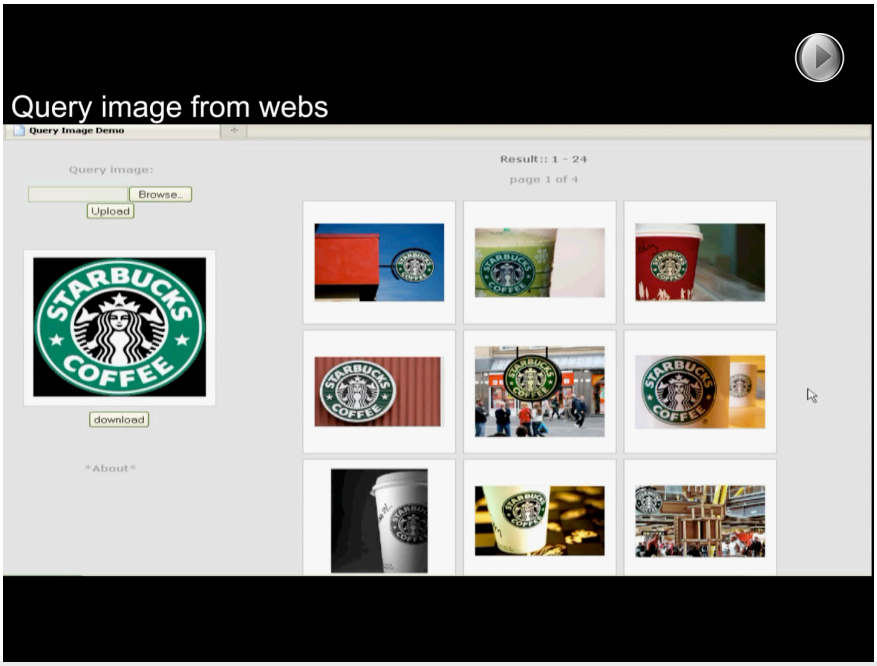
Exploiting Contextual Information for Visual Search

Yin-Hsi Kuo
Advisor: Prof. Winston H. Hsu

Graduate Institute of Networking and Multimedia,
National Taiwan University, Taiwan

National Taiwan University

Our Demo Video—Large-Scale Search Engine



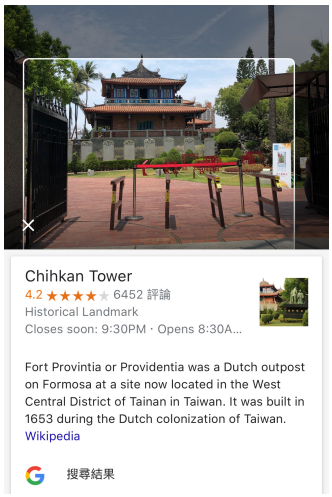
Yin-Hsi Kuo, Yi-Lun Wu, Kuan-Ting Chen, Yi-Hsuan Yang, Tzu-Hsuan Chiu, and Winston H. Hsu, "A Technical Demonstration of Large-Scale Image Object Retrieval by Efficient Query Evaluation and Effective Auxiliary Visual Feature Discovery," *ACM Multimedia*, pp. 1559-1562, 2010.

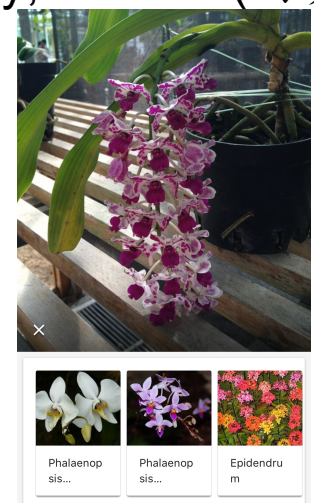
Applications for Visual (Product) Search

- Information inquiry by mobile phones
  - E.g., products, flowers, books
- Advertising by image matching
- Automatic tagging

Existing Applications for Visual Search

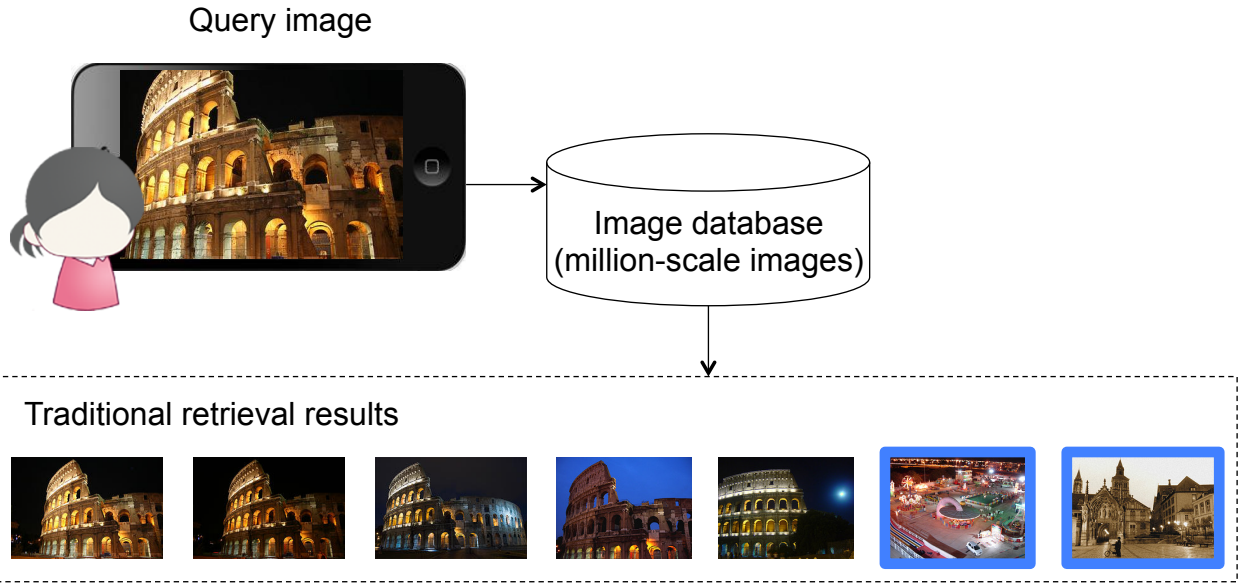
- Google Lens (Google Photos), Pinterest Lens, Amazon Firefly, Pailitao (淘寶, 拍立淘), etc.



