Web-scale Multimedia Search for Internet Video Content

Lu Jiang
Language Technologies Institute,
Carnegie Mellon University
April 27th 2017.

Thesis Committee:

Dr. Alex Hauptmann (Co-chair), Carnegie Mellon University
Dr. Teruko Mitamura (Co-chair), Carnegie Mellon University
Dr. Louis-Philippe Morency, Carnegie Mellon University
Dr. Tat-Seng Chua, National University of Singapore

Outline

- Introduction
- Review of Finished Work in Proposal
- Proposed Work:
  - Robust learning in weakly-labeled data
  - Deep visual query understanding & hybrid search
  - Visual MemexQA
- Conclusions

Outline

- **Introduction**
- Review of Finished Work in Proposal
- Proposed Work:
  - Robust learning in weakly-labeled data
  - Deep visual query understanding & hybrid search
  - Visual MemexQA
- Conclusions

Introduction

- We are living in an era of big multimedia data:
  - social media users are posting 12 million videos on Twitter every day;
  - video will account for 80% of all the world's internet traffic by 2019.

- Videos are becoming a valuable source for acquiring information and knowledge.

Introduction

80% personal photos or videos do not have textual metadata [Jiang e t al. 2017]

Introduction

- We are living in an era of big multimedia data:
  - social media users are posting 12 million videos on Twitter every day;
  - video will account for 80% of all the world's internet traffic by 2019.

- Videos are becoming a valuable source for acquiring information and knowledge.

How to acquire information or knowledge in video if there is no way to find it?

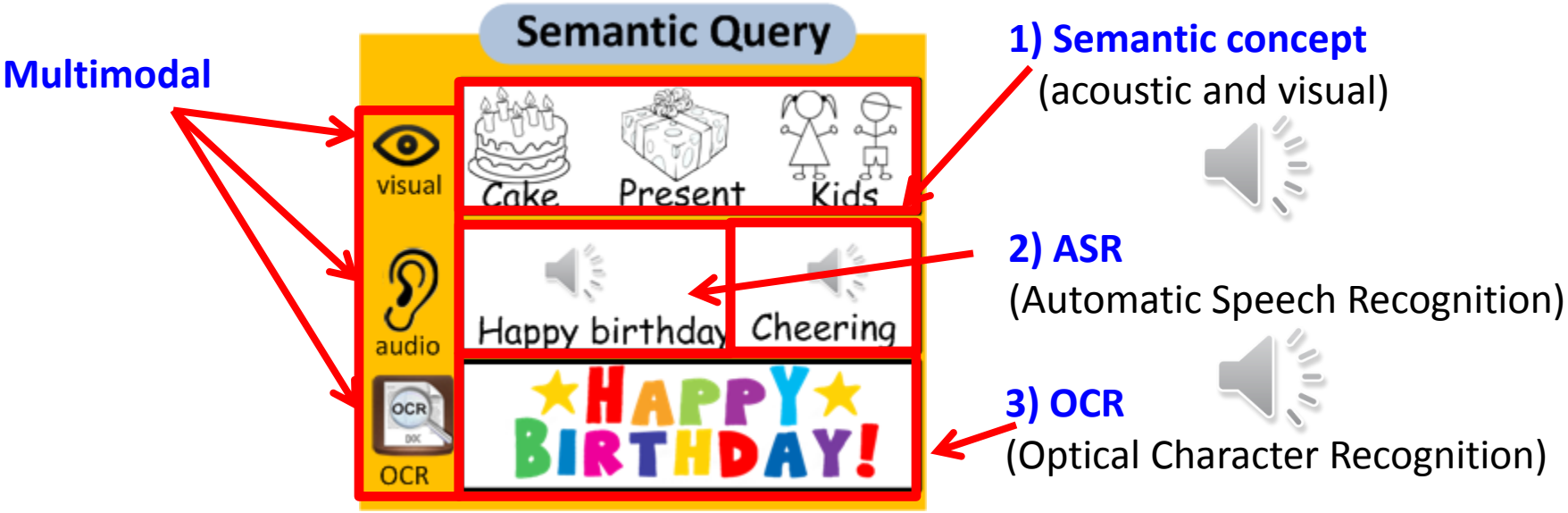
Introduction

- A fundamental research question:
  how to satisfy the information need about the **video content** at a very large scale?

- Two types of queries for video search: semantic query and hybrid query.

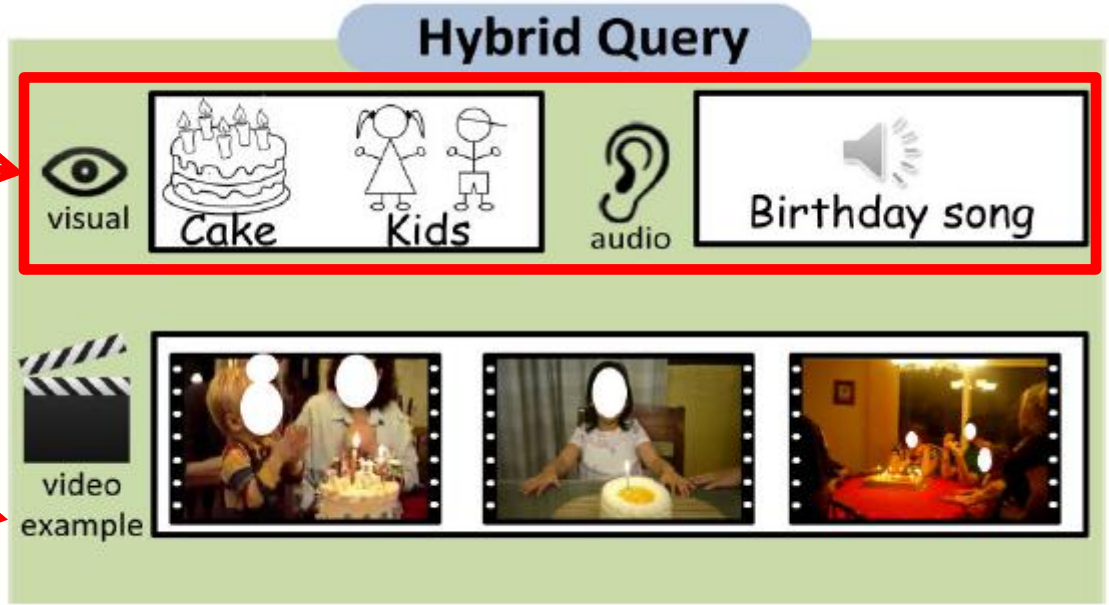
Semantic Query:

Information need:
Find videos about birthday party.



text-to-video search

Hybrid Query:



semantic query

video examples provided by users

Evaluation Benchmarks

- Problem: detect videos without metadata
- Initiated by a National Institute of Standards and Technology (NIST) task Multimedia Event Detection (MED) in 2012 (common evaluation benchmark).
  - Supported by 5-year multi-million IARPA project .
  - 20+ participants across the world → challenging problem.

From large-scale to web-scale

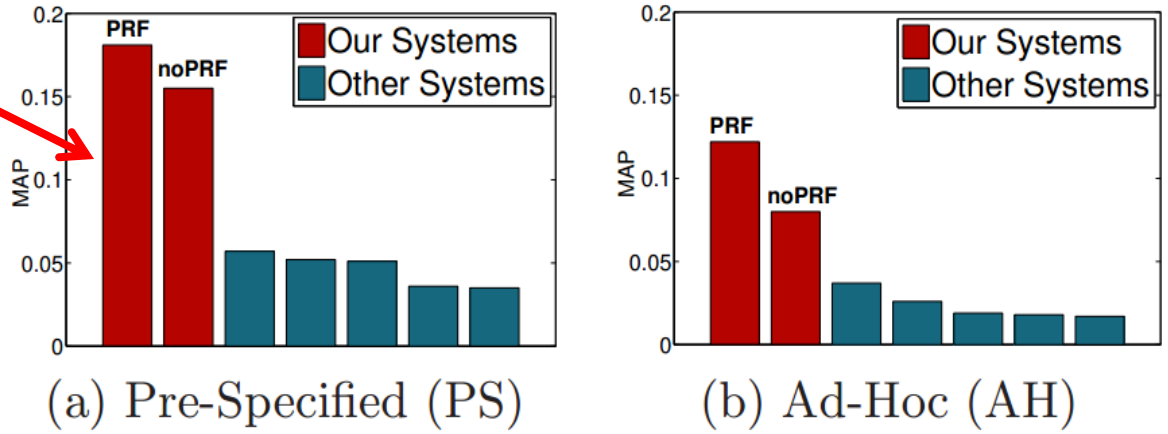
Let the above videos represent the upper-bound of the current largest dataset for this problem (200K videos)

(From Large-scale to Web-scale)

Result Overview

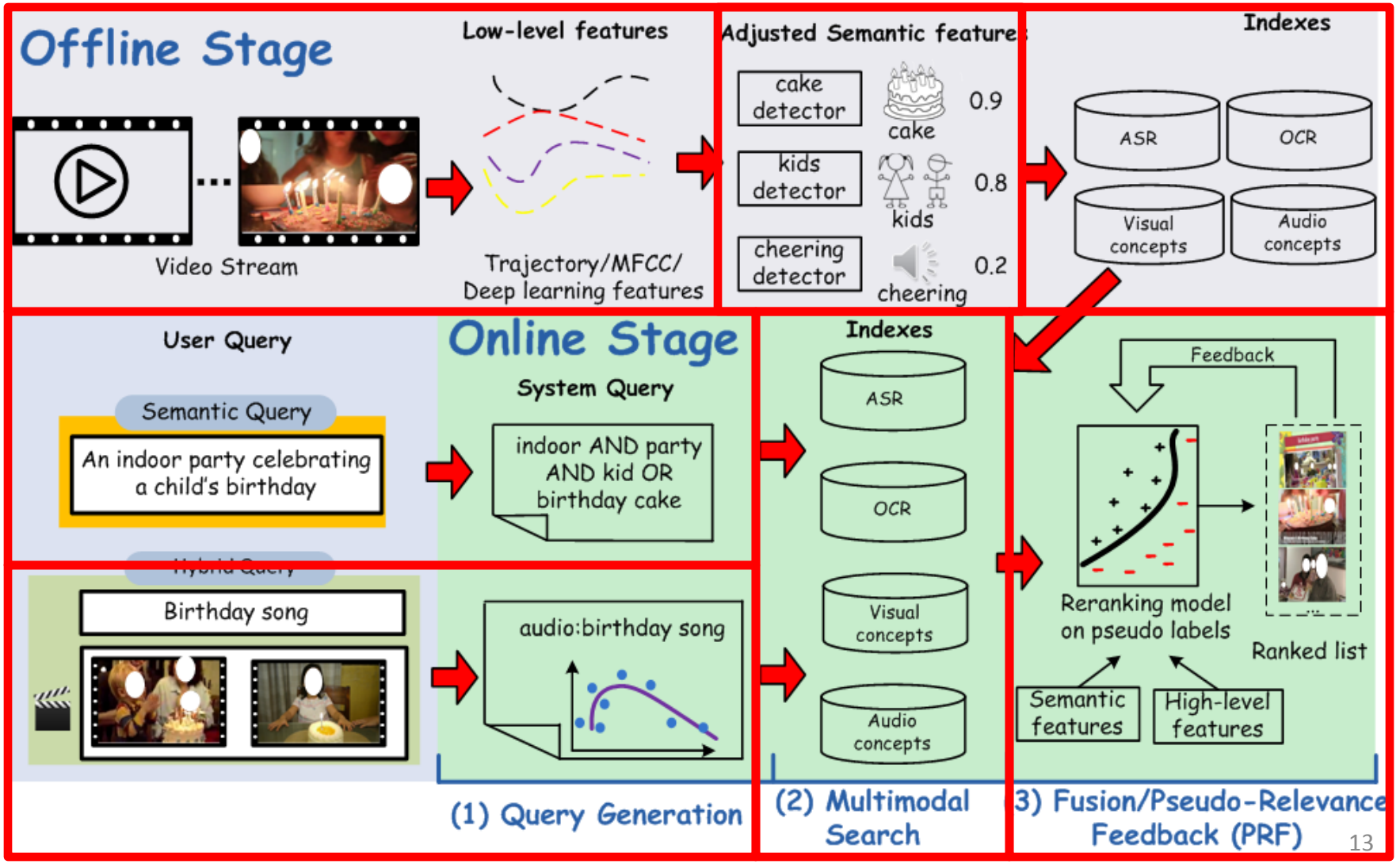
- We proposed a novel and practical solution that can:
  - substantially boost the state-of-the-art accuracy.
  - Scale up the search to hundreds of millions of Internet videos.
    - 0.2 second to process a semantic query on 100 million videos
    - Less than 1 second for hybrid query on millions of videos
- We implemented a large-scale multimedia engine for understanding and search the Internet videos:
  - Achieved the **best accuracy** in NIST TRECVID zero-example video search 2013, 2014 and 2015.

E-Lamp Lite



The official results released by NIST TRECVID on MED 2014 (on 200,000 videos)

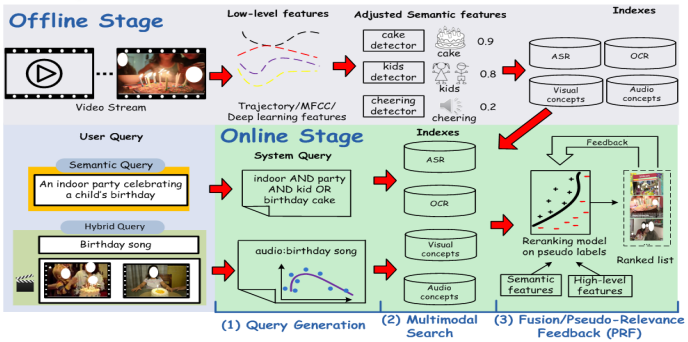
Framework


Offline Stage
Low-level features
Video Stream
Trajectory/MFCC/
Deep learning features
Adjusted Semantic features
cake detector
cake
0.9
kids detector
kids
0.8
cheering detector
cheering
0.2
Indexes
ASR
OCR
Visual concepts
Audio concepts
User Query
Semantic Query
An indoor party celebrating a child's birthday
Online Stage
System Query
indoor AND party AND kid OR birthday cake
Birthday song
audio:birthday song
Indexes
ASR
OCR
Visual concepts
Audio concepts
Feedback
Reranking model on pseudo labels
Ranked list
Semantic features
High-level features
(1) Query Generation
(2) Multimodal Search
(3) Fusion/Pseudo-Relevance Feedback (PRF)

Applications

in-video ads

content search and recommendation



Video Content Engine

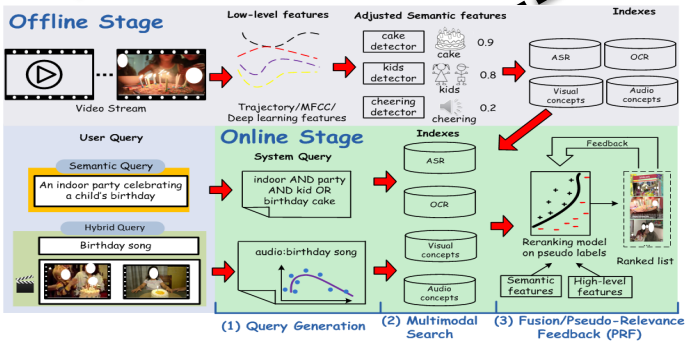
Applications

content search and recommendation

in-video ads

MemexQA

question answering



Video Content Engine

Key Contributions: First-of-its-kind Framework

- The first-of-its-kind framework for web-scale content-based search over hundreds of millions of Internet videos [ICMR'15]. The proposed framework supports text-to-video, video-to-video, and text&video-to-video search [MM'12, WSDM'17]. **(Chapter 1 and 5)**

[ICMR15] Lu Jiang, Shoou-I Yu, Deyu Meng, Teruko Mitamura, Alexander Hauptmann. Bridging the Ultimate Semantic Gap: A Semantic Search Engine for Internet Videos. In *ACM International Conference on Multimedia Retrieval (ICMR)*, 2015.
[MM12] Lu Jiang, Alexander Hauptmann, Guang Xiang. Leveraging High-level and Low-level Features for Multimedia Event Detection. *In ACM Multimedia (MM)*, 2012.
[WSDM17] Lu Jiang, Yannis Kalantidis, Liangliang Cao, Sachin, Farfade, Jiliang Tang, Alex Hauptmann. Delving Deep into Personal Photo and Video Search. *In In ACM International Conference on Web Search and Data Mining (WSDM)*, 2017.

