Supervised Set-to-Set Hashing in Visual Recognition

I-Hong Jhuo

CODAIT, IBM

Outline

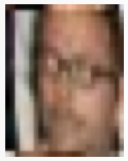
- Introduction
- Set-to-Set Hashing
- Results
- Summary

Outline

- Introduction
- Set-to-Set Hashing
- Results
- Summary

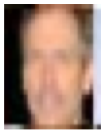
Point to Point (P2P) Hashing

Image (face) with real-valued feature vector

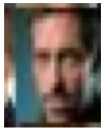


?

Similarity



0.1



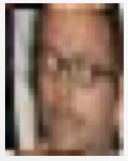
0.1

0.8 ✓

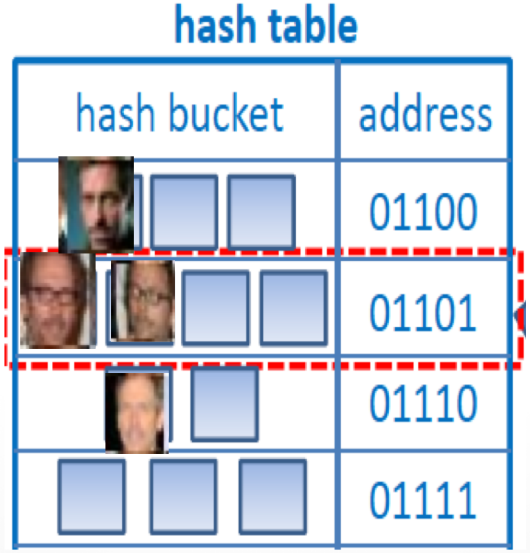
- time: O(n), increase with database size
- storage: high dimension of float number

x = [0.23, -0.2, 0.1, …, 0.8]

Feature vector to hash codes



?



✓

- time: O(1), independent of database size
- storage: low dimension of bits

x = [0 1 1 0 1]

Introduction

- P2P (ANN) search characteristics:
  - Computational time (sublinear search complexity)
  - Storage space

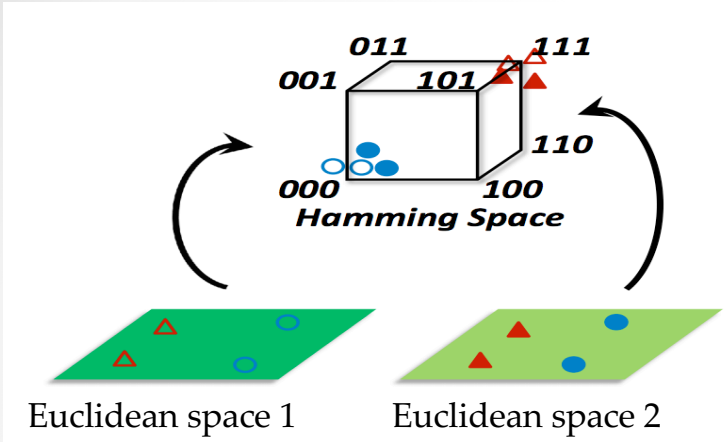
- What are the unsolved problems ?
  - How to measure distance between sets
    - Is there any useful information

