COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Visual Search: 3 Levels of Real-Time Feedback

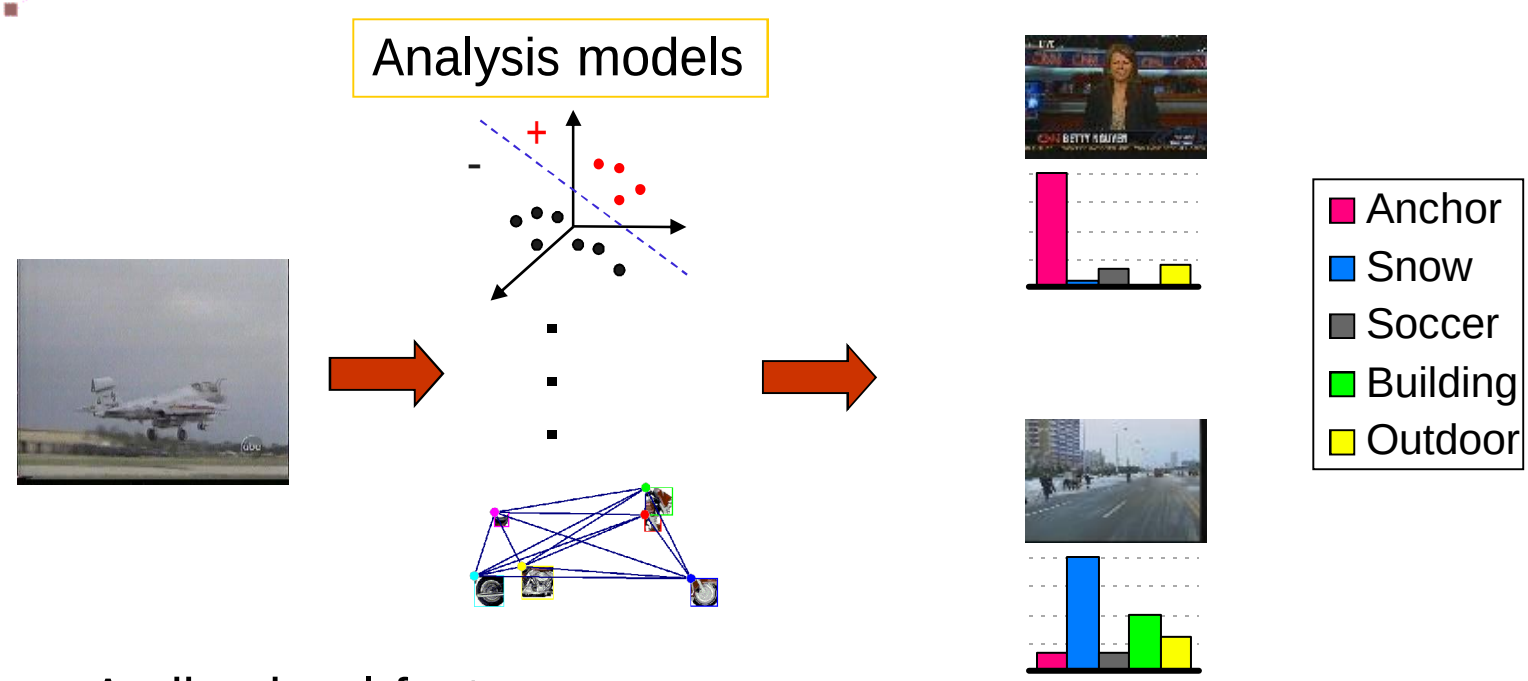
Prof. Shih-Fu Chang

Department of Electrical Engineering

Digital Video and Multimedia Lab

http://www.ee.columbia.edu/dvmm

A lot of work on Image Classification

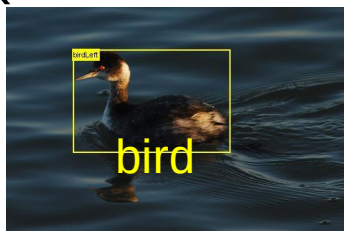


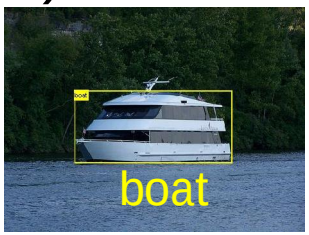
- Audio-visual features
- Geo, social, camera metadata
- User/mission context

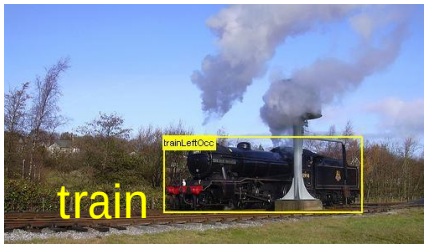
- Rich semantic labels

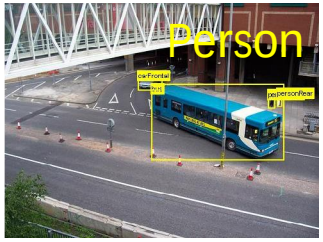
Object Detection

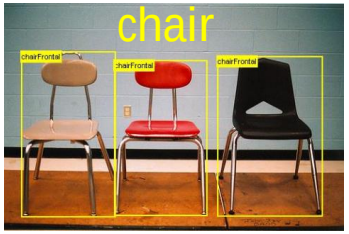
(PASCAL VOC)

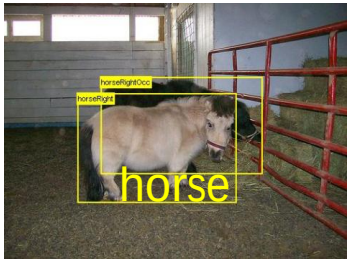


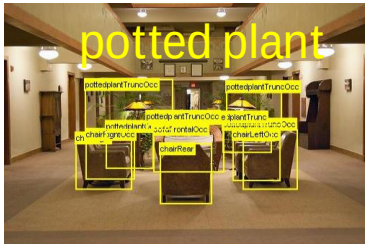


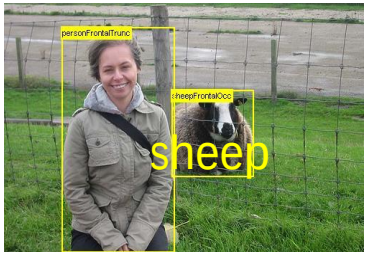












Research community growing fast!

(as of Nov. 2009)

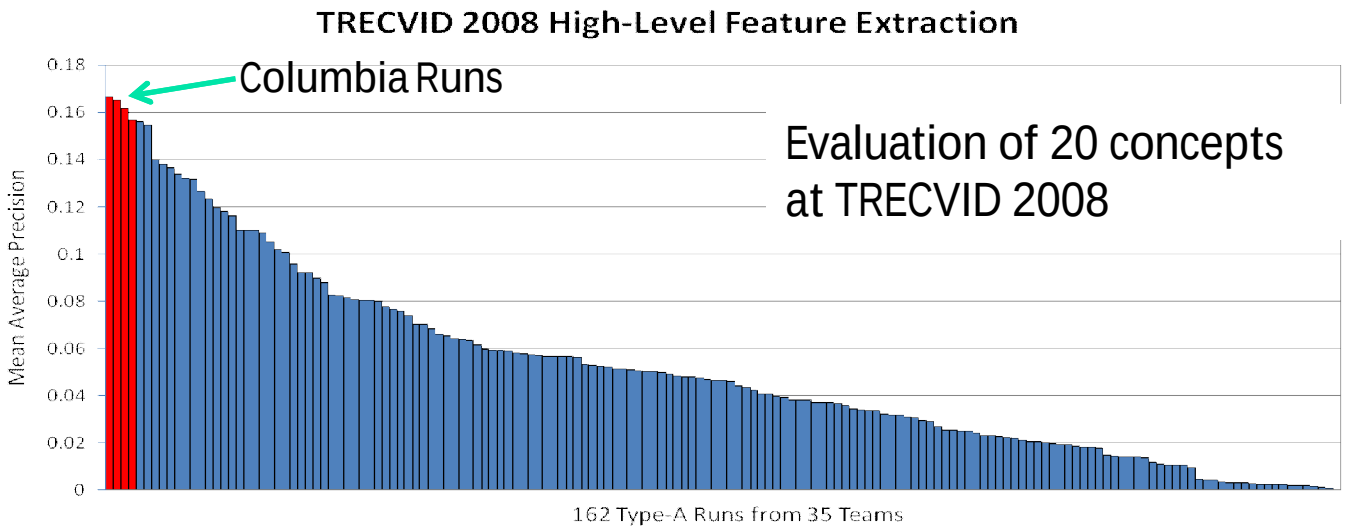
	Data domain	amount	types	Lexicon size
TRECVID	Broadcast news, documentary, surveillance, Internet, ...	400 hours Sound & Vision 170 hours Television News 100 hours BBC rushes (130,000+ subshots)	video shots, keyframes	10 (2004, 2005) 39 (2006, 2007) 20 (2008) 130 (2010)
LSCOM	Broadcast news	170 hours Broadcast News	video	1000+ concepts
CalTech256	Internet Images	30,607 images	images	256 classes
PASCAL	Internet Images	9,963 images 24,640 annotated objects	images, objects	20 classes
Tiny Image	Internet Images	80,000,000 tiny images (32x32)	images	75,378 WordNet nouns
LabelMe	Internet and UGC conten	30,369 images from 183 folders	images, keyframes	111,490 object labels
ImageNet	Internet images	9,386,073 images	images	14,847 WordNet synsets
Lotus Hill Dataset	Internet Images	500,000+ images and keyframes	images, keyframes	280 object classes

CuZero: 400+ visual classifier models

Columbia University in the City of New York
CuZero
digital video | multimedia laboratory

concept detection models:
objects, people, location, scenes, events, etc

airplane airplane_takeoff airport_or_airfield armed_person building car cityscape crowd desert dirt_gravel_road entertainment explosion_fire forest highway hospital insurgents landscape maps military military_base military_personnel mountain nighttime people-marching person powerplants riot river road rpg shooting smoke tanks urban vegetation vehicle waterscape_waterfront weapons weather



TRECVID: Concept Detection Examples

- Top five classification results

Classroom

Demonstration Or Protest

Cityscape

Airplane flying

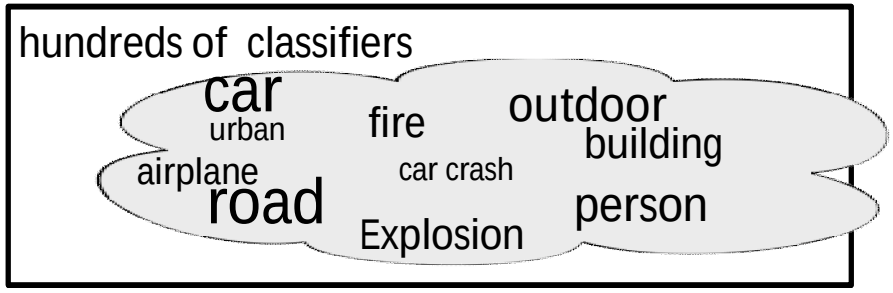
Singing

Problem: User Gap

- Given a new search target, users have difficulty in choosing appropriate concept classifiers

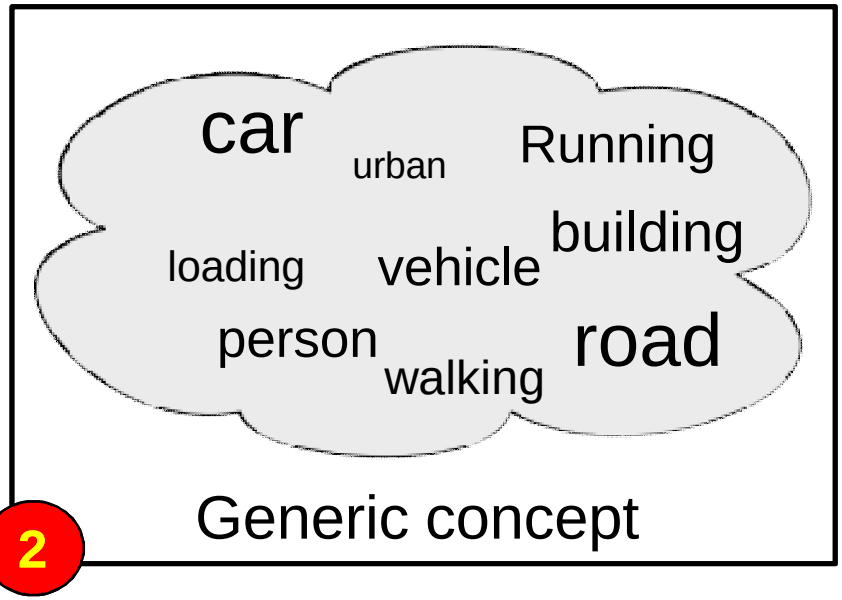
Find shots of something burning with flames visible

Which classifiers to use? Which classifiers work?