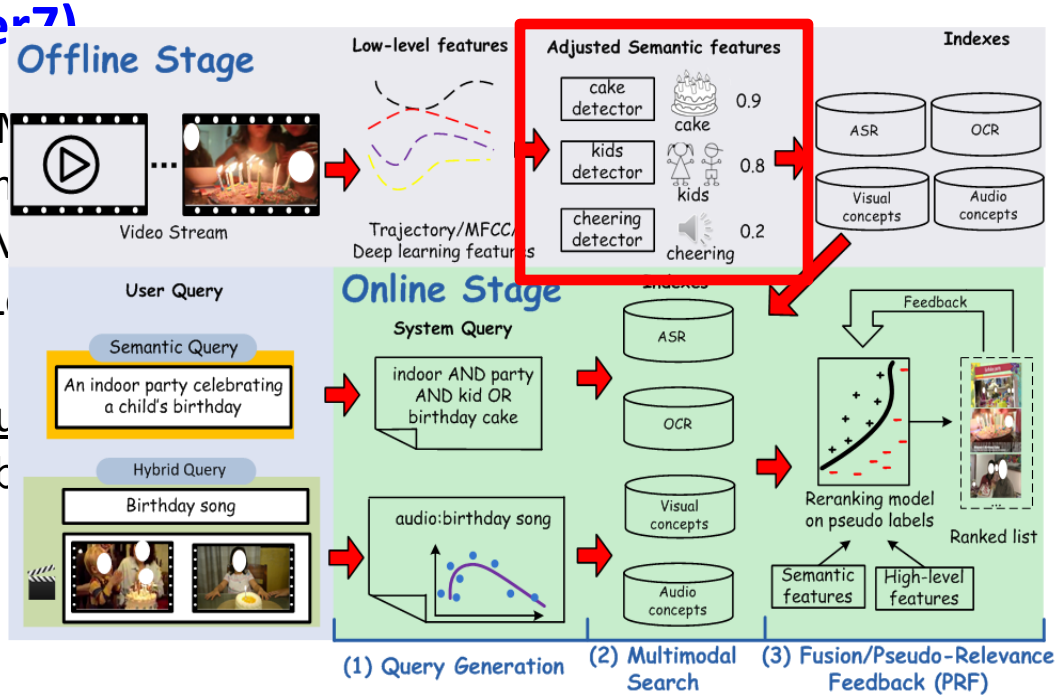
Key Contributions: Self-paced curriculums learning theory

- The first-of-its-kind framework for web-scale content-based search over hundreds of millions of Internet videos [ICMR'15]. The proposed framework supports text-to-video, video-to-video, and text&video-to-video search [MM'12, WSDM'17].
- A novel theory about self-paced curriculums learning and its application on robust concept detector training [NIPS'14, AAAI'15, IJCAI'16]. **(Chapter7)**

[AAAI15] Lu Jiang, Deyu Meng, Qian Zhao, Shiguang Shan, Alexander Hauptmann. Self-paced Curriculum Learning. *In Conference on Artificial Intelligence (AAAI)*, 2015.
[NIPS14] Lu Jiang, Deyu Meng, Shoou-I Yu, Zhen-Zhong Lan, Shiguang Shan, Alexander Hauptmann. Self-paced Learning with Diversity. *In Neural Information Processing Systems (NIPS)*, 2014.
[IJCAI16] Junwei Liang, Lu Jiang, Deyu Meng, Alexander Hauptmann. Learning to Detect Concepts from Webly-Labeled Video Data. *In Joint Conference on Artificial Intelligence (IJCAI)*, 2016.

Key Contributions: Self-paced curriculums learning theory

- The first-of-its-kind framework for web-scale content-based search over hundreds of millions of Internet videos [ICMR'15]. The proposed framework supports text-to-video, video-to-video, and text&video-to-video search [MM'12, WSDM'17].
- A novel theory about self-paced curriculums learning and its application on robust concept detector training [NIPS'14, AAAI'15, IJCAI'16]. **(Chapter 7)**



[AAAI15] Lu Jiang, Deyu M … ın. Self-paced Curriculum Learnir…

[NIPS14] Lu Jiang, Deyu M… xander Hauptmann. Self-paced L… *g Systems (NIPS)*, 2014.

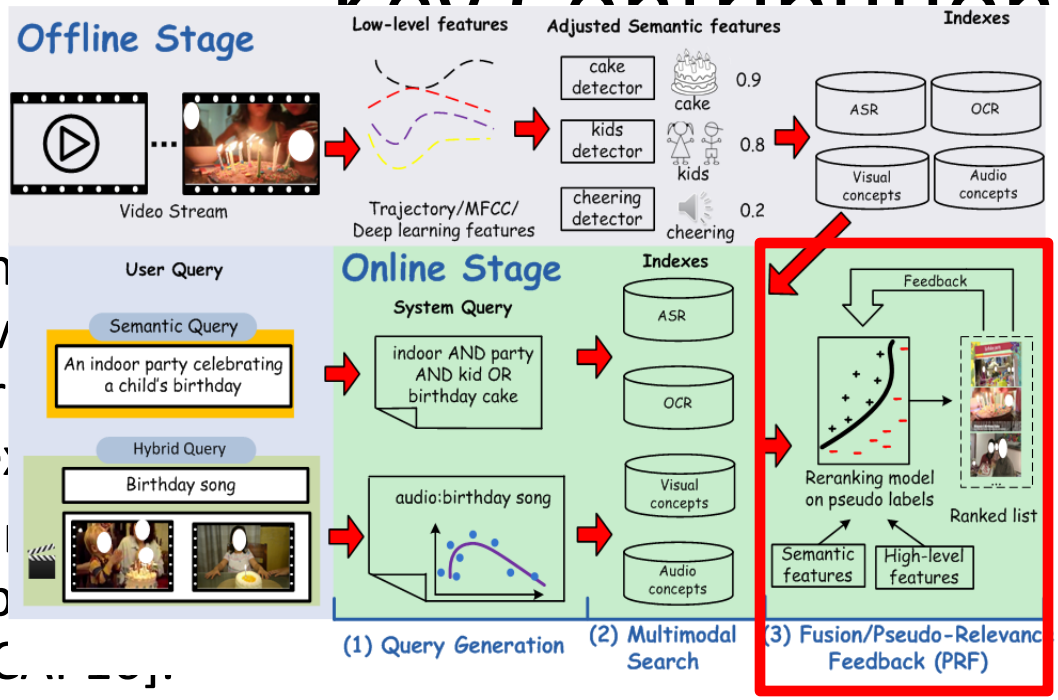
[IJCAI16] Junwei Liang, Lu… ɔ Detect Concepts from Webly-Lab… *igence (IJCAI)*, 2016.

Key Contributions: Reranking

- The first-of-its-kind framework for web-scale content-based search over hundreds of millions of Internet videos [ICMR'15]. The proposed framework supports text-to-video, video-to-video, and text&video-to-video search [MM'12, WSDM'17].
- A novel theory about self-paced curriculums learning and its application on robust concept detector training [NIPS'14, AAAI'15, IJCAI'16].
- Novel reranking algorithms for improving performance. They have concise mathematical objectives to optimize and useful properties that can be theoretically verified [MM'14, ICMR'14]. **(Chapter6)**

[MM14] Lu Jiang, Deyu Meng, Teruko Mitamura, Alexander Hauptmann. Easy Samples First: Selfpaced Reranking for Zero-Example Multimedia Search. *In ACM Multimedia (MM)*, 2014.
[ICMR14] Lu Jiang, Teruko Mitamura, Shoou-I Yu, Alexander Hauptmann. Zero-Example Event Search using MultiModal Pseudo Relevance Feedback. *In ACM International Conference on Multimedia Retrieval (ICMR)*, 2014.

Key Contributions:



- Th ontent-based search ov MR'15]. The pr deo-to-video, and te].
- A arning and its ap ; [NIPS'14, AAAI'15, IJC ..].
- Novel reranking algorithms for improving performance. They have concise mathematical objectives to optimize and useful properties that can be theoretically verified [MM'14, ICMR'14]. **(Chapter6)**

[MM14] Lu Jiang, Deyu Meng, Teruko Mitamura, Alexander Hauptmann. Easy Samples First: Selfpaced Reranking for Zero-Example Multimedia Search. *In ACM Multimedia (MM)*, 2014.

[ICMR14] Lu Jiang, Teruko Mitamura, Shoou-I Yu, Alexander Hauptmann. Zero-Example Event Search using MultiModal Pseudo Relevance Feedback. *In ACM International Conference on Multimedia Retrieval (ICMR)*, 2014.

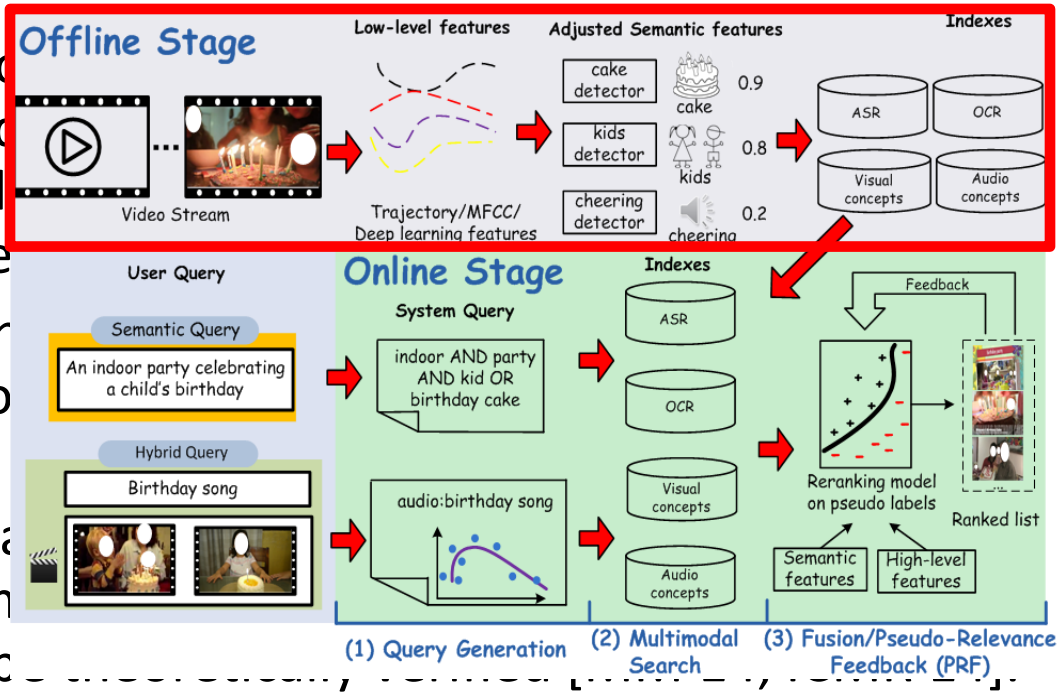
Key Contributions: Scalable Indexing Method

- The first-of-its-kind framework for web-scale content-based search over hundreds of millions of Internet videos [ICMR'15]. The proposed framework supports text-to-video, video-to-video, and text&video-to-video search [MM'12, WSDM'17].
- A novel theory about self-paced curriculums learning and its application on robust concept detector training [NIPS'14, AAAI'15, IJCAI'16].
- Novel reranking algorithms for improving performance. They have concise mathematical objectives to optimize and useful properties that can be theoretically verified [MM'14, ICMR'14].
- A concept adjustment method representing a video by a few salient and consistent concepts that can be efficiently indexed by the modified inverted index [MM'15] **(Chapter3)**

[MM15] Lu Jiang, Shoou-I Yu, Deyu Meng, Yi Yang, Teruko Mitamura, Alexander Hauptmann. Fast and Accurate Content-based Semantic Search in 100M Internet Videos. *In ACM Multimedia (MM)*, 2015

Key Contributions: Scalable Indexing Method

- The first-
- A novel th
- Novel rera



- A concept adjustment method representing a video by a few salient and consistent concepts that can be efficiently indexed by the modified inverted index [MM'15] **(Chapter3)**

[MM15] Lu Jiang, Shoou-I Yu, Deyu Meng, Yi Yang, Teruko Mitamura, Alexander Hauptmann. Fast and Accurate Content-based Semantic Search in 100M Internet Videos. *In ACM Multimedia (MM)*, 2015

Work Published After Proposal (October 15 2015)

- [1] Lu Jiang, Liangliang Cao, Yannis Kalantidis, Sachin, Farfade, Alex Hauptmann. Visual Memory QA: Your Personal Photo and Video Search Agent (Demo Paper). **AAAI**, 2017.
- [2] Lu Jiang, Yannis Kalantidis, Liangliang Cao, Sachin, Farfade, Jiliang Tang, Alex Hauptmann. Delving Deep into Personal Photo and Video Search. In Web Search and Data Mining(**WSDM**), 2017.
- [3] Lu Jiang. Web-scale Multimedia Search for Internet Video Content. In International Conference on World Wide Web (**WWW**), 2016.
- [4] Junwei Liang, Lu Jiang, Deyu Meng, Alexander Hauptmann. Learning to Detect Concepts from Webly-Labeled Video Data. In Joint Conference on Artificial Intelligence (**IJCAI**), 2016.

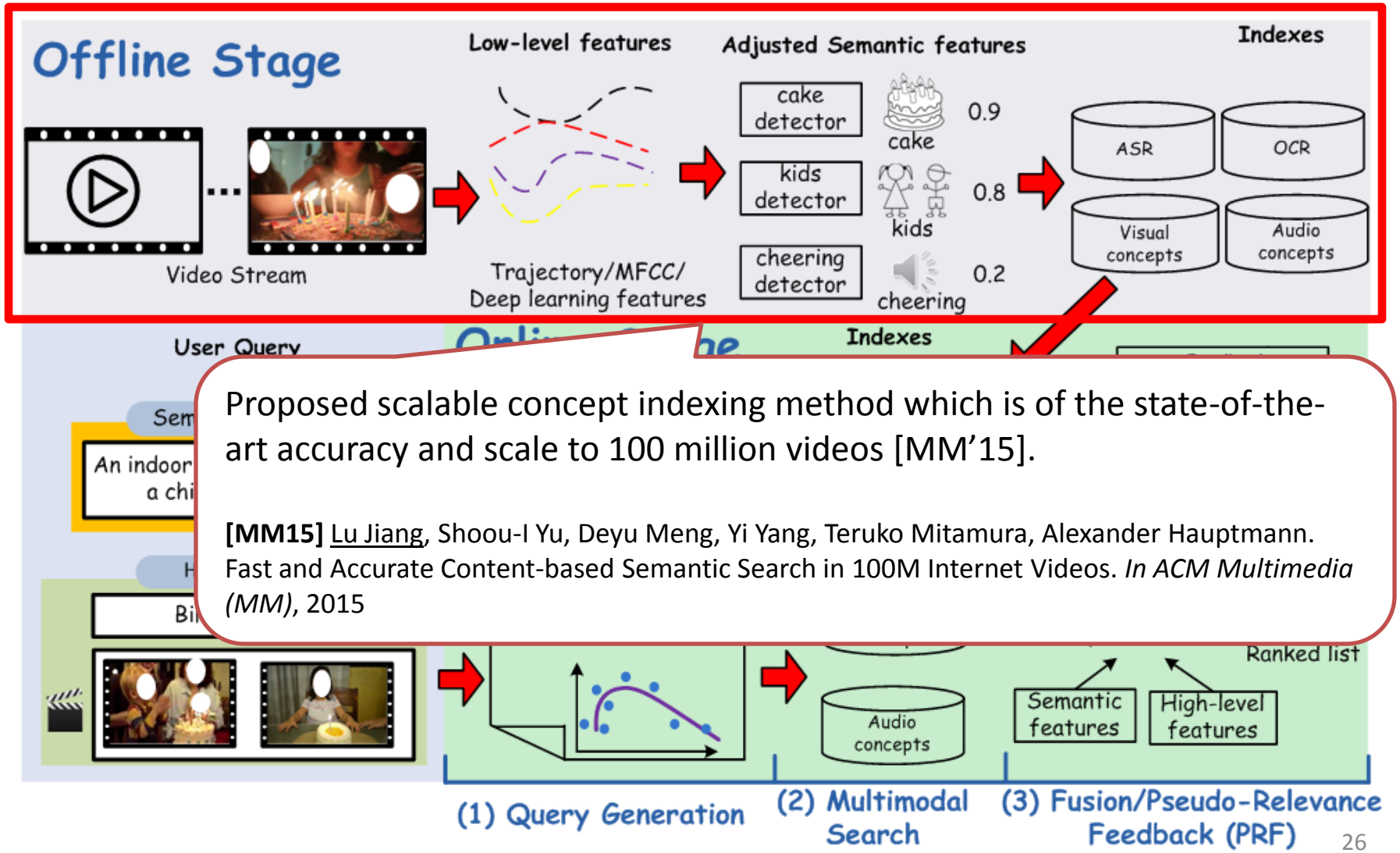
Thesis Statement

- In this thesis, we approach a fundamental problem of searching information in video content at a very large scale. We address the problem by proposing an accurate, efficient, and scalable method that can search the content of billions of videos by semantic visual/acoustic concepts, speech, visible texts, video examples, or any combination of these elements.

Outline

- Introduction
- **Review of Finished Work in Proposal**
- Proposed Work:
  - Robust learning in weakly-labeled data
  - Deep visual query understanding & hybrid search
  - Visual MemexQA
- Conclusions

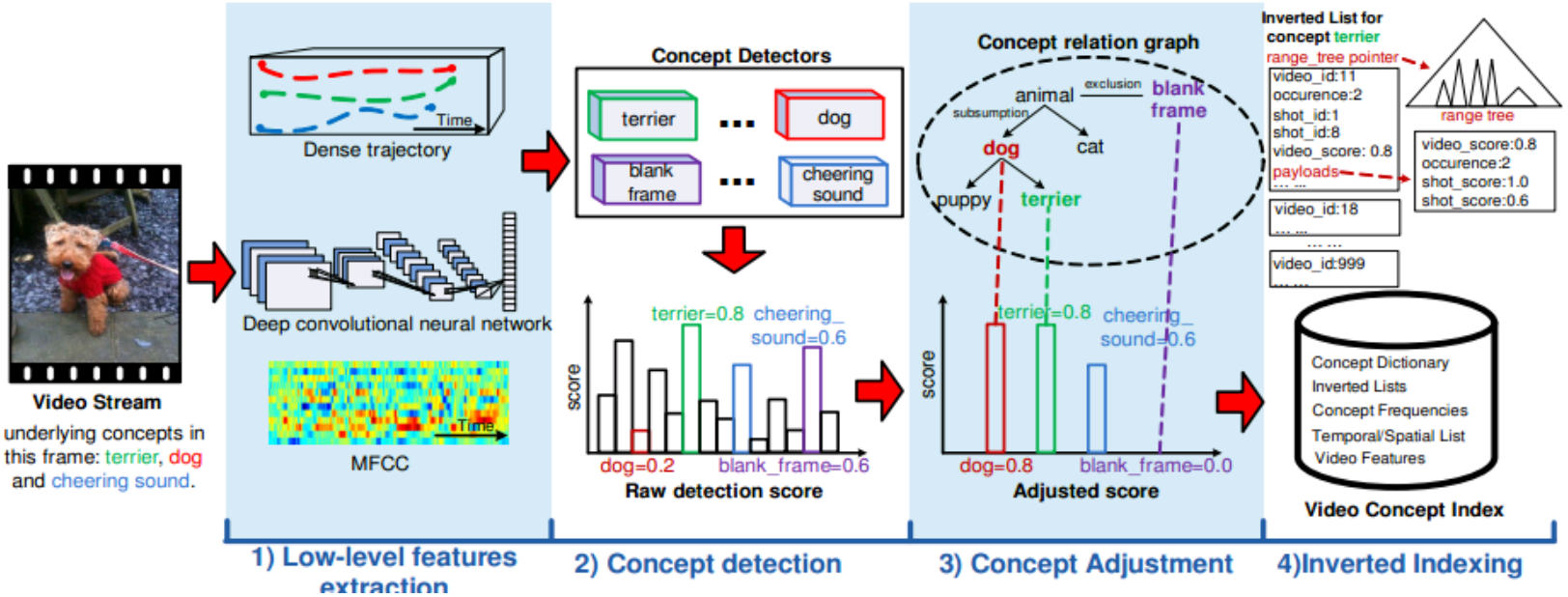
Framework



Proposed scalable concept indexing method which is of the state-of-the-art accuracy and scale to 100 million videos [MM'15].

