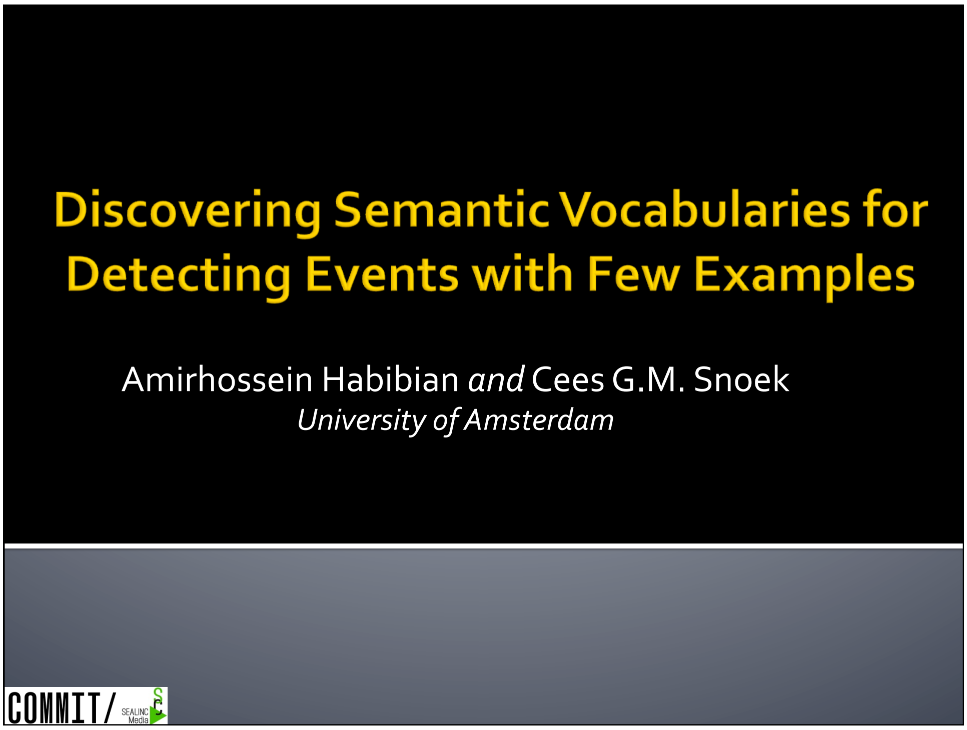
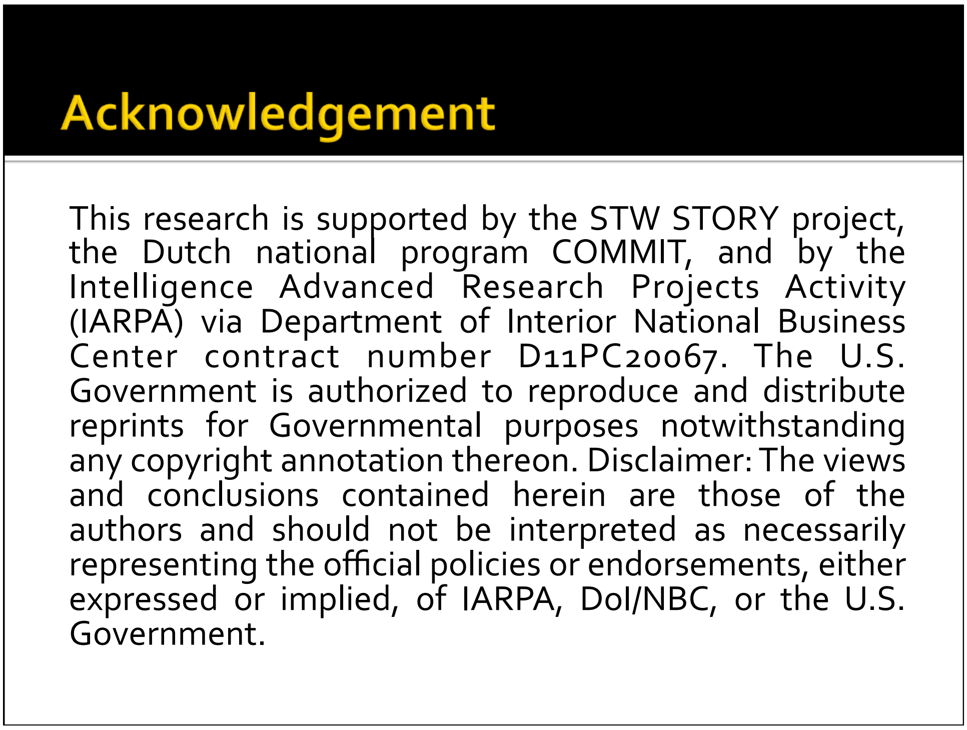


