Direct or Indirect Match?
Selecting Right Concepts for Zero-Example Case

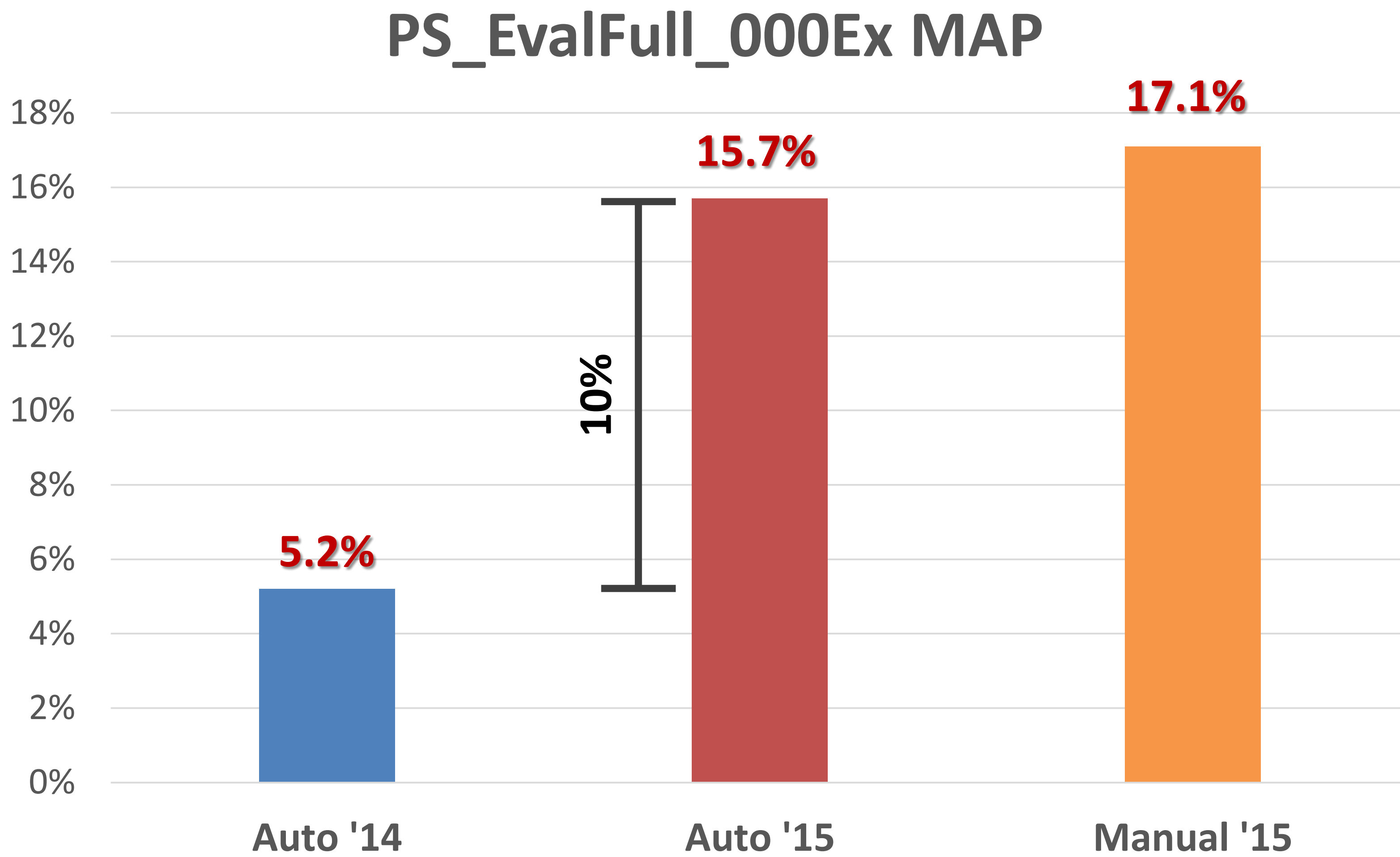
Speaker: Yi-Jie Lu

Yi-Jie Lu[1], Maaike de Boer[2,3], Hao Zhang[1],
Klamer Schutte[2], Wessel Kraaij[2,3], Chong-Wah Ngo[1]
[1]VIREO Group, City University of Hong Kong, Hong Kong
[2]Netherlands Organization for Applied Scientific Research (TNO), Netherlands
[3]Radboud University, Nijmegen, Netherlands