[MM15] Lu Jiang, Shoou-I Yu, Deyu Meng, Yi Yang, Teruko Mitamura, Alexander Hauptmann. Fast and Accurate Content-based Semantic Search in 100M Internet Videos. *In ACM Multimedia (MM)*, 2015

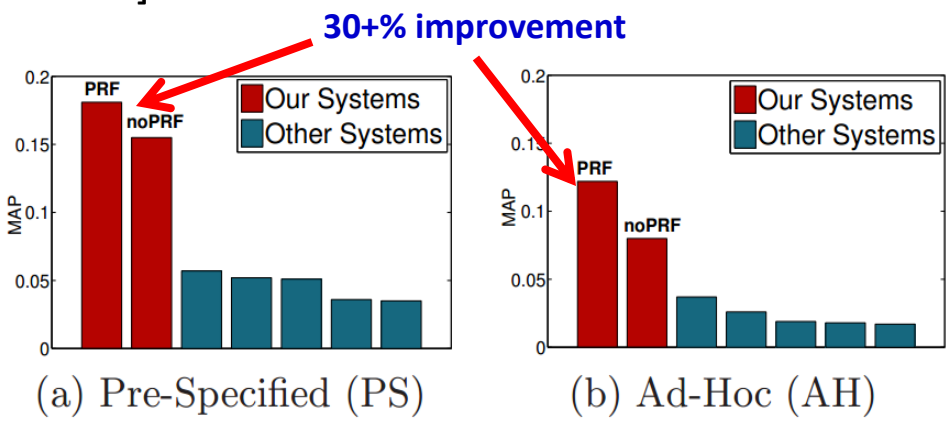
Method Overview



- concept adjustment represents a video by **a few salient and logically consistent visual/audio concepts**.
- Allows for fast search on **100M videos**.

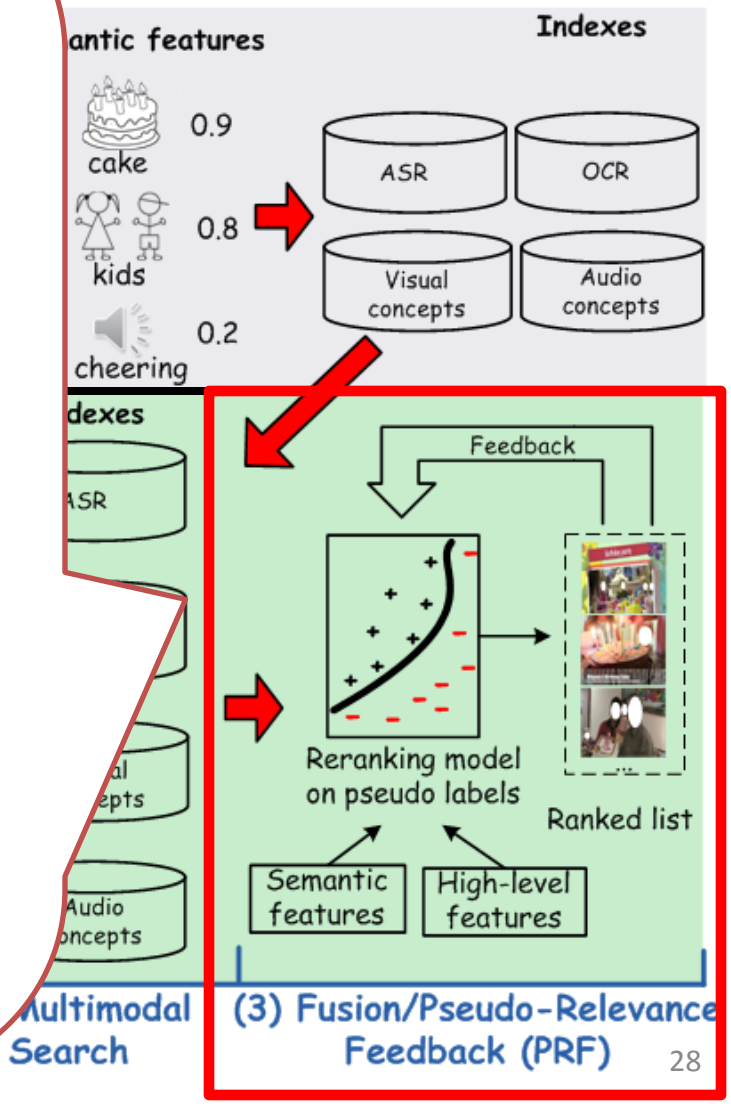
Framework

Novel reranking (pseudo relevance feedback) algorithms for improving performance [MM'14, ICMR'15].



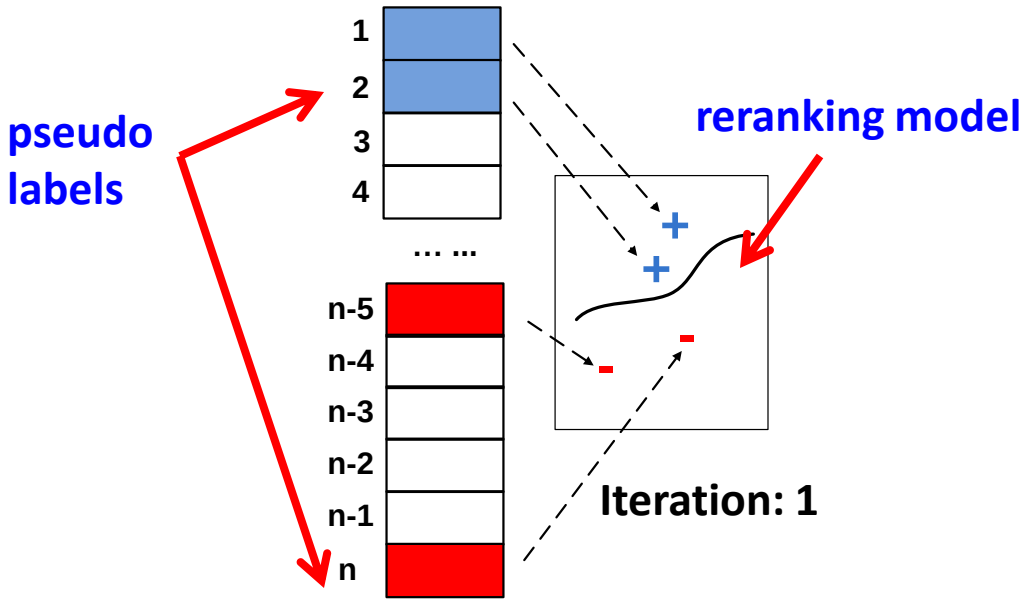
[ICMR15] Lu Jiang, Shoou-I Yu, Deyu Meng, Teruko Mitamura, Alexander Hauptmann. Bridging the Ultimate Semantic Gap: A Semantic Search Engine for Internet Videos. In ACM International Conference on Multimedia Retrieval (ICMR), 2015. **[best paper candidate]**

[MM14] Lu Jiang, Deyu Meng, Teruko Mitamura, Alexander Hauptmann. Easy Samples First: Selfpaced Reranking for Zero-Example Multimedia Search. In ACM Multimedia (MM), 2014.



Generic Reranking Algorithm

1: $t = 0$; //Iteration zero
2: Choose the initial pseudo labels and weights;
3: **while** $t \leq$ max iteration **do**
4: Train a reranking model on the fixed labels and weights;
5: Update the pseudo labels and weights;
6: **if** t is small **then** add more pseudo positives;
7: **end while**
8: **return** The list of samples after reranking;

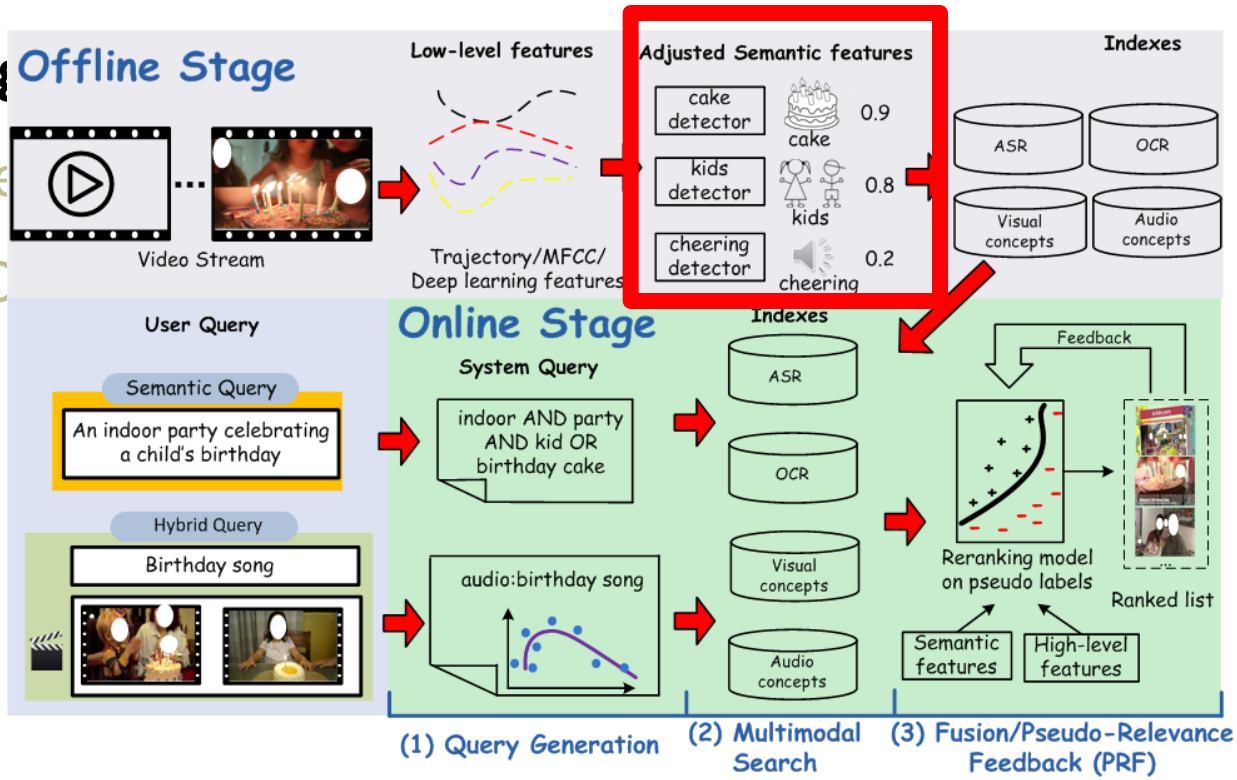


Outline

- Introduction
- Review of Finished Work in Proposal
- Proposed Work:
  - **Robust learning in weakly-labeled data**
  - Deep visual query understanding & hybrid search
  - Visual MemexQA
- Conclusions

Outline

- Introduction
- Review of Finished Work in Proposal
- Proposed Work:
  - **Robust learnin**
  - Deep visual que
  - Visual MemexC
- Conclusions



Motivation

- Training concept detectors need lots of labeled training data. Annotated video data are expensive to collect.

Funny Dogs - A Funny Dog Videos Compilation 2015
MashupZone
1 year ago • 8,732,858 views
Check out these funny videos of funny **dogs** and funny puppies. It ha...

Dogs, man's best and funniest friends - funny dog compilation
Tiger FunnyWorks
2 months ago • 1,974,967 views
It is never boring with our furry besties! They just don't fail to make...

- Our solution is to train detectors from weakly labeled video data (metadata) downloaded from the Internet.
  - Pros: no manual annotations. Cons: very noisy
- We are interested in approaching this problem in a principled way.

Curriculum Learning and Self-paced Learning

Learning philosophy [Bengio et al. 2009, Kumar et al. 2010]:

- Learning is an iterative process.
- Samples should be organized in a meaningful order (**called curriculum**).
- Model complexity increases in each iteration.

"dog" to learn earlier

Age

"dog" to learn later

Curriculum Learning and Self-paced Learning

- **Curriculum Learning (CL)**: assign learning priorities to training samples, according to prior knowledge or heuristics about specific problems [Bengio et al. 2009].
  - parsing from shorter sentences to longer sentences [Spitkovsky et al. 2009].
  - From clean to noisy background images for object detection [Chen et al. 2015]
- **Self-paced Learning (SPL):** the curriculum is determined by the learned models. Solving a joint optimization problem of the learning objective with the latent curriculum [Kumar, Packer, and Koller 2010].
  - Broadly used in many learning problems such as tracking[Supancicet al. 2013], domain adaptation [Tang et al. 2012], segmentation [Kumar et al. 2011], etc.

Self-paced Learning

- Formulated as an optimization problem (based on SPL).

$$\arg \min_{\mathbf{w},\mathbf{v}\in[0,1]^n} \sum_{i=1}^{n} v_i L(y_i, g(\mathbf{x}_i, \mathbf{w})) - \lambda \sum_{i=1}^{n} v_i$$

Learner
e.g. SVMs, neural network models

$\mathbf{w} \Rightarrow$ parameters in the off-the-shelf model

$L(y_i, g(\mathbf{x}_i, w)) \Rightarrow$ loss for the i*th* sample

Off-the-shelf model (SVM, neural networks etc.)

$\mathbf{v} = [v_1, \ldots, v_n] \Rightarrow$ latent weight vector for all samples

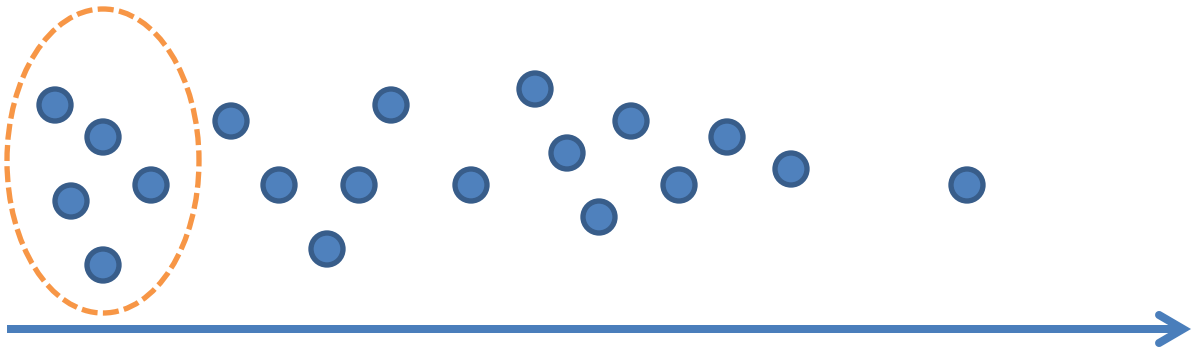
Optimization Algorithm

Algorithm 1: Self-paced Curriculum Learning.

input : Input dataset $\mathcal{D}$, predetermined curriculum γ, self-paced function f and a stepsize μ
output: Model parameter $\mathbf{w}$

1 Derive the curriculum region Ψ from γ;
2 Initialize $\mathbf{v}^*$, λ in the curriculum region;
3 **while** *not converged* **do**
4 Update $\mathbf{w}^* = \arg\min_{\mathbf{w}} \mathbb{E}(\mathbf{w}, \mathbf{v}^*; \lambda, \Psi)$;
5 Update $\mathbf{v}^* = \arg\min_{\mathbf{v}} \mathbb{E}(\mathbf{w}^*, \mathbf{v}; \lambda, \Psi)$;
6 **if** λ *is small* **then** increase λ by the stepsize μ;
7 **end**
8 **return** $\mathbf{w}^*$

Training a model



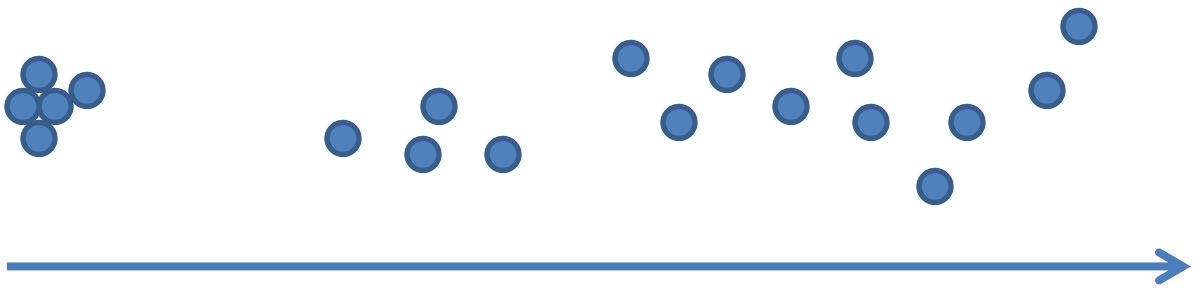
Optimization Algorithm

Algorithm 1: Self-paced Curriculum Learning.