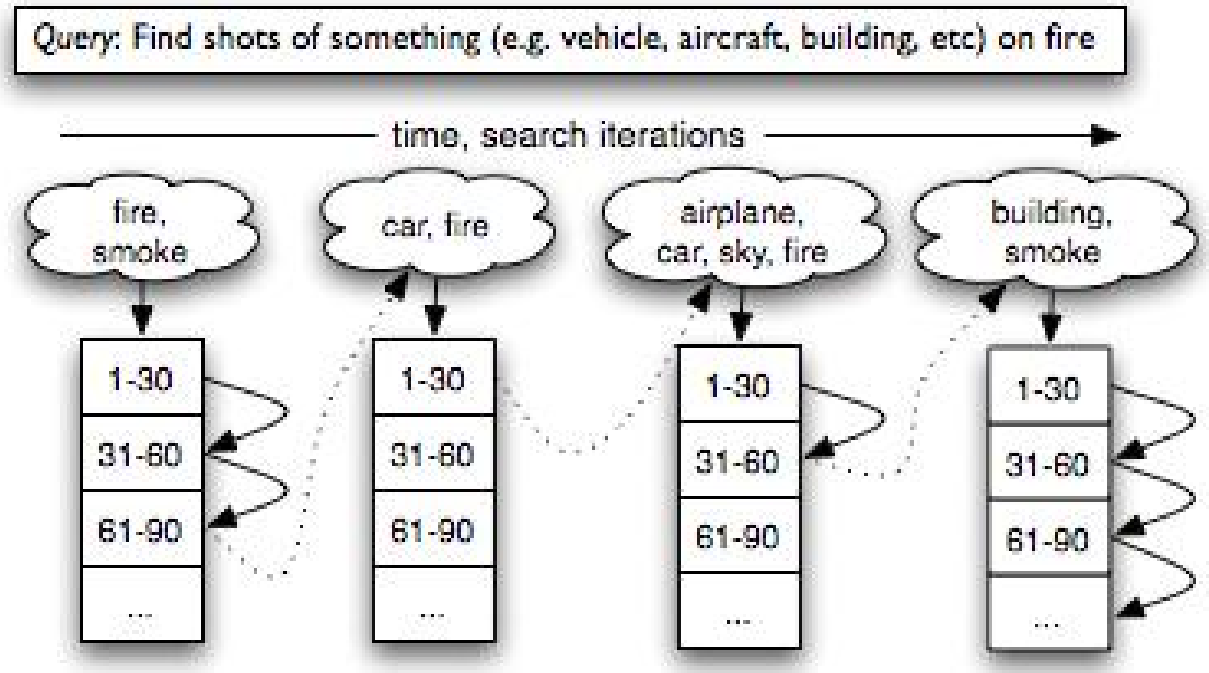
Problem:
Specific example
vs.
Generic concept

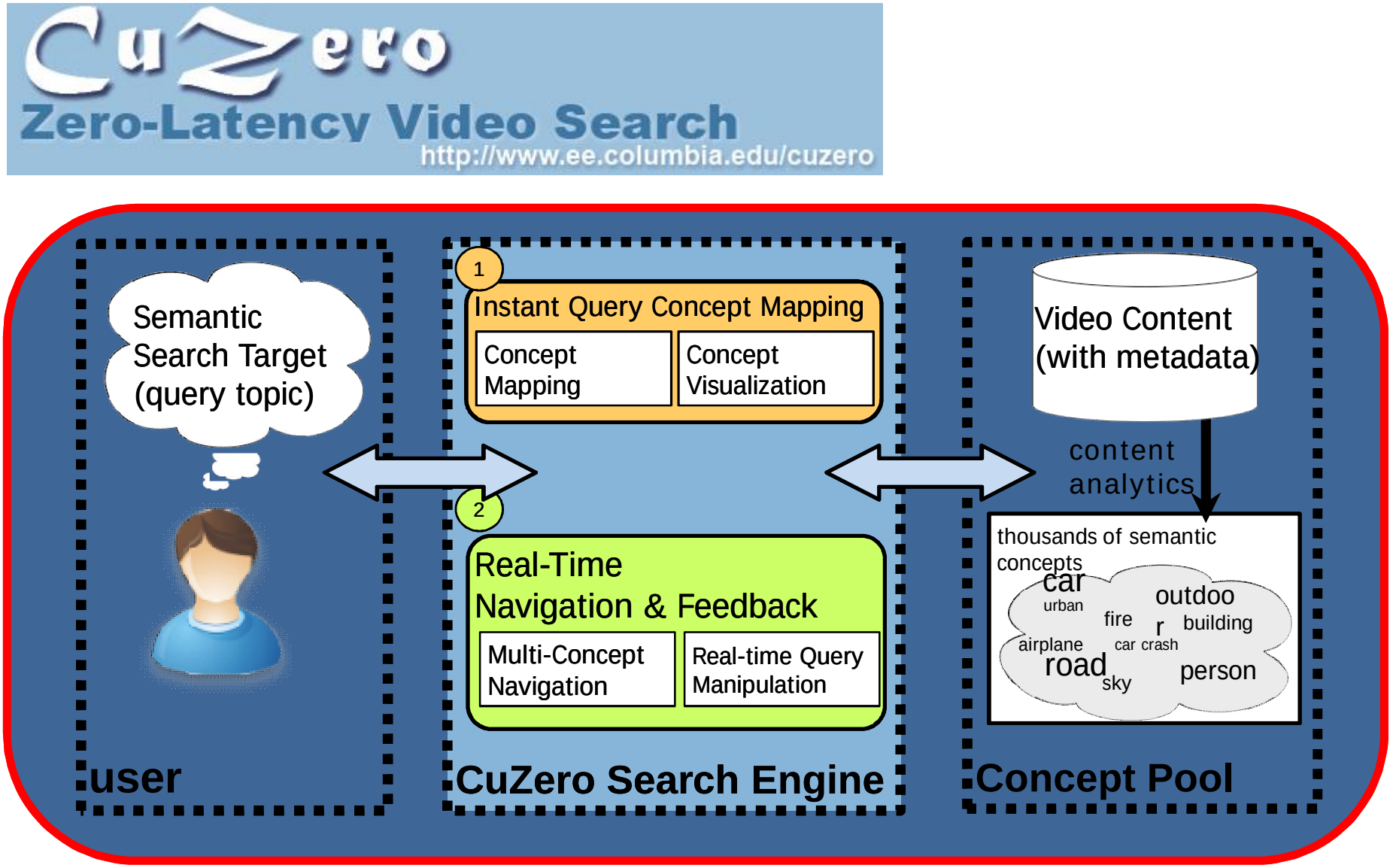
Query: “find person running around a building”



Pains of Frustrated Users

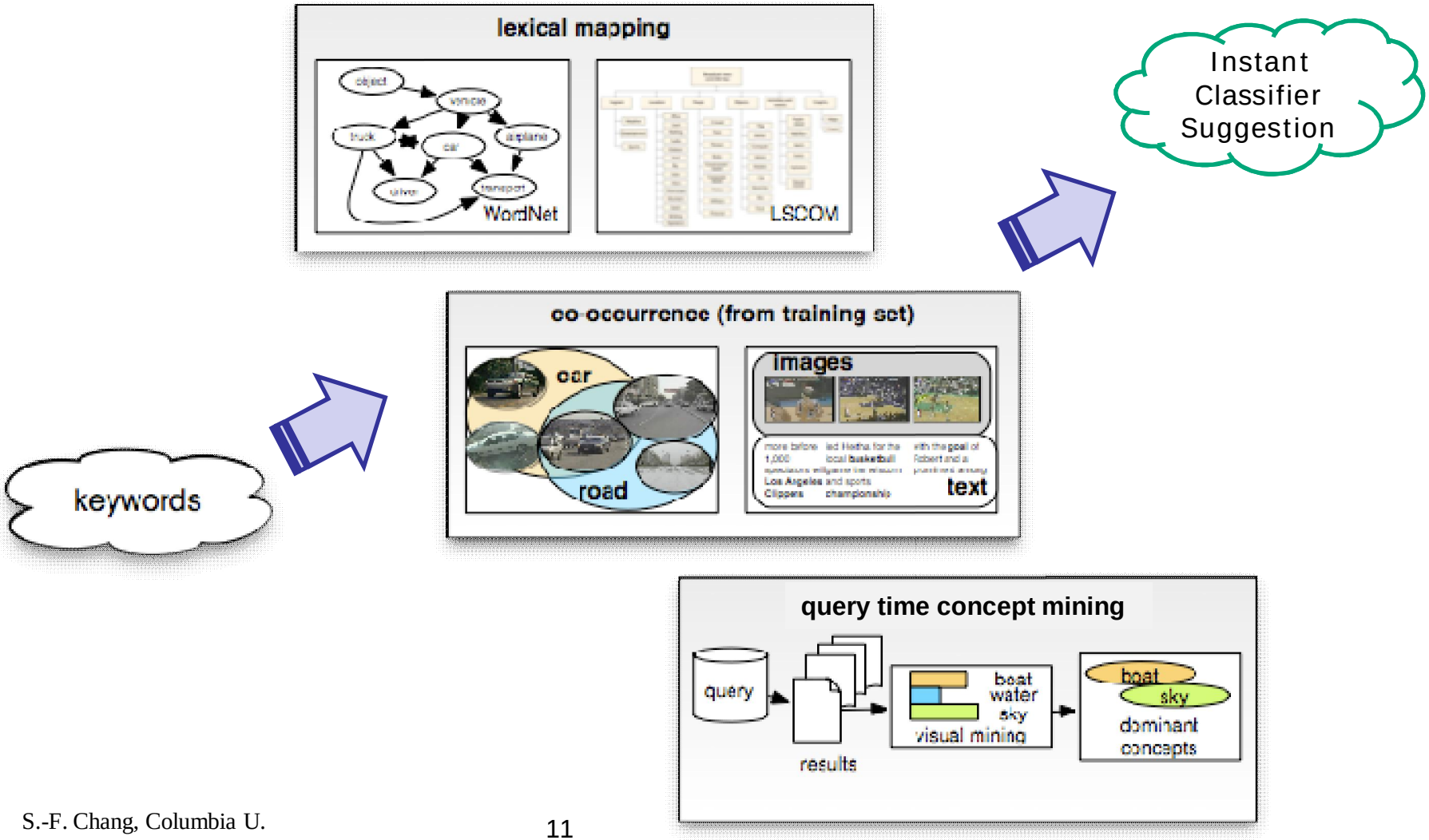
- User forced to take “one shot” searches, iterating queries with a trial and error approach…





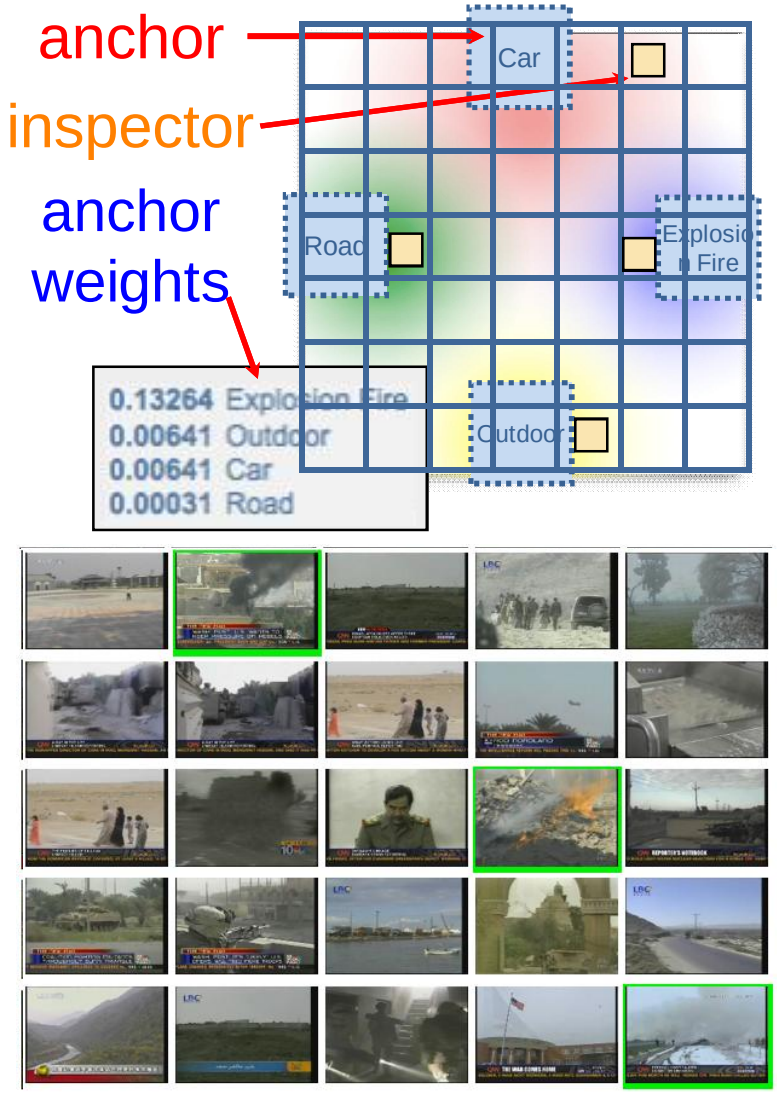
(Zavesky and Chang ,
Multimedia Info Retrieval MIR '08)

Instant Feedback:
Instant visual concept suggestion



Instant Feedback: Mapping → formulate query → refine

- Precise specification of search anchors
- Refine query by sliding inspector through the map
- Intuitive control:
  Closer to anchor is ‘more like it’
- Increases exploration breadth, instead of single list browsing
- Instant result display for each location, without new query computation



