11-2015

Multimedia event detection: Strong by integration

Hao ZHANG

Maaike DE BOER

Yi-Jie LU

Klamer SCHUTTE

Chong-wah NGO
Singapore Management University, cwngo@smu.edu.sg

See next page for additional authors

Citation

ZHANG, Hao; DE BOER, Maaike; LU, Yi-Jie; SCHUTTE, Klamer; NGO, Chong-wah; and NGO, Chong-Wah. Multimedia event detection: Strong by integration. (2015). *2015 TREC Video Retrieval Evaluation, TRECVID 2015, Gaithersburg, November 16-18*.
Available at: https://ink.library.smu.edu.sg/sis_research/6491

Multimedia Event Detection: Strong by Integration

Hao ZHANG[1], Maaike de Boer[2]
Yijie Lu[1], Klamer Schutte[2], Wessel Kraaij[2], Chong-Wah Ngo[1]

[1]City University of Hong Kong

[2]TNO and Radboud University

November 24, 2015

VIREO Video Retrieval Group

TNO innovation for life

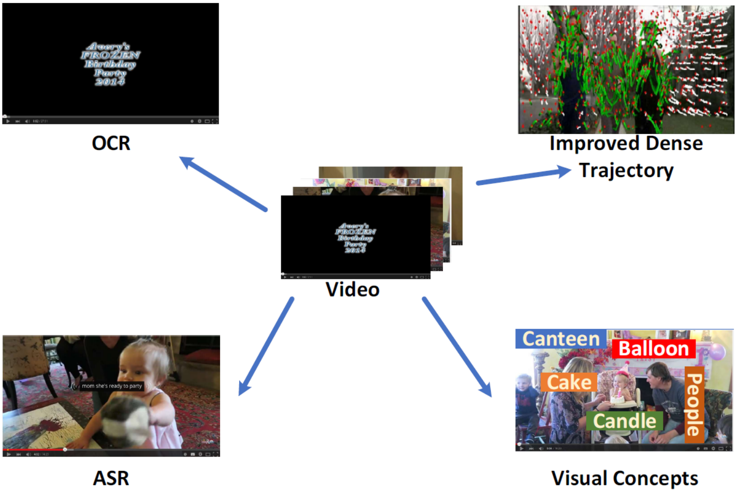
Radboud Universiteit Nijmegen

Overview

- Observations
- Modalities
- System
- Fusion: Joint Probability
- Fusion: Adding Zero-Shot
- Reranking: OCR/ASR
- Experiments: MED14_Test/MED15_Eval
- Conclusion

Observations

As is well known, multimedia event consists of multi-modalities: Audio, Motion, Visual, Texts ...



Observations

Multi-modalities: Audio, Motion, Visual, Texts ...

Observations

Multi-modalities: Audio, Motion, Visual, Texts ...

More efforts: single-modality.

Observations

Multi-modalities: Audio, Motion, Visual, Texts ...

More efforts: single-modality. e.g:

- Motion features: Dense Trajectories, Improved Dense Trajectories.
- Visual features: HOG, SIFT, Deep Features ...

Observations

Multi-modalities: Audio, Motion, Visual, Texts ...

More efforts: single-modality. e.g:

- Motion features: Dense Trajectories, Improved Dense Trajectories.
- Visual features: HOG, SIFT, Deep Features ...

Less efforts: integrate across modalities.

Problem:
Intergrating across modalities

Modalities

Problem:
Intergrating across modalities
Difficulties:

- Modalities have different meanings.
- Modalities have different precisions.

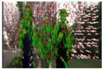
Modalities

OCR

ASR

Concepts

IDT



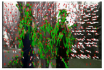
1. Semantics
2. High Precision
3. Relevant
4. Low Recall (rare)

