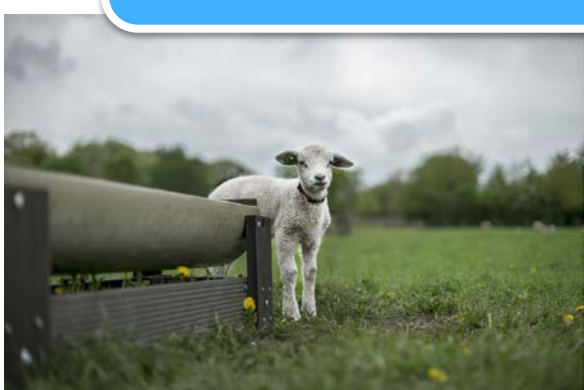
- ➢ 互联网上存在海量用户生成的弱标注信息

微博话题标签



购物网站用户评论的评分

用户生成的弱标注信息能一定程度上反映数据语义



弱标注信息

□ 良好的标注不能直接用于学习任务

Weakly-supervised Deep Learning for Customer Review Sentiment Classification

IJCAI 2016, IEEE TKDE 2017

- **Background & Problem Definition**
- Related Work
- Observation & Challenge & Solution
- Our approach
- Experiments
- Conclusion

Background

- **Sentiment classification on the product reviews**

Customers : Make decision

Merchants : Improve products or service

Problem Definition

- **A binary classification problem on <u>Sentence-Level</u>**

A review sentence: Bluetooth works half of the time.

Sentiment orientation prediction: **Positive** or **Negative**

- Background & Problem Definition
- **Related Work**
- Observation & Challenge & Solution
- Our approach
- Experiments
- Conclusion

Related works

- **Lexicon-based methods**
  1. Lexicon construction
  2. Rules design

Does not work when words out of lexicon

- **Machine Learning methods**

  (1) Traditional methods
  1. Feature engineering
  2. Feature set is lack of suitability

  (2) Deep Learning methods

  Need large scale training data [Bengio et al., 2013; Bengio, 2009]

Cost a lot of human efforts

An Intuition : we want to use **user-generated information** to train the deep model.

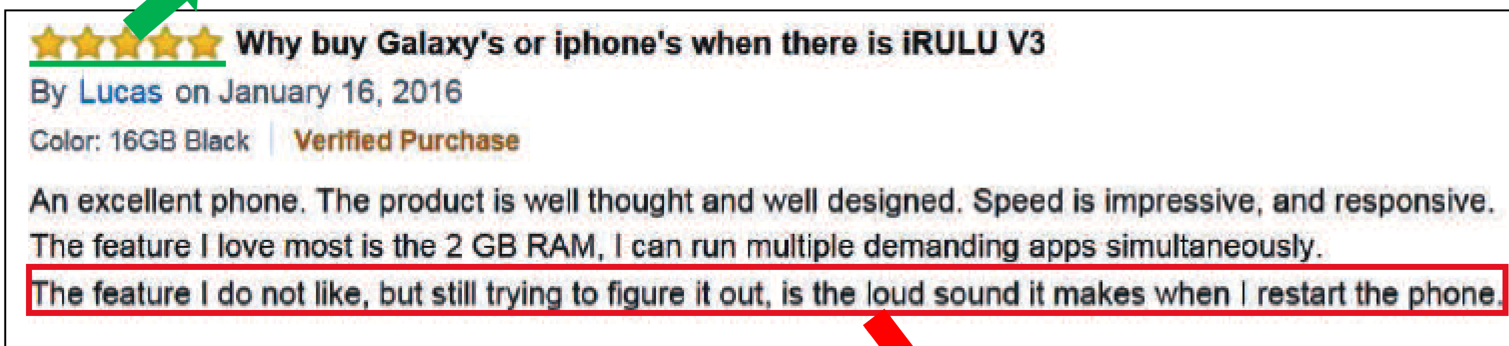
- Background & Problem Definition
- Related Work
- **Observation & Challenge & Solution**
- Our approach
- Experiments
- Conclusion

Observation

- Ratings reflect overall sentiment of customer reviews but has some wrong labeled sentences.

5-stars Review

A wrong labeled sentence by ratings

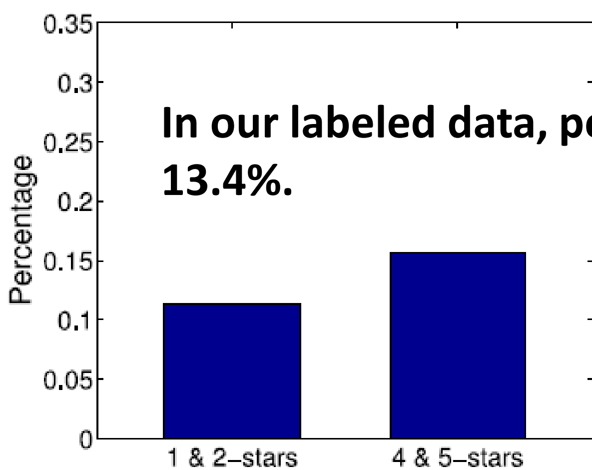

Why buy Galaxy's or iphone's when there is iRULU V3

By Lucas on January 16, 2016

Color: 16GB Black | Verified Purchase

An excellent phone. The product is well thought and well designed. Speed is impressive, and responsive. The feature I love most is the 2 GB RAM, I can run multiple demanding apps simultaneously. The feature I do not like, but still trying to figure it out, is the loud sound it makes when I restart the phone.

Negative Sentence



In our labeled data, percent of wrong-labeled sentences (noise) is 13.4%.

- **Weakly labeled data cannot be directly used to train our deep model.**

Challenge & Solution

- **Challenge** : We want to use sentences with ratings to train a deep model, but the ratings are weak labels which have noise.
- **Solution** : A 2-step training method.

Step1 : Embedding pre-training.

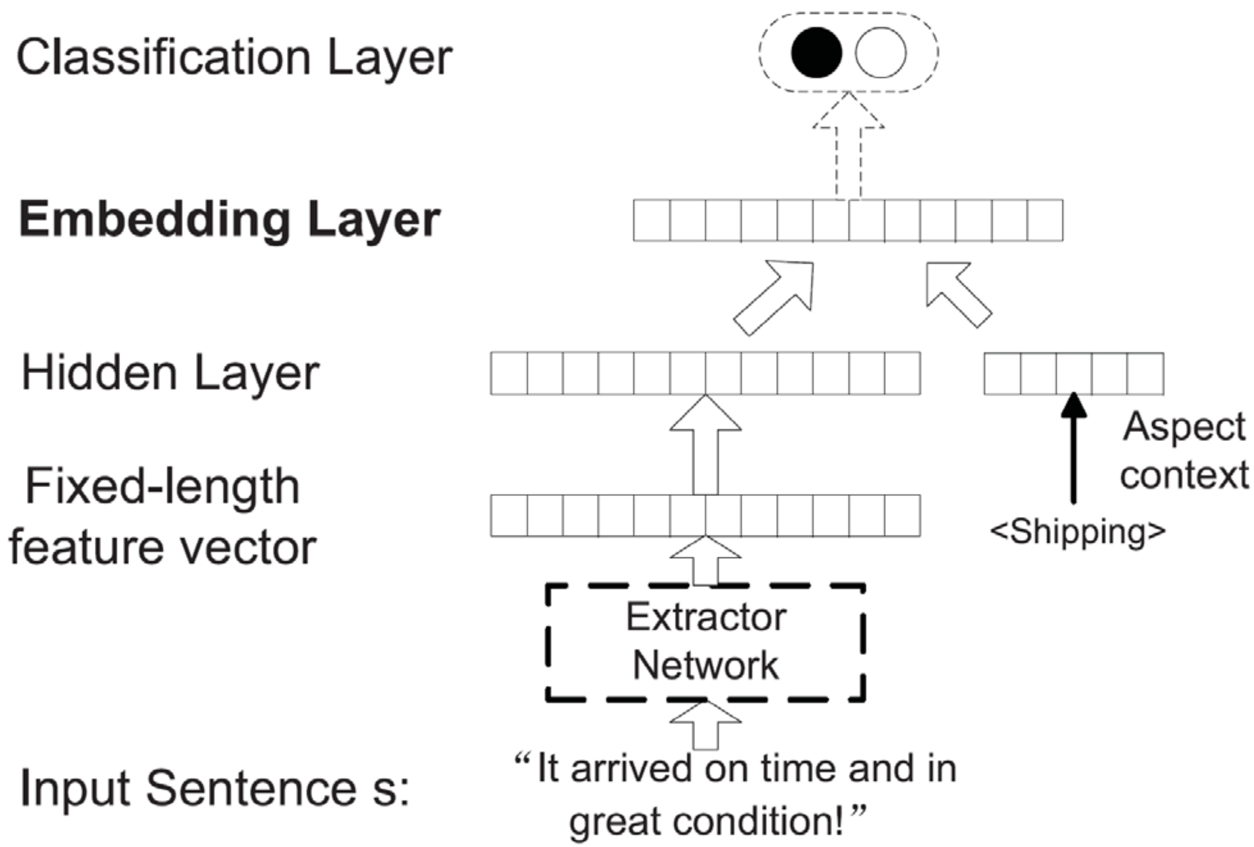
➡ Alleviate influence of the noise, learn a good embedding space which captures senti. distribution.

Step2 : Supervised fine-tuning.

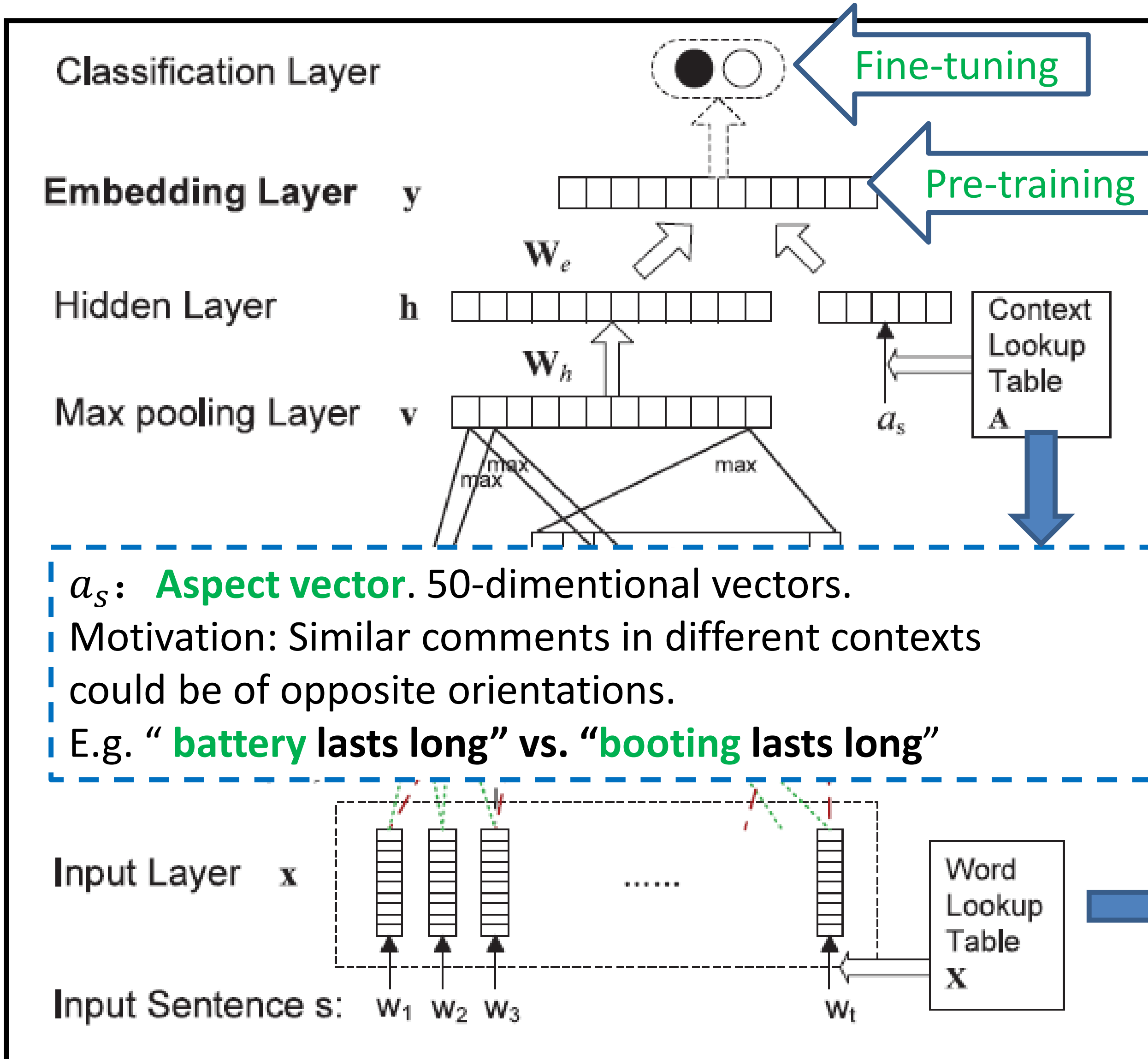
➡ Fine-tune the per-trained deep model.