Demos

- Find lake front buildings in the Central Park
- Find person walking around building
- Find a car on a road in a snowy condition
- Find urban explosion scenes in UAV videos

Instant Feedback of Query Manipulation

Multi-modal query input

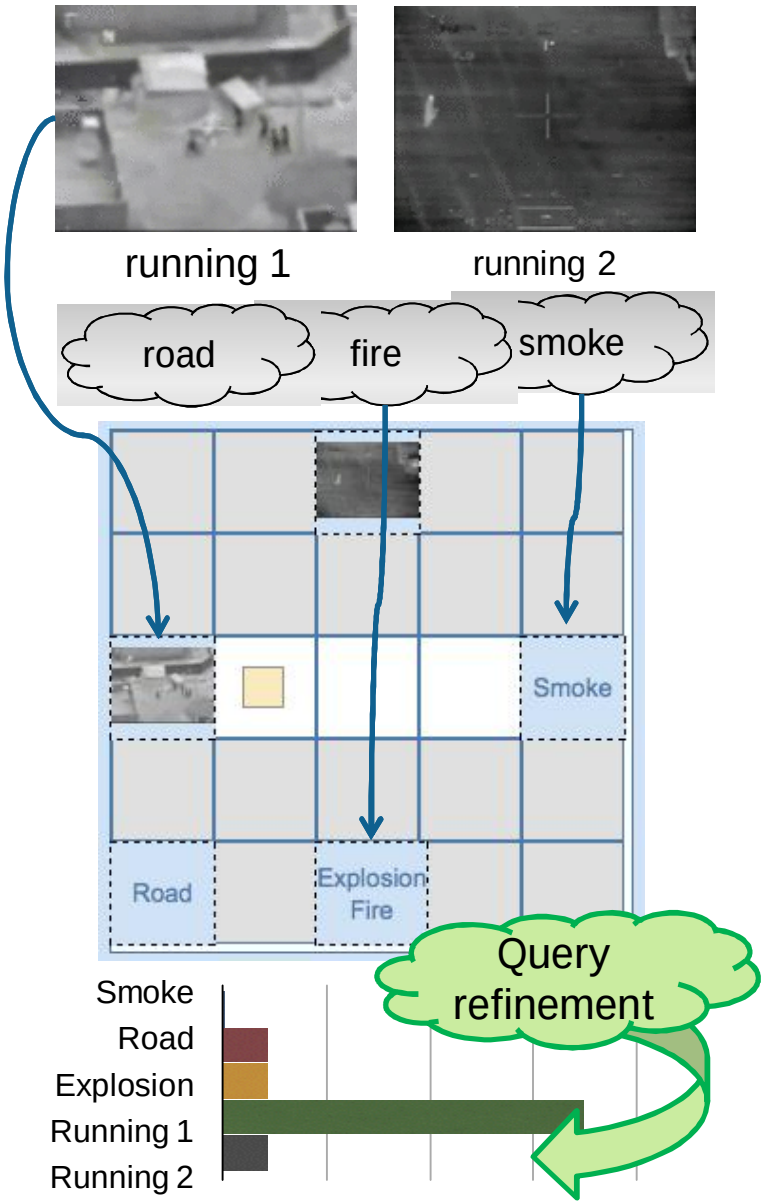
- Visual examples (motion track, object/image samples)
- >350 object/scene models
- Textual, geo-spatio-temporal cues

Real-Time Sliding Query Panel

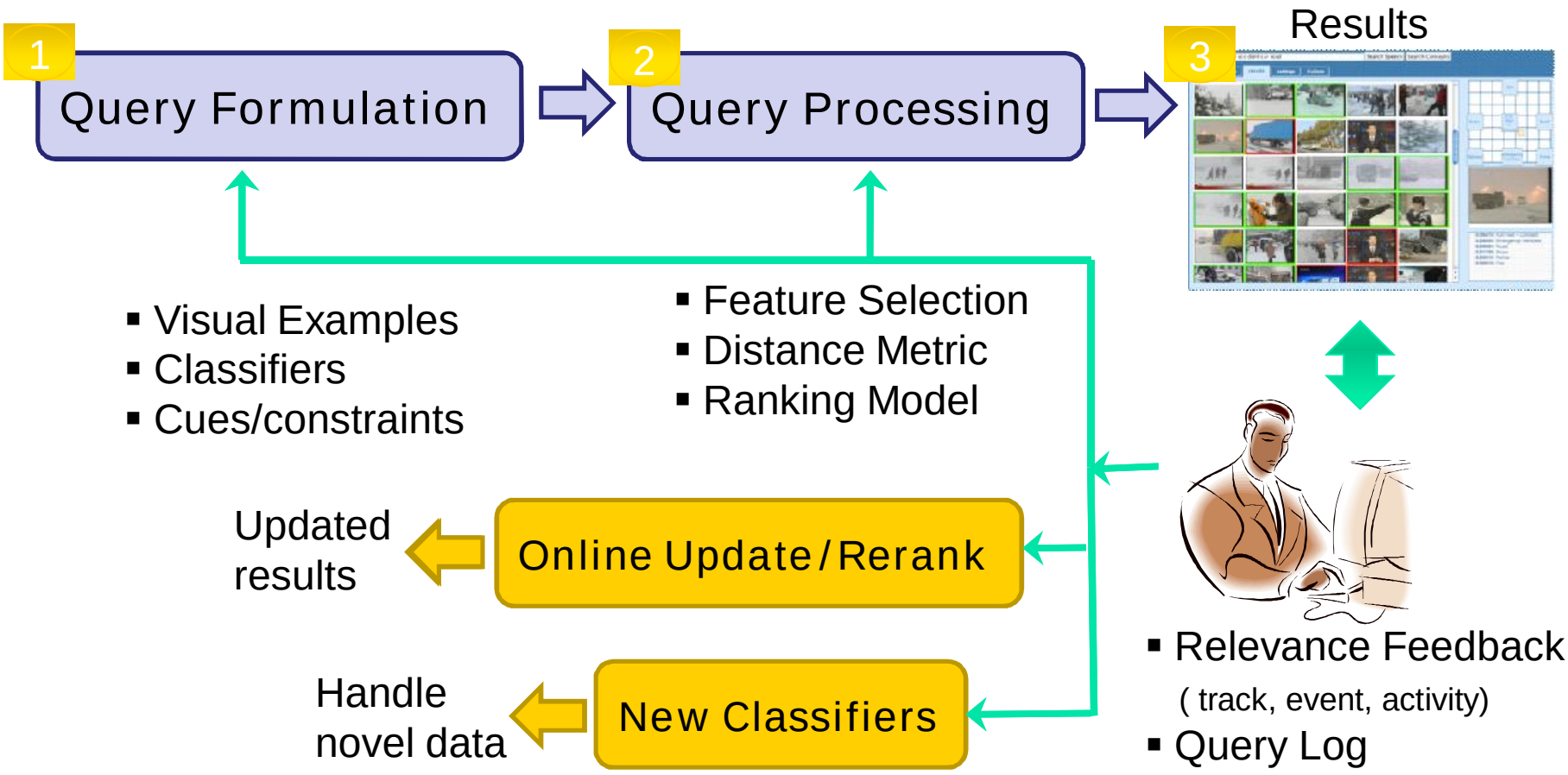
- 2D sliding panel for arbitrary query manipulation
- Instant query result update after refinement
- Intuitive feedback for query weight adjustment

Distributed, lightweight environment

- Distributed client-server approach for fast deployment with large data archive.
- Prototypesof 200+ hours of TRECVID videos and VIRAT videos

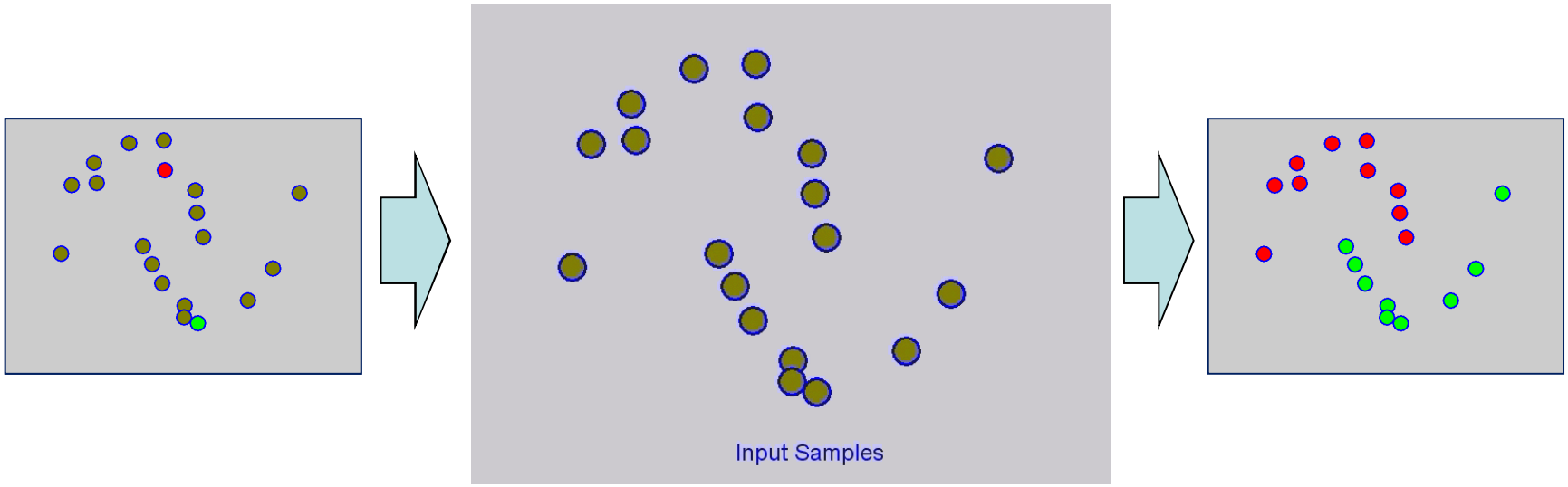


User feedback can be used in many other ways



Graph-based feedback propagation

- Propagate user feedback to larger collection



Input samples with sparse labels | Label propagation on graph | Label inference results

Positive — Negative

Unlabeled | Positive | Negative

$$f^* = \min_f Q(f, y, \mathcal{G}(V, W))$$

$\mathcal{G}$ -- graph
V -- graph node
W -- weight matrix
$\mathcal{Q}$ -- risk function
f -- classification
y -- label matrix

A hot topic in Machine Learning

- Given initial labels, Y, find classification function F over graph nodes

Label smoothness

Fit known labels

$$\begin{aligned} \mathcal{Q}(F) &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left\|\frac{F_{i.}}{\sqrt{D_{ii}}}-\frac{F_{j.}}{\sqrt{D_{jj}}}\right\|^2 + \mu\sum_{i=1}^{l}\|F_{i.}-Y_{i.}\|^2 \\ &= \mathrm{tr}\{F^\top LF + \mu(F-Y)^\top(F-Y)\} \end{aligned}$$

(Zhou, et al NIPS04)

- Gaussian fields & Harmonic functions (*Zhu et al ICML03*)

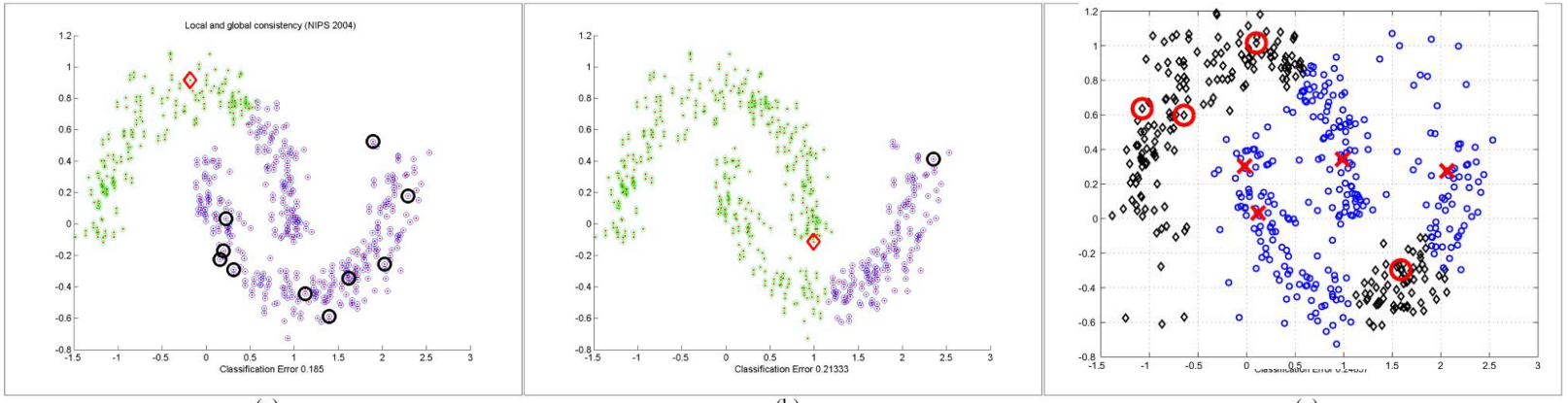
$$\mathcal{Q}(F) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\|F_{i.}-F_{j.}\|^2$$

1) $\triangle F = 0$ on unlabeled data, where $\triangle = D - W$ is the graph Laplacian;