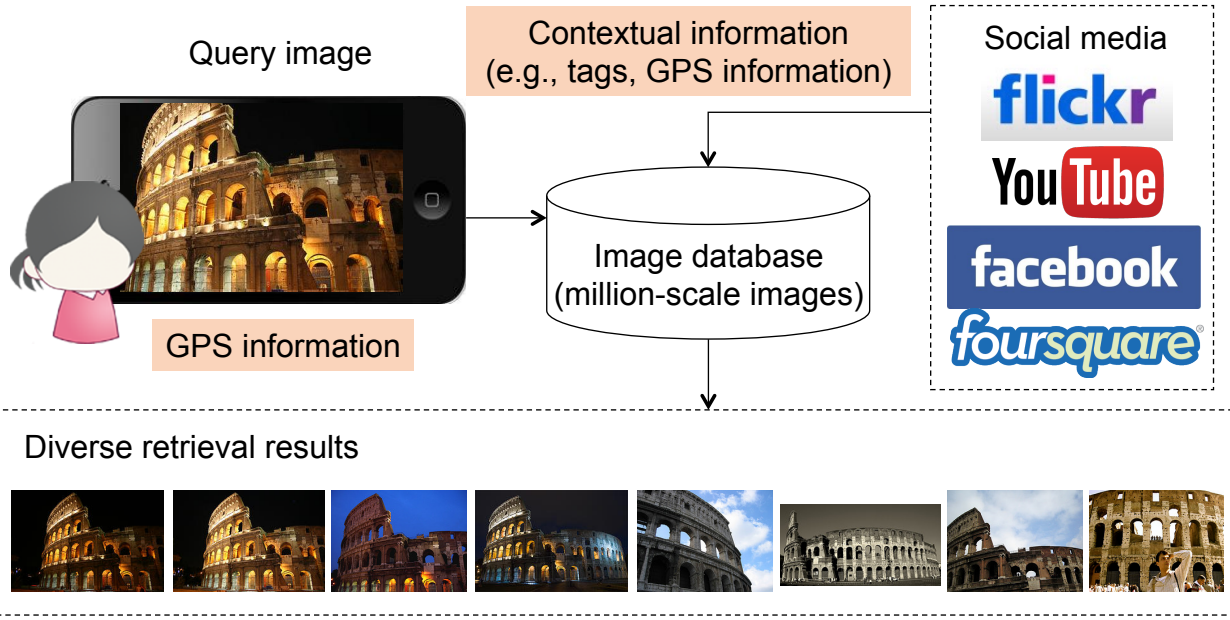


Amazon Press Conference: Jeff Bezos introduces Fire phone, the first smartphone designed by Amazon [https://youtu.be/w95kwXy_MOY?t=26m30s]
A. Zhai et al. Visual discovery at Pinterest. *WWW*, 2017.
Google I/O Keynote (Google I/O '17) [https://youtu.be/Y2VF8tmLFHw?t=11m35s]
Pailitao for Taobao [http://www.pailitao.com/] [demo video: https://youtu.be/bGNC6t9MTe0]

