1st AI4TV Workshop, ACM MM 2019,
Nice, France, October 2019

A Stepwise, Label-based Approach for Improving the Adversarial Training in Unsupervised Video Summarization

E. Apostolidis[1,2], A. I. Metsai[1], E. Adamantidou[1], V. Mezaris[1], I. Patras[2]

[1] CERTH-ITI, Thermi - Thessaloniki, Greece
[2] School of EECS, Queen Mary University of London, London, UK

Outline

- Problem statement
- Related work
- Limitations of existing approaches
- Developed approach
- Experiments
- Summarization examples
- Summary and next steps
- Key references

Problem statement

Video summary: a short visual summary that encapsulates the flow of the story and the essential parts of the full-length video

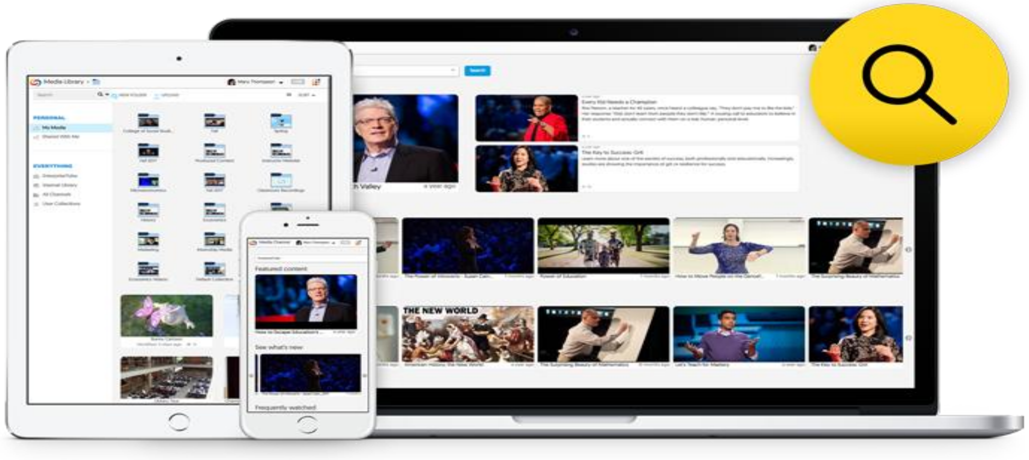
Original video

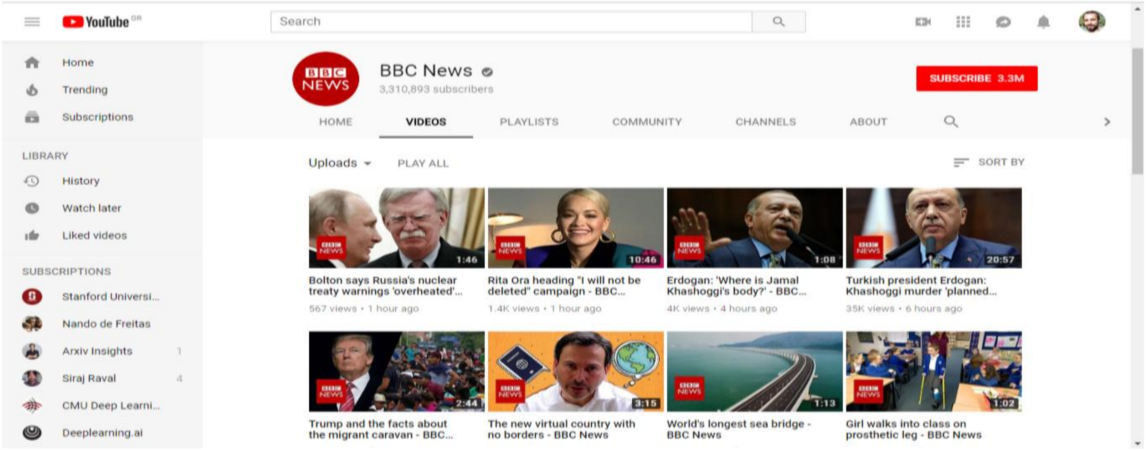
Video summary (storyboard)

Problem statement

General applications of video summarization

- Professional CMS: effective indexing, browsing, retrieval & promotion of media assets!
- Video sharing platforms: improved viewer experience, enhanced viewer engagement & increased content consumption!





- However, different distribution channels have different requirements / restrictions w.r.t. the **video content duration** (e.g. optimal Twitter videos < 45 sec.; ideal YouTube videos < 2 min.)

Related work

Deep-learning approaches

- CNN-based methods that extract and use video semantics to identify important video parts based on sequence labeling [49], self-attention networks [18], or video-level metadata [44, 46]
- RNN-based approaches that capture temporal dependency over the video frames, using
  - e.g. LSTMs for keyframe selection [64]
  - hierarchies of LSTMs to identify the video structure and select key-fragments [70, 71]
  - combinations of LSTMs with DTR units and GANs to capture long-range frame dependency [67]
  - attention-based encoder-decoders [22, 28], or memory-augmented networks [19]
- Unsupervised methods that do not rely on human-annotations, and build summaries
  - using adversarial learning objectives (GANs) to minimize the distance between training videos and a distribution of their summarizations [40], or to maximize mutual information between the created summary and the original video [60]
  - through a decision-making process that is learned via RL and reward functions [73]
  - by learning to extract key motions of appearing objects [68]
  - by learning a "video-to-summary" mapping from unpaired edited summaries available online [48]

Motivation and reasoning

Disadvantages of supervised learning

- Restricted amount of annotated data is available for supervised training of a video summarization method
- Highly-subjective nature of video summarization (relying on viewer's demands and aesthetics); there is no "ideal" or commonly accepted summary that could be used for training an algorithm

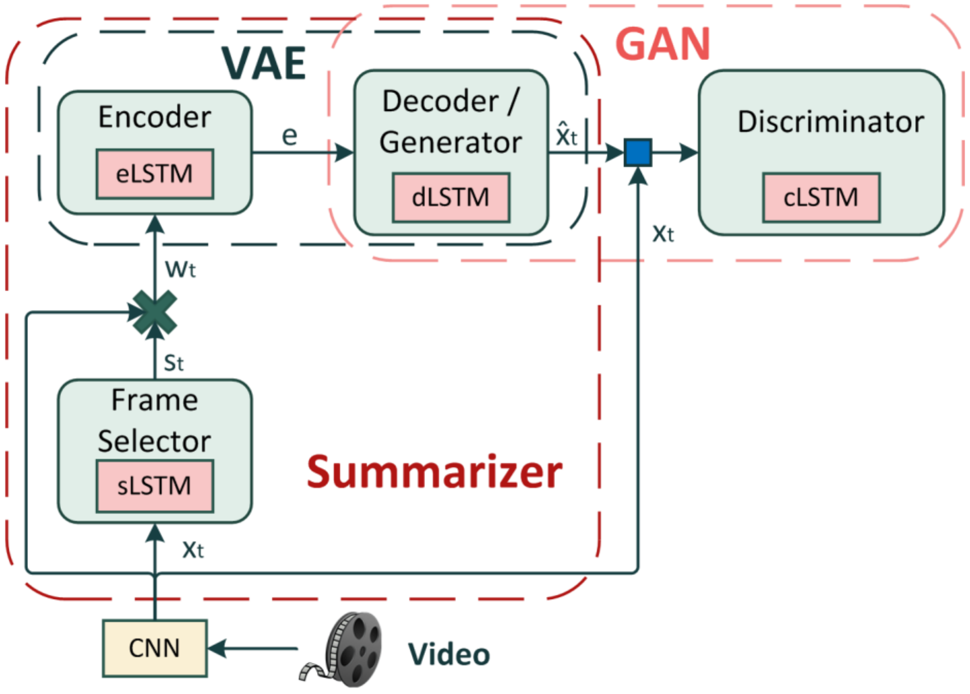
Advantages of unsupervised learning

- No need for learning data; avoid laborious and time-demanding labeling of video data
- Adaptability to different types of video; summarization is learned based on the video content