Outline

- Introduction
- Set-to-Set Hashing
- Results
- Summary

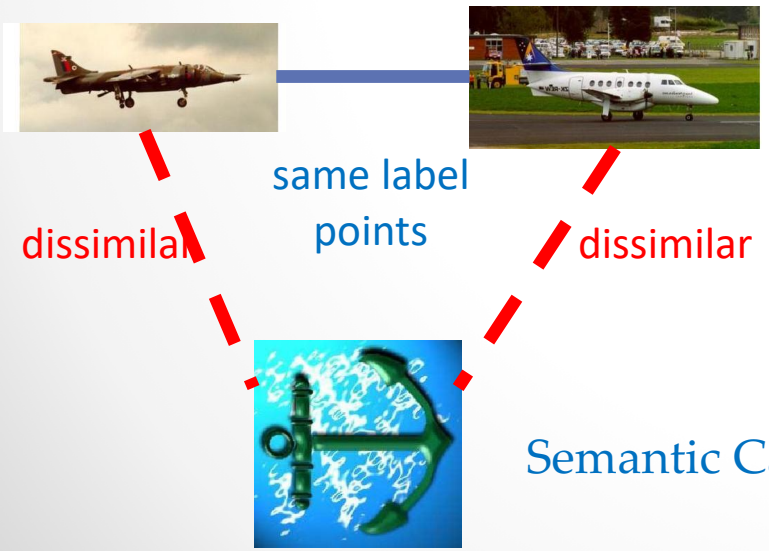
Set to Set (S2S) Hashing

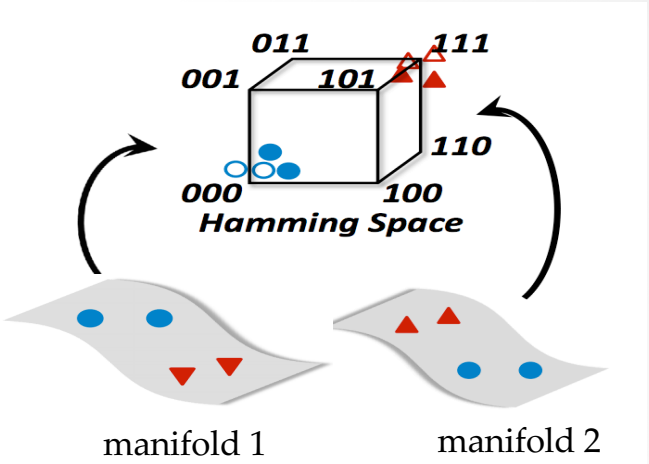
Supervised Information



Traditional hash learning methods

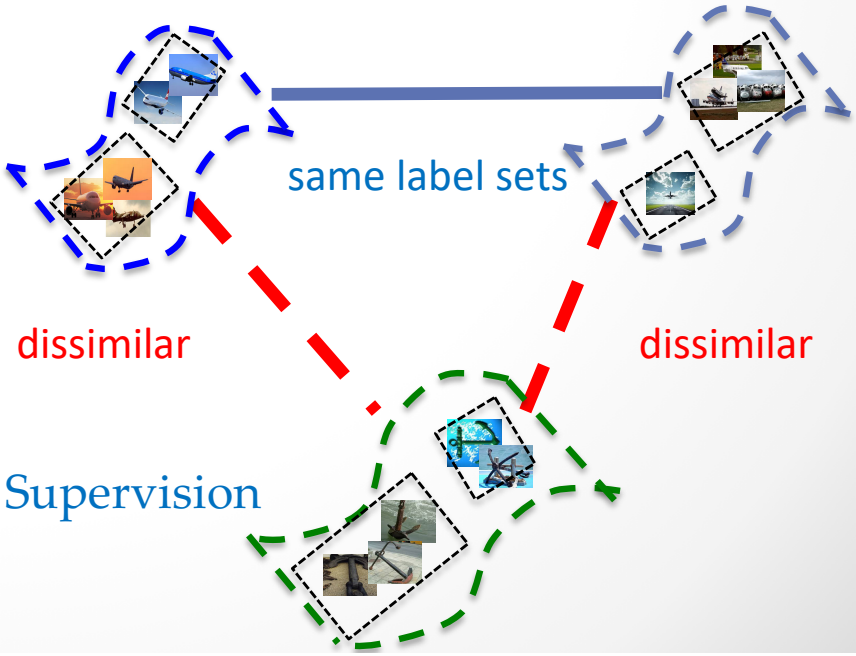
Point-to-point setting





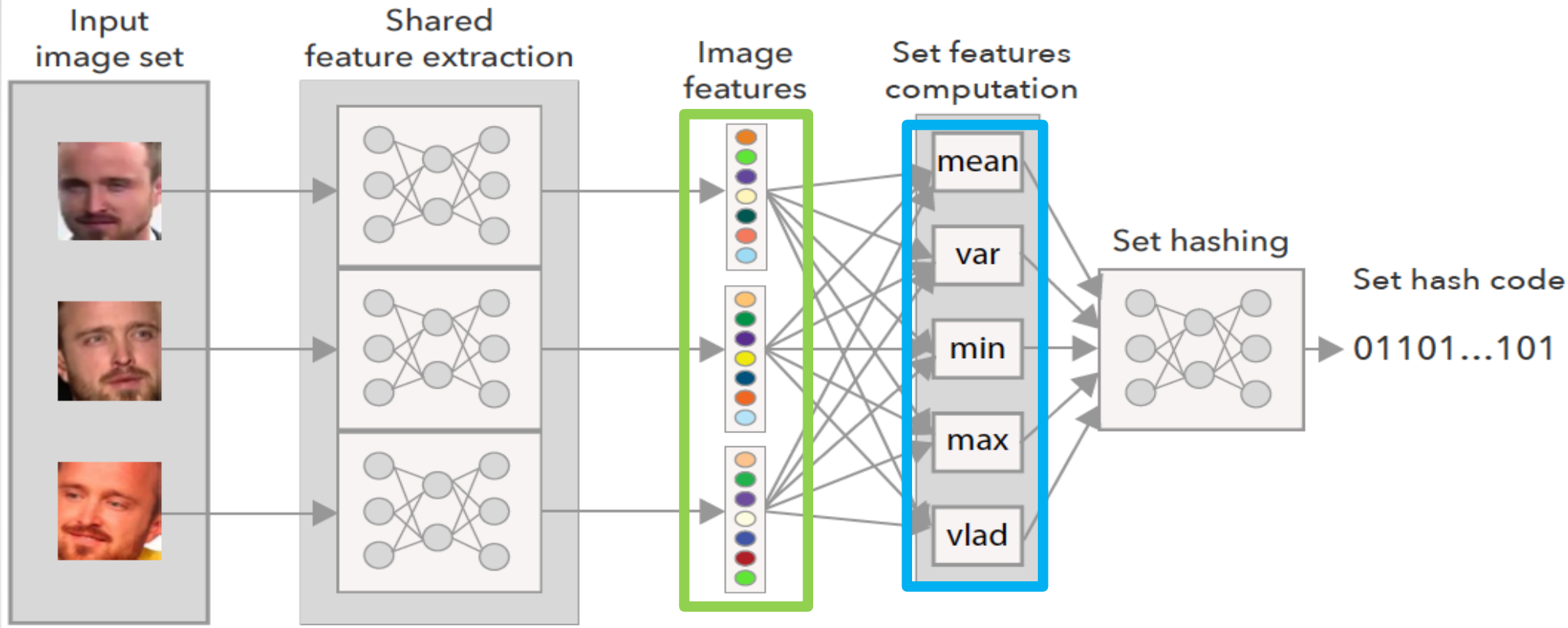
Proposed image set hash learning method

Set-to-set setting



Semantic Category Supervision

Image Set Hashing



Reference:

- Deep Image Set Hashing, [arXiv:1606.05381v2, 2016]
- End-to-End image Set hashing, but **No** structural Information

