Sequential Determinantal Point Processes (SeqDPPs) and Variations for *Supervised* Video Summarization

Boqing Gong

UCF Center for Research in Computer Vision

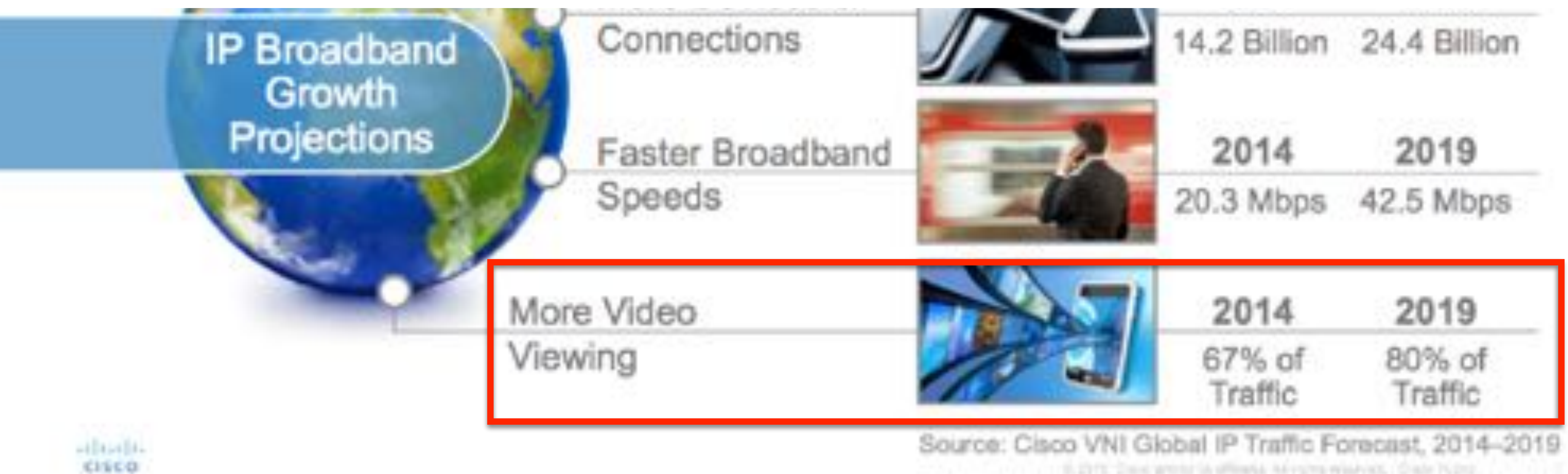
BGong@CRCV.ucf.edu

Big Video on the Internet

Big Video on the Internet

Big Video on the Internet

300 hours of video uploaded *per minute*



Big Video from surveillance

30 million CCTV cameras in US

Ineffective…

Big Video of “first person”

Law enforcement

Life logger

Robot exploring

Need for intelligent methods of video summarization!

Some use cases

Autoplay videos: good idea?
→ Autoplay highlights?

Adaptive fast-forwarding?

Some use cases

Autoplay videos: good idea?
→ Autoplay highlights?

Adaptive fast-forwarding?

HighlightHub

Some use cases

Autoplay videos: good idea?
→ Autoplay highlights?

Adaptive fast-forwarding?

HighlightHub gifs

Some use cases

Autoplay videos: good idea?
→ Autoplay highlights?

Adaptive fast-forwarding?

HighlightHub

gifs

NBA GAME TIME
CRACKLE
NETFLIX
epix
crunchyroll
vudu
hulu PLUS
amazon
NHL GAMECENTER
redbox instant

Video summarization

Extractive video summarization

Subset Selection problem

Compositional video summarization

Limited to well-controlled videos

[Pritch et al.'09]

Video summarization

Extractive video summarization

Subset Selection problem

Compositional video summarization

Limited to well-controlled videos

[Pritch et al.'09]

Two competing criteria

Extracting frames/shots

Individually **important**

Collectively **diverse**

[Wolf 1996, Vasconcelos and Lippman 1998, Aner and Kender 2002, Pal and Jojic 2005, Kang et al. 2006, Pritch et al. 2007, Jiang et al. 2009, Lee and Kwon 2012, Khosla et al. 2013, Kim et al. 2014, Song et al. 2015, Lee and Grauman 2015, …]

Output: a storyboard summary

Prior work

[Wolf 1996, Vasconcelos and Lippman 1998, Aner and Kender 2002, Pal and Jojic 2005, Kang et al. 2006, Pritch et al. 2007, Jiang et al. 2009, Lee and Kwon 2012, Khosla et al. 2013, Kim et al. 2014, Song et al. 2015, Lee and Grauman 2015, …]

Measuring **importance** of frames/shots

- Low-level visual cues, motion cues
- Weakly supervised Web images, texts
- Human labeled objects, attributes, etc.

Prior work

[Wolf 1996, Vasconcelos and Lippman 1998, Aner and Kender 2002, Pal and Jojic 2005, Kang et al. 2006, Pritch et al. 2007, Jiang et al. 2009, Lee and Kwon 2012, Khosla et al. 2013, Kim et al. 2014, Song et al. 2015, Lee and Grauman 2015, …]

Measuring **importance** of frames/shots

- Low-level visual cues, motion cues
- Weakly supervised Web images, texts
- Human labeled objects, attributes, etc.

Cons:

- Indirect cues
- System developers making decisions for users

Our goal:

Supervised video summarization

Learn video summarizer from ***user summaries***

Our goal:
Supervised video summarization

Learn video summarizer from ***user summaries***

What model constitutes a good video summarizer?

Model selection for ***Supervised*** video summarization

Determinantal Point Process (DPP)

Why DPP?

Modeling subset selection

Modeling diversity & importance

A generative probabilistic model

- Supervised video summarization
- Maximum likelihood & large-margin estimation

Effective for document summarization

This talk

DPP

SeqDPP

Variations

Lessons Learned

DPP

Large-margin DPP

BGong@CRCV.ucf.edu

Discrete point process

- N items (e.g., images or sentences):

$$\mathcal{Y} = \{1, 2, ..., N\}$$

- 2^N possible subsets
- Probability measure $\mathcal{P}$ over subsets $Y \subseteq \mathcal{Y}$

Discrete point process

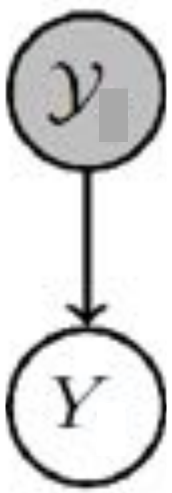
- N items (e.g., images or sentences):

$$\mathcal{Y} = \{1, 2, ..., N\}$$

- 2^N possible subsets
- Probability measure $\mathcal{P}$ over subsets $Y \subseteq \mathcal{Y}$

Vanilla DPP is a discrete point process.

Determinantal point process (DPP)



$$P(Y = \{2,4\})$$

$$\mathcal{Y} = \{1,2,3,4,5\}$$

$Y \subseteq \mathcal{Y}$: subset selection variable

Vanilla DPP is a discrete point process.

Determinantal point process (**DPP**)

$$
\begin{array}{c|ccccc}
 & 1 & 2 & 3 & 4 & 5 \\
\hline
1 & S_{11} & S_{12} & S_{13} & S_{14} & S_{15} \\
2 & S_{21} & S_{22} & S_{23} & S_{24} & S_{25} \\
3 & S_{31} & S_{32} & S_{33} & S_{34} & S_{35} \\
4 & S_{41} & S_{42} & S_{43} & S_{44} & S_{45} \\
5 & S_{51} & S_{52} & S_{53} & S_{54} & S_{55}
\end{array}
$$

$$P(Y = \{2, 4\})$$

$$\mathcal{Y} = \{1, 2, 3, 4, 5\}$$

$Y \subseteq \mathcal{Y}$: subset selection variable

Vanilla DPP is a discrete point process.

Determinantal point process (DPP)

$$
\begin{array}{c|ccccc}
 & 1 & \mathbf{2} & 3 & \mathbf{4} & 5 \\
\hline
1 & S_{11} & S_{12} & S_{13} & S_{14} & S_{15} \\
\mathbf{2} & S_{21} & \boxed{S_{22}} & S_{23} & \boxed{S_{24}} & S_{25} \\
3 & S_{31} & S_{32} & S_{33} & S_{34} & S_{35} \\
\mathbf{4} & S_{41} & \boxed{S_{42}} & S_{43} & \boxed{S_{44}} & S_{45} \\
5 & S_{51} & S_{52} & S_{53} & S_{54} & S_{55}
\end{array}
$$

$$
P(Y = \{2, 4\}) \propto \det \begin{pmatrix} S_{22} & S_{24} \\ S_{42} & S_{44} \end{pmatrix}
$$

$$\mathcal{Y} = \{1, 2, 3, 4, 5\}$$

$Y \subseteq \mathcal{Y}$: subset selection variable

Vanilla DPP is a discrete point process.

DPP models diversity & importance

Items 2 and 4

diverse, larger probability

important, larger probability

$$P(Y = \{2,4\})$$
$$\propto \det \begin{pmatrix} S_{22} & S_{24} \\ S_{42} & S_{44} \end{pmatrix}$$
$$= S_{22} \cdot S_{44} - S_{24} \cdot S_{42}$$

DPP models diversity & importance

	1	2	3	4	5
1	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}
2	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}
3	S_{31}	S_{32}	S_{33}	S_{34}	S_{35}
4	S_{41}	S_{42}	S_{43}	S_{44}	S_{45}
5	S_{51}	S_{52}	S_{53}	S_{54}	S_{55}

$$P(Y = \{2,4\})$$
$$\propto \det \begin{pmatrix} S_{22} & S_{24} \\ S_{42} & S_{44} \end{pmatrix}$$
$$= S_{22} \cdot S_{44} - S_{24} \cdot S_{42}$$

importance

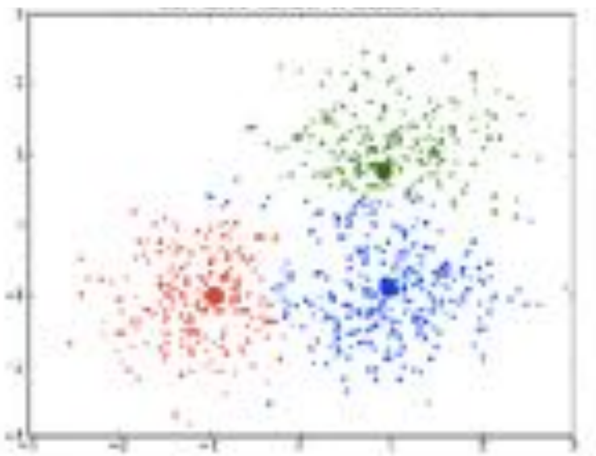
DPP models diversity & importance

$$
\begin{array}{c|ccccc}
 & 1 & 2 & 3 & 4 & 5 \\
\hline
1 & S_{11} & S_{12} & S_{13} & S_{14} & S_{15} \\
2 & S_{21} & S_{22} & S_{23} & S_{24} & S_{25} \\
3 & S_{31} & S_{32} & S_{33} & S_{34} & S_{35} \\
4 & S_{41} & S_{42} & S_{43} & S_{44} & S_{45} \\
5 & S_{51} & S_{52} & S_{53} & S_{54} & S_{55}
\end{array}
$$

$$
\begin{aligned}
&P(Y = \{2,4\}) \\
&\propto \det \begin{pmatrix} S_{22} & S_{24} \\ S_{42} & S_{44} \end{pmatrix} \\
&= S_{22} \cdot S_{44} - S_{24} \cdot S_{42}
\end{aligned}
$$

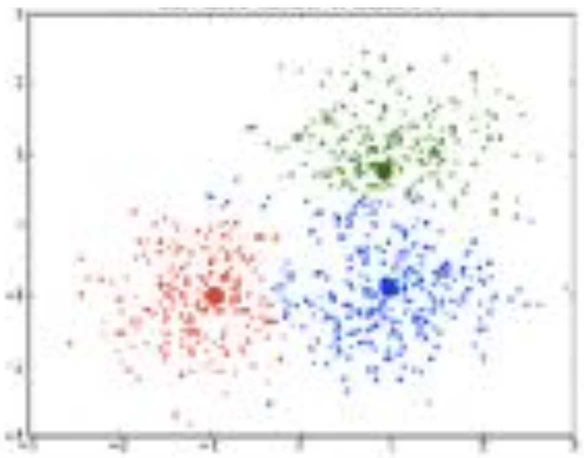
Diversity

Diversity

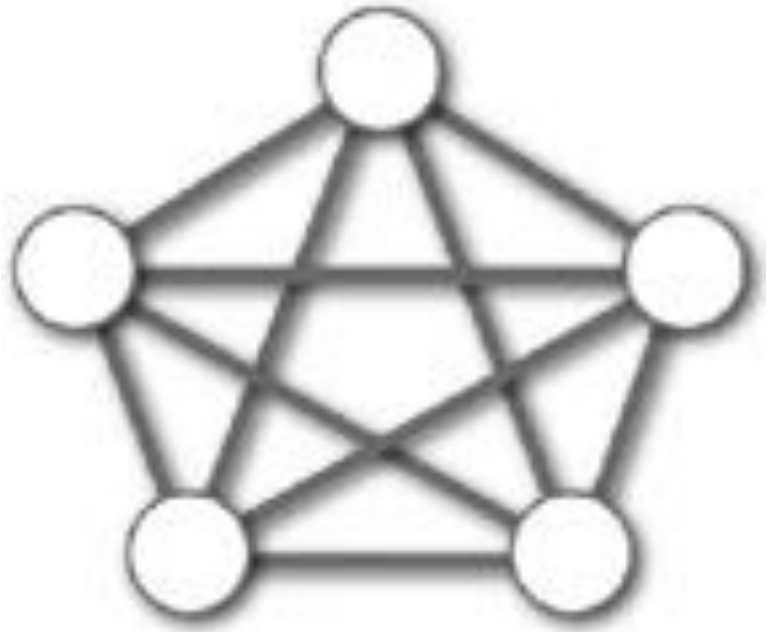


Clustering

Diversity

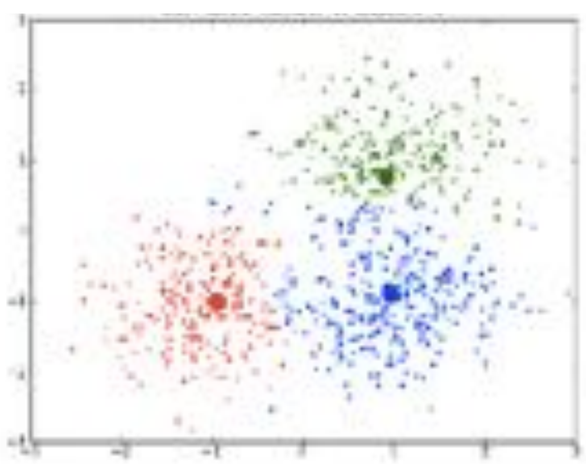


Clustering

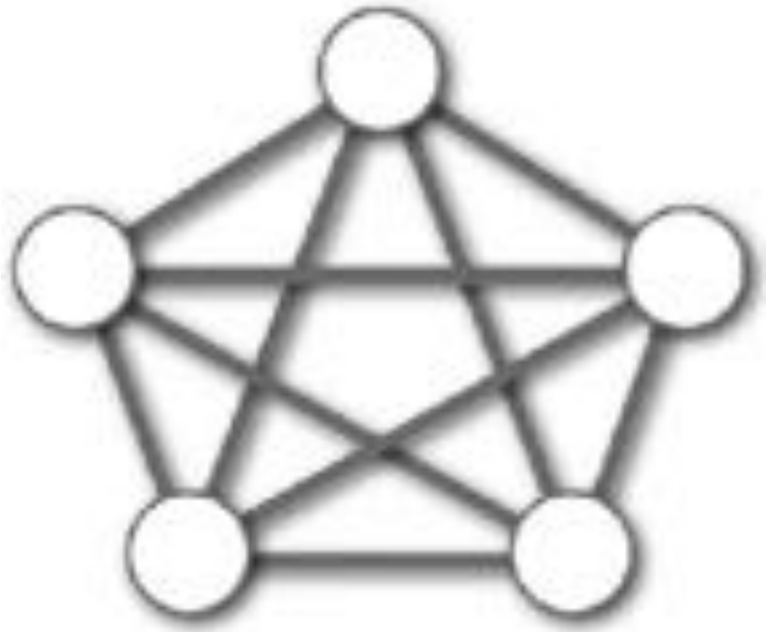


MRF

Diversity



Clustering



MRF

DPP

Diversity

	MRF	DPP
Inference	NP	Mostly tractable
MAP inference	NP	NP
Approx. MAP	Likewise NP	1/4

