Compact Hashing for Mixed Image-Keyword Query over Multi-Label Images

Xianglong Liu[1], Yadong Mu[2], Bo Lang[1] and Shih-Fu Chang[2]

[1]*Beihang University, Beijing, China*

[2]*Columbia University, New York, USA*

ICMR 2012!

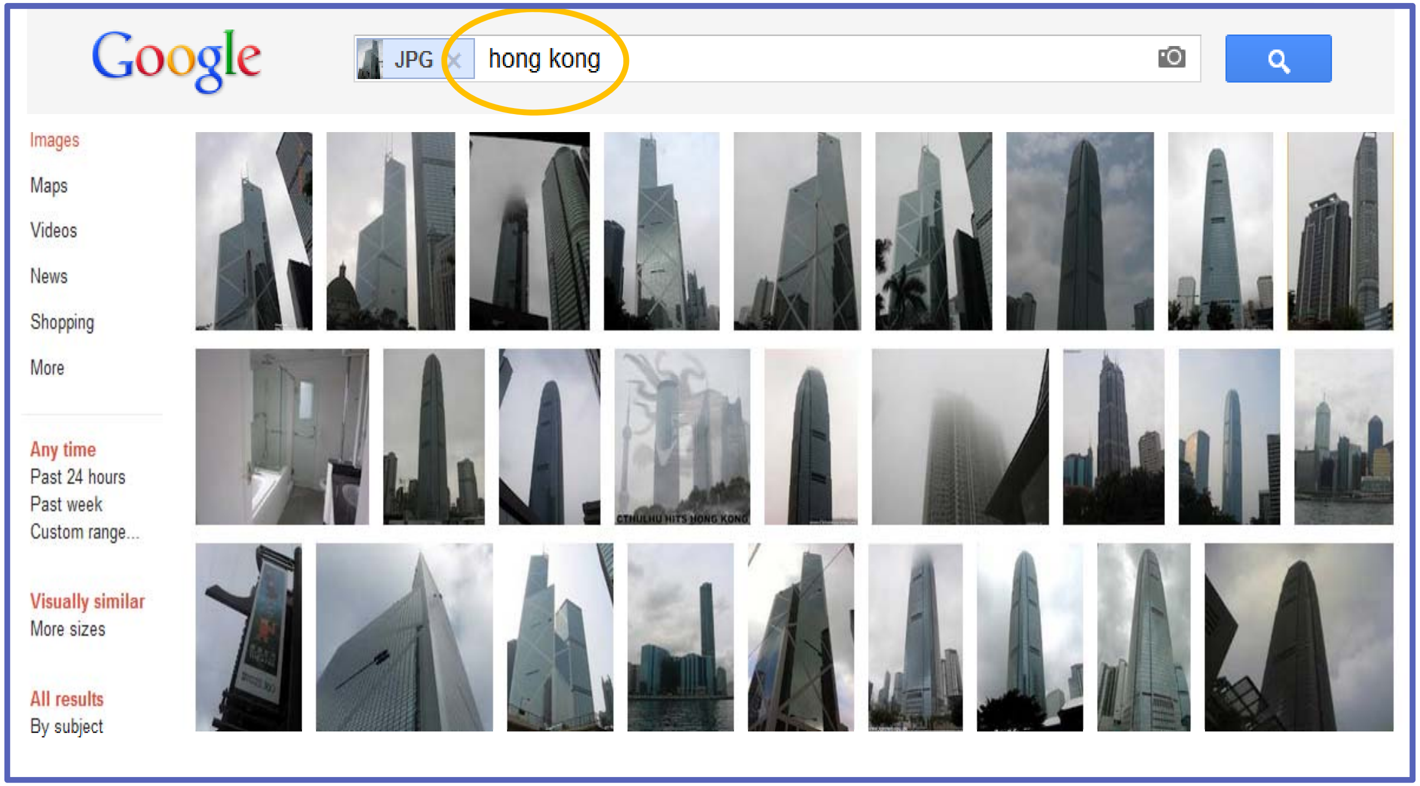
Outline

- **Introduction**
  - Motivation
  - Our Solution
- **Boosted Shared Hashing**
  - Formulation
  - Optimization
  - The Retrieval Stage
- **Experiments**
- **Conclusion**

“Image + Keyword” based Visual Search (1/4)

- **Yet another image search paradigm**
  - Query image provides content descriptor
  - Textual keywords greatly narrow the semantic gap!

Query Image



"Image + Keyword" based Visual Search (2/4)

- **Challenge-1: Noisy or unknown label information**
  - **Database Images**: labels are unknown and expensive to annotate
  - **Training Images**: a small set, and manually annotated
  - **Query**: Image + Keyword (or label)

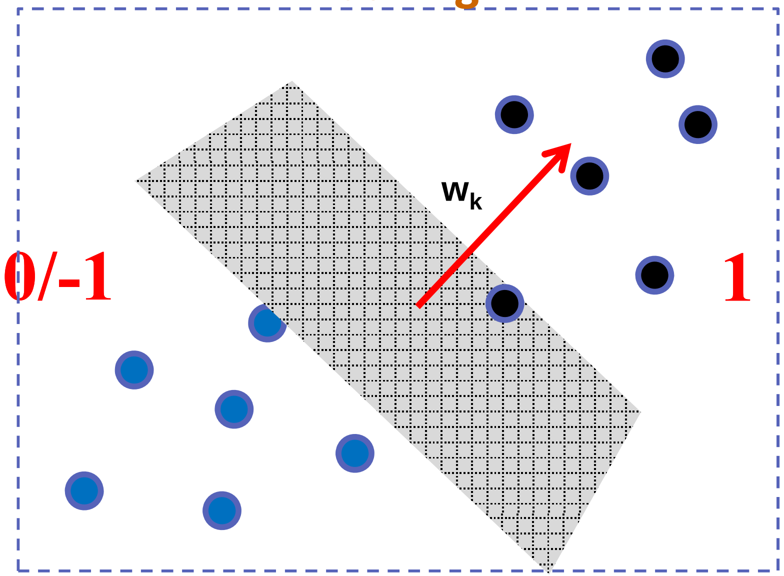
Problem Settings

	Visual Features	Labels
Database	√	
Training Set	√	√
Query	√	√

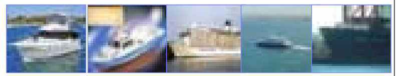
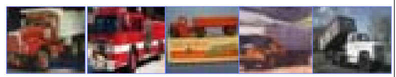
"Image + Keyword" based Visual Search (3/4)

- **Challenge-2: Scalability to Web-Scale Data**
  - Linear scan is infeasible
  - Approximate nearest neighbor (ANN)
    - balance the performance and computational complexity
    - Tree-based methods (KD tree, metric tree, ball tree, etc.)
    - **Hashing-based methods:** efficient index and search

