On the use of semantic features for the semantic indexing task

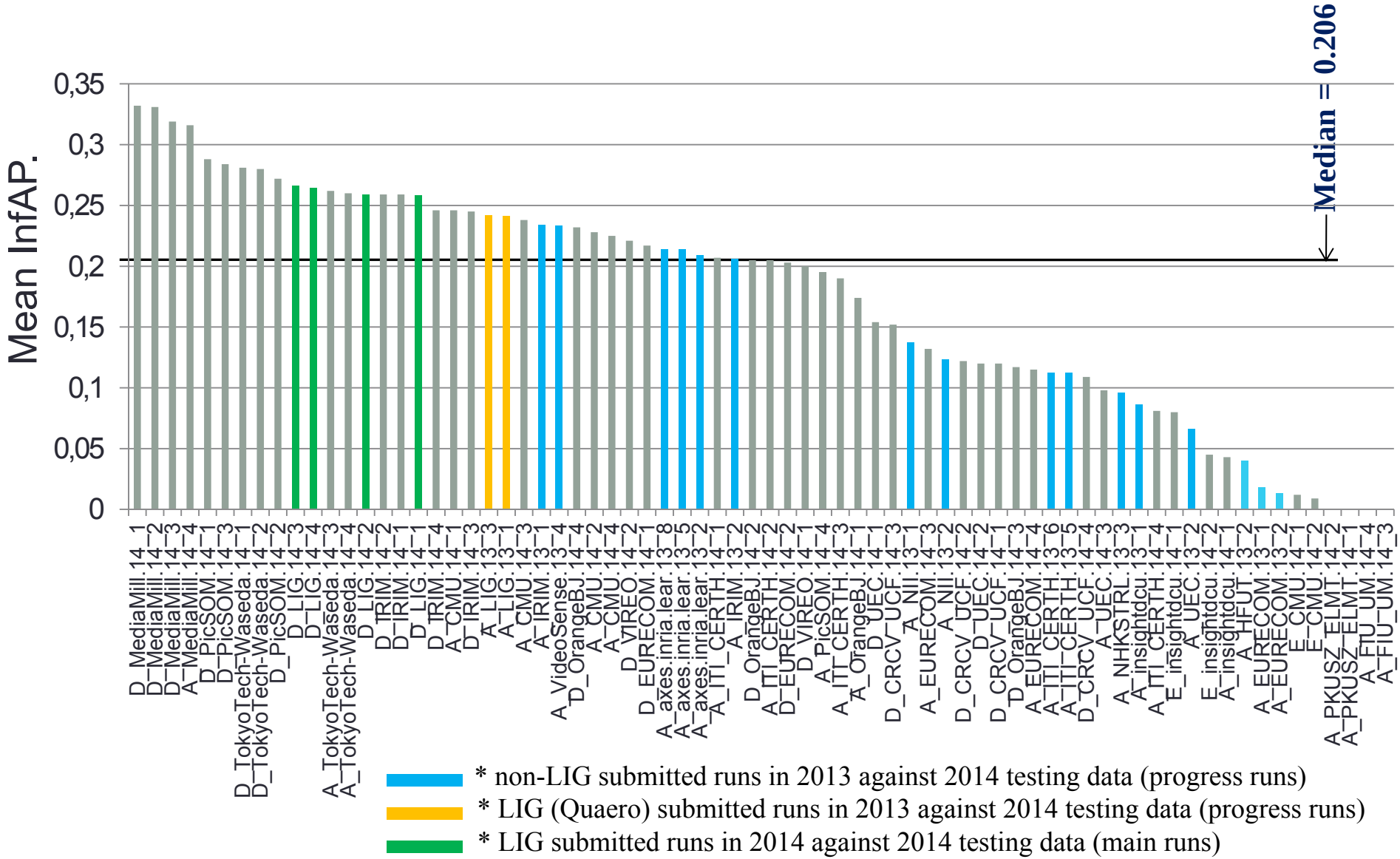
Bahjat Safadi, Nadia Derbas, Abdelkader Hamadi,
Mateusz Budnik, Philippe Mulhem and Georges Quénot
UJF-LIG
and
many other people from the IRIM group of GDR 720 ISIS

10 November 2014

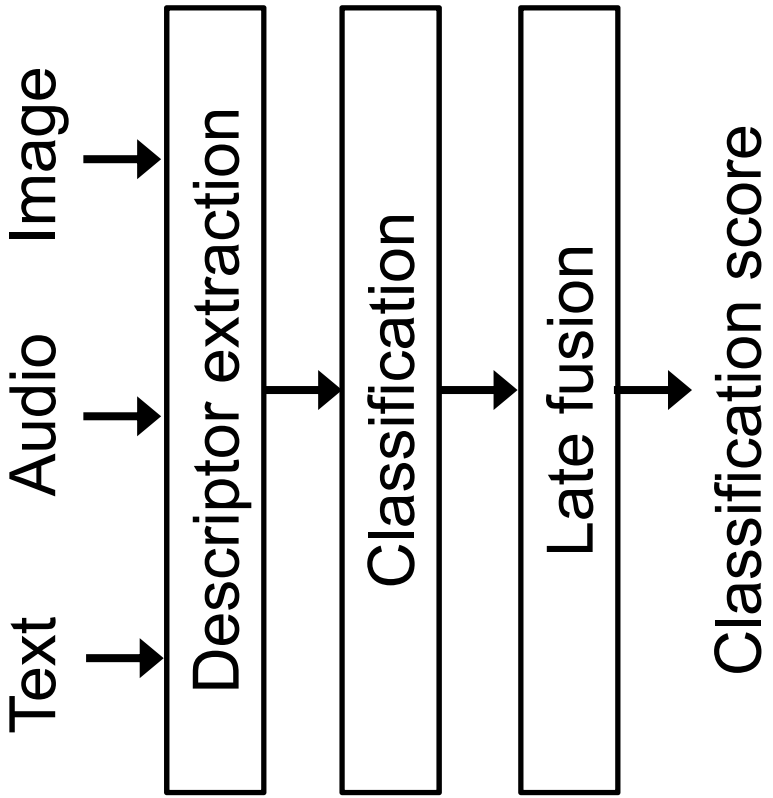
Outline

- System overview
- Semantic features
- Contrast experiments
- Engineered versus learned features
- Conclusion