DPP: some properties

Modeling subset selection, diversity, & importance

Log-submodular

- MAP inference is NP-hard
- 1/4-approximation under some constraints

Efficient sampling

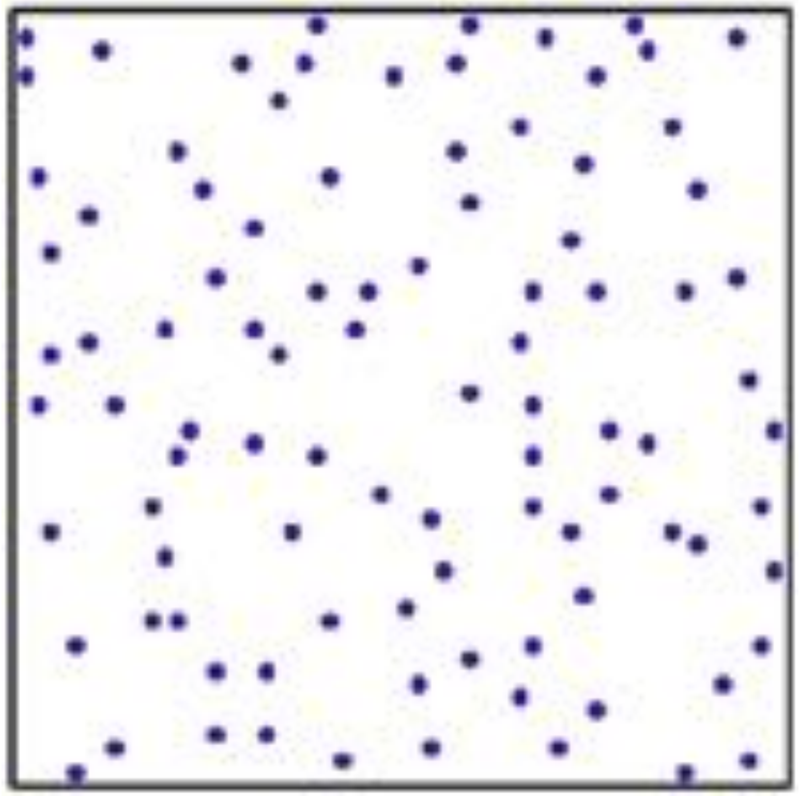
- Two-stage sampling, MCMC sampling

Closed-form marginalization & conditioning

The family of DPPs

- DPP

$$P(Y) \propto \det\left(L_Y\right)$$



The family of DPPs

- DPP $$P(Y) \propto \det\left(L_Y\right)$$
- k-DPP [Kulesza & Taskar, 2011] $$\text{s.t.} \quad \mathrm{CARD}(Y) = k$$

The family of DPPs

- DPP
- k-DPP [Kulesza & Taskar, 2011]
- Markov DPP [Affandi et al., 2012]

The family of DPPs

- DPP
- k-DPP [Kulesza & Taskar, 2011]
- Markov DPP [Affandi et al., 2012]
- Structured DPP [Kulesza & Taskar, 2010]

The family of DPPs

- DPP
- k-DPP [Kulesza & Taskar, 2011]
- Markov DPP [Affandi et al., 2012]
- Structured DPP [Kulesza & Taskar, 2010]
- Continuous DPP [Affandi et al., 2013]

The family of DPPs

- DPP
- k-DPP [Kulesza & Taskar, 2011]
- Markov DPP [Affandi et al., 2012]
- Structured DPP [Kulesza & Taskar, 2010]
- Continuous DPP [Affandi et al., 2013]
- **Sequential DPP** **[Gong et al., NIPS'14, UAI'15]**
  [ECCV'16, CVPR'17, ICML submitted]

This talk

DPP

SeqDPP

Variations

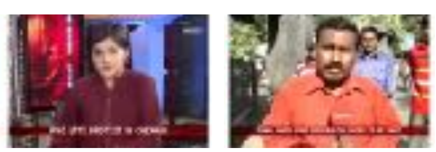
Lessons Learned

Vanilla DPP for supervised video summarization

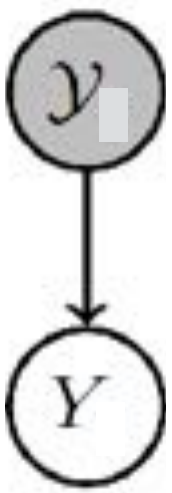
BGong@CRCV.ucf.edu

Video summarization by vanilla DPP

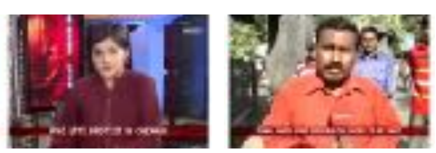
$\mathcal{Y} = \{1, 2, 3, 4, 5\} = \Big\{$  $\Big\}$

 $\leftarrow \arg\max_{y} P(Y = y)$

Video summarization by vanilla DPP



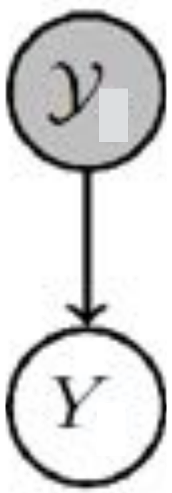
$$\mathcal{Y} = \{1, 2, 3, 4, 5\} = \Big\{$$  $$\Big\}$$



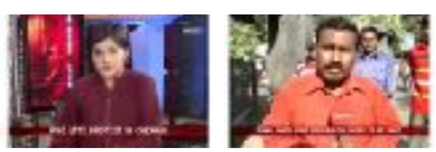
$$\leftarrow \arg\max_{y} P(Y = y)$$

	1	2	3	4	5
1	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}
2	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}
3	S_{31}	S_{32}	S_{33}	S_{34}	S_{35}
4	S_{41}	S_{42}	S_{43}	S_{44}	S_{45}
5	S_{51}	S_{52}	S_{53}	S_{54}	S_{55}

Video summarization by vanilla DPP



$\mathcal{Y} = \{1, 2, 3, 4, 5\} = \{$  $\}$

 $\leftarrow \arg\max_{y} P(Y = y)$

	1	2	3	4	5
1	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}
2	S_{21}			S_{24}	S_{25}
3	S_{31}			S_{34}	S_{35}
4	S_{41}	S_{42}	S_{43}	S_{44}	S_{45}
5	S_{51}	S_{52}	S_{53}	S_{54}	S_{55}

Parameterizing kernels for out-of-sample extension

$$L_{ij} = \langle f(\mathbf{x}_i), f(\mathbf{x}_j) \rangle$$

1-layer neural network: $f(\mathbf{x}) = W \tanh(U\mathbf{x})$

Linear: $f(\mathbf{x}) = W\mathbf{x}$

$$\begin{array}{c|c|c|c|c|c}
 & 1 & 2 & 3 & 4 & 5 \\
1 & S_{11} & S_{12} & S_{13} & S_{14} & S_{15} \\
\hline
2 & S_{21} & S_{22} & S_{23} & S_{24} & S_{25} \\
\hline
3 & S_{31} & S_{32} & S_{33} & S_{34} & S_{35} \\
\hline
4 & S_{41} & S_{42} & S_{43} & S_{44} & S_{45} \\
\hline
5 & S_{51} & S_{52} & S_{53} & S_{54} & S_{55}
\end{array}$$

Parameterizing kernels for out-of-sample extension

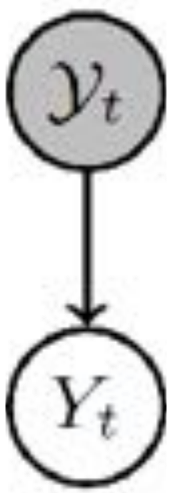
$$L_{ij} = \langle f(\mathbf{x}_i), f(\mathbf{x}_j) \rangle$$

1-layer neural network: $f(\mathbf{x}) = W \tanh(U\mathbf{x})$

Linear: $f(\mathbf{x}) = W\mathbf{x}$

	1	2	3	4	5
1	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}
2	S_{21}			S_{24}	S_{25}
3	S_{31}			S_{34}	S_{35}
4	S_{41}	S_{42}	S_{43}	S_{44}	S_{45}
5	S_{51}	S_{52}	S_{53}	S_{54}	S_{55}

Learning kernels by maximum likelihood estimation (MLE)

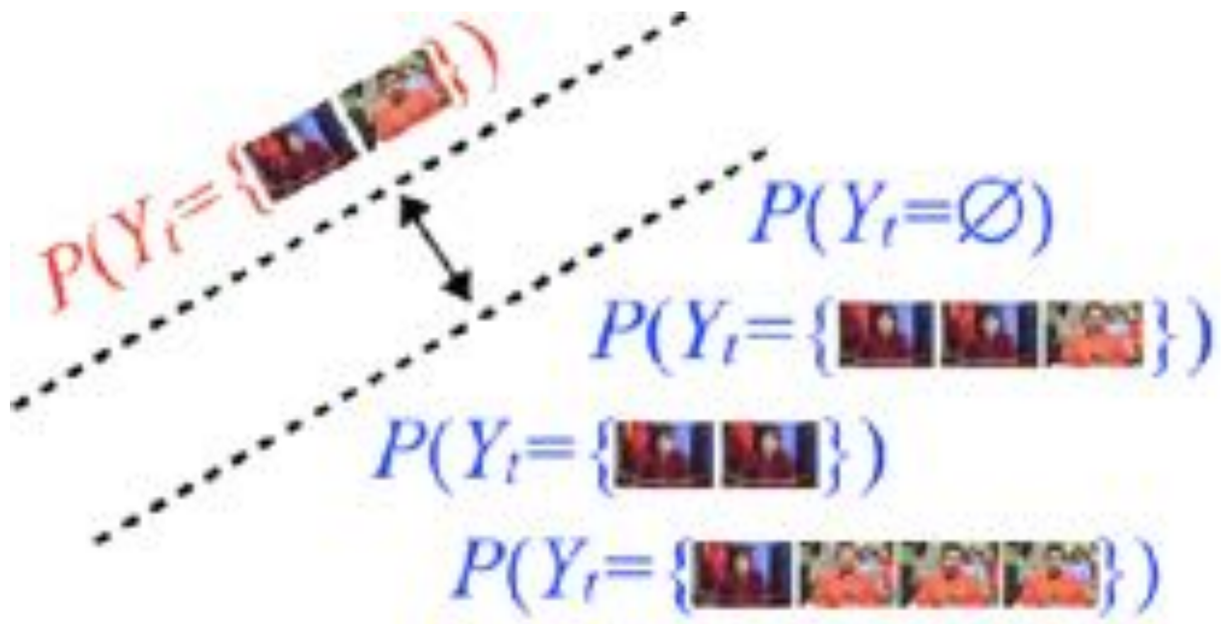


$$\begin{array}{c|ccccc} & 1 & 2 & 3 & 4 & 5 \\ \hline 1 & S_{11} & S_{12} & S_{13} & S_{14} & S_{15} \\ 2 & S_{21} & S_{22} & S_{23} & S_{24} & S_{25} \\ 3 & S_{31} & S_{32} & S_{33} & S_{34} & S_{35} \\ 4 & S_{41} & S_{42} & S_{43} & S_{44} & S_{45} \\ 5 & S_{51} & S_{52} & S_{53} & S_{54} & S_{55} \end{array} \longleftarrow$$

MLE

Image credits: [Lu and Grauman, CVPR'13]

Learning kernels by the large-margin principle

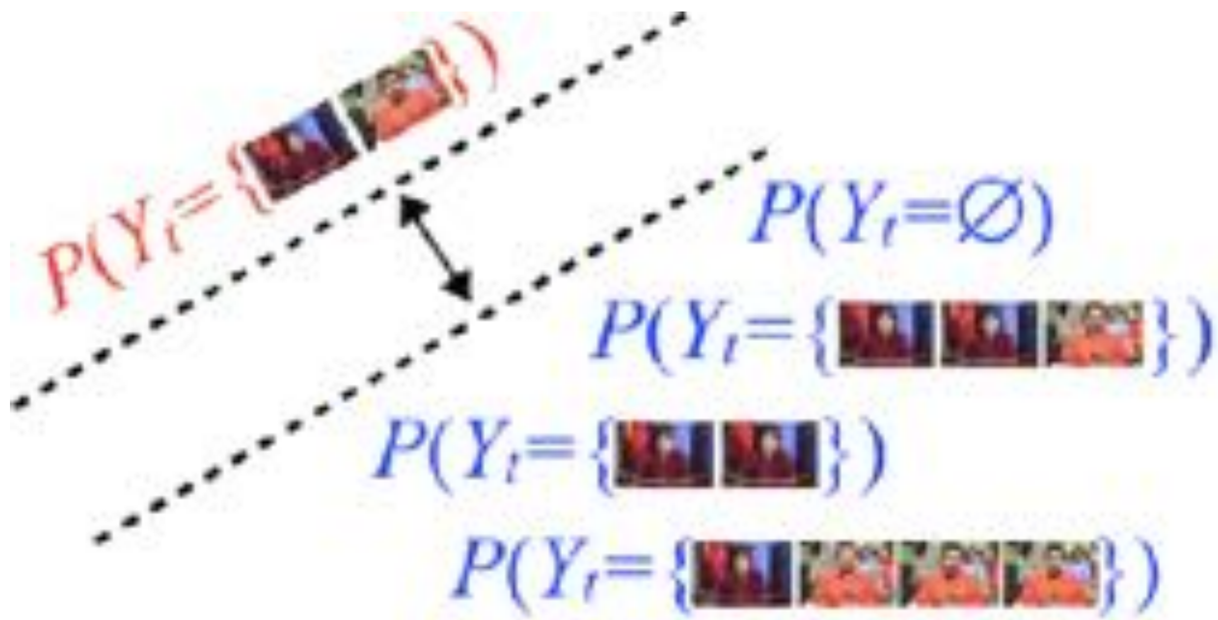


Advantages over MLE

Tracking errors

Accepting various margins (e.g., trade-off precision & recall)