- Background & Problem Definition
- Related Work
- Observation & Challenge & Solution
- **Our approach**
- Experiments
- Conclusion

Our approach: WDE



Our approach: WDE-CNN



a_s: **Aspect vector**. 50-dimentional vectors.
Motivation: Similar comments in different contexts could be of opposite orientations.
E.g. " **battery lasts long**" vs. "**booting lasts long**"

300-dimentional word vectors (generated by Word2Vec. [Mikolov et al., 2013])

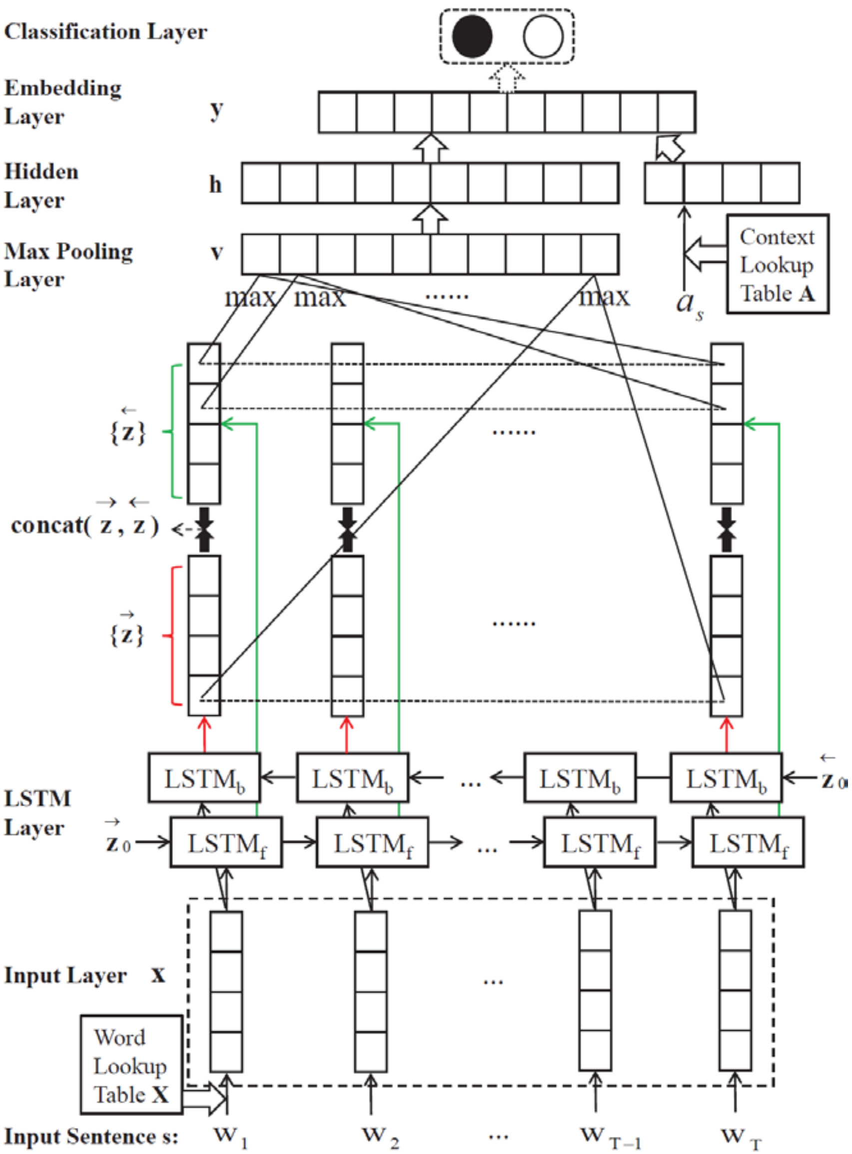
$$y = f(w_e \begin{matrix} h \\ a_s \end{matrix} + b_e)$$

$$h = f(w_h v + b_h)$$

$$v = max(u)$$

$$u = f(w_c x + b)$$

Our approach: WDE-LSTM

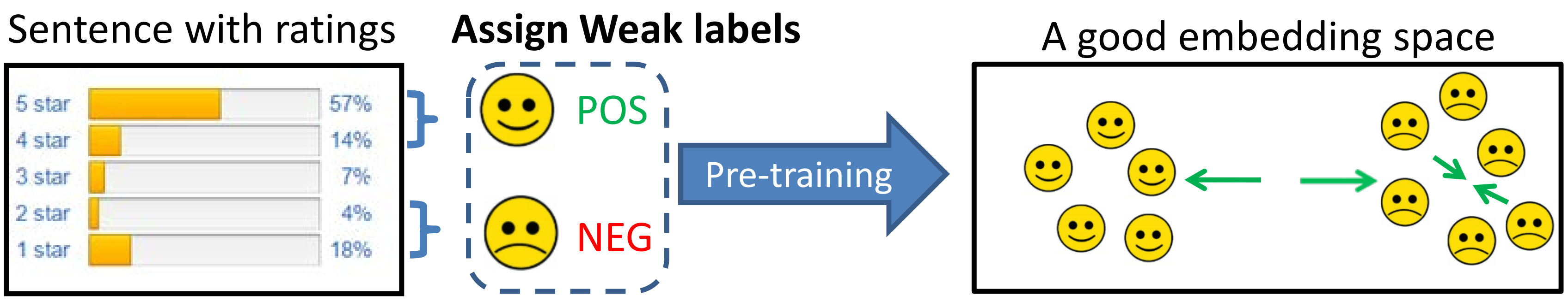


$$
\begin{aligned}
\overrightarrow{\mathbf{z}}_t &= \mathrm{LSTM_f}(\mathbf{x}_t, \overrightarrow{\mathbf{z}}_{t-1}) \\
\overleftarrow{\mathbf{z}}_t &= \mathrm{LSTM_b}(\mathbf{x}_t, \overleftarrow{\mathbf{z}}_{t+1}) \\
\mathbf{z}_t &= \mathrm{concat}(\overrightarrow{\mathbf{z}}_t, \overleftarrow{\mathbf{z}}_t),
\end{aligned}
$$

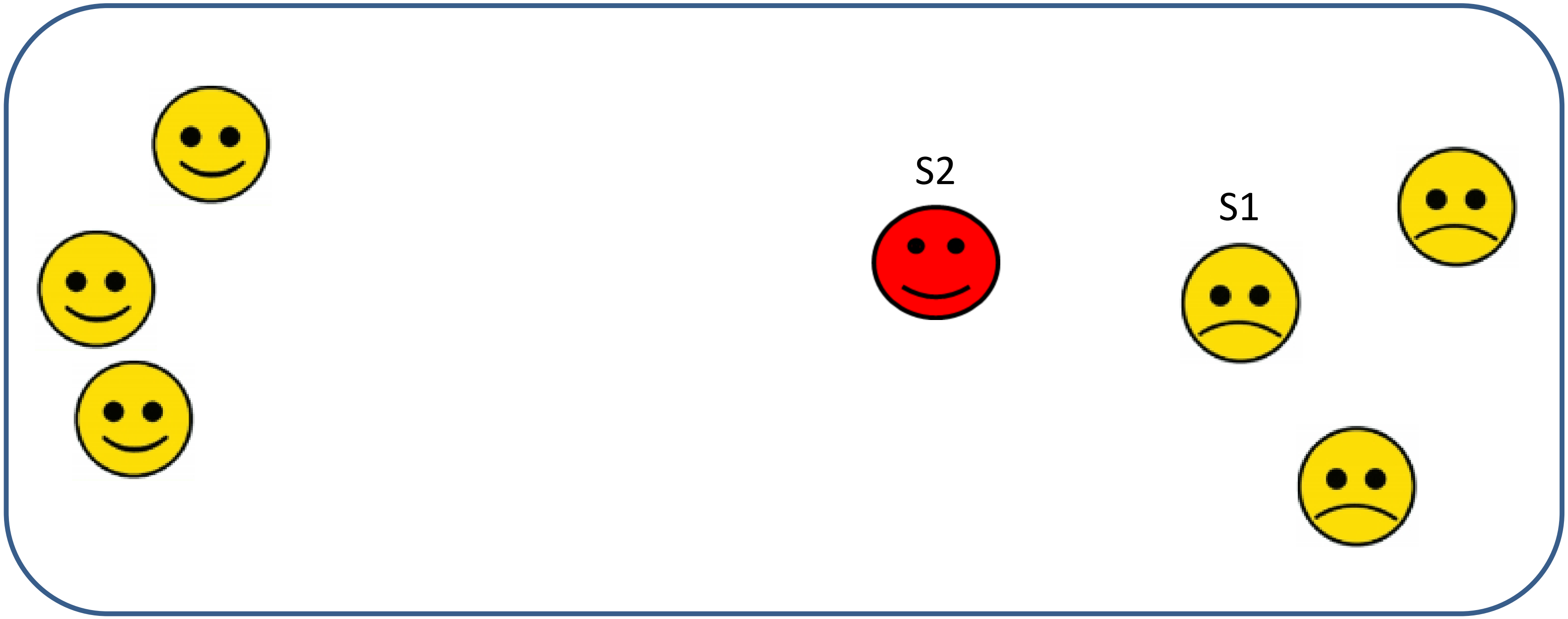
$$
\begin{aligned}
\mathbf{d}_t &= g(\mathbf{W_d}\mathbf{x}_t + \mathbf{U_d}\mathbf{z}_{t-1} + \mathbf{b_d}) \\
\mathbf{i}_t &= \sigma(\mathbf{W_i}\mathbf{x}_t + \mathbf{U_i}\mathbf{z}_{t-1} + \mathbf{b_i}) \\
\mathbf{f}_t &= \sigma(\mathbf{W_f}\mathbf{x}_t + \mathbf{U_f}\mathbf{z}_{t-1} + \mathbf{b_f}) \\
\mathbf{o}_t &= \sigma(\mathbf{W_o}\mathbf{x}_t + \mathbf{U_o}\mathbf{z}_{t-1} + \mathbf{b_o}) \\
\mathbf{c}_t &= \mathbf{i}_t \odot \mathbf{d}_t + \mathbf{f}_t \odot \mathbf{c}_{t-1} \\
\mathbf{z}_t &= \mathbf{o}_t \odot g(\mathbf{c}_t)
\end{aligned}
$$

Our approach : Pre-training

- **Train a good embedding space to capture general sentiment distributions of sentences.**

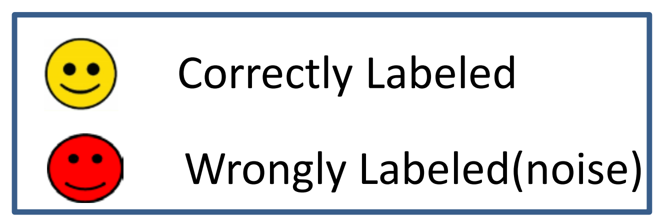


Our approach : Pair-based training



Pair-Based Training

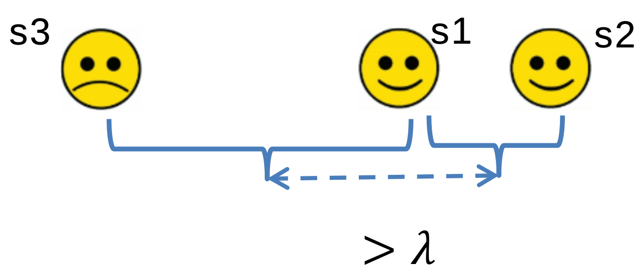
Sampled sentences pair(S1, S2)



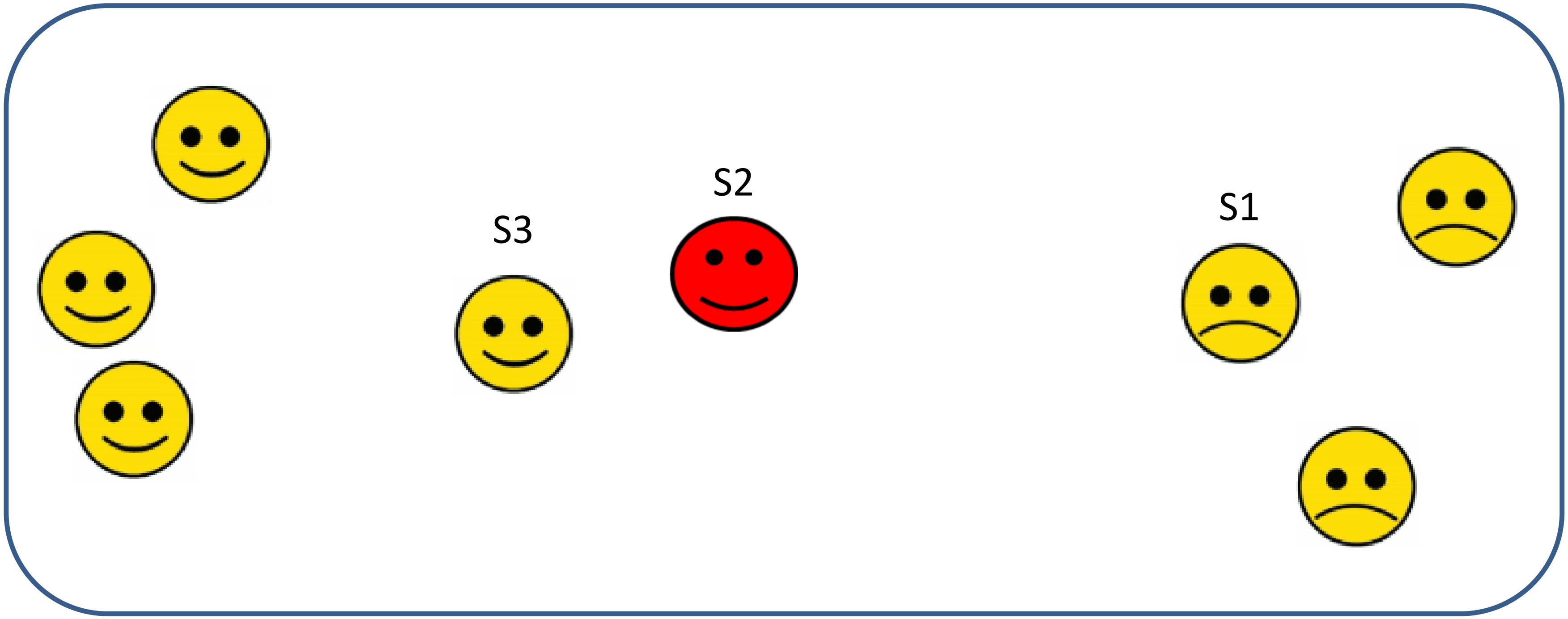
Our approach : Triplet-based training

- **Penalize relative distances for sentence triplets.**

$$l_{weak} = \sum_{\underbrace{<s_1, s_2, s_3>}_{\text{Triplet}}} \max(0, \underbrace{\lambda}_{\text{Margin}} - \underbrace{dis(s_1, s_3)}_{\text{Different labels}} + \underbrace{dis(s_1, s_2)}_{\text{Same label}}) \qquad \underbrace{dis(s_i, s_j) = \|y_i - y_j\|_2}_{\text{Distance}}$$