Discovering Semantic Vocabularies for Detecting Events with Few Examples

Amirhossein Habibian *and* Cees G.M. Snoek
University of Amsterdam

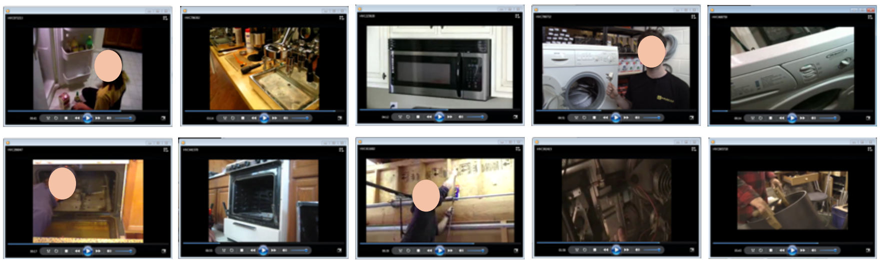


Acknowledgement

This research is supported by the STW STORY project, the Dutch national program COMMIT, and by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center contract number D11PC20067. 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 expressed or implied, of IARPA, DoI/NBC, or the U.S. Government.

Problem statement:
Few-shot Event Detection

- Detecting events from few positive examples
- Large variations in event examples
- Not enough training data to capture the variations



10 positive examples for "Repairing an appliance"

Related Work:
Event Representations

- Low-level Representation [Oneata'12] [Tamrakar'12] [Jiang'13]
  - Events as histogram of low-level features
  - BoW or Fisher encoding of descriptors
  - Applied to various types of descriptors: SIFT, MBH, MFCC …
- Semantic Representation [Merler'12] [Ma'13] [Younessian'12]
  - Events as histogram of detector responses
  - Applying pre-trained concept detectors on videos
  - More semantic representation

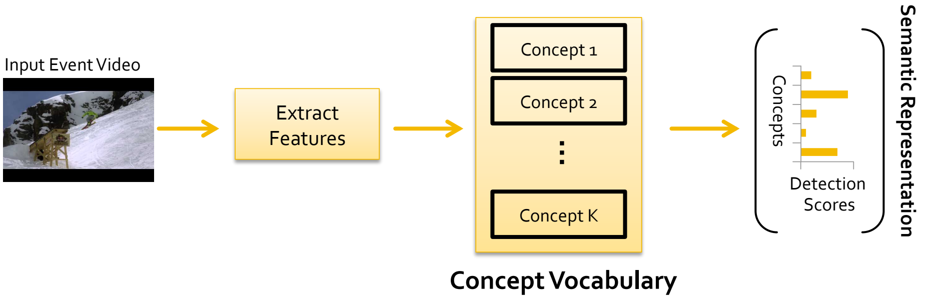
Related Work:
(Semantic) Few-shot Event Detection