Learning kernels by the large-margin principle



Main challenge:

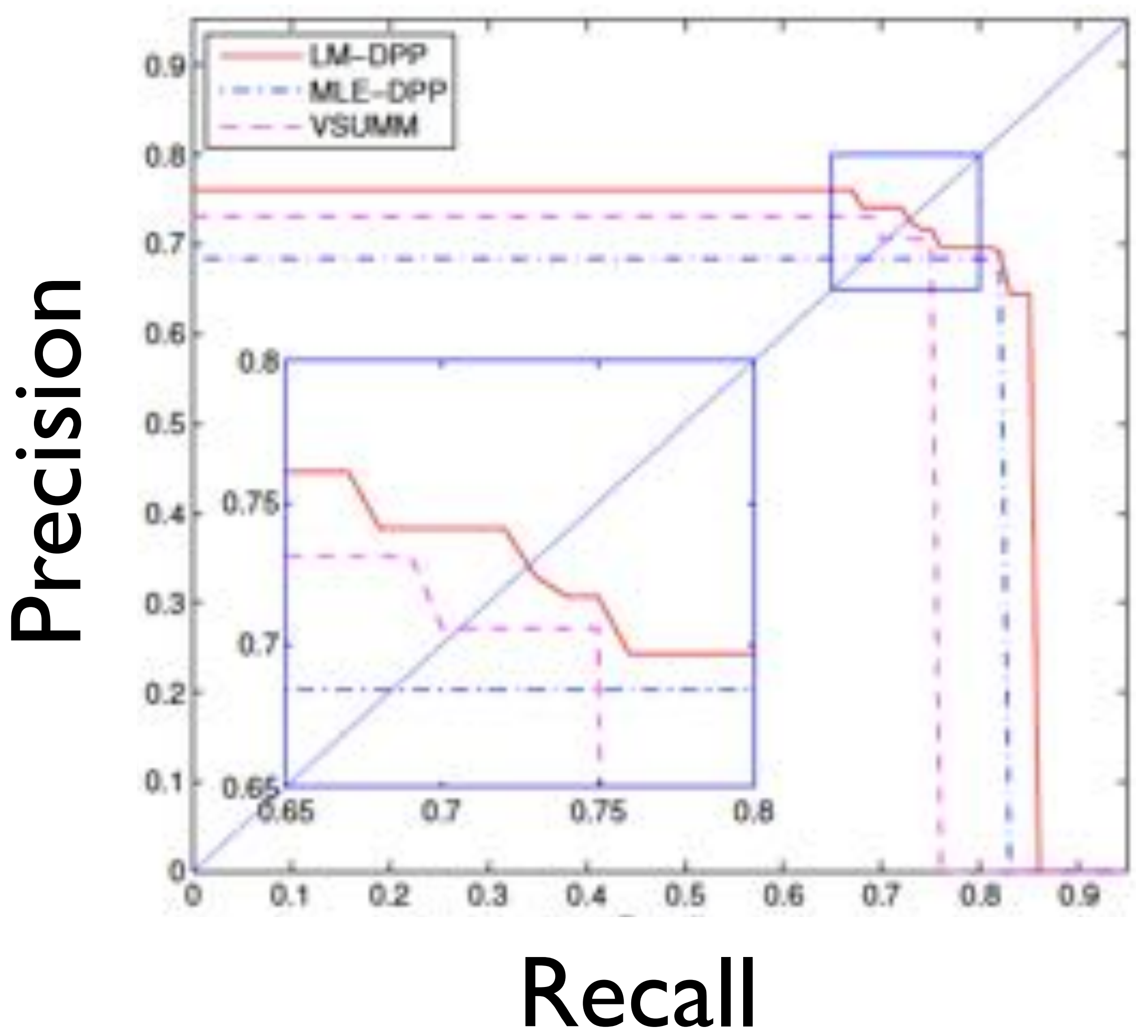
An exponential number of negative examples

Solution:

Multiplicative margin
Upper bound by softmax

[UAI'15]

Large-margin DPP better balances precision & recall



[UAI'15]

Video summarization by vanilla DPP: what's missing?

DPP fails to capture the ***temporal structure*** of videos

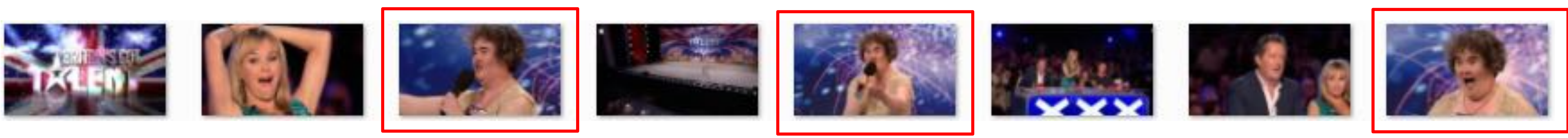
Video summarization by vanilla DPP: what's missing?

DPP fails to capture the ***temporal structure*** of videos

Susan Boyle performs in "Britain's Got Talent".

Video summarization by vanilla DPP: what's missing?

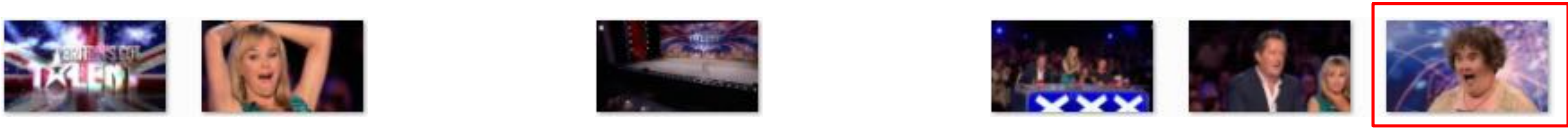
DPP fails to capture the ***temporal structure*** of videos



Susan Boyle performs in “Britain's Got Talent”.

Video summarization by vanilla DPP: what's missing?

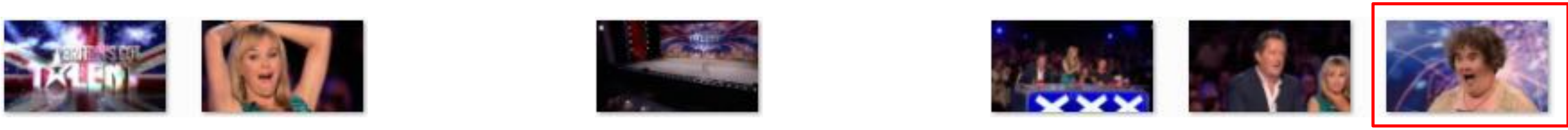
DPP fails to capture the ***temporal structure*** of videos



Susan Boyle performs in "Britain's Got Talent".

Video summarization by vanilla DPP: what's missing?

DPP fails to capture the ***temporal structure*** of videos



Susan Boyle performs in "Britain's Got Talent".

"Britain's Got Talent" … surprises a lady.

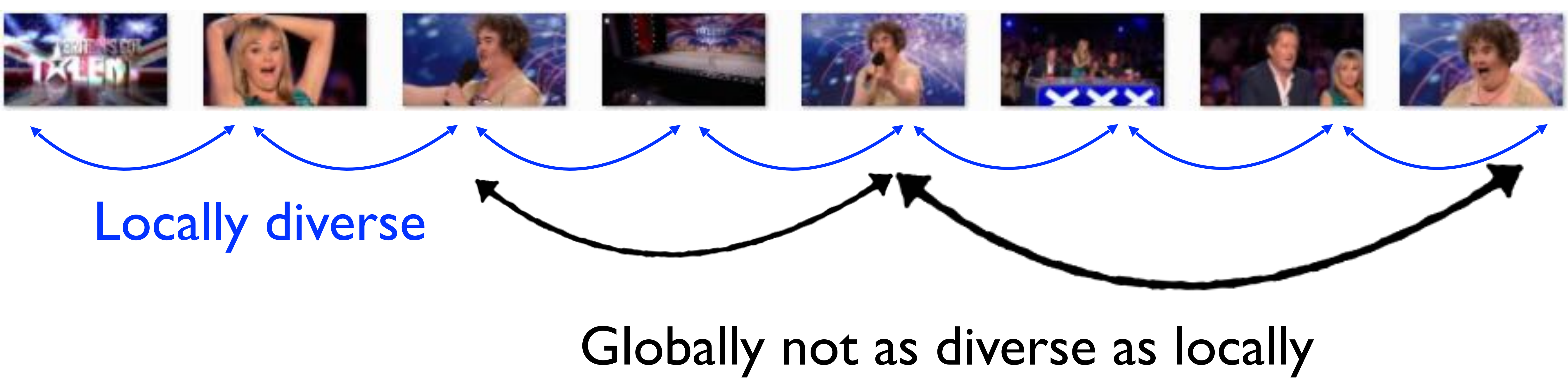
Video summarization by vanilla DPP: what's missing?

DPP fails to capture the ***temporal structure*** of videos

Susan Boyle performs in "Britain's Got Talent".

"Britain's Got Talent" … surprises a lady.

Need of a “sequential” DPP



This talk

DPP

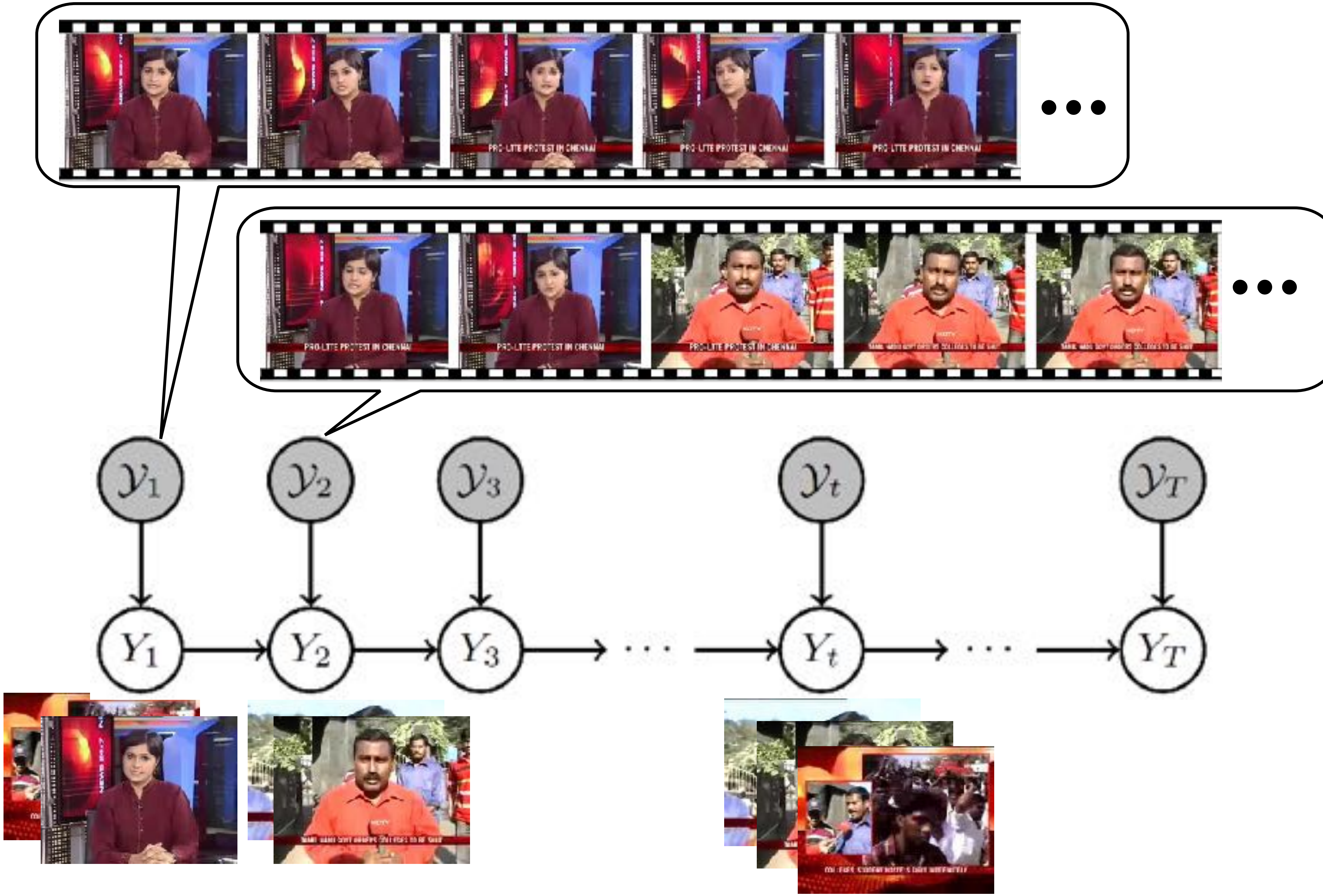
SeqDPP

Variations

Lessons Learned

Sequential DPP for supervised video summarization

BGong@CRCV.ucf.edu

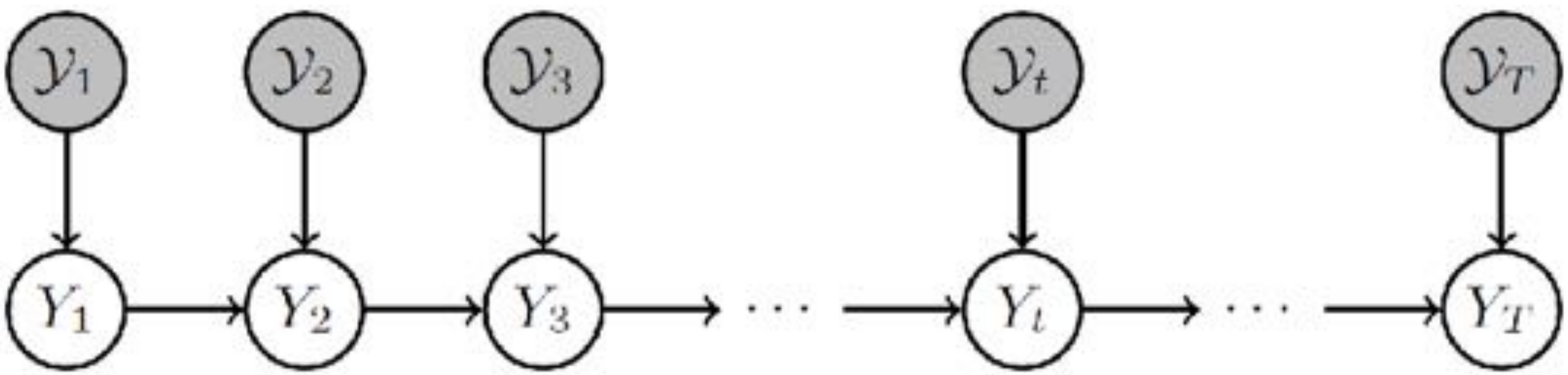

$\mathcal{Y}_1$
$\mathcal{Y}_2$
$\mathcal{Y}_3$
$\mathcal{Y}_t$
$\mathcal{Y}_T$
Y_1
Y_2
Y_3
Y_t
Y_T
PRO-LTTE PROTEST IN CHENNAI

