



ACM ICMR 2014

Event-Driven Semantic Concept Discovery by Exploiting Weakly Tagged Internet Images

Jiawei Chen, Yin Cui, Guangnan Ye, Dong Liu, Shih-Fu Chang

Columbia University

April 2, 2014

Acknowledgments: This work was supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center contract number D11PC20071. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either express or implied, of IARPA, Dol/NBC, or the U.S. Government.

What are events?

- Examples:

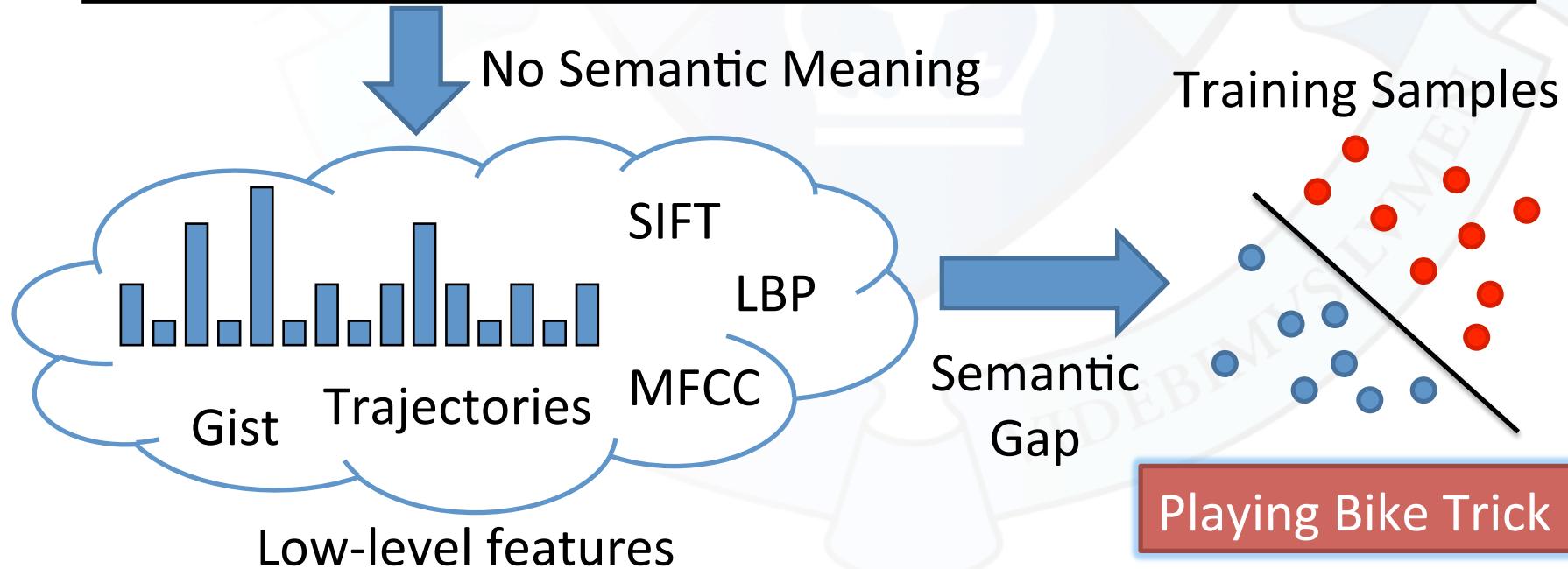


Image credits: CNN, Wikipedia

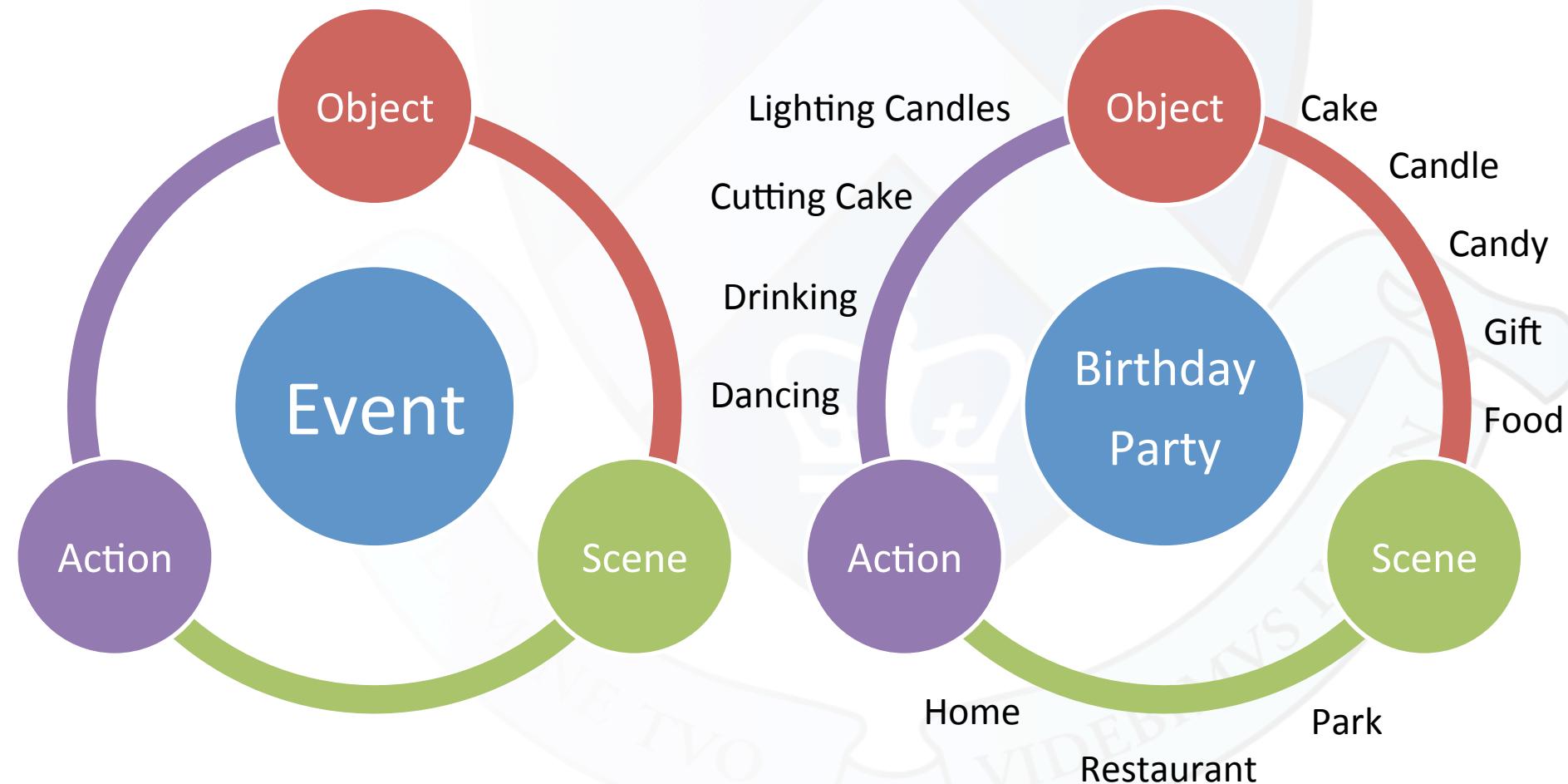
- Layman perspective:
 - People interact with other people and/or objects under certain scenes in a time period.