Developed approach

Building on adversarial learning

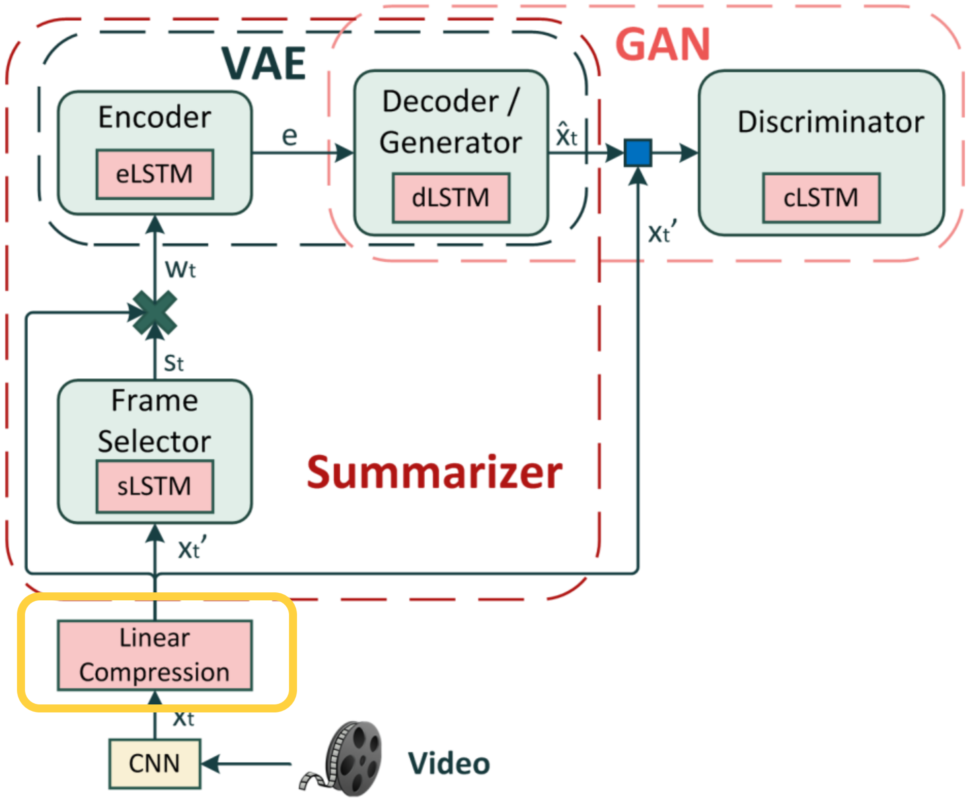
- Starting point: the SUM-GAN architecture
- Main idea: build a keyframe selection mechanism by minimizing the distance between the deep representations of the original video and a reconstructed version of it based on the selected key-frames
- Problem: how to define a good distance?
- Solution: use a trainable discriminator network!
- Goal: train the Summarizer to maximally confuse the Discriminator when distinguishing the original from the reconstructed video



Developed approach

Building on adversarial learning

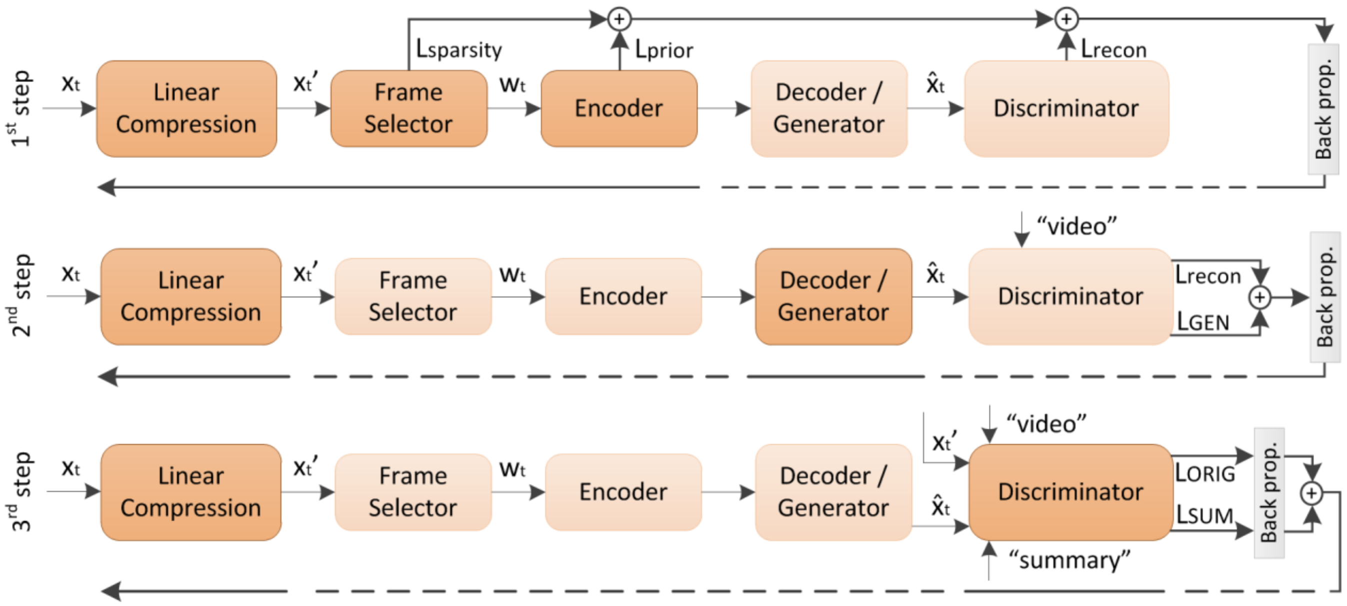
- New model: the SUM-GAN-sl architecture
  - Contains a linear compression layer that reduces:
    - The size of the CNN feature vectors from 1024 to 500
    - Thus also the number of trainable parameters