[NIPS'14]

Sequential DPP (seqDPP)

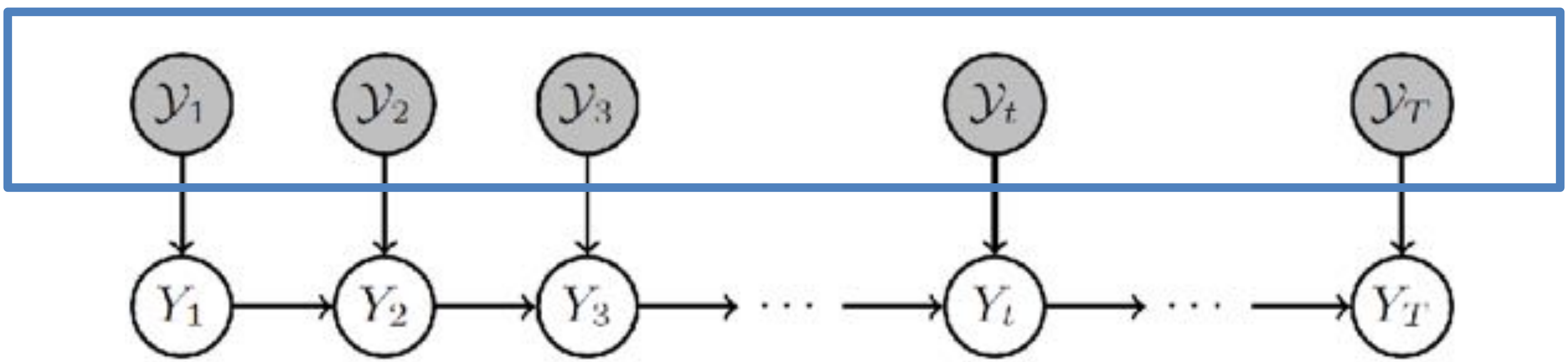
[NIPS'14]

Sequential DPP (seqDPP)



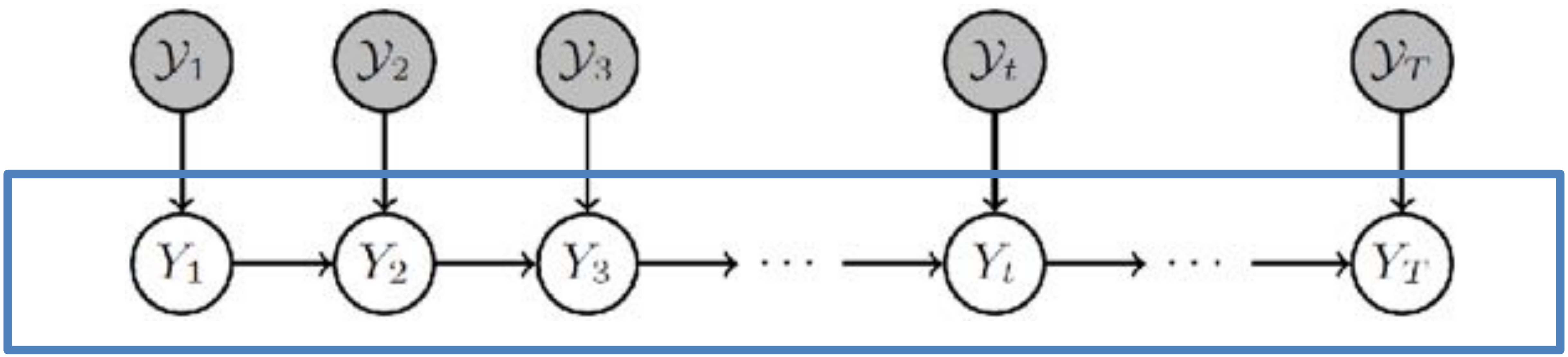
[NIPS'14]

Sequential DPP (seqDPP)



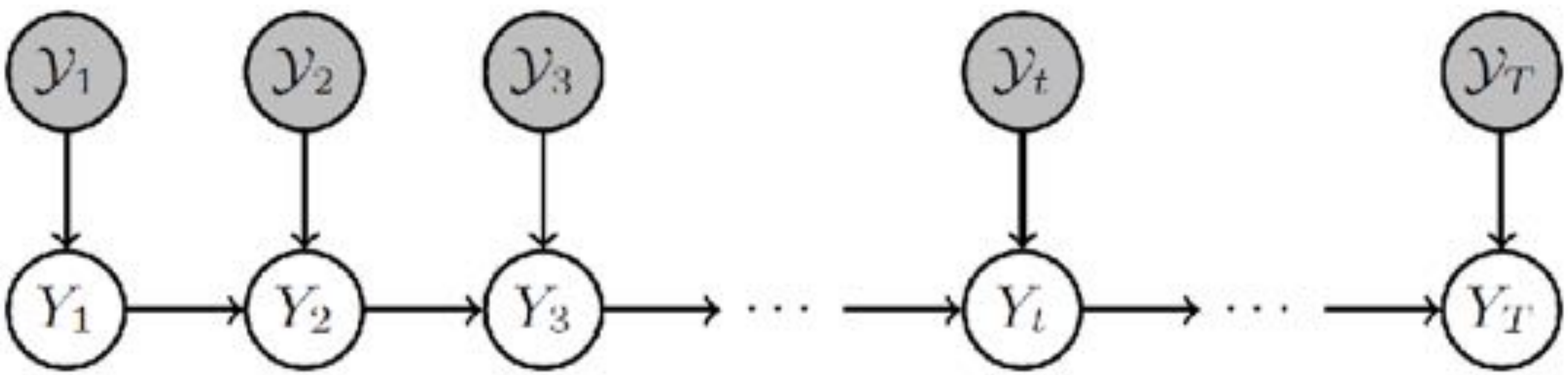
[NIPS'14]

Sequential DPP (seqDPP)



[NIPS'14]

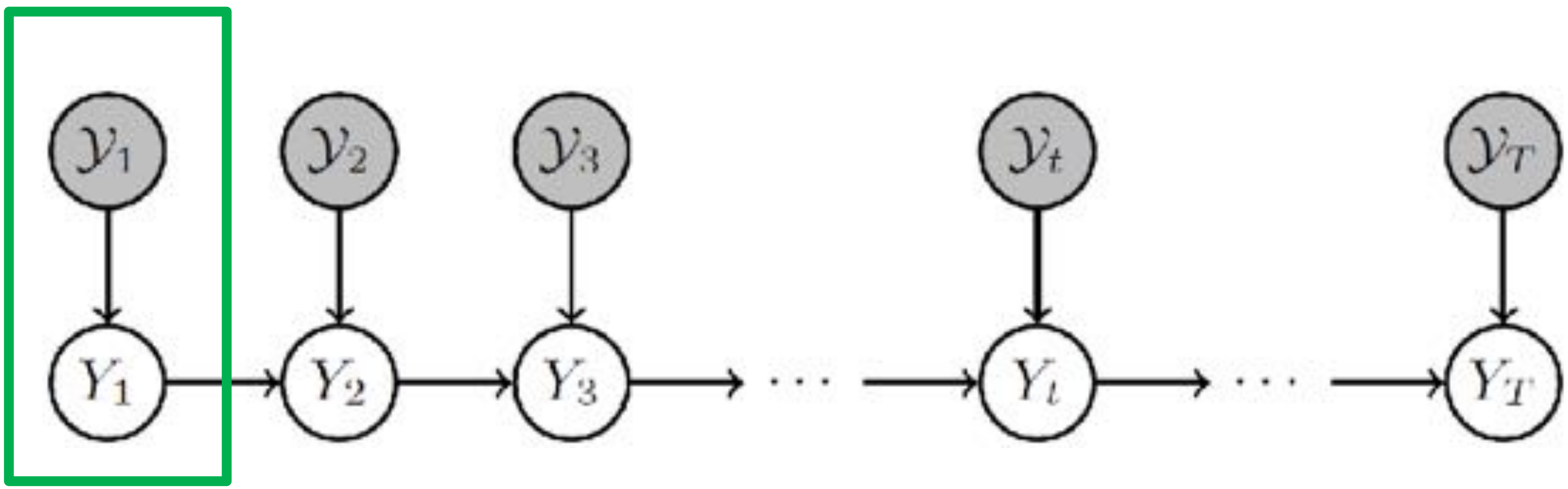
Sequential DPP (seqDPP)



$$P(Y_1 = \boldsymbol{y}_1, Y_2 = \boldsymbol{y}_2, \cdots, Y_T = \boldsymbol{y}_T) = P(Y_1 = \boldsymbol{y}_1) \prod_{t=2} P(Y_t = \boldsymbol{y}_t | Y_{t-1} = \boldsymbol{y}_{t-1})$$

[NIPS'14]

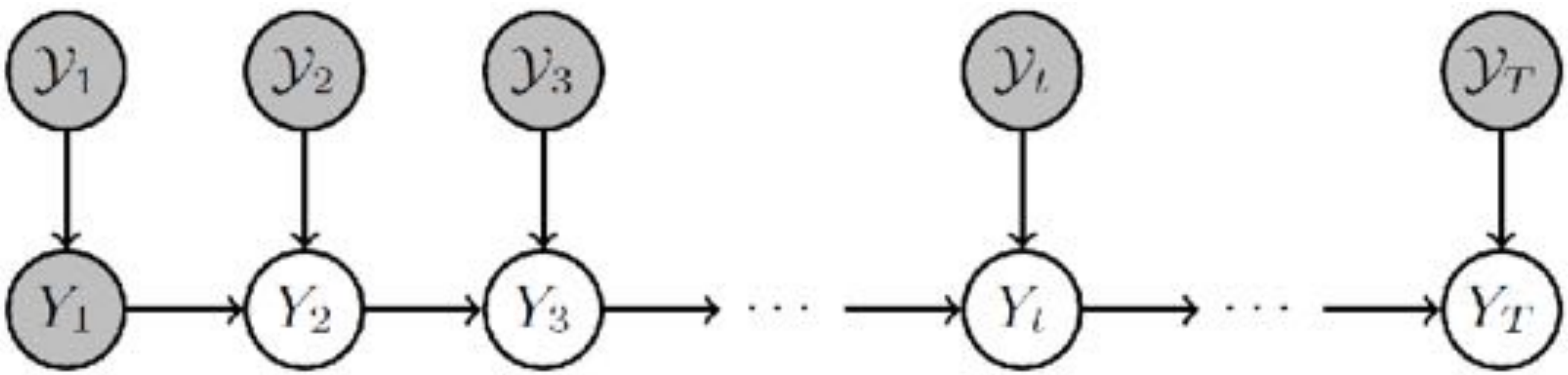
Sequential DPP (seqDPP)



$$P(Y_1 = \boldsymbol{y}_1, Y_2 = \boldsymbol{y}_2, \cdots, Y_T = \boldsymbol{y}_T) = \boxed{P(Y_1 = \boldsymbol{y}_1)} \prod_{t=2} P(Y_t = \boldsymbol{y}_t | Y_{t-1} = \boldsymbol{y}_{t-1})$$

[NIPS'14]

Sequential DPP (seqDPP)

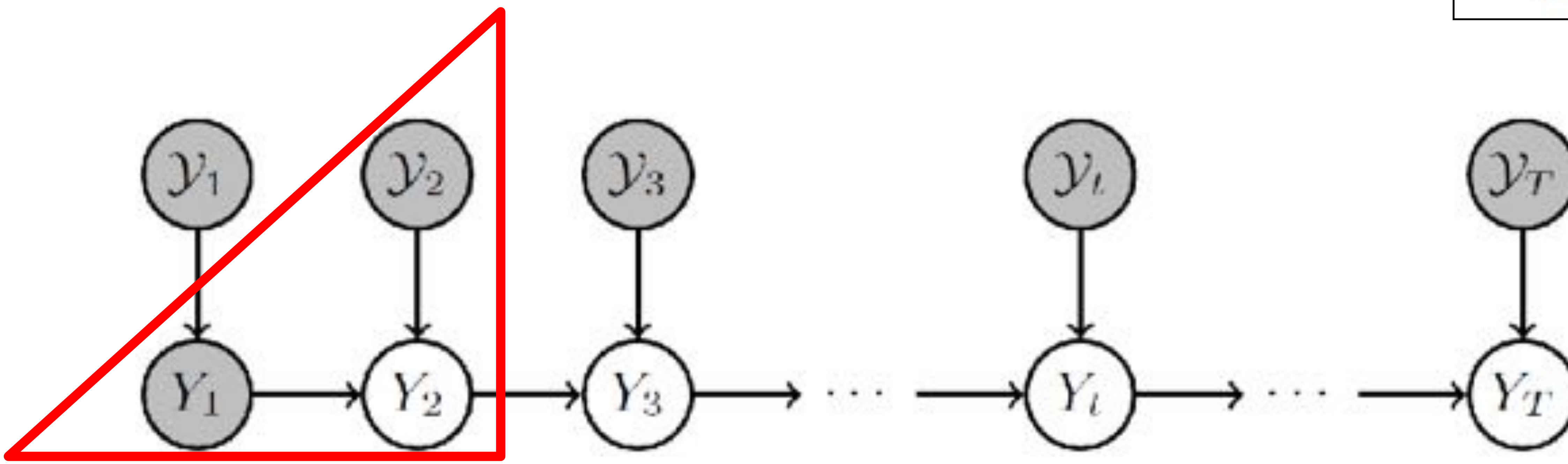


$$P(Y_1 = \boldsymbol{y}_1, Y_2 = \boldsymbol{y}_2, \cdots, Y_T = \boldsymbol{y}_T) = P(Y_1 = \boldsymbol{y}_1) \prod_{t=2} P(Y_t = \boldsymbol{y}_t | Y_{t-1} = \boldsymbol{y}_{t-1})$$

[NIPS'14]