How to represent an event?

- In Feature Space

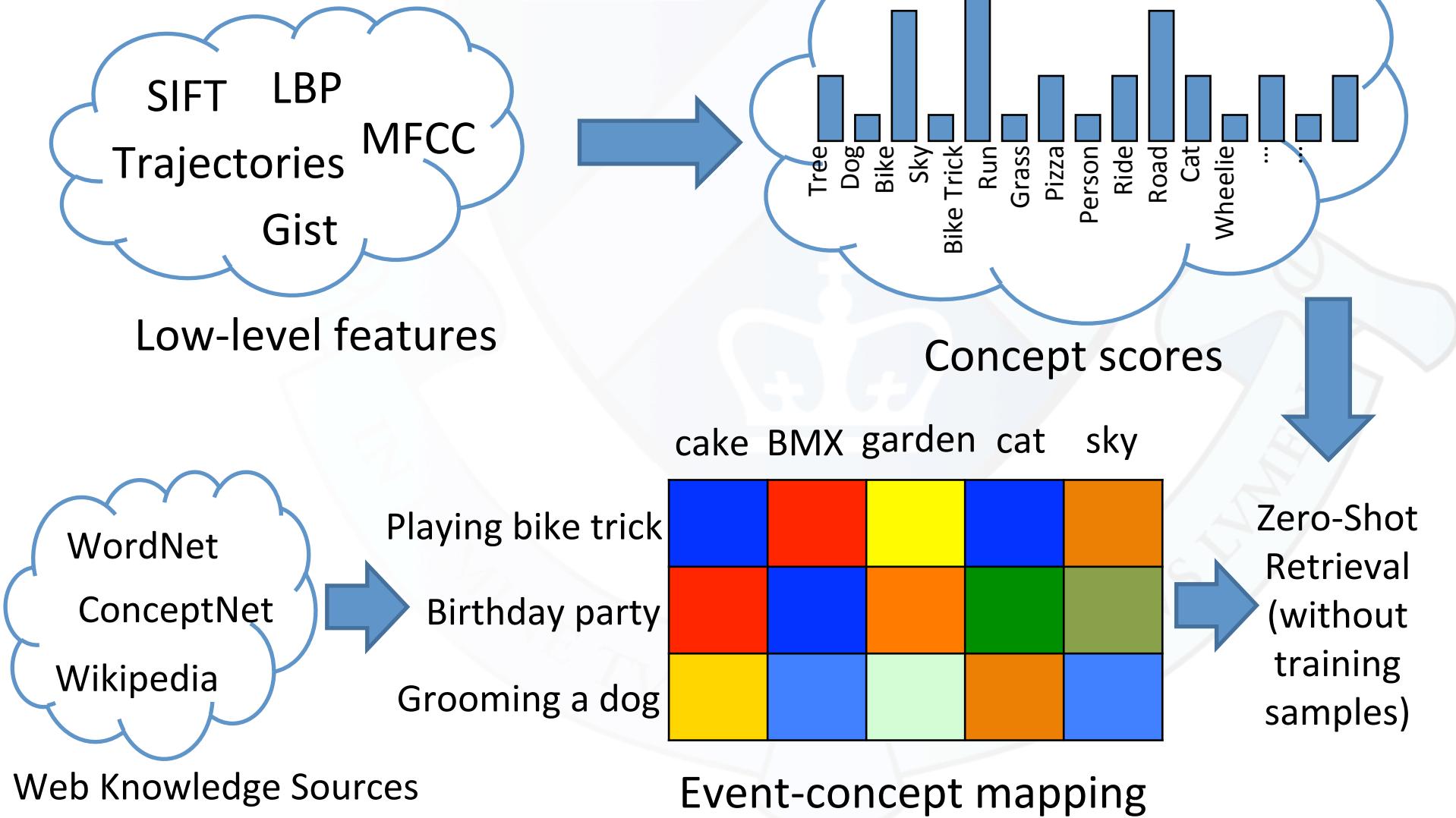


Decompose an event into concepts



How to represent an event?

