IERG 6130 Project
Reinforcement Learning for Video Summarization

Anyi Rao
Department of Information Engineering
The Chinese University of Hong Kong
Shatin, Hong Kong
ra018@ie.cuhk.edu.hk

Xudong Xu
Department of Information Engineering
The Chinese University of Hong Kong
Shatin, Hong Kong
xx018@ie.cuhk.edu.hk

Abstract

Despite the remarkable progress of computer vision, video summarization, which aims to select a subset of diverse and representative frames from a video, remains a challenging task due to its subjectiveness. Conventional approaches try to tackle this problem with supervised learning, which requires expensive human-annotated labels. Recently, unsupervised learning pipeline with pre-defined metrics is also explored, but the summarization results suffer from a lack of diversity. In this paper, we propose a novel end-to-end reinforcement learning framework which leverages the inherent structures of videos. It takes advantage of intra-video and cross-video structures to learn both diversity and representativeness, which are ensured by the proposed diversity and representativeness reward respectively. Reinforcing on the evaluation score, our model can even achieve better performance, which surpasses the supervised methods by a large margin. Since our model is *fully unsupervised*, it can be applied to large-scale unlabeled videos.

1 Introduction

Nowadays as the number of videos made available online is exploding, it becomes more and more difficult to query and search the video we want. Thus, video summarization techniques, which allow users to quickly digest the gist of the content by going through just a few short distilled clips, are of a great value.

A good video summary should be both *representative* and *diverse*. Being representative means the summary is able to cover the key contents of a long video using short clips, while being diverse means the summary should contain different contents to reflect different aspects or stages of the story. It remains challenging to balance the representativeness and the diversity for video summarization.

Based on whether user's annotations are used for model training, current video summarization methods can be categorized into supervised approaches [4, 5, 6, 13, 22] and unsupervised approaches [1, 8, 10, 11, 14, 15].

Supervised methods learn from the video summaries provided by human to determine the representative clips among a long video. However, the human summaries tend to be very subjective while a video can be summarized in various meaningful ways. The lack of consensus makes it difficult to learn from human summaries directly. Unsupervised approaches, instead, leverage some predefined metrics. The outputs of these methods tend to have similar patterns while the patterns of different video summaries should be varied in fact. Also, performance wise, there remains a gap between the scores achieved by unsupervised approaches and supervised ones. In addition, many of the aforementioned methods use fixed visual features from a pretrained convolutional network as the input, which might result in the loss of temporal information at the very beginning.

For a video dataset, there exist two types of structures, *intra-video* and *cross-video*. Particularly, the intra-video structure refers to the temporal structure existing within a video sequence. Generally, a long video usually consists of multiple clips – the clips tend to capture different semantic aspects when they are temporally far apart, which indicates there exist diverse summaries for a given video. The cross-video structure, on the other hand, refers to the relationships among videos. We found that given a general video collection, every video is *distinctive*. A good representation should be able to capture the distinctive aspects of individual videos, so that given a summary one can easily retrieve the corresponding video from a large collection. Therefore, the cross-video structure correspondeds to the representativeness of generated summaries.

Motivated by the observations above, we propose an end-to-end reinforcement learning approach for video summarization, which takes the intra-video and cross-video structures into consideration simultaneously. Specifically, in order to obtain more diversity reward, the RL model is encouraged to explore the intra-video structure to cover different semantic aspects when given a long video. Meanwhile, a representativeness reward is also proposed to facilitate the RL model to generate distinctive summaries. In terms of video summarization task itself, we formulate it as a sequential decision-making process, and each frame can be selected or not. Notice that the rewards are only presented when the whole summarization process is finished, reinforcement learning framework is suitable to such problem therefore.

2 Related Work

Traditional Methods. Traditional video summarization mostly relied on visual cues such as appearances, objects, and motions [10, 11, 15] to represent video frames or snippets, and used clustering algorithms to generate representative clusters. Auxiliary resources have also been exploited to aid the summarization process such as web images/videos [1, 8] and category information [16]. Most of such methods processed video frames independently, while ignoring the temporal structures.