input : Input dataset $\mathcal{D}$, predetermined curriculum γ, self-paced function f and a stepsize μ
output: Model parameter $\mathbf{w}$

1 Derive the curriculum region Ψ from γ;
2 Initialize $\mathbf{v}^*$, λ in the curriculum region;
3 **while** *not converged* **do**
4 | Update $\mathbf{w}^* = \arg\min_{\mathbf{w}} \mathbb{E}(\mathbf{w}, \mathbf{v}^*; \lambda, \Psi)$;
5 | Update $\mathbf{v}^* = \arg\min_{\mathbf{v}} \mathbb{E}(\mathbf{w}^*, \mathbf{v}; \lambda, \Psi)$;
6 | **if** λ *is small* **then** increase λ by the stepsize μ;
7 **end**
8 **return** $\mathbf{w}^*$

Recalculating the loss and select more examples.



smaller loss

bigger loss

Optimization Algorithm

Algorithm 1: Self-paced Curriculum Learning.

input : Input dataset $\mathcal{D}$, predetermined curriculum γ, self-paced function f and a stepsize μ
output: Model parameter $\mathbf{w}$

1 Derive the curriculum region Ψ from γ;
2 Initialize $\mathbf{v}^*$, λ in the curriculum region;
3 **while** *not converged* **do**
4 | Update $\mathbf{w}^* = \arg\min_{\mathbf{w}} \mathbb{E}(\mathbf{w}, \mathbf{v}^*; \lambda, \Psi)$;
5 | Update $\mathbf{v}^* = \arg\min_{\mathbf{v}} \mathbb{E}(\mathbf{w}^*, \mathbf{v}; \lambda, \Psi)$;
6 | **if** λ *is small* **then** increase λ by the stepsize μ;
7 **end**
8 **return** $\mathbf{w}^*$

$$\arg \min_{\mathbf{w},\mathbf{v}\in[0,1]^n} \sum_{i=1}^{n} v_i L(y_i, g(\mathbf{x}_i, \mathbf{w})) - \lambda \sum_{i=1}^{n} v_i$$

- While fixing w, the solution looks like:

$$v_i^* = \begin{cases} 1, & L(y_i, g(\mathbf{x}_i, \mathbf{w})) < \lambda, \\ 0, & \text{otherwise.} \end{cases} \qquad \lambda \Rightarrow \text{model age}$$

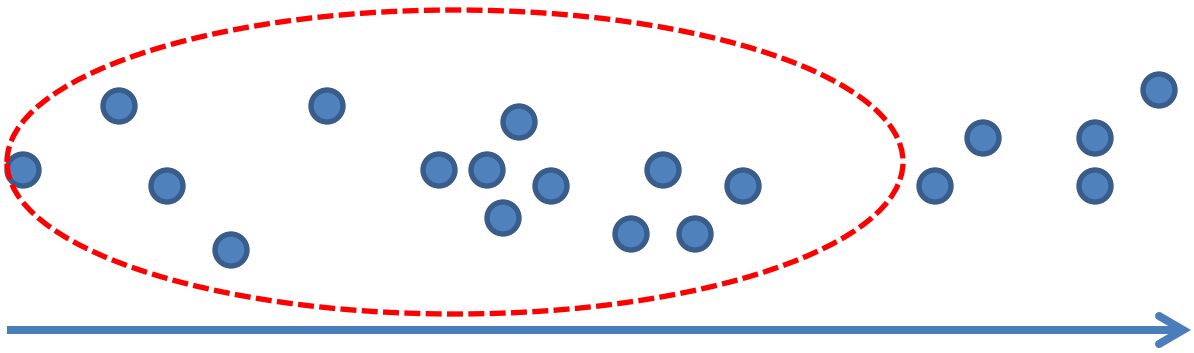
SPCL Algorithm

Algorithm 1: Self-paced Curriculum Learning.

input : Input dataset $\mathcal{D}$, predetermined curriculum γ, self-paced function f and a stepsize μ
output: Model parameter $\mathbf{w}$

1 Derive the curriculum region Ψ from γ;
2 Initialize $\mathbf{v}^*$, λ in the curriculum region;
3 **while** *not converged* **do**
4 Update $\mathbf{w}^* = \arg\min_{\mathbf{w}} \mathbb{E}(\mathbf{w}, \mathbf{v}^*; \lambda, \Psi)$;
5 Update $\mathbf{v}^* = \arg\min_{\mathbf{v}} \mathbb{E}(\mathbf{w}^*, \mathbf{v}; \lambda, \Psi)$;
6 **if** λ *is small* **then** increase λ by the stepsize μ;
7 **end**
8 **return** $\mathbf{w}^*$

Increase the model age to include more examples



smaller loss

bigger loss

Self-paced Curriculum Learning

- Proposed learning objectives:

Learning schemes

$$\arg \min_{\mathbf{w},\mathbf{v}\in[0,1]^n} \sum_{i=1}^{n} v_i L(y_i, g(\mathbf{x}_i, \mathbf{w})) + f(\mathbf{v}, \lambda)$$

$$\text{subject to } \mathbf{v} \in \Psi$$

$f(\mathbf{v}, \lambda) \Rightarrow$ regularizer determines the learning scheme

Generalize a single learning scheme to multiple learning schemes.
For different problems, we can use different learning schemes.

“Soft” Learning Schemes

$$\arg \min_{\mathbf{w},\mathbf{v}\in[0,1]^n} \sum_{i=1}^{n} v_i L(y_i, g(\mathbf{x}_i, \mathbf{w})) + f(\mathbf{v}, \lambda)$$

$$\text{subject to } \mathbf{v} \in \Psi$$

Learning schemes

Existing (Hard)

Binary weighting [Kumar et al 2010]

$$f(\mathbf{v}; k) = -\frac{1}{k}\|\mathbf{v}\|_1 = -\frac{1}{k}\sum_{i=1}^{n} v_i.$$

Proposed (Soft)

linear weighting

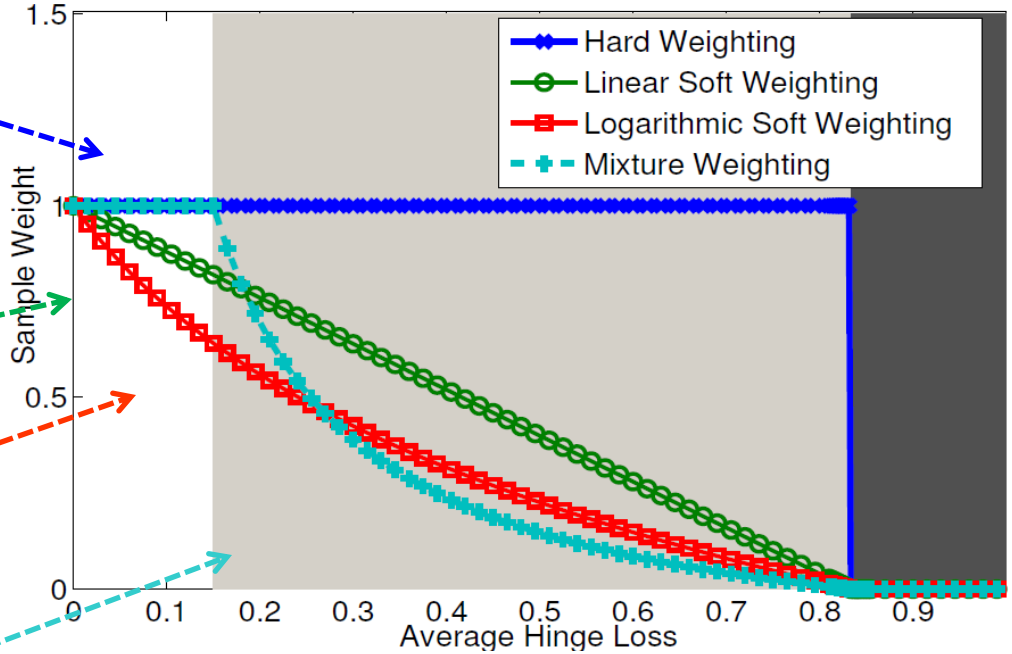
$$f(\mathbf{v}; k) = \frac{1}{k}(\frac{1}{2}\|\mathbf{v}\|_2^2 - \sum_{i=1}^{n} v_i).$$

Logarithmic weighting

$$f(\mathbf{v}; k) = \sum_{i=1}^{n}(\zeta v_i - \frac{\zeta^{v_i}}{\log \zeta}),$$

Mixture weighting

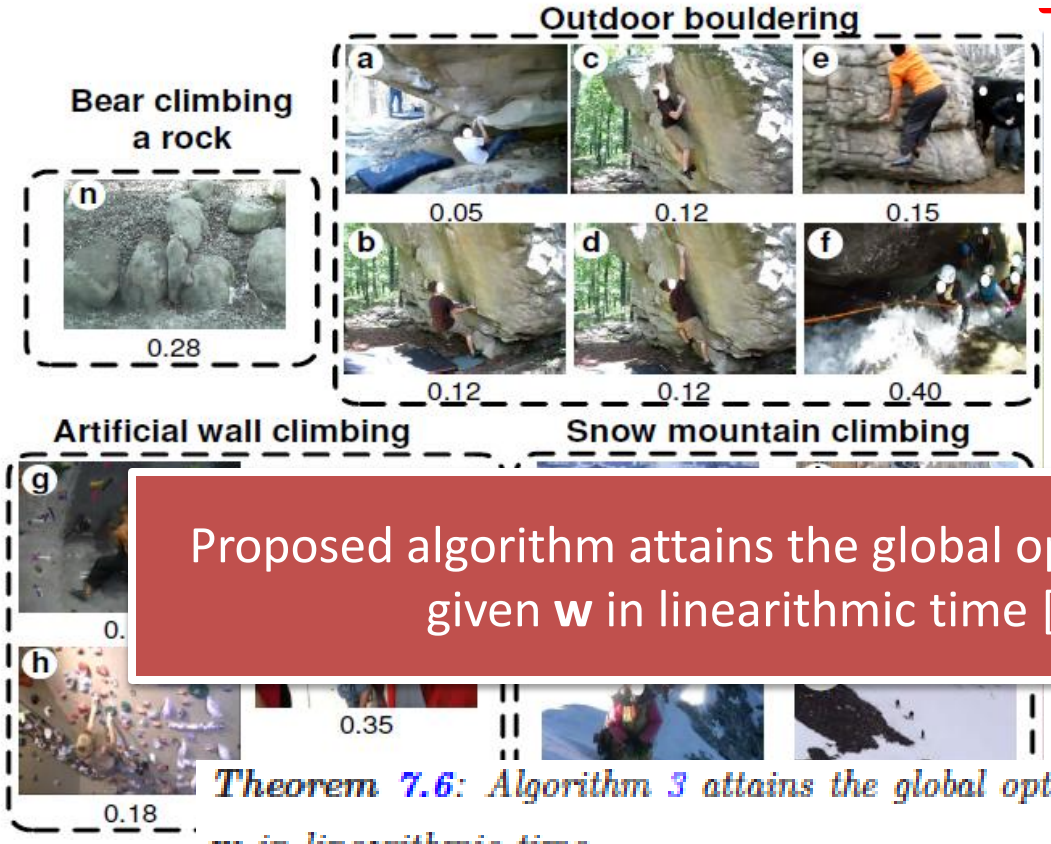
$$f(\mathbf{v}; k, k') = -\zeta \sum_{i=1}^{n} \log(v_i + \zeta k),$$

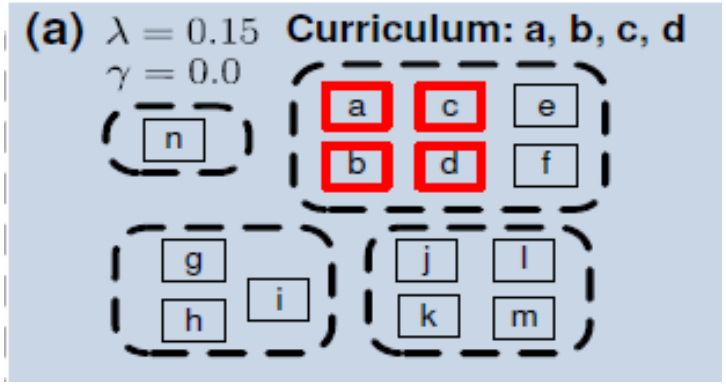


Learning Easy and Diverse Samples

$$f(\mathbf{v}, \lambda) = -\lambda \sum_{i=1}^{n} v_i$$

Favor diverse examples



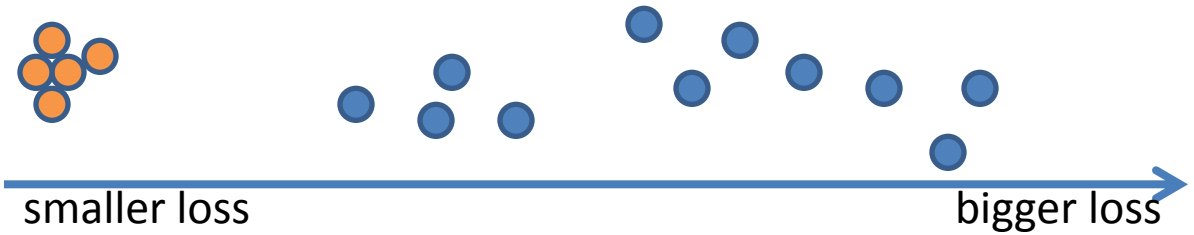


Proposed algorithm attains the global optimum solution of **v** for any given **w** in linearithmic time [Jiang et al. 2014]

Theorem 7.6: Algorithm 3 attains the global optimum to $\min_{\mathbf{v}} \mathbf{E}(\mathbf{w}, \mathbf{v})$ *for any given* $\mathbf{w}$ *in linearithmic time.*

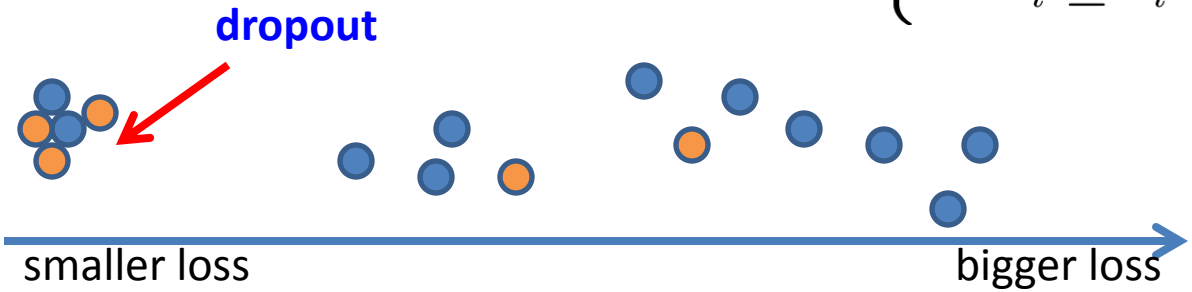
Learning scheme with Dropout

$$f_b(\mathbf{v};\lambda) = -\lambda\|\mathbf{v}\|_1 \quad \Rightarrow \quad v_i^* = \begin{cases} 1, & \ell_i < \lambda, \\ 0, & \text{otherwise.} \end{cases}$$



$$r_i(p) \sim \text{Bernoulli}(p) + \epsilon, (0 < \epsilon \ll 1)$$

$$f_b(\mathbf{v};\lambda,p) = -\lambda\|\mathbf{r}\cdot\mathbf{v}\|_1, \quad \Rightarrow \quad v_i^* = \begin{cases} 1 & \ell_i < r_i\lambda \\ 0 & \ell_i \geq r_i\lambda \end{cases}$$



Self-paced Curriculum Learning

- Proposed learning objectives:

Learning schemes

$$\arg \min_{\mathbf{w},\mathbf{v}\in[0,1]^n} \sum_{i=1}^{n} v_i L(y_i, g(\mathbf{x}_i, \mathbf{w})) + f(\mathbf{v}, \lambda)$$

$$\text{subject to } \mathbf{v} \in \Psi$$

- From difficult examples to easy examples in some problems?
  - Implicit self-paced function[Fan et al. 2017].
  - A self-paced function might not be expressed as a known function as long as it has a known closed-form solution.

Yanbo Fan, Ran He, Jian Liang and Baogang Hu . Self-Paced Learning: an Implicit Regularization Perspective. AAAI 2017

Self-paced Curriculum Learning

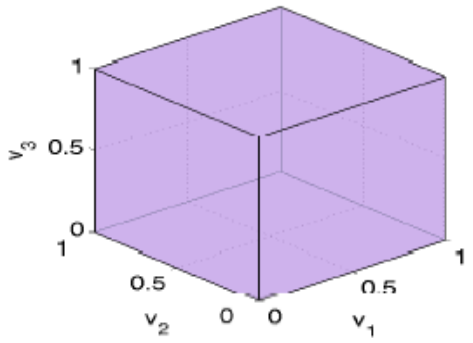
- Proposed learning objectives:

$$\arg \min_{\mathbf{w},\mathbf{v}\in[0,1]^n} \sum_{i=1}^{n} v_i L(y_i, g(\mathbf{x}_i, \mathbf{w})) + f(\mathbf{v}, \lambda)$$

$$\text{subject to } \mathbf{v} \in \Psi$$

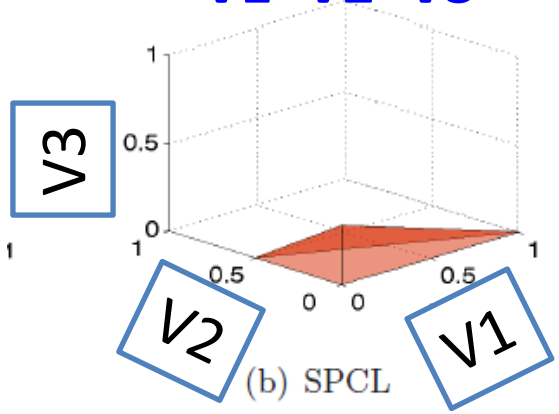
Prior knowledge in curriculum learning

- The shape of the feasible region weakly implies a prior learning sequence of samples.