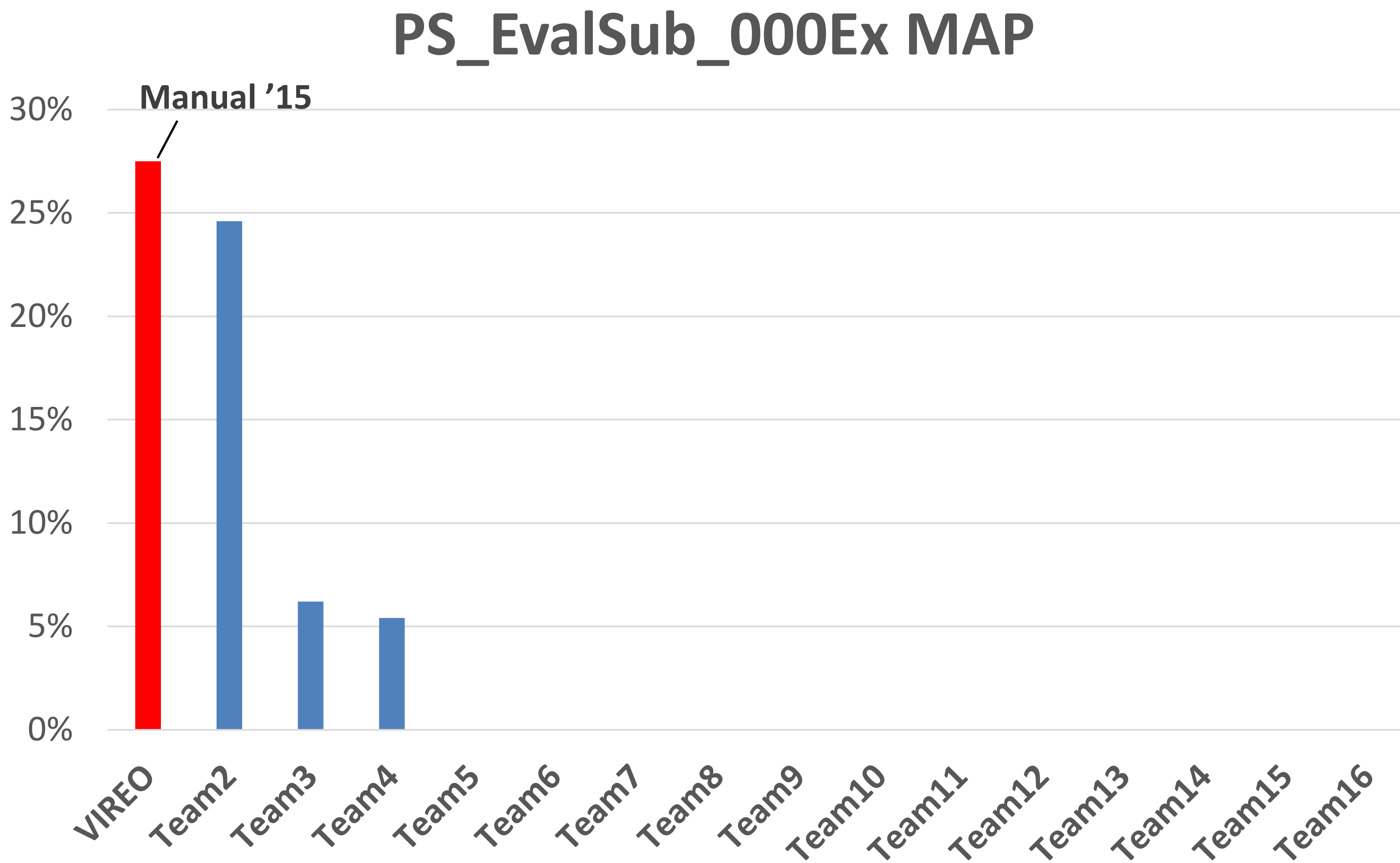
Outline

- **Introduce overall performance in 2015**
- **Difference with 2014 submission**
  - An enlarged concept bank
  - Strategy to pick up the right concepts from concept bank

Achievements in 2015

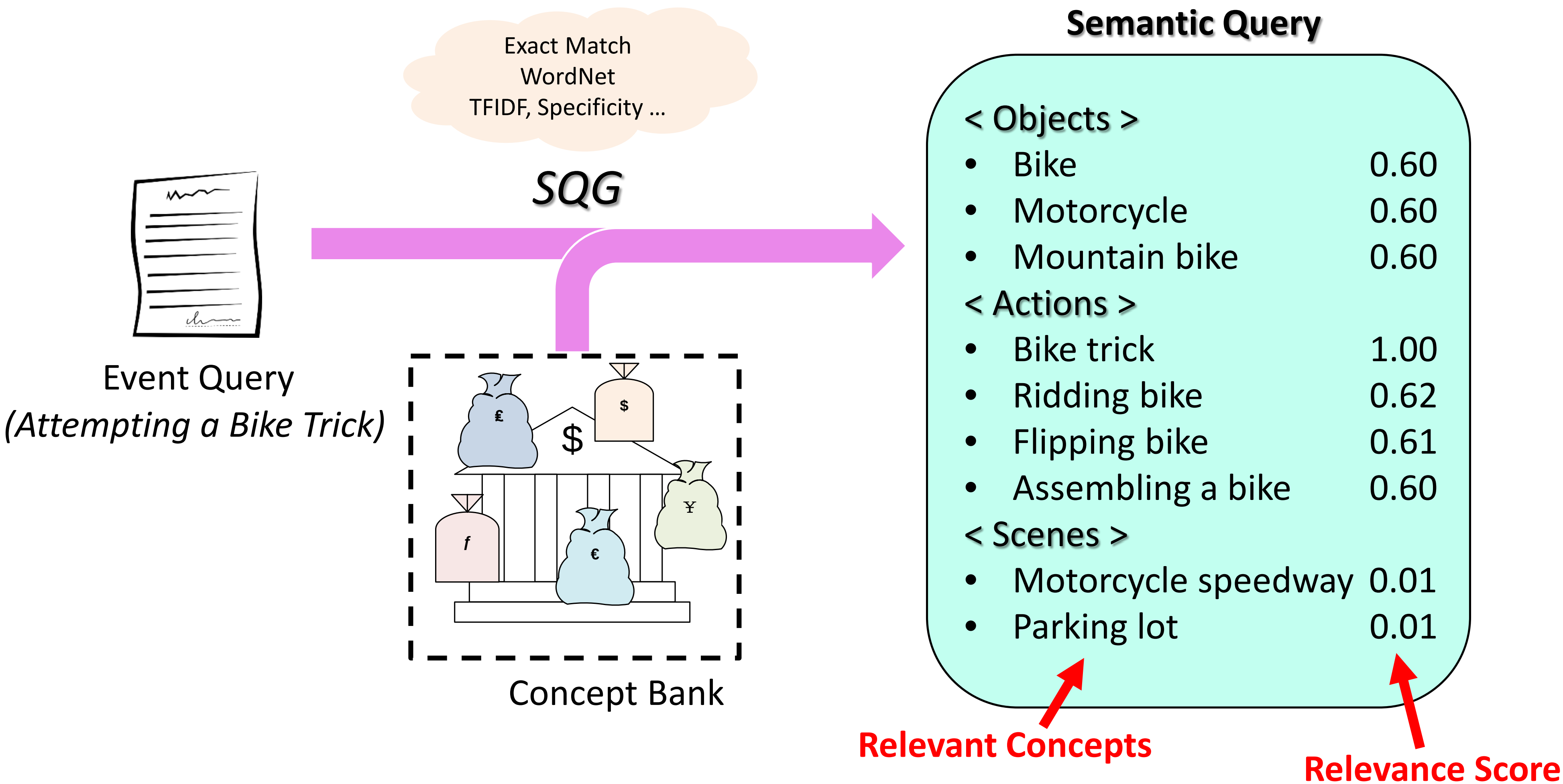


Achievements in 2015



Important changes from ’14?

Recall the Semantic Query Generation (SQG):