Traditional Approaches—Visually Similar Results



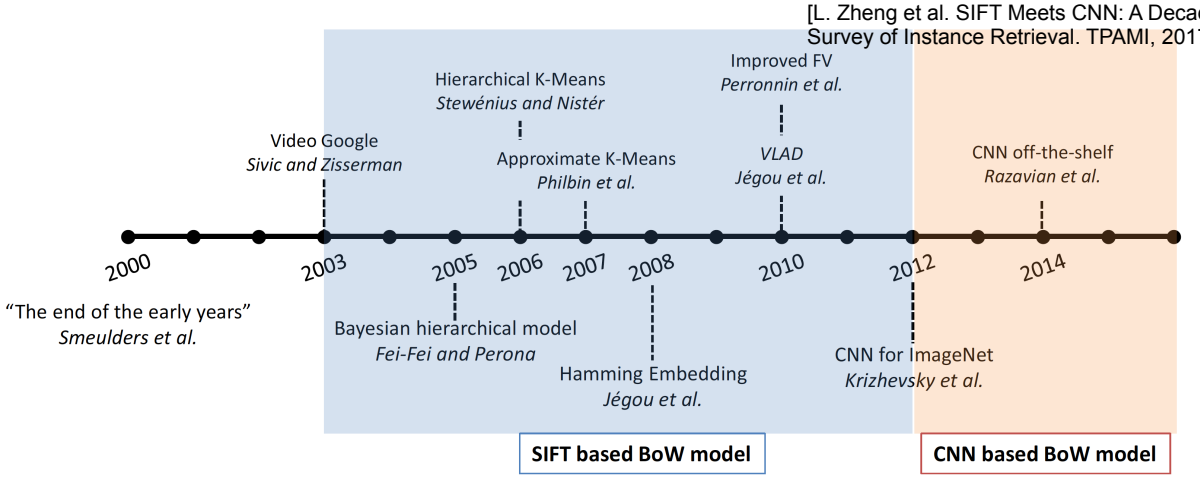
Goal—Improving Image Retrieval Results by Utilizing Contextual Information

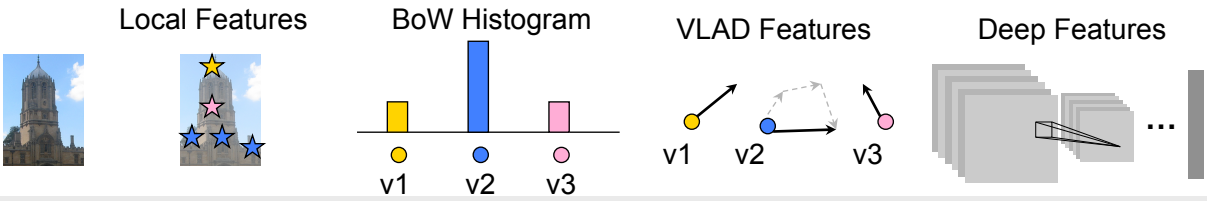


Changes in scaling, viewpoint, lighting condition, occlusion, etc.

State-of-the-Art Visual Features for Retrieval

[L. Zheng et al. SIFT Meets CNN: A Decade Survey of Instance Retrieval. TPAMI, 2017]