Hashing

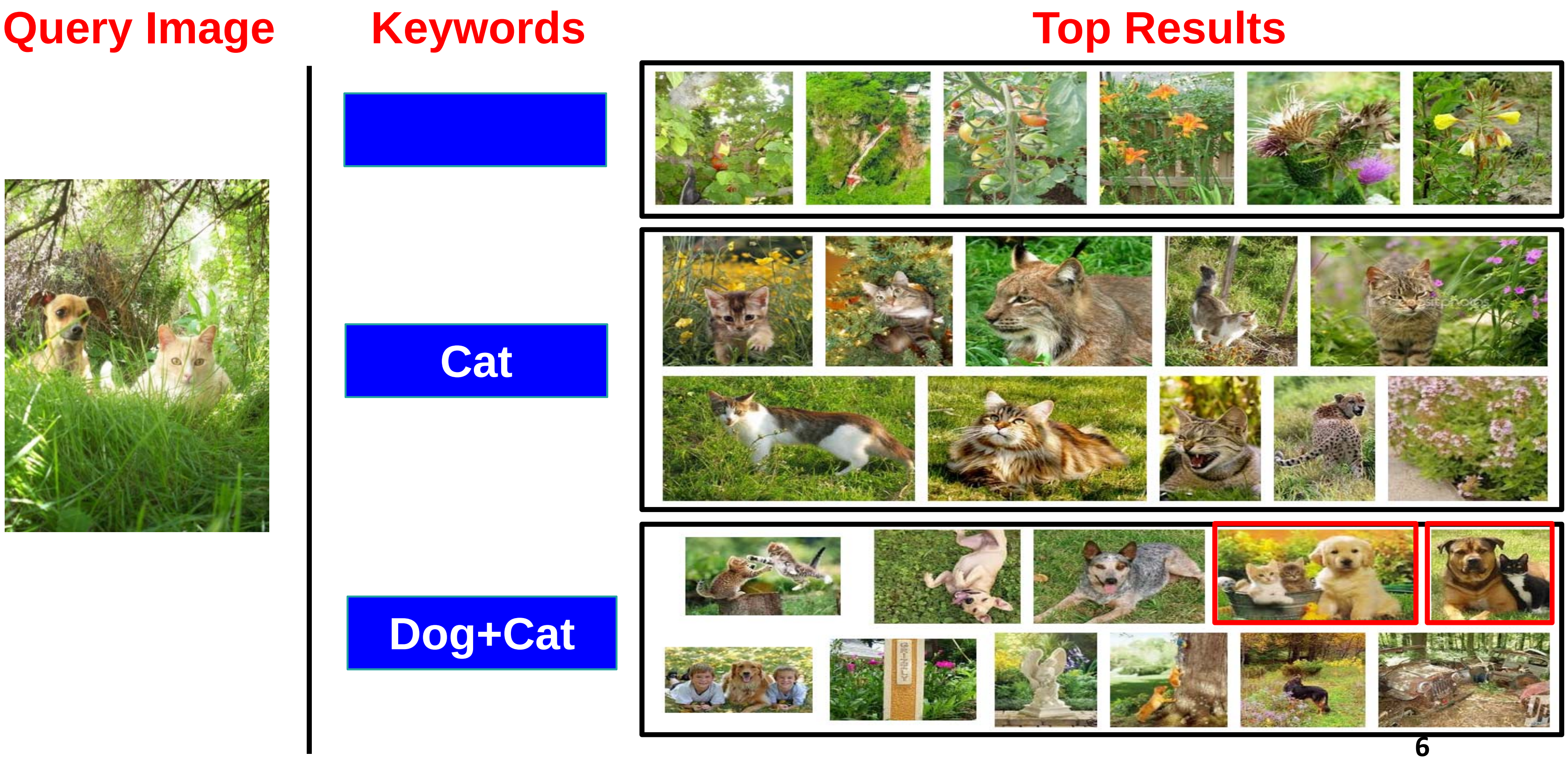


Hashing Tables

Bucket	Indexed Image
0010...	
0110...	
⋮	⋮
1111...	

“Image + Keyword” based Visual Search (4/4)

- **Challenge-3: Diverse Semantics**
  - User intention is ambiguous / diverse
  - **Query-adaptive hashing over multi-label data**



Related Works

- **Supervised Hashing**

Naïve solutions for Hashing with Multi-Label Data

Universal hash function:
Semi-Supervised Hash [Wang10a]
Sequential Projection Hash [Wang10b]

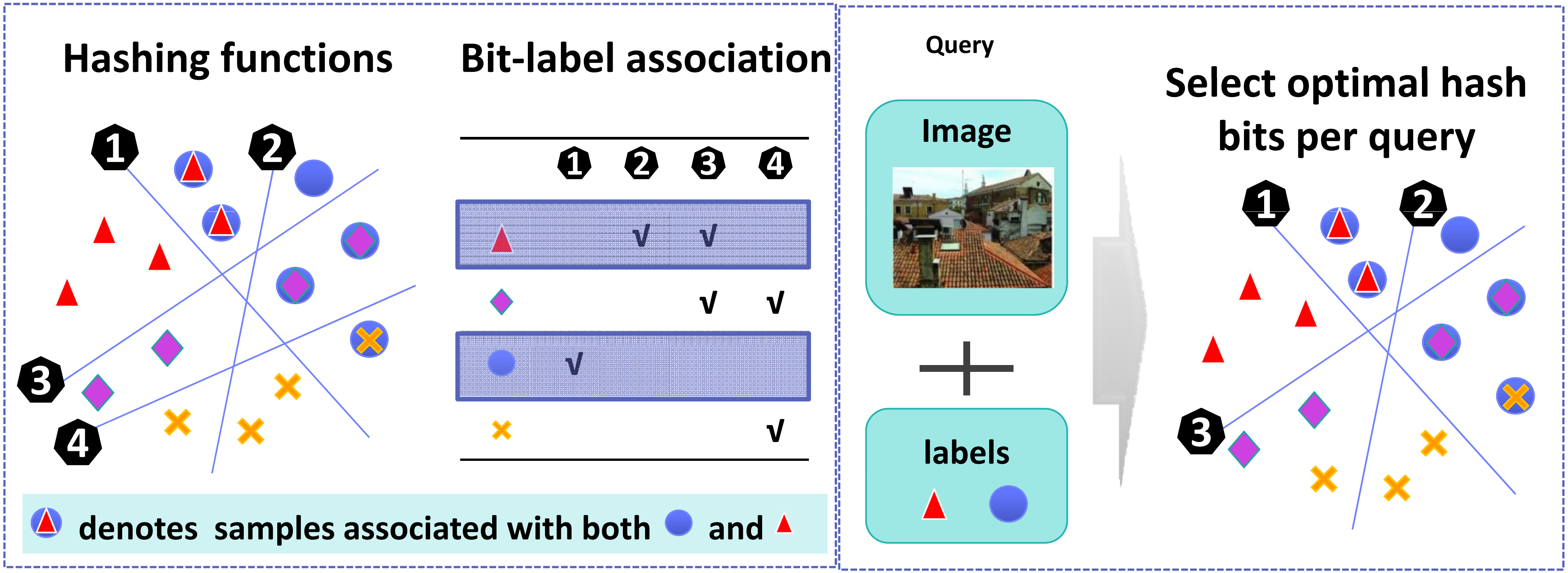
- **Universal ones: hard to learn**
- **Complicate bit selection**