- In Concept Space



Literature Review

- Representing event in feature space:
 - Many works have been done.
 - Cross-domain [Duan et al. 2010], latent temporal structure [Tang et al. 2012], Multimodal feature fusion [Natarajan et al. 2012], Deep Learning [Yang et al. 2012].
- Representing event in concept space:
 - Manually select concepts for each event [Izadinia et al. 2012, Liu et al. 2012].
 - Pre-built “concept” library not designed for events
 - LSCOM [Naphade et al. 2006], Classemes [Torresani et al. 2010], ImageNet [Deng et al. 2009], SUN Database [Xiao et al. 2010].

How to discover concepts and build
corresponding concept classifiers for each
event automatically?

Some Facts

- People like to share events by photos.
- Tags of an image reveal semantics in the image.
- There are more than 5 billion photos on Flickr.
- Every Flickr image has 8 tags in average.

To discover concepts for each event from a large tagged image collection



Cake, Birthday Party, Dinosaur,
Volcano, Gift, Boy



Cake,
Birthday,
Candles,
Lighting,
Shine,
Girl,
Grandchild,
Surprise

Overview of the proposed system

“grooming an animal”



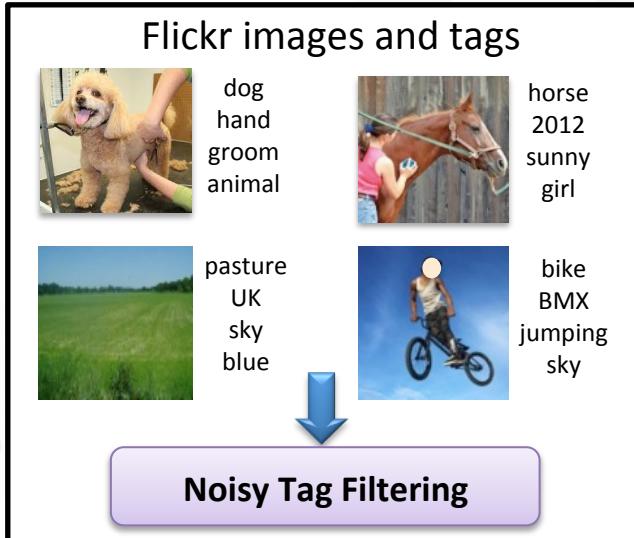
Keywords Extraction from Textual Event Definition



Keywords Combinations



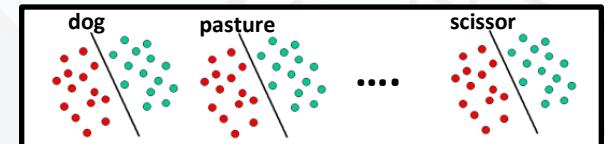
Query (Noun-Verb Pairs)
E.g. bike + jump



Concepts

KDE noisy image removal:
Sample training images from dense regions in feature space

Concept classifier training
After KDE



Concept Classifiers

Applications

Supervised Event Modeling

Zero-Shot Event Retrieval

Video Semantic Recounting

Concept Based Video Representation



We need data

- Step I: Flickr Image Crawling and Tag Extraction
 - 2013 TRECVID Multimedia Event Detection (MED) 20 pre-specified events (birthday party, parade, grooming a dog, attempting bike trick, etc.).
 - 20,000 Flickr images for each of 20 events.
 - Around 2,500,000 tags associated with images in total.
 - Under each event, rank their tags by TF-IDF value.

We need reliable tags

- Step II: Noisy Tag Filtering
 - Exclude non-meaningful tags (i.e., tags containing words outside WordNet)

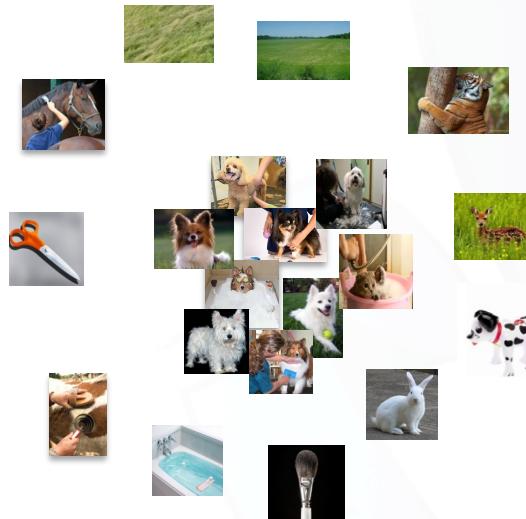


Candle
Charlie
Birthday Cake
Cake
Pets
Dogs
Cat
Party
Canon EOS
2011
Celebration
100Pictures

Concept images are noisy

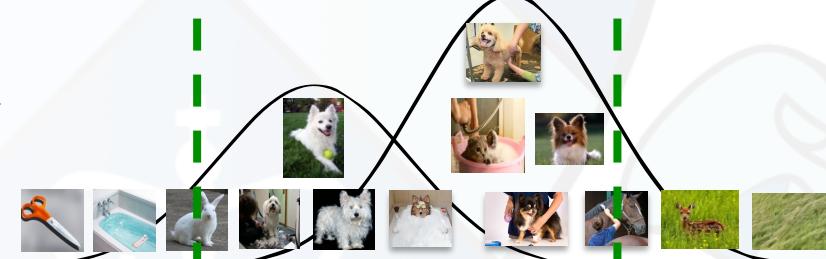
- Step III: Training Image Selection

Images annotated with “dog”



Kernel Density Estimation (KDE)

Outliers | Dense Regions | Outliers



Sampling from the dense region in the feature space.