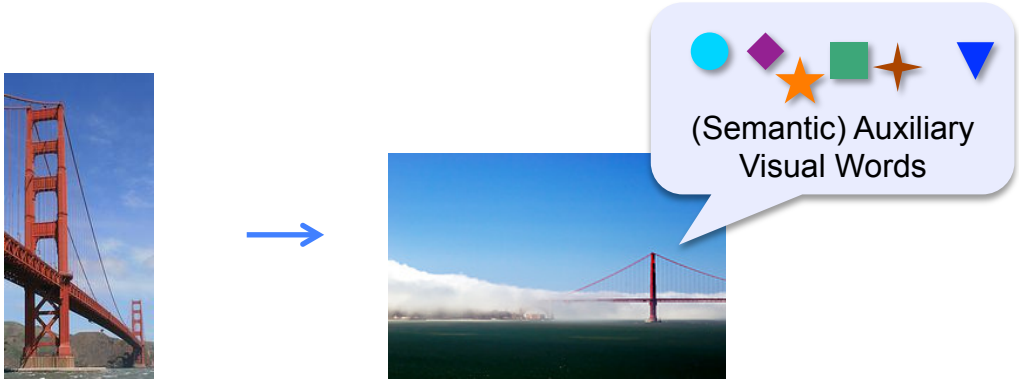




1

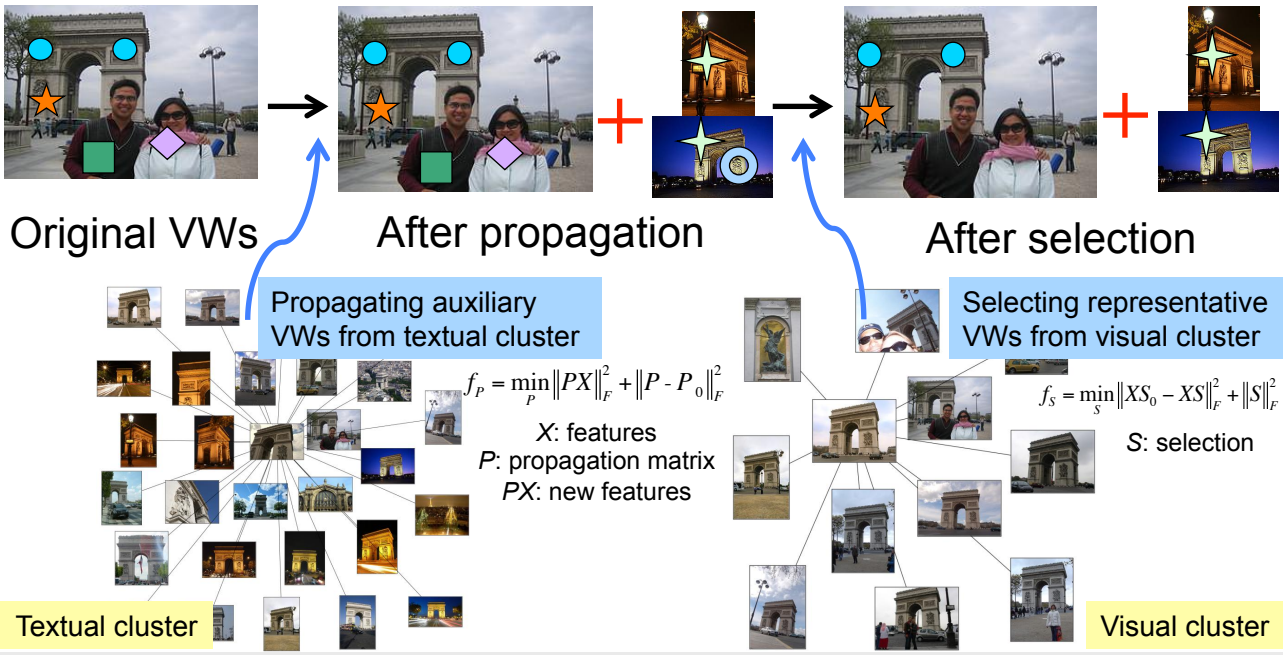
BoW for Image Retrieval

Unsupervised Semantic Feature Discovery



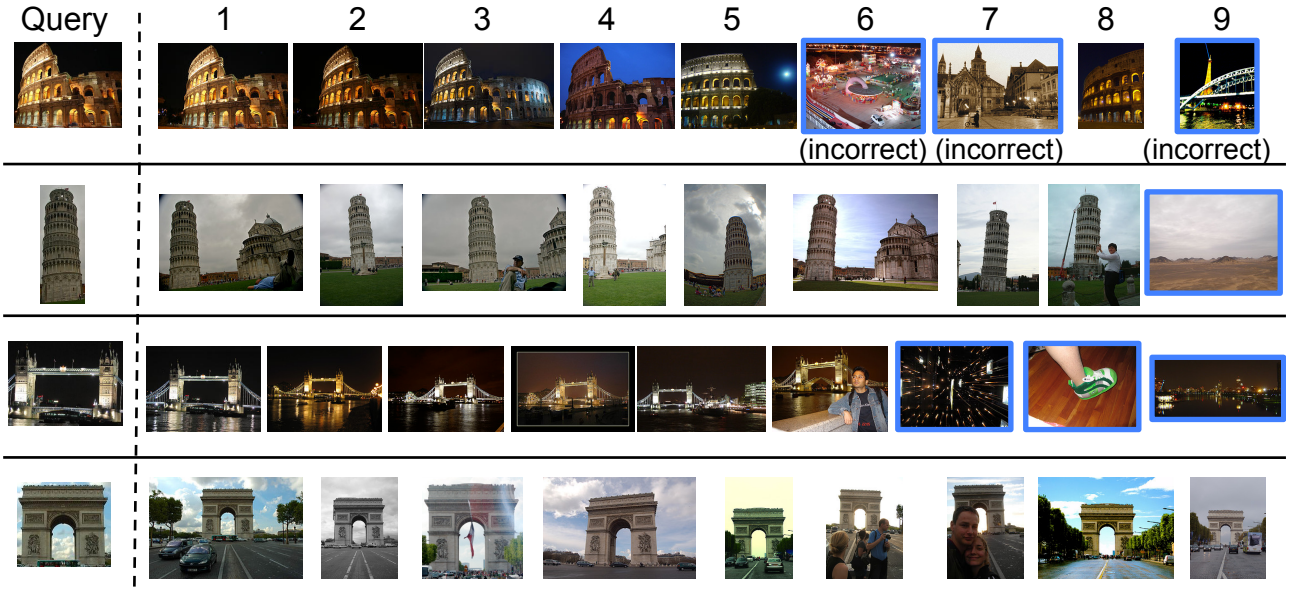
Yin-Hsi Kuo, Wen-Huang Cheng, Hsuan-Tien Lin, Winston H. Hsu, "Unsupervised Semantic Feature Discovery for Image Object Retrieval and Tag Refinement," IEEE Transactions on Multimedia, vol.14, no. 4, pp. 1079-1090, 2012.
Yin-Hsi Kuo, Hsuan-Tien Lin, Wen-Huang Cheng, Yi-Hsuan Yang, and Winston H. Hsu, "Unsupervised Auxiliary Visual Words Discovery for Large-Scale Image Object Retrieval," IEEE CVPR, pp. 905-912, 2011.