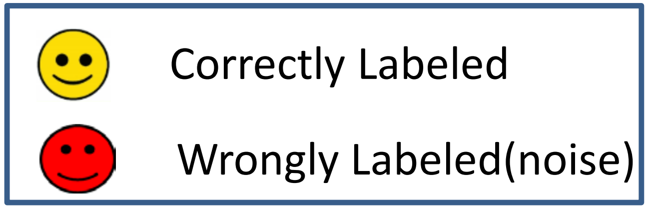


Our approach : Triplet-based training



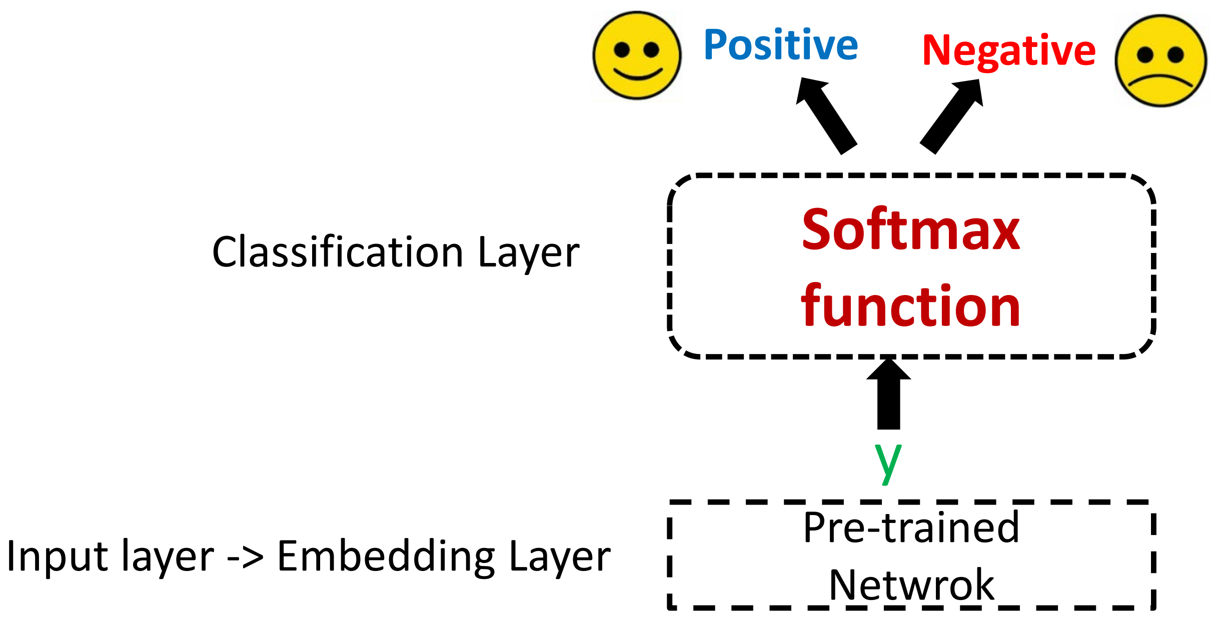
Triplet-Based Training

Sampled sentences triplet (S1, S2, S3)



Our approach : Fine-tuning

- Classification Layer: We use the pre-trained parameters to initialize the model and apply a softmax activation function to the prediction.
- Training: Use standard SGD to train the whole network.



- Background & Problem Definition
- Related Work
- Observation & Challenge & Solution
- Our approach
- **Experiments**
- Conclusion

Experiments : Data set

- **Statistics of the unlabeled dataset for pre-training**
  - Amazon Customer Review [McAuley et al., 2015]
  - 1.1M datasets include 1,143,721 sentences
  - Digital cameras, cell phones and laptops

- **Statistics of the labeled dataset for fine-tuning**

	Positive	Negative	Total
Subjective	3750	2024	5774
Objective	1860	4120	5980
Total	5610	6144	11754

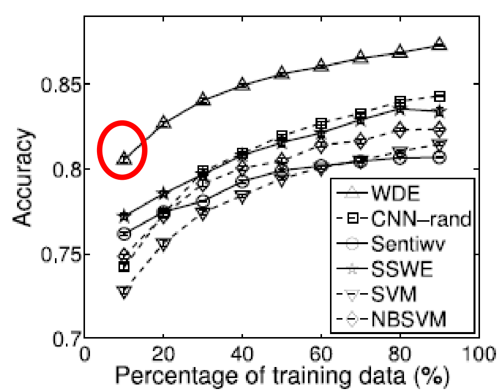
  - Training set: 50% , Validation set : 20%, Test set : 30%

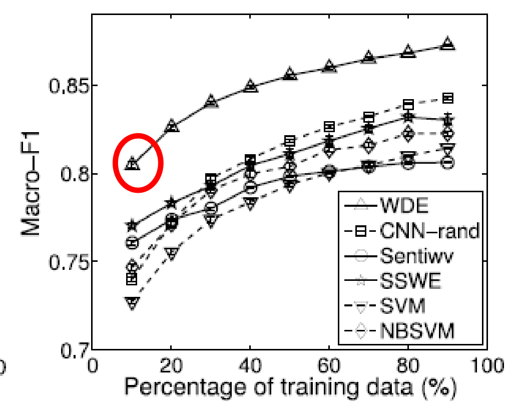
Experiments: Performance comparison

- **Performance comparison on Objective/Subjective/All**

Method	Accuracy			Macro-F1		
	All	Subj	Obj	All	Subj	Obj
Lexicon	.722	.827	.621	.721	.812	.613
SVM	.818	.838	.800	.818	.821	.765
NBSVM	.826	.844	.808	.825	.831	.773
SSWE	.835	.857	.815	.834	.826	.804
SentiWV	.808	.806	.809	.807	.786	.771
CNN-rand	.847	.861	.835	.847	.848	.802
CNN-weak	.771	.773	.770	.771	.755	.741
WDE	**.877**	**.886**	**.868**	**.876**	**.875**	**.843**

- **Performance comparison on different training data size (10%, 20%, 30%,… , 90%)**





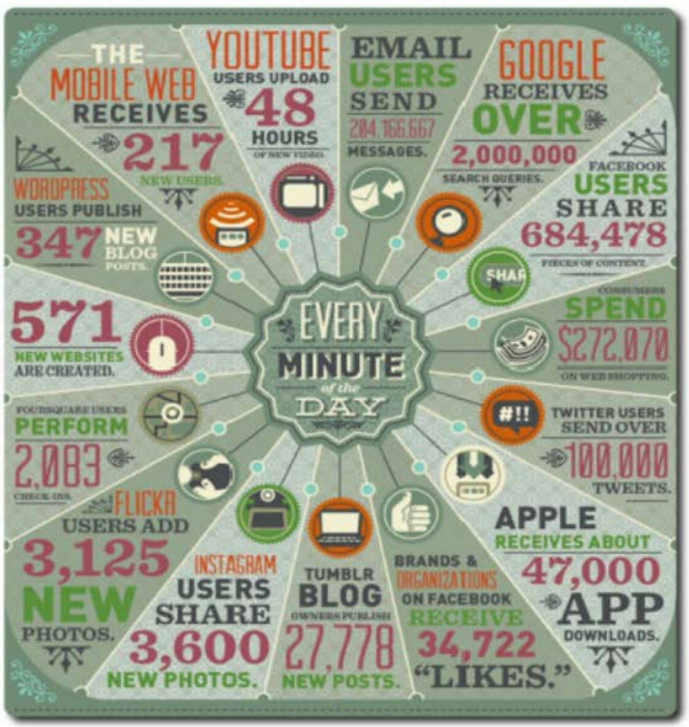
Deep Multiple Instance Hashing for Object-based Image Retrieval

IJCAI 2017

- **Background & Problem Definition**
- Related Work
- Challenge & Solution
- Our approach
- Experiments

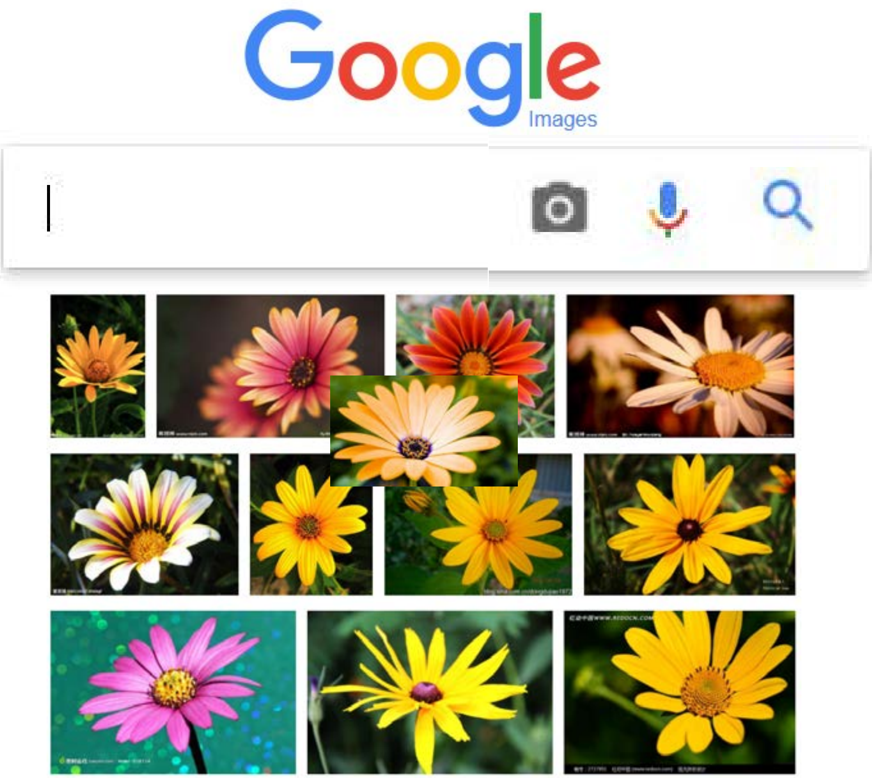
Background

Background



视频和图像占80%的非结构化数据，智能化管理与分享方面的核心技术亟待突破。

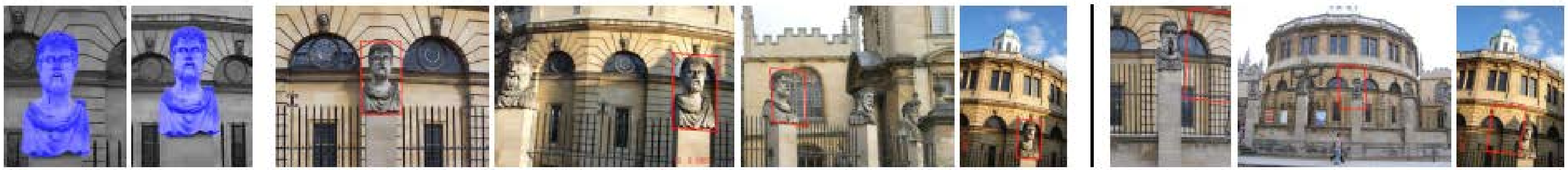
Background



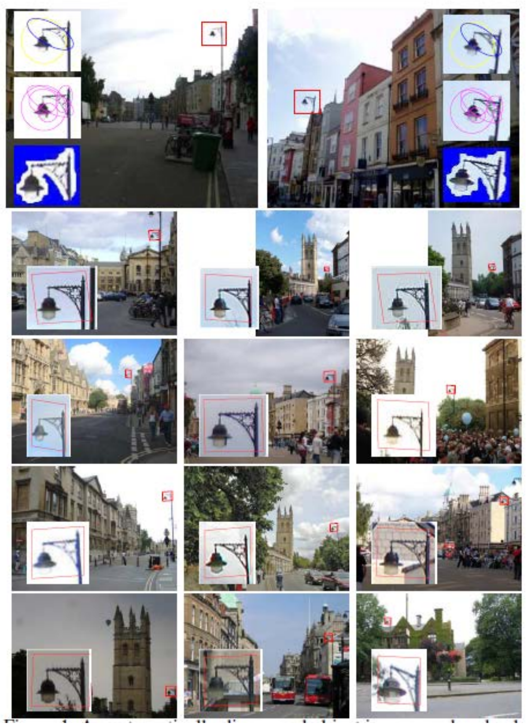
Convention Image Retrieval

Task: given a query image, find its nearest neighbors in an image repository.