Recall our 2014 findings

Missing key concepts

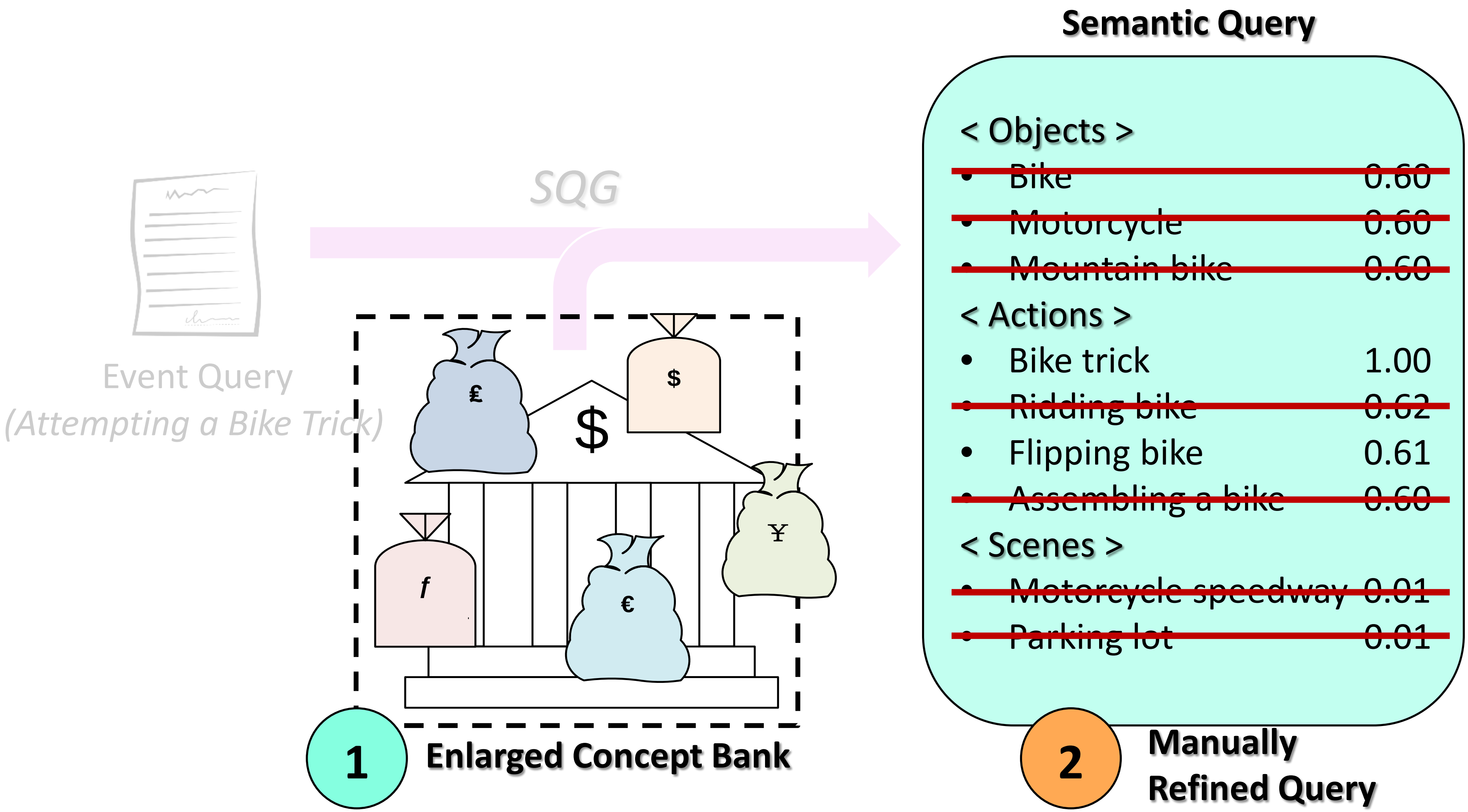
<u>Extinguishing a Fire</u>

[Fire extinguisher]

[Firefighter]

Exact match >> ~~WordNet/ConceptNet~~

- **What we do?**


Semantic Query
< Objects >
~~• Bike 0.60~~
~~• Motorcycle 0.60~~
~~• Mountain bike 0.60~~
< Actions >
• Bike trick 1.00
~~• Ridding bike 0.62~~
• Flipping bike 0.61
~~• Assembling a bike 0.60~~
< Scenes >
~~• Motorcycle speedway 0.01~~
~~• Parking lot 0.01~~
SQG
Event Query
(Attempting a Bike Trick)
$
£
$
f
€
¥
1
Enlarged Concept Bank
2
Manually
Refined Query

Enlarge the concept bank

2014

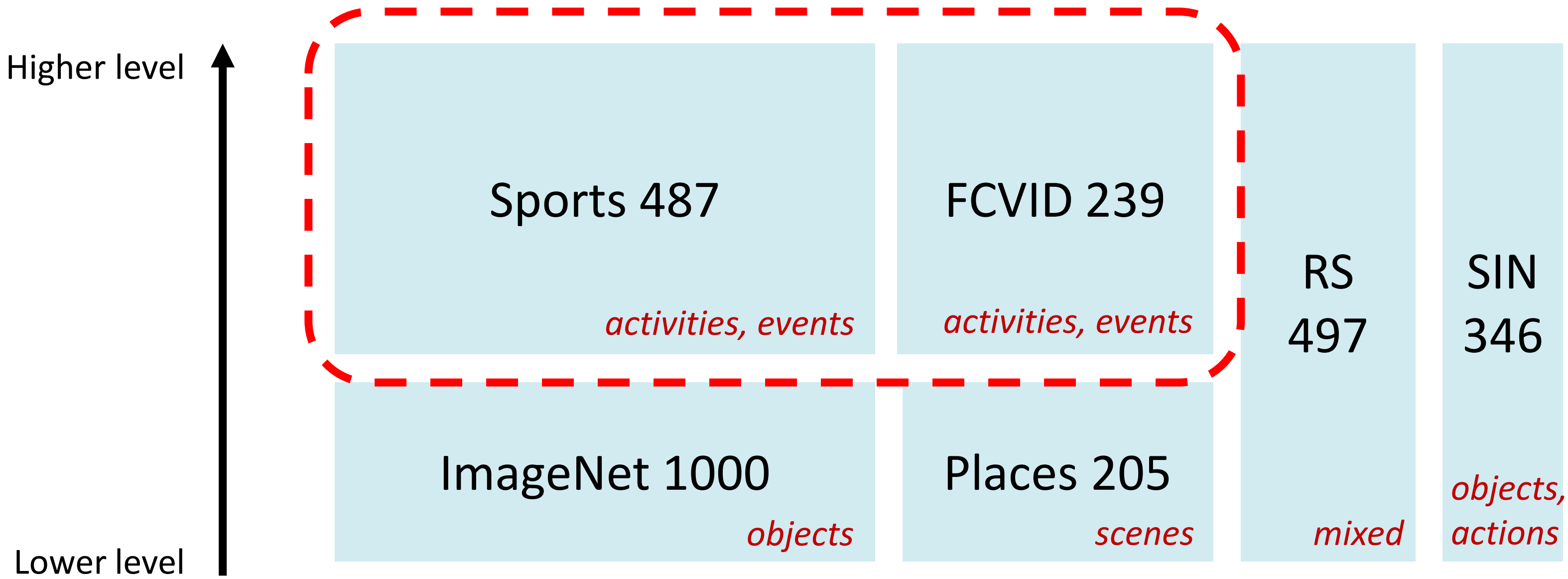
- Research set (497) **CNN**
- ImageNet ILSVRC (1000) **CNN**
- SIN (346) **CNN**