(a) SPL

V1>V2>V3



(b) SPCL

Self-paced Curriculum Learning

- Proposed learning objectives:

Learner

Learning schemes

$$\arg \min_{\mathbf{w},\mathbf{v}\in[0,1]^n} \sum_{i=1}^{n} v_i L(y_i, g(\mathbf{x}_i, \mathbf{w})) + f(\mathbf{v}, \lambda)$$

$$\text{subject to } \mathbf{v} \in \Psi$$

Prior knowledge in curriculum learning

- Has been used to solve a variety of problems:
  - Noisy Labels Training [Vembu et al. ECML'16, Zięba et al. AIIDB'16, Sangineto et al. archive'16]
  - Retrieval [Liang et al. SIGIR'16, Xu et al. MM'16, Habibian TPAMI'16]
  - Saliency detection [Zhang et al. ICCV'16]
  - Muti-task learning [Li et al. AAAI'16]
  - Deep Learning[Liang et al. PAMI, 17, Sangineto et al 2016, Wei PAMI 2016]
  - Matrix Factorization [Qian et al., AAAI'15]

Theoretical Discussions

- Let $v_i^*(\lambda, \ell_i) = \arg\min_{v_i \in [0,1]} v_i \ell_i + f(v_i, \lambda)$ represent the optimal value for v_i . The latent objective function has the following form (when the model age lambda is fixed) [Meng et al. 2015]:

$$F_\lambda(\ell_i) = \int_0^{\ell_i} v_i^*(\lambda; \ell_i) d\ell_i$$

- Incorporate the binary and linear scheme into the latent fun:

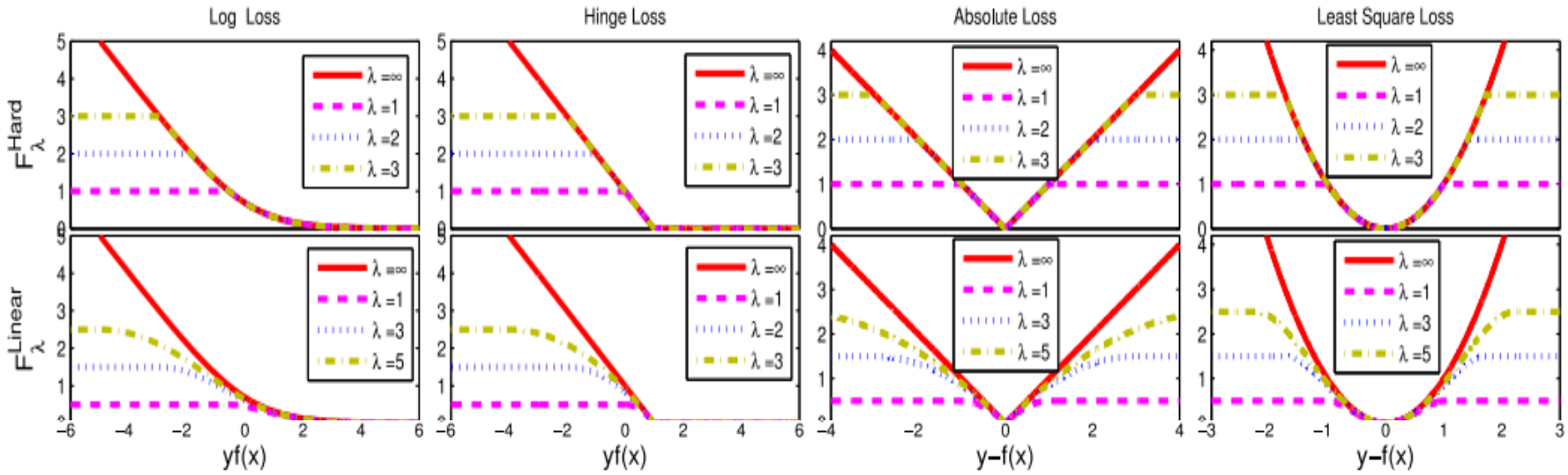
$$F_\lambda^b(\ell_i) = \min(\ell_i, \lambda)$$

Turns out the loss is the robust loss function called **Capped-Norm based Penalty(CNP)** and **Minimax Convex Plus (MCP)** [Zhang et al. 2010] popular penalties in the field of statistical machine learning [Liang et al. 2016]

Junwei Liang, Lu Jiang, Deyu Meng, Alexander Hauptmann. Learning to Detect Concepts from Webly-Labeled Video Data. In Joint Conference on Artificial Intelligence (IJCAI), 2016
Meng, Deyu, Qian Zhao, and Lu Jiang. "What objective does self-paced learning indeed optimize?." *arXiv preprint arXiv:1511.06049* (2015).
Zhang, Tong. "Analysis of multi-stage convex relaxation for sparse regularization." The Journal of Machine Learning Research 11 (2010): 1081-1107.
Zhang, Cun-Hui. "Nearly unbiased variable selection under minimax concave penalty." The Annals of statistics (2010): 894-942.

Theoretical Discussions

$$F_\lambda^b(\ell_i) = \min(\ell_i, \lambda)$$
$$F_\lambda^l(\ell_i) = \mathbf{I}(\ell_i \geq \lambda)\frac{\lambda}{2} + \mathbf{I}(\ell_i < \lambda)(\ell_i - \frac{\ell_i^2}{2\lambda})$$



Limitations of CL/SPL/SPCL

- We did observe 3 limitations of CL/SPL/SPCL:
  - Fundamental assumptions
    - learning to be conducted iteratively
    - samples organized in a meaningful sequence
    - models become more complex

Problems not applicable:
Learning over small or clean datasets, where samples are carefully selected. Examples: ImageNet, CIFAR-10. **May NOT outperform** traditional learning method [Avramova 2015].
Problems applicable:
Learning over noisy data or weakly-labeled data. YFCC100M, PASCAL VOC[Chen et al. ICCV 2015].

Avramova, Vanya. "Curriculum Learning with Deep Convolutional Neural Networks." (2015).
Chen, Xinlei, and Abhinav Gupta. "Webly supervised learning of convolutional networks." ICCC 2015

Limitations of CL/SPL/SPCL

- We did observe 3 limitations of CL/SPL/SPCL:
  - Fundamental assumptions
    - learning to be conducted iteratively
    - samples organized in a meaningful sequence
    - models become more complex
  - Unstable/Sensitive to starting values.

The curriculum needs to be meaningful so that it can provide supervision in the first few iteration.

We introduce curriculum constraints to incorporate prior knowledge about the problem. Defining curriculum is problem-specific.

Limitations of CL/SPL/SPCL

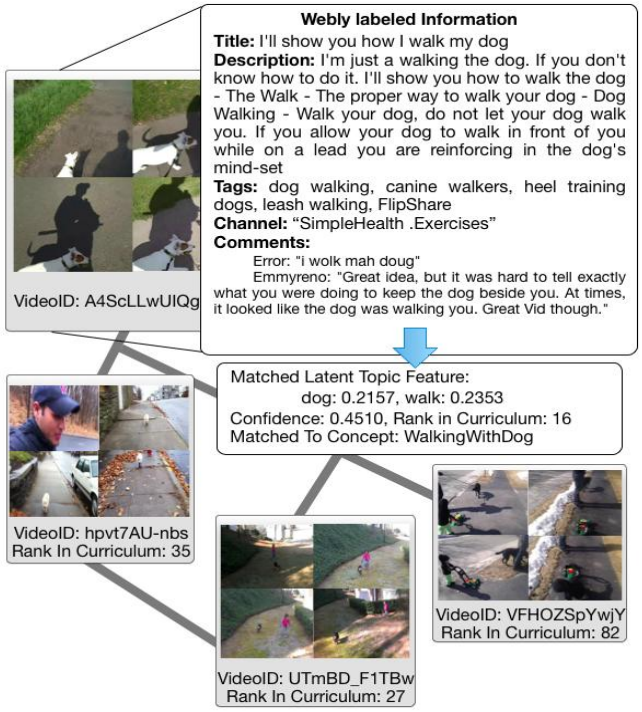
- We did observe 3 limitations of CL/SPL/SPCL:
  - Fundamental assumptions
    - learning to be conducted iteratively
    - samples organized in a meaningful sequence
    - models become more complex
  - Unstable/Sensitive to starting values.
  - Good validation set to tune the hyper-parameter.

Bad hyper-parameters “age” may hurt performance. Need a validation set to determine the timing of early stopping.

On noisy data, early stopping is a must.

Experiments on Big Noisy Data

- Fudan-Columbia Video Dataset (FCVID)
  - **91,223 YouTube videos** (4,232 hours) from 239 categories.
  - Only use the manual label in testing.
  - Feature VGG16 ideo-level feature.
  - Base learner: SVM hinge-loss + l2 norm.
- YFCC100M
  - 0.8 million personal videos on Flicker (creative common)
  - No manual labels available.
  - Derive web labels from the textual metadata for 101 concepts with the highest term frequency.
  - Evaluate on the NIST **dataset (23,000 Internet videos**) in terms of P@10.



Experiments on Big Noisy Data

Table 1: Performance comparison on FCVID.

Method	P@5	P@10	mAP
BatchTrain	0.782	0.763	0.469
Adaboost [Friedman, 2002]	0.211	0.173	0.08
SPL [Kumar *et al.*, 2011]	0.793	0.754	0.414
GoogleHNM [Varadarajan *et al.*, 2015]	0.781	0.757	0.472
BabyLearning [Liang *et al.*, 2015]	0.834	0.817	0.496
WELL (binary w/o dropout)	0.857	0.843	0.521
WELL (linear)	**0.893**	**0.877**	**0.566**
WELL (binary)	**0.893**	**0.878**	**0.567**

Proposed method

Table 3: Performance comparison on YFCC100M.

Method	P@3	P@5	P@10
BatchTrain	0.535	0.513	0.487
Adaboost [Friedman, 2002]	0.191	0.188	0.186
SPL [Kumar *et al.*, 2011]	0.485	0.463	0.454
GoogleHNM [Varadarajan *et al.*, 2015]	0.541	0.525	0.500
BabyLearning [Liang *et al.*, 2015]	0.548	0.519	0.466
WELL (binary w/o dropout)	0.607	0.608	0.589
WELL (linear)	**0.667**	**0.663**	**0.649**
WELL (binary)	**0.660**	**0.640**	**0.625**

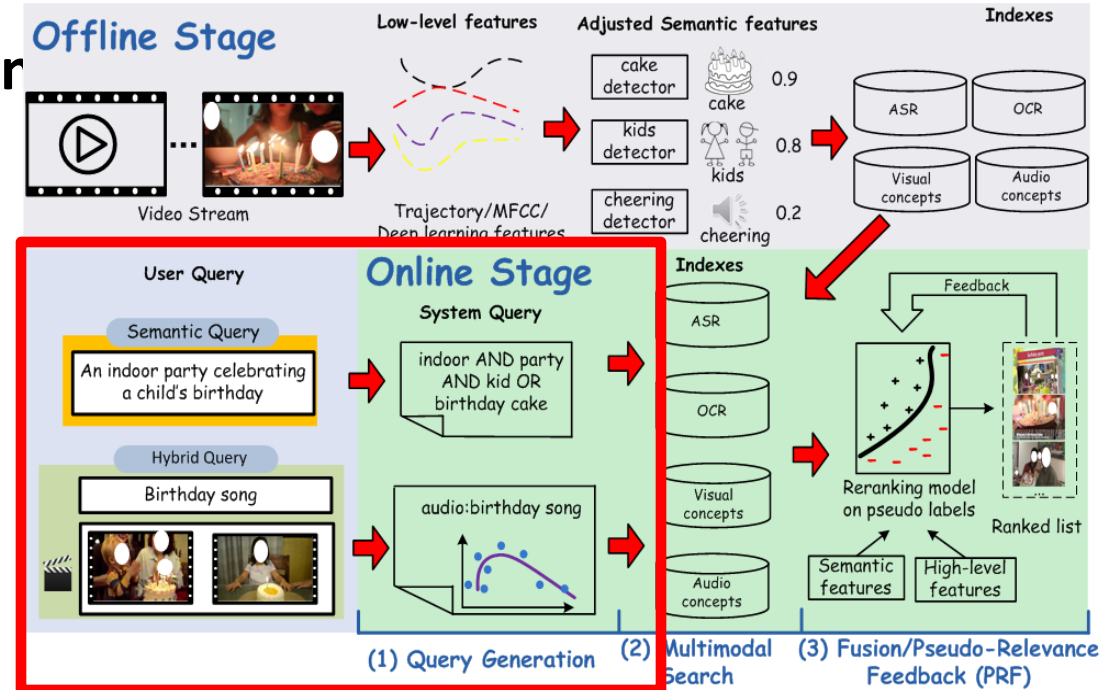
Proposed method

Outline

- Introduction
- Review of Finished Work in Proposal
- Proposed Work:
  - Robust learning in weakly-labeled data
  - **Deep visual query understanding & hybrid search**
  - Visual MemexQA
- Conclusions

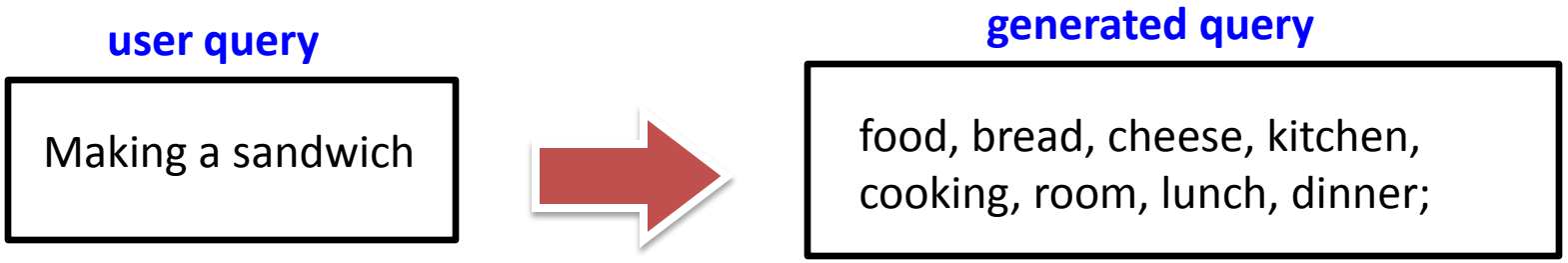
Outline

- Introduction
- Review of Finished Work in Proposal
- Proposed Work:
  - Robust learning in weakly-labeled data
  - **Deep visual query ur**
  - Visual MemexQA
- Conclusions



Semantic Query Understanding

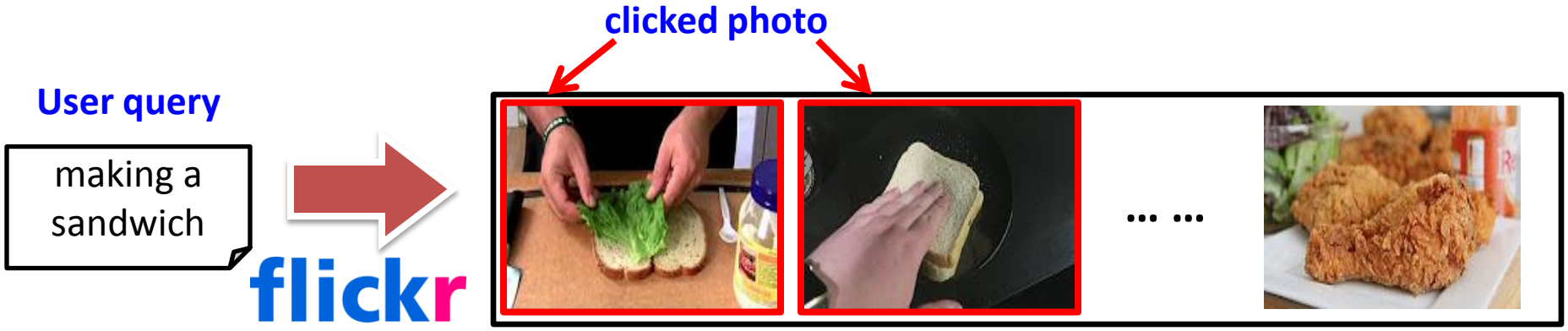
- Training detector about every query? Probably not.
- **Semantic query understanding**: how to map out-of-vocabulary query words to the concepts in our vocabulary?



- Challenging problem. Existing methods:
  - **Exact word matching**
  - **WordNet Similarity** [Miller, 1995]: structural depths in WordNet taxonomy.
  - **Word embedding mapping** [Mikolov et al., 2013]: word distance in a learned embedding space in Wikipedia by the skip-gram model (word2vec).

Deep Visual Query Embedding

- Our solution: learning deep visual query embedding mined from the Flickr search logs [Jiang et al. 2017].