Introduction

- **What are we proposing?**

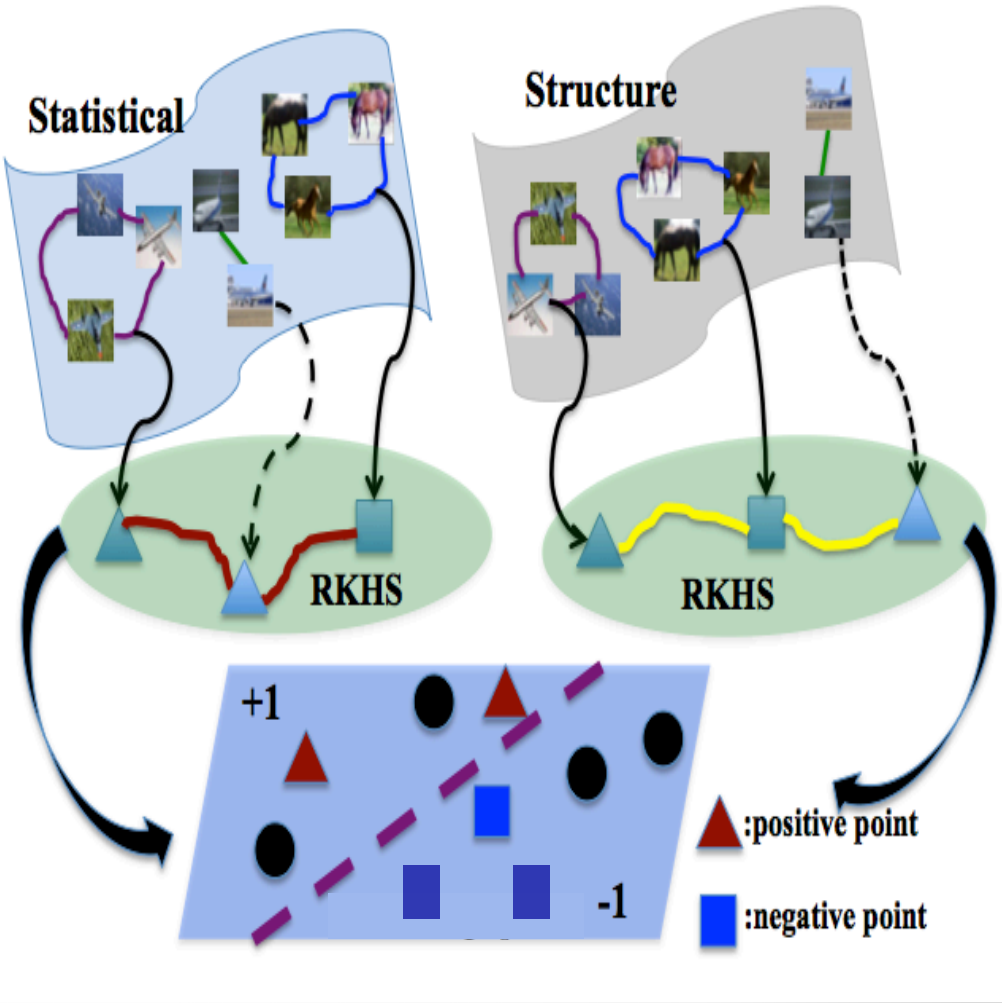
  - Set-to-Set Hashing => binary codes for a set

  - Multiple information => statistical & structural information to improve the retrieval performance