OCR

ASR

Concepts

IDT



1. Semantics
2. High Precision (wanted)
3. Relevant/Irrelevant
4. Low Recall (rare)

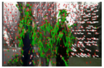
Modalities

OCR

ASR

Concepts

IDT



1. Semantics
2. Low Precision
4. Relevant/Irrelevant
5.High Recall (many)

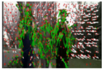
Modalities

OCR

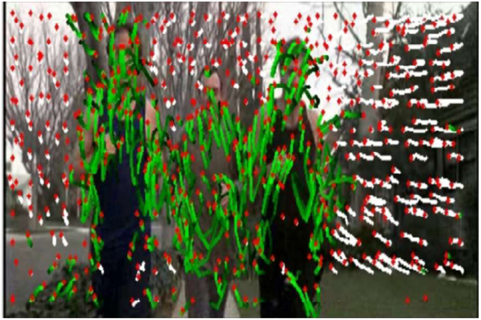
ASR

Concepts

IDT



1. Non-Semantics
2. Low precision
3. Relevant/Irrelevant
4. High Recall (many)



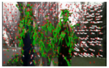
Modalities

OCR

ASR

Concepts

IDT



Reranking

1. Semantics
2. High Precision
3. Relevant
4. Low Recall

1. Semantics
2. High Precision
3. Relevant/Irrelevant
4. Low Recall

Fusion

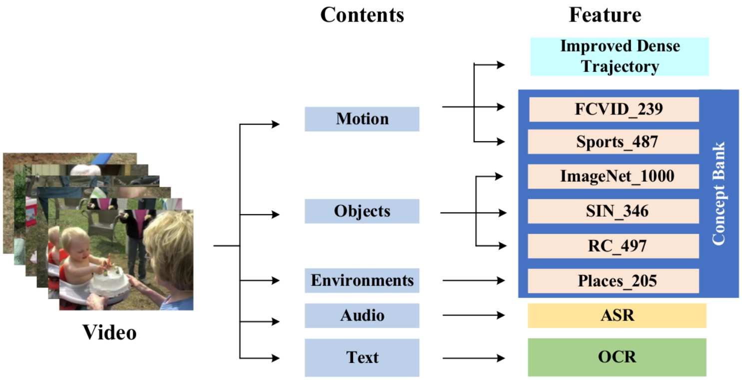
1. Semantics
2. Low Precision
3. Relevant/Irrelevant
4. High Recall

1. Non-Semantics
2. Low Precision
3. Relevant/irrelevant
4. High Recall

For Event Detection with 100Ex/10Ex:
An intergration system with multi-modalities.
We present 100Ex/10Ex as:

- Multi-modalities
- Different methods for different modalities
- Integration of modalities

Modalities



Video Retrieval Group

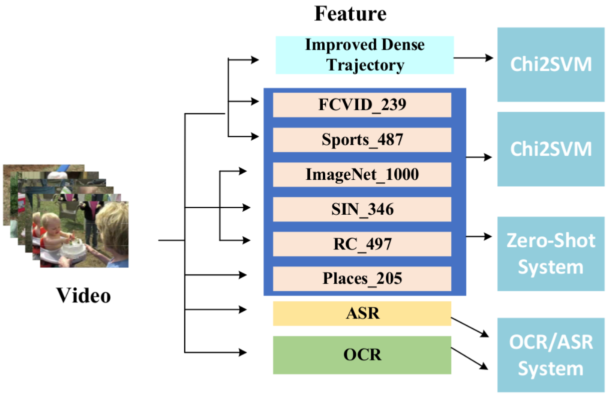
innovation for life

Radboud Universiteit Nijmegen

Concept Bank

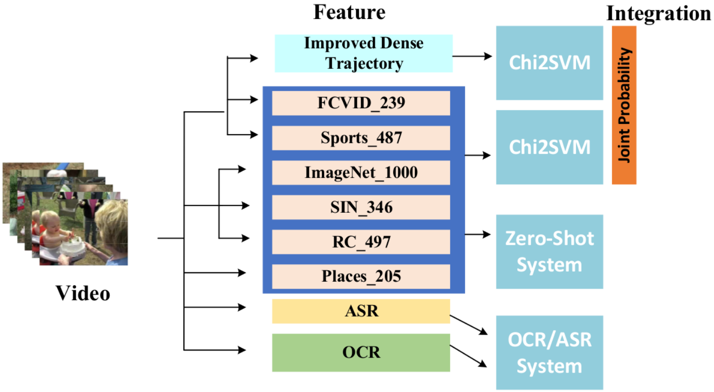
Feature	Dim	Structure	Dataset
Sports_487	487	3D-CNN	Sports-1M
ImageNet_1000	1000	DCNN	ImageNet
SIN_346	346	DCNN	TRECVID SIN
RC_487	487	DCNN	TRECVID Research Set
Places_205	205	DCNN	MIT Places
FCVID_239	239	SVM	Fudan-Columbia Dataset