2015

- + Sports (487) **CNN**
- + FCVID (239) **SVM**
- + Places (205) **CNN**

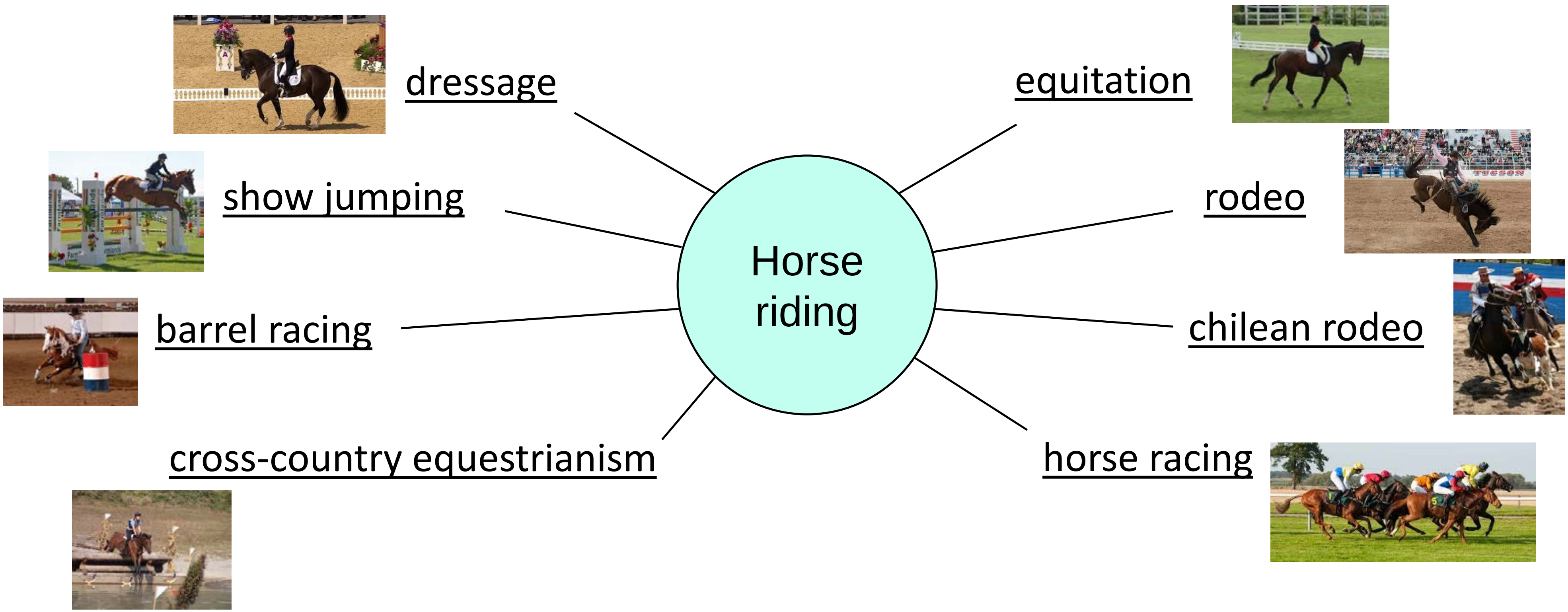
SFRISP (2774)

Concept Bank Review



Concept Bank Review

- **Sports (487) [1]**

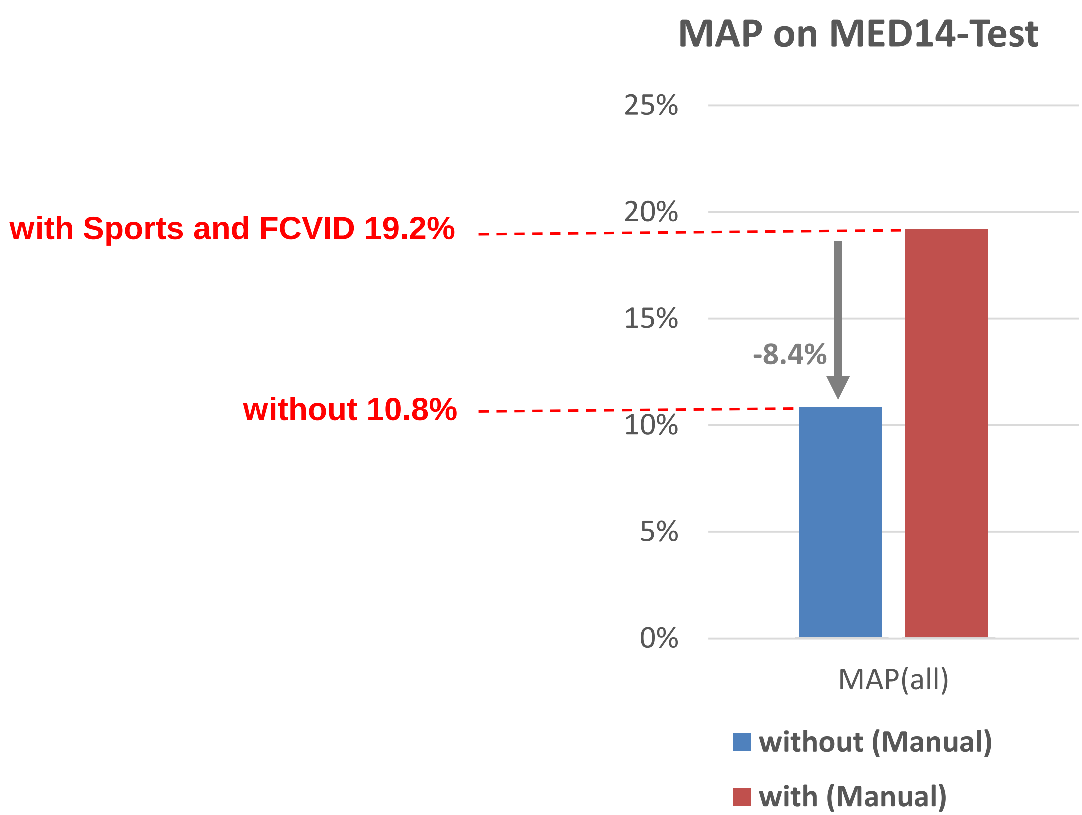


[1] L. Jiang, S.-I. Yu, D. Meng, T. Mitamura, and A. G. Hauptmann, "Bridging the ultimate semantic gap: A semantic search engine for internet videos," in *International Conference on Multimedia Retrieval*, 2015.