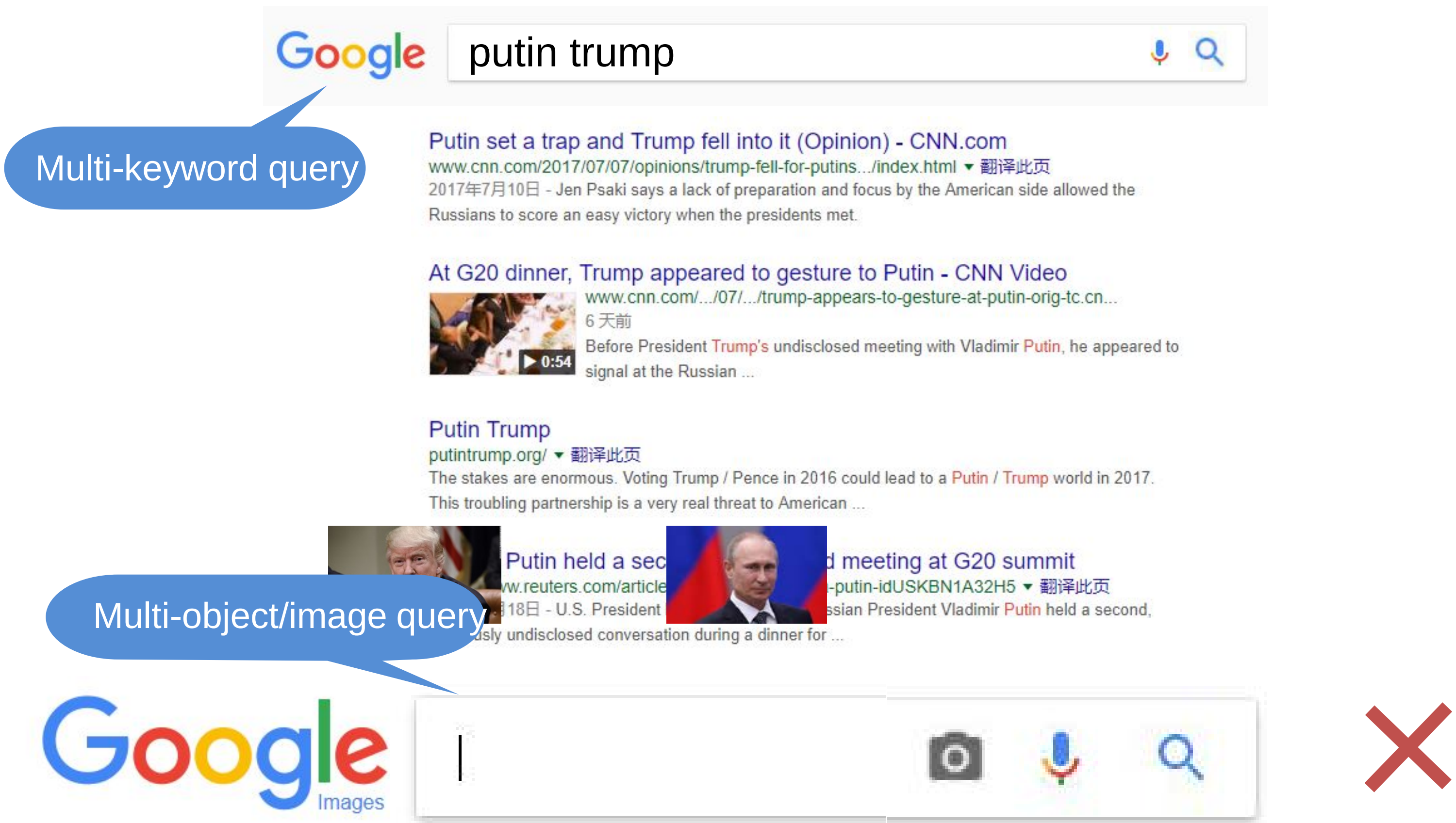
Background



Object-based
Image Retrival



Background


Google
putin trump
Multi-keyword query
Putin set a trap and Trump fell into it (Opinion) - CNN.com
www.cnn.com/2017/07/07/opinions/trump-fell-for-putins.../index.html ▾ 翻译此页
2017年7月10日 - Jen Psaki says a lack of preparation and focus by the American side allowed the Russians to score an easy victory when the presidents met.
At G20 dinner, Trump appeared to gesture to Putin - CNN Video
www.cnn.com/.../07/.../trump-appears-to-gesture-at-putin-orig-tc.cn...
6 天前
Before President Trump's undisclosed meeting with Vladimir Putin, he appeared to signal at the Russian ...
0:54
Putin Trump
putintrump.org/ ▾ 翻译此页
The stakes are enormous. Voting Trump / Pence in 2016 could lead to a Putin / Trump world in 2017. This troubling partnership is a very real threat to American ...
Putin held a sec
d meeting at G20 summit
-putin-idUSKBN1A32H5 ▾ 翻译此页
U.S. President
ssian President Vladimir Putin held a second,
undisclosed conversation during a dinner for ...
Multi-object/image query
Google
Images

Problem Definition

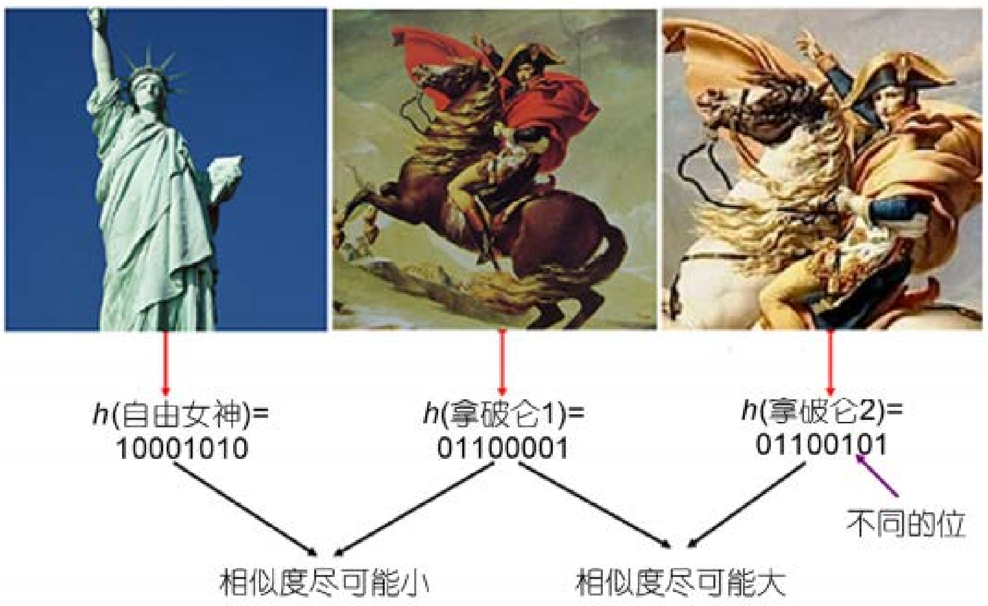
- **Querying about multiple objects**

Multi-object query

Retrieval Result

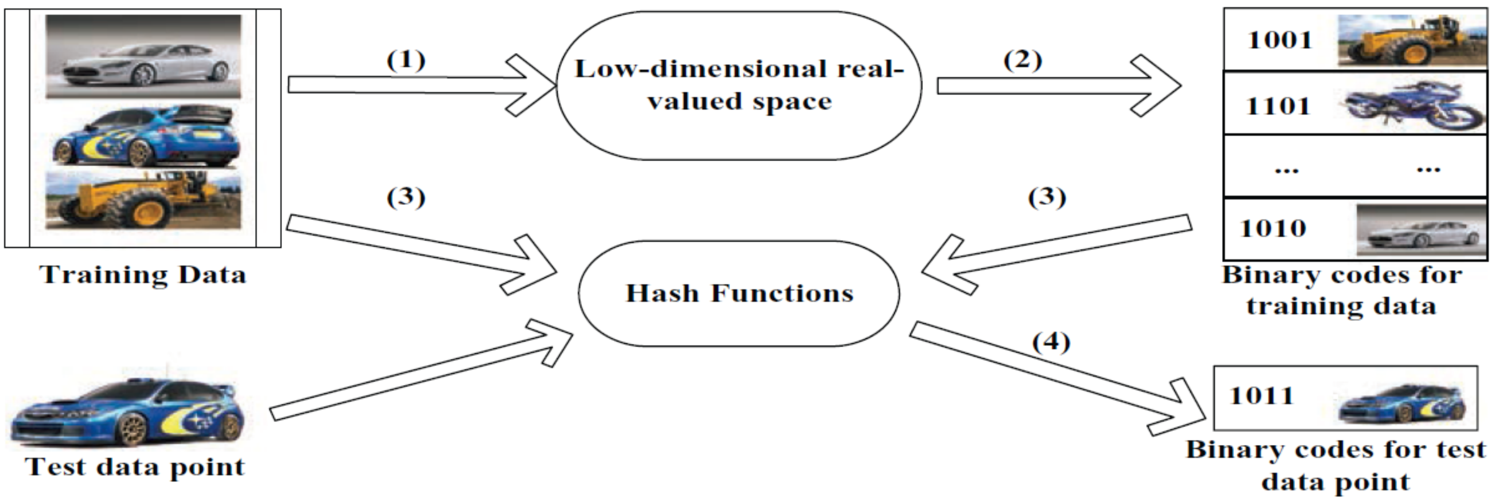
- Background & Problem Definition
- **Related Work**
- Challenge & Solution
- Our approach
- Experiments

选题背景及意义



哈希学习通过机器学习机制将数据映射成二进制码的形式，不仅能显著减少数据的存储和通信开销，还能降低数据维度，从而显著提高大数据检索系统的效率. 因此，哈希学习近年来成为大数据学习中的一个研究热点.

Hashing Learning



由于从原空间中的特征表示直接学习得到二进制的哈希编码是一个NP难问题.
现在很多的哈希学习方法都采用两步学习策略:

- 第一步, 先对原空间的样本采用度量学习(metric learning)进行降维, 得到1个低维空间的实数向量表示;
- 第二步, 对得到的实数向量进行量化(即离散化)得到二进制哈希码.

Related works

Unsupervised Hashing

SH[Weiss and Torralba. NIPS,2008]
LLH[Irie et al. CVPR,2014]

Previous hashing methods were focused on mapping whole images or fixed regions of images into the Hamming space

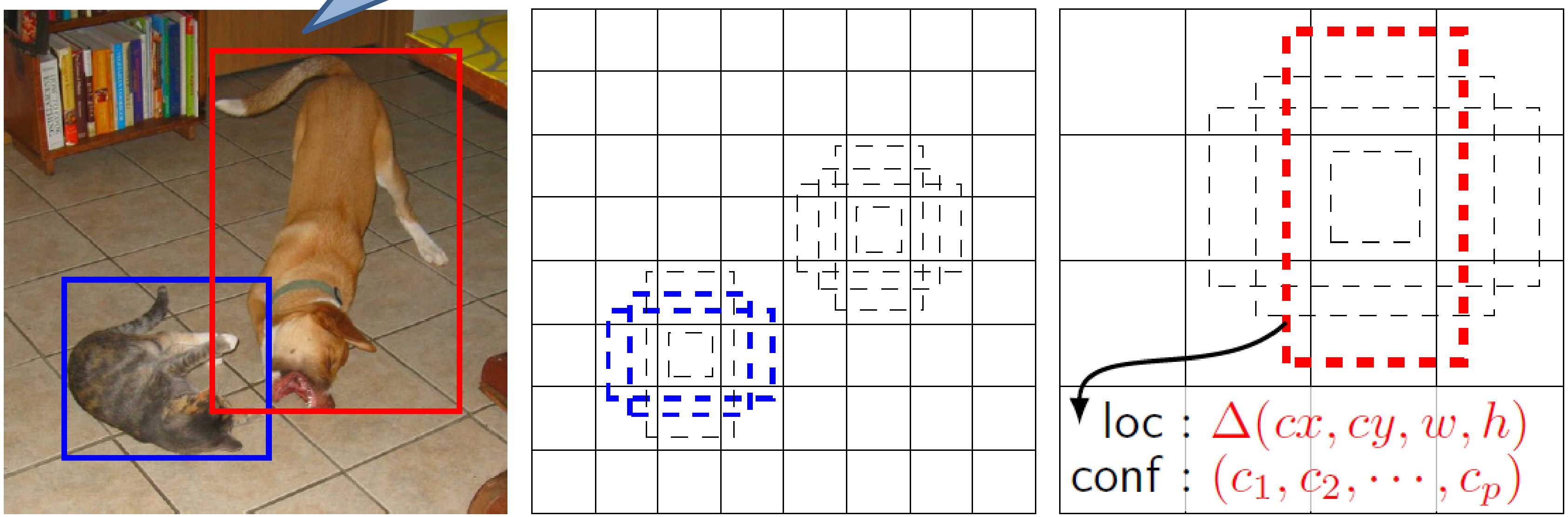
Supervised Hashing

BRE[Kulis and Darrell. NIPS,2009]
KSH[Liu et al. CVPR,2012]
DSRH[Zhao et al. CVPR,2015]

- Background & Problem Definition
- Related Work
- **Challenge & Solution**
- Our approach
- Experiments

Challenge & Solution

Current object detection methods require high quality GroundTruth boxes.



(a) Image with GT boxes (b) 8×8 feature map (c) 4×4 feature map

Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[J]. 2015.

Challenge & Solution

Supervised Learning Mode

Weakly Supervised Learning Mode

- **Challenge** : Our work goal is to build the end-to-end relation between a raw image and the binary hash codes of multiple objects in it by a weakly supervised learning mode.

Challenge & Solution

Solution : A multi-task training method.

➡ Task1 : Object prediction.

We cast the object detection of each object class as a binary MIL problem.