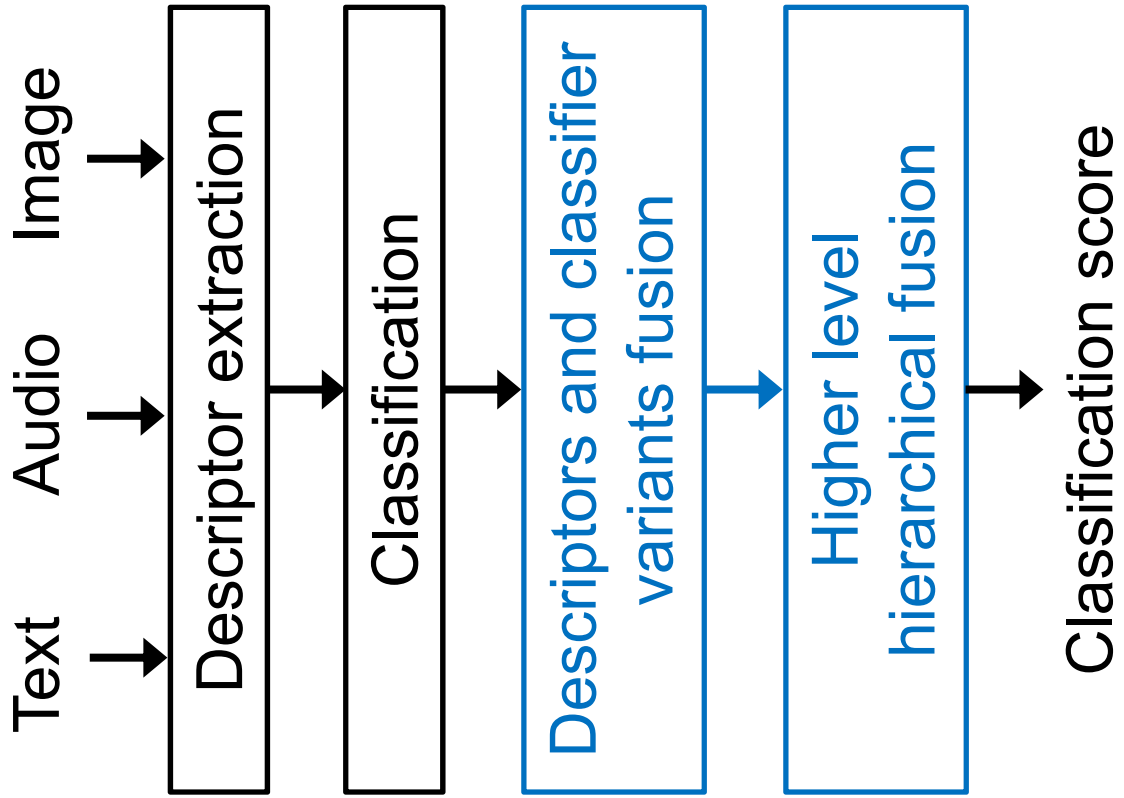
Main runs scores 2014 (from NIST)



Basic classification pipeline

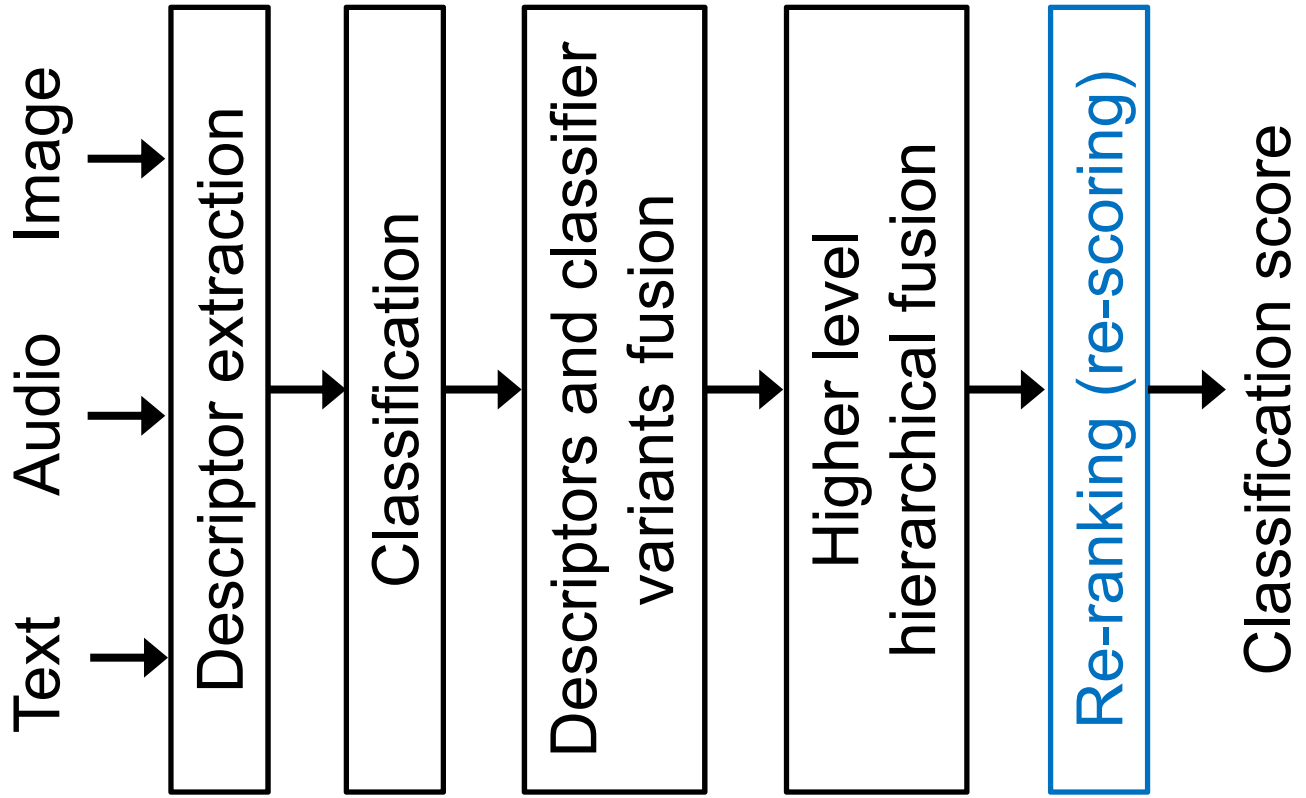


LIG/Quaero/IRIM classification pipeline



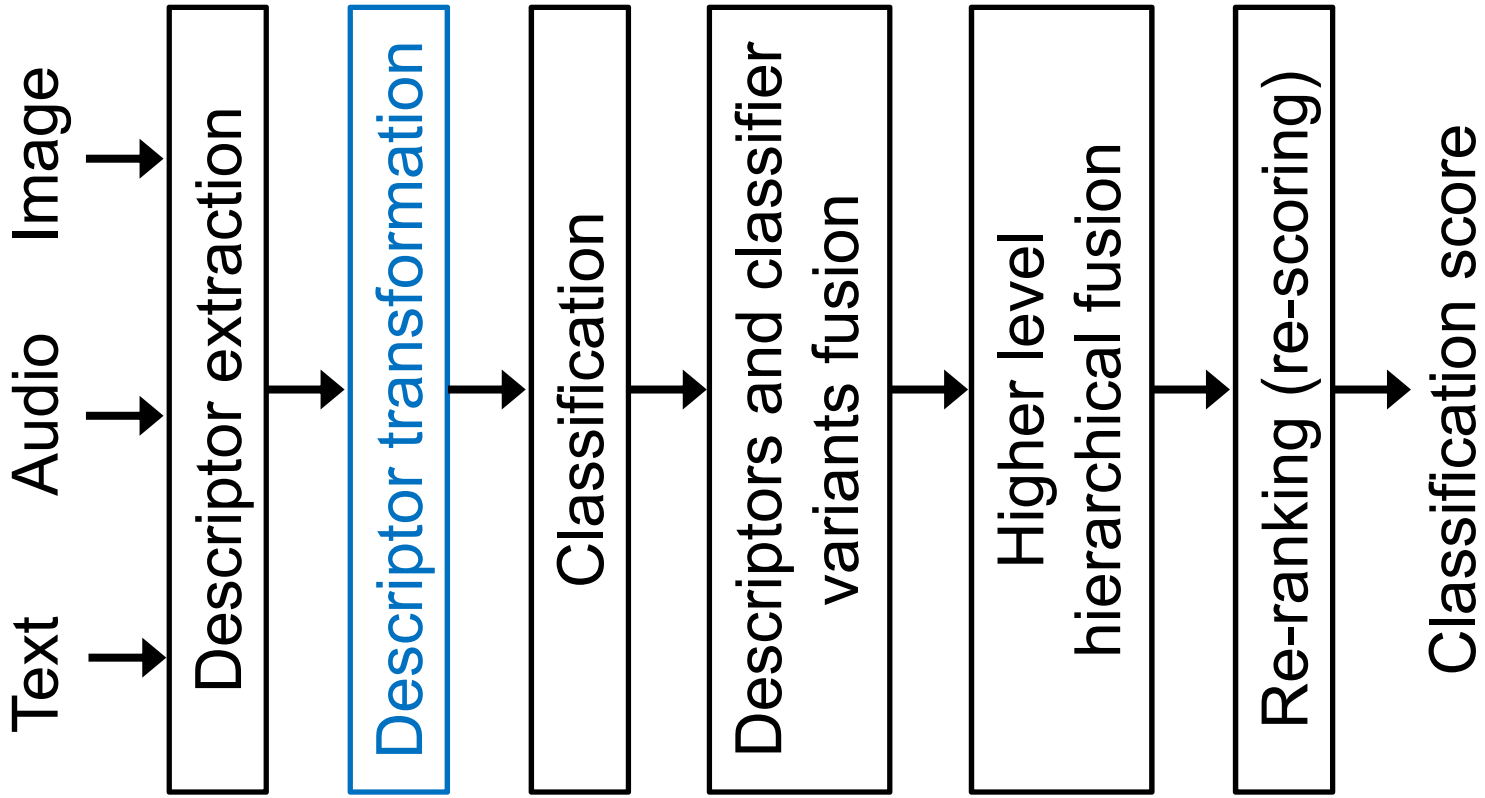
+ hierarchical fusion [Strat et al., ECCV/IFCVCR workshop 2012, Springer 2014]