Developed approach

Building on adversarial learning

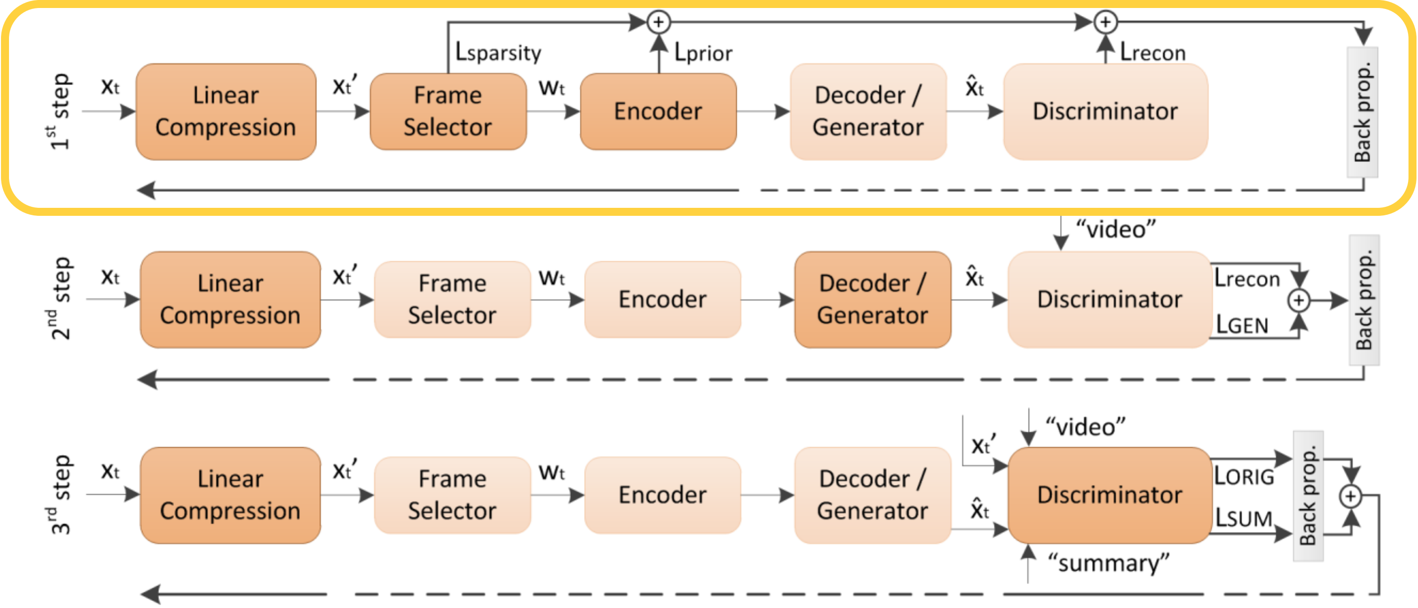
- New model: the SUM-GAN-sl architecture
  - Follows an incremental and fine-grained approach to training the model's components



Developed approach

Building on adversarial learning

- <u>New model:</u> the SUM-GAN-sl architecture
  - Follows an incremental and fine-grained approach to training the model's components



$$L_{reconst} = \left\|\varphi(\mathbf{x}') - \varphi(\hat{\mathbf{x}})\right\|^2$$

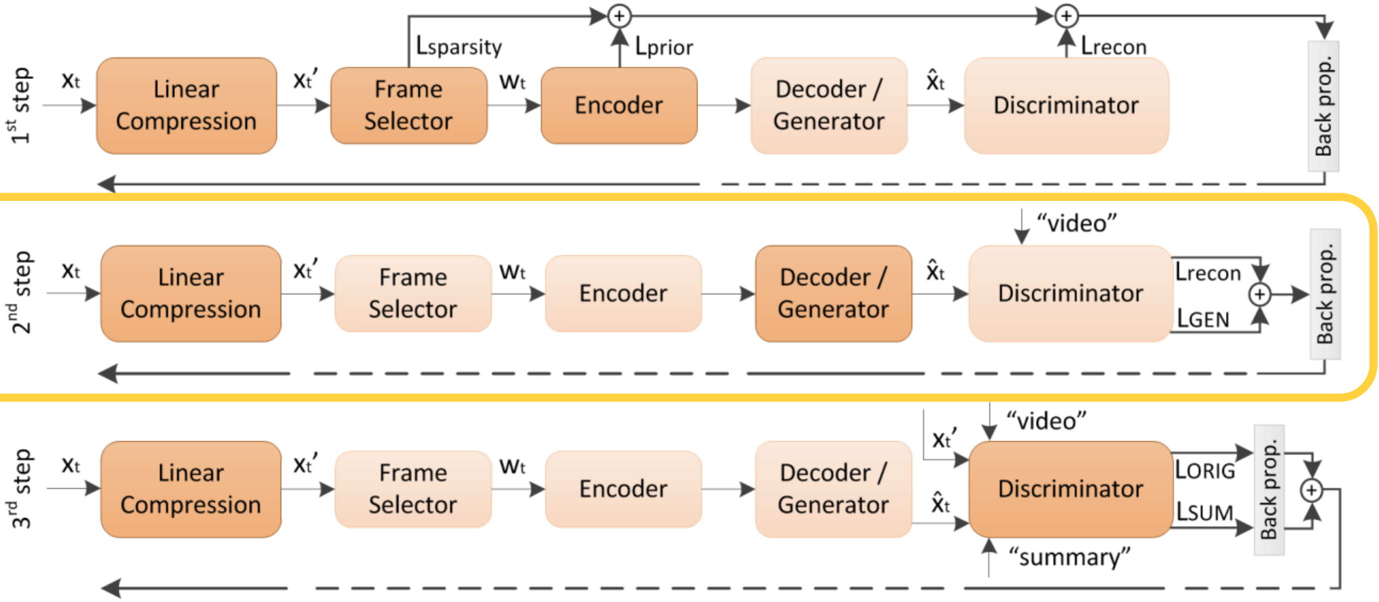
$$L_{prior} = D_{KL}(q(\mathbf{e}|\mathbf{x})||p(\mathbf{e}))$$

$$L_{sparsity} = \left\|\frac{1}{T}\sum_{t=1}^{T} s_t - \sigma\right\|^2$$

Developed approach

Building on adversarial learning

- New model: the SUM-GAN-sl architecture
  - Follows an incremental and fine-grained approach to training the model's components



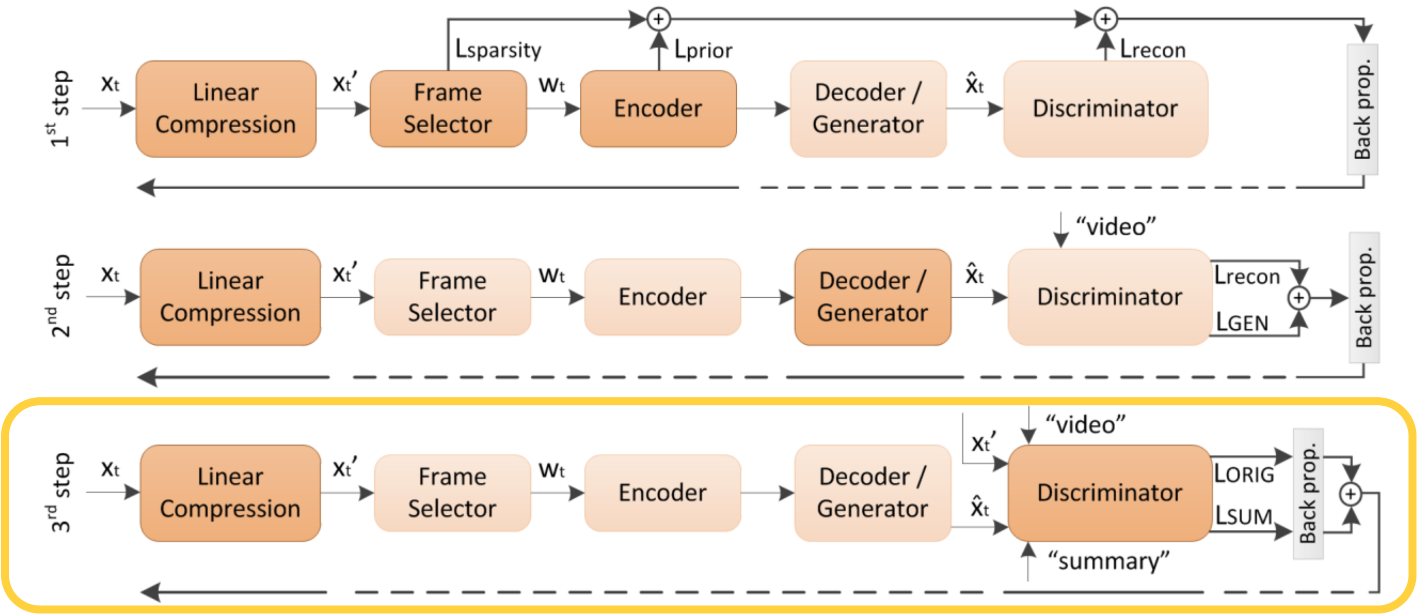
$$L_{reconst} = \|\varphi(\mathbf{x}') - \varphi(\hat{\mathbf{x}})\|^2$$