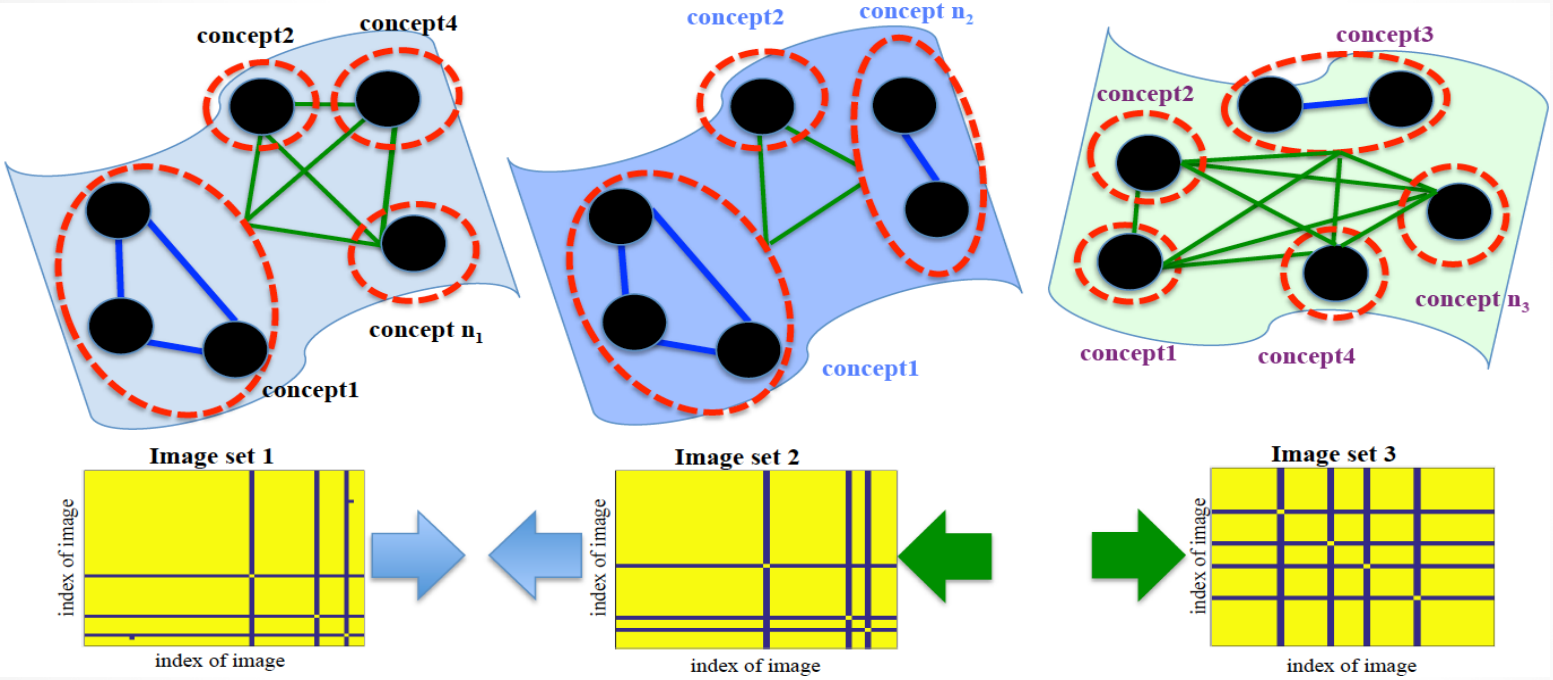
  - State-of-the-art results

Multiple Information for Image Set Hashing

- Statistical information:
  - Gaussian-logarithm kernel
- Structural information:
  - Graph kernel to capture topological (clique decomposition) similarity



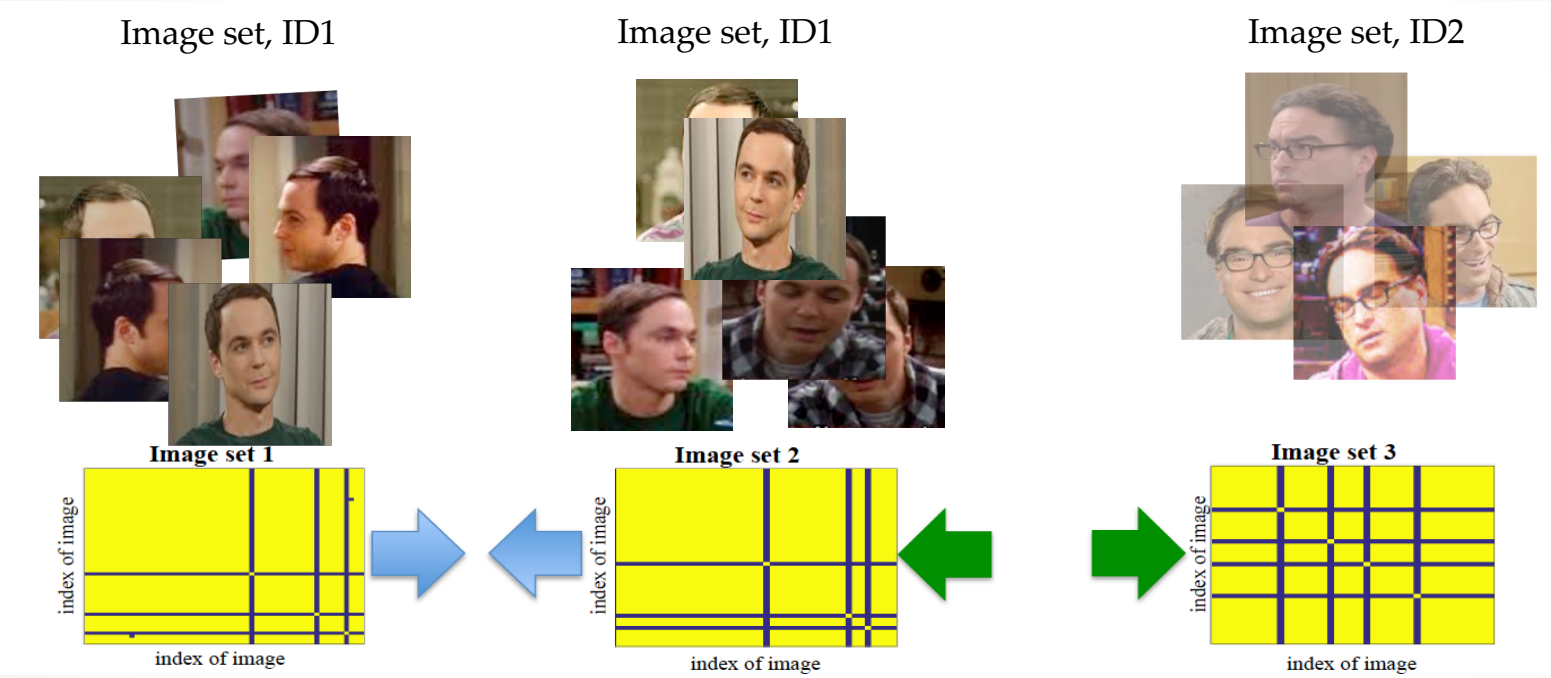
Structural Information and Similarity Measure



- Structural information: $K_g(\mathbf{x}_i, \mathbf{x}_j) = \frac{\sum_{p=1}^{n_i}\sum_{q=1}^{n_j} A_{ip}A_{jq}g(x_{ip}, x_{jq})}{\sum_{p=1}^{n_i} A_{ip}\sum_{q=1}^{n_j} A_{jq}}$