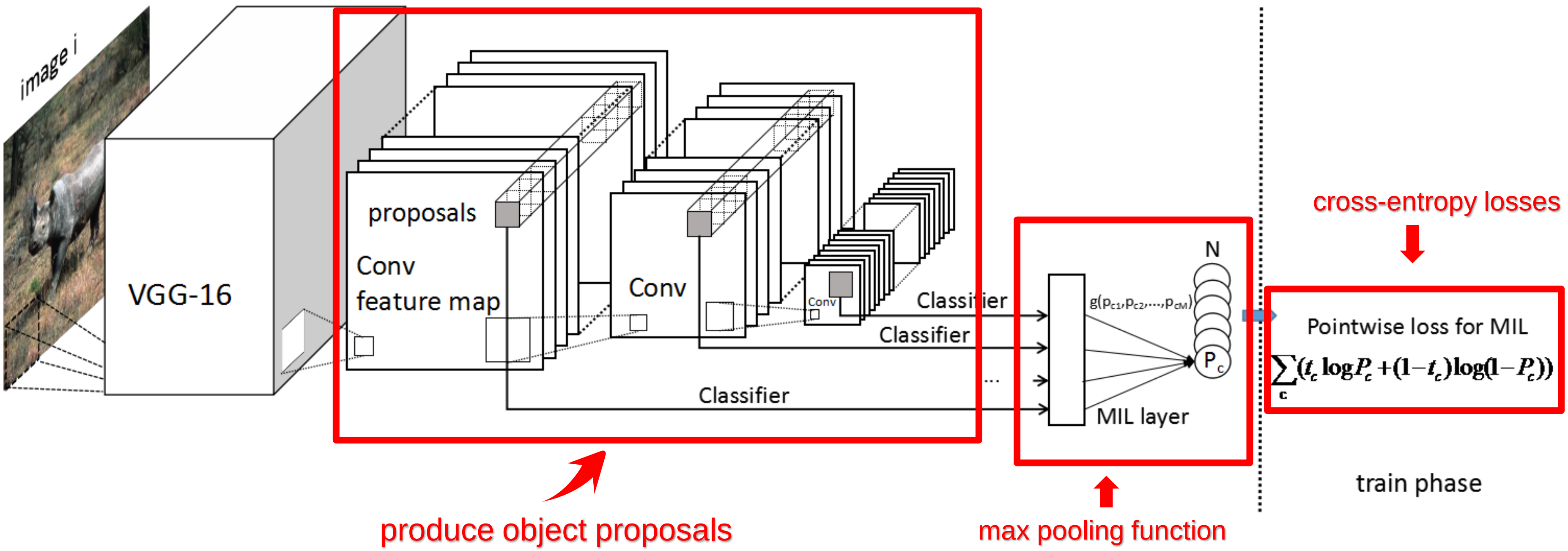
➡ Task2 : Hash Learning.

We pose the hashing learning as a problem of similarity learning with paired samples.

- Background & Problem Definition
- Related Work
- Challenge & Solution
- **Our approach**
- Experiments

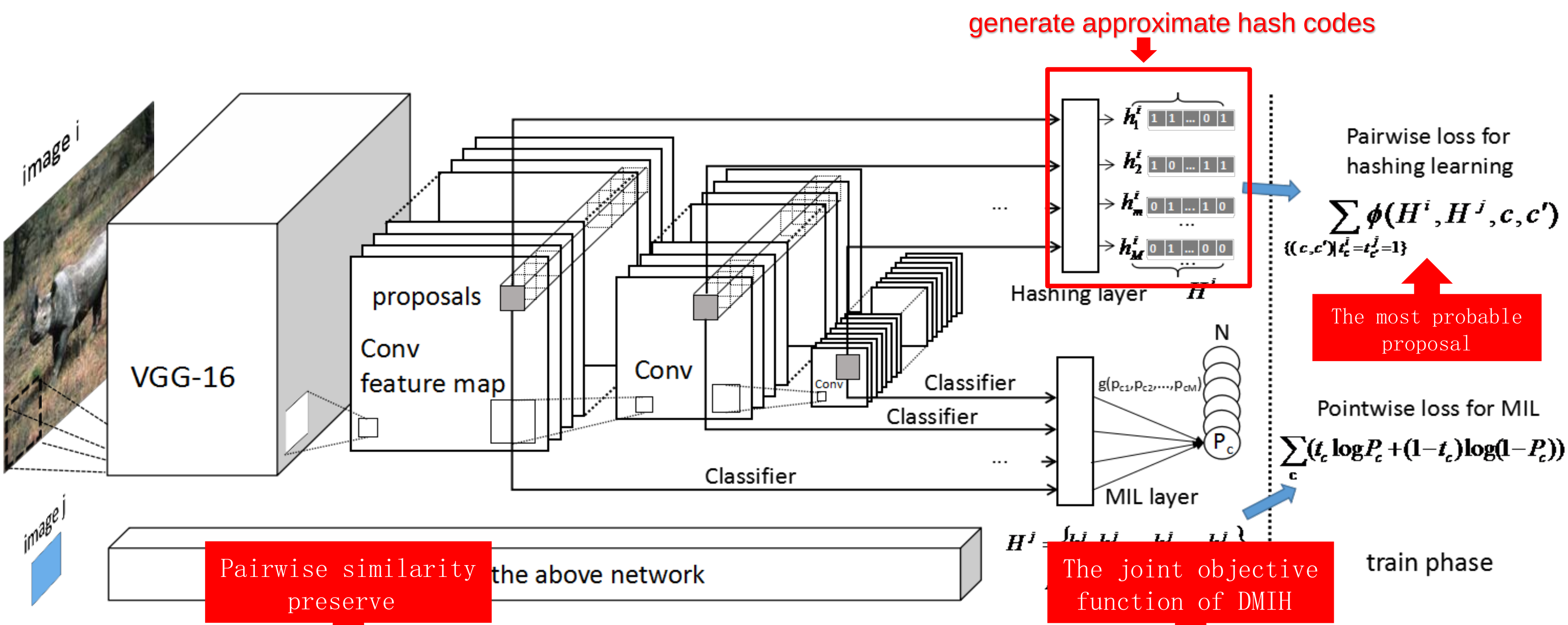
Our approach:

Object prediction based on multiple instance learning



Our approach:

Hashing Learning

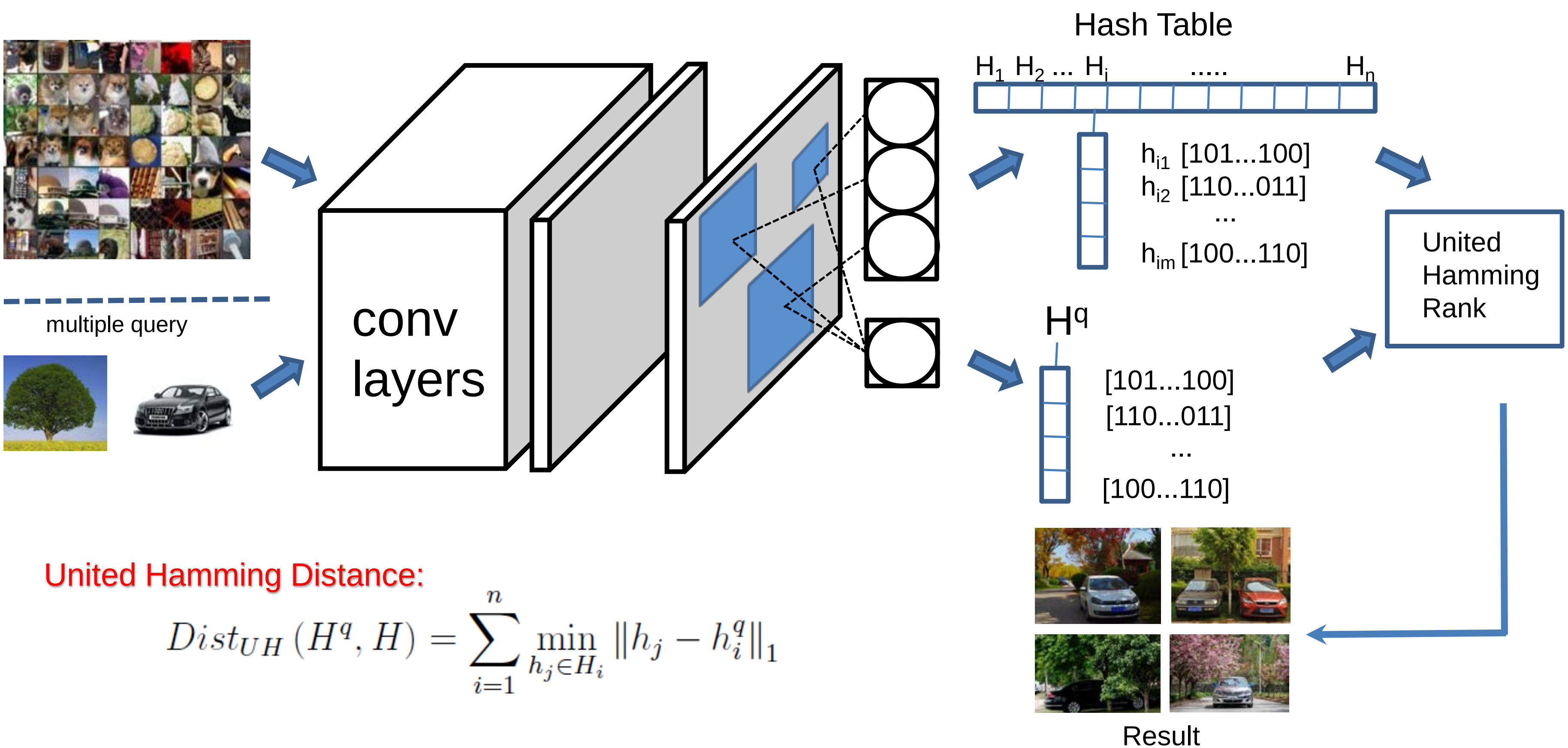


$$\Phi(H^i,H^j,c,c') = \begin{cases} Dst(\mathbf{h}^i_{Idx(P^i_c)},\mathbf{h}^j_{Idx(P^j_c)}) & if\ c=c' \\ \max(0,\beta - Dst(\mathbf{h}^i_{Idx(P^i_c)},\mathbf{h}^j_{Idx(P^j_c)})) & otherwise \end{cases}$$

$$J = \sum_{(i,j)} J_{pair-hash}(I_i, I_j) + \lambda J_{MIL}$$

Our approach:

Multiple query object-based Image Retrieval

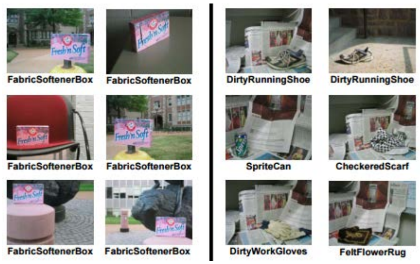


United Hamming Distance:

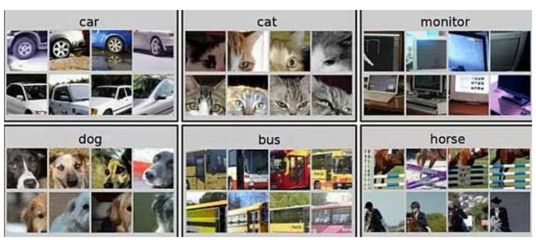
$$Dist_{UH}\left(H^{q}, H\right) = \sum_{i=1}^{n} \min_{h_j \in H_i} \|h_j - h_i^q\|_1$$

- Background & Problem Definition
- Related Work
- Challenge & Solution
- Our approach
- **Experiments**

Experiments : Data set



SIVAL： It consists of 25 different categories, and each category includes 60 images. The object can occur anywhere against highly diverse backgrounds in each image.

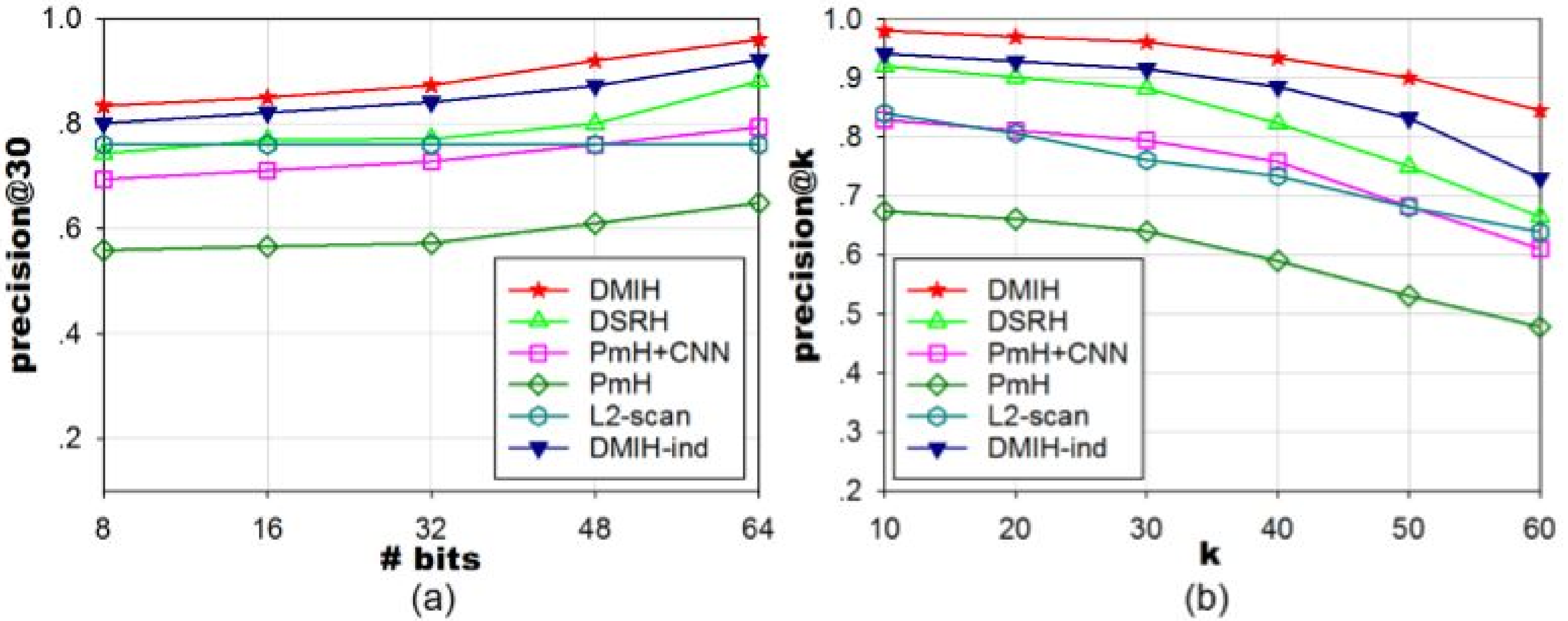


Pascal VOC 2007： It contains 9,963 images with 20 different object categories. we run experiments with multi-object queries on this dataset.