Auxiliary Visual Feature Discovery (Offline)

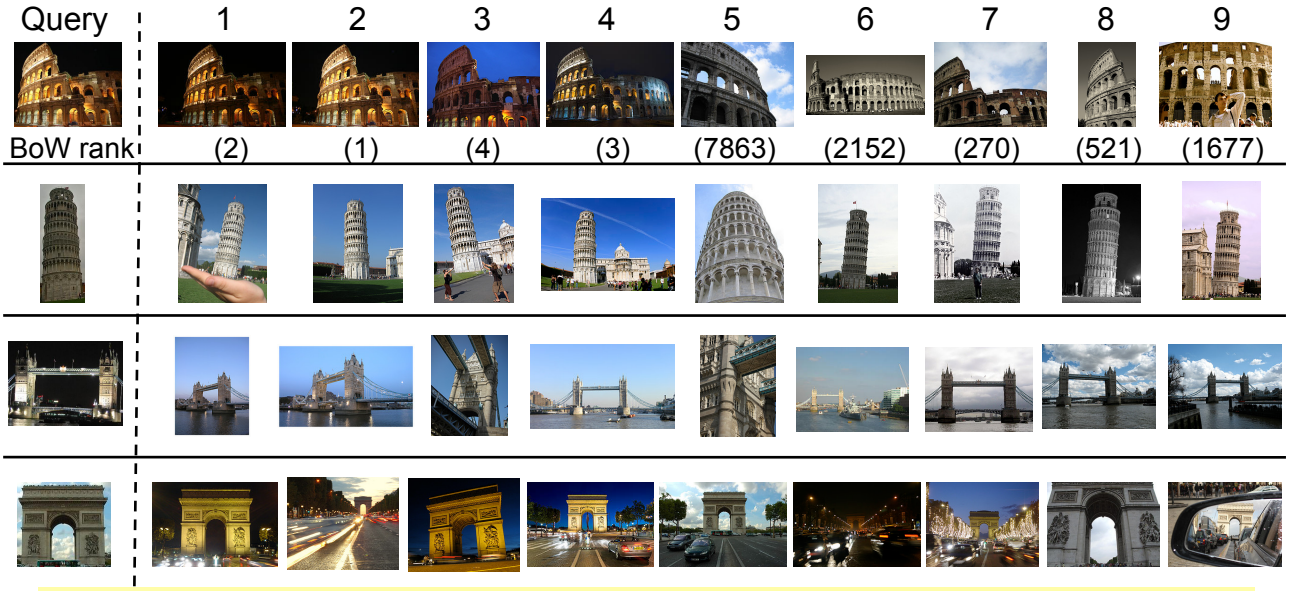


National Taiwan University

Original BoW Retrieval Results



Diverse and Semantically Related Results

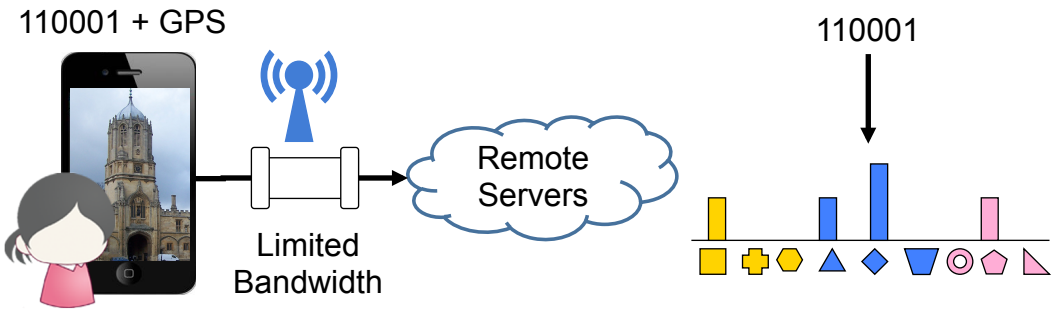


Improving the traditional bag-of-words model from 0.245 to 0.516 in MAP (+110.6%)

2

VLAD for Mobile Visual Search

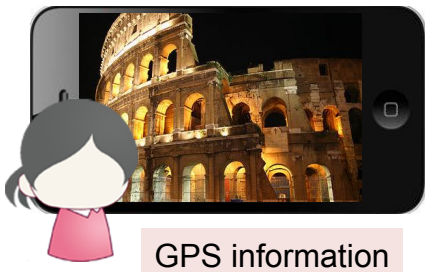
BoW reconstruction from binary codes (via VLAD)



Yin-Hsi Kuo and Winston H. Hsu, "De-Hashing: Server-Side Context-Aware Feature Reconstruction for Mobile Visual Search," IEEE TCSVT, vol. 27, no. 1, pp. 139-148, 2017.
Yin-Hsi Kuo and Winston H. Hsu, "Rank-Preserving and Unsupervised Hash Learning from Auxiliary Contextual Cues," ACM ICMR, pp. 399-402, 2014.

Challenges and Opportunities for Mobile Visual Search (MVS)

- Challenges
  - Limited bandwidth, memory, computing power
- Opportunities
  - Contextual information (e.g., GPS)
  - Lightweight computing (e.g., feature extraction)
  - Powerful computing on remote (cloud) servers



GPS information

Traditional approach: transmitting images

Mobile visual search: transmitting compact codes (e.g., 00010111011100)

