Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 01, January-2020

Digital Video Summarization Techniques: A Survey

Ashenafi Workie[1]
[1]MSc. Student
Adama Science and Technology University,
Department of Computer Science (CVR)

Rajesh Sharma[2]
[2]Assistant Professor
Adama Science and Technology University,
Department of Computer Science

Yun Koo Chung[3]
[3]Professor
Adama Science and Technology University,
Department of Computer Science

***Abstract*:- Video summarization which gives a short and precise representation of original video clips by showing the most representative synopsis is gaining more attention. The main objective of Video summarization is to provide a clear analysis of the video by removing redundant and extracting key frames contents from the video. The architecture in video summarization shows how a large video skims in to short and story contents. Many types of research were done in the past and ongoing until now. Therefore, multiple methods and techniques proposed by researchers from classical computer vision until the recent deep learning approaches. Most literature shows that most of the video generation and summarization approaches shift into deep generative models and variational auto encoders. These techniques may fall into summarized, unsupervised and deep reinforcement learning approaches. Video representation categorized in static and dynamic summarization ways. But video summarization still challenging with different problems, these are computational devices, complexity, and lack of dataset are some them. The effective implementation of video summarization applied in different real-world scenarios like movies tailor in the film industry, highlight in football soccer, anomaly detection video surveillance system.**

Keywords: Video summarization, supervised, unsupervised, dynamic summarization;"

1. INTRODUCTION

Video summarization is a mechanism of creating a short time original video while keeping main stories/content [1] from large *video dataset.* A wide range of applications can be achieved through video summarization. For example, if we have surveillance video at home events intentional intended to reduce a few minutes' meaningful illustrating anomaly events for easier understanding. In a sports video, the same thing is to so to summarize illustrating the most important events such as goals, penalty kicks, etc. In video summarization, the input is a video with whole content as such original. The aim of to choose small content of keyframe from the original input video to produce a summary video that can express with being explicitly watching the whole part and without losing important content [11]. In the long-duration video, viewers may not have enough time to watch the whole video. A viewer may interest to watch on the particular issue under important where the user is searching for [28]. Recently, it has been attracting much interest in extracting the representative visual elements from a video for sharing on social media, which aims to effectively express the semantics of the original lengthy video [24].

In today's digital world there are so many videos that were created and release over different stream media. Especially such videos uploaded to the internet or cloud, therefore it needs a high bandwidth network to browse it. Video summarization which gives a short and precise representation of original video clips by showing the most representative synopsis is gaining more attention. This is good practice to save multiple resources time, storage and other network and multimedia infrastructure [2]. In fact, there are two types of video summaries: 1) *static video abstract*, which is a sequence of keyframes and 2) *dynamic video skimming*, which is a collection of dynamically-composed audio-video sub-clips, and in both cases, the aim is to collect the most interesting or important video segments that show the essence of the original clips.In the real-world Even though we have plenty of video with large content, all the possible frames may not equality important or some of the content is redundant or irrelevant content. But working on video summarization is quite a difficult task while finding the potential information to create an interesting short video with either no repetitive or missed information from the whole content provided as an input. Many types of research have done in video summarization. In problem is under-constraint since this summery is hag of the subjectivity of understanding [24].

This research explains both classical computer vision techniques as well as the recent deep learning approaches to summarize potential relevant information apart from the whole video contents given as input. It also explores methods or techniques used under video summarization while working with its application area in different scenarios. Literature review and surveys on the main objective, gaps or limitations and in-line with its method and contribution. This paper organizes as follows section I architecture, Section III related works, Section IV Application area, and section V conclusion.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 01, January-2020

2. ARCHITECTURE IN VIDEO SUMMARIZATION

In the video summarization is a process that explains how large video content will summarize into short and concise information. The videos in small computation and storage resources regardless of losing an important section of the content. The mapping between the ground truth (original video) and the summarize one also important since. The following figure 1 shows the basic architecture of video summarization in line with the mapping function between the inputs (a large chunk of frame sequences) and summarizes (short and selected frame sequence).