LIG/Quaero/IRIM classification pipeline



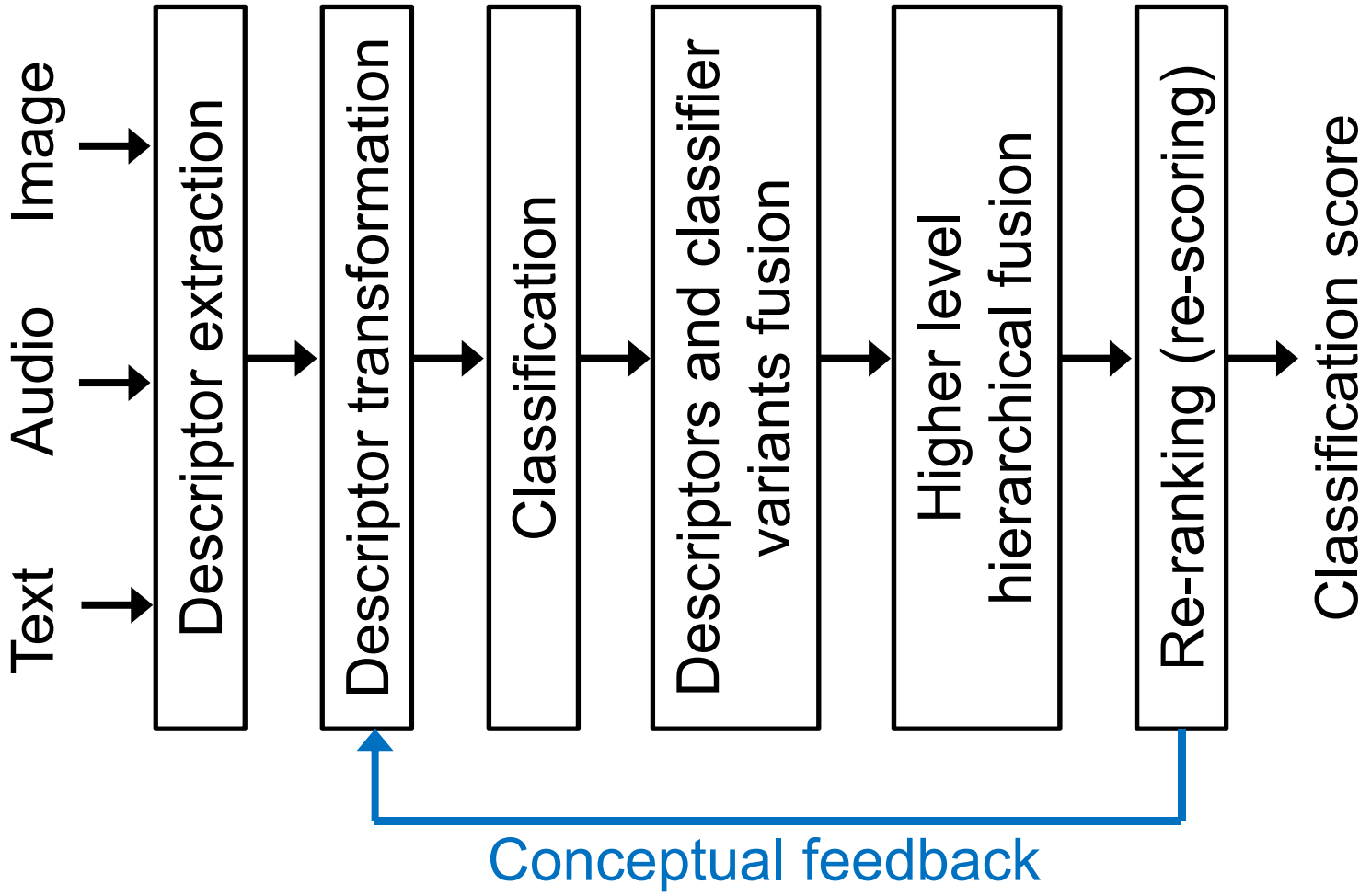
+ Temporal re-ranking [Safadi et al., CIKM 2011; Wang et al, TV 2009]: update shot scores considering other shots' scores for a same concept