$$L_{GEN} = (1 - p(\hat{\mathbf{x}}))^2$$

Developed approach

Building on adversarial learning

- New model: the SUM-GAN-sl architecture
  - Follows an incremental and fine-grained approach to training the model's components



$$L_{ORIG} = (1 - p(\mathbf{x}'))^2$$

$$L_{SUM} = (p(\hat{\mathbf{x}}))^2$$

Experiments

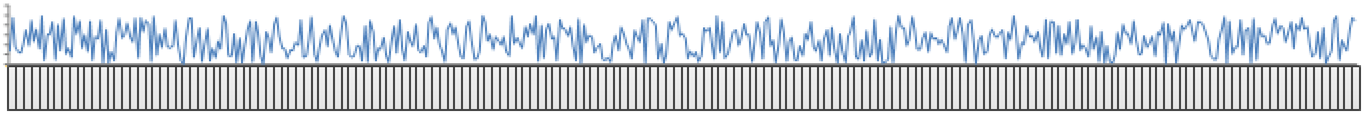
Datasets

- SumMe (https://gyglim.github.io/me/vsum/index.html#benchmark)
  - 25 videos capturing multiple events (e.g. cooking and sports)
  - video length: 1.6 to 6.5 min
  - annotation: fragment-based video summaries

Ground-truth summary in SumMe: video key-fragments

- TVSum (https://github.com/yalesong/tvsum)
  - 50 videos from 10 categories of TRECVid MED task
  - video length: 1 to 5 min
  - annotation: frame-level importance scores

Ground-truth summary in TVSum: frame-level importance scores



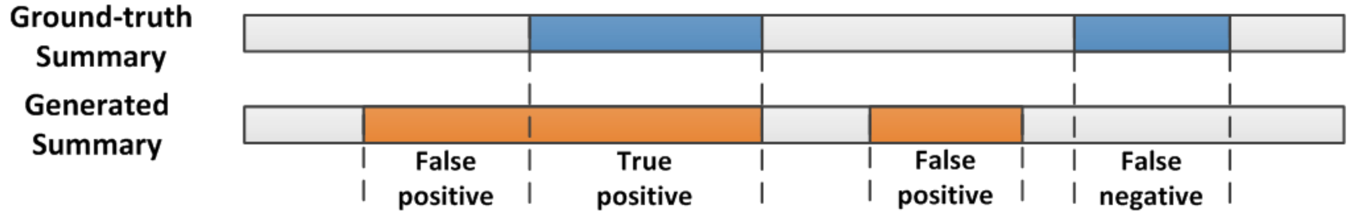
Experiments

Evaluation protocol

- The generated summary should not exceed 15% of the video length
- Similarity between automatically generated (A) and ground-truth (G) summary is expressed by the F-Score (%), with (P)recision and (R)ecall measuring the temporal overlap (∩) (|| || means duration)

$$F = 2 \times \frac{P \times R}{P + R} \times 100, \quad \text{with} \quad P = \frac{A \cap G}{||A||} \quad \text{and} \quad R = \frac{A \cap G}{||G||}$$

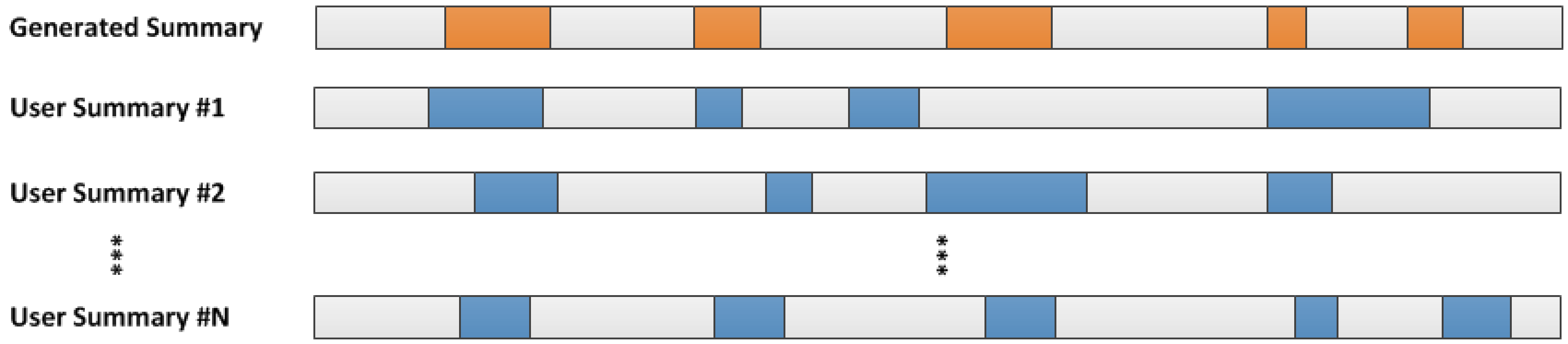
- Typical metrics for computing Precision and Recall at the frame-level



Experiments

Evaluation protocol

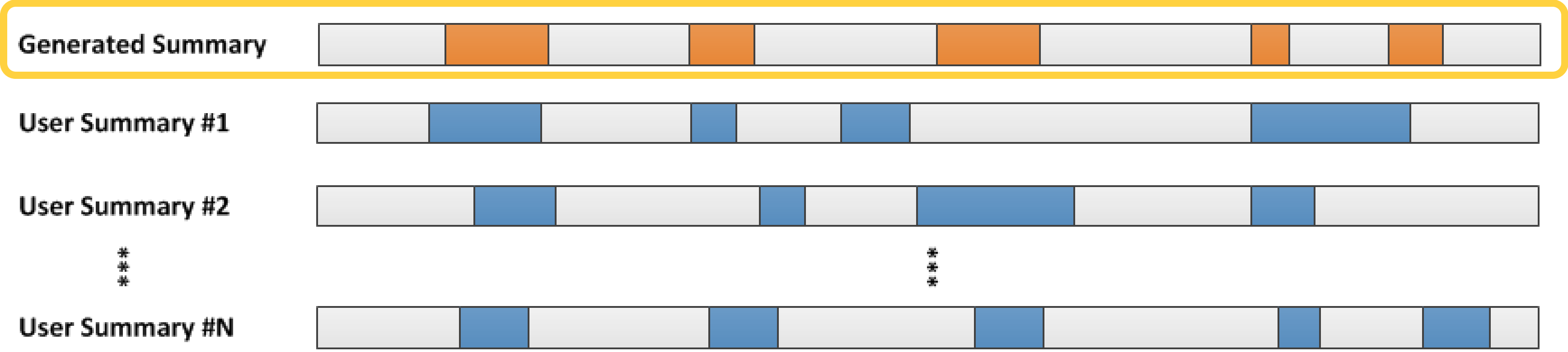
- Slight but important distinction w.r.t. what is eventually used as ground-truth summary
- Most used approach (by [16, 18, 19, 29, 48, 49, 64, 61, 70, 71, 73, 74])