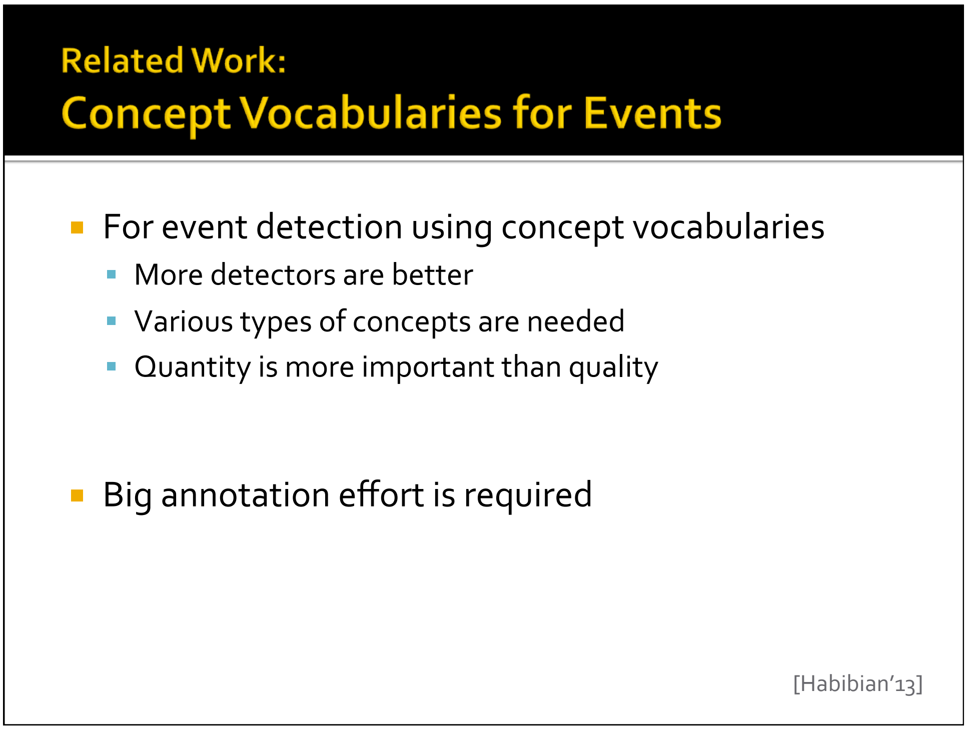
- Semantic representation for few-shot event detection
- Knowledge transfer by using pre-trained detectors
- Outperforms low-level representation
- Effective even with simple classifiers

[Mazloom'13a]

Related Work:
Semantic Representation Pipeline

- Requires a vocabulary of concept detectors
- Pre-trained on annotated image or video datasets
  - ImageNet, TRECVID SIN, Object Bank …
  - [Merler'12] [Mazloom'13b] [Althoff'12]

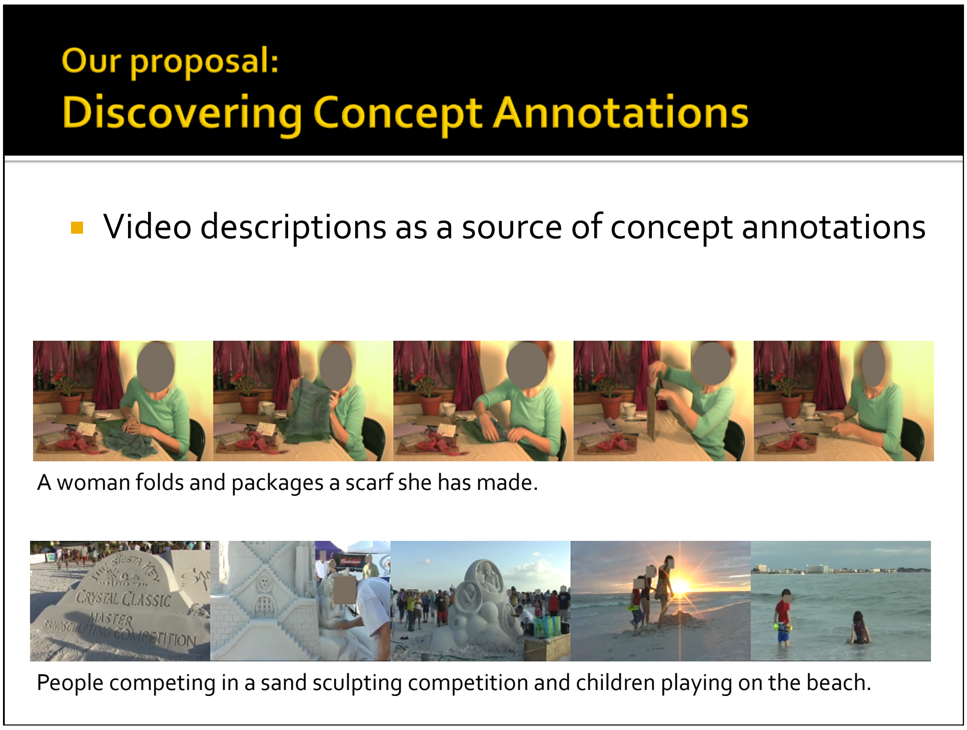




Related Work: Concept Vocabularies for Events

- For event detection using concept vocabularies
  - More detectors are better
  - Various types of concepts are needed
  - Quantity is more important than quality
- Big annotation effort is required

[Habibian'13]