LIG/Quaero/IRIM classification pipeline



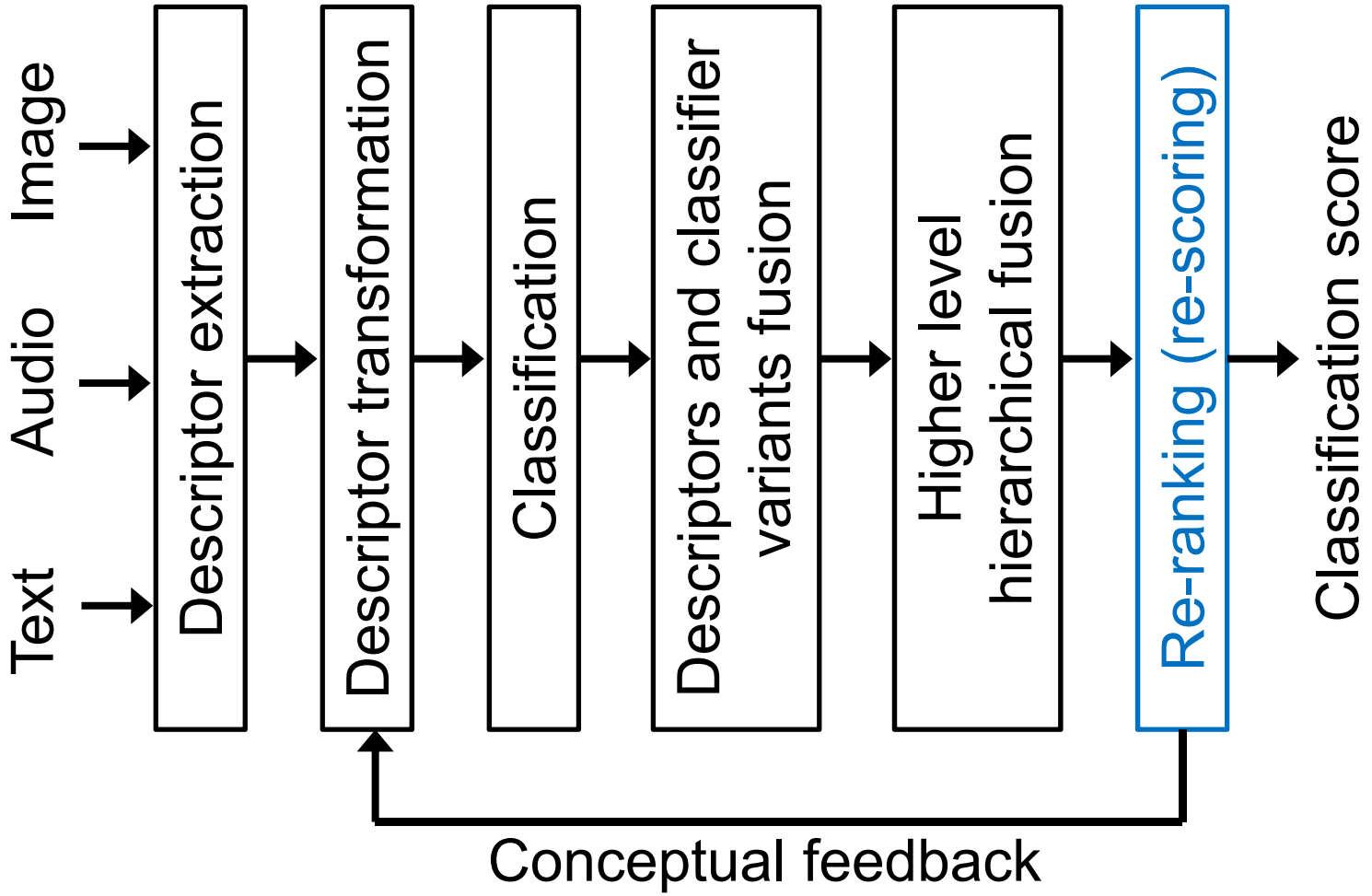
+ Descriptor optimization [Safadi et al., MTAP 2014]: combination of PCA-based dimensionality reduction and pre- and post- power transformations

LIG/Quaero/IRIM classification pipeline



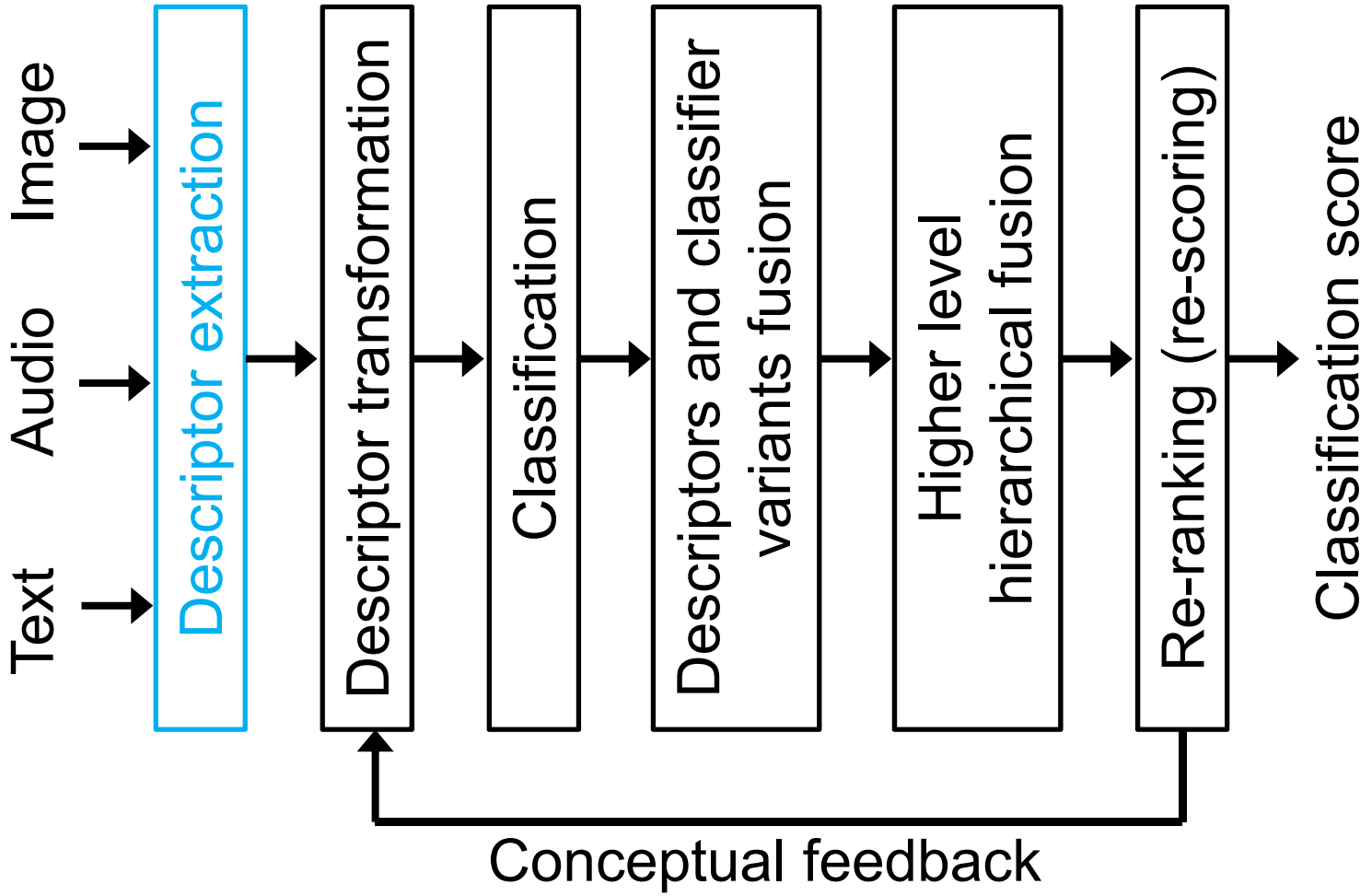
+ conceptual feedback [Hamadi et al., MTAP, 2014]