- Statistical information: $K_s(\mathbf{x}_i, \mathbf{x}_j) = \phi(C_i)^\top \phi(C_j)$
$= exp(-||log(C_i) - log(C_j)||_F^2 / 2\gamma_s^2)$

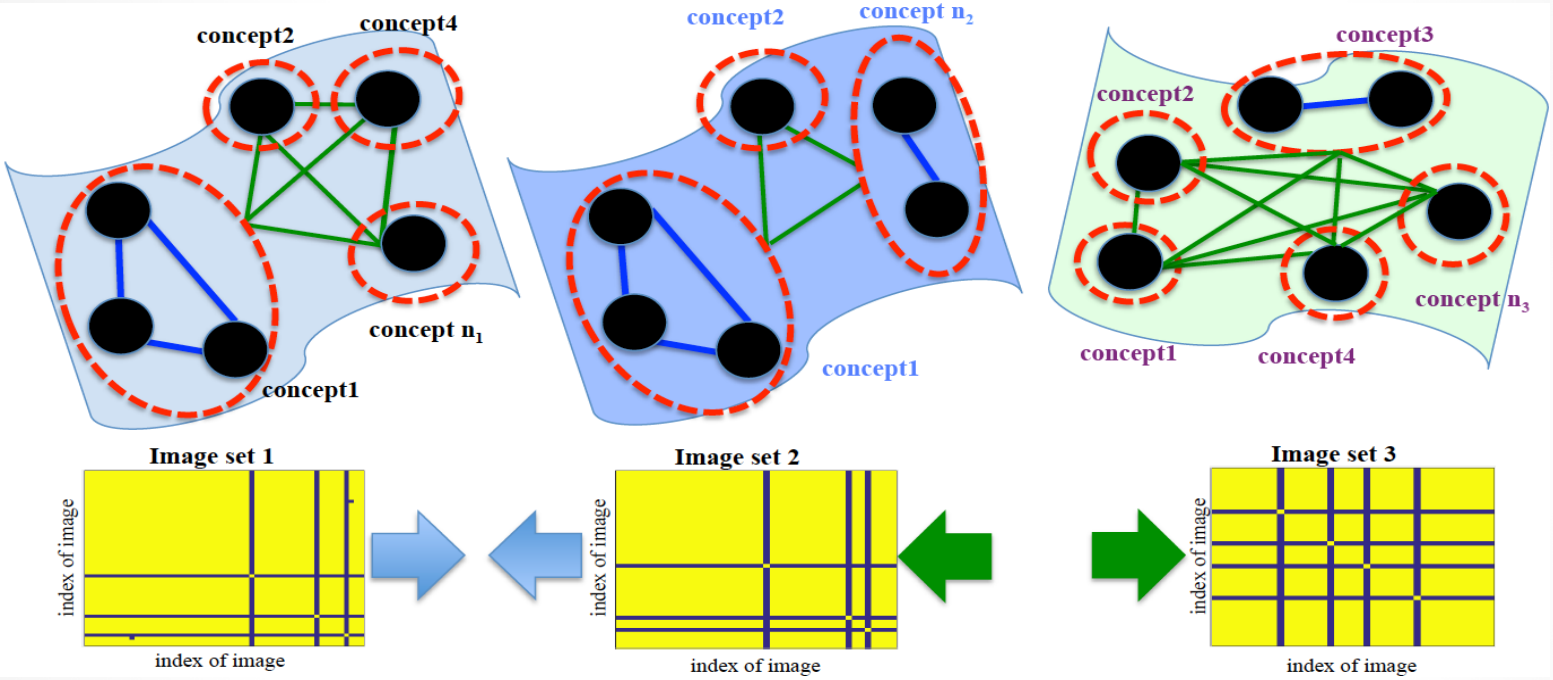
Structural Information and Similarity Measure



- Structural information: $K_g(\mathbf{x}_i, \mathbf{x}_j) = \frac{\sum_{p=1}^{n_i}\sum_{q=1}^{n_j} A_{ip}A_{jq}g(x_{ip}, x_{jq})}{\sum_{p=1}^{n_i} A_{ip}\sum_{q=1}^{n_j} A_{jq}}$

- Statistical information: $K_s(\mathbf{x}_i, \mathbf{x}_j) = \phi(C_i)^\top \phi(C_j)$
$= exp(-||log(C_i) - log(C_j)||_F^2 / 2\gamma_s^2)$

Structural Information and Similarity Measure



- Structural information: $K_g(\mathbf{x}_i, \mathbf{x}_j) = \frac{\sum_{p=1}^{n_i}\sum_{q=1}^{n_j} A_{ip}A_{jq}g(x_{ip}, x_{jq})}{\sum_{p=1}^{n_i} A_{ip}\sum_{q=1}^{n_j} A_{jq}}$

- Statistical information: $K_s(\mathbf{x}_i, \mathbf{x}_j) = \phi(C_i)^\top \phi(C_j)$
$= exp(-||log(C_i) - log(C_j)||_F^2 / 2\gamma_s^2)$