Remote Servers

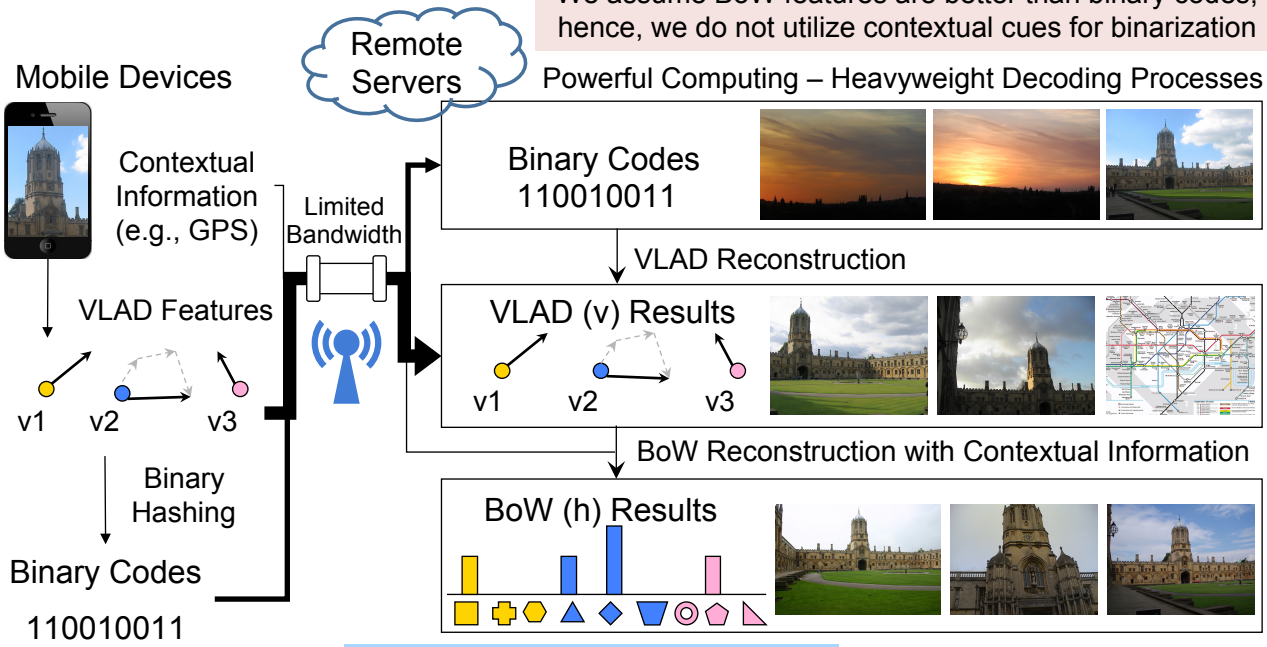
Generating discriminative features before matching

CMLab Communications & Multimedia Lab.

Leveraging mobile devices and cloud servers for balancing effectiveness and efficiency of MVS

Context-Aware Feature Reconstruction

We assume BoW features are better than binary codes; hence, we do not utilize contextual cues for binarization



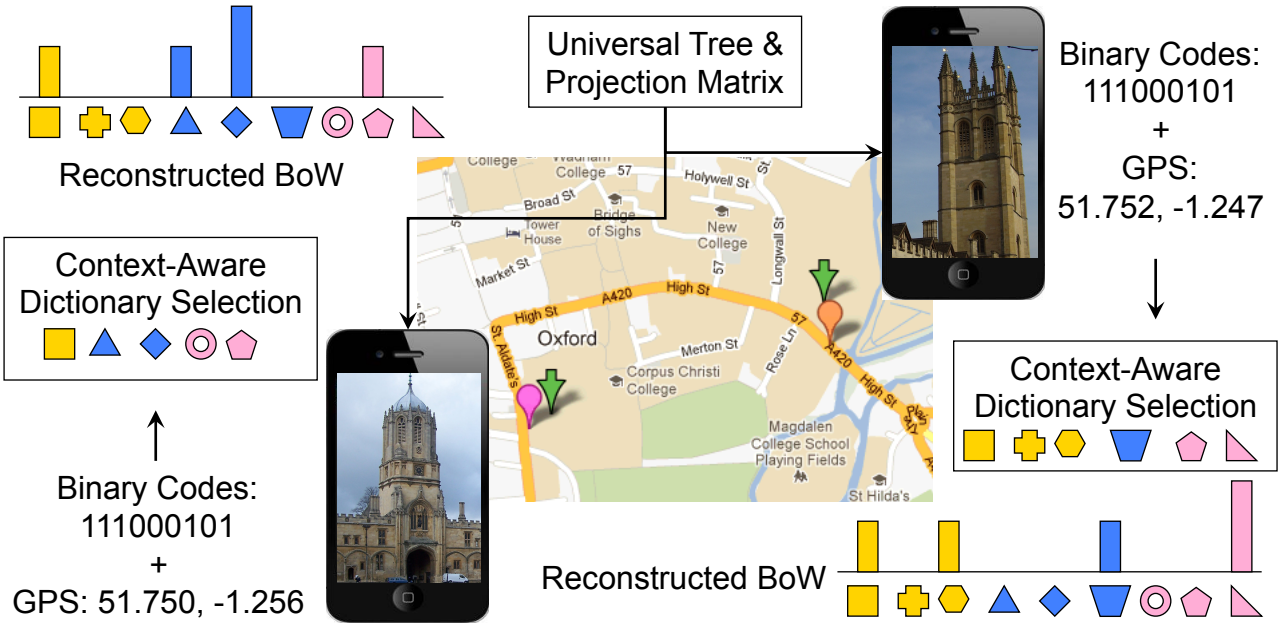
CMLab Communications & Multimedia Lab.