LIG/Quaero/IRIM classification pipeline



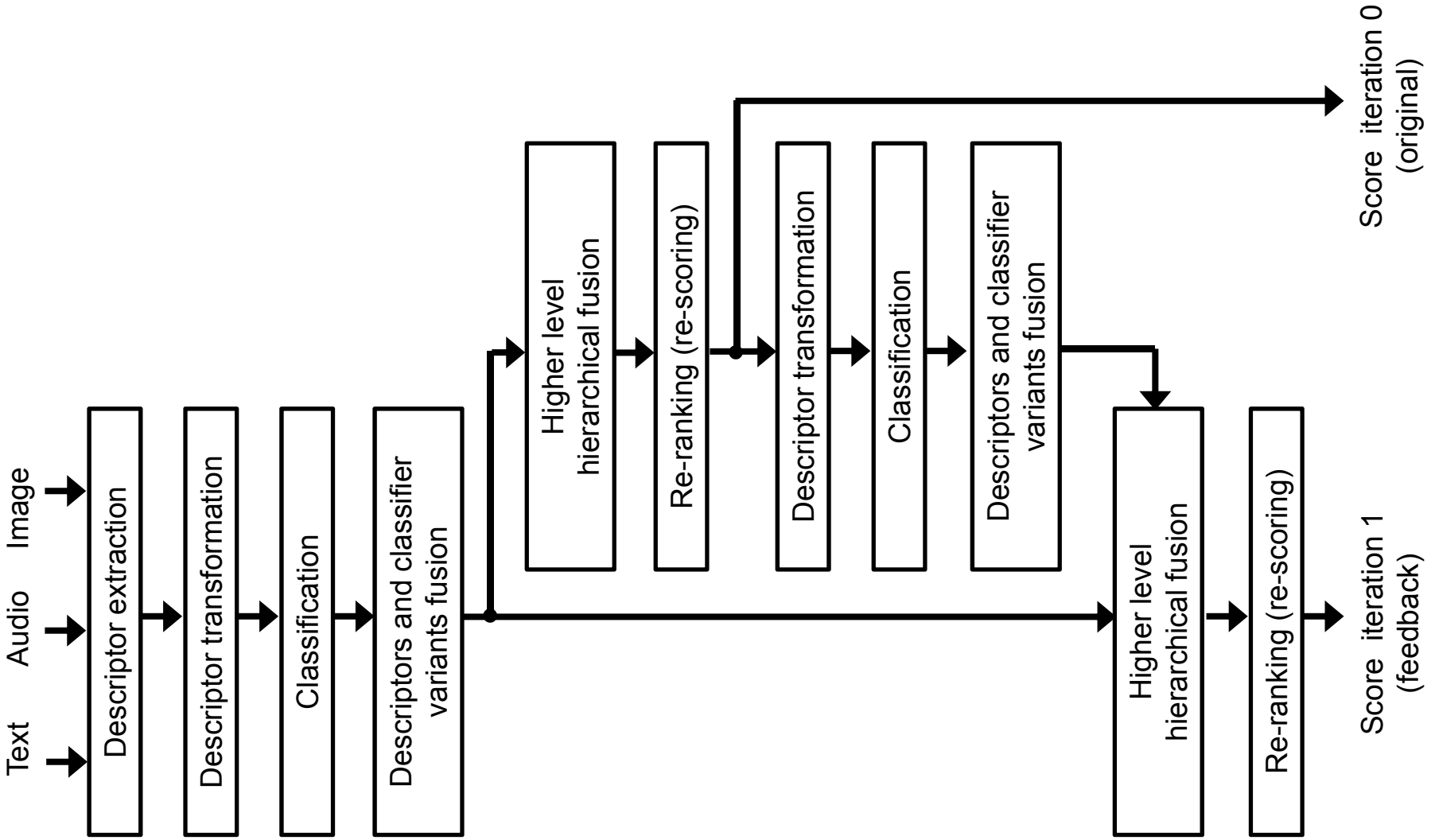
+ conceptual re-ranking [Hamadi et al., MTAP, 2014] update concept scores considering other concepts' scores for a same shot

LIG/Quaero/IRIM classification pipeline

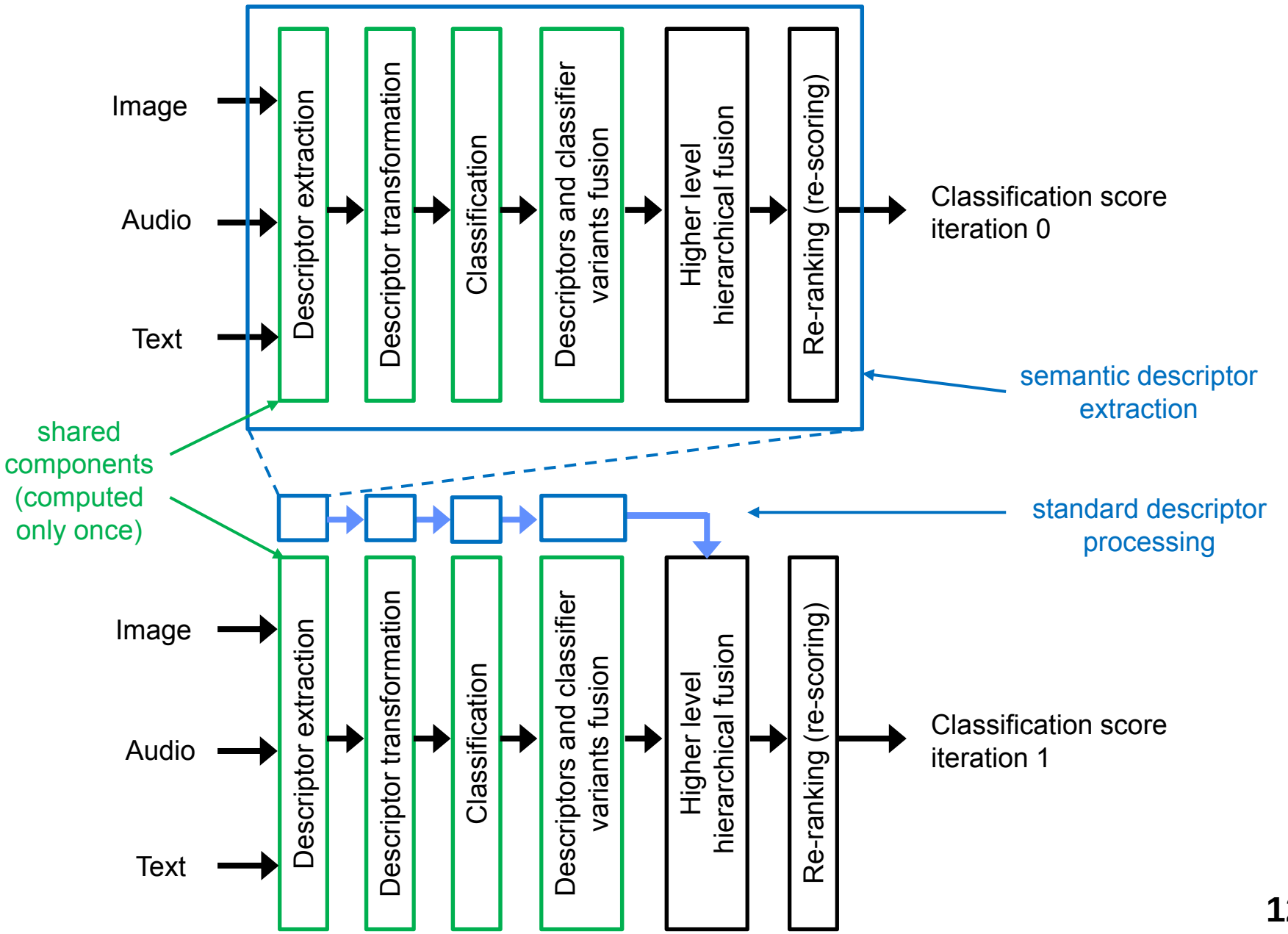


+ semantic descriptors [TRECVid 2013 and 2014]