Table : Concept Bank



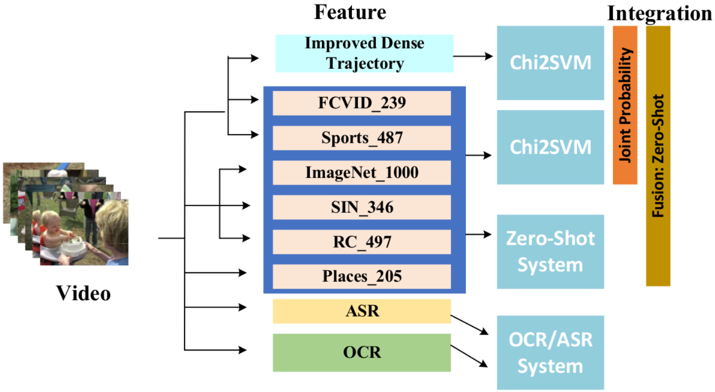
We propose three stages of fusion strategy, which can improve event detection step-by-step.

System



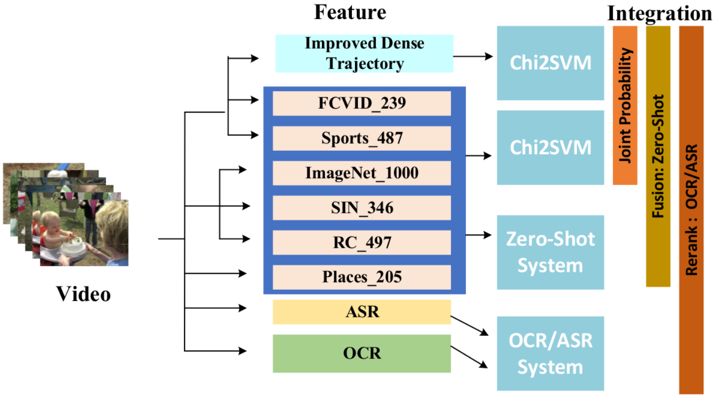
VIREO Video Retrieval Group   TNO innovation for life   Radboud Universiteit Nijmegen

System



VIREO Video Retrieval Group
TNO innovation for life
Radboud Universiteit Nijmegen

System

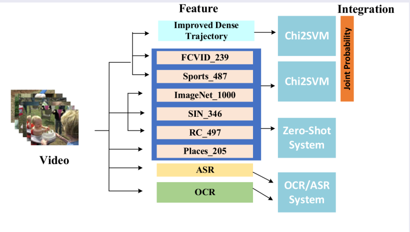


VIREO Video Retrieval Group

TNO innovation for life

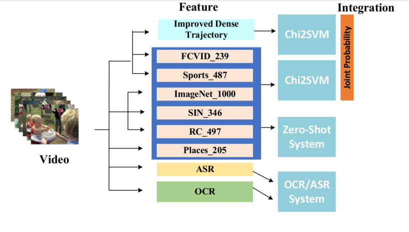
Radboud Universiteit Nijmegen

Fusion: Joint Probability

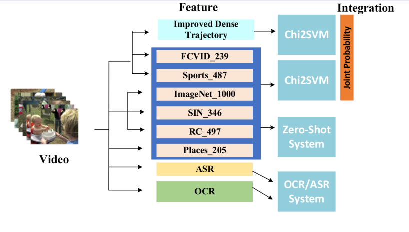


Classification:
Two classifiers make predicts independently.

Fusion: Joint Probability



Average:
A low score of one type of classifier downgrades a possibly relevant video.



Average:
A low score of one type of classifier downgrades a possibly relevant video.

Joint Probability:
Only videos that receive a low score from both classifiers will be put at the bottom of the ranking list.