Rank all images based on their confidence scores derived from *KDE*.



clean dog images

Noisy images or outliers

SVM with RBF kernel

Concept classifier for dog

Event-specific concepts we discovered

Event: Attempting a bike trick

Size of each concept indicates its TF-IDF value.
Top 30 concepts are shown here.

Analysis & Applications

Why our concepts are good?

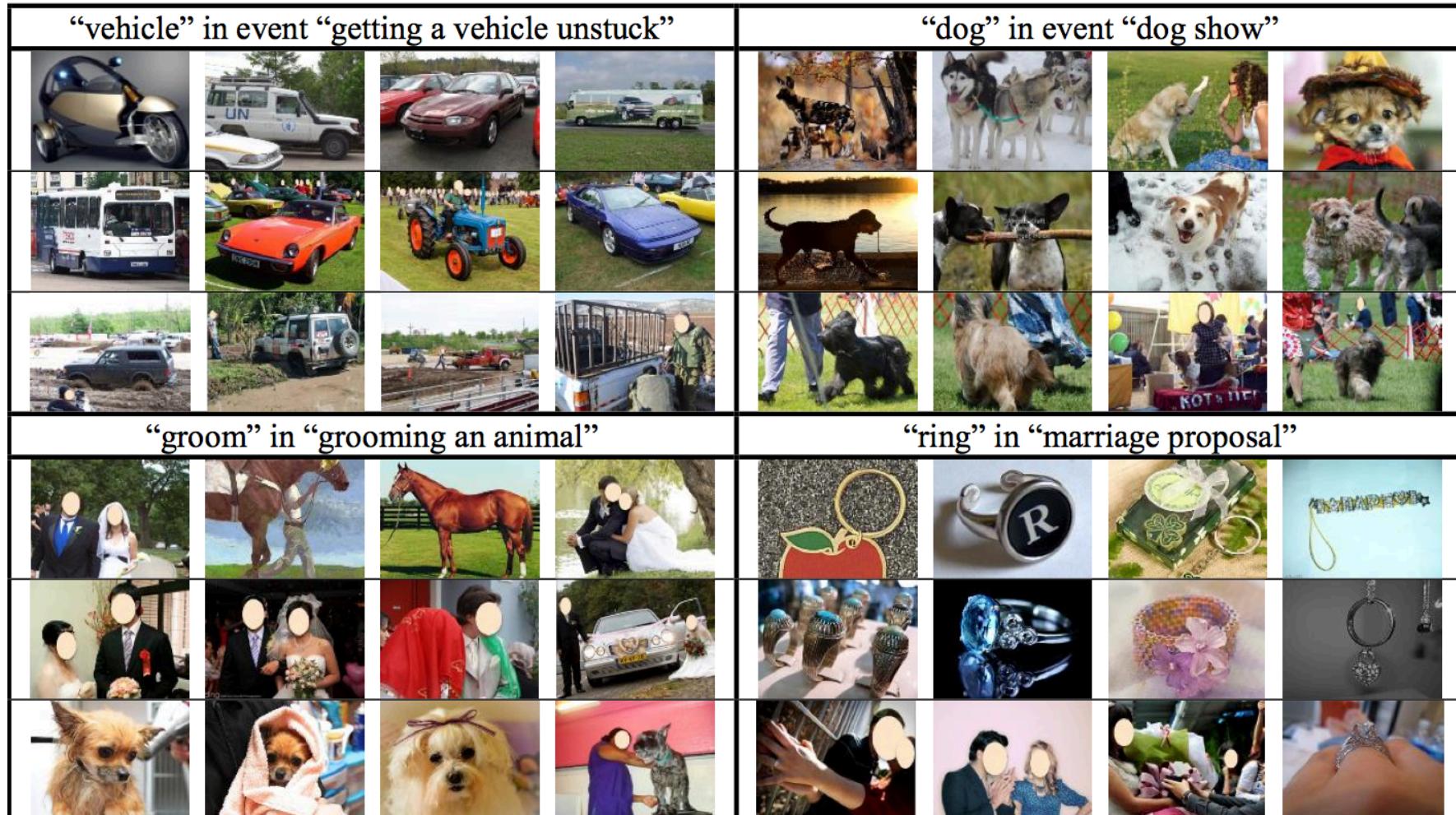
- Concepts discovered from Flickr are more semantic relevant to the event

Event Name	Concepts Discovered from Different Resources	
getting a vehicle unstuck	Classemes	air transportation vehicle, all terrain vehicle, amphibious vehicle, armed person, armored fighting vehicle, armored recovery vehicle, armored vehicle, armored vehicle heavy, armored vehicle light, command vehicle
	ImageNet	vehicle, bumper car, craft, military vehicle, rocket, skibob, sled, steamroller, wheeled vehicle, conveyance
	Ours	tire, car, snow, stick, stuck, winter, vehicle, truck, night, blizzard
grooming an animal	Classemes	adult animal, animal, animal activity, animal blo, animal body part, animal body region, animal cage, animal container, animal pen, animal shelter
	ImageNet	groom, animal, invertebrate, homeotherm, work animal, darter, range animal, creepy-crawly, domestic animal, molter
	Ours	dog, pet, grooming, cat, animal, bath, cute, canine, puppy, water
making a sandwich	Classemes	baking dish, cafe place, classroom setting, collection display setting, cutting device, dish drying rack, food utensil, hair cutting razor, hdtv set, hole making tool
	ImageNet	sandwich, open-face sandwich, butty, reuben, ham sandwich, gyro, chicken sandwich, hotdog, club sandwich, wrap
	Ours	sandwich, food, bread, cooking, cheese, spice, baking, pan, kitchen, breakfast

Top 10 concepts discovered from different resources

Why our concepts are good?