ILSVRC 2013： It contains nearly 400K images in 200 object categories. We use this dataset to primarily evaluate the runtime efficiency of DMIH and baseline algorithms.

Experiments: Performance comparison

- **Performance comparison on SIVAL for single-object query.**



Experiments: Performance comparison

- **Performance comparison on Pascal VOC 2007 for multi-object query.**

Multi-object Query	PASCAL VOC 2007 (MAP %)			
	DMIH	DSRH	PmH+ CNN	PmH
bottle + tv	78.6	67.1	61.2	54.3
horse + person	81.2	69.3	64.3	56.6
bus + car	83.1	71.4	66.0	58.4
dog + cat	79.9	64.2	57.8	58.1
bottle + chair + tv	85.7	72.4	69.1	56.1
dog + cat + person	84.3	71.1	67.9	55.2
bus + car + bike	87.1	73.5	70.5	58.4
horse + person + car + dog	91.8	79.2	74.4	61.2
chair + plant + sofa + tv	93.4	80.5	74.9	62.5
Average	85.7	72.6	68.9	57.1

Experiments

- **Comparison of the average query time on ILSVRC 2013**

Method	TIME(ms)			MAP(%)
	O	H+S	F+O+H+S	64 bits
DMIH	-	-	18.32	74.06
DSRH	6.32	-	27.76	68.12
PmH+CNN	-	18.45	22.97	62.26
PmH	-	19.76	297.43	52.22
ℓ_2-scan	-	-	975.91	66.51

Experiments:

Case studies on PASCAL VOC 2007 for queries with different combinations of object types.

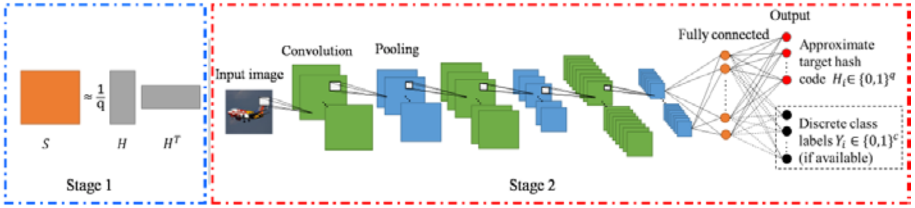
(a) queries (b) DMIH (c) BING+DSRH (d) PmH+CNN (e) PmH (f) l2-scan

Tag-based Weakly-supervised Hashing for Image Retrieval

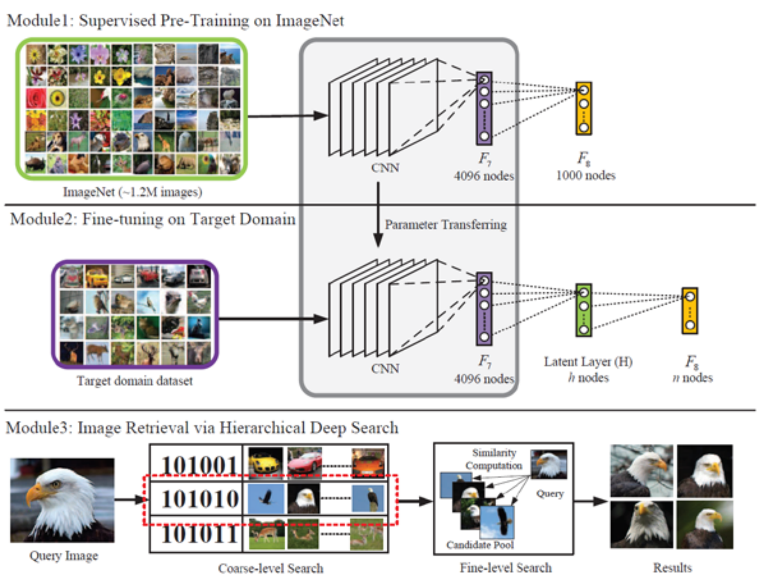
IJCAI 2018

Supervised Deep CNN based Hashing Methods

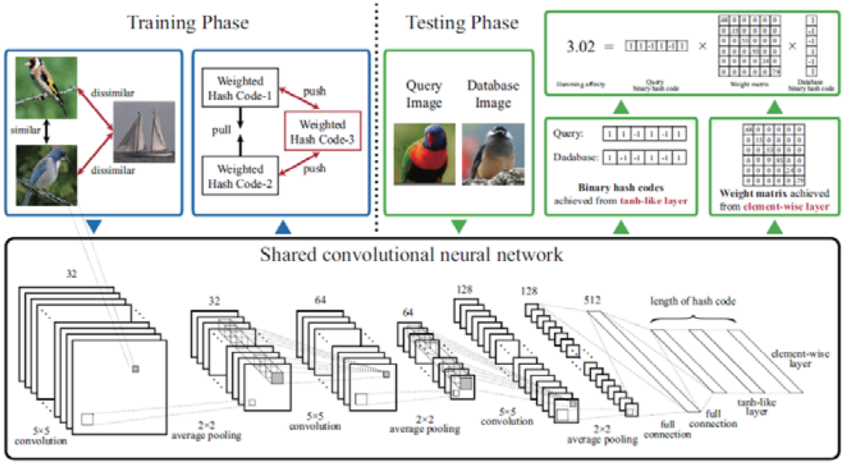
The supervised approaches based on deep models have achieved the state-of-the-art results with the help of labeled images.



R Xia, et al. "Supervised Hashing for Image Retrieval via Image Representation Learning" AAAI 2012.



K Lin, et al. "Deep learning of binary hash codes for fast image retrieval." CVPR 2015.



R Zhang, et al. "Bit-Scalable Deep Hashing with Regularized Similarity Learning for Image Retrieval and Person Redentification." IEEE TIP, 2015, 24 (12) :4766-4779.

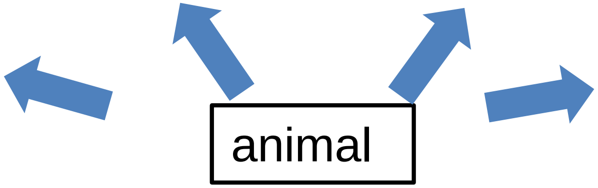
Issues

Two Issues:

- Training an effective and generalized hashing model requires a large quantity of labeled images, whereas the labeling work is very tedious and expensive.

- The controlled label set usually only includes coarse-grained concept labels in images, which cannot well characterize the fine-grained similarity relationships between images.



Motivation

Massive and growing amount
of digital images with user tags:

– Flickr and YouTube
– Audiovisual archives (BBC, INA)
– Personal collections

Benefits and Challenges

				
User Provided Tags	• rue4 • tricycle • bravo • driver • taxi ...	• factory • muscle car • mustang • cougar • firefly ...	• hawker • KFC • jump jet • uss essex • 2016 ...	• bug • állat • spider • WOW • animal ...
Manually Labeled Tags	✓vehicle	✓cars ✓sky ✓house	✓airport ✓plane ✓person ✓sky	✓animal

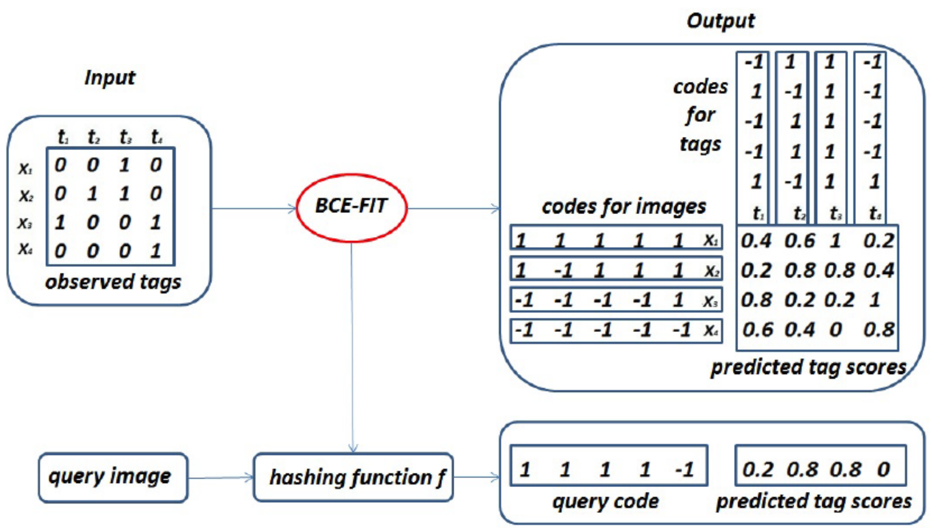
Benefits:

- abundant training data
- richer semantics

Challenges:

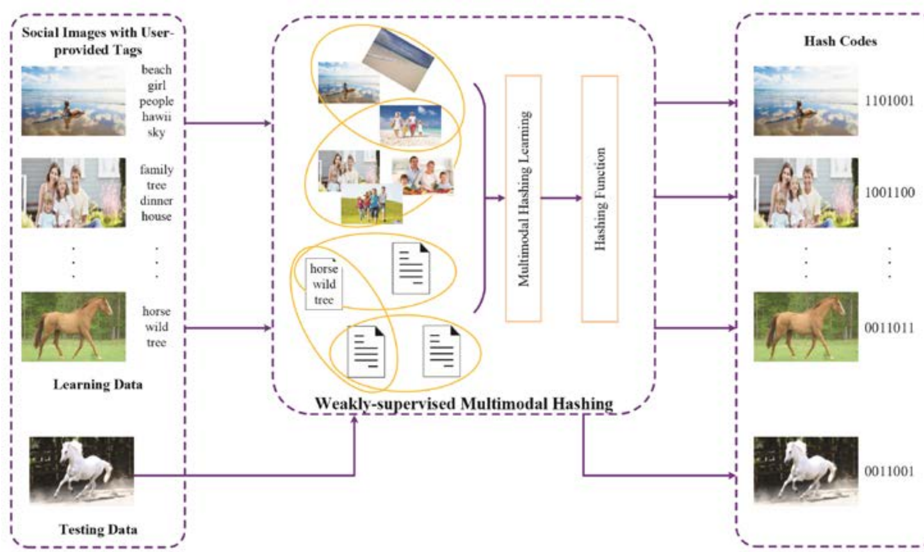
- noises
- vagueness
- incompleteness

Related Works



Wang, Q., Shen, B., Wang, S., Li, L., & Si, L., Binary codes embedding for fast image tagging with incomplete labels. ECCV' 14.

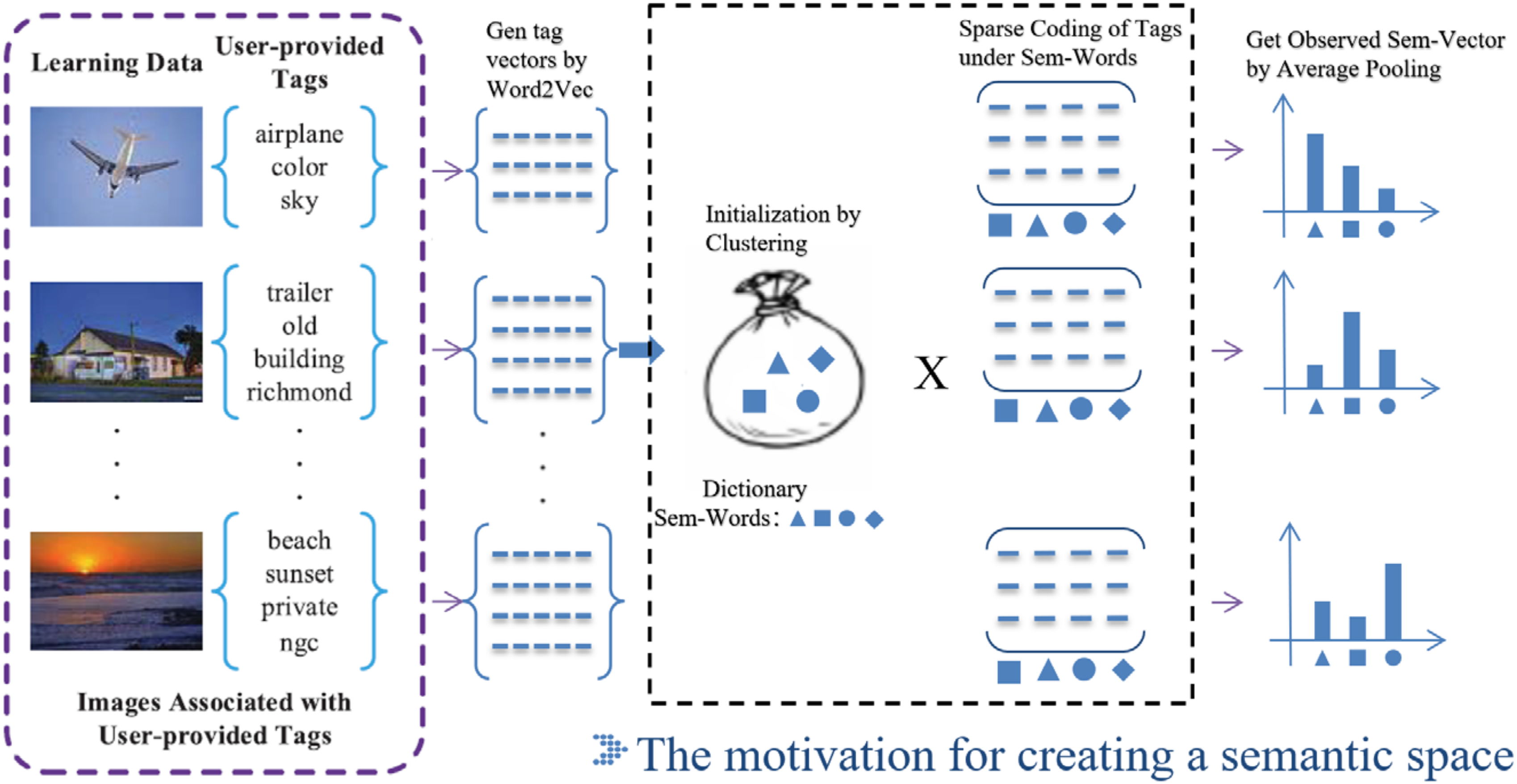
Tang, J. and Li, Z., 2017. Weakly-supervised multimodal hashing for scalable social image retrieval. IEEE Trans. Circuits and Systems for Video Technology.