$$JP = 1 - (1 - P_{CB}) \times (1 - P_{IDT})$$

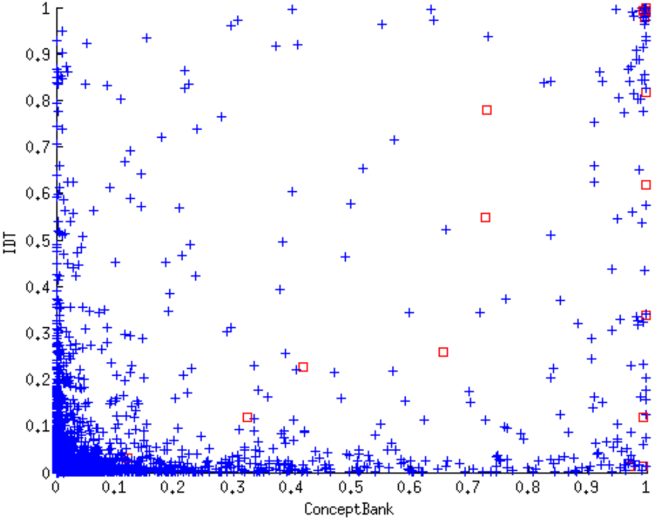
Video Retrieval Group

innovation for life

Radboud Universiteit Nijmegen

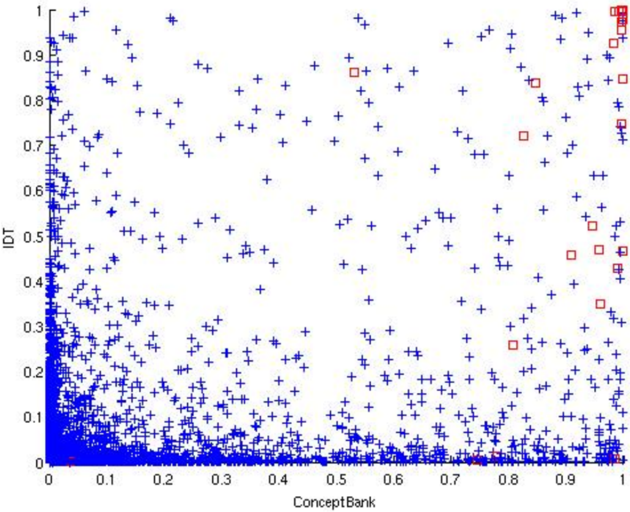
Fusion:Joint Probability

E021-SVM Prediction Scores with Concept feature and Improved Dense Trajectory



Fusion:Joint Probability

E039-SVM Prediction Scores with Concept feature and Improved Dense Trajectory



rsiteit Nijmegen

Fusion:Joint Probability

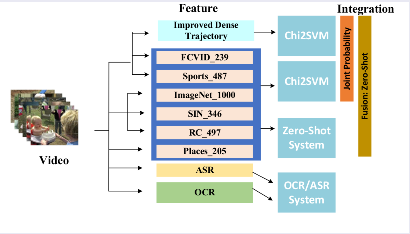
Contour Map

Average Fusion

Joint Probability

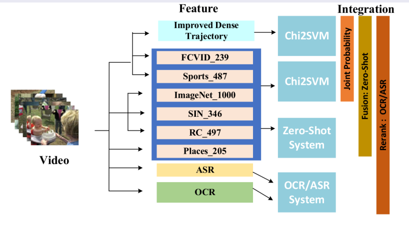
Fusion:Joint Probability

- Joint Probability is our first try to fuse two kinds of prediction scores by distributions of predicted scores.
- Based on the distributions of predicted scores, there might be more powerful unsupervised distribution-based fusion strategy.



Adding Zero-Shot:
We averaged scores predicted by the Zero-Shot system (the other PPT) with scores predicted by the event detectors (SVM).

Reranking: OCR/ASR



"Re-ranking":
Design high precision ASR and OCR systems for reranking.

Reranking: OCR/ASR

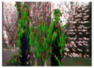
Recall OCR

OCR

ASR

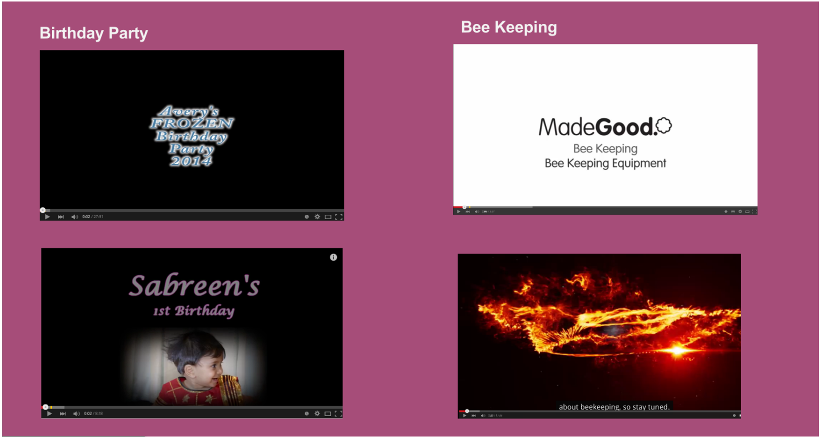
Concepts

IDT



1. Semantics
2. High Precision
3. Relevant
4. **Low Recall** (rare)

OCR Observations:



VIREO Video Retrieval Group

TNO innovation for life

Radboud Universiteit Nijmegen

Reranking: OCR/ASR

OCR Observations:

Tuning Musical Instruments

Bike Trick

Rock Climbing

Reranking: OCR/ASR

OCR Observations and Strategy:

- Parts of relevant videos were post-producted (include titles).