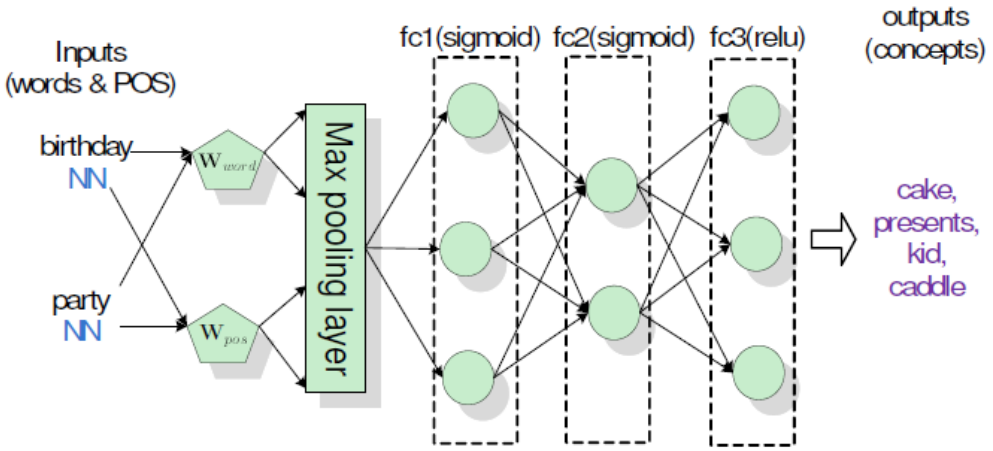


- Training dat

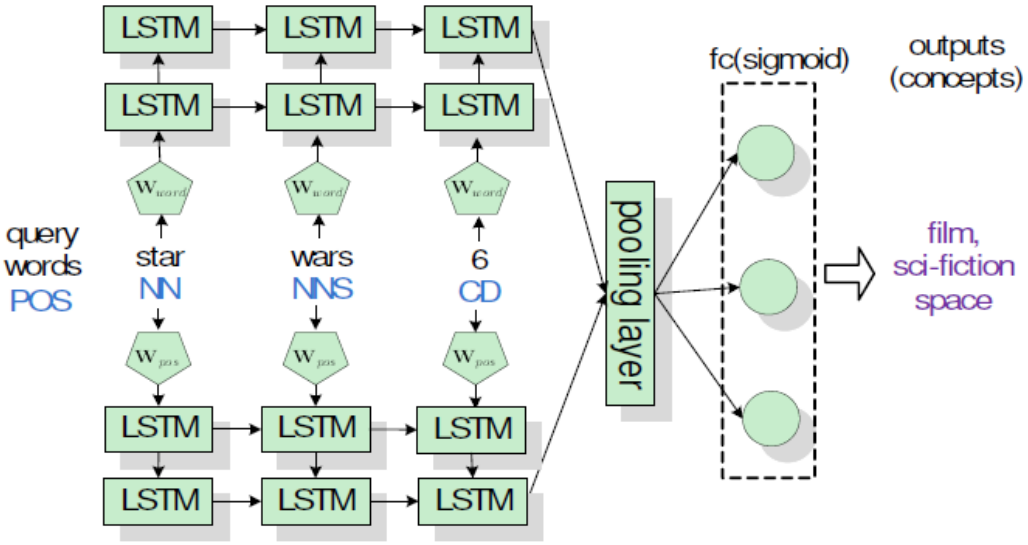
User queries	Related Visual Concepts
jaguar →	sports car, road
playa →	coast, ocean
bluebell →	flower, purple
tiger →	carnivore, big cat, tiger
andromeda →	empty, dreamlike, fire, bonfire
zoo →	people, animal, primate, dog, monkey

(3 users, top 30 results) (mutual information)

Deep Visual Query Embedding



(a) Max-Pooled MLP



(b) Two-channel RNN

Experimental Results

- Goal: rank clicked photos closer to the top, evaluated by the mean average precision and Recall@K.
  - Training: 20,600 queries from **3,978 users**
  - Test on 2,443 queries from **1,620 users** over about 148,000 personal photos.

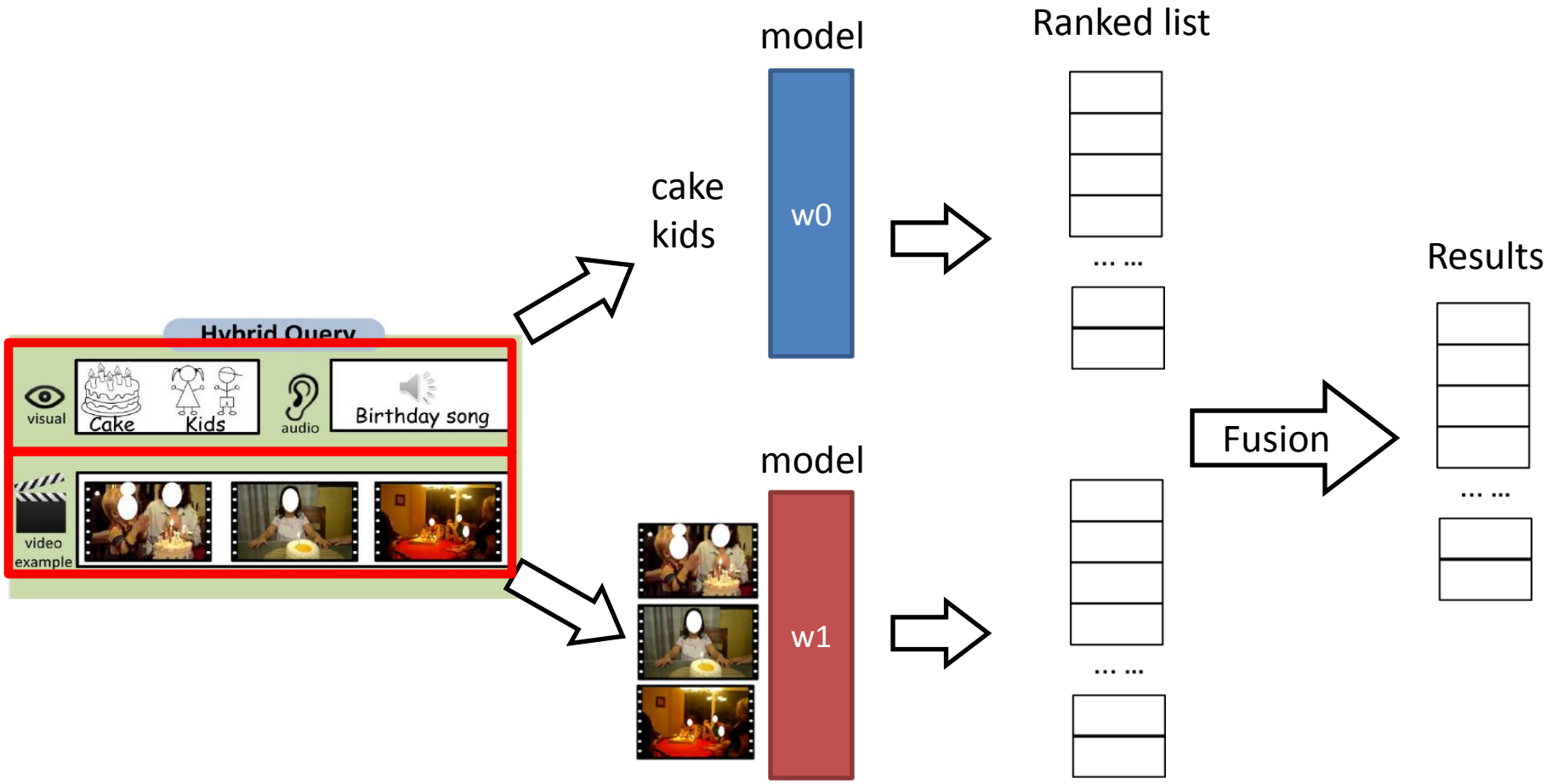
Word2vec embedding

Method	mAP	R@1	R@3	R@5
Exact Match [19]	0.231	0.209	0.086	0.067
VQE (MaxMLP)	**0.390**	**0.524**	**0.374**	**0.289**

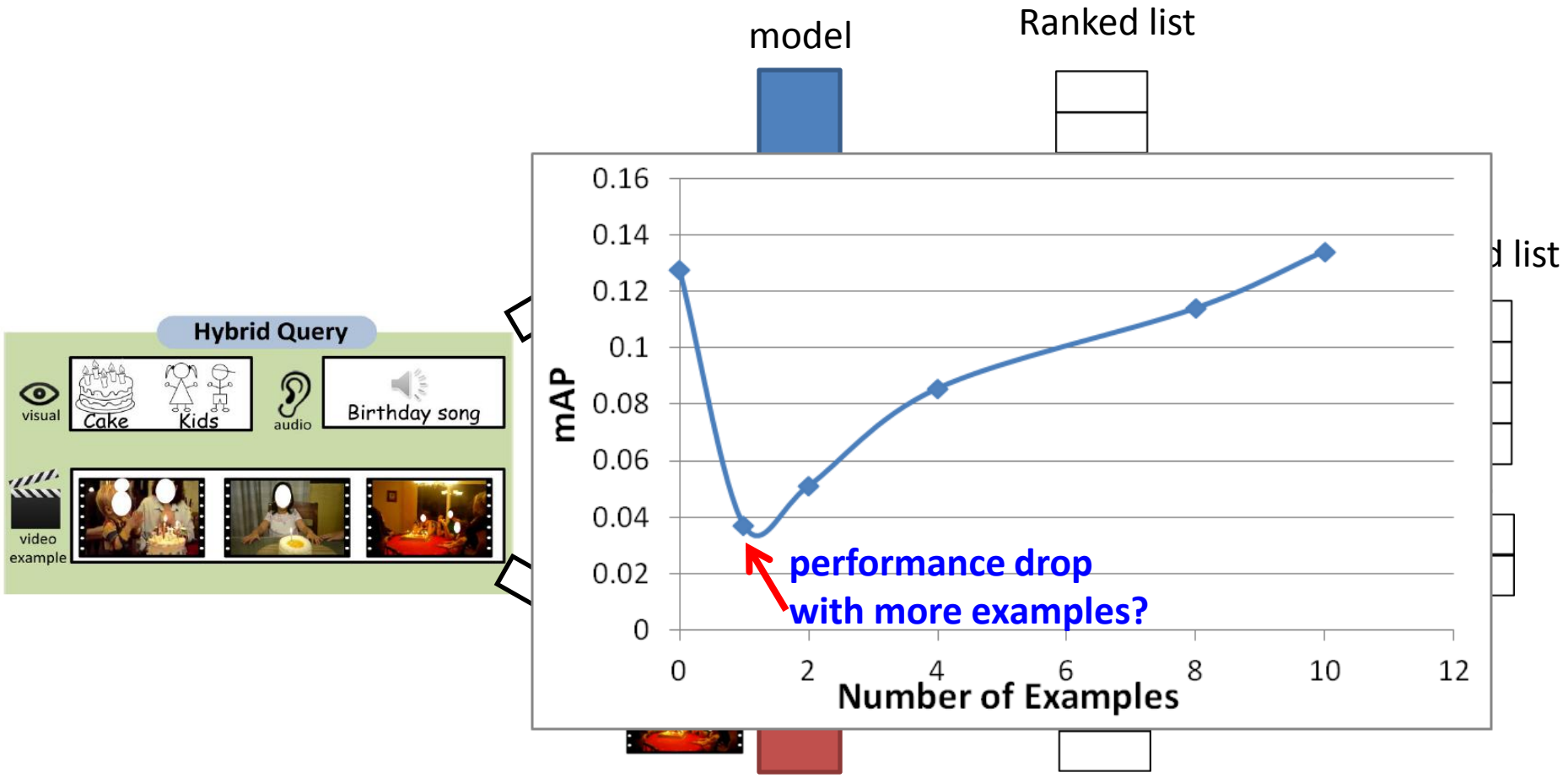
More results can be found in Section 5.3.3

Proposed model (45% relative improvement)

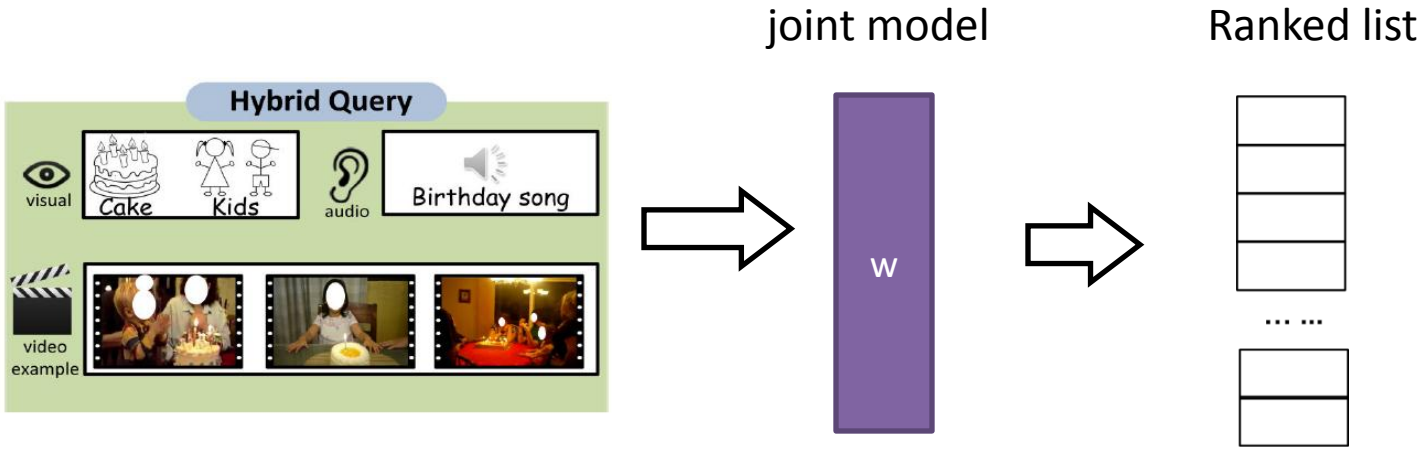
Hybrid Search



Hybrid Search

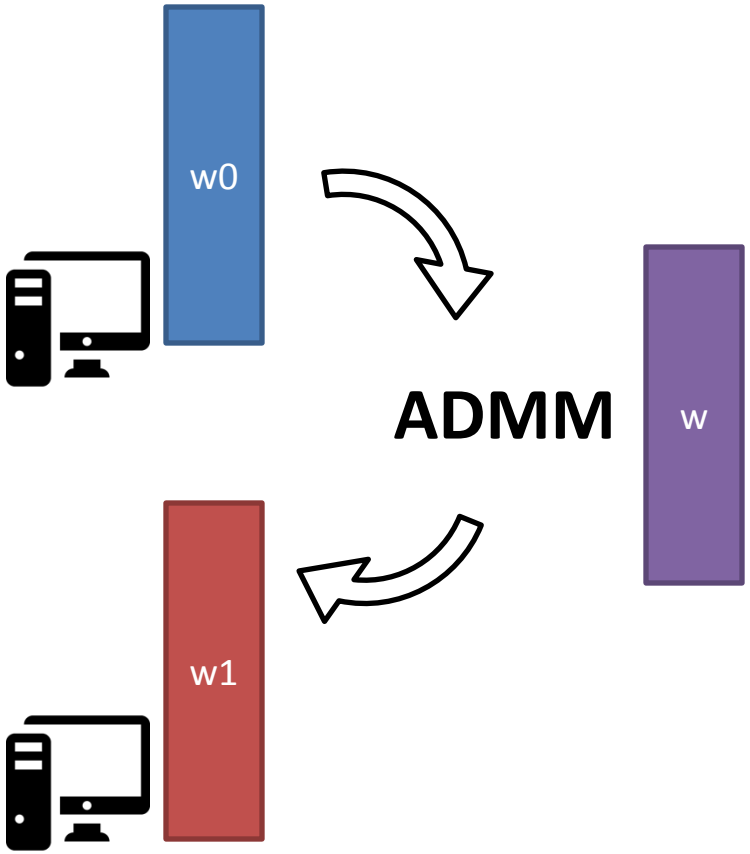


Our Solution



Hybrid Search

Optimize model on multiple machines



Alternating Direction Method of Multipliers (ADMM)

ADMM Algorithm:

$$\mathbf{w}_1^{(k+1)} = \arg\min_{\mathbf{w}_1^{(k)}} L_p(\mathbf{X}, \mathbf{y}; \mathbf{w}_0^{(k)}, \mathbf{w}_1^{(k)}, \lambda^{(k)})$$

$$\mathbf{w}_0^{(k+1)} = \arg\min_{\mathbf{w}_0^{(k)}} L_p(\mathbf{X}, \mathbf{y}; \mathbf{w}_0^{(k)}, \mathbf{w}_1^{(k+1)}, \lambda^{(k)})$$

$$\lambda^{(k+1)} = \lambda^{(k)} + (p/2+1)(\mathbf{w}_1^{(k+1)} - \mathbf{w}_0^{(k+1)})$$

Model trained w/. examples.

Model trained w/o. examples.

In our problem, we assume users do not change the semantic query. When ADMM converges, we have the following objective function:

$$\text{minimize}_{\mathbf{w}_1} L(\mathbf{X}, \mathbf{y}; \mathbf{w}_1, \mathbf{w}_0) + p\|\mathbf{w}_1 - \mathbf{w}_0\|_2^2$$

Hybrid Search Experiments

Dataset: MED14Test (around 25,000 videos on 20 events).