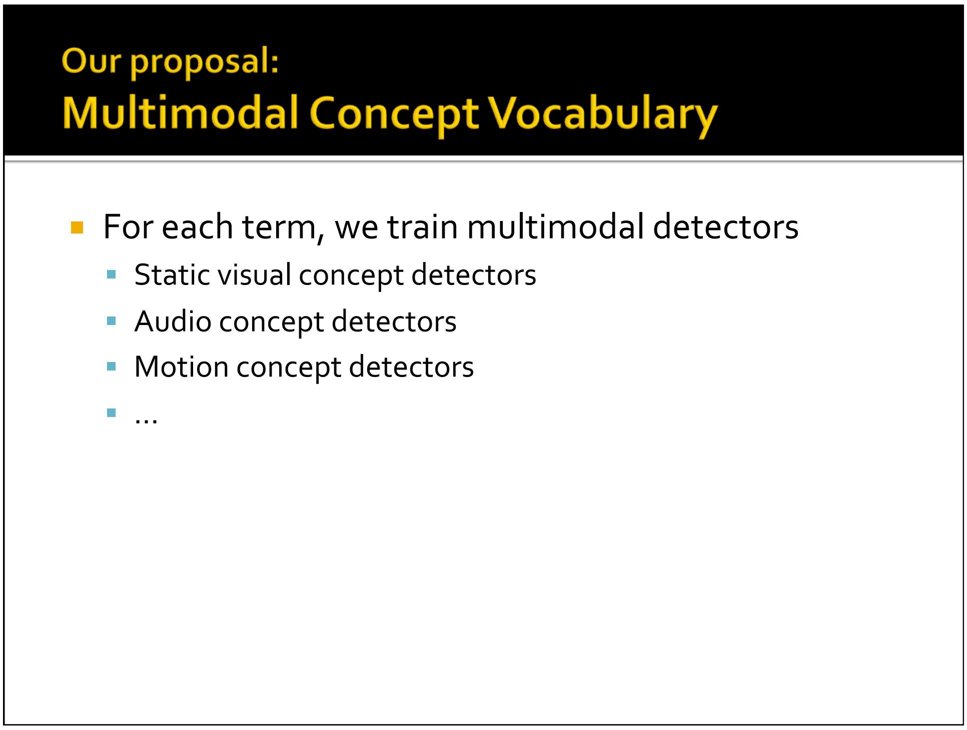


Our proposal: Discovering Concept Annotations

- Video descriptions as a source of concept annotations

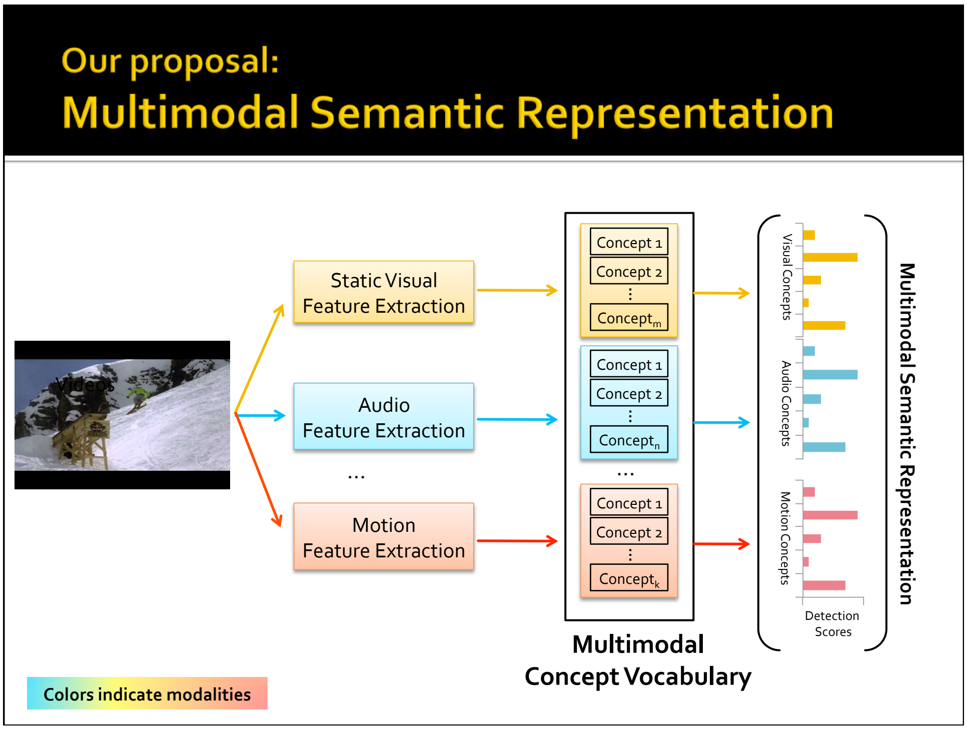
A woman folds and packages a scarf she has made.