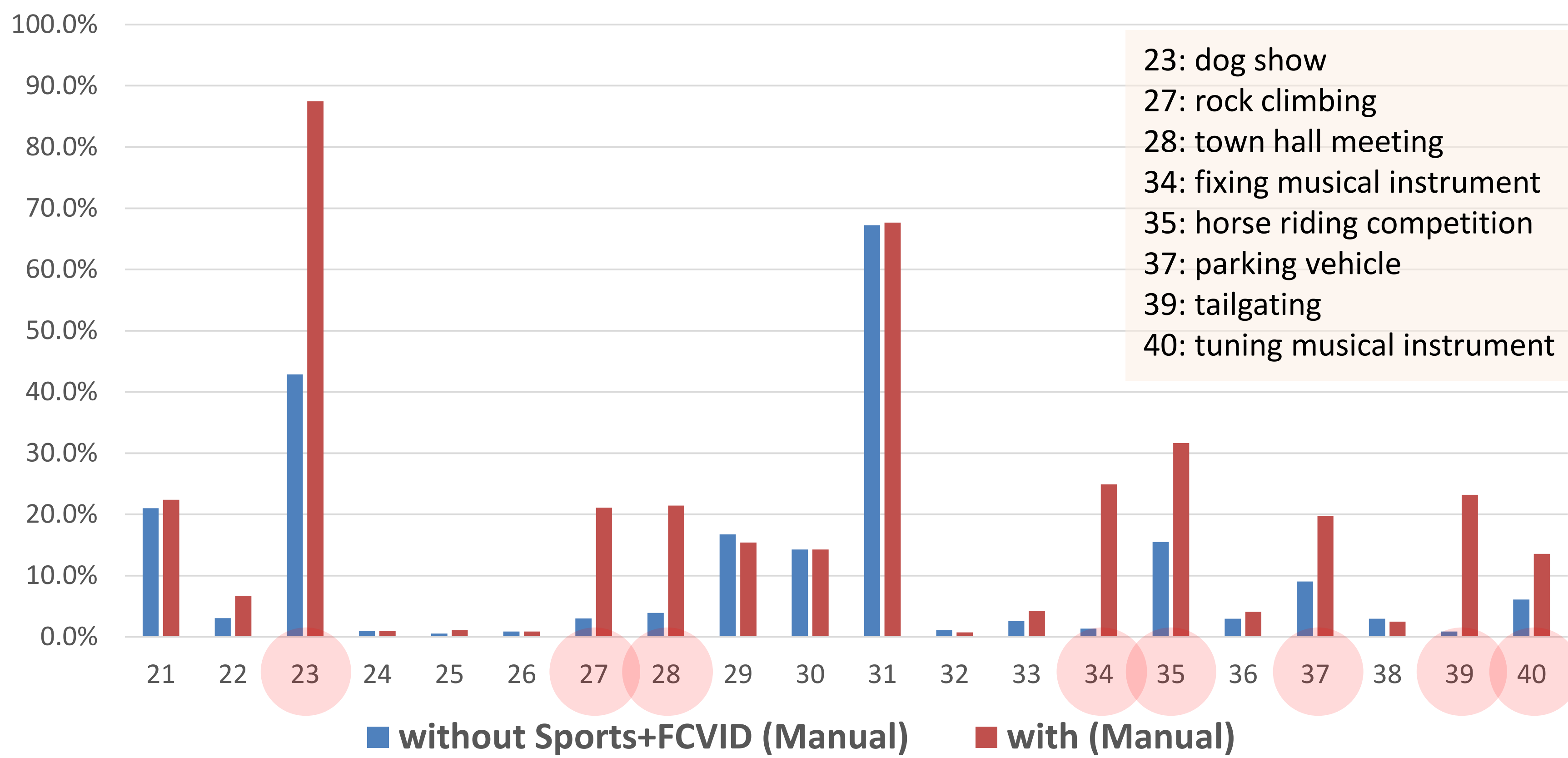
Concept Bank Review

- **FCVID (239)**
  - A large dataset contains high-level activities/events
    - *accordion performance*
    - *American football professional*
    - *bungee jumping*
    - *car accidents*
    - *fire fighting*
    - *playing frisbee with dog*
    - *rock climbing*
    - *wedding ceremony*

Contributions of Sports and FCVID



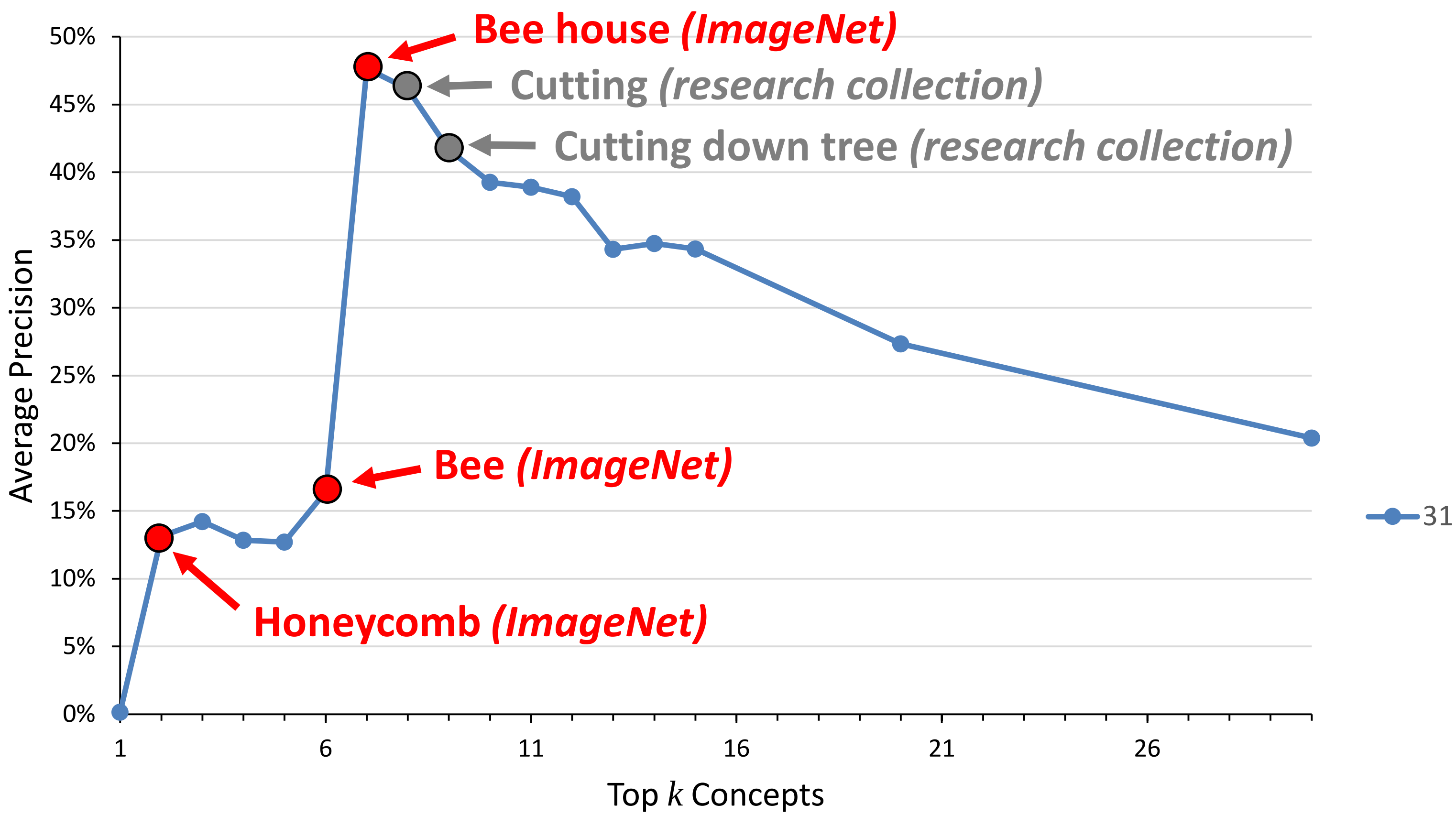
Contribution of Sports+FCVID (726 concepts) on MED14-Test