OCR Observations and Strategy:

- Parts of relevant videos were post-producted (include titles).
- Pick out these video by matching OCR and Query.

Reranking: OCR/ASR

OCR Observations and Strategy:

- Parts of relevant videos were post-producted (include titles).
- Pick out these video by matching OCR and Query.
- Rerank these videos with extra-bonus score, boosting their ranks.

Reranking: OCR/ASR

OCR Observations and Strategy:

- Parts of relevant videos were post-producted (include titles).
- Pick out these video by matching OCR and Query.
- Rerank these videos with extra-bonus score, boosting their ranks.

Same strategy is used for ASR,

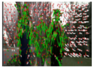
Drawbacks of ASR:

OCR

ASR

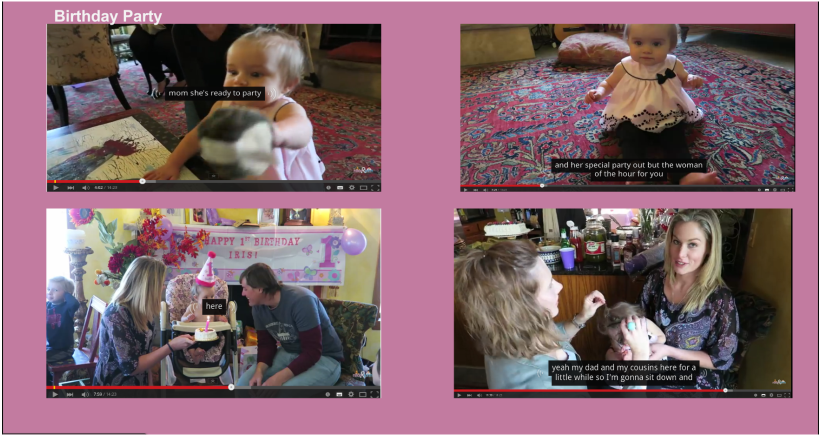
Concepts

IDT



1. **Semantics**
2. **High Precision** (wanted)
3. **Relevant/Irrelevant**
4. **Low Recall** (rare)

ASR Observations:



VIREO Video Retrieval Group TNO innovation for life Radboud Universiteit Nijmegen

ASR Observations:

Video Retrieval Group

innovation for life

Radboud Universiteit Nijmegen

Reranking: OCR/ASR

ASR Observations:

- The portion of relevant ASR results is small.

Reranking: OCR/ASR

ASR Observations:

- The portion of relevant ASR results is small.
- The portion of irrelevant ASR resuts is large.

Reranking: OCR/ASR

ASR Observations:

- The portion of relevant ASR results is small.
- The portion of irrelevant ASR resuts is large.
- Mining event relevance with ASR is still an open topic.

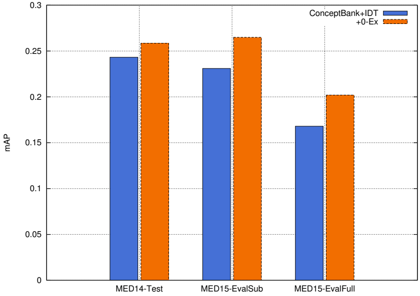
The indexing and search tool Lucene is used for the OCR and ASR data. High precision is retrieved by:

- OCR: manually defining a Boolean Query using the event description and Wikipedia and some information on known common mistakes from the Tesseract tool (e.g. zero (0) and O).
- ASR: manually defining a Boolean Query and adding a PhraseQuery so the words in the query do not occur more than five words from each other. Only the words specific for the event are added.