People competing in a sand sculpting competition and children playing on the beach.



Our proposal: Multimodal Concept Vocabulary

- For each term, we train multimodal detectors
  - Static visual concept detectors
  - Audio concept detectors
  - Motion concept detectors
  - …



Our proposal: Multimodal Semantic Representation

Colors indicate modalities

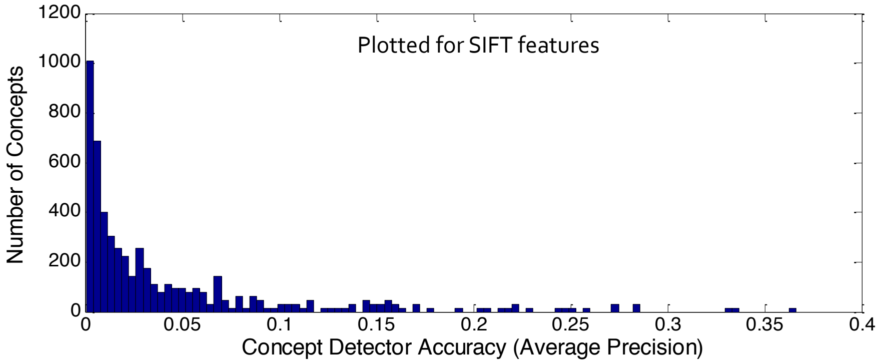
Experimental Setup

- Concept Vocabulary
  - Training data: *research* collection
  - Features: SIFT, HOG, MBH, MFCC, High-level
  - Classifiers: linear SVM
- Event Detection
  - Training data: *medtrain* collection
  - Features: outputs of vocabulary concept detectors
  - Classifiers: nonlinear SVM with HIK kernel

Results:

Discovered Concepts

- We discover around 5,500 unique and frequent terms
- Most of them have low detection accuracies
- 1,500 most accurate detectors are used per modality

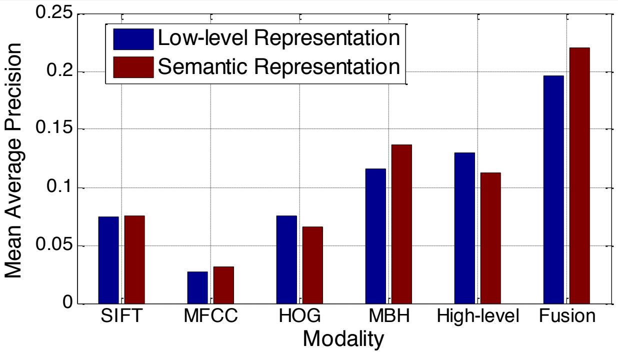


Results:
Examples of Discovered Concepts

- Most accurate concept detectors per modality

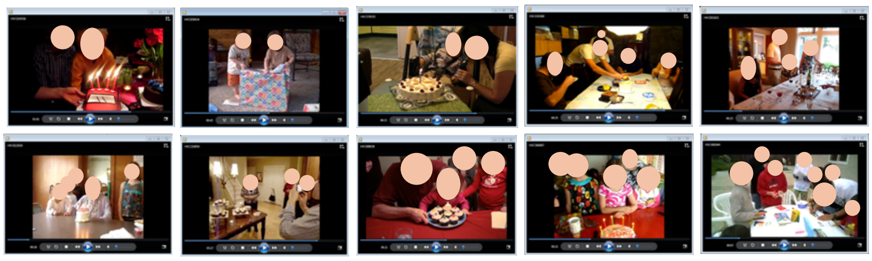
SIFT	MFCC	HOG	MBH	High-level
Dance	Celebrate	People	Dance	People
Bike	Sing	Girl	Climb	Dog
Snow	Demonstrate	Baby	Celebrate	Car
Climb	Baby	Woman	People	Snow
Sack	Bath	Dog	Meeting	Outdoor
Driver	Hall	Cheese	Demonstrate	Woman
Baby	Traffic	Church	Walk	Adult
Car	Play	Microwave	Woman	March
People	Show	School	Flash	Mountain