In combination of 6 different resources:

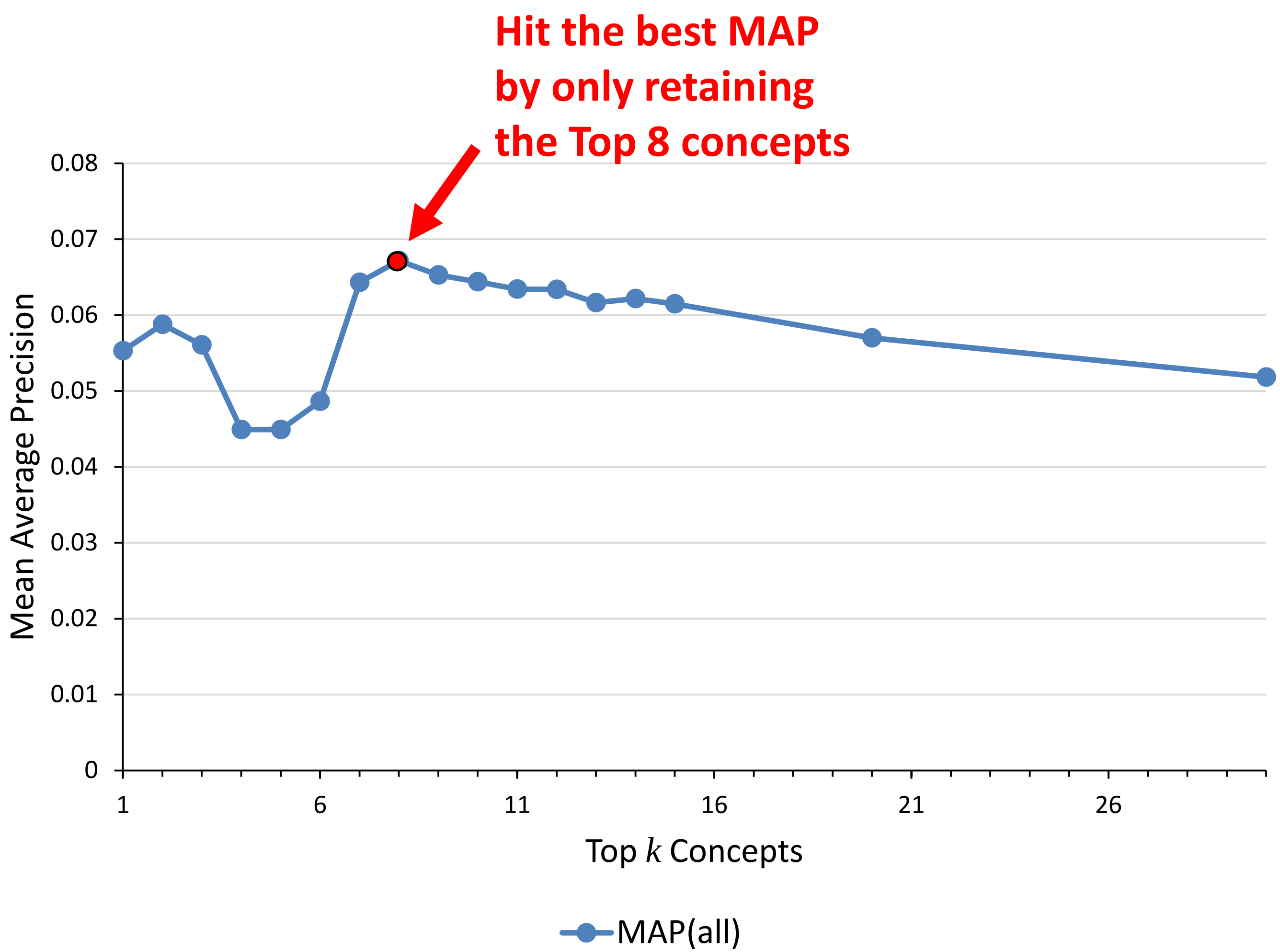
How to wisely choose the right concepts?

Recall an important finding in the last year



Event 31: Beekeeping

Strategies for automatic SQG last year



What we got?

- The top few concepts might have already achieved a good performance
- Adding concepts that are *less relevant* tends to decrease the performance

Per-dataset performance by using *best-k* concepts (MED14-Test)

EventID	EventName	Research497 (Top 2)	ILSVRC1000 (Top 3)	SIN346 (Top 5)	Places205 (Top 2)	FCVID239 (Top 1)	Sports487 (Manual)
21	attempting_bike_trick	0.132	0.109	0.059	0.007	0.003	0.196
22	cleaning_appliance	0.012	0.019	0.005	0.009	0.062	0.002
23	dog_show	**0.430**	0.011	0.012	0.004	0.004	**0.777**
24	giving_direction_location	0.006	0.003	0.003	0.007	0.001	0.003
25	marriage_proposal	0.005	0.002	0.006	0.002	0.010	0.006
26	renovating_home	0.007	0.003	0.003	0.003	0.001	0.006
27	rock_climbing	0.022	0.004	0.001	0.004	0.065	**0.288**
28	town_hall_meeting	0.024	0.001	0.016	0.008	**0.148**	0.001
29	winning_race_vehicle	0.147	0.005	0.001	0.006	0.011	0.016
30	working_metal_craft_project	0.144	0.009	0.002	0.001	0.005	0.001
31	beekeeping	0.003	**0.648**	0.002	0.002	0.262	0.001
32	wedding_shower	0.009	0.003	**0.022**	0.002	0.005	0.003
33	non-motorized_vehicle_repair	0.026	0.002	0.005	0.002	0.008	**0.450**
34	fixing_musical_instrument	0.016	0.002	0.011	0.004	**0.146**	0.001
35	horse_riding_competition	0.013	0.022	0.071	**0.234**	0.115	**0.278**
36	felling_tree	0.022	0.004	0.018	**0.051**	0.018	0.001
37	parking_vehicle	0.026	0.057	0.037	0.022	**0.215**	0.002
38	playing_fetch	0.002	**0.032**	0.010	0.017	0.008	0.020
39	tailgating	0.002	0.001	0.001	0.007	**0.232**	0.001
40	tuning_musical_instrument	0.008	**0.048**	0.001	0.002	**0.050**	0.001
MAP(all)		0.053	0.049	0.014	0.020	0.071	**0.103**
MAP(21-30)		0.093	0.017	0.011	0.005	0.037	0.130
MAP(31-40)		0.013	0.082	0.018	0.034	0.106	0.076

Finding: If a good match can be found, high-level concepts far overwhelm componential concepts such as objects and scenes.