Per-Label hash function:
Bit-Selection Hash [Mu 11]

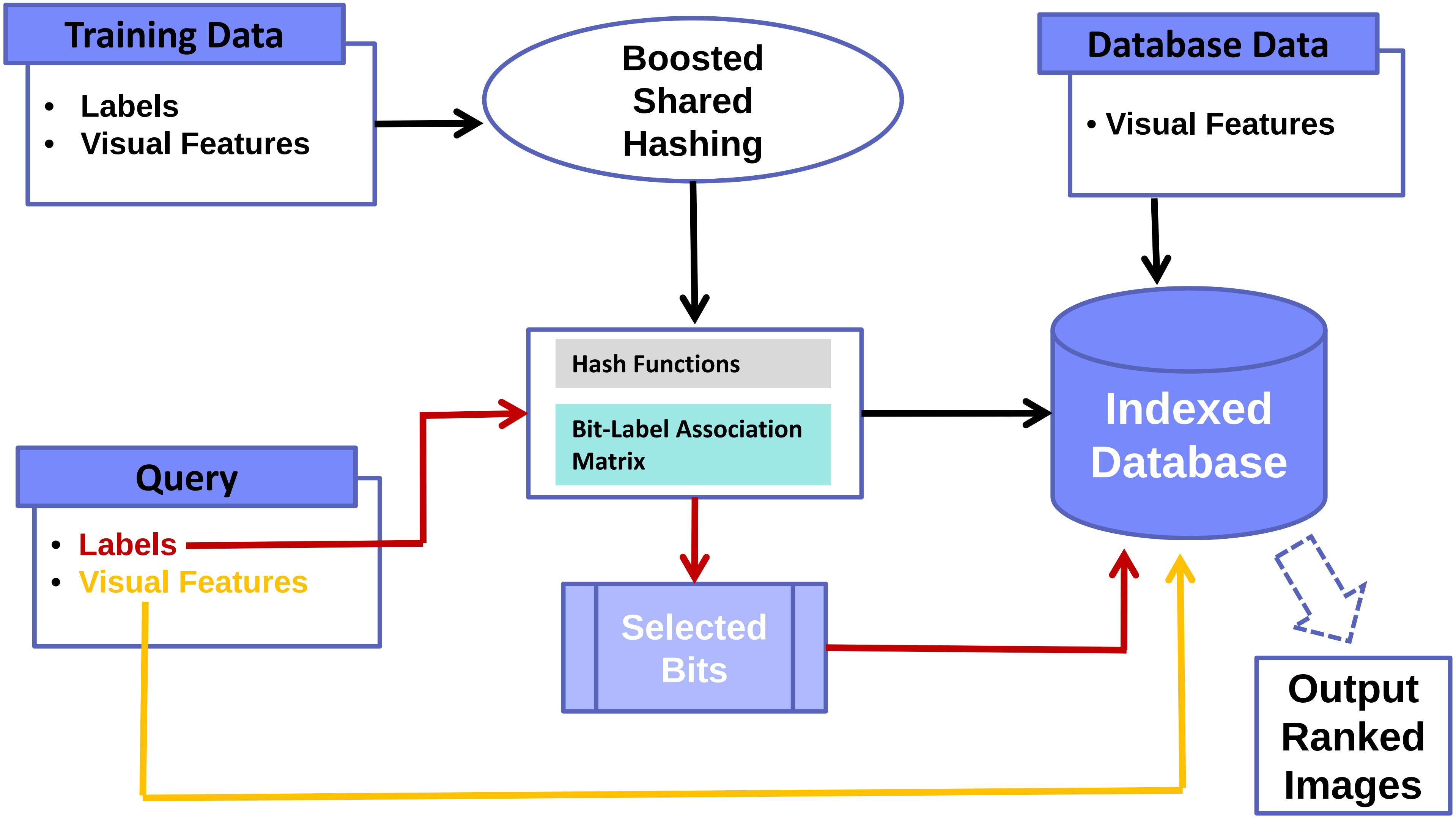
- **Redundancy**

Overview of Our Solution (1/2)

- **Key idea:** to encourage sparse association between hashing functions and labels by exploiting shared subspaces among the labels



Overview of Our Solution (2/2)