The Mapping Function

Weak Learners:

$$f(\mathbf{x}) = \text{sign}(K_m(\mathbf{x}_a, \mathbf{x}) - K_m(\mathbf{x}_b, \mathbf{x}) + \varepsilon)$$

positive — The m_{th} kernel — negative

A strong split (bit):

$$F(\mathbf{x}) = \text{sign}(\sum_{t=1}^{T} \lambda^t f^t(\mathbf{x}))$$

The best weak learner
at iteration t

Image Set Hashing via Multiple Information

- Problem Formulation

$$\min_{H_q, H_r} \underbrace{\alpha D_s}_{\text{Discriminative property}} + \underbrace{\beta D_c}_{\text{Refining hash code}} + \underbrace{\nu_1 \sum_{\substack{r' \in \{1:R\} \\ i \in \{1:N\}}} \Omega}_{\text{MKL}}$$

$$s.t. \quad H_q^{r'i} = sgn\left(\sum_{t=1}^{T} \lambda_t f_q^{tr'}(\mathbf{x}_i)\right), \forall i \in \{1:N\}, \forall r' \in \{1:R\}$$

$$H_r^{r'i} = sgn\left(\sum_{t=1}^{T} \lambda_t f_r^{tr'}(\mathbf{x}_i)\right), \forall i \in \{1:N\}, \forall r' \in \{1:R\}$$

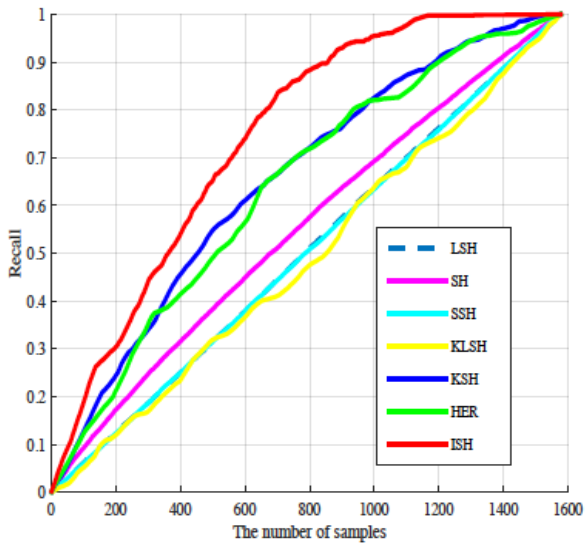
Experiment Settings

- Dataset:
  - CIFAR-10: 60, 000 images are randomly selected to 195 image sets for training process; 100 image sets for test query, the left images are around 50,000 for test database (around 1700 image sets)
  - Big Bang Theory video (image set) benchmark (BBT): 3,341 videos with 12 characters, 100 image sets for training, 100 image sets for test query, 3041 image sets for test database

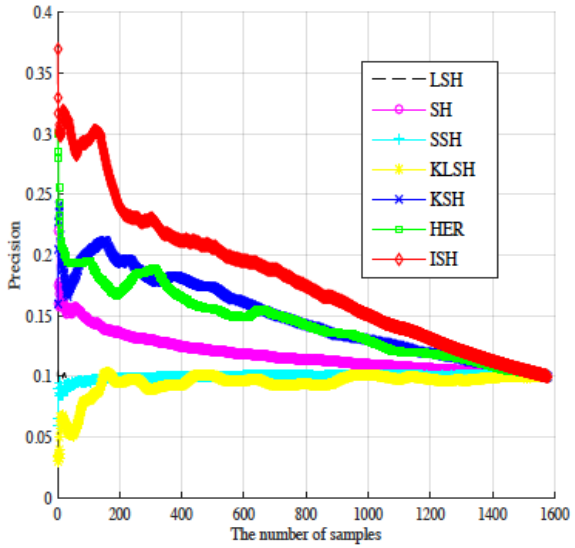
Retrieval Performance on CIFAR-10

	Compared Methods						Our Method
bits	LSH [13]	SH [49]	SSH [41]	KLSH [16]	KSH [21]	HER [18]	ISH
8bits	0.1063	0.1279	0.1165	0.1067	0.2363	0.2284	**0.2822**
12bits	0.1073	0.1330	0.1416	0.1210	0.2486	0.2672	**0.3069**
24bits	0.1086	0.1317	0.1512	0.1420	0.2680	0.2772	**0.3159**
32bits	0.1194	0.1322	0.1574	0.1501	0.2818	0.2937	**0.3292**
48bits	0.1105	0.1352	0.1629	0.1622	0.3003	0.3154	**0.3334**

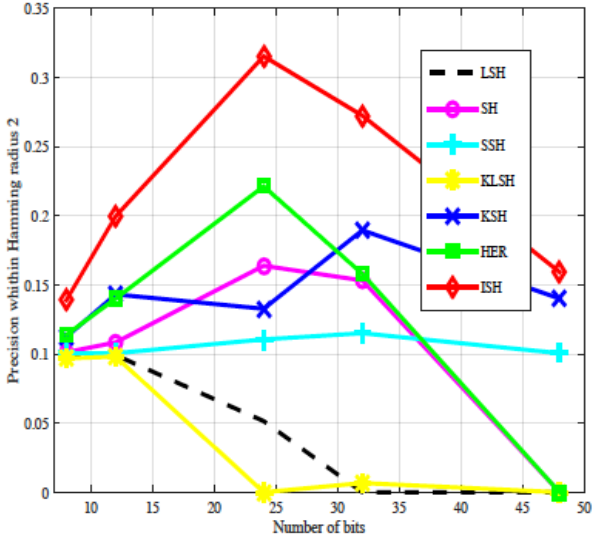
Performance on CIFAR-10. KLSH, KSH, HER, ISH use image sets as input