Sequential DPP (seqDPP)

$$\boxed{L_{Y_1 \cup \mathcal{Y}_2}}$$

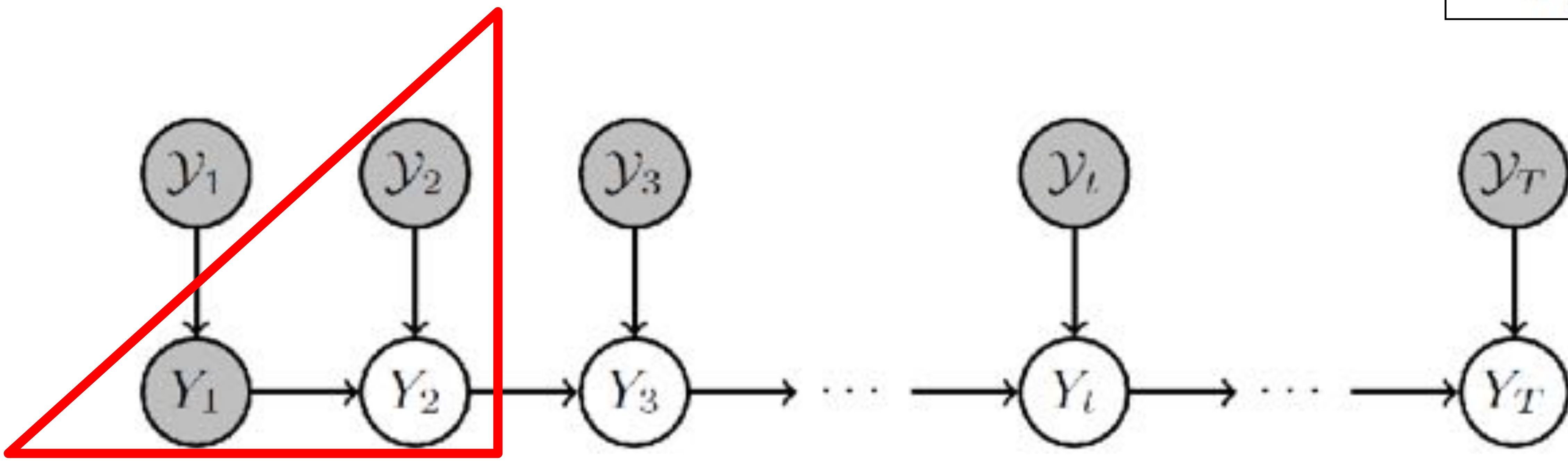


$$P(Y_1 = \boldsymbol{y}_1, Y_2 = \boldsymbol{y}_2, \cdots, Y_T = \boldsymbol{y}_T) = P(Y_1 = \boldsymbol{y}_1) \prod_{t=2} \boxed{P(Y_t = \boldsymbol{y}_t | Y_{t-1} = \boldsymbol{y}_{t-1})}$$

[NIPS'14]

Sequential DPP (seqDPP)

$$\boxed{L_{Y_1 \cup \mathcal{Y}_2}}$$

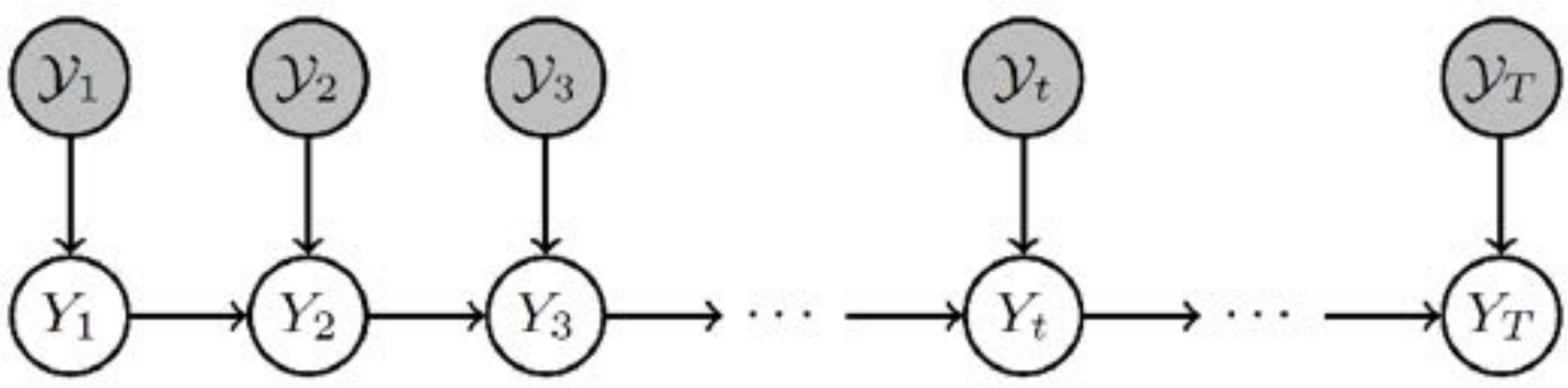


$$P(Y_1 = \boldsymbol{y}_1, Y_2 = \boldsymbol{y}_2, \cdots, Y_T = \boldsymbol{y}_T) = P(Y_1 = \boldsymbol{y}_1) \prod_{t=2} \boxed{P(Y_t = \boldsymbol{y}_t | Y_{t-1} = \boldsymbol{y}_{t-1})}$$

Conditional probability: still a DPP !

[NIPS'14]

Advantages of SeqDPP



Modeling importance, diversity, and ***sequential*** structure

More efficient inference: $O(2^N) \rightarrow O(M \cdot 2^{N/M})$

Summarizing streaming videos on the fly

Experimental study

Three benchmark datasets:

Open video project, Youtube (50), Kodak

Preprocessing: down-sampling 1 frame/sec

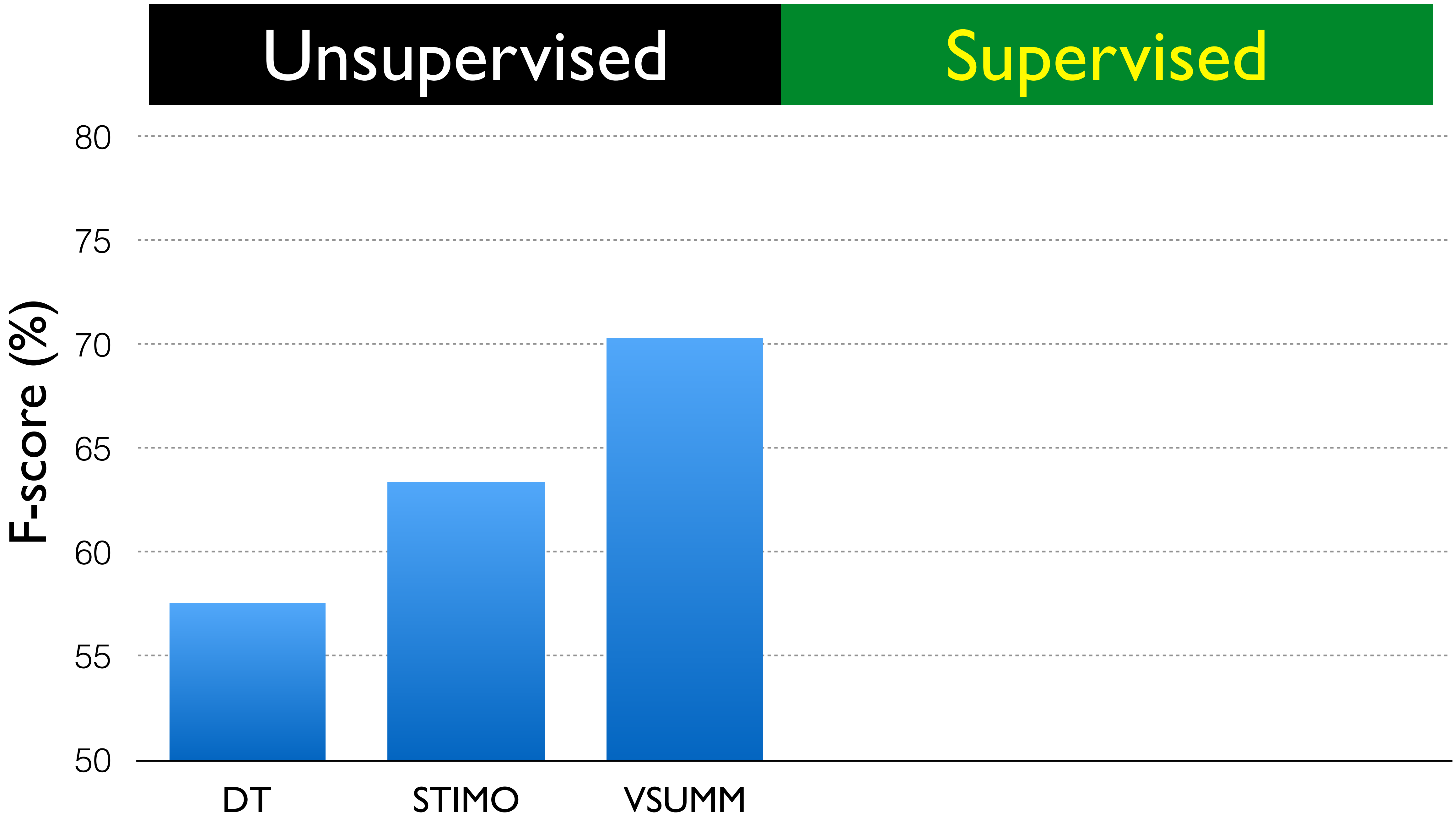
Features: saliency, Fisher vectors, context

Evaluation:

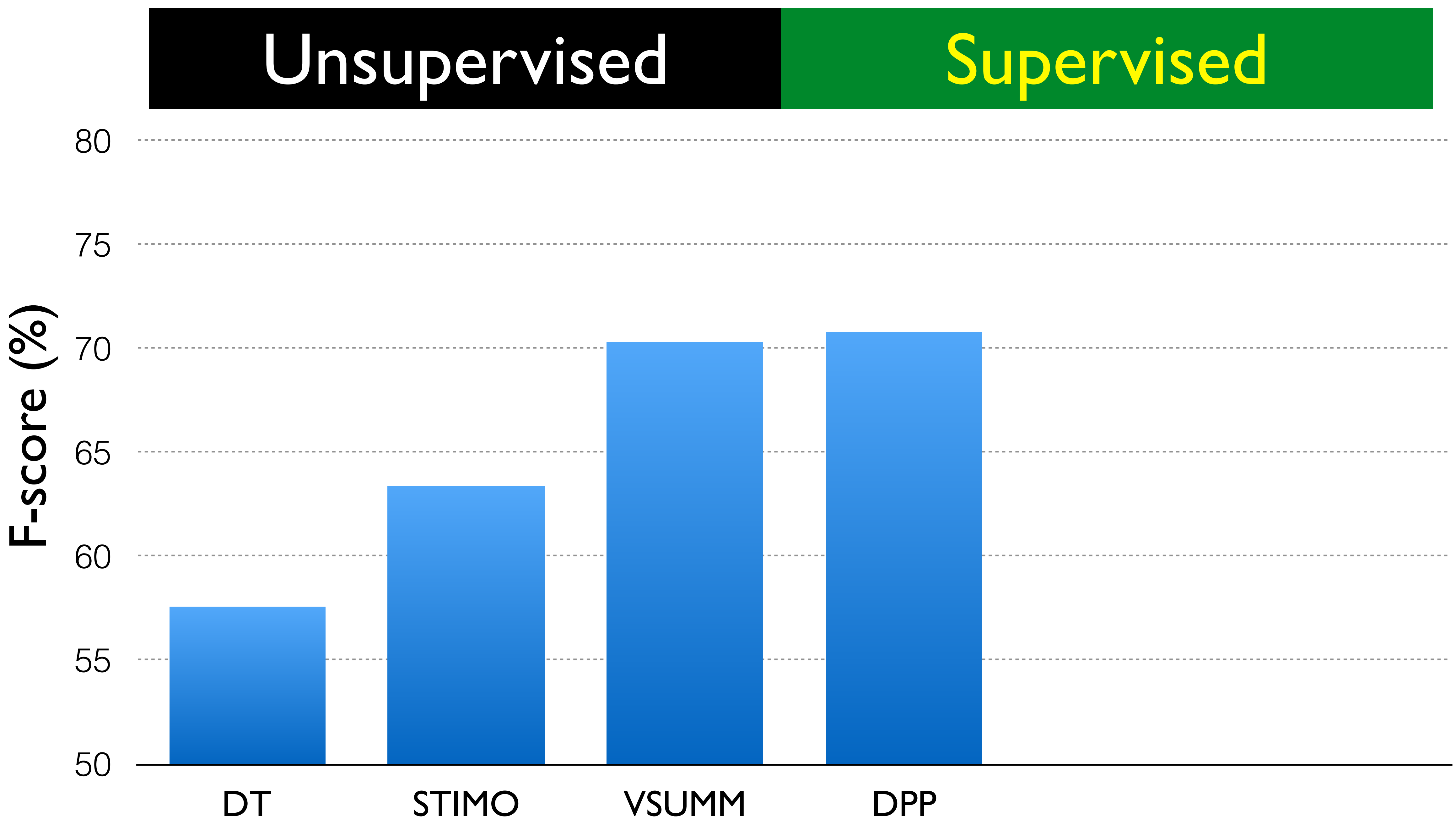
Precision, recall, F-score by the VSUMM package

[Avila et al.'10]