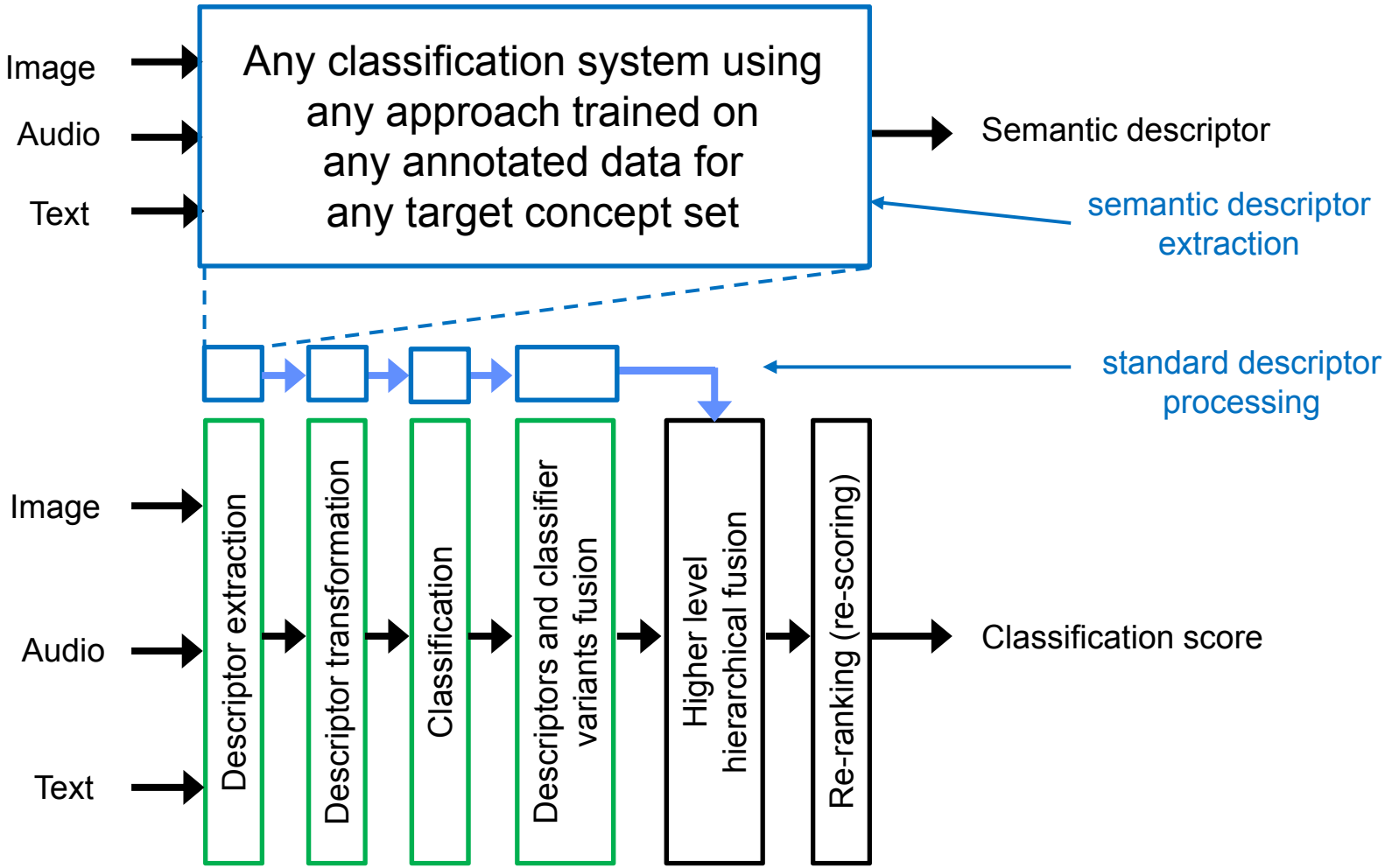
Conceptual feedback: unfolded graph



Conceptual feedback: semantic descriptor



Semantic descriptor: general case



Model vectors [Smith et al. ICME 2003]

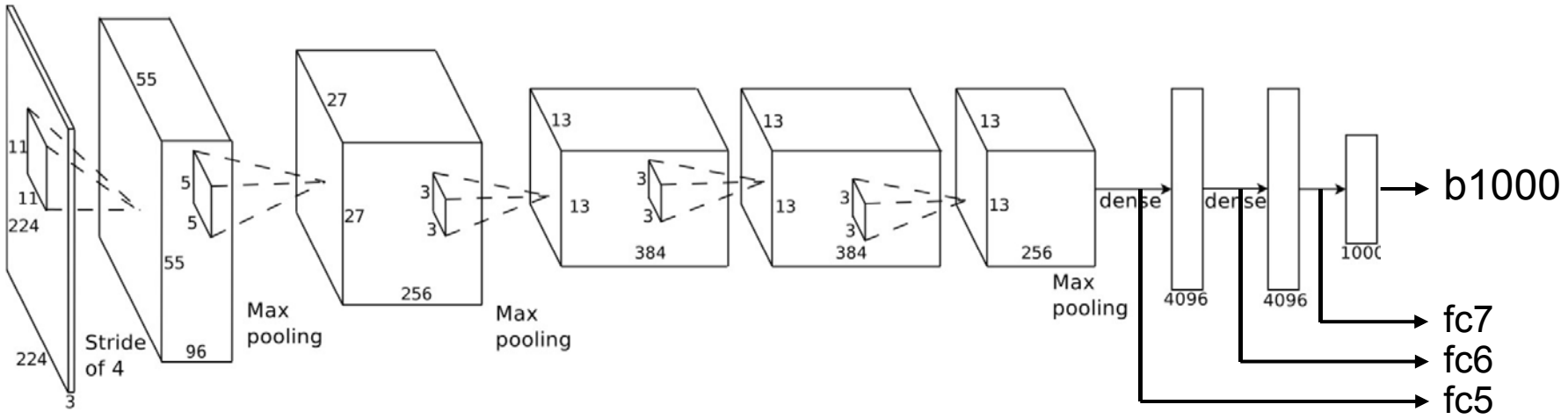
Semantic descriptors trained on ImageNet

- Fisher Vector based descriptor [Perronnin, IJCV 2013]:
  - XEROX/ilsvrc2010: vectors of 1000 scores trained on ILSVRC10 and applied to key frames, kindly produced by Florent Perronnin from Xerox (XRCE)
  - XEROX/imagenet10174: same with10274 concepts scores trained ImageNet
- Deep learning based descriptors, computed by Eurecom and LIG using Berkeley caffe tool [Jia et al, 2013]:
  - EUR/caffe1000: vectors of 1000 scores trained on ILSVRC12 and applied to key frames, fusing outputs for 10 variants of each input image
  - LIG/caffe1000b: same with a different version of the tool and using only one variant of each input image

“Quasi-semantic” descriptors from deep learning and ImageNet

[Krizhevsky et al., 2012]

- 7 hidden layers, 650K units, 630 M connections, 60M parameters
- GPU implementation (50× speed-up over CPU)
- Trained on two GPUs for a week



“Quasi-semantic” descriptors from deep learning and ImageNet

- Deep learning based descriptors, computed by LIG using Berkeley caffe tool [Jia et al, 2013]:
  - LIG/caffe_fc7b_4096: 4096 values of the last hidden layer (non convolutional)
  - LIG/caffe_fc6b_4096: 4096 values of the last but one hidden layers (non convolutional)
  - LIG/caffe_fc5b_43264: 43264 values of the last but two hidden layers (convolutional, 13×13×256)
- Not strictly semantic as not classification scores, close to the semantic level however
- Expected to perform better than the last layer:
  - No (or les) information loss due to the targeting of different and/or unrelated target concepts