Strategies for manual concept screening

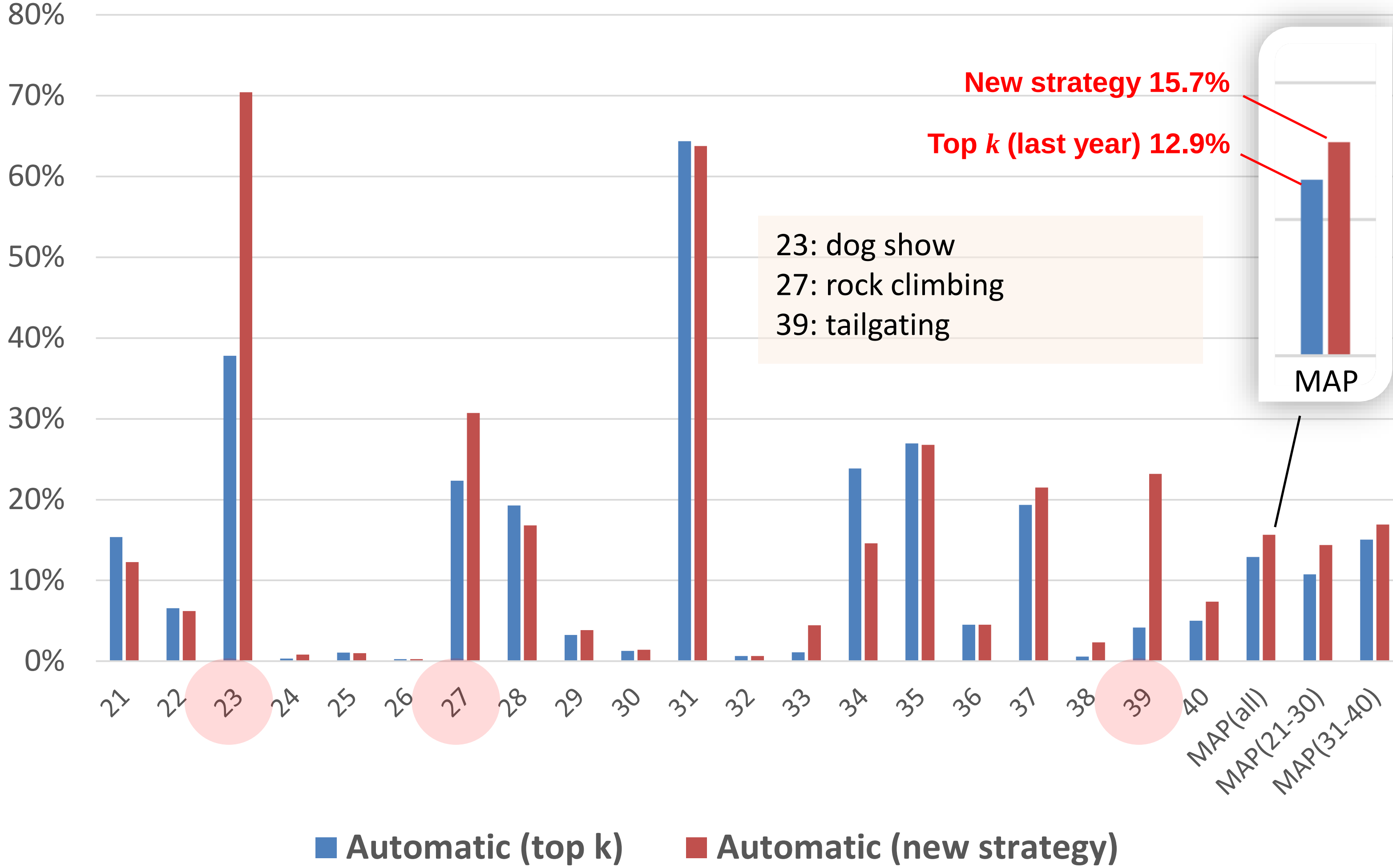
- Only carefully include concepts that are *distinctive* to an event if we find a concept detector *semantically same* as the event
- Remove *false positives* by screening the names of concepts
- Remove concepts for which training videos appear in *very different context* based on human's common sense

- *Rock climbing, bouldering, sport climbing, artificial rock wall*	**Relevant**
- *Rope climbing, climbing, rock*	**Not distinctive**
- *Rock fishing, rock band performance*	**False positive**
- *Stone wall, grabbing rock*	**Different context**