Supervised Methods. Supervised methods learn from human video summaries to select the key video clips. Gygli *et al* [5] used a linear regressor to select key frames with the highest interestingness scores. Gygli *et al* [6] learned a linear combination of objectives from user summaries. Zhang *et al* [21] transfered the structures of annotated summaries to new videos with a nonparametric approach. Zhao *et al* developed Hierarchical RNNs [23, 24] to segment videos before summarization. The video temporal segmentation [17] can be approched with multi-semantic elements. Gong *et al* [4] employed the determinantal point process (DPP) to encourage diverse selection. Following this work, a number of variants have been developed, including dppLSTM [22], GANdpp [14] and SeqGDPP [19], and DySeqDPP [13]. Our work is not the first to apply Reinforcement Learning to video summarization. Existing methods [25, 26] applied RL to training a summarization network for selecting category-specific keyframes. Their learning framework requires keyframe-labels and category information of training videos. While these supervised methods show promising results on benchmarks, they tend to overfit, given that the user annotations are rather limited compared to the sheer diversity of video contents. Also, the problem of subjectivness, namely a video can be summarized in different meaningful ways, has not been effectively tackled.

Unsupervised Methods. Recently, there has been great progress on developing unsupervised learning methods for video summarization. Generally, different unsupervised methods are based on different assumptions. GANdpp [14] assumes that the learned features of the summary and that of the original video should be similar. Through adversarial learning, it tries to bridge the distribution of the summary features and that of the original video features. This method considers the representation of each video as a whole and focuses on the feature distribution. It does not explicitly leverage the temporal structures within a video.

Compared to the efforts mentioned above, our work differs essentially. Instead of trying to develop new methods to deal with feature distributions, we take a step back and revisit the inherent structures of the video data and develop a new reward function with a new RL pipeline. Our study reveals that these structures play a crucial role in learning a good model for video summarization.

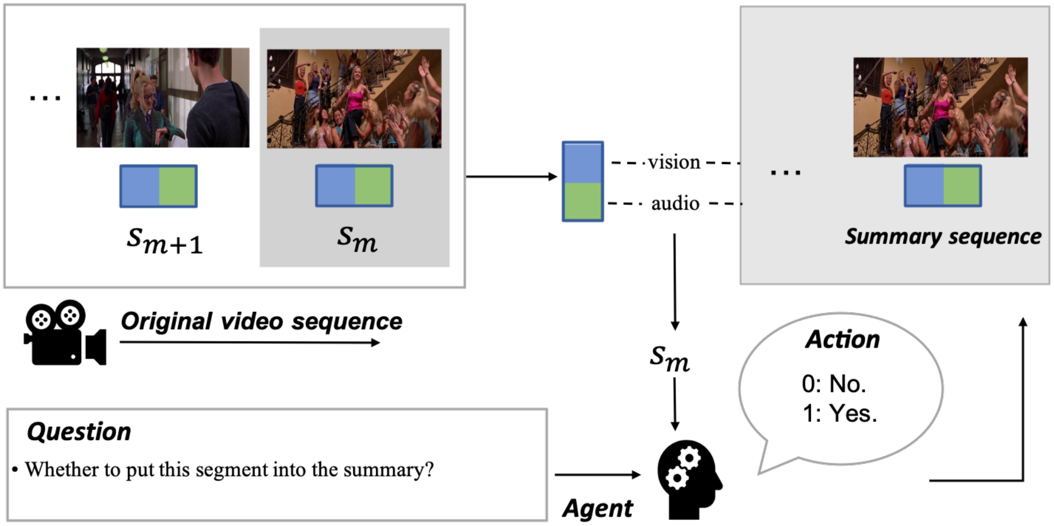

Figure 1: Basic pipeline for the reinforcement learning video summarization.

3 Reinforcement Learning

We will introduce how to model the video summary problem as a Markov Decision Process (MDP) with some basic notations first. And then, four solutions and their corresponding derivations are elaborated in detail. They are 1) REINFORCE, 2) A2C, 3) PPO, 4) Reinforce on evaluation metric.