- Concept training images reveal context information of the concept

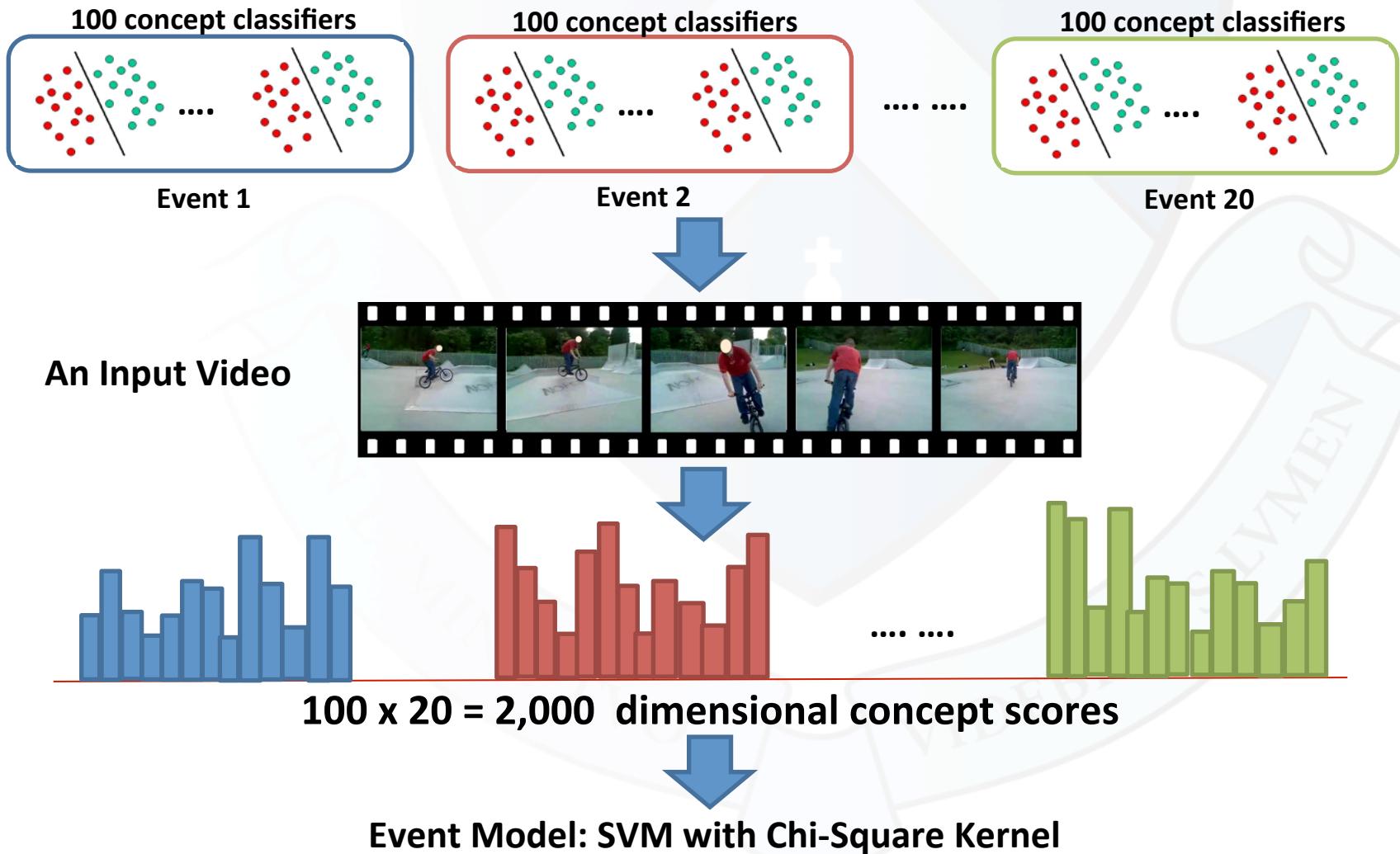


CC: Clasemes

IN: ImageNet

I: Supervised Event Modeling

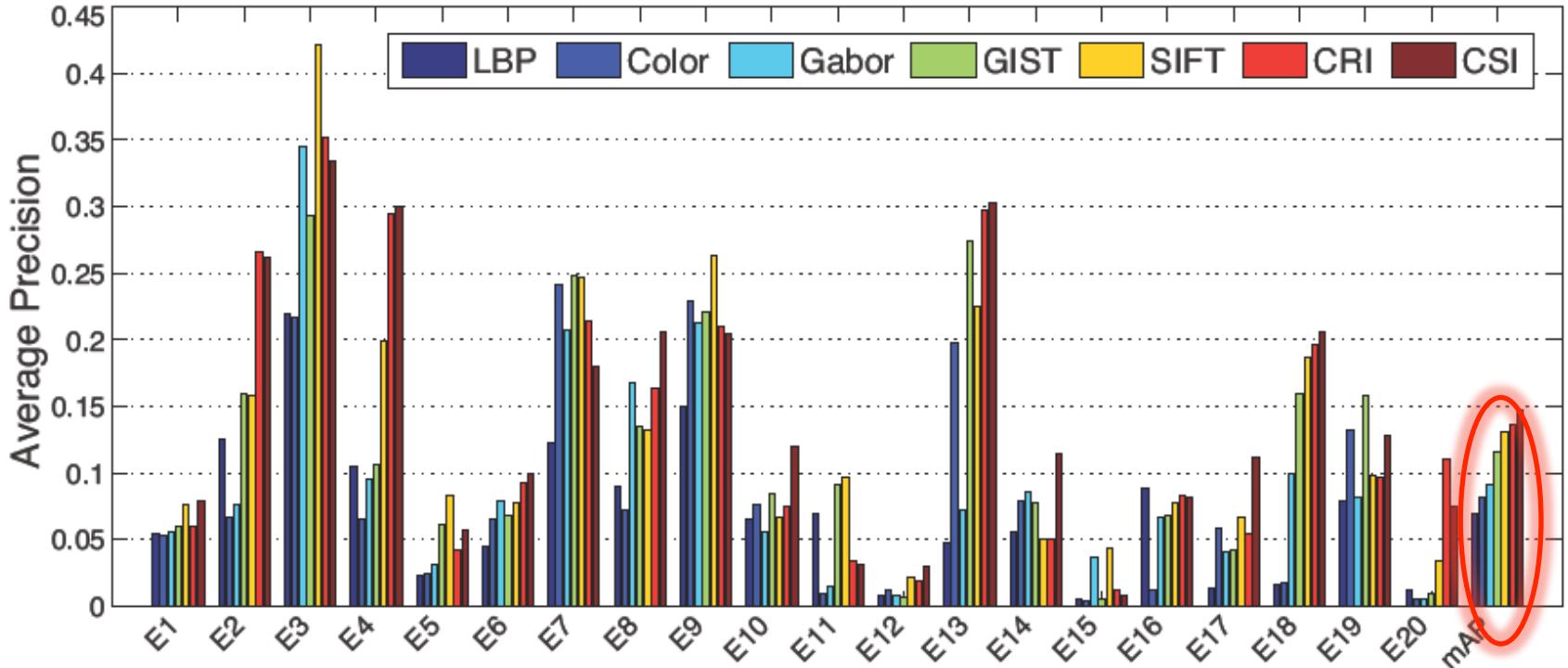
- Use concept scores as video feature representation
 - For each event, use the discovered 2,000 concepts (100 for each event).



Experiment Setup

- 20 Pre-specified events in TRECVID MED
 - 100 positive videos for each of the 20 events.
 - Use 2,000 dimensional concept score vector as feature representation.
 - Evaluation metric : full-length Mean Average Precision (MAP).

Performance comparison



Different low-level features: LBP, Color, Gabor, GIST, SIFT (2,000 dimension, same as ours).

CRI : Concept classifier with Randomly subsampled concept Images (without KDE).

CSI : Concept classifier with Selected representative Images by KDE.

- Our CSI concepts outperform other low-level features.
- Our CSI concepts beat the CRI method since it uses more reliable training images.
- Random Guess mAP: 0.003

E1: birthday party, E2: changing a vehicle tire, E3: flash mob gathering, E4: getting a vehicle unstuck, E5: grooming an animal, E6: making a sandwich, E7: parade, E8: parkour, E9: repair an appliance, E10: working on a sewing project, E11: attempting a bike trick, E12: cleaning an appliance, E13: dog show, E14: giving directions to a location, E15: marriage proposal, E16: renovating a home, E17: rock climbing, E18: town hall meeting, E19: winning a race without a vehicle, E20: working on a metal crafts project

II: Zero-Shot Event Retrieval

- Given an event name as textual query without any training videos, rank all videos.
 - Different concepts have different weights:
 - “cake” is more relevant and useful than “home ”in detecting “birthday party”.
 - Give higher weights to more relevant concepts.