Recall curve @24 bits



Precision curves @24 bits

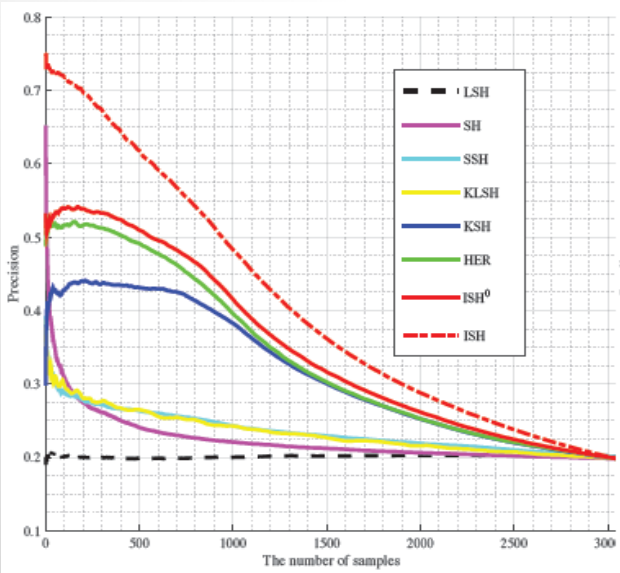


Hamming radius 2

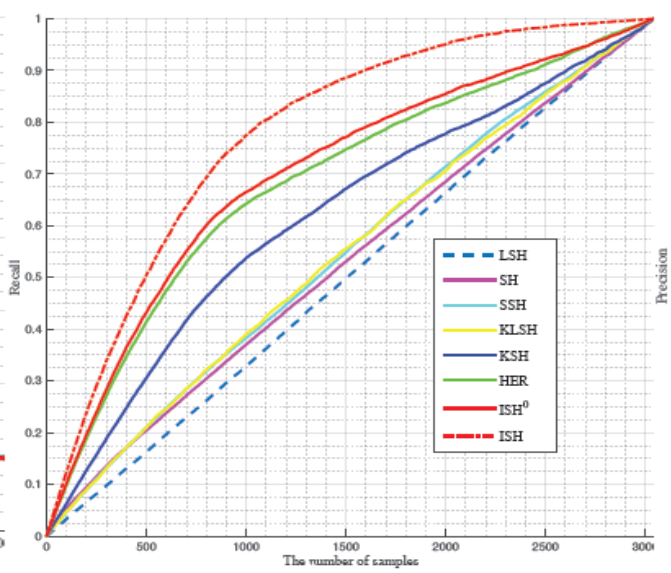
Performance on BBT TV-series

	TV drama: the Big Bang Theory				
Method	8 bits	16 bits	32 bits	64 bits	128 bits
LSH [13]	0.2109	0.2086	0.2092	0.1963	0.1994
ITQ [10]	0.2935	0.3025	0.2989	0.3029	0.3060
SH [49]	0.2377	0.2652	0.2665	0.2623	0.2673
DBC [30]	0.4489	0.4495	0.4235	0.4005	0.3867
SSH [42]	0.2716	0.2855	0.2662	0.2584	0.3003
MM-NN [25]	0.3752	0.3955	0.4664	0.5124	0.4922
KLSH [16]	0.2450	0.2498	0.2381	0.2256	0.2325
KSH (point)	0.4090	0.4366	0.4454	0.4567	0.4604
KSH [21] (set)	0.4590	0.4619	0.4534	0.4685	0.4631
HER [18]	0.4606	0.5049	0.5227	0.5490	0.5539
ISH^0	0.4833	0.5279	0.5359	0.5501	0.5712
ISH	**0.5018**	**0.5592**	**0.5864**	**0.6007**	**0.6280**

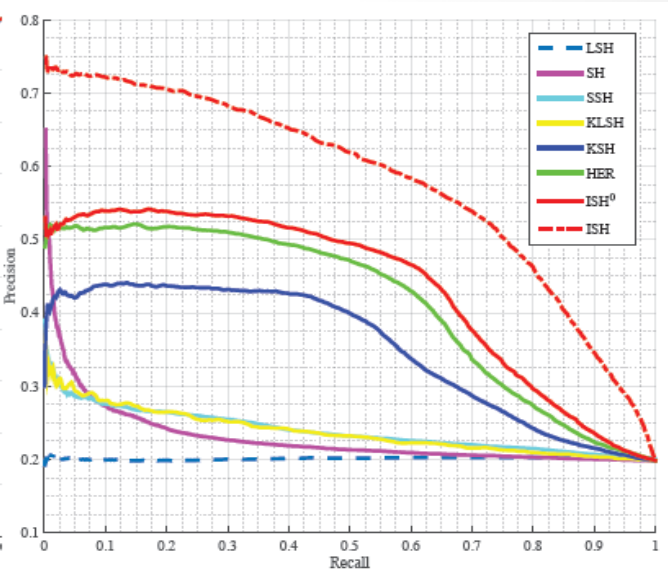
Performance on BBT. KLSH, KSH, HER, ISH^0, ISH use image sets as input



Mean recall curves for Hamming ranking using 128 bits



Mean precision curves for Hamming ranking using 128 bits



Mean precision-recall curves for Hamming ranking using 128 bits

Summary

- S2S provides the promising applications/results comparing to the traditional hashing methods (P2P)

- S2S hashing improves the retrieval performance by utilizing multiple (structural and statistical) information