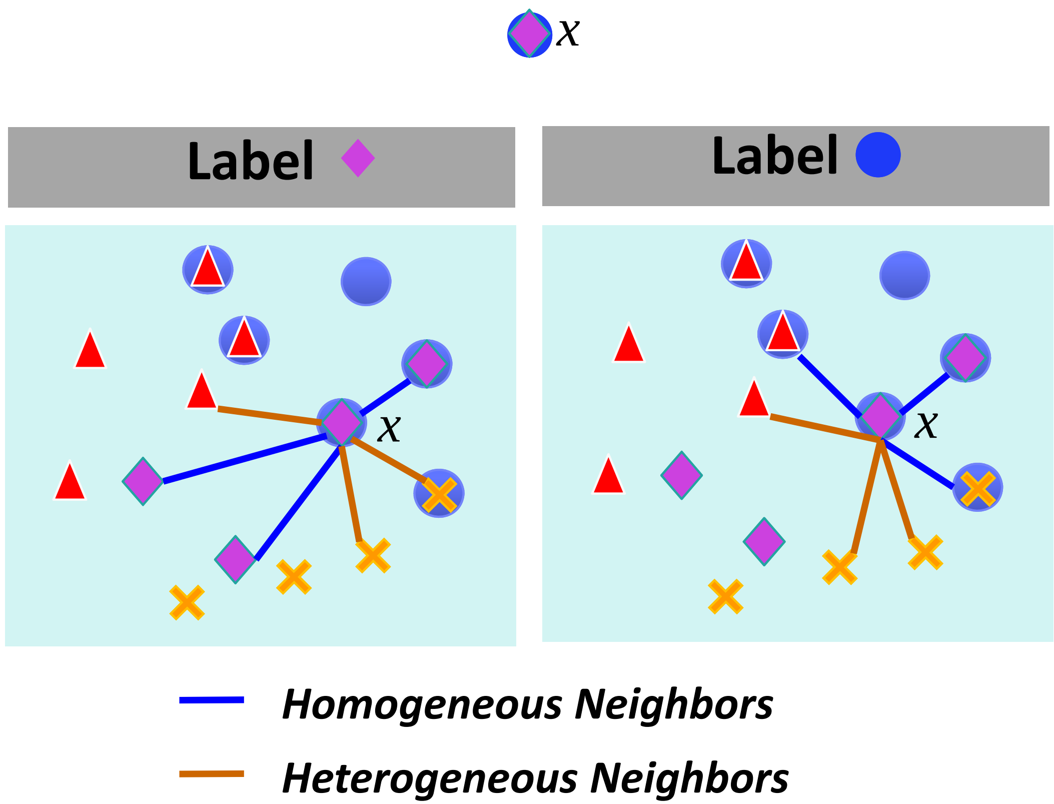
Data Structure

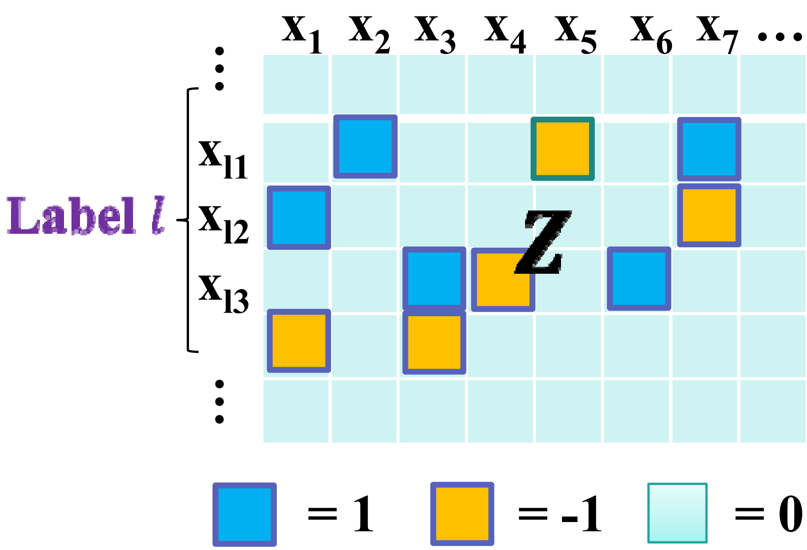
Multi-label data $\langle x_i, l_i \rangle \in \mathbb{R}^D \times \{0,1\}^L,\ i = 1 \dots N$

$l_i(k) = 1$: x_i is associated with the k-th label

Neighbor graph homogeneous neighbors: with the same label

heterogeneous neighbors: with different labels





Objective Function

- **Encourage prediction f of hashing function h on neighbor pairs to have the same sign with neighbor matrix Z :**

$$(x_i, x_j, k)\begin{cases} \textbf{Homogeneous}: \; z_{ij}=1, \; \text{expect } h_b(x_i)\, h_b(x_j)=1 \\ \textbf{Heterogeneous}: \; z_{ij}=-1, \; \text{expect } \underline{h_b(x_i)}\, \underline{h_b(x_j)}=-1 \end{cases} \Bigg\} \; z_{ij}\, f(x_i,x_j,k)=1$$

Ex

Hashing prediction $$f^{(b)}(x_i, x_j, k) = \delta\left[k \in \mathcal{S}(b)\right] \cdot h_b(x_i) \cdot h_b(x_j)$$

$$F(x_i, x_j, k) = \sum_{b=1}^{B} f^{(b)}(x_i, x_j, k)$$

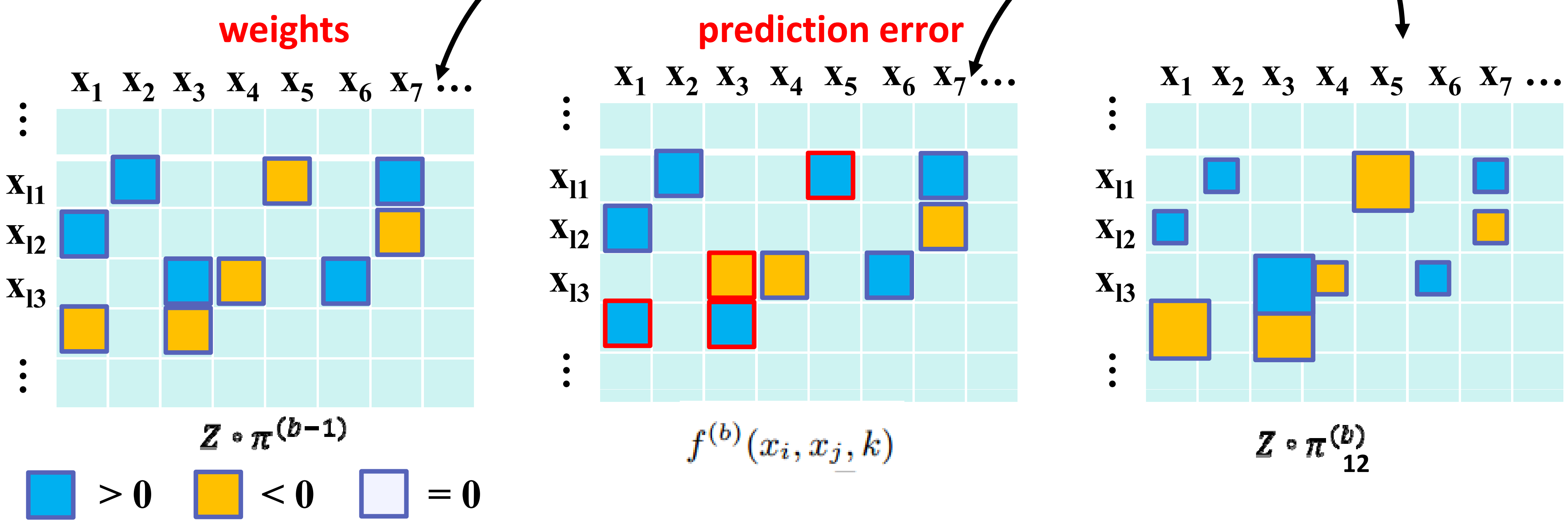
Active Label Set $S(b)$: labels associated with the b-th bit

Sequential Learning: Boosting

- **Boosting style:** to learn a hashing function that tries to correct the previous mistakes by updating weights on neighbor pairs

$$\pi^{(b)}(i,j,k) = \pi^{(b-1)}(i,j,k) \cdot e^{-z_{ij} f^{(b)}(x_i,x_j,k)}$$

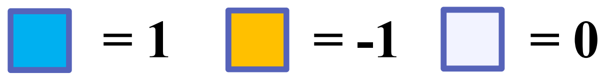
$$\pi^{(b-1)}(x_i,x_j,k)$$



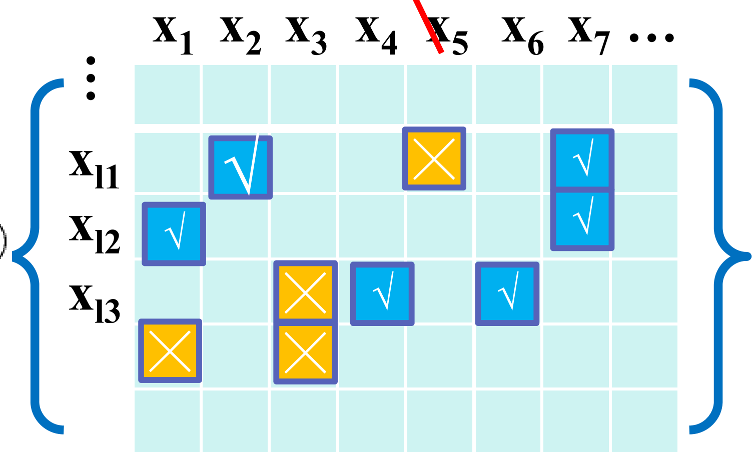
Optimize: hashing function

- **Taylor expansion**

$$e^{-z_{ij}F^{(b)}(x_i,x_j,k)} \approx -\ \pi^{(b-1)}(i,j,k)\cdot \boxed{z_{ij}f^{(b)}(x_i,x_j,k)}$$



$$\mathcal{J} \approx -\sum_{(i,j,k)\in\mathcal{I},k\in\mathcal{S}(b)} \pi^{(b-1)} \odot$$