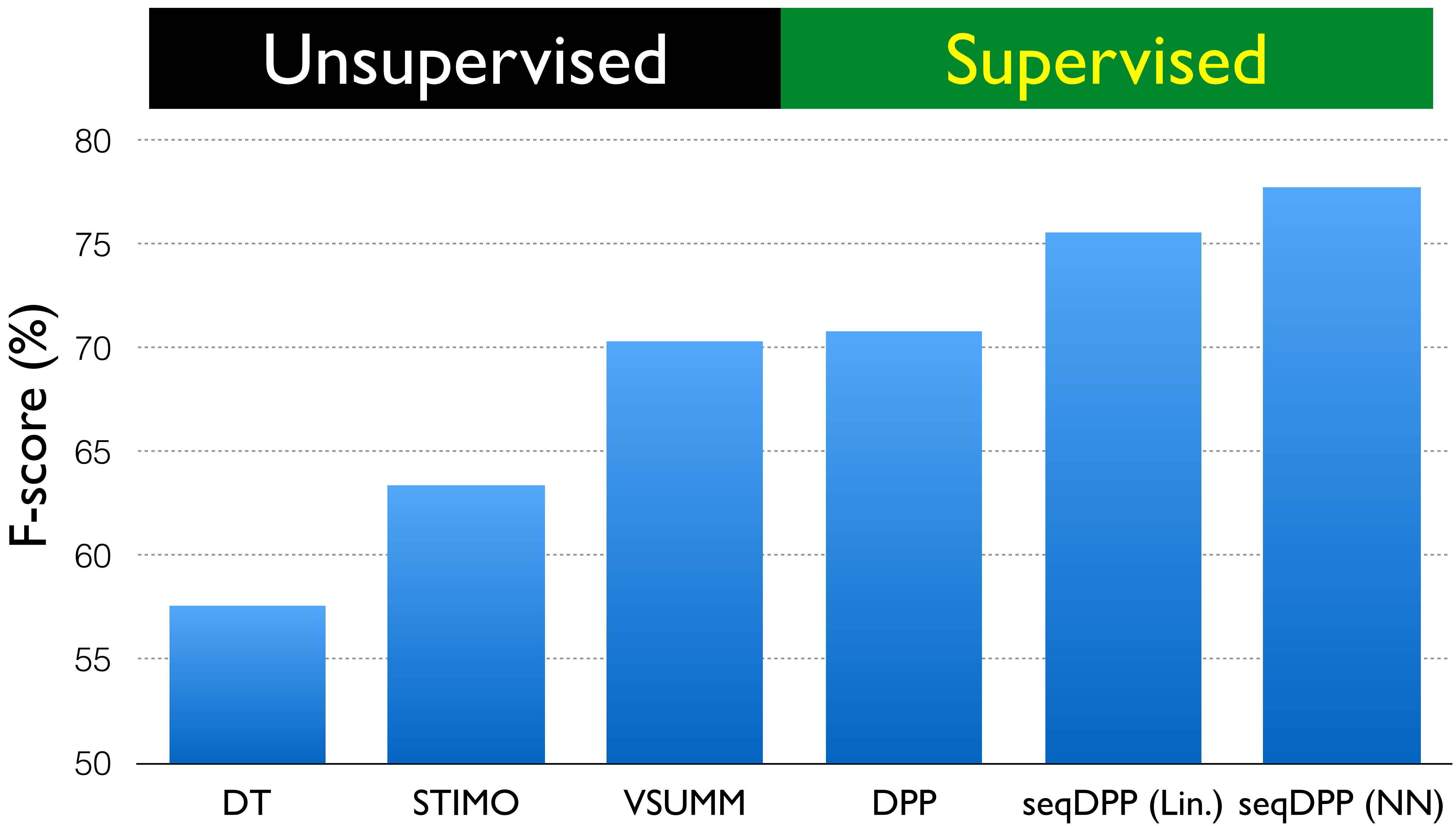
Experimental results



Experimental results



Experimental results



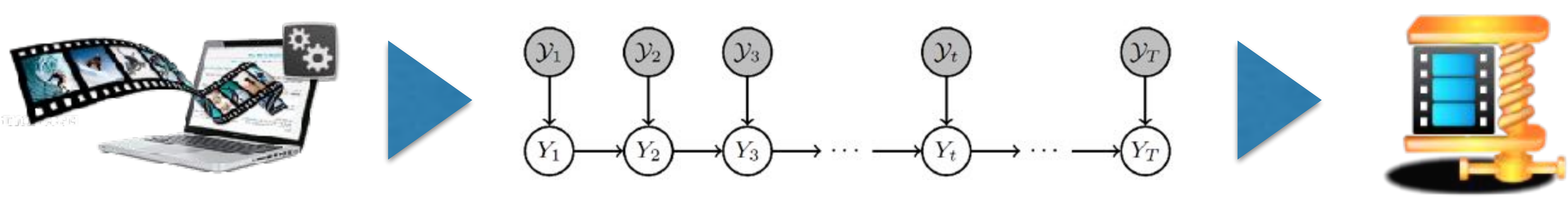
Thus far,

Supervised *video summarization*

DPP: MLE & large-margin

Sequential DPP

Experimental results & analysis



Thus far,

Supervised *video summarization*

DPP: MLE & large-margin

Sequential DPP

Experimental results & analysis

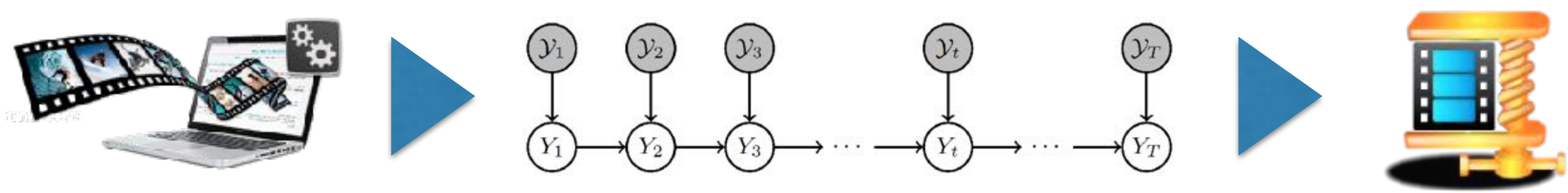
Lessons learned

Video summarization is ***subjective***

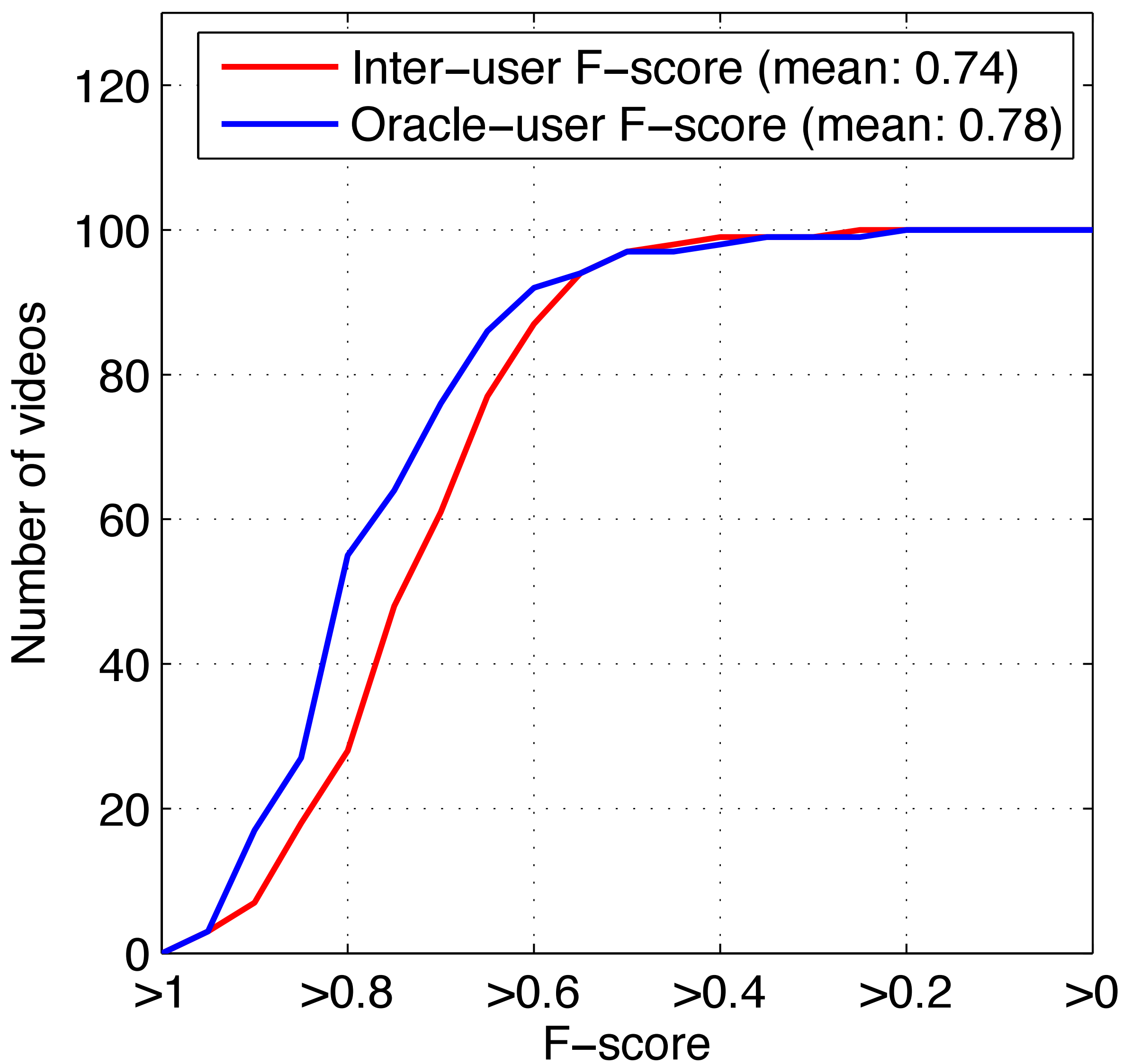
1. <u>Personalization</u>

System needs a channel to infer user's preference

2. <u>Evaluation is hard</u>



Inter-user agreement



100 videos

Five summaries per video

*No "**groundtruth**" summary*

Fairly high inter-user agreement

This talk

DPP

SeqDPP

Variations

Lessons Learned

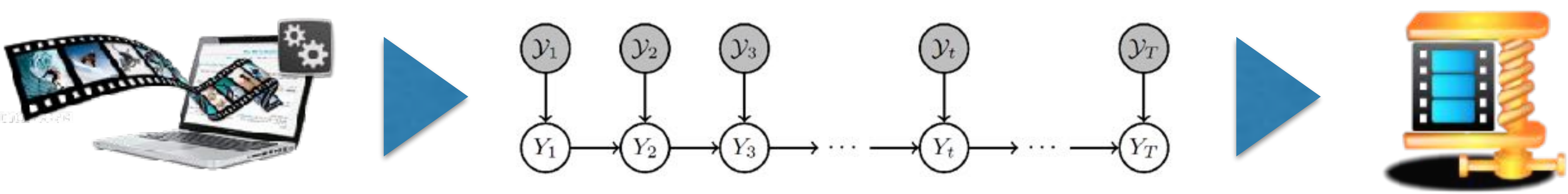
User-subjectivity

1. Personalizing video summarizers
2. An improved evaluation metric

BGong@CRCV.ucf.edu

Query-focused video summarization

(a) Input: Video & Query **(b) Algorithm**: Sequential & Hierarchical Determinantal Point Process (SH-DPP) **(c) Output**: Summary



[ECCV'16, CVPR'17]

Query-focused video summarization

Decision to include a frame/short in summary

- Relevance to query (*be responsive to user input*)
- Importance in the context (*maintain story flow*)
- Collective diversity

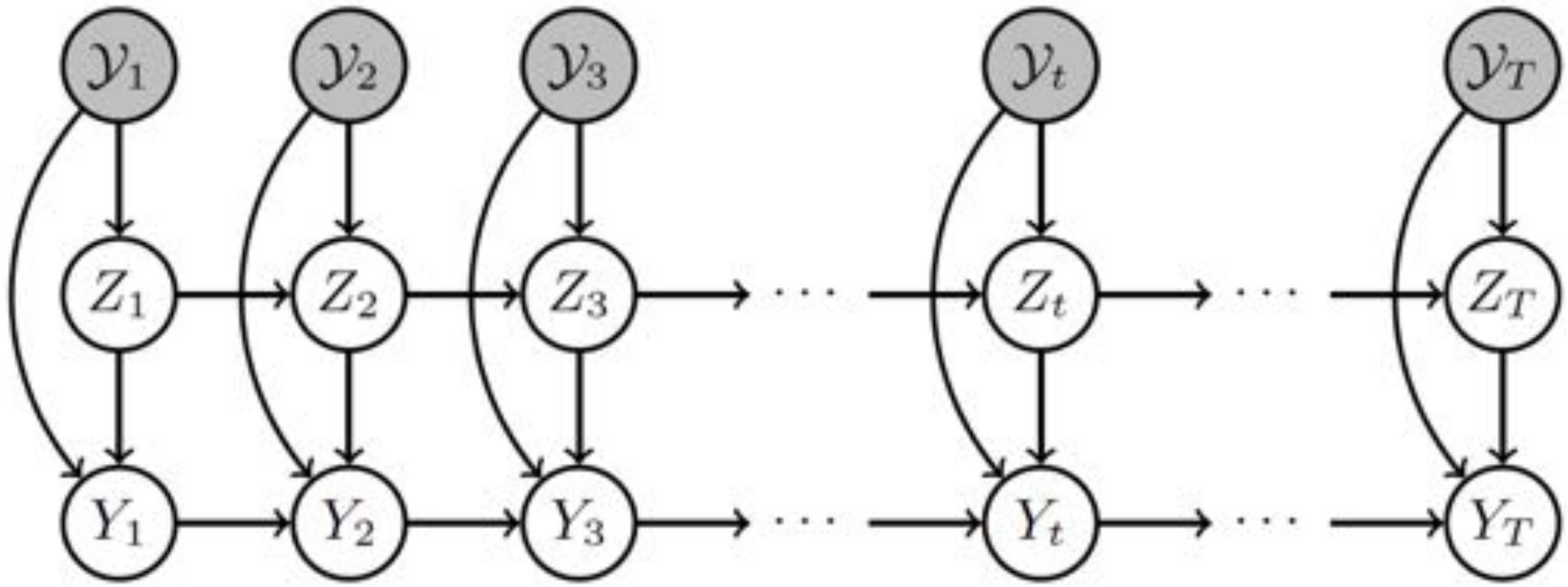
Query-focused video summarization

Decision to include a frame/short in summary

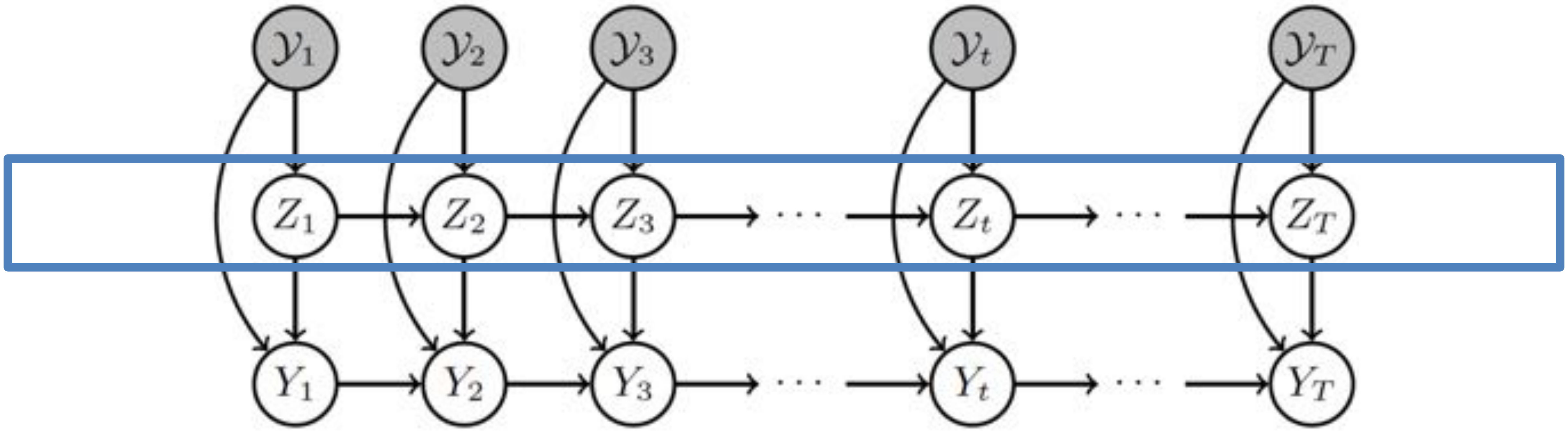
- Relevance to query (*be responsive to user input*)
- Importance in the context (*maintain story flow*)
- Collective diversity

Two levels of summarization granularity.