VLAD as an intermediate feature

$$f_h = \min_h \|v - Dh\|_2^2 + \lambda \|h\|_1$$

D: the difference between the centers of BoW and VLAD

Context-Aware Dictionary Selection



8% ~ 12% relative improvement

CMLab
Communications & Multimedia Lab.

National Taiwan University

3

Deep Features for Product Search

Rank-based candidate selection for feature learning

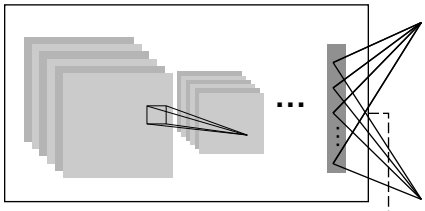
Query image → The same or similar products

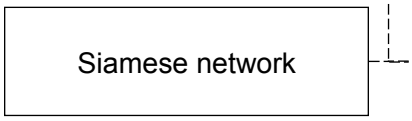
E-commerce datasets

CMLab
Communications & Multimedia Lab.

Yin-Hsi Kuo and Winston H. Hsu, "Feature Learning with Rank-Based Candidate Selection for Product Search," ICCV Workshop, 2017.

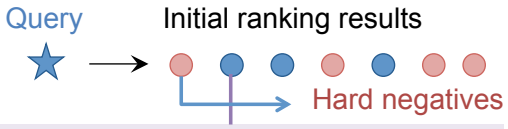
End-to-End Learning for Deep Features



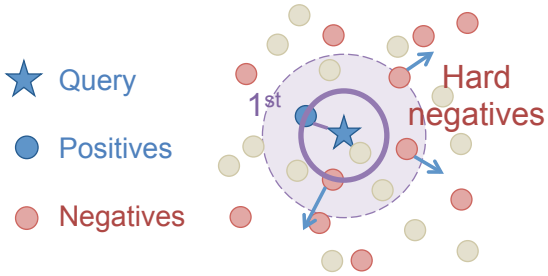


Contrastive loss (target, positive/negative)
Triplet loss (target, positive, negative)

Hard negative mining



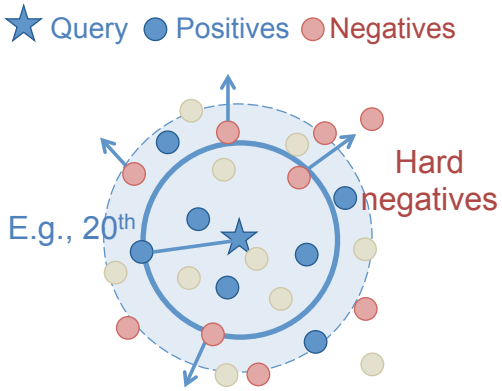
Pushing hard negatives away from the closest positive

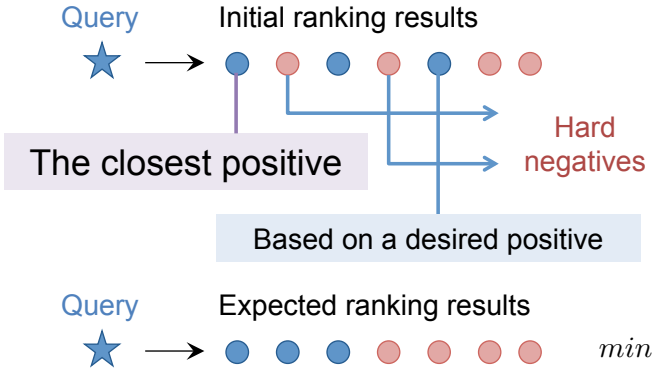


R. Arandjelović et al. NetVLAD: CNN architecture for weakly supervised place recognition. CVPR, 2016.
S. Chopra, R. Hadsell and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," CVPR, 2005.