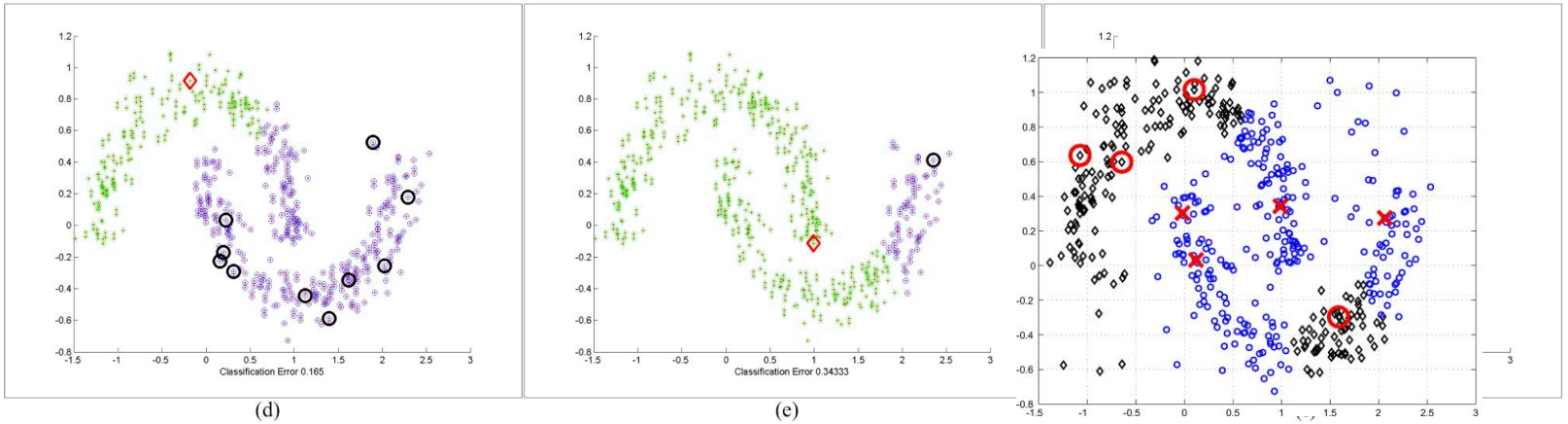
2) $F_{i.} = Y_{i.}$ on labeled data.

Many Challenging Issues

LGC Method



GFHF Method



Unbalanced Labels

Bad Label Locations

Noisy Labels

Graph Transduction via Alternating Minimization (GTAM)

(Wang, Jebara, Chang, ICML08) (Wang and Chang, CVPR09)

-- Bivariate Optimization over Labels (Y) and Prediction (F)

$$\mathcal{Q}(\mathbf{F},\mathbf{Y}) = \frac{1}{2}\mathrm{tr}\left\{\mathbf{F}^T\mathbf{L}\mathbf{F} + \mu(\mathbf{F}-\mathbf{V}\mathbf{Y})^T(\mathbf{F}-\mathbf{V}\mathbf{Y})\right\}$$

- Propagation Step
  - Given label (Y), propagate over graph, predict F

$$\frac{\partial \mathcal{Q}}{\partial \mathbf{F}^*} = 0 \Rightarrow \mathbf{F}^* = (\mathbf{L}/\mu + \mathbf{I})^{-1}\mathbf{V}\mathbf{Y} = \mathbf{P}\mathbf{V}\mathbf{Y}$$

- Label Selection Step
  - Iteratively add good labels or remove bad labels

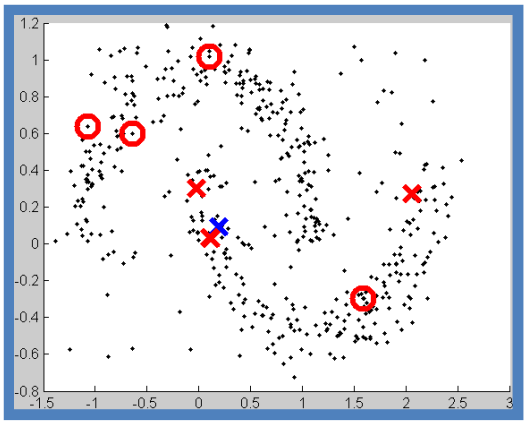
$$\mathcal{Q}(\mathbf{Y}) = \frac{1}{2}\mathrm{tr}\left(\mathbf{Y}^T\mathbf{V}^T\left[\mathbf{P}^T\mathbf{L}\mathbf{P} + \mu(\mathbf{P}^T-\mathbf{I})(\mathbf{P}-\mathbf{I})\right]\mathbf{V}\mathbf{Y}\right)$$

(TAG demo)

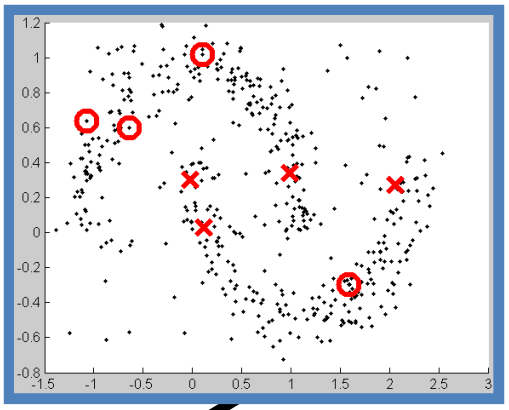
Label Diagnosis and Self Tuning

(Wang and Chang CVPR'09)