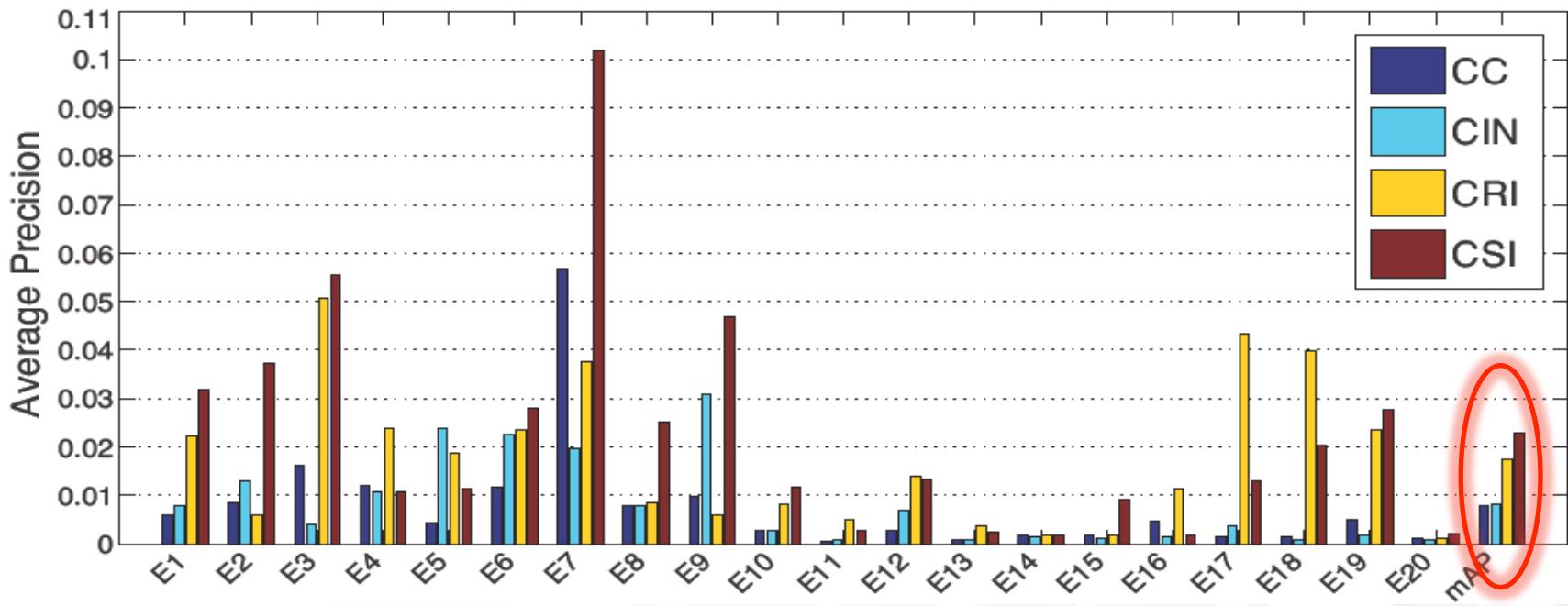
$$RankingScore = \sum_{i=1}^T sim(concept_i, Event) \times score_i$$

T : # of concepts discovered for this event,
we set $T=100$ in our work.

Experiment Setup

- 20 Pre-specified MED events
 - No training video for each event.
 - Use 100 dimension concept score vector for each event as feature representation.
 - Evaluation metric : full-length Mean Average Precision (MAP).

Performance Comparison



CC: Classeme concepts (top 100 most relevant concepts computed by WordNet similarity).

CIN: ImageNet concepts (top 100 most relevant concepts computed by WordNet similarity).

CRI: Concept learned with Randomly subsampled concept Images (without KDE).

CSI: Concept learned with Selected Representative images from KDE.

- Our CSI concepts have a relative performance gain as high as 228%.
- Our CSI concepts outperform the existing concept library Classemes.
- Our CSI concepts beat CIN method which discovers concepts from ImageNet.
- Random Guess mAP: 0.003

E1: birthday party, E2: changing a vehicle tire, E3: flash mob gathering, E4: getting a vehicle unstuck, E5: grooming an animal, E6: making a sandwich, E7: parade, E8: parkour, E9: repair an appliance, E10: working on a sewing project, E11: attempting a bike trick, E12: cleaning an appliance, E13: dog show, E14: giving directions to a location, E15: marriage proposal, E16: renovating a home, E17: rock climbing, E18: town hall meeting, E19: winning a race without a vehicle, E20: working on a metal crafts project

III: Video Semantic Recounting

Top ranked concepts are considered as the recounting concepts for each video.

Event Exemplar 1: Working on a sewing project



Top Discovered Concepts for MER:

“woman”, “hand”, “machine”, “fabric”, “place”, “girl”, “work”, “sewing”,...