Our method: roadmap

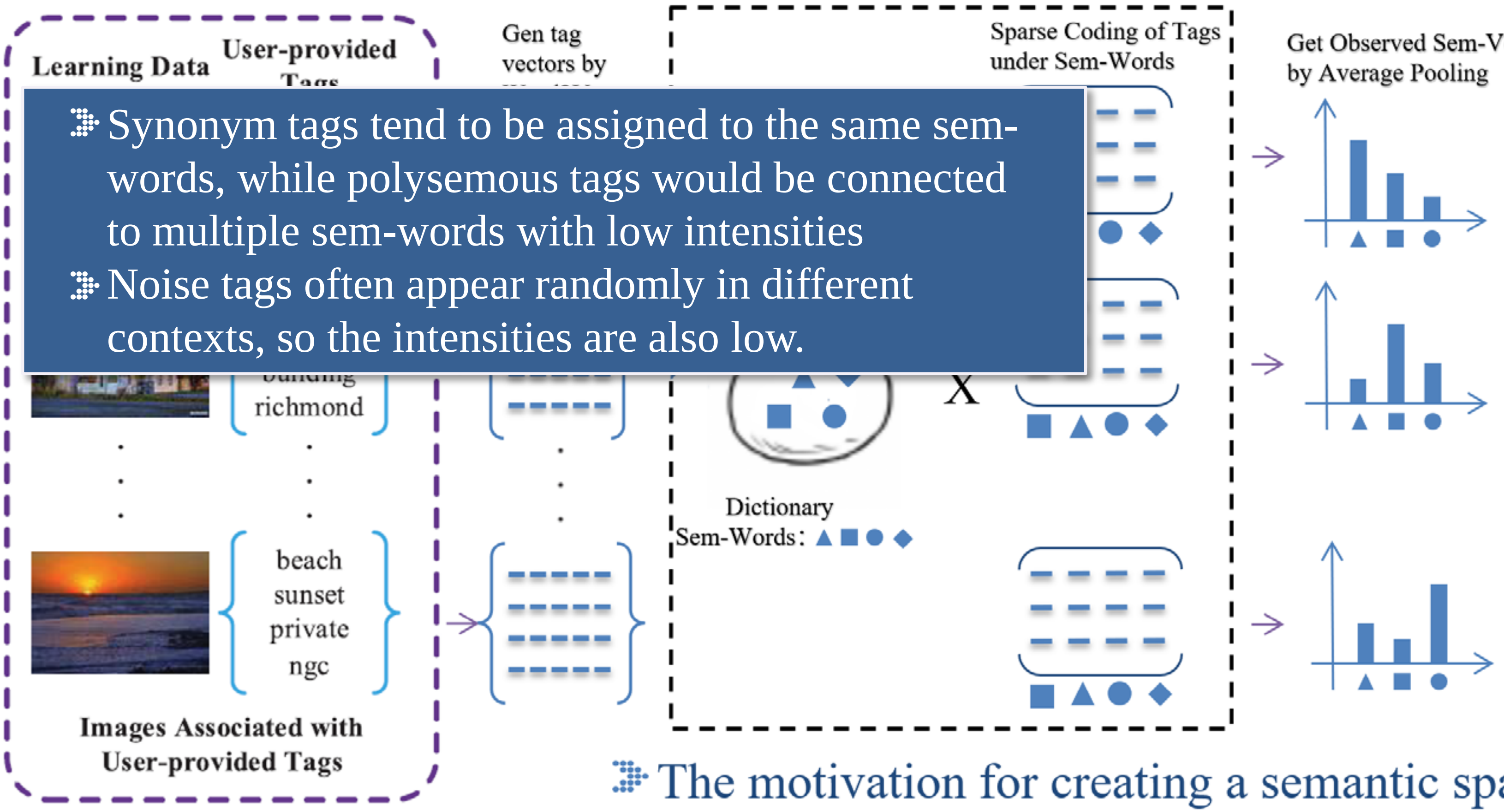
Weakly-supervised pre-training → Supervised fine-tuning

Our Method: Semantic Space Construction



- The motivation for creating a semantic space: alleviate the vagueness and noise issue of tags.

Our Method: Semantic Space Construction



- Synonym tags tend to be assigned to the same sem-words, while polysemous tags would be connected to multiple sem-words with low intensities
- Noise tags often appear randomly in different contexts, so the intensities are also low.

- The motivation for creating a semantic space: alleviate the vagueness and noise issue of tags.

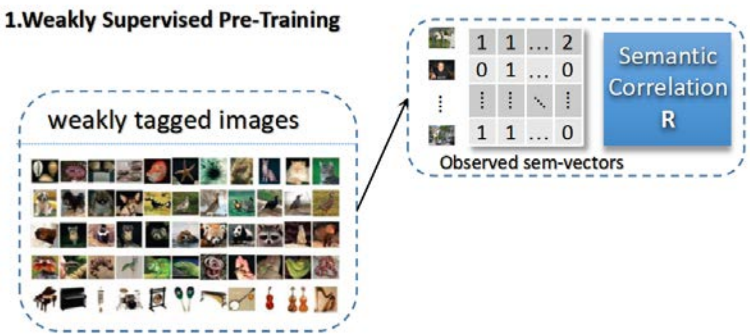
Our Method: Weakly Supervised Pre-Trainning

Then, a unified optimization problem is established where we synthesize tagging information and image content for learning sem-vectors and hash codes, trying to alleviate the **incompleteness** issues.

Including 3 sub-objectives:

First: $\min \|\hat{\mathbf{z}}_i - \mathbf{z}_i\|^2$ The the sem-vector $\hat{\mathbf{z}}_i$ should not deviate too much from observed sem-vector $\mathbf{z}_i$

Our Method: Weakly Supervised Pre-Trainning



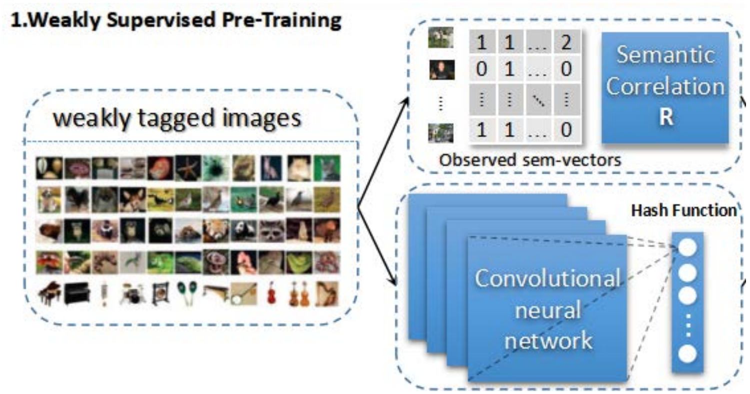
Including 3 sub-objectives:

First: $\min \|\hat{\mathbf{z}}_i - \mathbf{z}_i\|^2$

Second: $\min \frac{1}{2}\sum_{k=1}^{M}\sum_{q=1}^{M} R_{kq}(\hat{z}_{ik} - \hat{z}_{iq})^2 = \hat{\mathbf{z}}_i^\top \mathbf{L}\hat{\mathbf{z}}_i \quad R_{kq} = \frac{f_{k,q}}{f_k + f_q - f_{k,q}}$

We exploit tag correlation to cope with the incompleteness issue.

Our Method: Weakly Supervised Pre-Trainning



Including 3 sub-objectives:

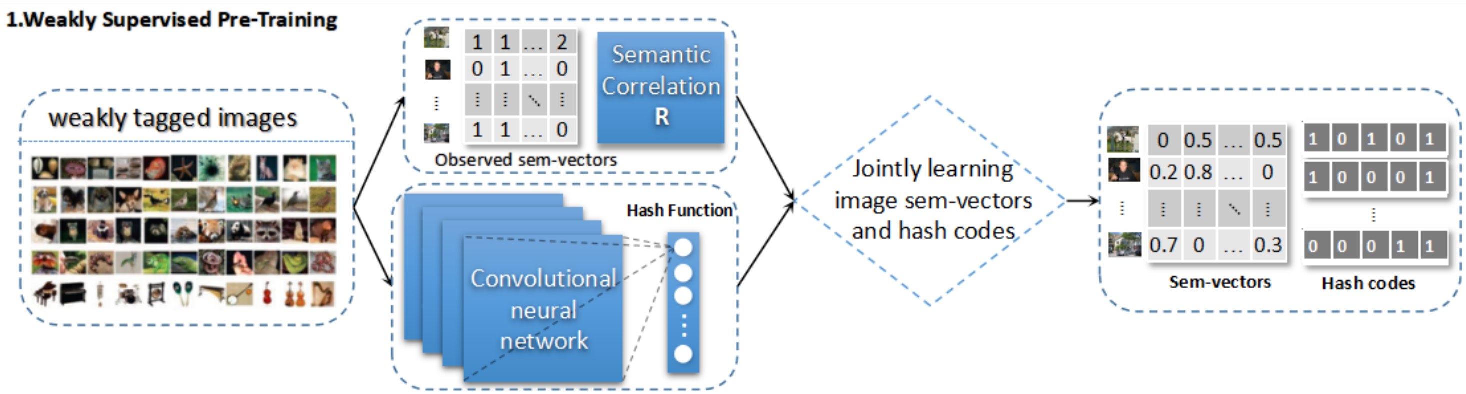
First: $\min \|\hat{\mathbf{z}}_i - \mathbf{z}_i\|^2$

Second: $\min \frac{1}{2} \sum_{k=1}^{M} \sum_{q=1}^{M} R_{kq} (\hat{z}_{ik} - \hat{z}_{iq})^2 = \hat{\mathbf{z}}_i^\top \mathbf{L} \hat{\mathbf{z}}_i$

Third: $\min \sum_{i,j} \|\hat{\mathbf{z}}_i^\top \hat{\mathbf{z}}_j - \mathbf{h}_i^\top \mathbf{h}_j\|^2$

Forcing the similarities of images estimated using sem-vectors to comply with those obtained from image features (hash codes).

Our Method: Weakly Supervised Pre-Trainning



① $\min \|\hat{\mathbf{z}}_i - \mathbf{z}_i\|^2$

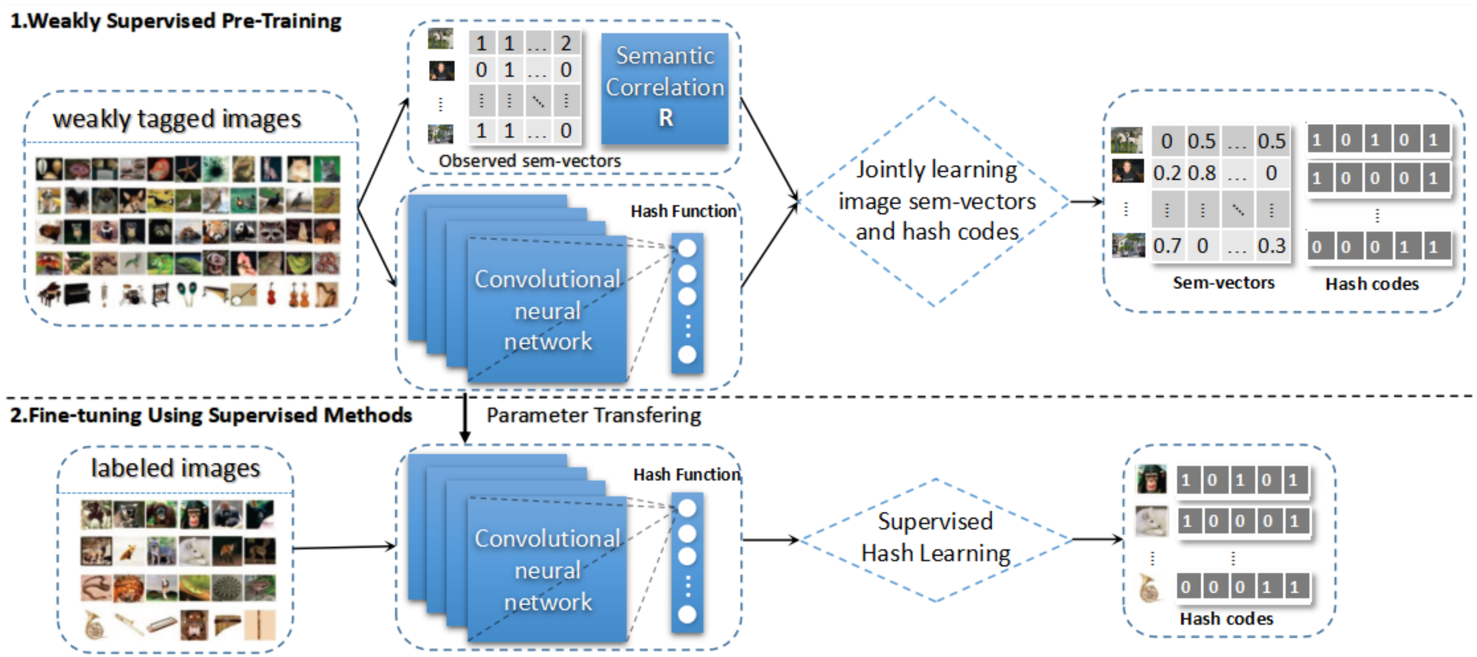
② $\min \frac{1}{2}\sum_{k=1}^{M}\sum_{q=1}^{M} R_{kq}(\hat{z}_{ik} - \hat{z}_{iq})^2 = \hat{\mathbf{z}}_i^\top \mathbf{L}\hat{\mathbf{z}}_i$