Evaluation metric: mean Average Precision, Mean Reciprocal Rank, Precision@20, mAP@20

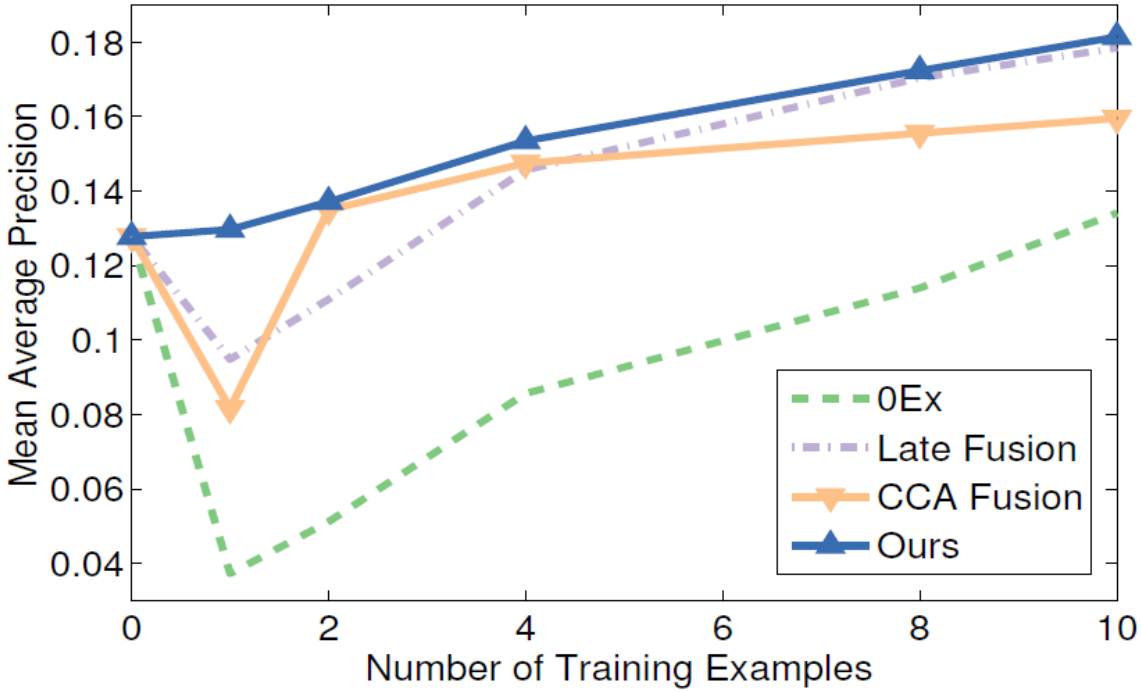
Proposed model

Methods	mAP	P@20	MRR	mAP@20
0Ex	0.1278	0.1600	0.4130	0.1099
	1 Examples			
1Ex	0.0372	0.0675	0.2957	0.0312
1Ex+0Ex Fusion	0.0948	0.1325	0.4341	0.0761
2Ex	0.0512	0.0800	0.2753	0.0393
2Ex+0Ex Fusion	0.1109	0.1350	0.4076	0.0865
2Ex+0Ex CCA	0.1350	0.1600	**0.4977**	0.1110
2Ex+0Ex Ours	**0.1363**	**0.1608**	0.4794	**0.1131**

More results can be found in Section 5.4.2

Hybrid Search Experiments

Dataset: MED14Test (around 25,000 videos on 20 events).



1. **Our method significantly outperforms others given 1-5 examples.**
2. **Fusion yields reasonable results given enough examples.**

Hybrid Search Experiments

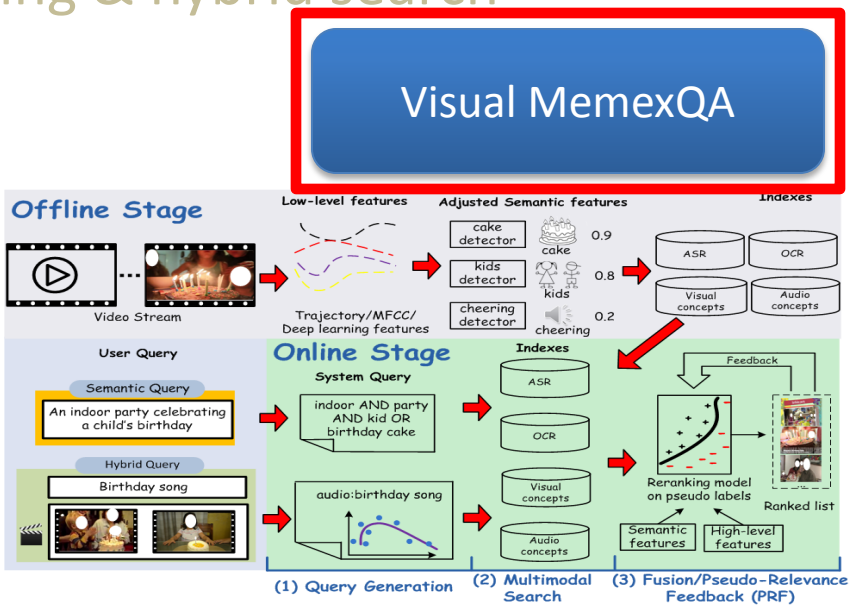
Underlined concepts are the ones in the user semantic query

ID	Event Name	Late Fusion		Ours	
		mAP	concepts	mAP	concepts
E028	Town hall meeting	0.121	space, concert, disgust, male reporter, reporter	0.228	graduation, speech, talking, cheering, reporter, boredom
E035	Horse riding competition	0.283	equestrianism, endurance riding, kalaripayattu, forest	0.338	show jumping, horse, barrel racing, steeplechase, dressage
E039	Tailgating	0.074	bill, armored vehicle, trip	0.156	tent, truck, van, team, stadium

- Our method finds more **relevant** concepts.
- It assigns **appropriate weights** to the concepts in the user query.
- It may find concepts not existing in user queries.
- Models are **sparser** → allows for fast searching over millions of videos.

Outline

- Introduction
- Review of Finished Work in Proposal
- Proposed Work:
  - Robust learning in weakly-labeled data
  - Deep visual query understanding & hybrid search
  - **Visual MemexQA**
- Conclusions



Characteristics

What's the difference when users search their own photos versus the photos on the web?

- Understanding user needs by exploring the Flickr search logs.

❑Show me the photo of my dog.

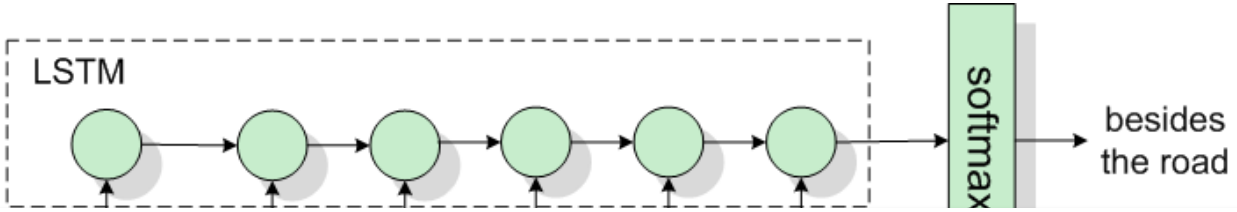
❑What was the last time we went hiking?

❑What's the name of that amazing hotel in our last trip to SF?

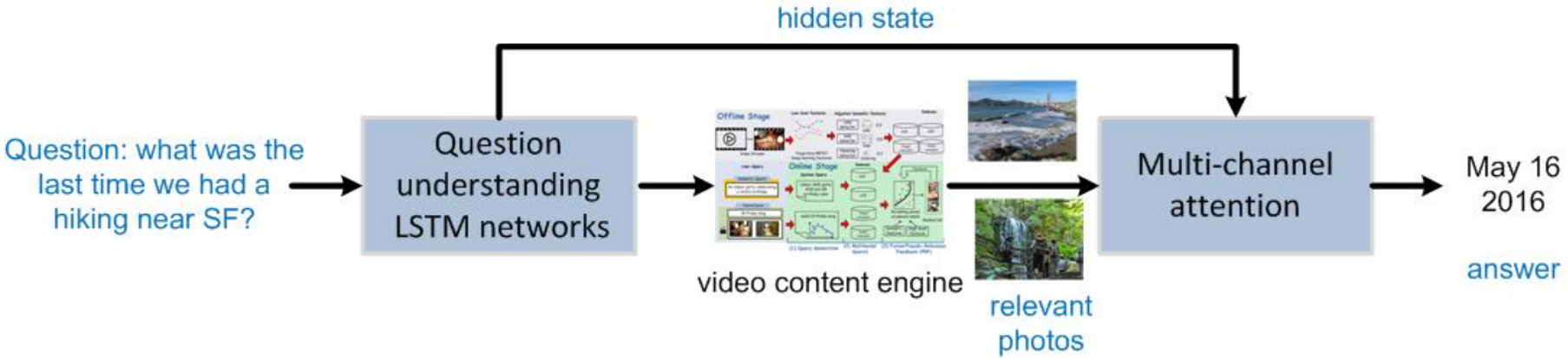
❑Where was my brother's graduation ceremony in 2013?

❑How many times have I had sushi last week?

Visual QA and Visual MemexQA



- Major difference to Visual MemexQA:
  - ❑ Ask question to a large collection of photos and videos.
  - ❑ Closed-domain. Questions to recall visual memory.
  - ❑ And **more useful**!



Pre-train each network on its own task then fine tune them on the entire dataset.

Demo

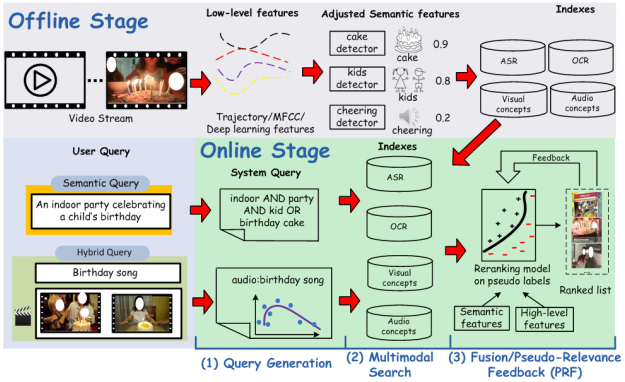
- Demo
- Demo Video1 Demo Video2
- Benchmark dataset will be ready in weeks.
  - 20K question answers.
  - 14k Flickr photos.
  - 101 real users.

Outline

- Introduction
- Review of Finished Work in Proposal
- Proposed Work:
  - Robust learning in weakly-labeled data
  - Deep visual query understanding & hybrid search
  - Visual MemexQA
- **Conclusions**

Conclusions

- In this talk, we discussed:
  - ✓ A state-of-the-art web-scale content-based search and learning framework over hundreds of millions of Internet videos [MM'12, MM'14, MM'15, ICMR'15, WWW'16].



[MM12] Lu Jiang, Alexander Hauptmann, Guang Xiang. Leveraging High-level and Low-level Features for Multimedia Event Detection. In ACM Multimedia (MM), 2012.

[MM14] Lu Jiang, Deyu Meng, Teruko Mitamura, Alexander Hauptmann. Easy Samples First: Selfpaced Reranking for Zero-Example Multimedia Search. In ACM Multimedia (MM), 2014.

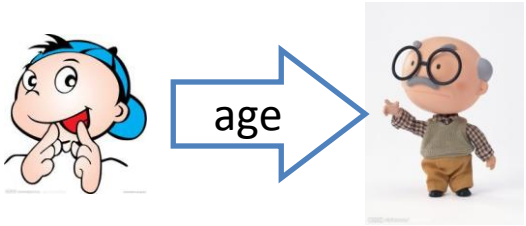
[MM15] Lu Jiang, Shoou-I Yu, Deyu Meng, Yi Yang, Teruko Mitamura, Alexander Hauptmann. Fast and Accurate Content-based Semantic Search in 100M Internet Videos. In ACM Multimedia (MM), 2015.

[ICMR15] Lu Jiang, Shoou-I Yu, Deyu Meng, Teruko Mitamura, Alexander Hauptmann. Bridging the Ultimate Semantic Gap: A Semantic Search Engine for Internet Videos. In ACM International Conference on Multimedia Retrieval (ICMR), 2015. **[best paper candidate]**

[WWW16] Lu Jiang. Web-scale Multimedia Search for Internet Video Content In World Wide Web (WWW), 2016

Conclusions

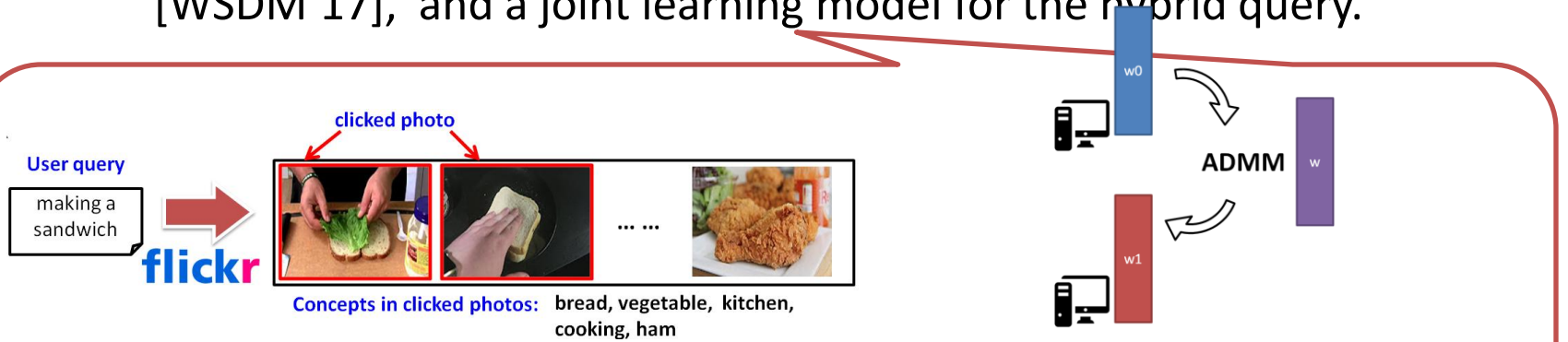
- In this talk, we discussed:
  - ✓ A state-of-the-art web-scale content-based search and learning framework over hundreds of millions of Internet videos [MM'12, MM'14, MM'15, ICMR'15, WWW'16].
  - ✓ A novel theory about self-paced curriculums learning framework, and its successful application on concept learning on weakly labeled data[NIPS'14, AAAI'15, IJCAI'16].



[NIPS14] Lu Jiang, Deyu Meng, Shoou-I Yu, Zhen-Zhong Lan, Shiguang Shan, Alexander Hauptmann. Self-paced Learning with Diversity. In Neural Information Processing Systems (NIPS), 2014.
[AAAI15] Lu Jiang, Deyu Meng, Qian Zhao, Shiguang Shan, Alexander Hauptmann. Self-paced Curriculum Learning. In Conference on Artificial Intelligence (AAAI), 2015.
[IJCAI16] Junwei Liang, Lu Jiang, Deyu Meng, Alexander Hauptmann. Learning to Detect Concepts from Webly-Labeled Video Data. In Joint Conference on Artificial Intelligence (IJCAI), 2016.

Conclusions

- In this talk, we discussed:
  - ✓ A state-of-the-art web-scale content-based search and learning framework over hundreds of millions of Internet videos [MM'12, MM'14, MM'15, ICMR'15, WWW'16].
  - ✓ A novel theory about self-paced curriculums learning framework, and its successful application on concept learning on weakly labeled data[NIPS'14, AAAI'15, IJCAI'16].
  - ✓ An deep visual query embedding model to understand user queries [WSDM'17], and a joint learning model for the hybrid query.



Lu Jiang, Yannis Kalantidis, Liangliang Cao, Sachin, Farfade, Jiliang Tang, Alex Hauptmann. Delving Deep into Personal Photo and Video Search, WSDM. 2017.

Conclusions

- In this talk, we discussed:
  - ✓ A state-of-the-art web-scale content-based search and learning framework over hundreds of millions of Internet videos [MM'12, MM'14, MM'15, ICMR'15, WWW'16].
  - ✓ A novel theory about self-paced curriculums learning framework, and its successful application on concept learning on weakly labeled data[NIPS'14, AAAI'15, IJCAI'16].
  - ✓ An deep visual query embedding model to understand user queries [WSDM'17], and a joint learning model for the hybrid query.
  - ✓ A visual MemexQA to answer questions about users' daily lives captured by personal photos and videos [Jiang et al. 2017].

Lu Jiang, Liangliang Cao, Yannis Kalantidis, Sachin, Farfade, Jiliang Tang, Alex Hauptmann. Visual MemoryQA your personal photo and video search agent, under review AAAI 2017.

Published Papers on Thesis topic

[AAAI 17] Lu Jiang, Liangliang Cao, Yannis Kalantidis, Sachin, Farfade, Alex Hauptmann. Visual Memory QA: Your Personal Photo and Video Search Agent (Demo Paper). AAAI, 2017.

[WSDM 17] Lu Jiang, Yannis Kalantidis, Liangliang Cao, Sachin, Farfade, Jiliang Tang, Alex Hauptmann. Delving Deep into Personal Photo and Video Search. In Web Search and Data **Mining(WSDM**), 2017.

[WWW 16] Lu Jiang. Web-scale Multimedia Search for Internet Video Content. In International Conference on World Wide Web (WWW), 2016.

[IJCAI 16] Junwei Liang, Lu Jiang, Deyu Meng, Alexander Hauptmann. Learning to Detect Concepts from Webly-Labeled Video Data. In Joint Conference on Artificial Intelligence (IJCAI), 2016.

[MM15] Lu Jiang, Shoou-I Yu, Deyu Meng, Yi Yang, Teruko Mitamura, Alexander Hauptmann. Fast and Accurate Content-based Semantic Search in 100M Internet Videos. In ACM Multimedia (MM), 2015.

[ICMR15] Lu Jiang, Shoou-I Yu, Deyu Meng, Teruko Mitamura, Alexander Hauptmann. Bridging the Ultimate Semantic Gap: A Semantic Search Engine for Internet Videos. In ACM International Conference on Multimedia Retrieval (ICMR), 2015. **[best paper candidate]**

[AAAI15] Lu Jiang, Deyu Meng, Qian Zhao, Shiguang Shan, Alexander Hauptmann. Self-paced Curriculum Learning. In Conference on Artificial Intelligence (AAAI), 2015.

[NIPS14] Lu Jiang, Deyu Meng, Shoou-I Yu, Zhen-Zhong Lan, Shiguang Shan, Alexander Hauptmann. Self-paced Learning with Diversity. In Neural Information Processing Systems (NIPS), 2014.