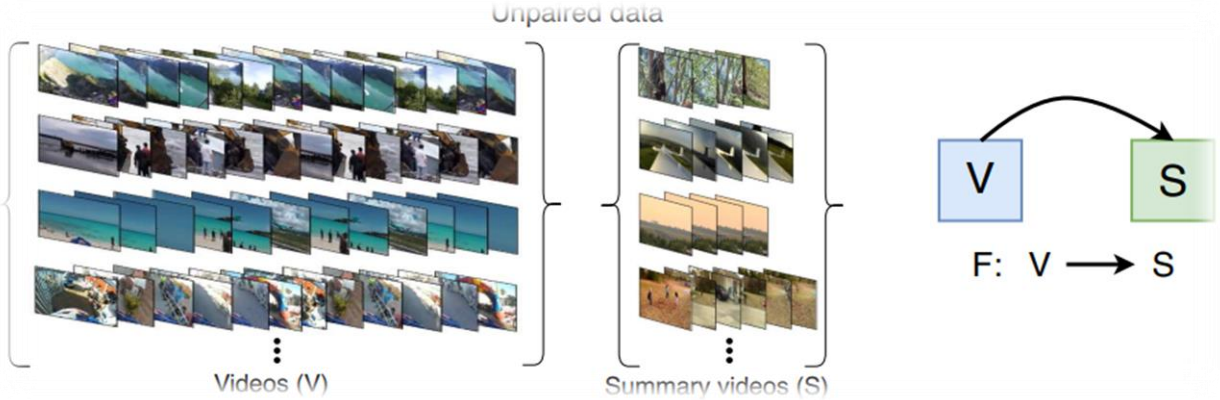

Figure 1 Learn video summarization from the whole contents by mapping F: V → S (right) linking two different domains V and S. [11].

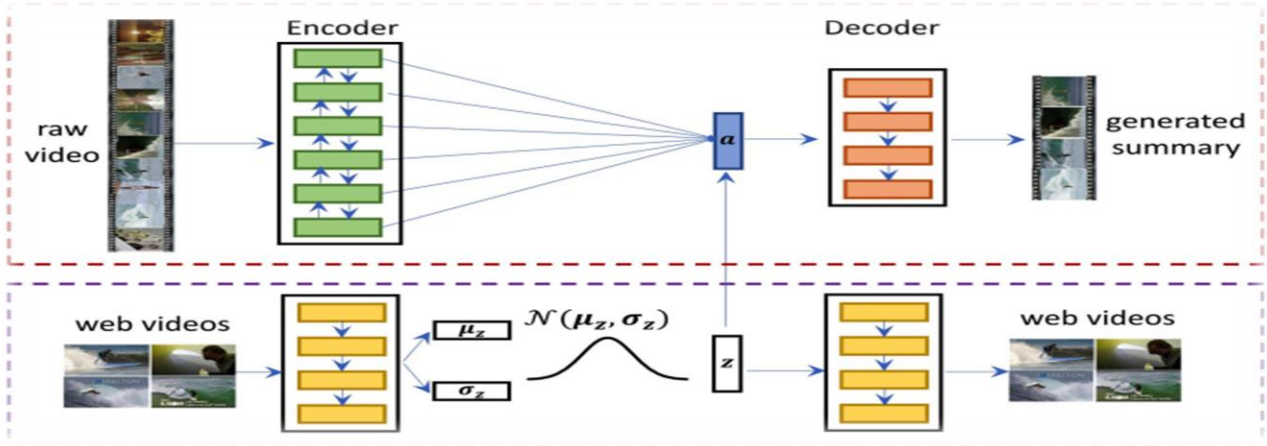

Figure 2 GAN based framework for video summarize using VAE variational auto encoder

3. VIDEO SUMMARIZATION

Multiple types of research done in video summarization potential methods and application areas in keyframe scenarios like a high-light football game. The proposed methods outlined by the researchers so far were both supervised and unsupervised approaches.in supervised methods. But recently research the reinforcement learning mechanism also applied to it.

A. Supervised Methods

In a supervised learning approach video, summarization learns from labelled data by consisting of videos and along with ground-truth summary videos. Getting an annotated data is quite expensive, difficult and costly even in some way it becomes impossible [24] [11]. Due to its requirement of human-annotated video-summary pairs or per frame the training label is guided to summarize the video accordingly.

To address under the supervision of human-annotated video to produce a subset of contents. Selection problem. This annotation training sample along with the original source video that can teach how summarization will works while selecting informative subsets [14]. The target label annotation which is a user -created summaries that help by teachers for selecting the best video frames directed on how the algorithm to summarize in accordance with the guidance of user input fashion. Much work has been proposed to measure shot importance through supervised learning.