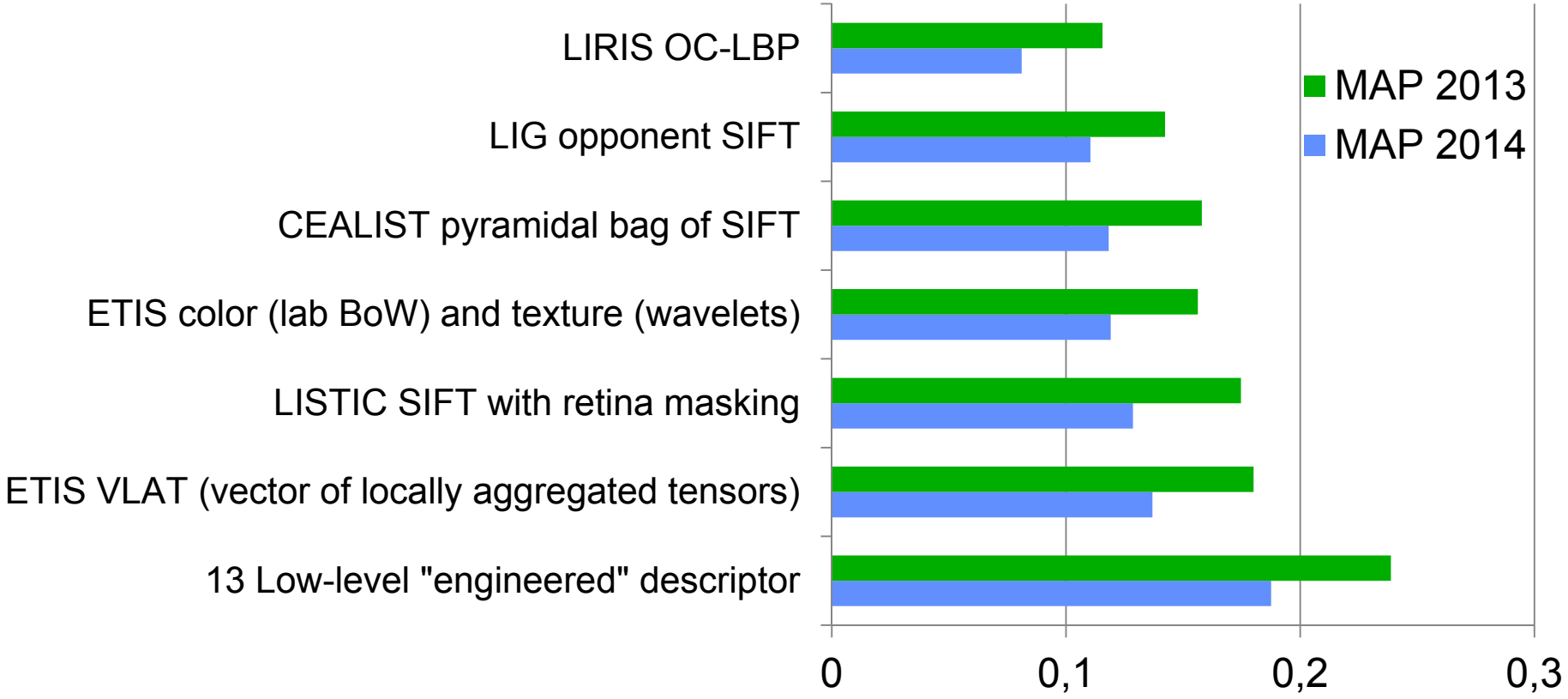
Local semantic descriptors trained on TRECVid 2003

- Scores for 15 TRECVid 2003 concepts (sky, building, water, greenery ...) on image patches trained using local annotations [Ayache et al., IVP 2007]
  - LIG/percepts*: computed at various resolutions in a pyramidal way, aggregated by concatenation
  - Computed using local color and texture descriptors
- No longer state of the art

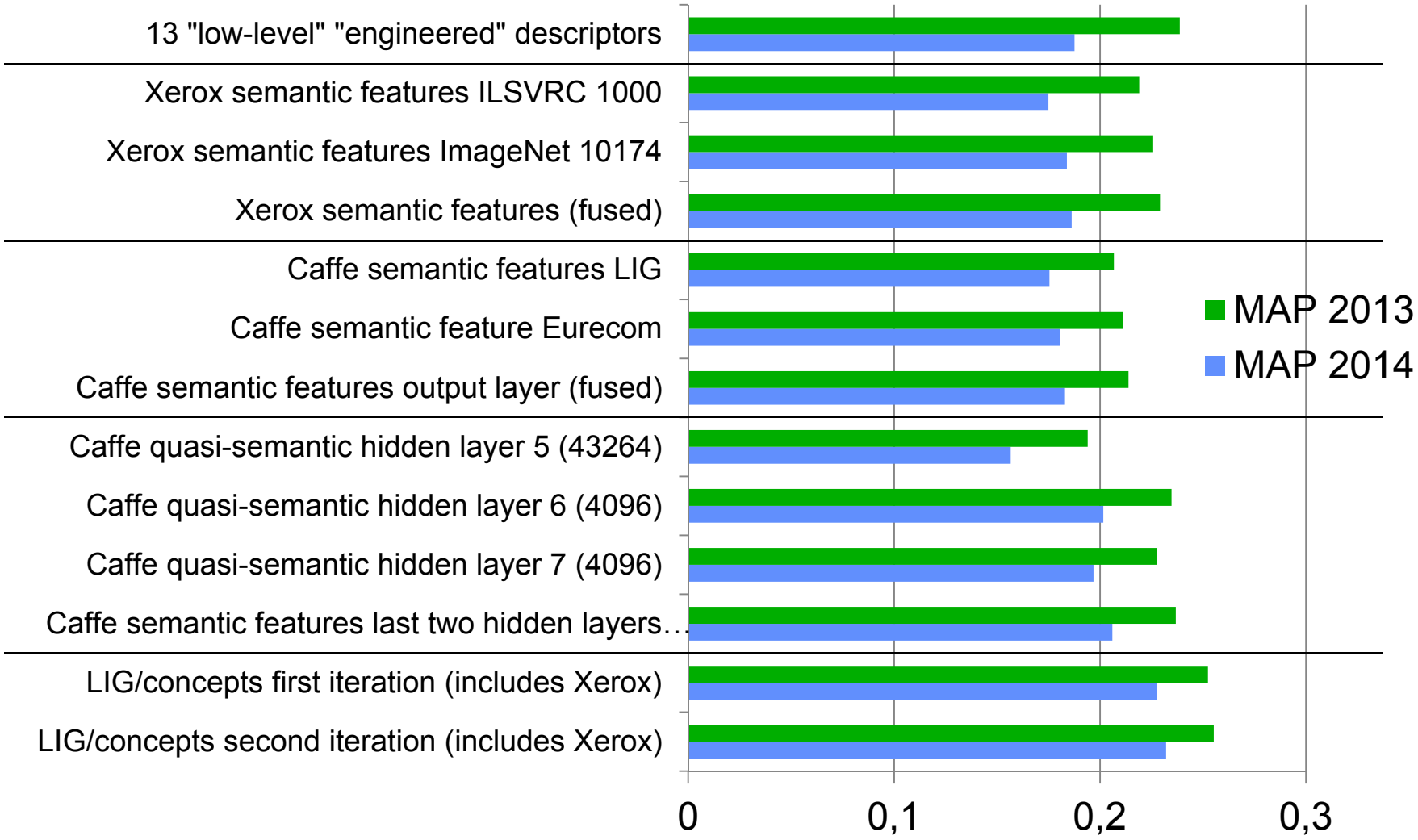
Experiments

- Use of SIN 2013 development data only (no tuning on SIN 2013 test data) and various components using ImageNet annotated data → D type submissions
- Evaluation on SIN 2013 and 2014 test data
- Use of a combination of kNN and MSVM for classification [Safadi, RIAO 2010]
- Use of uploader information: multiplicative factor at the video level, weighted at 10%, provided by Eurecom [Niaz, TV 2012]