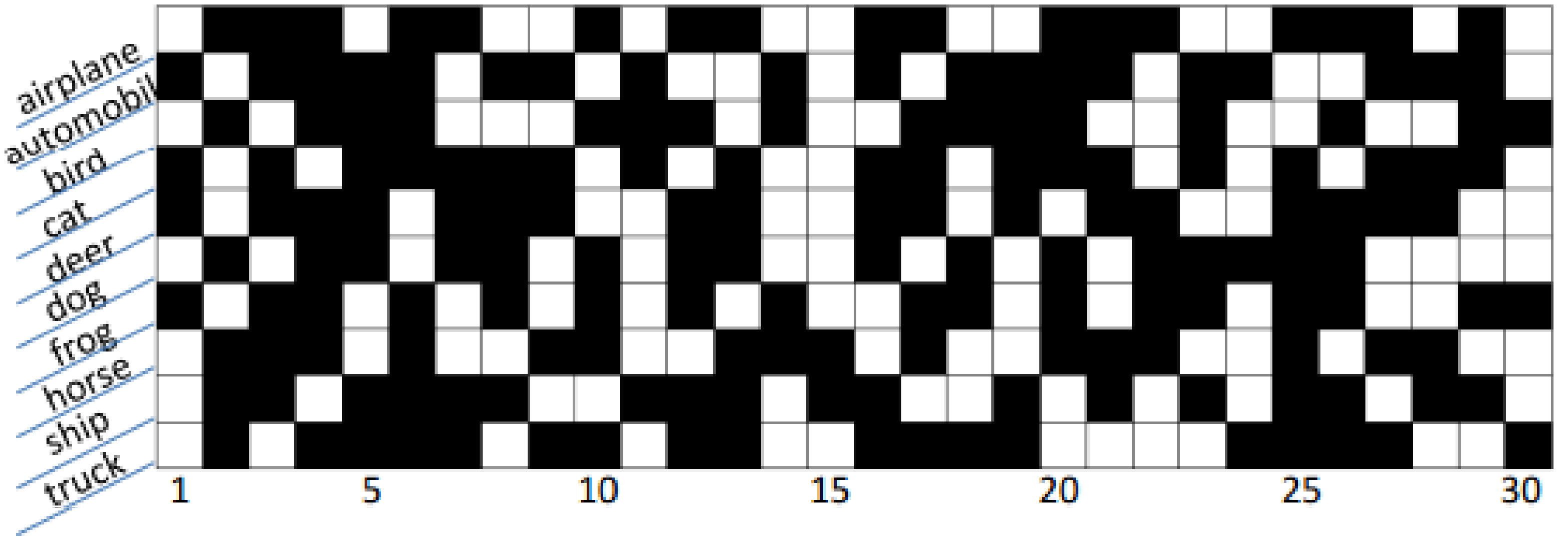
- **Relaxation of sign function**

$$\mathcal{J} \approx \frac{1}{2}w^T X R X^T w$$

efficiently solved by eigen-decomposition

Optimize: active label set

- **Find a label subset $S(b)$ that gives minimum loss**
  - Intuitive way: exhaustively compare all possible 2^L subsets
  - A greedy selection $O(L^2)$:
    - Initialize $S(b)$: the label giving minimum loss;
    - Expand $S(b)$: add label giving the most loss decrease among all rest labels
    - Terminated when the gain is incremental (<5%)



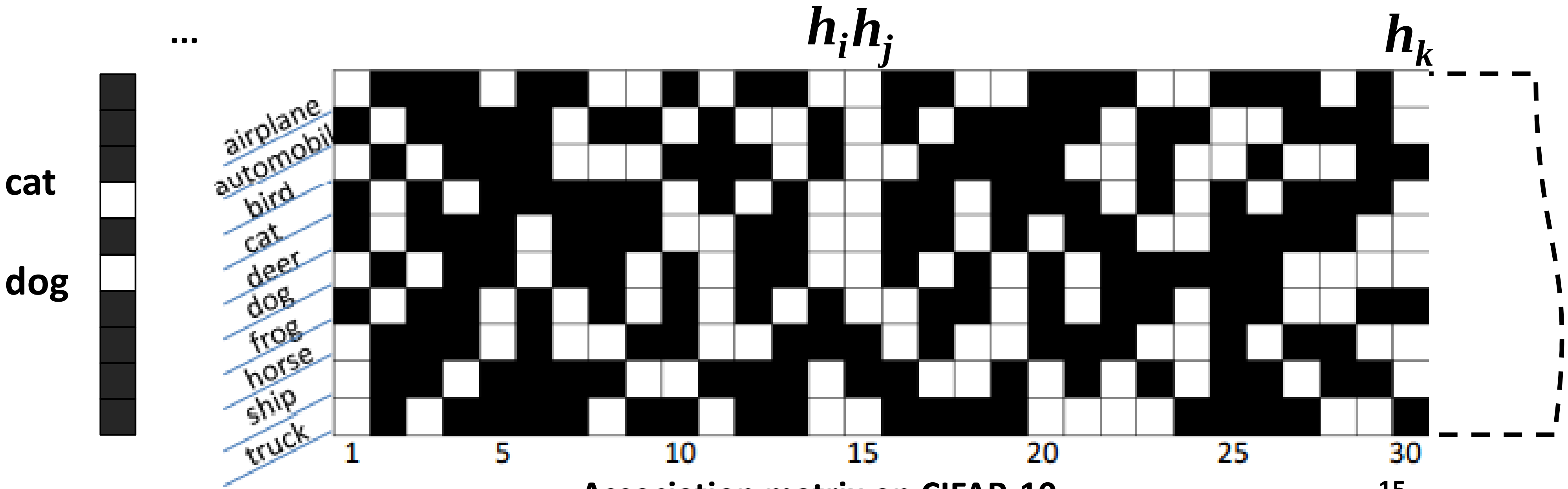
Association matrix on CIFAR-10

Query-Adaptive Search

- **Bit selection based on matching-score**
  - Select bits that are most confident across all query labels l_q
  - Measured by Jaccard index: computed between a (any column of matrix A) and query labels l_q:

$$s_J(a, l_q) = \frac{|a \cap l_q|}{|a \cup l_q|}$$

$A \in \{0,1\}^{L \times B}$ is the learned bit-label association matrix



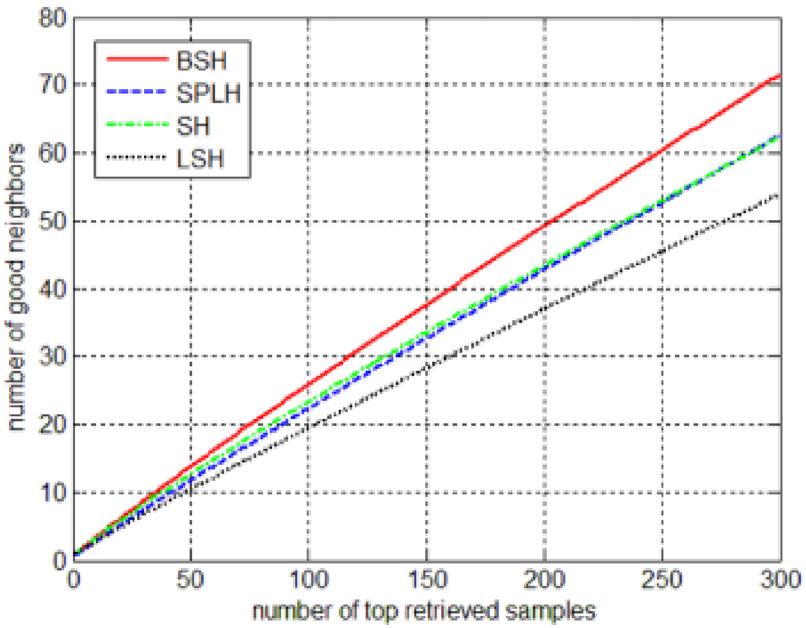
Association matrix on CIFAR-10

Experiments

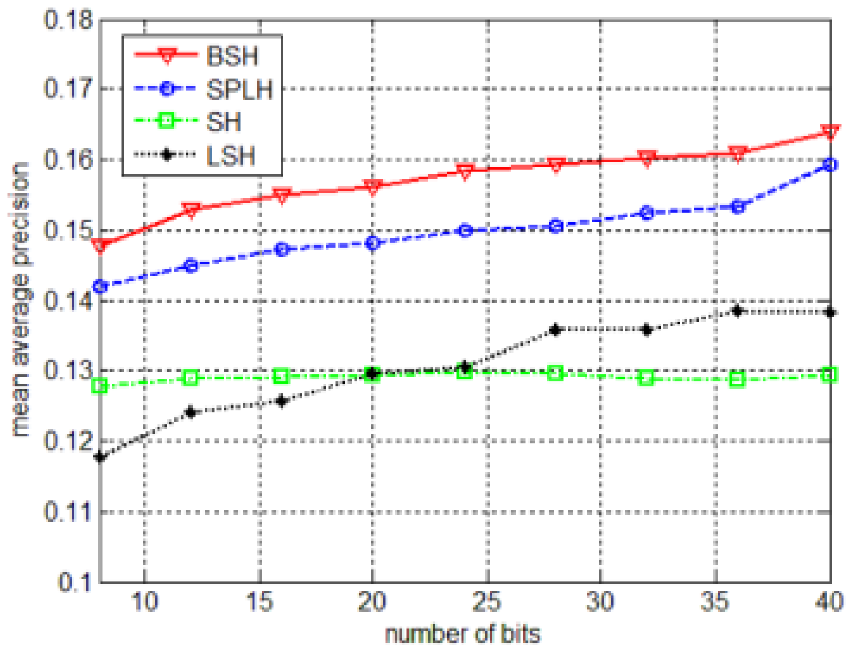
- **Datasets**
  - Multi-category: **CIFAR-10** (60K)
  - Multi-label: **NUS-WIDE** (270K)
- **Baselines:**
  - SPLH [Wang 10a], SH [Weiss 08], and LSH [Indyk 98]
- **Setting:**
  - 15 homogeneous and 30 heterogeneous neighbors without tuning.
  - same # bits per query for all methods
  - Average performance of 10 independent runs

CIFAR-10

- 32x32 color images, 10 semantic categories (e.g., airplane, frog, truck etc.)
- 3,000 images as training data
- 1,000 random samples as the queries
- 384-D GIST features

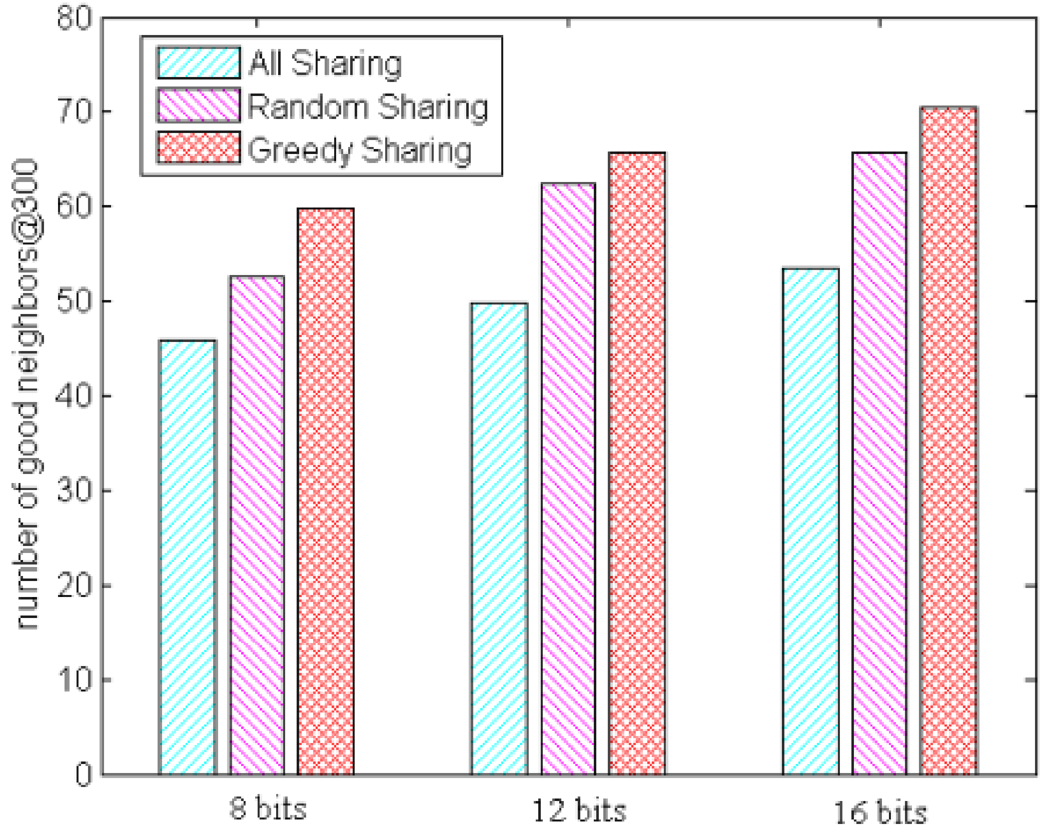


Good neighbors under 24 hash bits



Mean-Average-Precision (0-1)