Experiments

Evaluation protocol

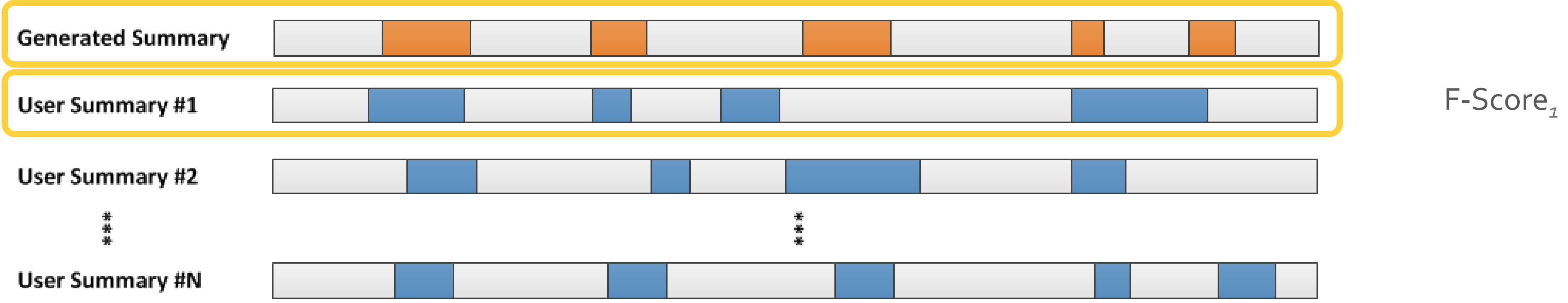
- Slight but important distinction w.r.t. what is eventually used as ground-truth summary
- Most used approach (by [16, 18, 19, 29, 48, 49, 64, 61, 70, 71, 73, 74])



Experiments

Evaluation protocol

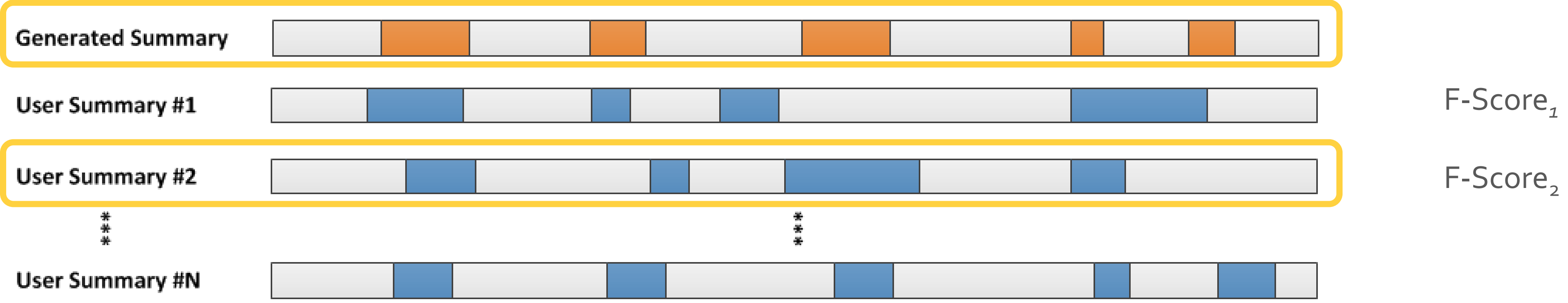
- Slight but important distinction w.r.t. what is eventually used as ground-truth summary
- Most used approach (by [16, 18, 19, 29, 48, 49, 64, 61, 70, 71, 73, 74])



Experiments

Evaluation protocol

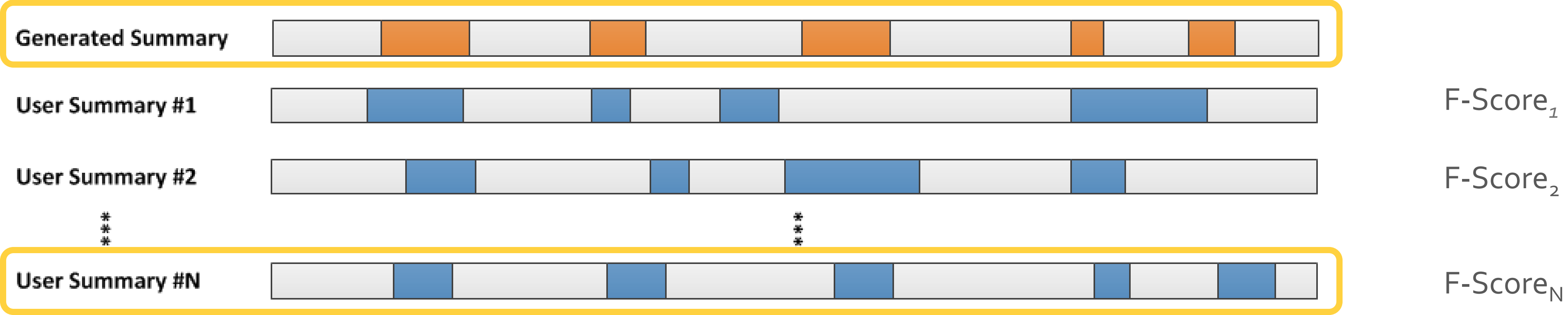
- Slight but important distinction w.r.t. what is eventually used as ground-truth summary
- Most used approach (by [16, 18, 19, 29, 48, 49, 64, 61, 70, 71, 73, 74])



Experiments

Evaluation protocol

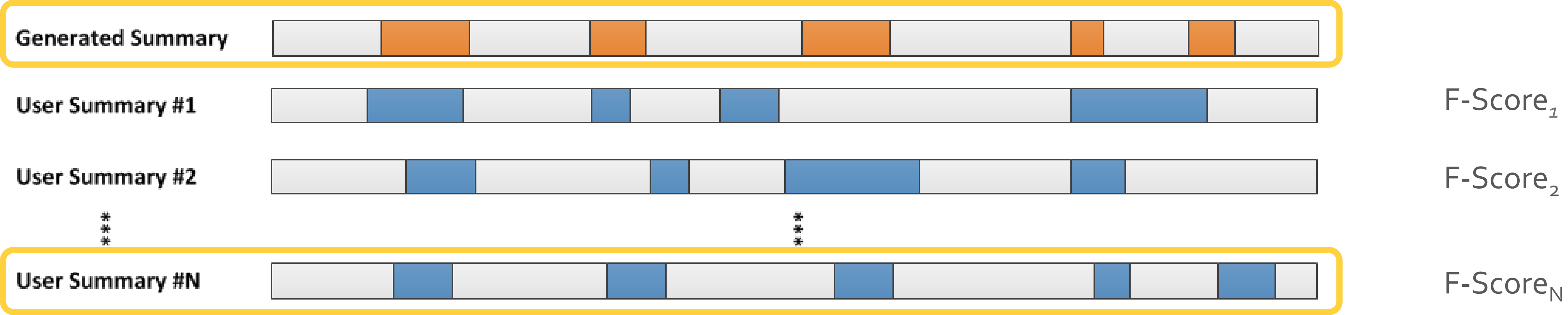
- Slight but important distinction w.r.t. what is eventually used as ground-truth summary
- Most used approach (by [16, 18, 19, 29, 48, 49, 64, 61, 70, 71, 73, 74])