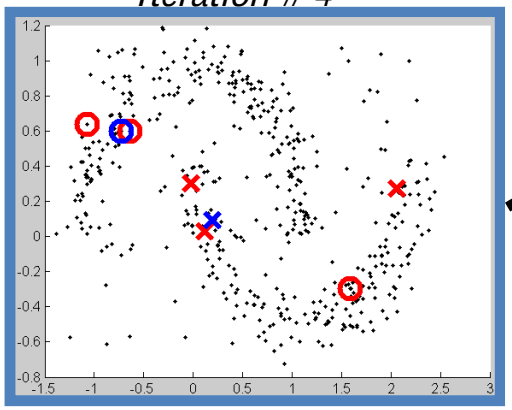
Iteration # 2



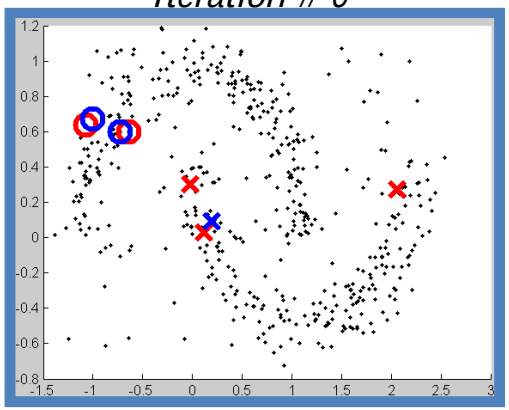
Initial Labels

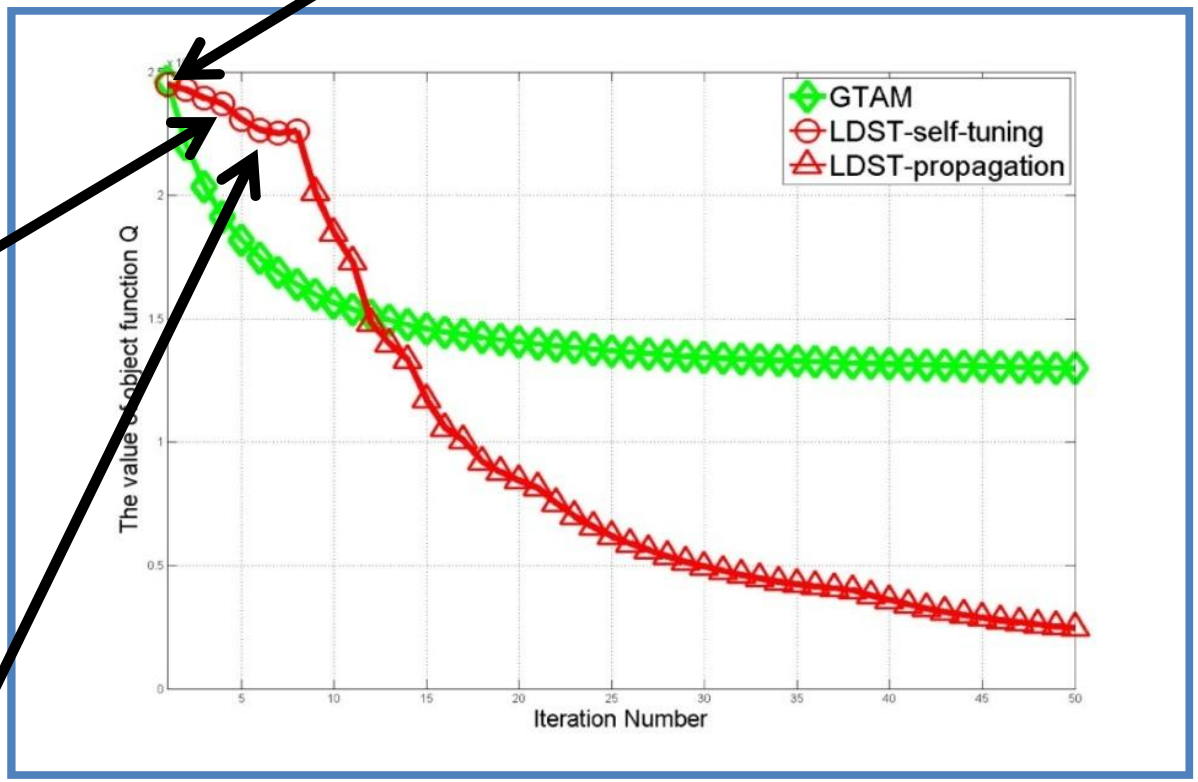


Iteration # 4



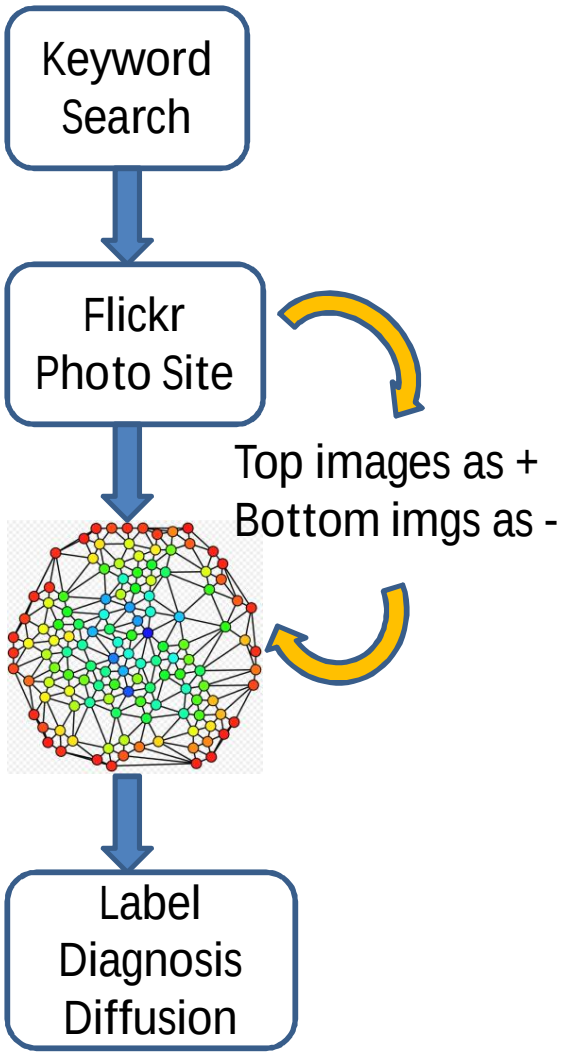
Iteration # 6



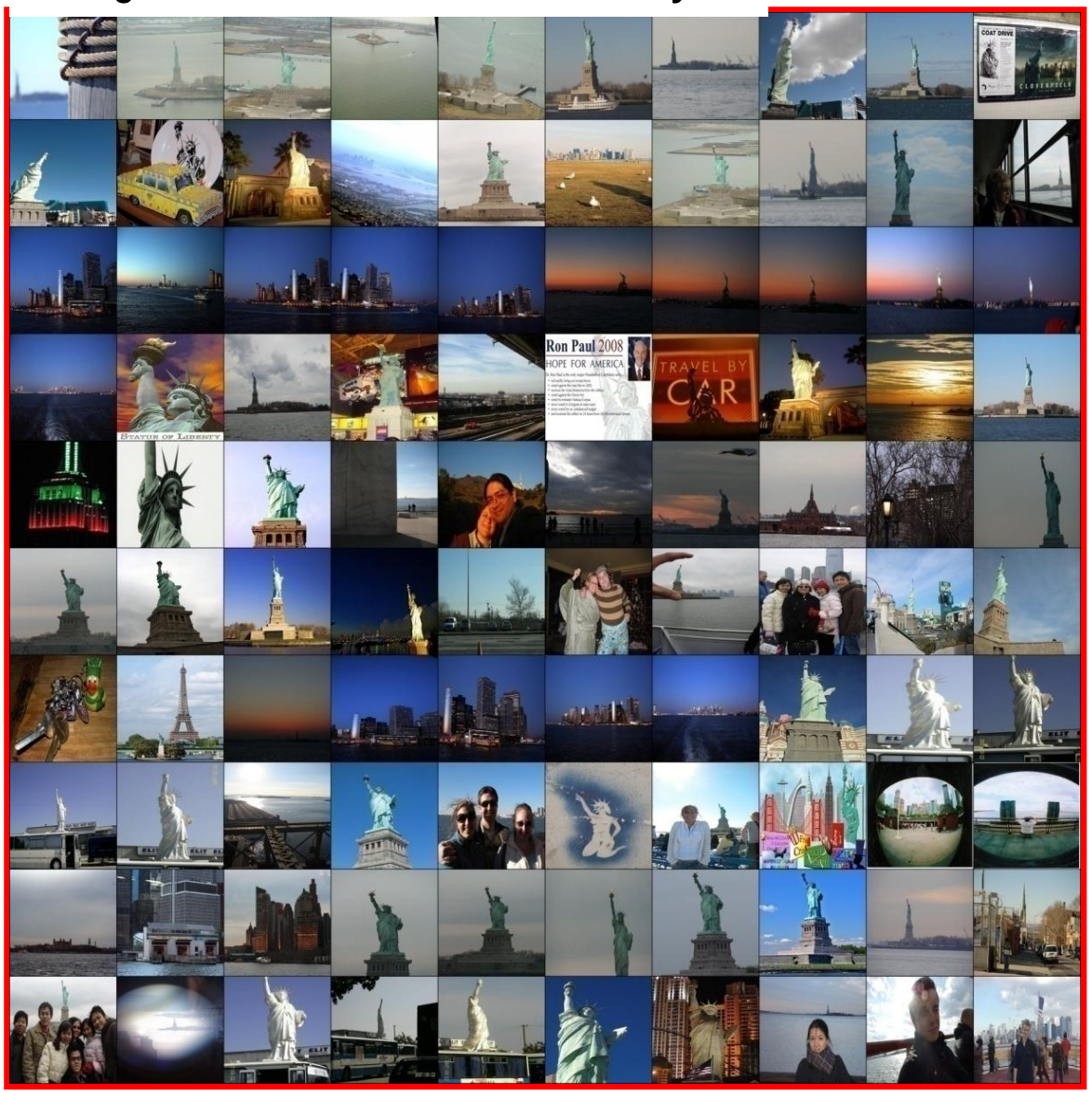


The values of the cost function Q during optimization procedure of LDST and GTAM methods.

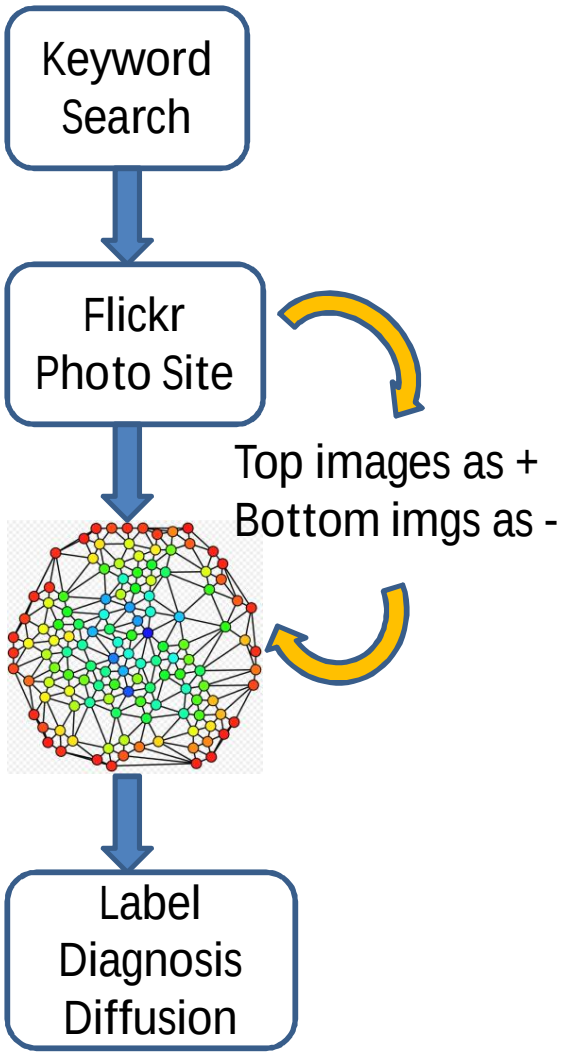
Use Pseudo-Feedback when user not available



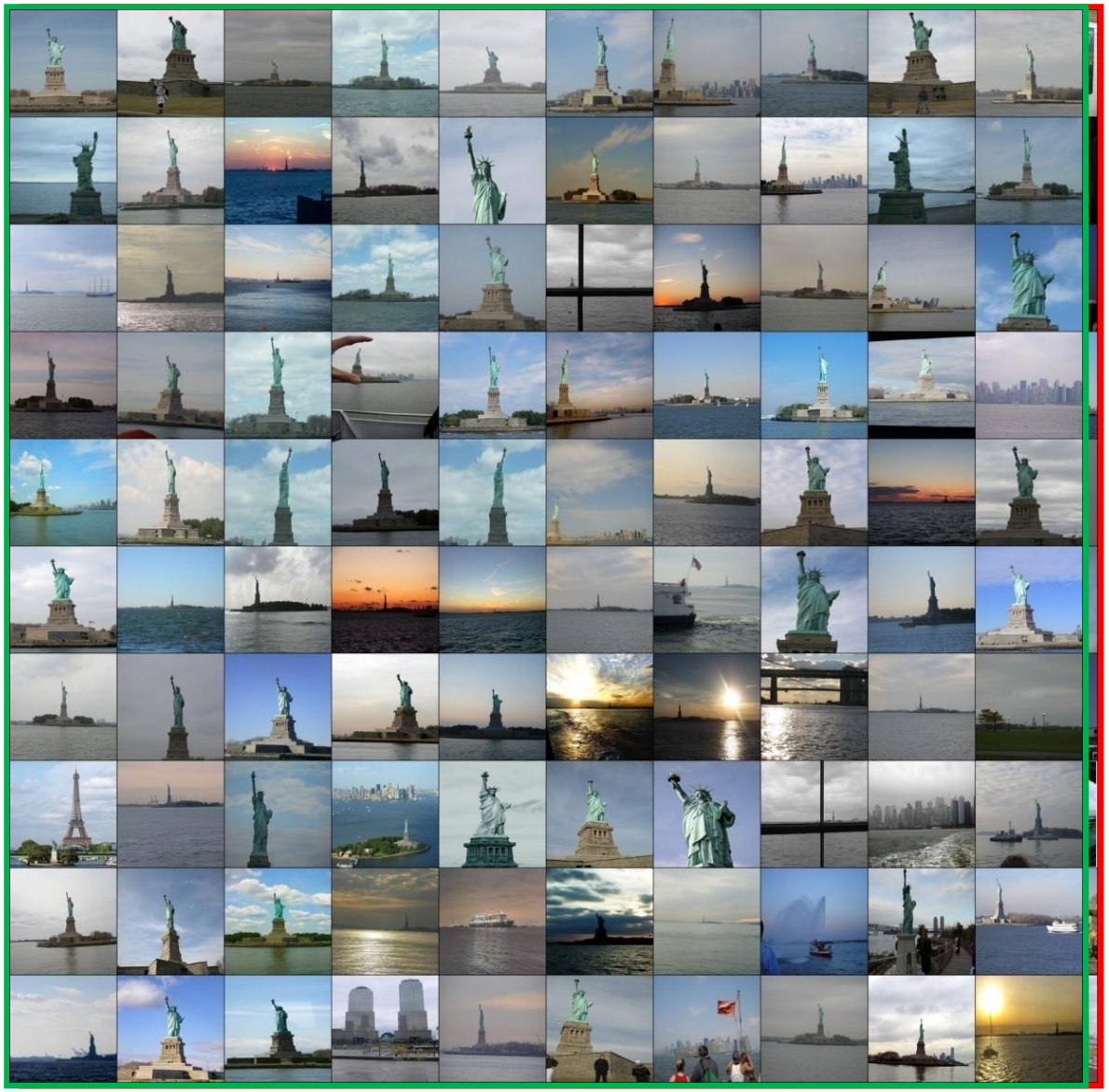
Google Search “Statue of Liberty”



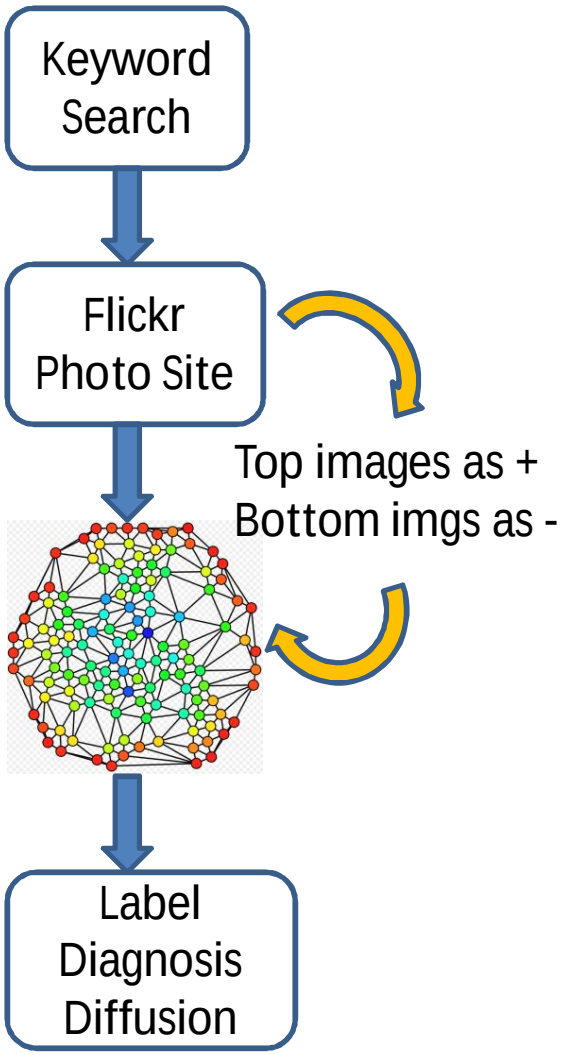
Use Pseudo-Feedback when user not available



Rerank

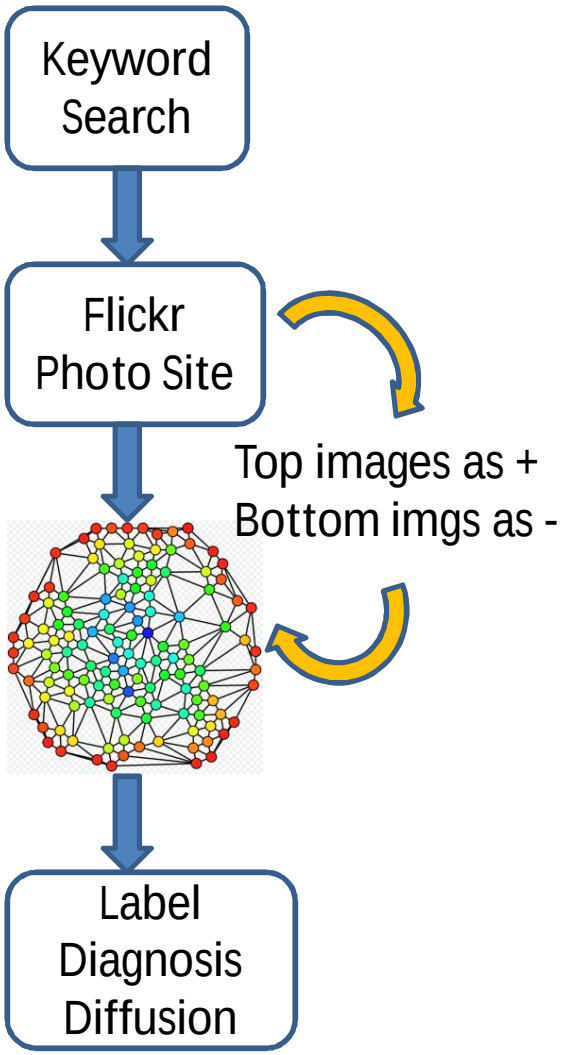


Use Pseudo-Feedback when user not available



Google Search "Tiger"

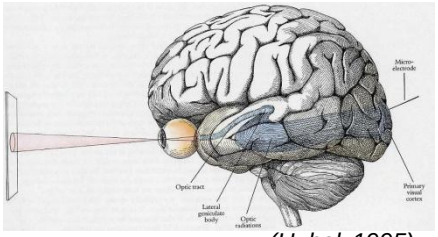
Use Pseudo-Feedback when user not available



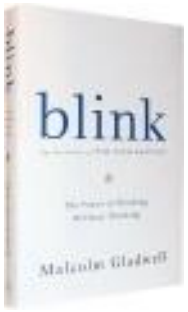
Rerank

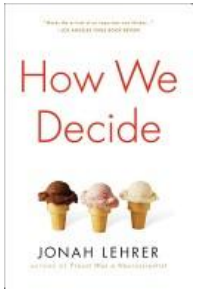
Feedback at a faster time scale via Brain State Signaling