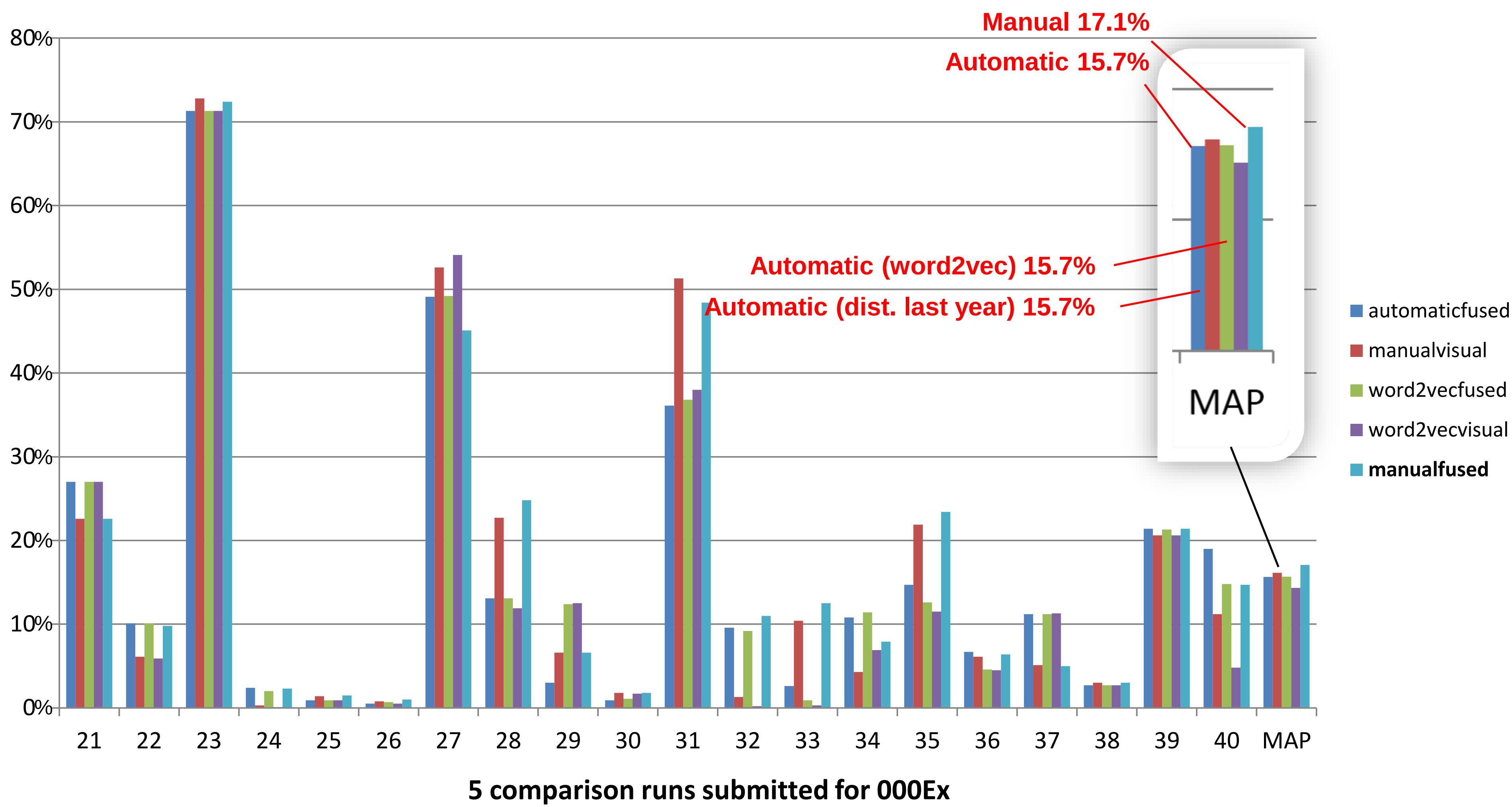
Strategies for automatic SQG

- If a concept detector with the *same name* of the event can be found, simply choose that detector and discard anything else
- Otherwise, choose the top k concepts according to the relevance score
- k is found to be optimized at around 10, and kept the same for all events

Automatic SQG top *k* vs. new strategy (MED14-Test)



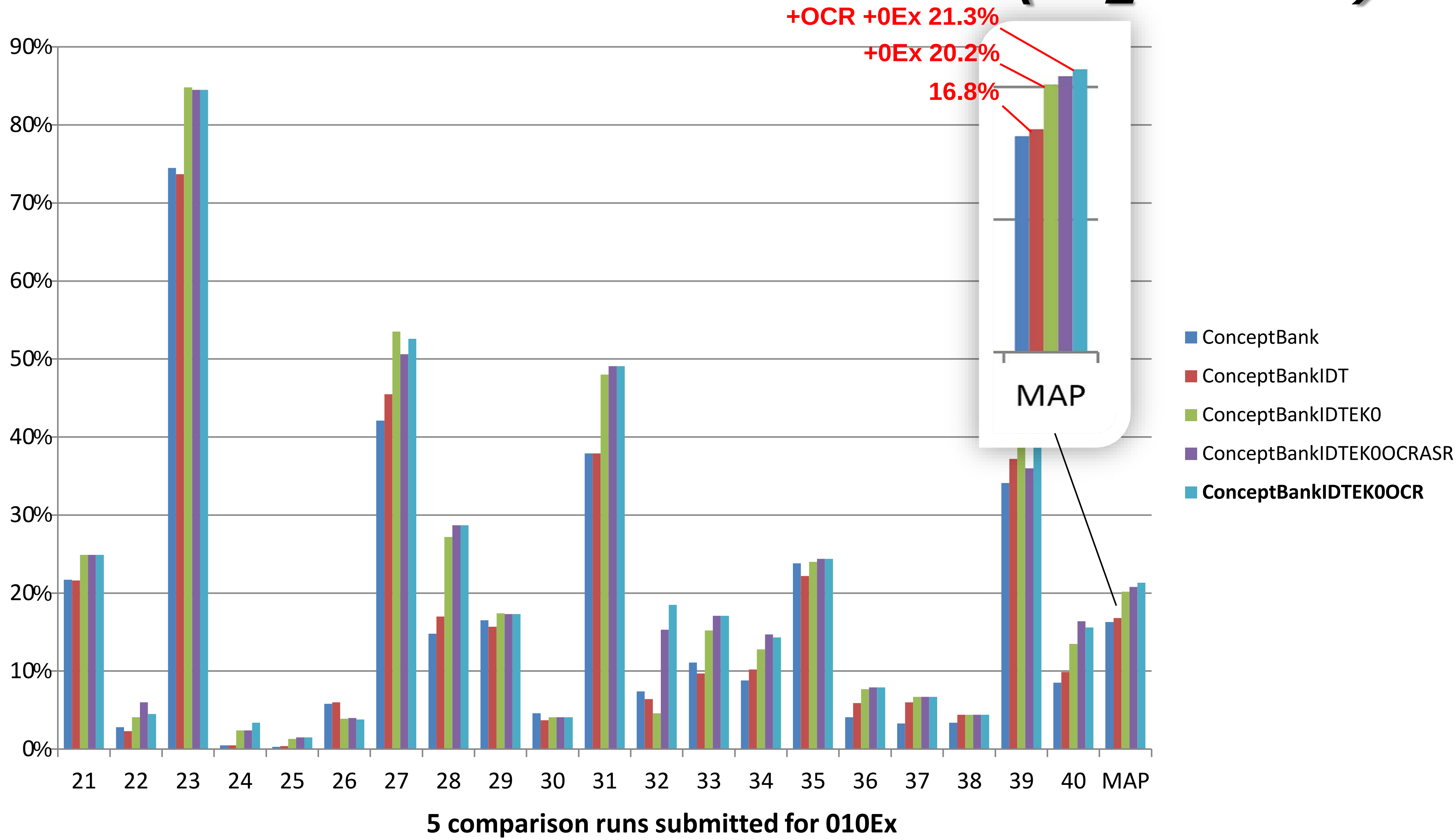
Manual vs. Automatic *(PS_EvalFull)*



VIREO Video Retrieval Group

TNO innovation for life

Contribution of 0Ex in 10Ex task *(PS_EvalFull)*



Summary

- An enlarged concept bank involving *high-level concepts* such as activities and events does great help for event detection
- A wise strategy for picking up the right concepts given a large concept bank is key to the detection performance

Thank you!