Experiments

Evaluation protocol

- Slight but important distinction w.r.t. what is eventually used as ground-truth summary
- Most used approach (by [16, 18, 19, 29, 48, 49, 64, 61, 70, 71, 73, 74])



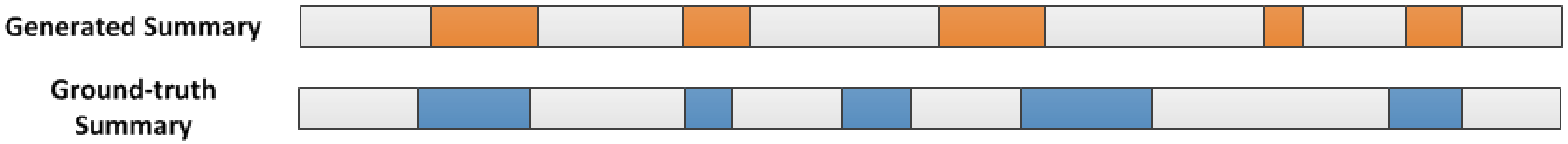
SumMe: $F\text{-}Score = \max\{F\text{-}Score\}_1^N$

TVSum: $F\text{-}Score = \dfrac{\sum_{i=1}^{N} F\text{-}Score_i}{N}$

Experiments

Evaluation protocol

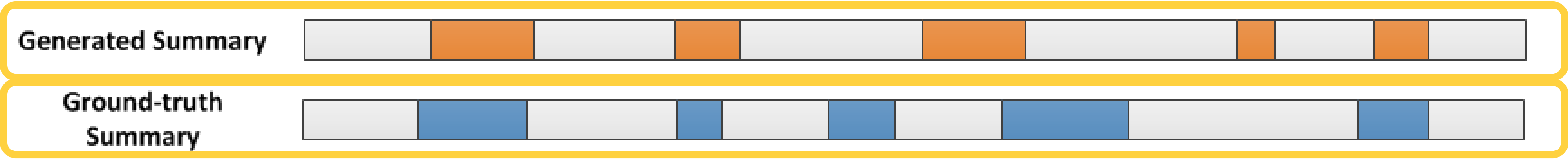
- Slight but important distinction w.r.t. what is eventually used as ground-truth summary
- Alternative approach (used in [22, 28, 40, 58, 60, 67])



Experiments

Evaluation protocol

- Slight but important distinction w.r.t. what is eventually used as ground-truth summary
- Alternative approach (used in [22, 28, 40, 58, 60, 67])



F-Score

Experiments

Preliminary study on summarization datasets

- To get insights about the efficiency of the used datasets and metrics, we examined:
  - the efficiency of a randomly generated summary (frames' importance scores were defined based on a uniform distribution of probabilities, and the experiment was performed 100 times)
  - the human performance, i.e. how well a human annotator would perform based on the preferences of the remaining annotators
  - the highest performance on TVSum according to the best user summary (with the highest overlap) for each video of the dataset (best for SumMe is 100%)

	SumMe		TVSum	
	Average	Max	Average	Max
Random	18.1	39.9	53.9	75.5
Human Summaries	31.3	55.1	53.8	77.5
Best Possible	44.7	100.0	64.7	100.0

- Results consistent with the findings of a recent CVPR paper ([41])

Experiments

Evaluation outcomes

- Step 1: Assessing the impact of regularization factor σ

	SumMe	TVSum
$\sigma = 0.05$	44.7	58.2
$\sigma = 0.1$	**47.3**	58.0
$\sigma = 0.15$	46.6	58.6
$\sigma = 0.3$	46.4	**58.8**
$\sigma = 0.5$	42.7	58.6

Findings

- Value of σ affects the models' performance and needs fine-tuning
- Too small and too large values lead to reduced efficiency, and only a specific range of values results in good performance
- Fine-tuning is dataset-dependent as the highest performance is achieved for different values of σ in each dataset

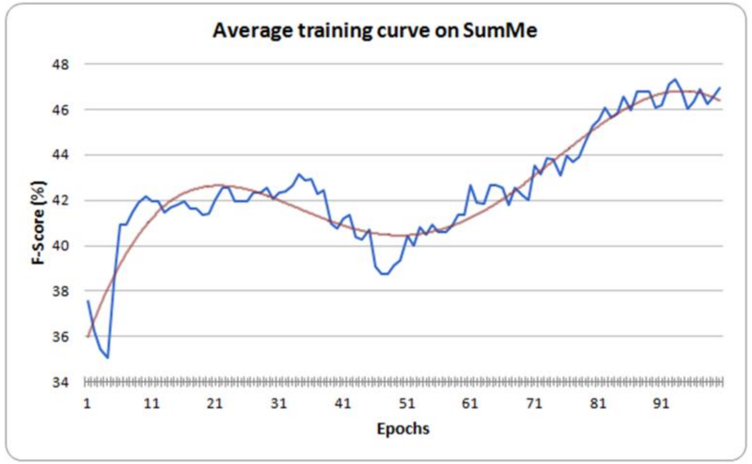
Experiments

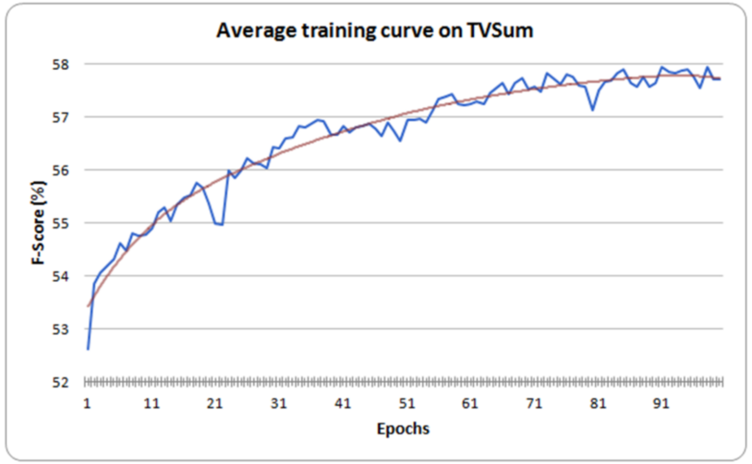
Evaluation outcomes

- Step 2: Selecting the best configuration

	SumMe	TVSum
$\sigma = 0.05$	44.7	58.2
$\sigma = 0.1$	**47.3**	58.0
$\sigma = 0.15$	46.6	58.6
$\sigma = 0.3$	46.4	**58.8**
$\sigma = 0.5$	42.7	58.6

Learning curves





Experiments

Evaluation outcomes

- Step 3: Comparison with SoA unsupervised approaches based on multiple user summaries

	SumMe	TVSum
Random summary	39.9 (–)	53.9 (–)
Tessellation [30]	41.4 (–)	**64.1** (+)
DR-DSN [74]	41.4 (–)	57.6 (–)
Online Motion-AE [69]	37.7 (–)	51.5 (–)
UnpairedVSN [49]	47.5 (–)	55.6 (–)
SUM-GAN-sl	47.3	58.0

Findings

- A few SoA methods are comparable (or even worse) with a random summary generator
- Best method on SumMe (UnpairedVSN) performs slightly better than our method, and it is less competitive on TVSum
- Best method on TVSum shows random-level performance on SumMe (seems to be dataset-tailored)
- SUM-GAN-sl performs consistently well in both datasets and is the most competitive one

Experiments

Evaluation outcomes

- Step 4: Comparison with SoA **supervised** approaches based on multiple user summaries