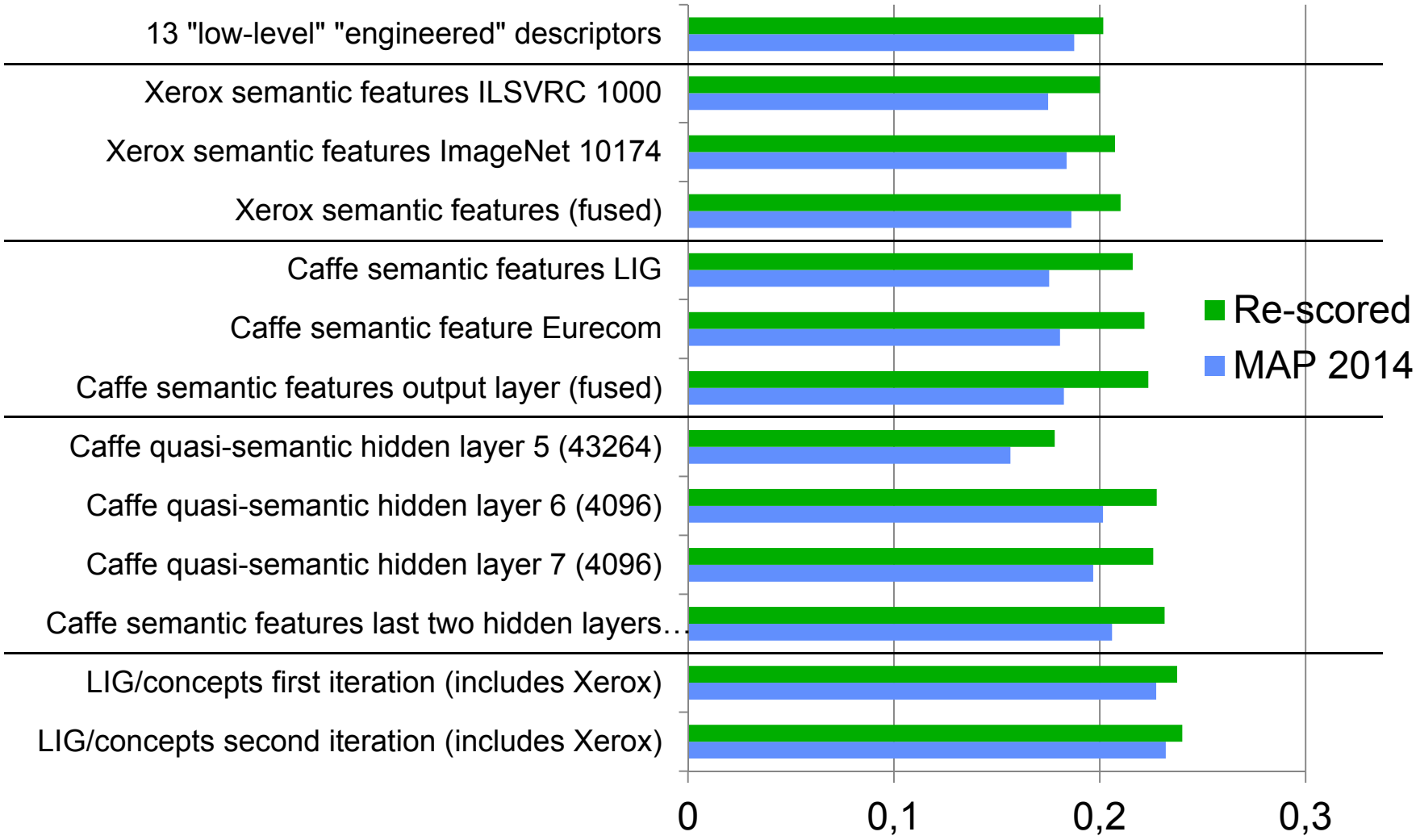
Performance of “low-level” descriptors



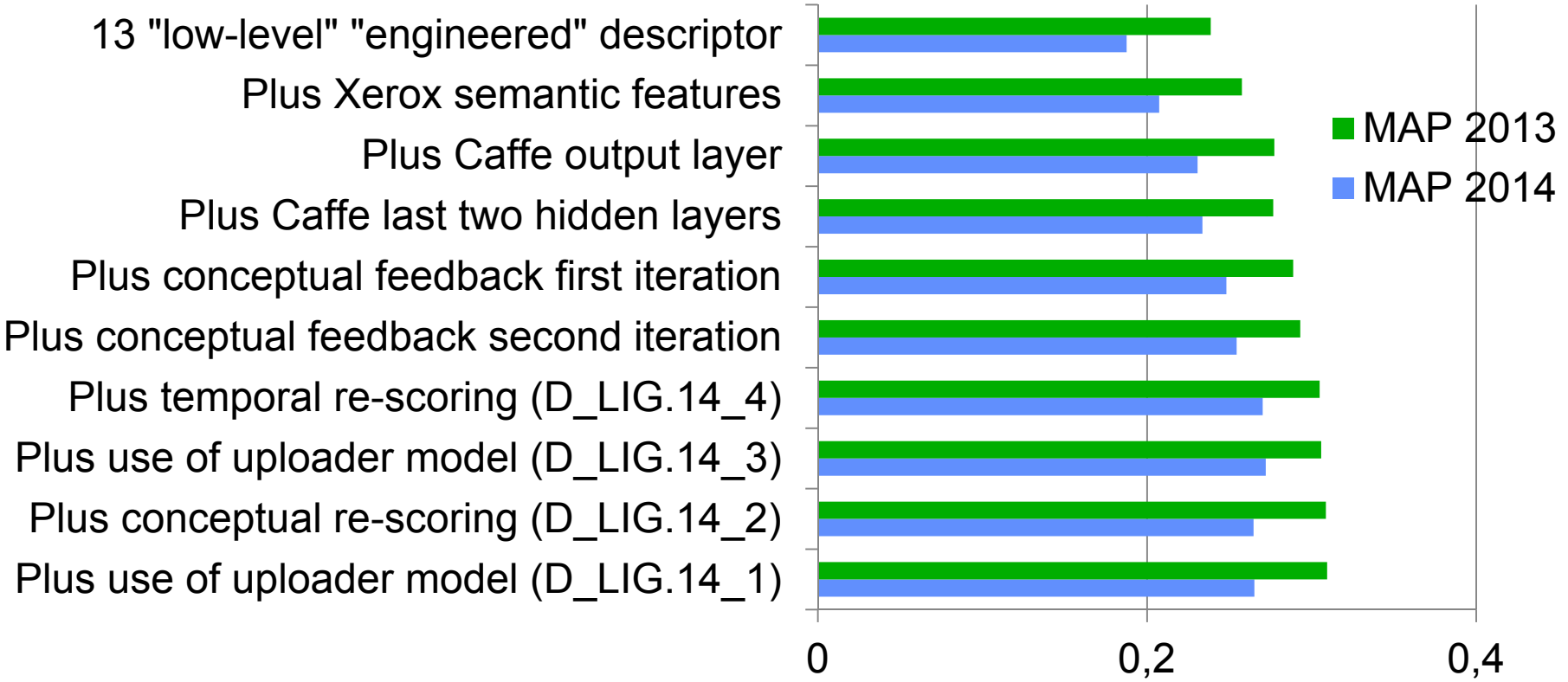
Performance of semantic descriptors



Temporal re-scoring on semantic descriptors



Combination of improvement methods



The relative gain brought by each improvement method depends upon the order in which they are applied.

Final conceptual re-scoring did not further improve on 2014 data

Use of semantic features for the semantic indexing task

- Fisher vectors based descriptors on par with deep learning based descriptors
- Both on par with a combination of 13 low-level engineered descriptors types, some of which being state of the art
- Any single engineered descriptor performs significantly lower than any sematic descriptor → why? maybe a question of training data (more and cleaner in ImageNet)
- Conceptual feedback based semantic descriptors better than all (even when not including other semantic ones)
- Fusion and combination with other methods (e.g. temporal re-scoring) further improves
- Direct application of FV and deep learning on SIN training data on-going but unlikely to compete
- Very small gain from the uploader field

Thanks