- Human Vision:
  Superb by quick “gist”
  in the “Blink of an Eye”



(Hubel, 1995)

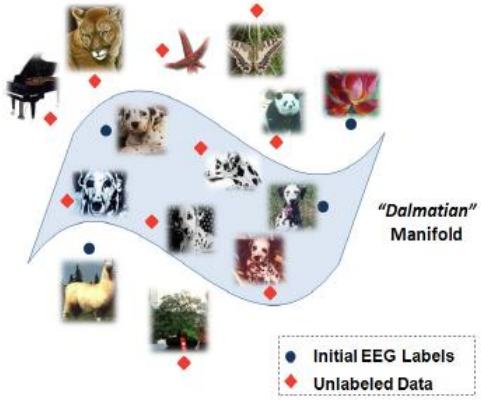




Joint work with Paul Sajda’s group

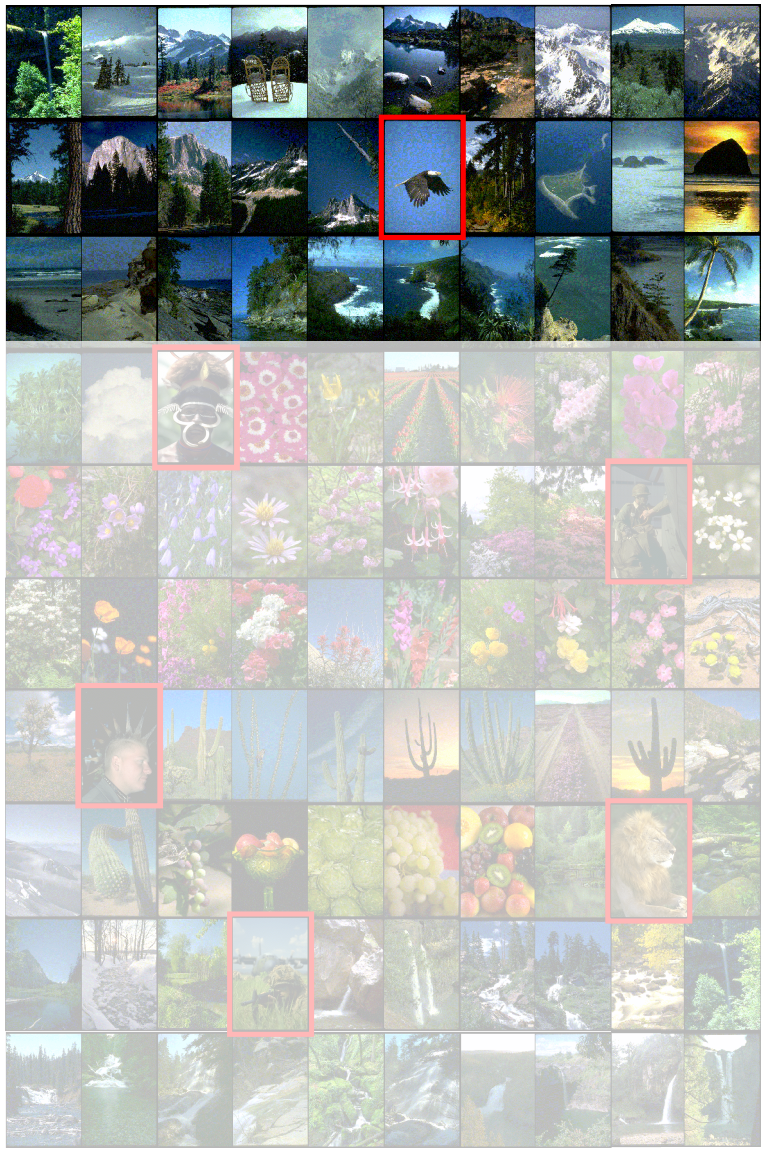
Hybrid Computer-Human Vision System

- Goal: optimally integrate neuro-vision and computer vision/machine learning to maximize information throughput and retrieval accuracy of image content



Graph-Based Visual Pattern Discovery

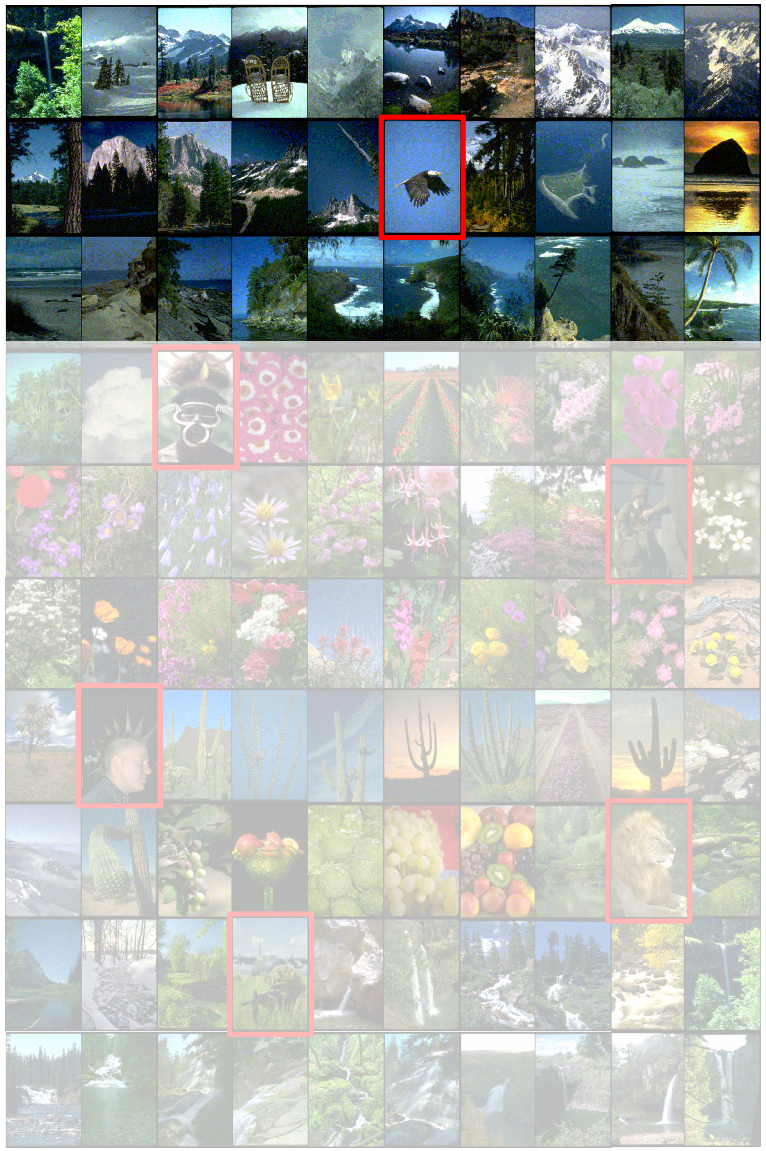
The Intention Readout Experiment



Database (any target that may interest users)

User thinks about what he/she wants to search

The Paradigm

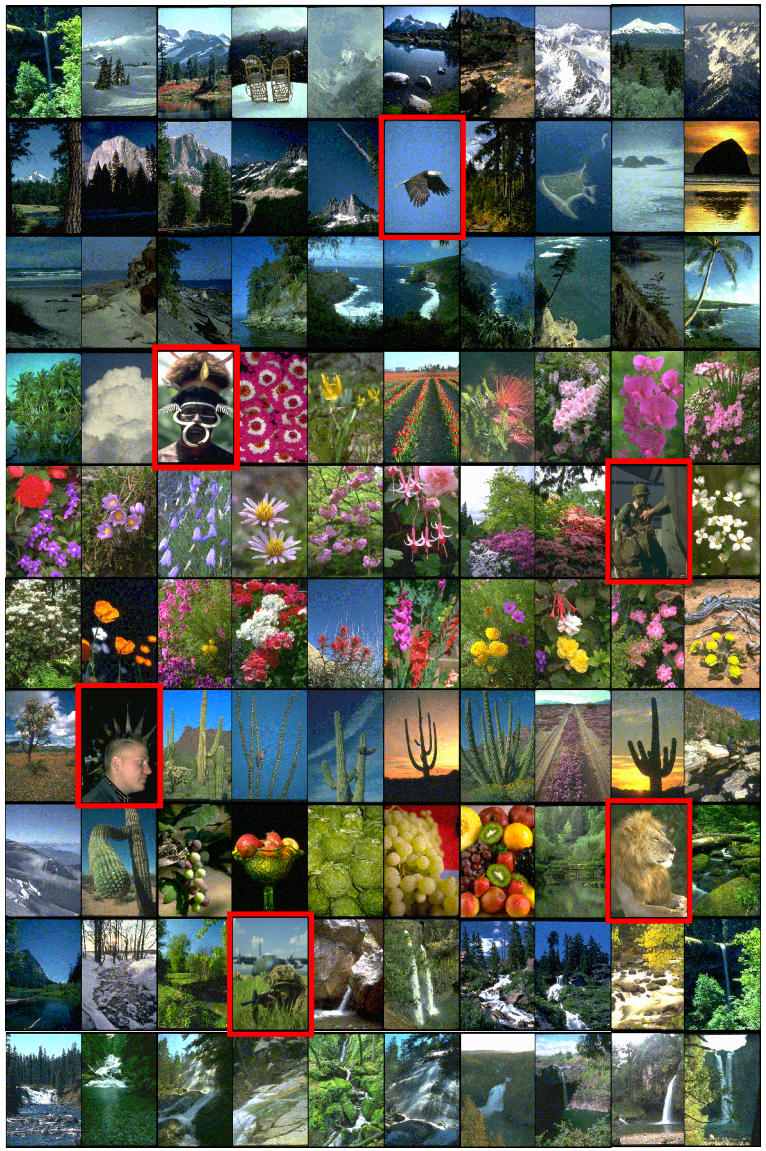


Database

Neural (EEG) decoder

Interest-scores

The Paradigm



Database

Neural (EEG) decoder

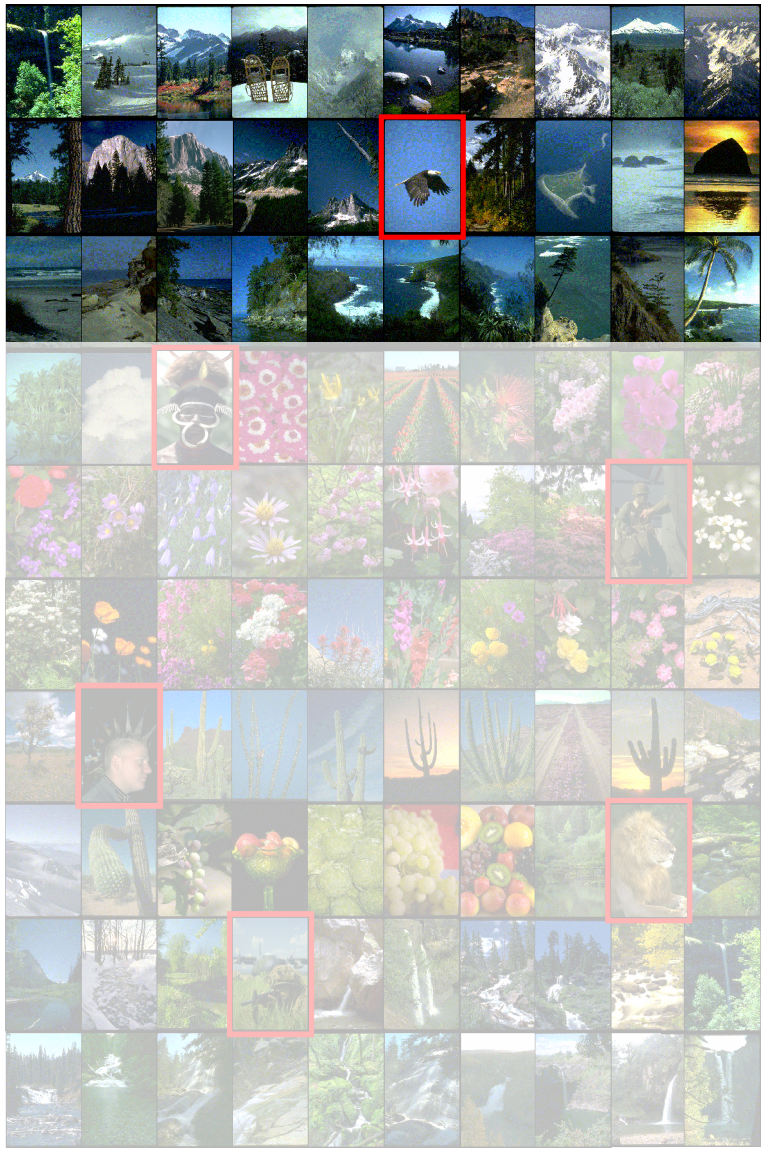
Exemplar labels (noisy)