3.1 Problem Modeling

Suppose a video $\mathcal{V}$ is composed of many segments $\mathcal{V} = \{s_1, s_2, s_3, \cdots, s_L\}$. The segments can be achieved from shot detection or scene detection [17]. To generate a summary, the method needs to make a decision for each segment, *i.e.*, it will decide to keep the segment in the output summary or drop it totally.

Since video summarization is subjective, $V_s' = \{s_2, s_5, \cdots, s_L\}$ and $V_s'' = \{s_1, s_5, \cdots, s_{L-1}\}$ may be both good. it is not appropriate to apply a method that requires per segment supervision signals. Based on these facts, we decide to model the video summary problem as a Markov Decision Process (MDP), which could be solved with reinforment learning. Specifically, the decision maker is regarded as an agent and its action will be 0 or 1, where 1 means putting the current segment into the summary and vice versa. In each step, this MDP proposes a segment and the agent will make its decision accordingly utill the whole process ends.

For example, if we make a decision $\mathcal{A} = \{a_1, a_2, \cdots, a_L\} = \{0, 1, 0, \cdots, 1\}$, the candidate summary would be $V_s = \{s_2, s_5, \cdots, s_L\}$. Note that, the candidate summary is based on the segment level and always exceed the longest time limits, which is usually 15% of the original input video length. Choosing the frames based on candidate summary to achieve maximum rewards is a NP-hard problem, dynamic programming [20] is implemented to obtain a near-optimal solution.

3.2 REINFORCE

For simplicity, we choose a random image x_t from each shot s_t to represent the whole shot. Meanwhile, a ResNet152 [7] is used to extract feature for every image. To ensure the diversity and representativeness of final video summary, we follow [25] to compute the final reward with the proposed equation $R(S) = R_{div} + R_{rep}$.

Since the desision of keeping or not for each shot is a random policy, it's more appropriate to select policy gradient to train our agent. Start from the baisc REINFORCE algorithm, the objective function can be formulated as follows,

$$J(\theta) = E_{\tau \sim p(\tau)} \left[\sum_{t=1}^{T} r\left(s_t, a_t\right) \right],$$

where s_t is the current state at timestep t, a_t is the action taken at timestep t, τ is the trajectory, *i.e.* an ordered sequence $s_1, a_1, s_2, a_2, \ldots, s_t, a_t$ when executing a policy.

The objective is the expectation over all different possible paths an agent takes of the sum of its rewards. The derivative of the obejctive is,

$$\begin{aligned}
\nabla_\theta E_{\tau \sim p(\tau)} \left[\sum_{t=1}^{T} r\left(s_t, a_t\right) \right] &= \int \nabla_\theta p(\tau) \left(\sum_{t=1}^{T} r\left(s_t, a_t\right) \right) d\tau \\
&= \int p(\tau) \nabla_\theta \log p(\tau) \left(\sum_{t=1}^{T} r\left(s_t, a_t\right) \right) d\tau \\
&= E_{\tau \sim p(\tau)} \left[\nabla_\theta \log p(\tau) \sum_{t=1}^{T} r\left(s_t, a_t\right) \right].
\end{aligned}$$

According to the Markov property of our problem,

$$p(\tau) = \prod_{t=1}^{T} \pi_\theta(a_t|s_t) p(s_{t+1}|s_t, a_t),$$

Noting that θ is parameters of the policy $\pi_\theta\left(a_t|s_t\right)$ and has no relation with $p(s_{t+1}|s_t, a_t)$. Therefore,

$$\nabla_\theta \log p(\tau) = \sum_{t=1}^{T} \nabla_\theta \log \pi_\theta\left(a_t|s_t\right).$$

The derivative can be written as,

$$\nabla_\theta E_{\tau \sim p(\tau)} \left[\sum_{t=1}^{T} r\left(s_t, a_t\right) \right] = E_{\tau \sim p(\tau)} \left[\left(\sum_{t=1}^{T} \nabla_\theta \log \pi_\theta\left(a_t|s_t\right) \right) \left(\sum_{t=1}^{T} r\left(s_t, a_t\right) \right) \right].$$

In the real implement, we approximate the gradient with the average gradient of N episodes on the same video

$$\nabla_\theta J(\theta) \approx \frac{1}{N} \sum_{n=1}^{N} \left(\sum_{t=1}^{T} \nabla_\theta \log \pi_\theta\left(a_{n,t}|s_{n,t}\right) \right) \left(\sum_{t=1}^{T} r\left(s_{n,t}, a_{n,t}\right) \right).$$

3.3 A2C

Owing to the large variance of basic REINFORCE algorithm for policy gradient, the training process will be unstable and hard to control, which even makes the network diverge. To circumvent this hurdle, we adopt the classical Advantage Actor-Critic (A2C) algorithm to reduce the variance during training.