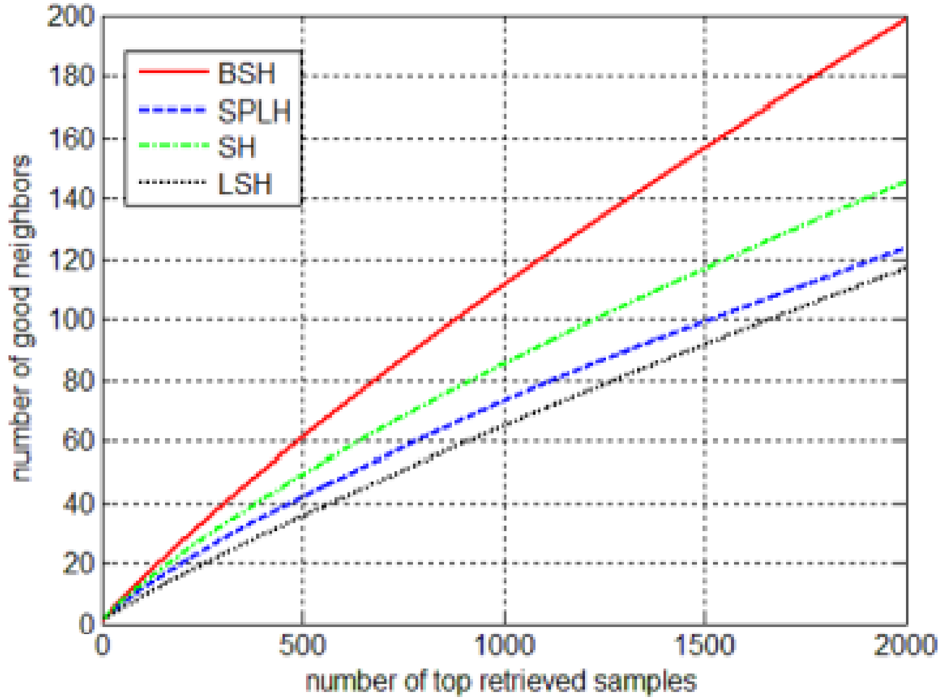
Impact of Sharing



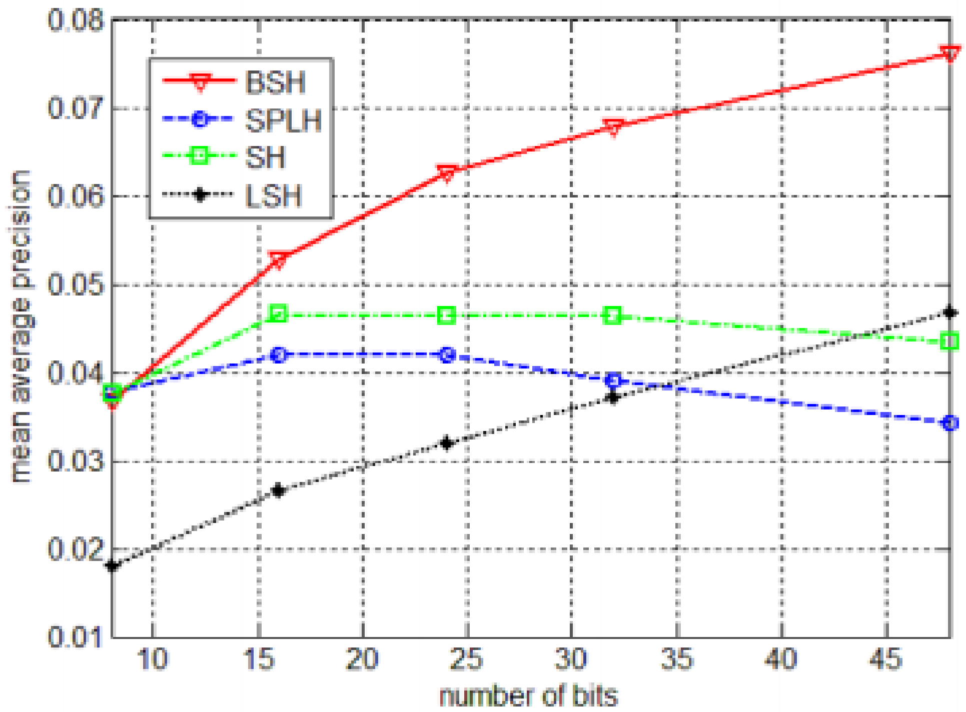
- Greedy sharing: $S(b)$
- All sharing: each hashing function is universal for all labels
- Random sharing: uniformly sample specific number (the averaged size of $S(b)$) of labels to be active

NUS-WIDE

- Select 25 most-frequent tags (“sky”, “clouds”, “person”, etc.) from 81 tags
- 5,000 images as training set
- 1,000 images with two randomly selected labels as the query set
- Groundtruth for each query: images with both (1) the same labels; and (2) the closest distances of their visual features
- Concatenate 500-D Bow (SIFT) and 225-D block-wise color moment