Semi-supervised computer vision retrieval

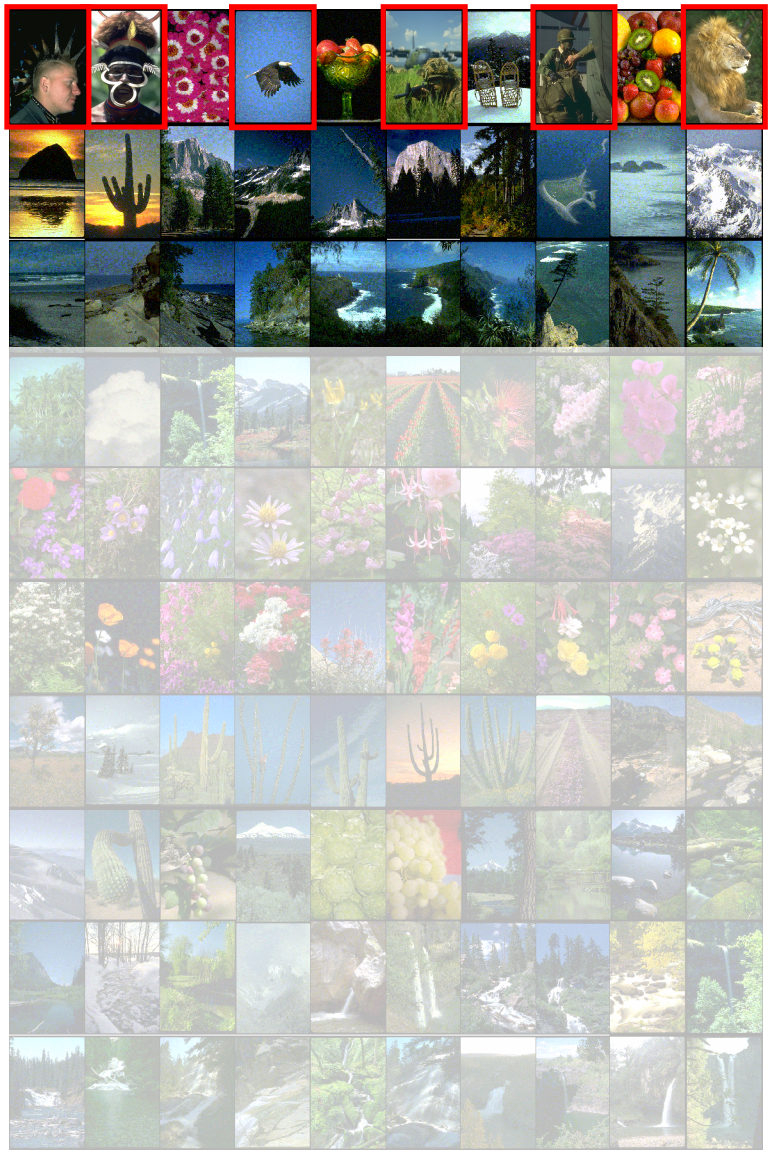
Features from the entire DB

prediction score

The Paradigm

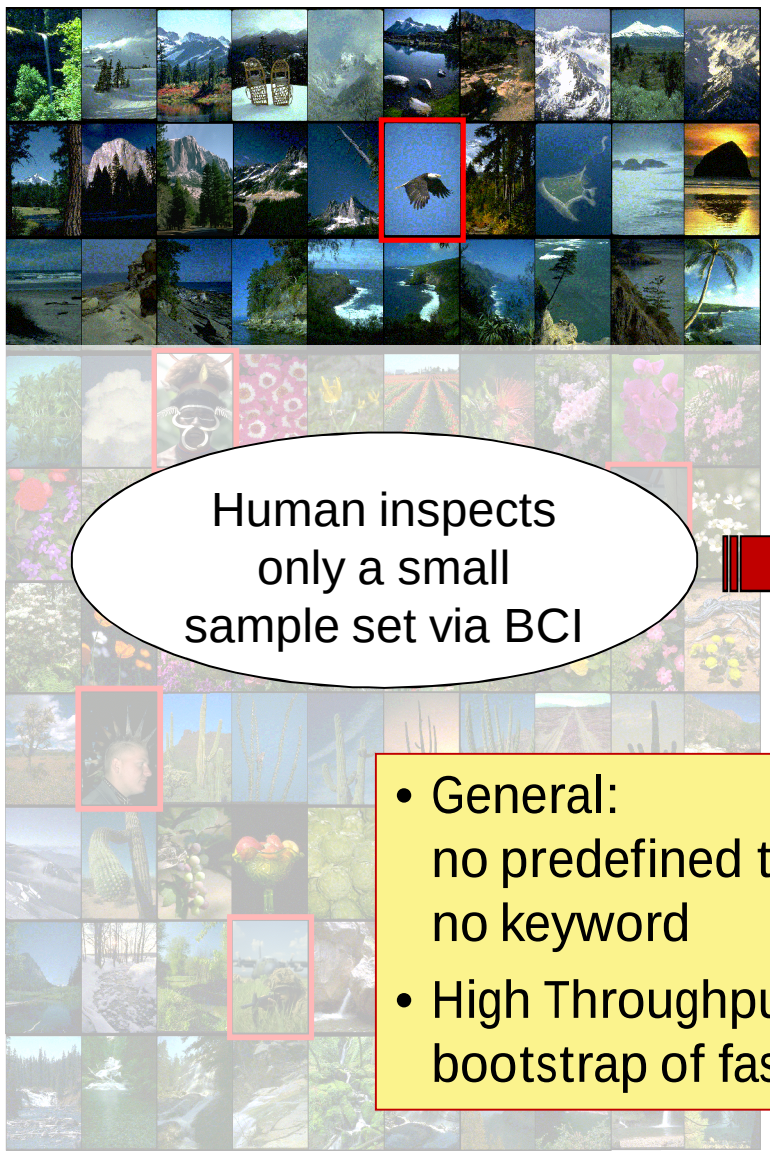


Pre-triage

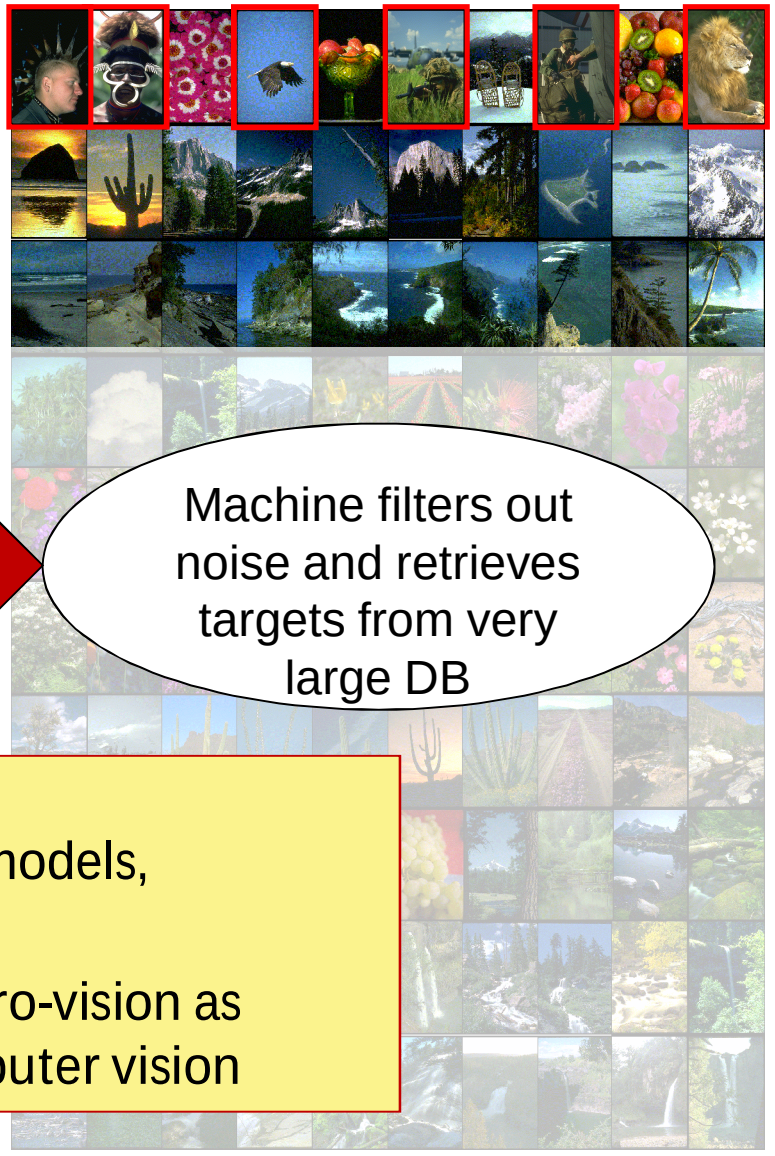


Post-triage

The Paradigm



Human inspects only a small sample set via BCI



Machine filters out noise and retrieves targets from very large DB

- General: no predefined target models, no keyword
- High Throughput: neuro-vision as bootstrap of fast computer vision

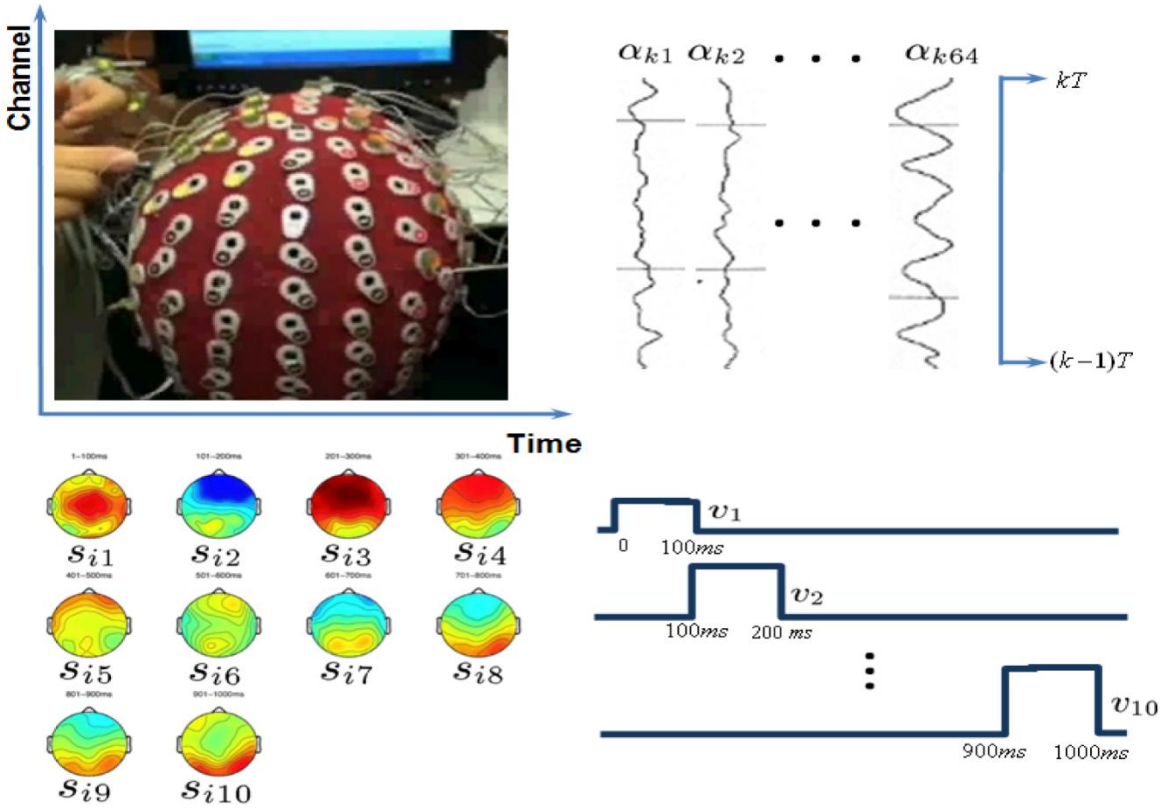
Pre-triage

Post-triage

Identifying Discriminative Components in the EEG Using Single-Trial Analysis

LDA or Logistic Regression is used to learn the contributions of EEG signal components at different spatial-temporal locations

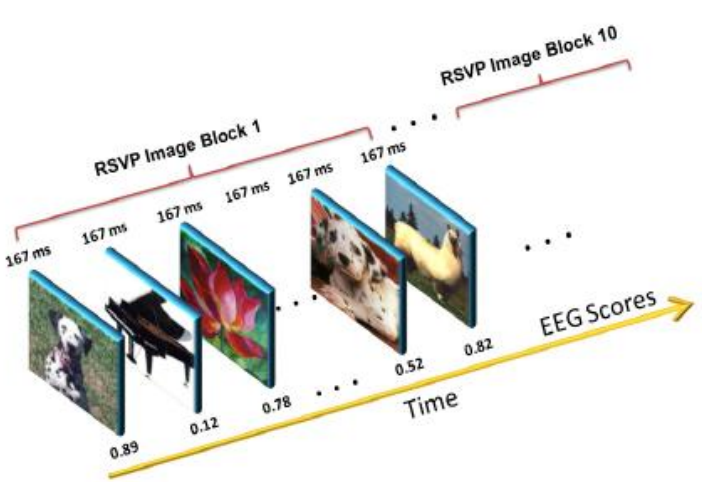
(Parra, Sajda et al. 2002, 2003)