	SumMe	TVSum		SumMe	TVSum
Random summary	39.9 (–)	53.9 (–)	MAVS [19]	40.3 (–)	**66.8** (+)
vsLSTM [64]	37.6 (–)	54.2 (–)	SUM-FCN [49]	47.5 (–)	56.8 (–)
dppLSTM [64]	38.6 (–)	54.7 (–)	SUM-DeepLab [49]	48.8 (=)	58.4 (=)
H-RNN [70]	41.1 (–)	57.7 (–)	DR-DSNsup [73]	42.1 (–)	58.1 (–)
Tessellationsup [29]	37.2 (–)	63.4 (+)	ActionRanking [16]	40.1 (–)	56.3 (–)
HSA-RNN [71]	44.1 (–)	59.8 (+)	UnpairedVSNpsup [48]	48.0 (–)	56.1 (–)
DQSN [74]	-	58.6 (+)	VASNet [18]	**49.7** (+)	61.4 (+)
DSSE [61]	-	57.0 (–)	SUM-GAN-sl	47.3 (–)	58.0 (–)

Findings

- Best methods in TVSum (MAVS & Tessellationsup) are highly-adapted to this dataset, as they exhibit random-level performance on SumMe
- Only a few supervised methods clearly surpass the performance of a randomly-generated summary on both datasets, with VASNet being the best among them
- Their performance ranges in [44.1 - 49.7] on SumMe, and in [56.1 - 61.4] on TVSum
- The performance of SUM-GAN-sl makes our **unsupervised** method comparable with SoA **supervised** techniques for video summarization

Experiments

Evaluation outcomes

- Step 5: Comparison with SoA approaches based on single ground-truth summaries
- Impact of regularization factor σ

	σ	SumMe	TVSum
SUM-GAN	0.3	38.7	50.8
SUM-GAN-sl	0.1	38.1	61.0
	0.3	45.2	62.4
	0.5	**46.8**	**65.3**

Findings

- Model's performance is affected by the value of σ in a way similar to the one reported in [40]
- The effect of this hyper-parameter depends on the used evaluation approach (best performance when using multiple human summaries was observed for σ = 0.1)
- SUM-GAN-sl clearly outperforms the original SUM-GAN model on both datasets

Experiments

Evaluation outcomes

- Step 5: Comparison with SoA approaches based on single ground-truth summaries

	SumMe	TVSum
* SUM-GAN [40]	38.7 (–)	50.8 (–)
* SUM-GANdpp [40]	39.1 (–)	51.7 (–)
SUM-GANsup [40]	41.7 (–)	56.3 (–)
SASUM [58]	45.3 (–)	58.2 (–)
DTR-GAN [67]	44.6 (–)	59.1 (–)
A-AVS [28]	43.9 (–)	59.4 (–)
M-AVS [28]	44.4 (–)	61.0 (–)
AALVS [22]	46.2 (–)	63.6 (–)
* Cycle-SUM [60]	41.9 (–)	57.6 (–)
* SUM-GAN-sl	46.8	**65.3**

Unsupervised approaches marked with an asterisk

Findings

- Best version of SUM-GAN-sl (observed for σ = 0.5) exceeds the performance of **all** other techniques (both supervised and unsupervised ones) that follow this evaluation protocol

Summarization examples

Full video

Generated summary

Summary and next steps

- Summary
  - Conducted a study to assess and advance the effectiveness of unsupervised video summarization based on adversarial learning
  - Focused on the SUM-GAN model and suggested a new training approach to advance the learning efficiency of the adversarial module of the architecture
  - Examined the evaluation protocols and metrics, and made estimates on the possible performance on two datasets and the suitability of the used metrics
  - Comparative evaluations with SoA showed that our model is among the best unsupervised methods and comparable with supervised algorithms too
- Next steps
  - Further improve our model by exploiting the efficiency of attention networks and the training capacity of reinforcement learning approaches
  - Extend our model with mechanisms that capture the temporal structure of the video to support multi-level video summarization
  - Investigate methods for video summarization tailored to the targeted audience and distribution channel (apply rules about the content of the summary; integrate a human-critic in the pipeline)

Key references

1. J. Almeida, et al. Vison: Video summarization for online applications. Pat. Recogn. Lett., 33(4):397-409, Mar. 2012
2. R. Anirudh, et al. Diversity promoting online sampling for streaming video summarization. ICIP 2016, pp 3329-3333, Sept 2016
3. H. Bahuleyan, et al. Variational attention for sequence-to-sequence models. 27th Int. Conf. on Computational Linguistics, pp 1672-1682, USA, Aug. 2018
4. J. Calic, et al. Efficient layout of comic-like video summaries. IEEE Transactions on Circuits and Systems for Video Technology, 17(7):931-936, July 2007
5. P. P. K. Chan, et al. A novel method to reduce redundancy in adaptive threshold clustering key frame extraction systems. ICMLC 2011, vol 4, pp 1637-1642
6. J. Cho. PyTorch Implementation of SUM-GAN from "Unsupervised Video Summarization with Adversarial LSTM Networks", 2017
7. K. Cho, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. EMNLP 2014 , pp. 1724-1734, Qatar, Oct. 2014
8. W. Chu, et al. Video co-summarization: Video summarization by visual co-occurrence. CVPR 2015, pp 3584-3592, June 2015
9. Y. Cong, et al. Sparse reconstruction cost for abnormal event detection. CVPR 2011, pp 3449-3456, June 2011
10. K. Darabi, et al. Personalized video summarization based on group scoring. ChinaSIP 2014, pp 310-314, July 2014
11. K. Davila, et al. Whiteboard video summarization via spatio-temporal conflict minimization. ICDAR 2017, vol 1, pp 355-362, Nov. 2017
12. S. E. F. de Avila, et al. Vsumm: A mechanism designed to produce static video summaries and a novel evaluation method. Pat. Rec. Lett., 32(1):56-68, Jan. 2011
13. M. S. Drew, et al. Clustering of compressed illumination-invariant chromaticity signatures for efficient video summarization. Im. Vis. Comp., 21:705-716, 2003
14. N. Ejaz, et al. Feature aggregation based visual attention model for video summarization. Computers & Electrical Engineering, 40(3):993-1005, 2014
15. N. Ejaz, et al. Adaptive key frame extraction for video summarization using an aggregation mechanism. J. of Vis. Comm. & Image Repr., 23(7):1031-1040, 2012
16. M. Elfeki, et al. Video summarization via actionness ranking. WACV 2019, Waikoloa Village, HI, USA, pp 754-763, Jan 2019
17. E. Elhamifar, et al. See all by looking at a few: Sparse modeling for finding representative objects. CVPR 2012, pp 1600-1607, June 2012
18. J. Fajtl, et al. Summarizing videos with attention. ACCV 2018 Workshops, pp 39-54, Cham, 2019. Springer International Publishing
19. L. Feng, et al. Extractive video summarizer with memory augmented neural networks. ACM MM 2018, pp 976-983, New York, NY, USA, 2018. ACM
20. S. Feng, et al. Online content-aware video condensation. CVPR 2012, pp 2082-2087, June 2012
21. M. Fleischman, et al. Temporal feature induction for baseball highlight classification. ACM MM 2007, pp 333-336, New York, NY, USA, 2007. ACM
22. T. Fu, et al. Attentive and adversarial learning for video summarization. WACV 2019, Waikoloa Village, HI, USA, pp 1579-1587, January 2019
23. M. Furini, et al. Stimo: Still and moving video storyboard for the web scenario. Multimedia Tools Appl., 46(1):47-69, Jan. 2010