Rank-Based Candidate Selection—Hard Negatives

- Enforcing top-ranked results with more relevant images

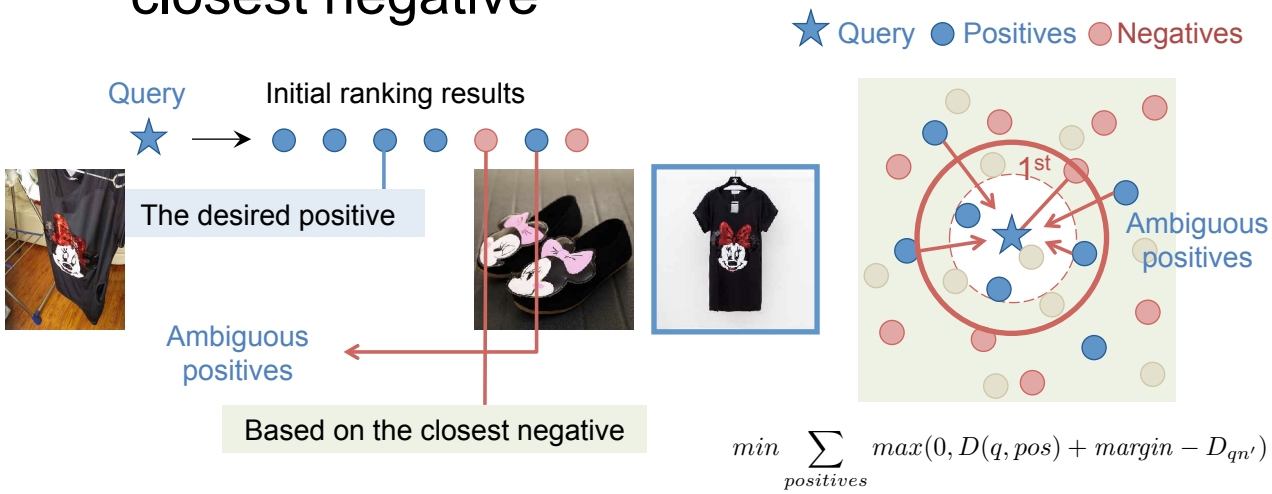




$$min \sum_{negatives} max(0, D_{qp'} + margin - D(q, neg))$$

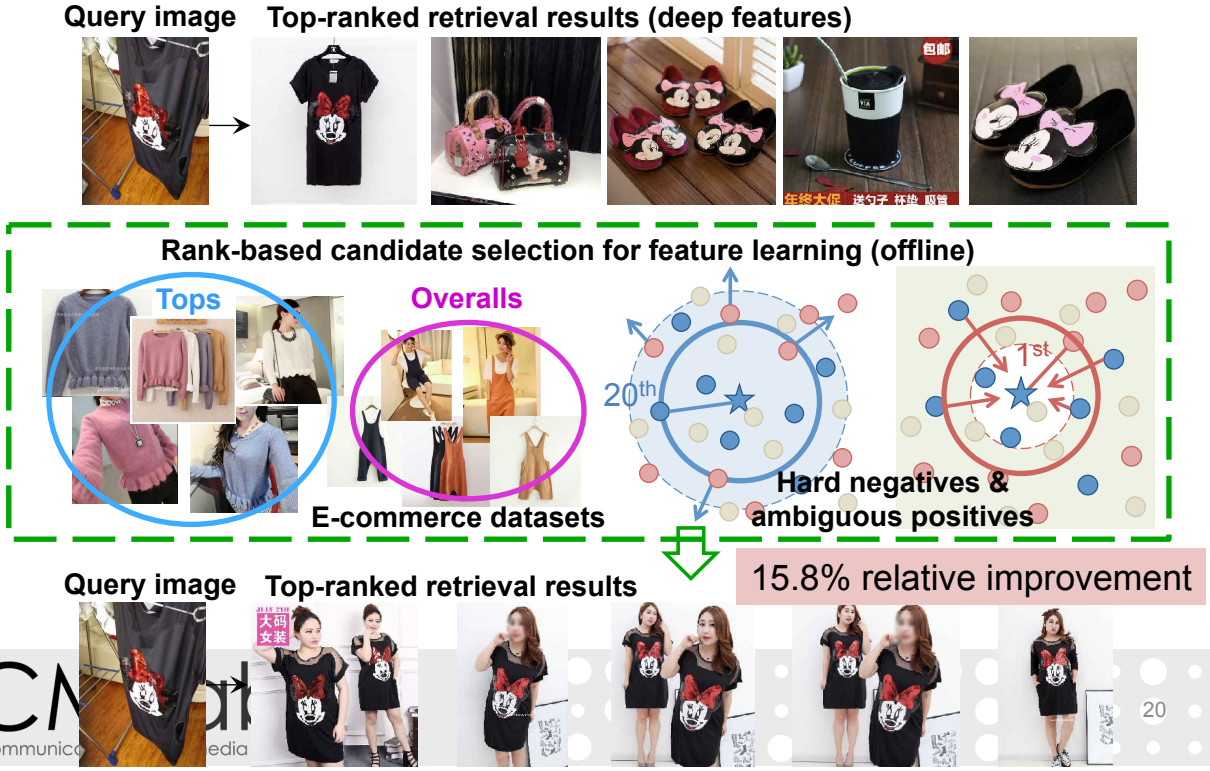
Rank-Based Candidate Selection—Ambiguous Positives

- Pulling relevant images away from the closest negative

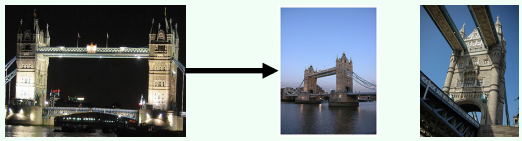
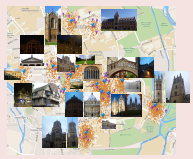


$$min \sum_{positives} max(0, D(q, pos) + margin - D_{qn'})$$

Feature Learning by Rank-Based Candidate Selection for Product Search



Exploiting Different Contextual Information with State-of-the-art Visual Features

	Offline Learning / Training	Online Ranking via visual features
1 **BoW** BoW (bag-of-words) [Sivic and Zisserman, ICCV'03]	Tags Tower bridge, London, Bridge, England	
2 **VLAD** VLAD (vector of locally aggregated descriptors) [Jégou et al. CVPR'10]	GPS information 	
3 **Deep features** Deep (convolutional) features [Razavian et al. CVPRW'14]	Product information Tops Skirts 	

National Taiwan University

Thank you for your attention!

Q & A