Instead of using the Monte-Carlo reward to evaluate the policy, A2C uses a neural network as the critic to estimate the action-value function, *i.e.*, the Q function $Q^\pi(s, a)$. Meanwhile, an extra neural network is applied to fit the value function $V^\pi(s)$, which seves as a great baseline for the critic. Hence, the advantage function can be represented as follows:

$$A^\pi(s, a) = Q^\pi(s, a) - V^\pi(s),$$

where $V^\pi(s)$ is approximated with a parametic neural network $V_{\mathbf{v}}(s)$ and $Q\pi(s, a)$ is approximated with $Q_{\mathbf{w}}(s, a)$ similarly. Replacing the MC reward with the advantage term, the derivative of objective function can be rewritten as,

$$\nabla_\theta J(\theta) \approx \frac{1}{N} \sum_{n=1}^{N} \left(\sum_{t=1}^{T} \nabla_\theta \log \pi_\theta\left(a_{n,t}|s_{n,t}\right) \right) \left(\sum_{t=1}^{T} Q_{\mathbf{w}}(s_{n,t}, a_{n,t}) - V_{\mathbf{v}}(s_{n,t}) \right).$$

Note that, here we use a shared neural networks with just two convolution layers to approximate $Q(s, a)$ and value function $V(s)$ respectively.

3.4 PPO

Policy gradient is the steepset ascend in parameter space, while we want to achieve steepest ascend in the distribution space with KL-divergence, which is called as natural policy gradient. Proximal policy optimization algorithm (PPO) is a simple off-policy RL algorithm for natural policy gradient and has the advantages of stability and reliability. Specifically, the objective funtion value will be maximized within a trust region, which can be represented as follows:

$$\begin{aligned} &\max_{\theta} \mathbb{E}\left[\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)} A_t\right] \\ &s.t. \ \mathbb{E}_t\Big[KL\big[\pi_{\theta_{old}}(\cdot|s_t), \pi_\theta(\cdot|s_t)\big]\Big] \leq \delta, \end{aligned}$$

where A_t is the advantage function introduced in the above section.

Meanwhile, a clipping trick used in PPO algotithm will avoid the policy change dramatically and lead to more stable training process. Finally, the derivate of objective function can be represented as follows:

$$\nabla_\theta J(\theta) \approx \frac{1}{N} \sum_{n=1}^{N} \sum_{t=1}^{T} \Big[\min \big(\frac{\pi_\theta(a_{n,t}|s_{n,t})}{\pi_{\theta_{old}}(a_{n,t}|s_{n,t})} A_{n,t}, \text{clip}(\frac{\pi_\theta(a_{n,t}|s_{n,t})}{\pi_{\theta_{old}}(a_{n,t}|s_{n,t})}, 1-\epsilon, 1+\epsilon) A_{n,t} \big) \Big],$$

where ϵ is set as 0.2 by default.

3.5 Reinforce on Evaluation Metric

The goal of video summarization is to generate a concrete summary and its quality is reflected by the overall evaluation score. Instead of using equation $R(S) = R_{div} + R_{rep}$ to get the final reward, we target to take the evaluation score as the agent's reward in this stage. However, the quality score is estimated on the frame-level, while the agent decisions are based on the segment-level, which means the candidate summary couldn't be evaluated directly.