Event Exemplar 2: Getting a vehicle unstuck

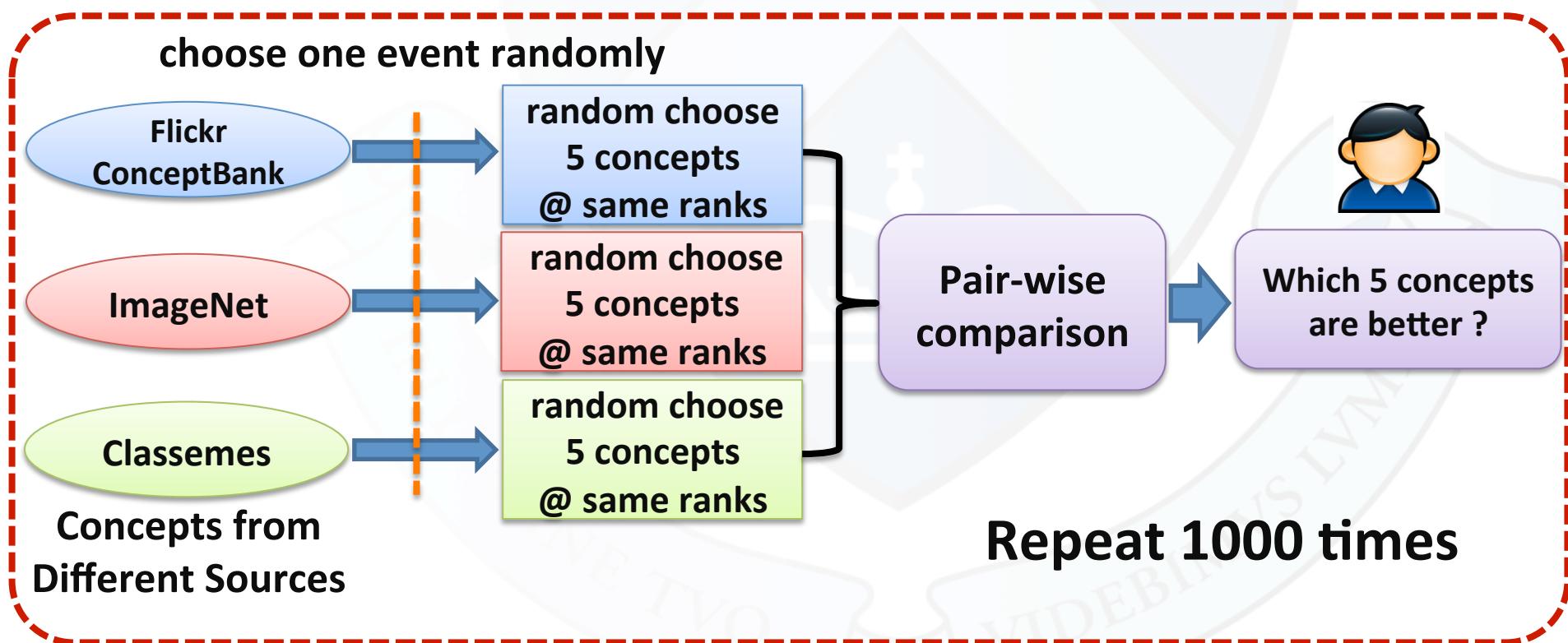


Top Discovered Concepts for MER:

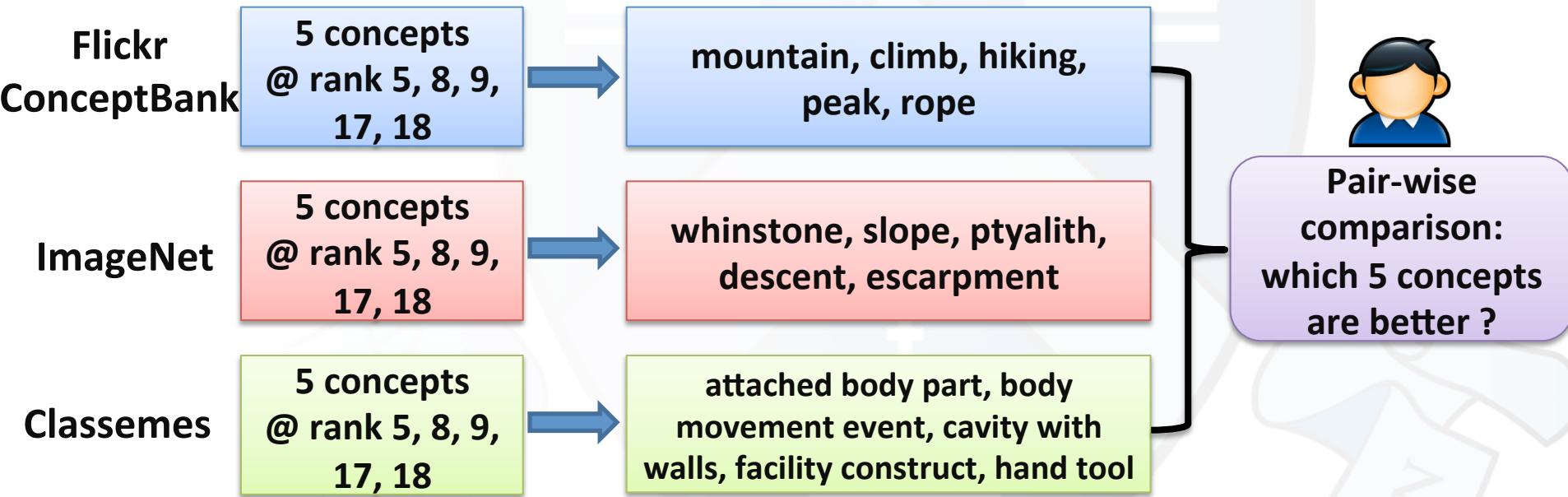
“car”, “snow”, “stuck”, “blizzard”, “winter”, “vehicle”, “truck”, “storm”,...

Human Evaluation

- For each event, choose top 100 concepts based on semantic similarity between concept name and event name.



Human Evaluation: an example



Event: Rock climbing

Result:

- Our Flickr ConceptBank is better than others with **81.29%** chance.
- ImageNet concept is better than others with **34.19%** chance.
- Classemes concept is better than others with **32.46%** chance.

Summary

- Develop a novel systematic framework for automatic concept discovery.
 - 2,000 concepts easily discovered for 20 events.
- Demonstrate significant performance improvements.
 - Outperforms 5 commonly used low-level features (SIFT, LBP, Gabor, Gist, Color-histogram).
 - Relatively 228% improvement over concepts from ImageNet and Classemes in zero-shot retrieval.
- Successfully reveals the underlying semantic clues in videos.
- 2,000 concept classifiers will be available to public soon.

Ongoing Work

- Expand the concept library beyond 20 MED events:
 - Mine 631 events from Wikihow crowdsourcing website.
 - Build 4,876 event-specific concept classifiers for all events.
 - For novel events, use concepts from similar events.

Events	Bank of Event-Specific Concept Classifiers								
<i>Surfing</i>	Surfboard		Beach		Jumping		...	Wave	
<i>Bicycling</i>	Mountain Bike		Jumping		Wheelie		...	Road	
<i>Dog Grooming</i>	Bath		Brush		Ear		...	Tools	

- In arXiv:

Yin Cui, Dong Liu, Jiawei Chen, Shih-Fu Chang
Building A Large Concept Bank for Representing Events in Video