Key references

24. B. Gong, et al. Diverse sequential subset selection for supervised video summarization. NIPS 2014, pp 2069-2077, Cambridge, MA, USA, 2014. MIT Press
25. M. Gygli, et al. Video summarization by learning sub-modular mixtures of objectives. CVPR 2015, pp 3090-3098, June 2015
26. M. Gygli, et al. Creating summaries from user videos. ECCV 2014, pp 505-520, Cham, 2014. Springer International Publishing
27. S. Hochreiter et al. Long short-term memory. Neural computation, 9(8):1735-1780, 1997
28. Z. Ji, et al. Video summarization with attention-based encoder-decoder networks. IEEE Trans. on Circuits and Systems for Video Technology, 2019
29. D. Kaufman, G. Levi, T. Hassner, and L. Wolf. Temporal tessellation: A unified approach for video analysis. ICCV 2017, pp 94-104, Oct 2017
30. A. Khosla, et al. Large-scale video summarization using web-image priors. CVPR 2013, pp 2698-2705, June 2013
31. C. Kim, et al. Object-based video abstraction for video surveillance systems. IEEE TCSVT, 12(12):1128-1138, Dec 2002
32. J.-L. Lai, et al. Key frame extraction based on visual attention model. Journal of Visual Communication and Image Representation, 23(1):114-125, 2012
33. S. Lee, et al. A Memory Network Approach for Story-based Temporal Summarization of 360 Videos. CVPR 2018
34. Y. J. Lee, et al. Discovering important people and objects for egocentric video summarization. CVPR 2012, pp 1346-1353, June 2012
35. X. Li, et al. A general framework for edited video and raw video summarization. IEEE Trans. on Image Processing, 26(8):3652-3664, Aug 2017
36. Y. Li, et al. An overview of video abstraction techniques. Hewlett Packard, Technical Reports, 01 2001
37. S. Lu, et al. A bag-of-importance model with locality-constrained coding based feature learning for video summarization. IEEE TMM, 16(6):1497-1509, Oct 2014
38. Z. Lu, et al. Story-driven summarization for egocentric video. CVPR 2013, pp 2714-2721, June 2013
39. Y.-F. Ma, et al. A generic framework of user attention model and its application in video summarization. IEEE Trans. on Multimedia, 7(5):907{919, Oct 2005
40. B. Mahasseni, et al. Unsupervised video summarization with adversarial LSTM networks. CVPR 2017, pp 2982-2991
41. M. Otani, et al. Rethinking the evaluation of video summaries. CVPR 2019
42. M. Otani, et al. Video summarization using deep semantic features. ACCV 2016
43. M. Otani, et al. Video summarization using textual descriptions for authoring video blogs. Multimedia Tools and Applications, 76(9):12097-12115, May 2017
44. R. Panda, et al. Weakly supervised summarization of web videos. ICCV 2017, pp 3677-3686, Oct 2017
45. J. Peng, et al. Keyframe-based video summary using visual attention clues. IEEE MultiMedia, 17(2):64-73, April 2010
46. D. Potapov, et al. Category-specific video summarization. ECCV 2017, pages 540-555, Zurich, Switzerland, Sept. 2014. Springer

Key references

47. A. Radford, et al. Unsupervised representation learning with deep convolutional generative adversarial networks. ICLR 2016
48. M. Rochan, et al. Video summarization by learning from unpaired data. CVPR 2019
49. M. Rochan, et al. Video summarization using fully convolutional sequence networks. ECCV 2018, pp 358-374, Cham, 2018. Springer International Publishing
50. A. Sharghi, et al. Query-focused extractive video summarization. ECCV 2016, pp 3-19, Cham, 2016. Springer International Publishing
51. Y. Song, et al. Tvsum: Summarizing web videos using titles. CVPR 2015, pp 5179-5187, June 2015
52. M. Sun, et al. Ranking highlights in personal videos by analyzing edited videos. IEEE Trans. on Image Processing, 25(11):5145-5157, Nov 2016
53. C. Szegedy, et al. Going deeper with convolutions. CVPR 2015, pp 1-9, June 2015
54. B. T. Truong, et al. Video abstraction: A systematic review and classification. ACM Trans. Multimedia Comput. Commun. Appl., 3(1), Feb. 2007
55. A. B. Vasudevan, et al. Query-adaptive video summarization via quality-aware relevance estimation. ACM MM 2017, pp 582-590, NY, USA, 2017. ACM
56. T. Wang, et al. Video collage: A novel presentation of video sequence. ICME 2007, pp 1479-1482, July 2007
57. X. Wang, et al. User-specific video summarization. ICMSP 2011, vol 1, pp 213-219, May 2011
58. H. Wei, et al. Video summarization via semantic attended networks. AAAI 2018
59. T. Yao, et al. Highlight detection with pairwise deep ranking for first-person video summarization. CVPR 2016, pp 982-990, June 2016
60. L. Yuan, et al. Cycle-sum: Cycle-consistent adversarial LSTM networks for unsupervised video summarization. AAAI 2019
61. Y. Yuan, et al. Video summarization by learning deep side semantic embedding. IEEE TCSVT, 29(1):226-237, Jan 2019
62. H. Zhang, et al. An integrated system for content-based video retrieval and browsing. Pattern Recognition, 30:643-658, 1997
63. K. Zhang, et al. Summary transfer: Exemplar-based subset selection for video summarization. CVPR 2016, pp 1059-1067, 2016
64. K. Zhang, et al. Video summarization with long short-term memory. ECCV 2016, pp 766-782, Cham, 2016. Springer International Publishing
65. S. Zhang, et al. Context-aware surveillance video summarization. IEEE Trans. on Image Processing, 25(11):5469-5478, Nov 2016
66. X.-D. Zhang, et al. Dynamic selection and effective compression of key frames for video abstraction. Pattern Recogn. Lett., 24(9-10):1523-1532, June 2003
67. Y. Zhang, et al. DTR-GAN: Dilated temporal relational adversarial network for video summarization. ACM TUR-C 2019: 89:1-89:6
68. Y. Zhang, et al. Unsupervised object-level video summarization with online motion auto-encoder. Pattern Recognition Letters, 2018
69. Y. Zhang, et al. Motion-state-adaptive video summarization via spatiotemporal analysis. IEEE TCSVT, 27(6):1340-1352, June 2017

Key references

70. B. Zhao, et al. Hierarchical recurrent neural network for video summarization. ACM MM 2017, pp 863-871, New York, NY, USA, 2017. ACM
71. B. Zhao, et al. HSA-RNN: Hierarchical structure-adaptive RNN for video summarization. CVPR 2018
72. B. Zhao, et al. Quasi real-time summarization for consumer videos. CVPR 2014, pp 2513-2520, June 2014
73. K. Zhou, et al. Deep reinforcement learning for unsupervised video summarization with diversity-representativeness reward. AAAI 2018
74. K. Zhou, et al. Video summarization by classification with deep reinforcement learning. BMVC 2018
75. G. Zhu, et al. Trajectory based event tactics analysis in broadcast sports video. ACM MM 2007, pp 58-67, New York, NY, USA, 2007. ACM

Thank you for your attention!
Questions?

Vasileios Mezaris, bmezaris@iti.gr

Code and documentation publicly available at:
https://github.com/e-apostolidis/SUM-GAN-sl

This work was supported by the EUs Horizon 2020 research and innovation programme under grant agreement H2020-780656 ReTV