Results on med13testval:
Event Detection



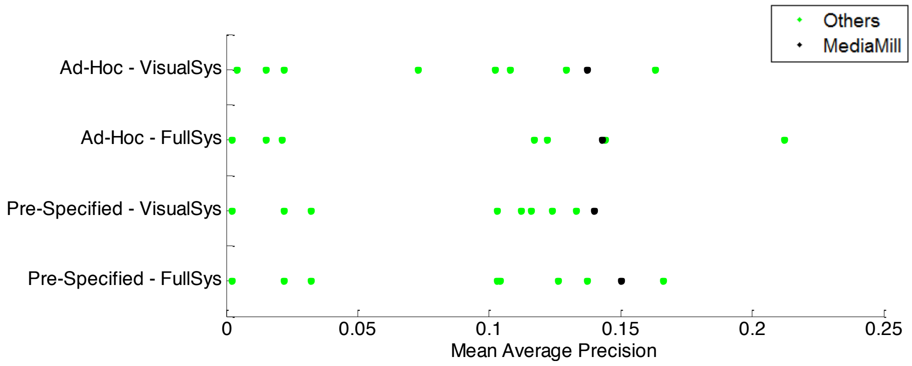
- Low-level representations fused by late fusion (averaging)
- Semantic representations fused by early fusion (concatenation)

Results on med13testval:
Top 10 results for "*Birthday party*"

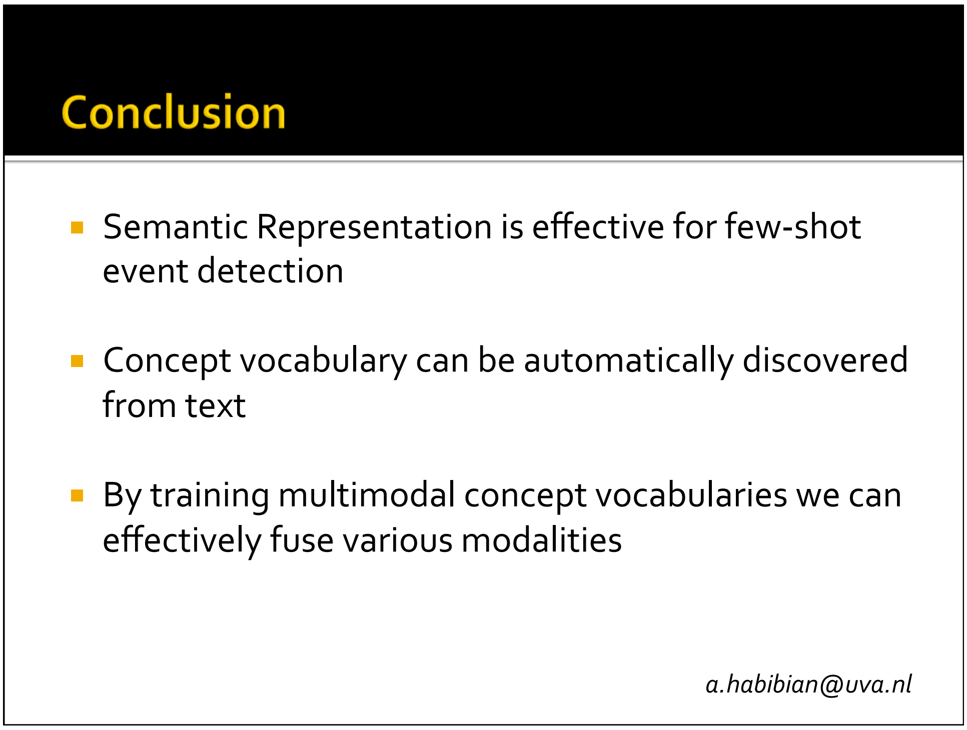


- 10 Most effective concepts for "Birthday party"
  - People (High-level)
  - Girl (MBH)
  - Sing (MFCC)
  - Celebrate (MBH)
  - Boy(High-level)
  - Surprise (MBH)
  - Kid (SIFT)
  - Group (SIFT)
  - Party (MFCC)
  - Indoor (High-level)

Results on progress set:
Event Detection



- Best visual event detection with just ten examples



Conclusion

- Semantic Representation is effective for few-shot event detection
- Concept vocabulary can be automatically discovered from text
- By training multimodal concept vocabularies we can effectively fuse various modalities

a.habibian@uva.nl