Sequential and hierarchical DPP (SH-DPP)



Sequential and hierarchical DPP (SH-DPP)



***Z*-layer summarizes query-relevant video shots/frames.**

$\boldsymbol{Z}$-layer: responsive to user query q

$\cong$ SeqDPP: Markov process with DPP

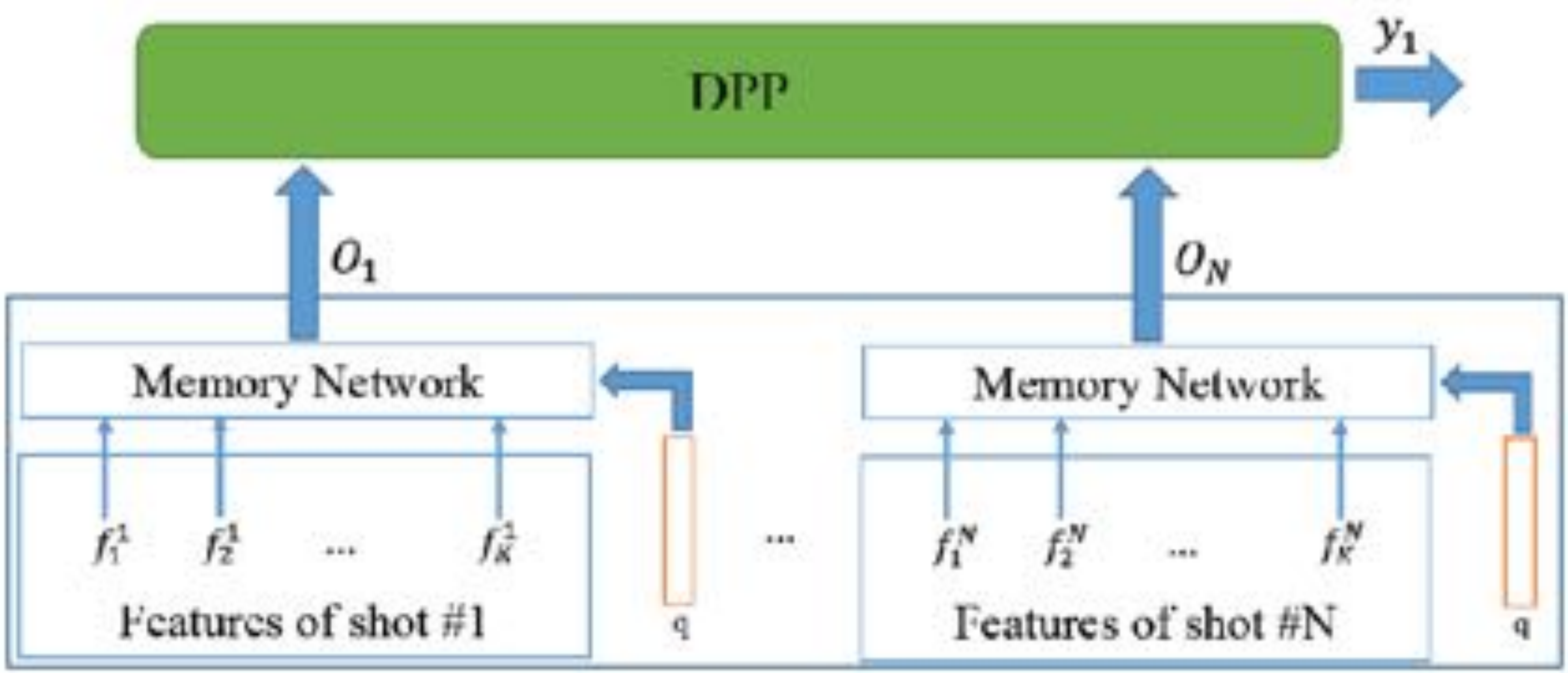
Summarizes shots/frames relevant to query

The DPP kernel is thus query-dependent

$$\boldsymbol{\Omega}_{ij} = [\boldsymbol{f}_i(q)]^T W^T W [\boldsymbol{f}_j(q)]$$

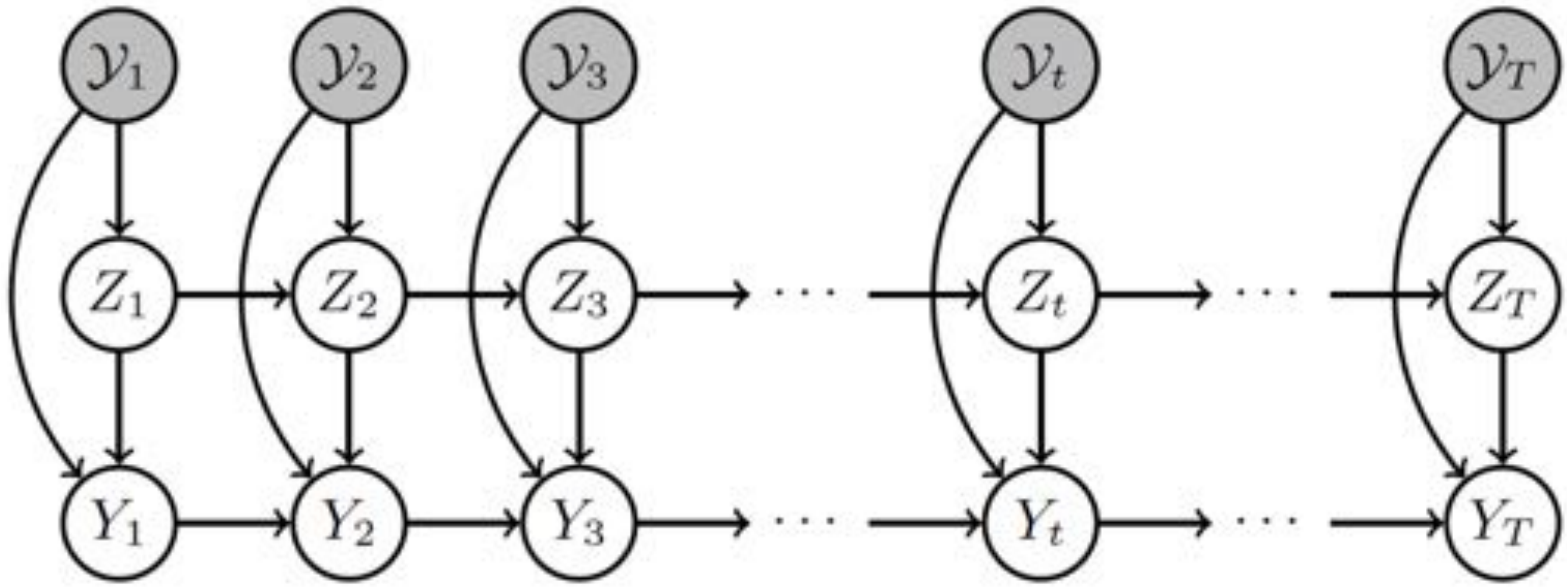
$\boxed{\boldsymbol{Z}\text{-layer summarizes query-relevant video shots/frames.}}$

$\boldsymbol{Z}$-layer: responsive to user query q

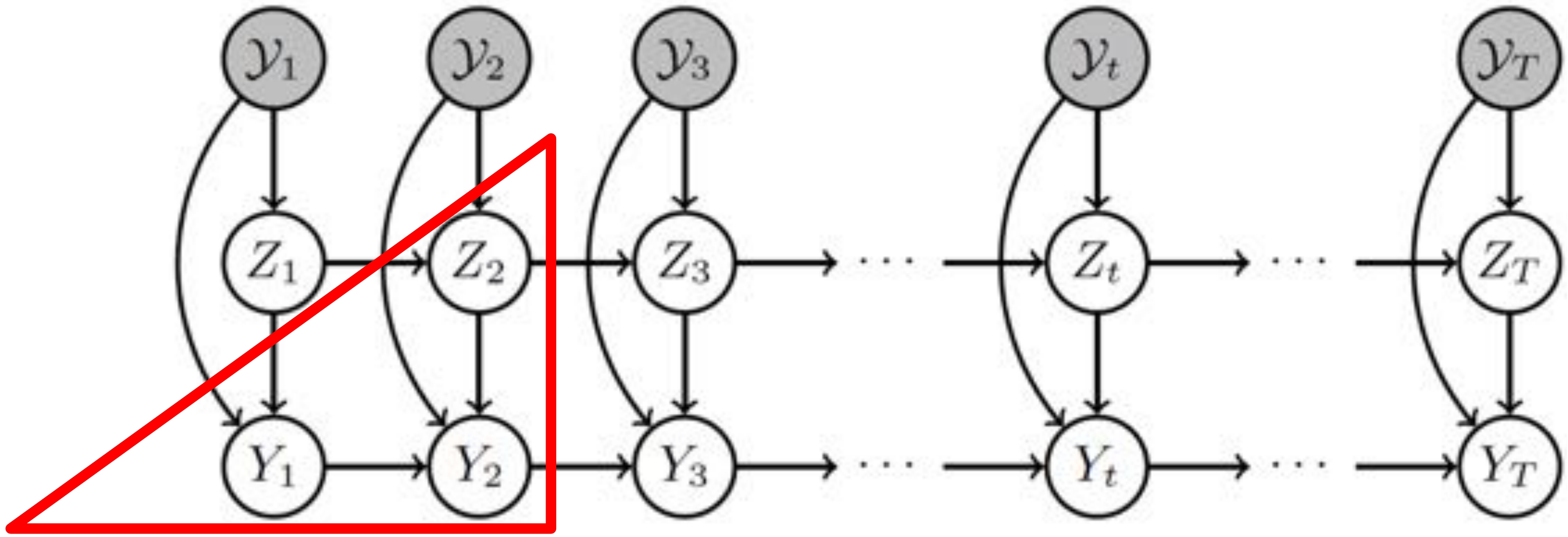


[CVPR'17]

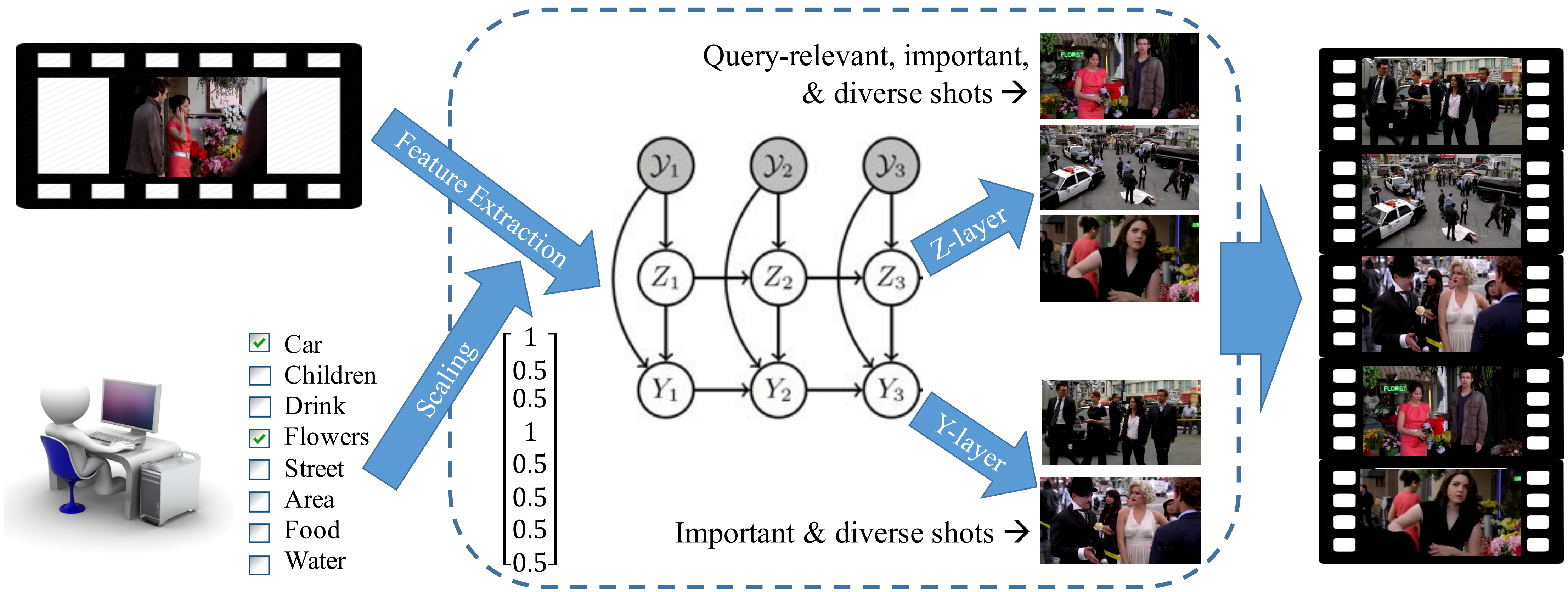
Y-layer: summ. remaining video (*maintain story flow*)



$\boldsymbol{Y}$-layer: summ. remaining video (*maintain story flow*)



Query-focused video summarization



(a) Input: Video & Query **(b) Algorithm**: Sequential & Hierarchical Determinantal Point Process (SH-DPP) **(c) Output**: Summary

Experimental results

Query: CAR+PHONE

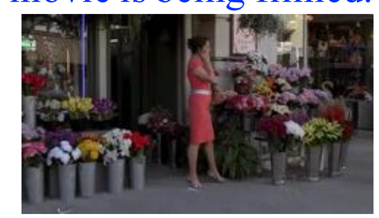
Relevant to query

Cho and Lisbon examine Hanson's CAR

Lisbon and Rigsby speak on the PHONE.

...

Felicia Scott speaks to Sydney on the PHONE, while the movie is being filmed.



Experimental results

Query: CAR+PHONE

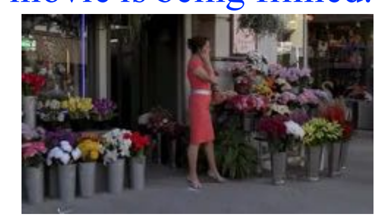
Relevant to query

Cho and Lisbon examine Hanson's CAR

Lisbon and Rigsby speak on the PHONE.

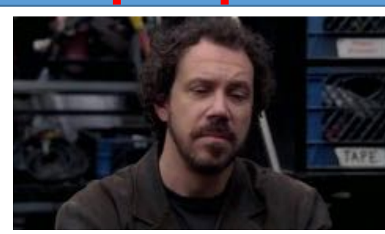
…

Felicia Scott speaks to Sydney on the PHONE, while the movie is being filmed.



Jane finishes his conversation with the policeman.