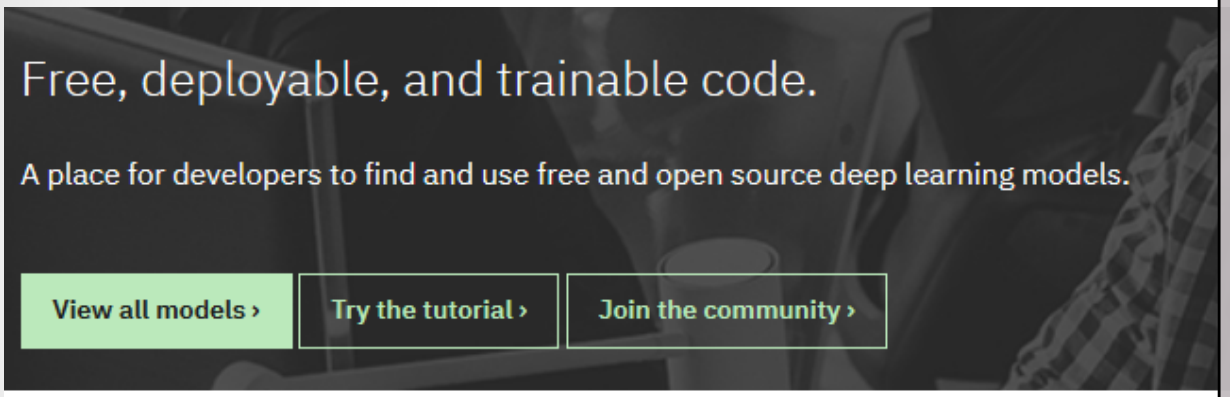
- State-of-the-art results in retrieval datasets

- No complex deep learning frameworks

MAX for Developers

Find model asset → Deploy pre-trained model asset → Use model asset

- Audio classification
- Image classification
- Text classification
- Object detection
- Facial recognition
- Image-to-image translation
- Image-to-text translation
- …



Deployable | Facial Recognition

Facial Emotion Classifier

Detect faces in an image and predict the emotional state of each person

View model »

Artificial intelligence | Deep learning +

Deployable | Object Detection In Images

Image Segmenter

Identify objects in an image, additionally assigning each pixel of the image to a particular object.

View model »

Deployable | Object Detection In Images

Object Detector

Localize and identify multiple objects in a single image.

View model »

Artificial intelligence | Deep learning +

ibm.biz/model-exchange

IBM Data Asset eXchange (DAX)

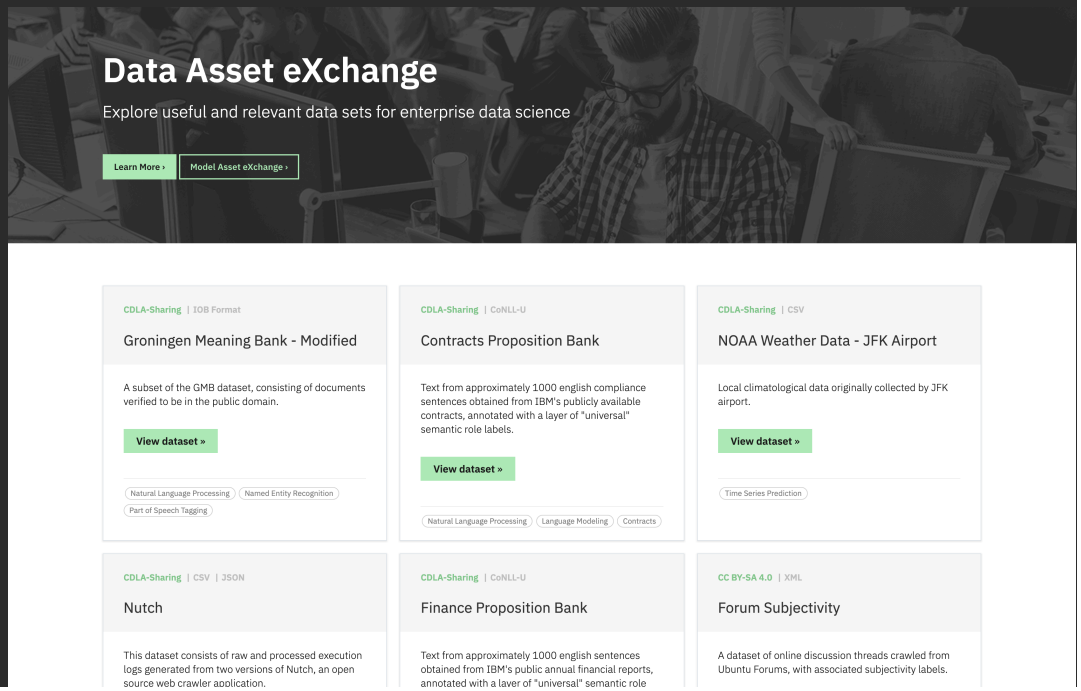
- Curated free and open datasets under open data licenses
- Standardized dataset formats and metadata
- Ready for use in enterprise AI applications
- Complement to the **M**odel **A**sset e**X**change (**MAX**)

Data **A**sset e**X**change

ibm.biz/data-asset-exchange

Model **A**sset e**X**change

ibm.biz/model-exchange




THANK
YOU
GRACIAS
ARIGATO
SHUKURIA
DANKSCHEEN
JUSPAXAR
KOMAPSUMNIDA
GOZAIMASHITA
EFCHARISTO
TASHAKKUR ATU
YAQHANYELAY
SUKSAMA
EKHMET
MEHRBANI
GRAZIE
MAAKE
BOLZÏN
MERCI
BÏYAN
SHUKRIA
TINGKI
SPASSIBO
SNACHALHUYA
NUHUN
CHALTU
MERASTAWHY
GAEJTTHO
AGUYJE
FAKAAUE
TAVTAPUCH
MEDAWAGSE
SANCO
LAH
ATTO
WABEEJA
MAITEKA
PALDIES
SPASIBO
DENKAUJA
NENACHALHYA
UNALCHEESH
GUI
HATUR
EKOJU
SIKOMO
MAKETAI
MIMMONCHAR

CODAIT, IBM