③ $\min \sum_{i,j} \|\hat{\mathbf{z}}_i^\top \hat{\mathbf{z}}_j - \mathbf{h}_i^\top \mathbf{h}_j\|^2$

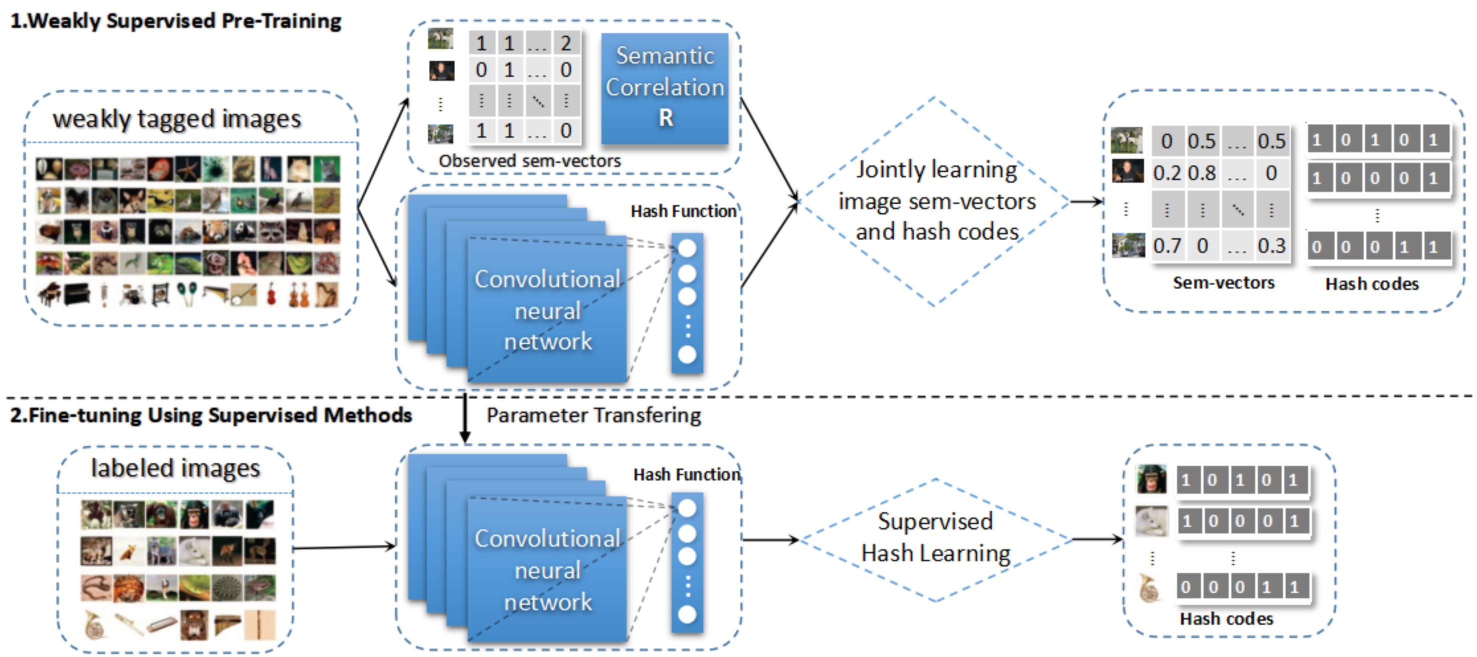
$$\mathcal{L}(\{\hat{\mathbf{z}}\},\{\mathbf{h}\}) = \frac{1}{2}\sum_{i=1}^{N}\|\hat{\mathbf{z}}_i - \mathbf{z}_i\|^2 + \frac{\beta}{2}\sum_{i=1}^{N}\hat{\mathbf{z}}_i^\top \mathbf{L}\hat{\mathbf{z}}_i + \frac{\delta}{2}\sum_{i,j}\|\hat{\mathbf{z}}_i^\top \hat{\mathbf{z}}_j - \mathbf{h}_i^\top \mathbf{h}_j\|^2$$

Overall Loss Function

Our Method: Fine-tuning Using Supervised Methods



Our Method: Fine-tuning Using Supervised Methods



Proposed framework that can incorporate any state-of-art deep hashing models.

Our Method: Optimization

alernatively update$\{\hat{\mathbf{z}}\}$and$\{\mathbf{h}\}$by two steps:

Step 1: Fix$\{\mathbf{h}\}$, Update$\{\hat{\mathbf{z}}\}$.

$$\nabla_{\hat{\mathbf{z}}_i}\mathcal{L}(\{\hat{\mathbf{z}}\},\{\mathbf{h}\}) = (\hat{\mathbf{z}}_i - \mathbf{z}_i) + \beta\mathbf{L}\hat{\mathbf{z}}_i + \delta\sum_{j\in\mathcal{S}}(\hat{\mathbf{z}}_i^\top\hat{\mathbf{z}}_j - \mathbf{h}_i^\top\mathbf{h}_j)\hat{\mathbf{z}}_\mathbf{j}$$

Step 2: Fix$\{\hat{\mathbf{z}}\}$, Update$\{\mathbf{h}\}$.

$$\nabla_{\mathbf{h}_i}\mathcal{L}(\{\hat{\mathbf{z}}\},\{\mathbf{h}\}) = \delta\sum_{j\in\mathcal{S}}(\hat{\mathbf{z}}_i^\top\hat{\mathbf{z}}_j - \mathbf{h}_i^\top\mathbf{h}_j)\mathbf{h}_j$$

Our Method: Optimization

Algorithm 1 Weakly-supervised hashing learning

Input: A set of user-tagged training images
Parameters: λ, β
Output: Pre-trained hashing model
Compute the observed sem-vectors $\{\mathbf{z}\}$ for the images.
Initialize $\hat{\mathbf{z}}$ and $\mathbf{h}$ (CNN model).
repeat
Initialize the updating frequencies: l, m
for 1,...,l **do**
Update $\{\hat{\mathbf{z}}\}$ according to Step 1.
end for
for 1,...,m **do**
Update $\{\mathbf{h}\}$ (CNN model) according to Step 2.
end for
until the solution converges

locally optimal solution:

- At the beginning, the ratio l : m is empirically set to 10:1.
- Afterwards, the ratio will gradually turn to 1:10 to emphasize hashing model training.

Experiments: Data set

	NUS-WIDE	MIR Flickr-1M
# Images	269,648	897,500
# Tags	5018	1000
Avg./max # T/I	7.9/201	12.7/76
Avg./max # I/T	677.1/20,140	416.5/76,890
# of manually labeled images	269,648	25,000
# manual labels	81	38

Experiments: Compared Algorithms

Supervised Methods:

Kernel-based Supervised Hashing(KSH) [Liu et al., CVPR, 2012]

Deep Supervised Hashing(DSH) [Liu et al., CVPR, 2016]

Deep Regularized Similarity Comparison Hashing(DRSCH) [Zhang et al., CVPR, 2015]

Weakly-supervised Pre-trained (WP) versions of Supervised methods:

WP-KSH

WP-DSH

WP-DRSCH

Unsupervised Methods:

Locality-Sensitive Hashing(LSH) [Gionis et al., VLDB, 2000]

Unsupervised CNN-based hashing(DeepBit) [Lin et al., 2016]

Experiments: Result

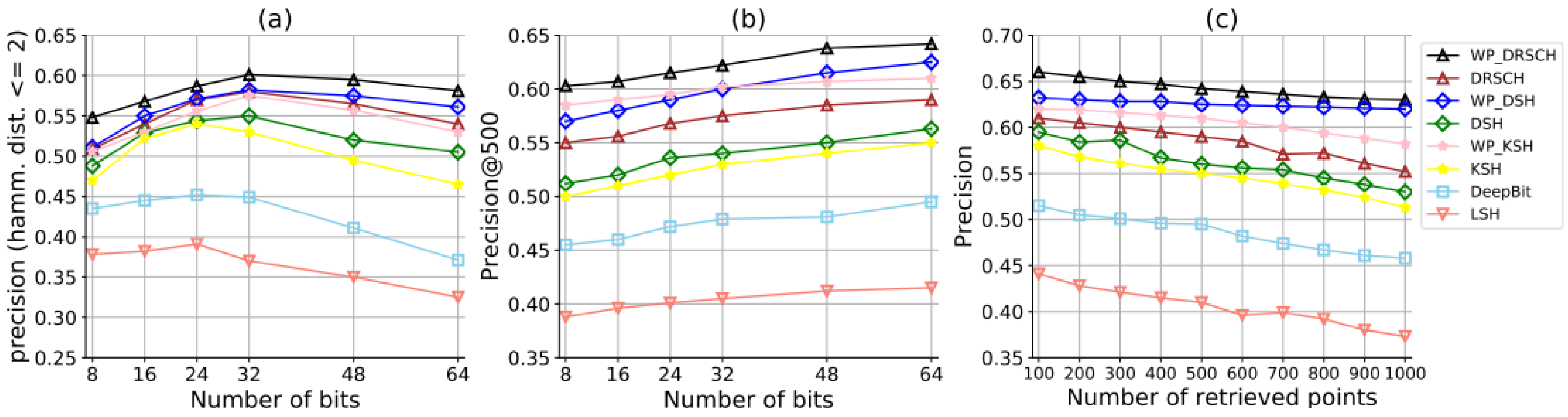


Figure 3: Performance comparison on the NUS-WIDE dataset. (a) Precision within Hamming radius 2; (b) Precision@500; (c) Precision curves with 64 hashing bits.

Experiments: Result

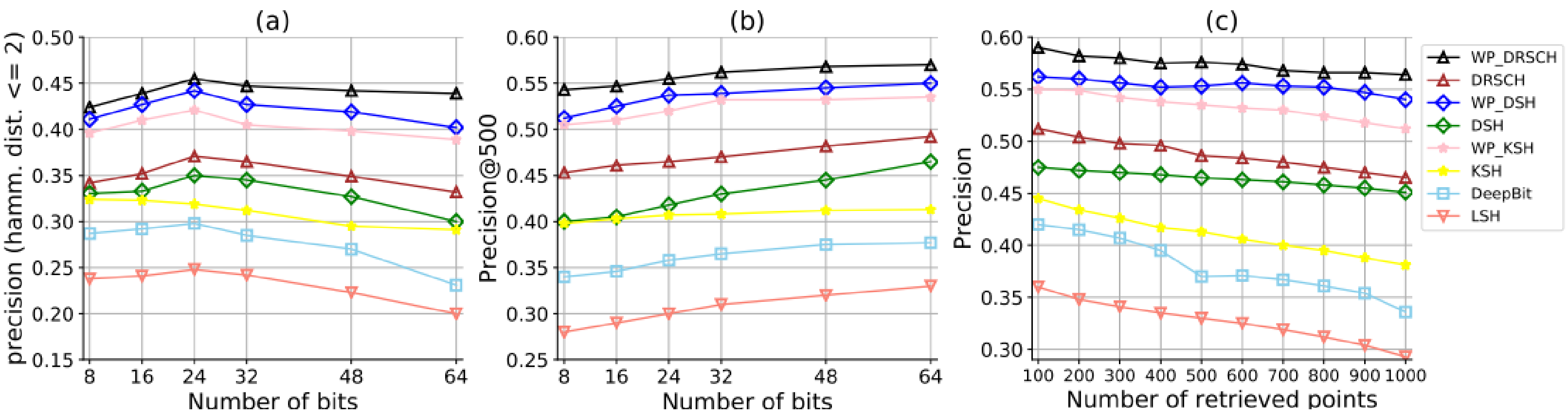


Figure 4: Performance comparison on the MIR Flickr-1M dataset. (a) Precision within Hamming radius 2; (b) Precision@500; (c) Precision curves with 64 hashing bits.

Experiments: Result

Method	Hamming ranking (MAP, %)			
	8 bits	16 bits	32 bits	64 bits
KSH^-	12.63	15.32	17.11	18.32
$WP\text{-}KSH^-$	**40.64**	**43.55**	**46.76**	**49.34**
DSH^-	15.12	18.45	20.15	23.34
$WP\text{-}DSH^-$	**44.02**	**47.72**	**49.76**	**52.96**
$DRSCH^-$	18.87	21.56	25.45	28.97
$WP\text{-}DRSCH^-$	**50.09**	**53.21**	**55.50**	**56.43**

Table 2: Comparison of the proposed weakly-supervised hashing learning algorithm to supervised methods on training with tagging data only. We only report results on NUS-WIDE due to space limitation. The minus sign means the fine-tuning process is removed (for WP- methods) or we apply the method directly on tagging data (for supervised methods).

Experiments: Result

Figure 5: Retrieval results obtained by (a) WP-DRSCH and (b) DRSCH in MIR Flickr-1M for two queries. Each query is shown on the left, with top ranked images shown to the right.

谢谢！