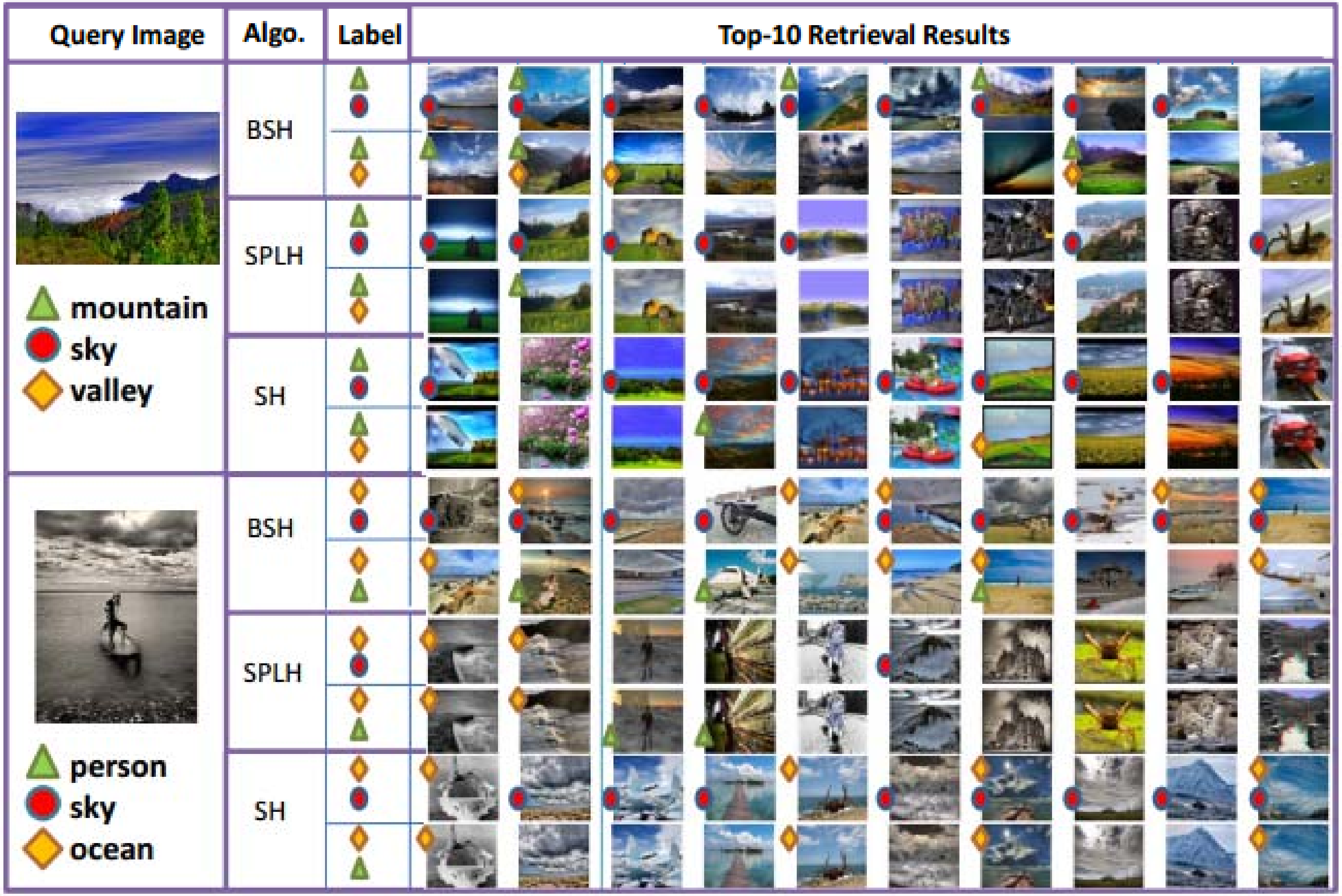


Good neighbors: two-label query



MAP: two-label query

Examples

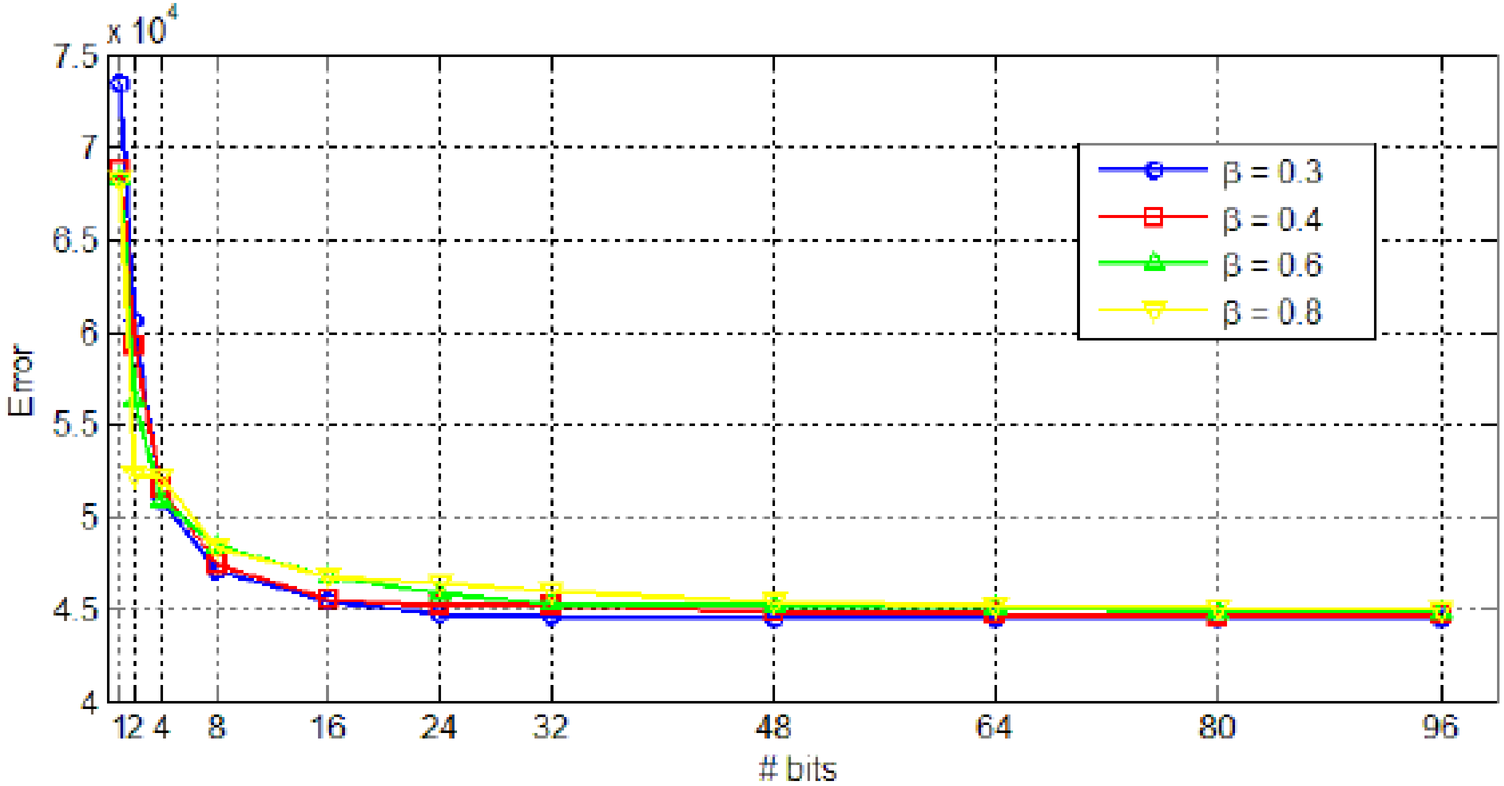

Query Image
Algo.
Label
Top-10 Retrieval Results
BSH
SPLH
SH
BSH
SPLH
SH
mountain
sky
valley
person
sky
ocean

Summary and Conclusion

- **Summary and contributions**
  - the first compact hashing technique for mixed image-keyword search over multi-label images
  - an efficient Boosting-style algorithm to sequentially learn the hashing functions and active label set for multi-label images
  - A simple hashing function selection adaptive to query labels
- **Future work**
  - Theoretical analysis of performance guarantee
  - Extension to non-linear and reweighted hashing

Thank you!

Convergency



Sharing Rate