An effective countermeasure proposed in this paper is to select frames randomly from the candidate summary, and all the chosen frames form the pseudo-summary, whose length can't exceed length limit of course. Specifically, the pseudo-summary $\mathbf{S}_p$ will compare with the ground-truth summaries $\mathbf{S}_{gt}$ innotated by 20 people, where the average similarity score is considered as the final reward. Hence, quality score is formulated as follows:

$$\text{QS} = \frac{1}{20} \sum_{i=1}^{20} \left[\frac{\mathbf{S}_p * \mathbf{S}_{gt}}{\text{len}(\mathbf{S}_{gt})} \right],$$

where $\mathbf{S}_{gt}$ and $\mathbf{S}_p$ are both binary vector for keeping or dropping on frame-level. Actually, the item $\mathbf{S}_p * \mathbf{S}_{gt}$ stands for the temporal overlap between pseudo summaries and ground truth summaries.

During training, we leverage PPO algorithms to reinforce on evaluation metric and the advantage function $A_{n,t}$ in PPO will be improved by our proposed quality score. Besides, other hyper-parameters and training procedures keep the same settings as the PPO section.

4 Experiments

4.1 Setup

Dataset Our frameworks are evaluated on SumMe [5] for the video summarization. The SumMe [5] dataset comes from user generated videos covering holidays, events and sports. It consists of 25 raw or minimally edited user videos and each video is annotated by 15 to 18 different people. The length of the videos ranges from 1 to 6 minutes.

Method	SumME [5]
Video-MMR [12]	26.6
Co-archetypal [20]	26.6
Uniform sampling [3]	29.3
K-medoids [3]	33.4
Vsumm [2]	33.7
Dictionary selection [3]	37.8
GANdpp [14]	39.1
DR-DSN [25]	41.4
Ours (REINFORCE)	**42.4**
Ours (A2C)	**43.9**
Ours (PPO)	**44.6**
Ours (Reinforce on Evaluation Score)	**44.9**

Table 1: Video summarization quantitaive results (%) on SumMe [5]. Our approach achieves the state-of-the-art results.

Evaluation settings We take F-score as our evaluation metrics according to the temporal overlap between predicted summaries and ground truth summaries. Since there are multiple ground truth summaries for one video, we take the average score achieved from all the ground truthes. 5-fold cross validation is used in the experiments *i.e.* 80% of videos for training and the rest 20% for testing.

Implementation details Videos are downsampled to 2fps. ResNet50 [7] is used to extract representation for each shot. And a 5-layer CNN is taken as the agent. SGD is used as the optimizer and the learning rate is set to 0.01.

4.2 Results and Comparison

Quantitative Results We compare our method with other video summarization approaches, including unsupervised methods Video-MMR [12], Co-archetypal [20], Uniform sampling, K-medoids, Vsumm [2], Dictionary selection [3] and supervised approches GANdpp [14], DR-DSN [25].

1) **Overall analysis.** As shown in Table 1, our approaches outperform unsupervised methods by a large margin (∼10 absolutely and ∼30% relatively better). Traditional methods such as Video-MMR, K-medoids, Vsumm are basically based on low-level visual cues and are not able to capture high-level semantic features and do not achieve promising results. Our RL methods also achieves better results (∼5 absolutely and ∼13% relatively better) than existing learning based methods, *e.g.* GANdpp and DR-DSN, since they fail to completely consider video inherent structures. The superior performance of the method demonstrates the effectiveness of the usage of Reinforcement Learning into the methods.

2) **Our methods analysis.** As shown in the bottom block of the Table 1, compared to REINFORCE, the RL method using A2C [9] improves from 42.4 to 43.9 and the RL method using PPO [18] further improves to 44.6. These shows that PPO is better than A2C and REINFORCE in terms of reinforcement video summarization. Futhermore, when we apply supervison on PPO algorithm *i.e.* Reinforce on Evaluation Metric, the performance gets 0.3 absolutely boost. This shows that our methods can also incoporate supervision from human beings.

Qualitative Evaluation Two video summarization examples are shown in Figure 2. We compare our method with dppLSTM [22] and DR-DSN [25]. In the first example, our method captures much more diverse objects and scenes. In the second example video, our approach generates a more diverse summary than the other two by selecting distinctive clips. Note that our method is unsupervised and does not use the video title *Statue of liberty*, it is interesting in the first example our method selects lots of statue of liberty. The reason may contribute to that the statue of liberty is the most distinguishable object compared with the objects in other video instances.