…



Mitch Cavenaugh enters the RV, and explains the drugs are his

…

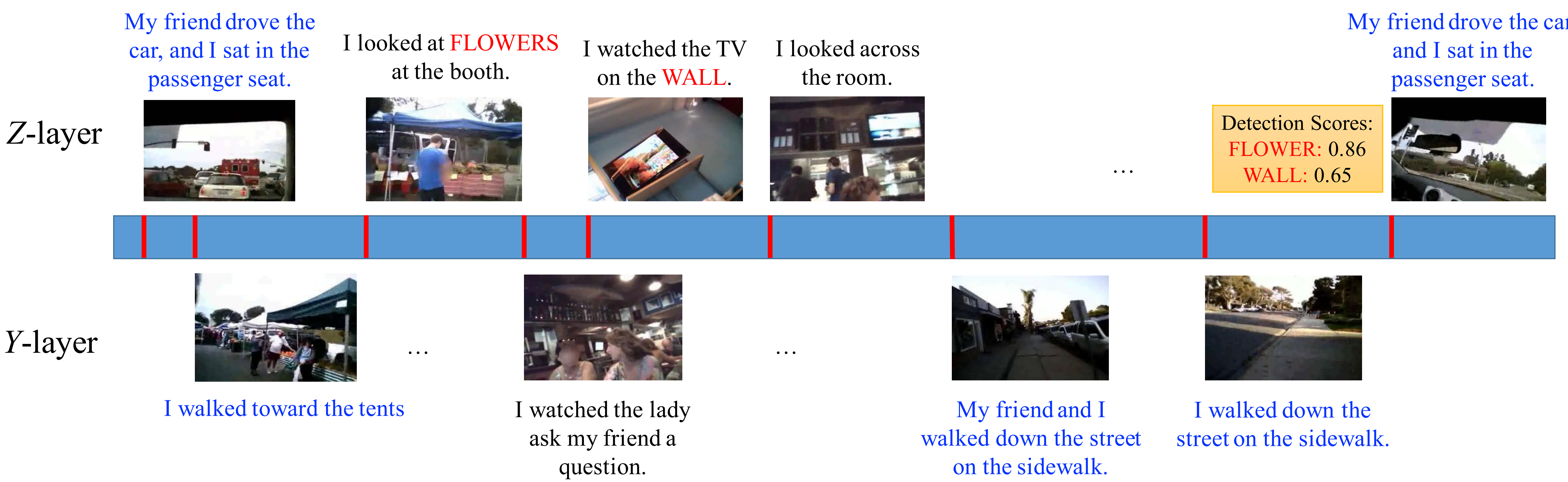
Jane speaks to Felicia Scott about how well she is acting.

Important in context

(maintain story flow)

Experimental results

Query: FLOWER+WALL



Ground-truth Summary

My friend drove the car, and I sat in the passenger seat. I got out of the car. I walked toward the tents. I looked at the fruit at the booth. My friend and I walked through the market. My friend and I looked at FLOWERS at the booth. My friend drove the car, and I sat in the passenger seat.

I sat with my friend and looked over at the TV on the WALL. I sat at the table while my friend drank. I ate pizza with my friend and we looked at the TV. I looked at the TV on the WALL and then looked back at my friend. I watched the TV on the WALL's at the restaurant.

I walked out the shop with my friend. My friend and I walked down the street on the sidewalk. I walked on the side walk.

This talk

DPP

SeqDPP

Variations

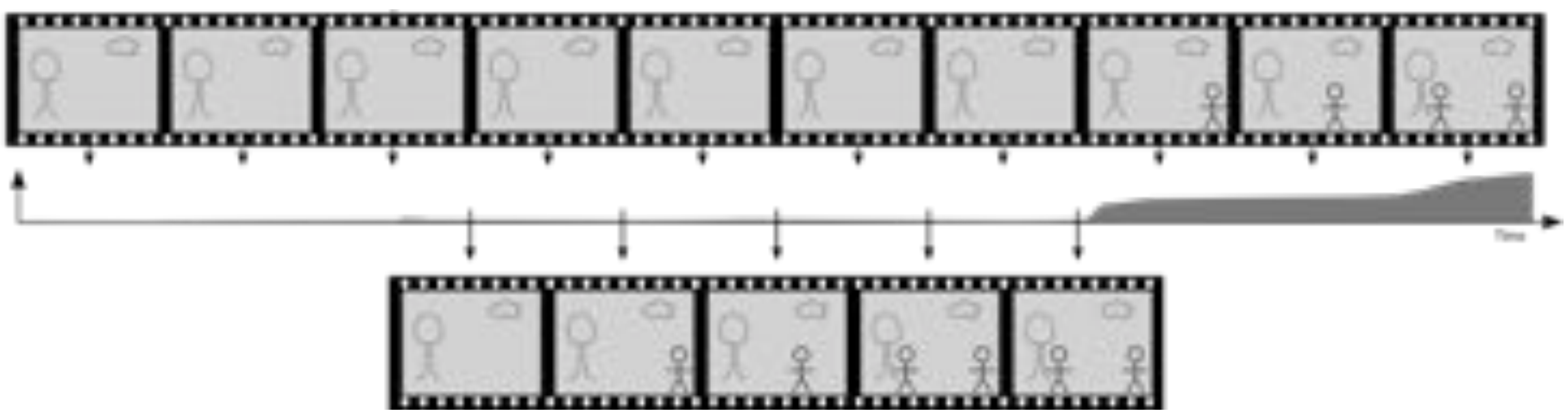
Lessons Learned

User-subjectivity

1. Personalizing video summarizers
2. An improved evaluation metric

BGong@CRCV.ucf.edu

Let user control the summary length / granularity



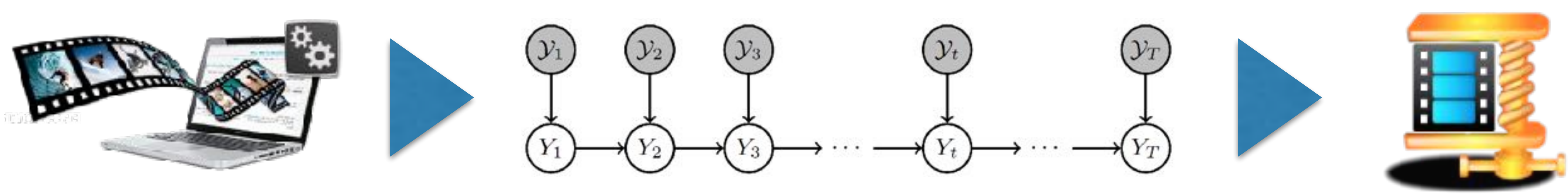


Image credit: http://www.vis.uni-stuttgart.de/index.php?id=1351

[Under review]

This talk

DPP

SeqDPP

Variations

Lessons Learned

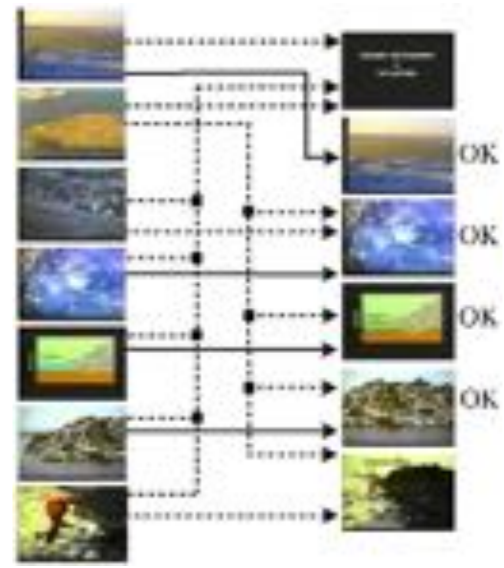
User-subjectivity

1. Personalizing video summarizers
2. An improved evaluation metric

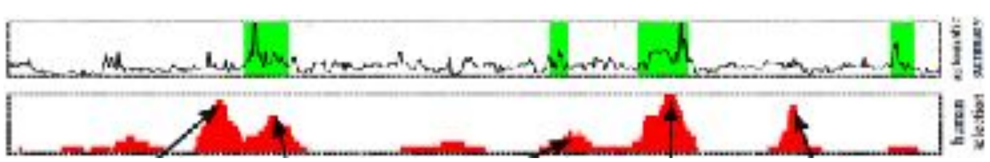
BGong@CRCV.ucf.edu

What makes a good evaluation for video summarization?

A/B test



Bipartite matching
[Avila et al. 2011]

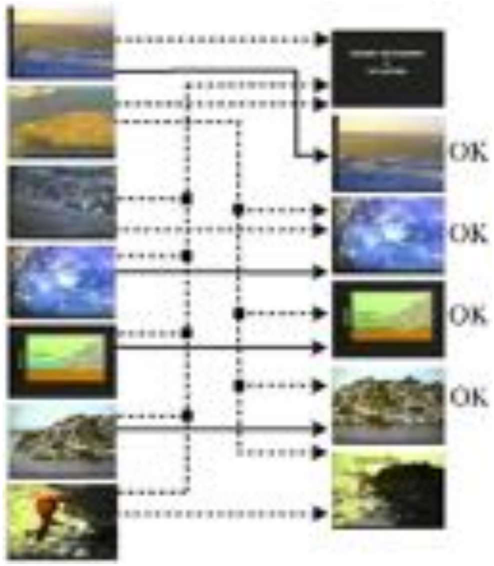


Time overlap
[Gygli et al. 2014]

Video → text
[Yeung et al. 2014]

What makes a good evaluation for video summarization?

A/B test



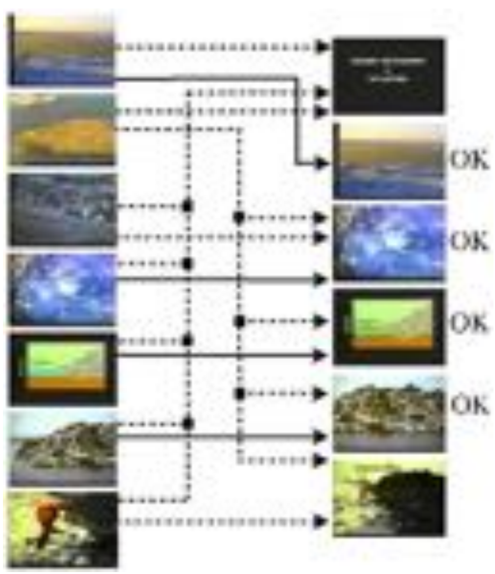
Bipartite matching
[Avila et al. 2011]

Time overlap
[Gygli et al. 2014]

Video → text
[Yeung et al. 2014]

Captions per video shot
➡ Dense concepts

What makes a good evaluation for video summarization?



Bipartite matching [Avila et al. 2011]

Bipartite matching ***of concept vectors***

[Lady, Man, Phone, Cab, Street, Building, Restaurants, …]

This talk

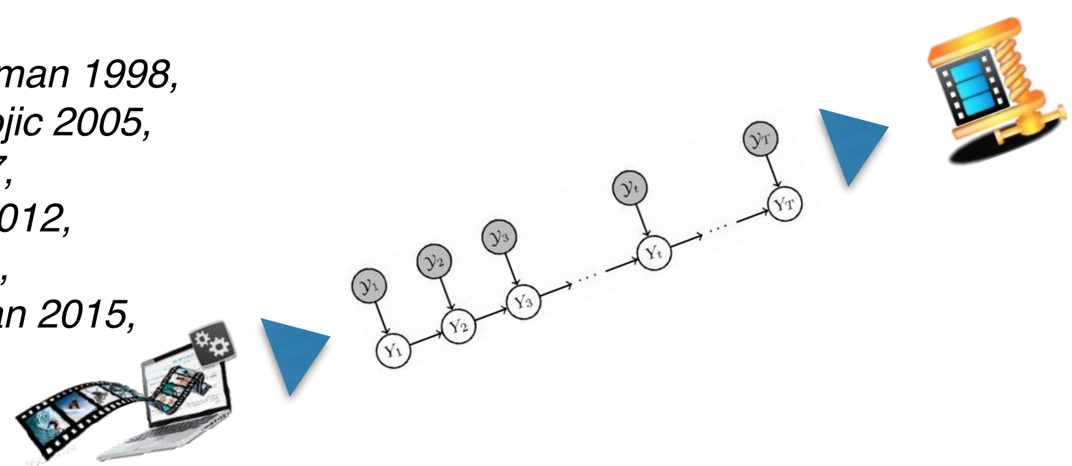
- DPP
- SeqDPP
- Variations
- Lessons Learned

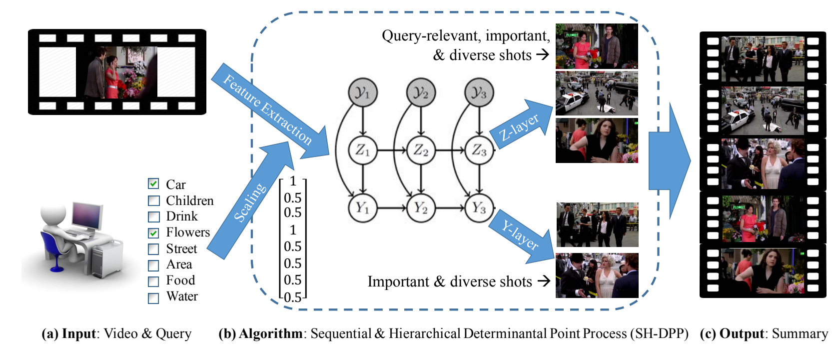
What makes a good video summarizer?

Video summarization: a ***subjective*** process

Prior: unsupervised	*SeqDPP: average user*	*SH-DPP: “the” user*

[Wolf 1996, Vasconcelos and Lippman 1998, Aner and Kender 2002, Pal and Jojic 2005, Kang et al. 2006, Pritch et al. 2007, Jiang et al. 2009, Lee and Kwon 2012, Khosla et al. 2013, Kim et al. 2014, Song et al. 2015, Lee and Grauman 2015, …]





Challenges in *Supervised* video summarization

Extremely lengthy videos

Videos of hours, days, or months ➡ minutes

Heterogenous content

Party time flies; coding is boring and slow

Transcending content

Summarizers independent of content?

Challenges (continued) in *Supervised* video summarization

User-subjectivity

Evaluation is the killer

Different users prefer distinct summaries

Granularities / lengths

Patient vs. impatient users, 15" vs. iPhone, etc.

Multiple videos of the same event

Anti-Trumpt vs. Pro-Trump

etc.

Undergoing and future work

DPPs

- Deep DPP: end(video)-to-end(summary) learning
- Recurrent DPPs: Markov dependency is limited

Video summarization

- Personalization & domain adaptation
- Video summarization for the first person

  (*Egocentric videos from life-loggers, police, sports, etc.*)

Acknowledgements

U. Southern California

Fei Sha, Wei-Lun Chao

U. Texas at Austin: Kristen Grauman

U. Central Florida

Mubarak Shah, Aidean Sharghi

MIT: Chengtao Li

U. Iowa: Tianbao Yang

Large-Margin Determinantal Point Processes

Wei-Lun Chao* Boqing Gong* Kristen Grauman Fei Sha
U. of Southern Califor
Los Angeles, CA 900

[UAI 2015]

Diverse Sequential Subset Selection for Supervised Video Summarization

Boqing Gong*
Department of Computer Science
University of Southern California
Los Angeles, CA 90089
boqinggo@usc.edu

Wei-Lun Chao*
Department of Computer Science
University of Southern California
Los Angeles, CA 90089
weilunc@usc.edu

Kristen Grauman
Department of
University of
Austin,
grauman@c

Fei Sha

[NIPS 2014]

Query-Focused Extractive Video Summarization

Aidean Sharghi, Boqing Gong, Mubarak Shah
aidea

[ECCV 2016]

Query-Focused Video Summarization: Dataset, Evaluation, and A Memory Network Based Approach

Aidean Sharghi†, Jacob Laurel‡, and Boqing Gong†

[CVPR 2017]

Code of SeqDPP:
https://github.com/pujols/Video-summarization

BGong@CRCV.ucf.edu