Based on internal test, we have the following settings for MED 2015 Submission:

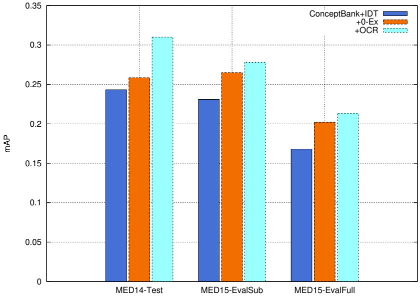
- 10 Exemplars:
  Adding Zero-Shot, Reranking by OCR/ASR
- 100 Exemplars:
  Joint Probability, Reranking by OCR/ASR

MED PS_10-Ex: Mean AP of fusion strategies on MED14-Test/EvalSub/Full



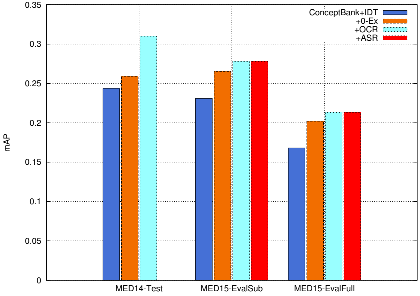
For 10 exemplars, adding the results of Zero-Shot case does really improve performance (more than 3%)

MED PS_10-Ex: Mean AP of fusion strategies on MED14-Test/EvalSub/Full



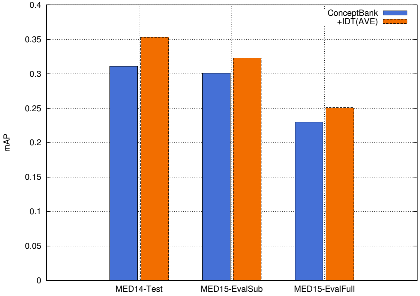
OCR gives an improvement of 1.2%.

MED PS_10-Ex: Mean AP of fusion strategies on MED14-Test/EvalSub/Full



ASR slightly decreases performance in the Evaluation Set. This is probably because the precision of our ASR system is not as high as our OCR system.

MED PS_100-Ex: Mean AP of fusion strategies on MED14-Test/EvalSub/Full



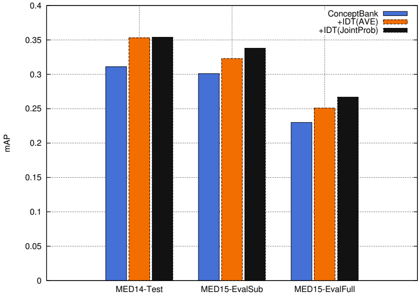
IDT increases overall performance (2%-4%)

Video Retrieval Group

innovation for life

Radboud Universiteit Nijmegen

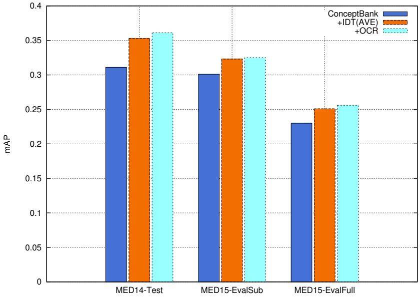
MED PS_100-Ex: Mean AP of fusion strategies on MED14-Test/EvalSub/Full



Joint Probability is better than average fusion, providing for an additional improvement of 1%.

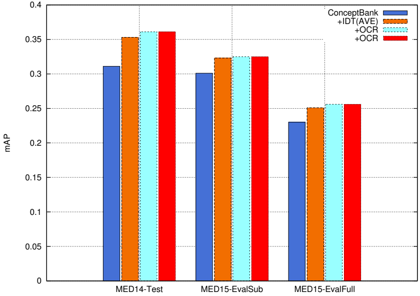
VIREO Video Retrieval Group TNO innovation for life Radboud Universiteit Nijmegen

MED PS_100-Ex: Mean AP of fusion strategies on MED14-Test/EvalSub/Full



Adding OCR gives a small improvement.

MED PS_100-Ex: Mean AP of fusion strategies on MED14-Test/EvalSub/Full



ASR slightly decreases performance as with the 10Ex Experiments.

Conclusion

- For the 10 Ex case, fusion of the system trained on 10 examples and the zero-shot case improves a lot.
- Fusion with OCR slightly improves performance in all runs. Because the precision of ASR system is not as high as OCR system, performance drops a bit by adding ASR.
- Improved Dense Trajectory improves performance, especially with more training data (100 Ex VS 10 Ex).
- Using Joint Probability of concept features and IDT improves performance on 100 Ex task.

Video Retrieval Group

innovation for life

Radboud Universiteit Nijmegen