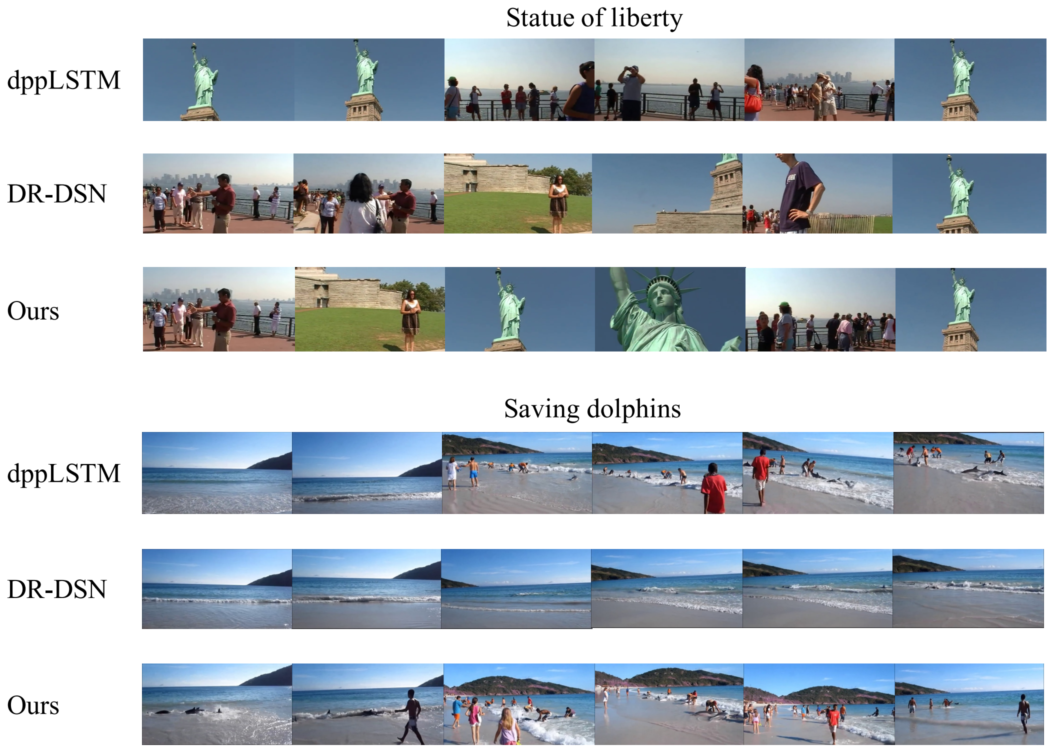

Figure 2: Qualitative Results of two videos *Satue of liberty* and *Saving dolphins* in SumMe [5].

5 Conclusion

In this work, we propose an end-to-end reinforcement learning method for video summarization, which takes the two video structures into consideration simultaneously *i.e.* intra-video and cross-video. Training with reinforcement learning, our summarization model improves a lot on both quantitative and qualitative results across four different RL methods, *i.e.*, REINFORCE, A2C, PPO and reinforce on evaluation metric. The superior performance illustrated in the experiments demenstrates that it's more appropriate to model video summarization to a sequential decision making process. Noting that our model is fully unsupervised, it can get rid of the expensive human annotation and be applied to large-scale unlabeled online videos.

References

[1] W.-S. Chu, Y. Song, and A. Jaimes. Video co-summarization: Video summarization by visual co-occurrence. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3584–3592, 2015. 1, 2

[2] S. E. F. De Avila, A. P. B. Lopes, A. da Luz Jr, and A. de Albuquerque Araújo. Vsumm: A mechanism designed to produce static video summaries and a novel evaluation method. *Pattern Recognition Letters*, 32(1):56–68, 2011. 6

[3] E. Elhamifar, G. Sapiro, and R. Vidal. See all by looking at a few: Sparse modeling for finding representative objects. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1600–1607. IEEE, 2012. 6

