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

Speaker: Yi-Jie Lu

Yi-Jie Lu¹, Maaike de Boer²,³, Hao Zhang¹,
Klamer Schutte², Wessel Kraaij²,³, Chong-Wah Ngo¹

¹VIREO Group, City University of Hong Kong, Hong Kong
²Netherlands Organization for Applied Scientific Research (TNO), Netherlands
³Radboud 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

PS_EvalFull_000Ex MAP

- Auto '14: 5.2%
- Auto '15: 15.7%
- Manual '15: 17.1%
Achievements in 2015

PS_EvalSub_000Ex MAP

Manual ’15
Important changes from ’14?
• Recall the Semantic Query Generation (SQG):

**Event Query**

(Attempting a Bike Trick)

**Concept Bank**

**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

Exact Match
WordNet
TFIDF, Specificity ...
Recall our 2014 findings

Extinguishing a Fire

Missing key concepts

[ Fire extinguisher ]  [ Firefighter ]

Exact match >> WordNet/ConceptNet
What we do?

**Event Query**
(Attempting a Bike Trick)

**SQG**

**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

**Enlarged Concept Bank**

1. Manually
2. Refined Query
**Enlarge the concept bank**

**2014**
- Research set (497)
- ImageNet ILSVRC (1000)
- SIN (346)

**2015**
- CNN + Sports (487)
- CNN + FCVID (239)
- CNN + Places (205)

---

**SFRISP (2774)**
Concept Bank Review

Higher level

Sports 487 (activities, events)

ImageNet 1000 (objects)

FCVID 239 (activities, events)

Places 205 (scenes)

RS 497 (mixed)

SIN 346 (objects, actions)

Lower level
• Sports (487) [1]

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

MAP on MED14-Test

- with Sports and FCVID: 19.2%
- without: 10.8%

- without (Manual): 10.8%
- with (Manual): 19.2% (-8.4%)
Contribution of Sports+FCVID (726 concepts) on MED14-Test

- 23: dog show
- 27: rock climbing
- 28: town hall meeting
- 34: fixing musical instrument
- 35: horse riding competition
- 37: parking vehicle
- 39: tailgating
- 40: tuning musical instrument

The diagram compares the contribution of Sports+FCVID with and without the manual approach. The concepts are visualized with bars indicating their contribution percentage.
In combination of 6 different resources:

**How to wisely choose the right concepts?**
Recall an important finding in the last year:

Event 31: Beekeeping

- Bee house (ImageNet)
- Cutting (research collection)
- Cutting down tree (research collection)
- Bee (ImageNet)
- Honeycomb (ImageNet)
Strategies for automatic SQG last year

Hit the best MAP by only retaining the Top 8 concepts

Mean Average Precision

Top k Concepts

MAP(all)
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)

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

**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
- Rope climbing, climbing, rock
- Rock fishing, rock band performance
- Stone wall, grabbing rock
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)

<table>
<thead>
<tr>
<th>Week</th>
<th>Activity</th>
<th>Automatic (top k)</th>
<th>Automatic (new strategy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td></td>
<td>New strategy 15.7%</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
<td>Top $k$ (last year) 12.9%</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td></td>
<td></td>
<td>23: dog show</td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
<td>27: rock climbing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>39: tailgating</td>
</tr>
</tbody>
</table>

MAP

Automatic (top k) vs. Automatic (new strategy)
Manual vs. Automatic (PS_EvalFull)

- Manual: 17.1%
- Automatic: 15.7%
- Automatic (word2vec): 15.7%
- Automatic (dist. last year): 15.7%

5 comparison runs submitted for 000Ex
Contribution of 0Ex in 10Ex task (PS_EvalFull)

5 comparison runs submitted for 010Ex

+OCR +0Ex 21.3%
+0Ex 20.2%
16.8%

MAP

ConceptBank
ConceptBankIDT
ConceptBankIDTEK0
ConceptBankIDTEK0OCRASR
ConceptBankIDTEK0OCR
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!