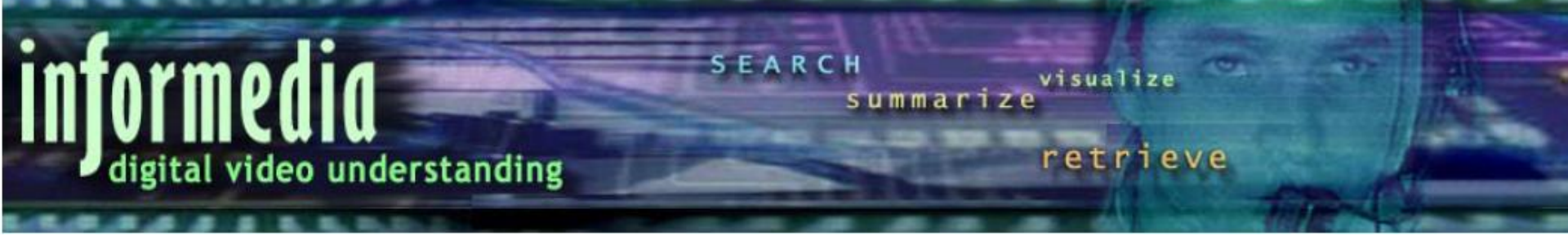
[MM14] Lu Jiang, Deyu Meng, Teruko Mitamura, Alexander Hauptmann. Easy Samples First: Selfpaced Reranking for Zero-Example Multimedia Search. In ACM Multimedia (MM), 2014.

[ICMR14] Lu Jiang, Teruko Mitamura, Shoou-I Yu, Alexander Hauptmann. Zero-Example Event Search using MultiModal Pseudo Relevance Feedback. In ACM International Conference on Multimedia Retrieval (ICMR), 2014.

[ICMR14] Lu Jiang, Wei Tong, Deyu Meng, Alexander Hauptmann. Towards Efficient Learning of Optimal Spatial Bag-of-Words Representations. In ACM International Conference on Multimedia Retrieval (ICMR). 2014. **[best paper candidate]**

[SLT14] Yajie Miao, Lu Jiang, Hao Zhang, Florian Metze. Improvements to Speaker Adaptive Training of Deep Neural Networks. In IEEE Spoken Language Technology (SLT), 2014. **[best poster]**

[MM12] Lu Jiang, Alexander Hauptmann, Guang Xiang. Leveraging High-level and Low-level Features for Multimedia Event Detection. In ACM Multimedia (MM), 2012.



Key Contributions:

•The first-of-its-kind framework for web-scale content-based search over hundreds of millions of Internet videos [ICMR'15]. The proposed framework supports text-to-video, video-to-video, and text&video-to-video search [MM'12, WSDM'17].

•A novel theory about self-paced curriculums learning and its application on robust concept detector training [NIPS'14, AAAI'15, IJCAI'16].

•Novel reranking algorithms for improving performance [MM'14, ICMR'14].

•A concept adjustment method representing a video by a few salient and consistent concepts that can be efficiently indexed by the modified inverted index [MM'15]

THANK YOU.
QUESTIONS?

References

- A. G. Hauptmann, M. G. Christel, and R. Yan. Video retrieval based on semantic concepts. Proceedings of the IEEE, 96(4):602–622, 2008.
- Baptist Vandersmissen, Fr´ederic Godin, Abhineshwar Tomar, Wesley De Neve,and Rik Van de Walle. The rise of mobile and social short-form video: an indepth measurement study of vine. In ICMR Workshop on Social Multimedia and Storytelling, 2014.
- E.M. Voorhees. Proceedings of the 8th Text Retrieval Conference. TREC-8 Question Answering Track Report. 1999
- Ehsan Younessian, Teruko Mitamura, and Alexander Hauptmann. Multimodal knowledge-based analysis in multimedia event detection. In ICMR, 2012.
- G. A. Miller. Wordnet: a lexical database for english. Communications of the ACM, 38(11):39-41, 1995.
- Hyungtae Lee. Analyzing complex events and human actions in" in-the-wild" videos. In UMD Ph.D Theses and Dissertations, 2014.
- J. Supanˇciˇc III and D. Ramanan. Self-paced learning for long-term tracking. In CVPR, 2013.
- Jeffrey Dalton, James Allan, and Pranav Mirajkar. Zero-shot video retrieval using content and concepts. In CIKM, 2013.
- Jia Deng, Nan Ding, Yangqing Jia, Andrea Frome, Kevin Murphy, Samy Bengio, Yuan Li, Hartmut Neven, and Hartwig Adam. Large-scale object classification using label relation graphs. In ECCV, 2014.
- Kevin Tang, Vignesh Ramanathan, Li Fei-Fei, and Daphne Koller. Shifting weights: Adapting object detectors from image to video. In NIPS, 2012.
- Kevin Tang, Vignesh Ramanathan, Li Fei-Fei, and Daphne Koller. Shifting weights: Adapting object detectors from image to video. In NIPS, 2012
- L. Jiang, T. Mitamura, S.-I. Yu, and A. G. Hauptmann. Zero-example event search using multimodal pseudo relevance feedback. In ICMR, 2014

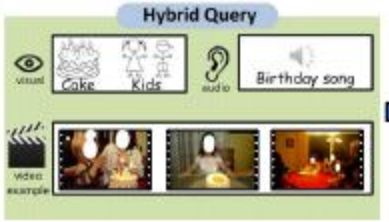
References

- L. Nie, S. Yan, M. Wang, R. Hong, and T.-S. Chua. Harvesting visual concepts for image search with complex queries. In Multimedia, 2012.
- M. Kumar, H. Turki, D. Preston, and D. Koller. Learning specific-class segmentation from diverse data. In ICCV, 2011.
- M. P. Kumar, B. Packer, and D. Koller. Self-paced learning for latent variable models. In NIPS, pages 1189–1197, 2010.
- Masoud Mazloom, Xirong Li, and Cees GM Snoek. Few-example video event retrieval using tag propagation. In ICMR, 2014.
- Noah Simon, Jerome Friedman, Trevor Hastie, and Robert Tibshirani. A sparsegroup lasso. Journal of Computational and Graphical Statistics, 22(2):231–245,2013.
- R. Yan, A. G. Hauptmann, and R. Jin. Multimedia search with pseudo-relevance feedback. In CVIR, 2003.
- Shuang Wu, Sravanthi Bondugula, Florian Luisier, Xiaodan Zhuang, and Pradeep Natarajan. Zero-shot event detection using multi-modal fusion of weakly supervised concepts. In CVPR, 2014.
- T. Mikolov and J. Dean. Distributed representations of words and phrases and their compositionality. 2013.
- V. I. Spitkovsky, H. Alshawi, and D. Jurafsky. Baby steps: How less is more in unsupervised dependency parsing. In NIPS, 2009.
- W. H. Hsu, L. S. Kennedy, and S.-F. Chang. Video search reranking through random walk over document-level context graph. In Multimedia, 2007.
- X. Tian, L. Yang, J. Wang, Y. Yang, X. Wu, and X.-S. Hua. Bayesian video search reranking. In Multimedia, 2008.
- X. Tian, Y. Lu, L. Yang, and Q. Tian. Learning to judge image search results. In Multimedia, 2011.
- Y. Bengio, J. Louradour, R. Collobert, and J. Weston. Curriculum learning. In ICML, 2009.
- Y. Liu, T. Mei, X.-S. Hua, J. Tang, X. Wu, and S. Li. Learning to video search rerank via pseudo preference feedback. In ICME, 2008.

References

- Tsvetkov, Yulia, et al. "Learning the Curriculum with Bayesian Optimization for Task-Specific Word Representation Learning." arXiv preprint arXiv:1605.03852 (2016).
- Vembu, Shankar, and Sandra Zilles. "Interactive Learning from Multiple Noisy Labels." Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer International Publishing, 2016.
- Zięba, Maciej, Jakub M. Tomczak, and Jerzy Świątek. "Self-paced Learning for Imbalanced Data." Asian Conference on Intelligent Information and Database Systems. Springer Berlin Heidelberg, 2016.
- Xu, Dan, et al. "Academic Coupled Dictionary Learning for Sketch-based Image Retrieval." Proceedings of the 2016 ACM on Multimedia Conference. ACM, 2016.
- Habibian, Amirhossein, Thomas Mensink, and Cees GM Snoek. "Video2vec Embeddings Recognize Events when Examples are Scarce.“
- Sangineto, Enver, et al. "Self Paced Deep Learning for Weakly Supervised Object Detection." arXiv preprint arXiv:1605.07651 (2016).
- Liang, Jian, et al. "Self-paced cross-modal subspace matching."Proceedings of the 39th International ACM conference on Research and Development in Information Retrieval. ACM, 2016.
- Lin, Liang, et al. "Active self-paced learning for cost-effective and progressive face identification." IEEE Transactions on Pattern Analysis and Machine Intelligence (2017).

Appendix

Problem	Query	Result
Content-based Image Search		 
Copy Detection		 
Semantic Concept Detection	Cat	 
Ours		

More complicated Semantic Query:

Information need:
people running away after an explosion in urban areas.

Query: **Boolean logical operator**

urban_scene
AND (marathon OR running)
OR fire OR smoke
OR audio:explosion
TBefore(audio:explosion, police)

Temporal operators

Applications

30 In-video ads. Average top 20 precision is 0.81

[MM15] Lu Jiang, Shoou-I Yu, Deyu Meng, Yi Yang, Teruko Mitamura, Alexander Hauptmann. Fast and Accurate Content-based Semantic Search in 100M Internet Videos. In ACM Multimedia (MM), 2015.

CL/SPL/SPCL following work(1)

Task Curricula via Minimum Feature Selection: a Case Study in Boolean Networks

Abstract

We consider the effect of introducing a curriculum of tasks when training Boolean models on supervised Multi Label Classification (MLC) problems. In particular, we consider how to order tasks in the absence of prior knowledge, and how such a curriculum may be enforced when using meta-heuristics to train discrete non-linear models.

We show that hierarchical dependencies between tasks can be exploited by enforcing an appropriate task order using a hierarchical loss function. On several multi output circuit-inference problems with known tasks difficulties, Feedforward Boolean Networks (FBNs) trained with such a loss function achieve significantly lower out-of-sample error, up to 10% in some cases. This improvement increases as the loss places more emphasis on task order.

JMLR Submission 2017

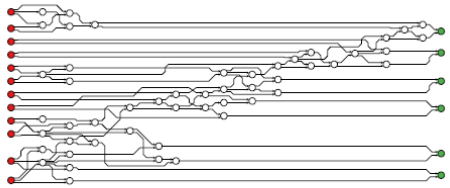


Figure 1: A 56-node FBN resulting from our learning procedure, which correctly implements the 6-bit addition function. Each node takes 2 inputs and computes the NAND function as its output. Inputs (far left) have been coloured red and outputs (far right) green. Note the ripple-carry style flow of information between sub-networks.

CL/SPL/SPCL following work(2)

CNN fine tuning using self-paced selection

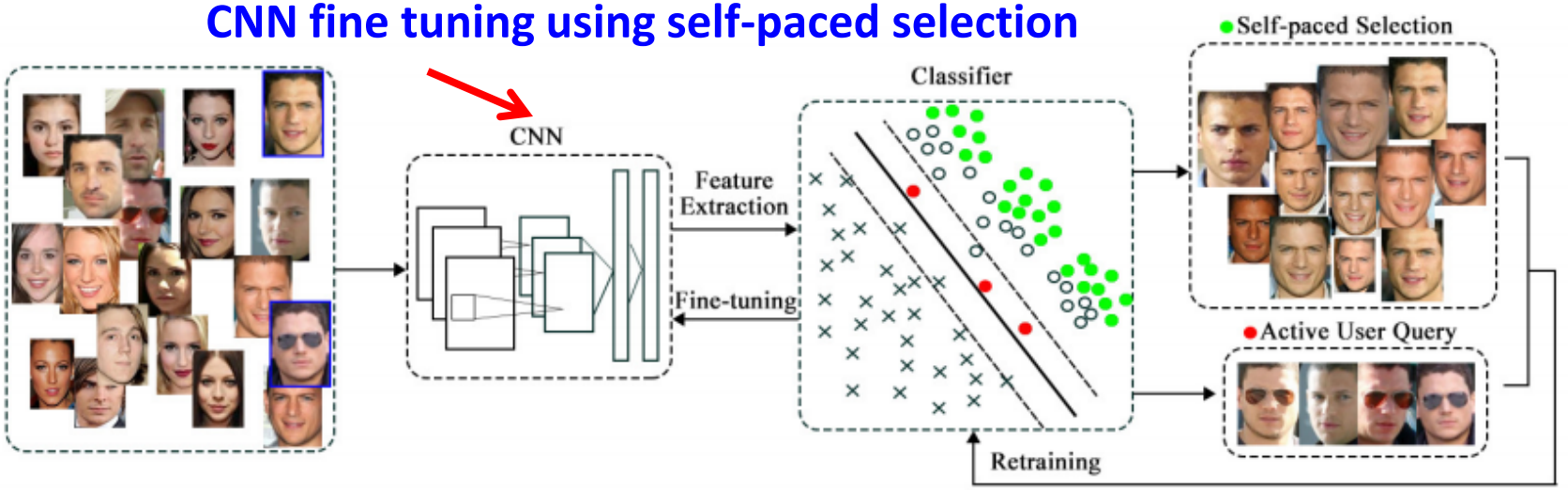
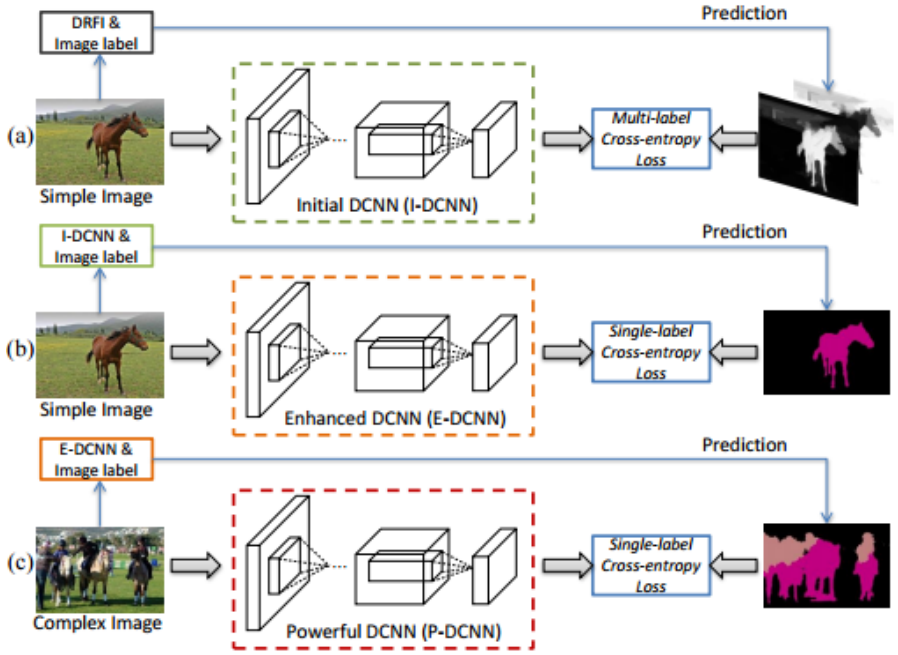


Fig. 2. Illustration of our proposed cost-effective framework. The pipeline includes stages of CNN and model initialization; classifier updating; high-confidence sample labeling by the SPL, low-confidence sample annotating by AL and CNN fine-tuning, where the arrows represent the workflow. The images highlighted by blue in the left panel represent the initially selected samples.

Lin, Liang, et al. "Active self-paced learning for cost-effective and progressive face identification." *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2017).

CL/SPL/SPCL following work(3)



Wei, Yunchao, et al. "STC: A simple to complex framework for weakly-supervised semantic segmentation." IEEE Transactions on Pattern Analysis and Machine Intelligence (2016).

CL/SPL/SPCL following work(4)

Self Paced Deep Learning for Weakly Supervised Object Detection

Enver Sangineto[*1] **Moin Nabi**[*1] **Dubravko Culibrk**[2] **Nicu Sebe**[1]
[1]DISI, University of Trento, Italy [2]FTS, University of Novi Sad, Serbia
{enver.sangineto,moin.nabi,niculae.sebe}@unitn.it,{dculibrk}@uns.ac.rs

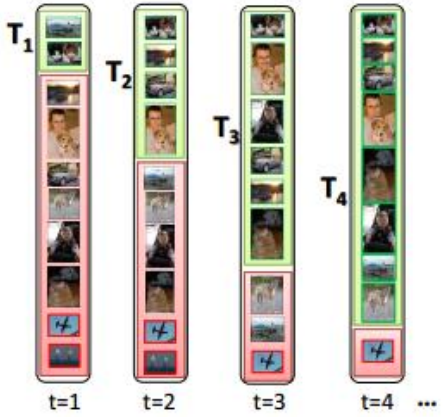


Figure 1: A schematic illustration of how the training dataset T_t of our deep net evolves depending on t and on the progressively increasing recognition skills of the trained net.

Sangineto, Enver, et al. "Self Paced Deep Learning for Weakly Supervised Object Detection." *arXiv preprint arXiv:1605.07651* (2016).

CL/SPL/SPCL following work(5)

Learning What Data to Learn

Yang Fan [1] Fei Tian [2] Tao Qin [2] Jiang Bian [2] Tie-Yan Liu [2]

Abstract

Machine learning is essentially the sciences of playing with data. An adaptive data selection strategy, enabling to dynamically choose different data at various training stages, can reach a more effective model in a more efficient way. In this paper, we propose a deep reinforcement learning framework, which we call ***Neural Data Filter*** (**NDF**), to explore automatic and adaptive data selection in the training process. In particular, NDF takes advantage of a deep neural network to adaptively select and filter important data instances from a sequential stream of training data, such that the future accumulative reward (e.g., the convergence speed) is maximized. In contrast to previous studies in data selection that is mainly based on heuristic strategies, NDF is quite generic and thus can be widely suitable for many machine learning tasks. Taking neural network training with stochastic gradient descent (SGD) as an example, comprehensive experiments with respect to various neural network modeling (e.g., multi-layer perceptron networks, convolutional neural networks and recurrent neural networks) and several applications (e.g., image classification and text understanding) demonstrate that NDF powered SGD can achieve comparable accuracy with standard SGD process by using less data and fewer iterations.

Fan, Yang, et al. "Learning What Data to Learn." *arXiv preprint arXiv:1702.08635* (2017).