[4] B. Gong, W.-L. Chao, K. Grauman, and F. Sha. Diverse sequential subset selection for supervised video summarization. In *Advances in Neural Information Processing Systems*, pages 2069–2077, 2014. 1, 2

[5] M. Gygli, H. Grabner, H. Riemenschneider, and L. Van Gool. Creating summaries from user videos. In *European conference on computer vision*, pages 505–520. Springer, 2014. 1, 2, 5, 6, 7

[6] M. Gygli, H. Grabner, and L. Van Gool. Video summarization by learning submodular mixtures of objectives. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3090–3098, 2015. 1, 2

[7] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 3, 6

[8] A. Khosla, R. Hamid, C.-J. Lin, and N. Sundaresan. Large-scale video summarization using web-image priors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2698–2705, 2013. 1, 2

[9] V. R. Konda and J. N. Tsitsiklis. Actor-critic algorithms. In S. A. Solla, T. K. Leen, and K. Müller, editors, *Advances in Neural Information Processing Systems 12*, pages 1008–1014. MIT Press, 2000. 6

[10] J. Kwon and K. M. Lee. A unified framework for event summarization and rare event detection. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1266–1273. IEEE, 2012. 1, 2

[11] Y. J. Lee, J. Ghosh, and K. Grauman. Discovering important people and objects for egocentric video summarization. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1346–1353. IEEE, 2012. 1, 2

[12] Y. Li and B. Merialdo. Multi-video summarization based on video-mmr. In *Image Analysis for Multimedia Interactive Services (WIAMIS), 2010 11th International Workshop on*, pages 1–4. IEEE, 2010. 6

[13] Y. Li, L. Wang, T. Yang, and B. Gong. How local is the local diversity? reinforcing sequential determinantal point processes with dynamic ground sets for supervised video summarization. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 151–167, 2018. 1, 2

[14] B. Mahasseni, M. Lam, and S. Todorovic. Unsupervised video summarization with adversarial lstm networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, 2017. 1, 2, 6

[15] C.-W. Ngo, Y.-F. Ma, and H.-J. Zhang. Automatic video summarization by graph modeling. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 104–109. IEEE, 2003. 1, 2

[16] D. Potapov, M. Douze, Z. Harchaoui, and C. Schmid. Category-specific video summarization. In *European conference on computer vision*, pages 540–555. Springer, 2014. 2

[17] A. Rao, L. Xu, Y. Xiong, G. Xu, Q. Huang, B. Zhou, and D. Lin. A local-to-global approach to multi-modal movie scene segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 2, 3

[18] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017. 6

[19] A. Sharghi, A. Borji, C. Li, T. Yang, and B. Gong. Improving sequential determinantal point processes for supervised video summarization. In *European Conference on Computer Vision*, pages 533–550. Springer, 2018. 2

[20] Y. Song, J. Vallmitjana, A. Stent, and A. Jaimes. Tvsum: Summarizing web videos using titles. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5179–5187, 2015. 3, 6

[21] K. Zhang, W.-L. Chao, F. Sha, and K. Grauman. Summary transfer: Exemplar-based subset selection for video summarization. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1059–1067. IEEE, 2016. 2

[22] K. Zhang, W.-L. Chao, F. Sha, and K. Grauman. Video summarization with long short-term memory. In *European conference on computer vision*, pages 766–782. Springer, 2016. 1, 2, 6

[23] B. Zhao, X. Li, and X. Lu. Hierarchical recurrent neural network for video summarization. In *Proceedings of the 2017 ACM on Multimedia Conference*, pages 863–871. ACM, 2017. 2

[24] B. Zhao, X. Li, and X. Lu. Hsa-rnn: Hierarchical structure-adaptive rnn for video summarization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7405–7414, 2018. 2

[25] K. Zhou, Y. Qiao, and T. Xiang. Deep reinforcement learning for unsupervised video summarization with diversity-representativeness reward. *arXiv:1801.00054*, 2017. 2, 3, 6

[26] K. Zhou, T. Xiang, and A. Cavallaro. Video summarisation by classification with deep reinforcement learning. *arXiv preprint arXiv:1807.03089*, 2018. 2