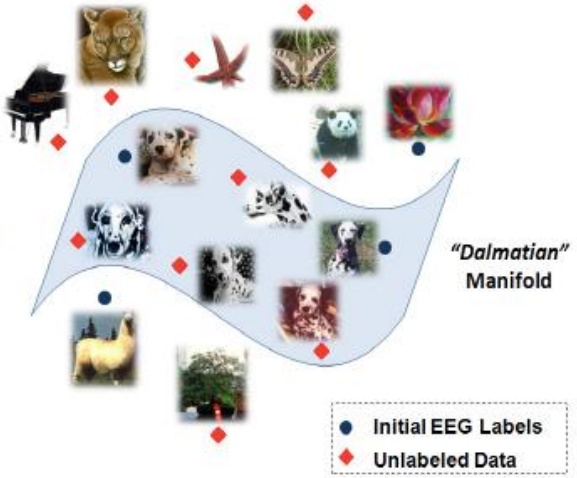
Optimal spatial filtering across electrodes within each short window (e.g., 100ms)

Optimal temporal filtering over time windows after onset

EEG "Feedback" and Manifold Learning



Rapid Serial Presentation of Caltech 101 Images



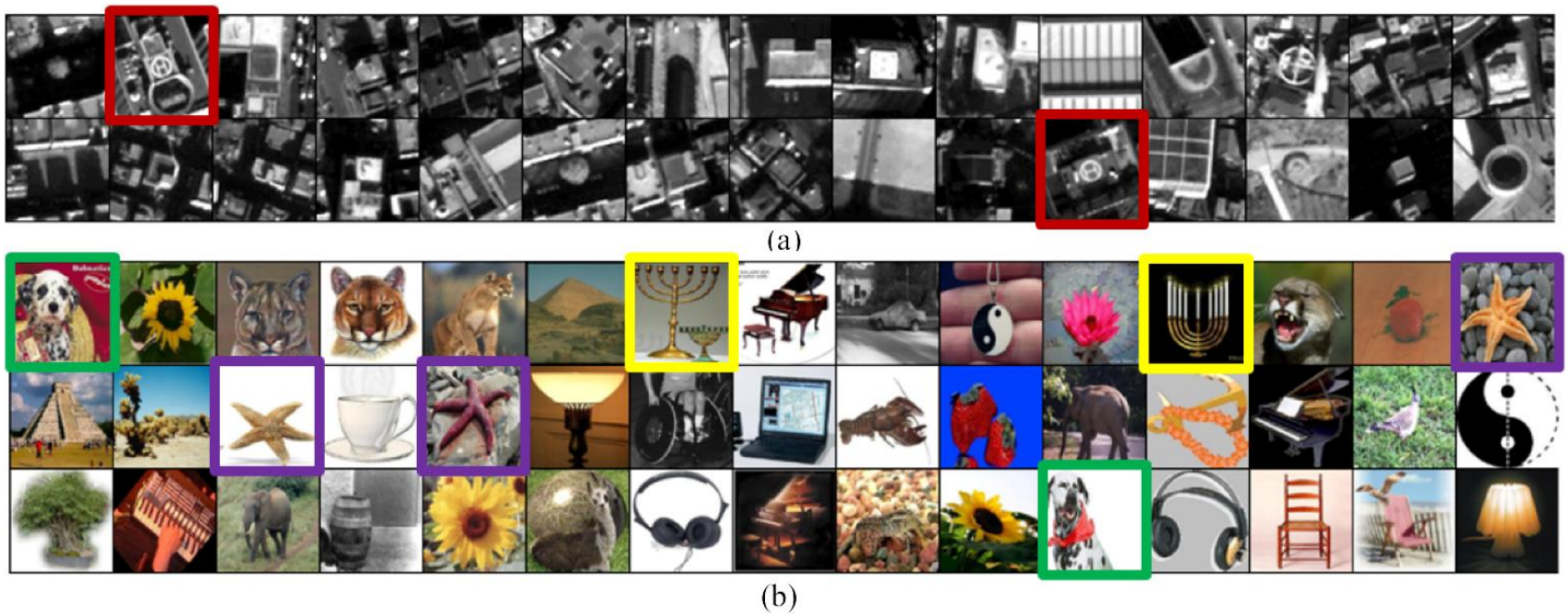
Graph-Based Visual Pattern Discovery

Opportunities and Issues:

- EEG results used as exemplar feedback indicating user interest
- Propagate "interest" scores over manifolds in the image space
- Challenge: EEG labels are noisy and limited
- No prior knowledge about target models

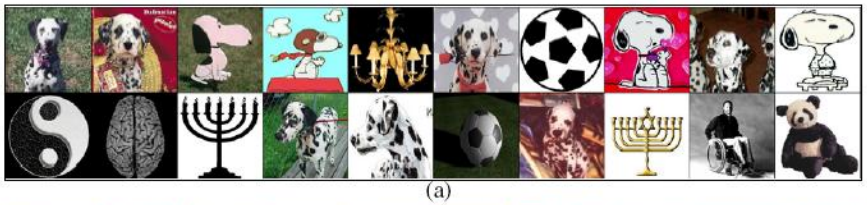
Experiments

- CalTech101: 3798 images from 62 categories
  Satellite images
- Generic neural decoder trained per user using images *(Soccer Ball* or *Baseball Gloves)* from Caltech256
- A subset images randomly sampled to construct 6-Hz RSVP sequence
- Initial Trials: 4 subjects, 3 targets (*Dalmatian*, *Chandelier/Menorah, & Starfish)*

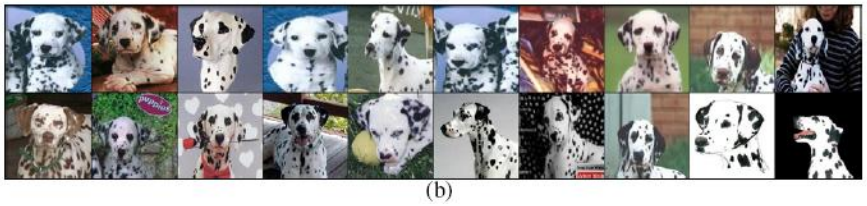


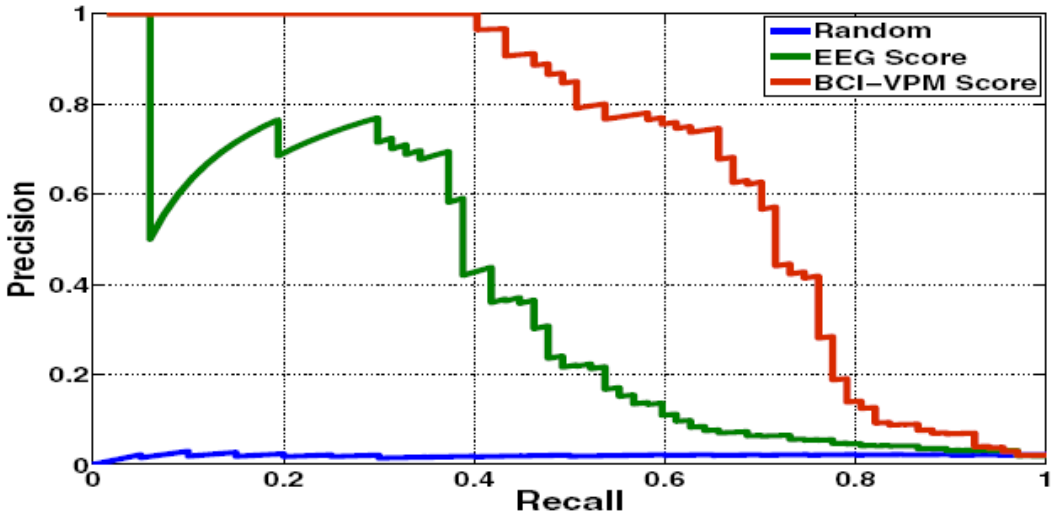
Example results

Top 20 results of
Neural EEG detection

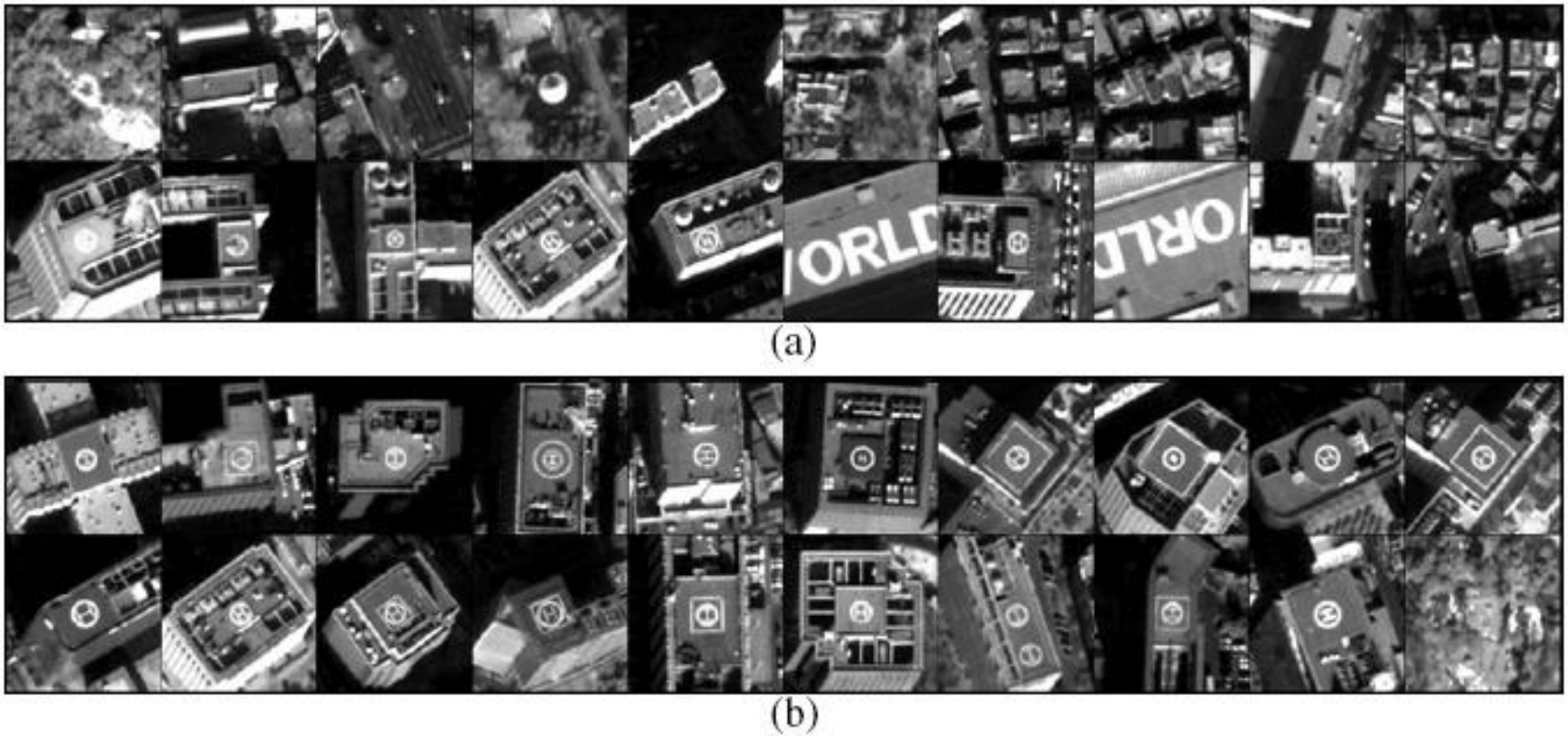


Top 20 results of
Hybrid System (BCI-VPM)





Retrieval on Satellite Imagery



The experimental results of "*helipad" target RSVP, showing the top 20 ranked images* . a) ranking by original EEG scores; b) ranking by the BCI-VPM refined interest score.

Conclusions

- Instant feedback and refinement tools improve the utility of *small imperfect* classifier pools
  - As shown in CuZero search system
- Relevance feedback techniques maximizes the utility of *limited imperfect* user input
  - As shown in graph-based propagation
- Other forms of relevance feedback
  - Pseudo feedback from Web search
  - Feedback from neuro state decoding

References

(many papers can be found at http://www.ee.columbia.edu/dvmm)

- E. Zavesky, S.-F. Chang, "**CuZero**: Embracing the Frontier of Interactive Visual Search for Informed Users," ACM Multimedia Information Retrieval, 2008.
- Y.-G. Jiang, A.Yanagawa, S.-F.Chang, C.-W.Ngo, "**CU-VIREO374**: Fusing Columbia374 and VIREO374 for Large Scale Semantic Concept Detection," ADVENT Technical Report #223-2008-1 Columbia University, August 2008. (http://www.ee.columbia.edu/ln/dvmm/CU-VIREO374/)
- J. Wang, E. Pohlmeyer, B. Hanna, Y.-G. Jiang, P. Sajda, and S.-F. Chang, "**Brain State Decoding** for Rapid Image Retrieval," ACM Multimedia Conference, 2009.
- Jun Wang, Yu-Gang Jiang, Shih-Fu Chang, "**Label Diagnosis** through Self Tuning for Web Image Search," CVPR 2009.
- Jun Wang, Sanjiv Kumar, Shih-Fu Chang, "**Semi-Supervised Hashing** for Scalable Image Retrieval", CVPR 2010.
- Junfeng He, Wei Liu, Shih-Fu Chang, "Scalable Similarity Search with **Optimized Kernel Hashing**," ACM SIGKDD, 2010.
- Wei Liu, Junfeng He, Shih-Fu Chang, "**Large Graph Construction** for Scalable Semi-Supervised Learning," ICML 2010.