B. Unsupervised Methods

Unsupervised video summarization in Spatio-temporal feature and reduction with clustering methods. Unlike supervised methods without including annotated video summarize it is possible to create an unsupervised way. In Egocentric video summarization methods have also used unsupervised learning to categorize sports actions [12]. However, the video is challenging problems because of that placement of camera shows in a video a great variation in object vanishing points or angle, illumination conditions, and movement. The author used Alex Net which is a convolutional neural network to filter the key-frames (frames where camera wearer interacts closely with the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 01, January-2020

people) while finding a subset abstract story from whole contents.

C. *Reinforcement Methods*

based on the statistical probability of a given frame in a given video sequence. Use an end to end learning for training so needs high computational resources.

Reinforcement deep learning without a label
or as an unsupervised video summarization approach works in the sequential process [28]. In this paper, the author used a deep summary network that can predict

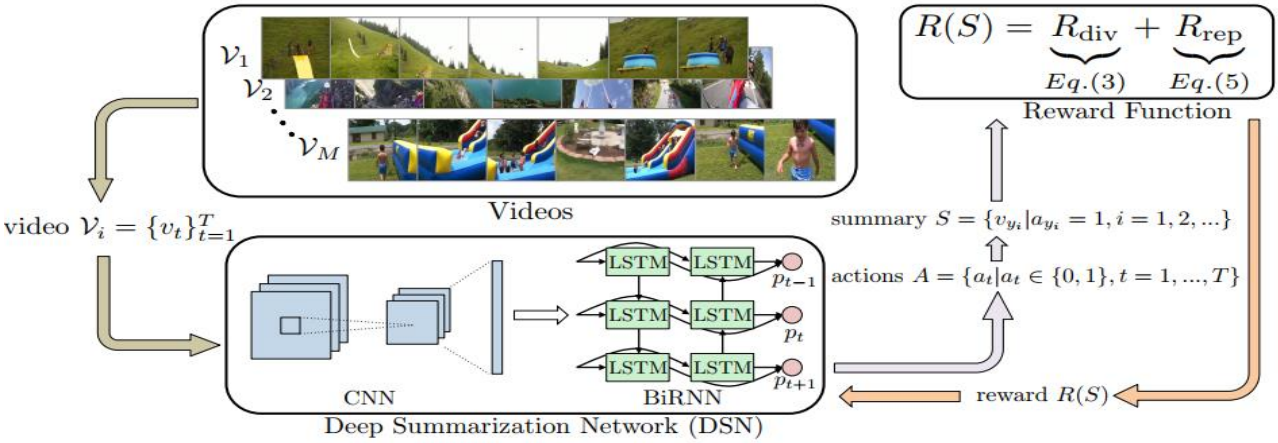

Figure 3

Training deep summarization network (DSN) via reinforcement learning. DSN receives a video *Vi* and takes actions [28].

4. VIDEO REPRESENTATION

Video representation is an important problem in video pre-processing. A good video representation should include the key point and useful information for discrimination by discarding unnecessary information. Generally, in this video processing, video frames are usually represented as a matrix. In this paper author methods method, use the luminance information to keep the data in every single frame. In video summarization, mainly focus on creating a video summary that can finish watching within a short period of time.
During this process, the generating mechanism for creating video frame contents may be static or dynamic approaches. The static video summarization is also known as R frame. Still, images consist of 3 types of classification. These are sampling, shot segmentation and scene-based classification where keyframes are extracted pre sampling in uniform as well as in a random manner. On the other hand, the dynamic video summarization during producing summarize informative video contents.

5. LITERATURE REVIEW

Taru et al., 2017 [31] propose a new method of video summarization into text. This approach takes an input of consumed video. The authors are motivated while the viewer of a video may not have time to go through full content. The aim of this research is a user can easily understand the summary in text format. Srinivas et al., 2016 [30] proposed an improved method for video summarization techniques. The aim is to get a summary content of a video which is interesting to the viewer and representing the whole video. The result is better while it compares with other methods.

Anomaly author, 2020 [29] proposed ILS-SUMM which an iterative local search for unsupervised video summarization. Its objective is to create automatically a short summary of the whole contents. Moreover, to indicate the high scalability of ILS-SUMM, the authors introduce a new dataset consisting of videos of various lengths. Zhou et al., 2018[28] develop a deep summarization network (DSN) to summarize videos for predicts each video frame probabilities. The training is an end to end reinforcement learning .so the result is better than that of supervised approaches. SARMADI et al., 2017[27] proposed a general video summarization method that is divided into static and dynamic; Static Summary done through a keyframe.

Cai et al., 2018[24] proposed a generative modeling framework to learn representation with a variational autoencoder. Encoder-decoder attention for saliency estimation of raw video for generating the summary.

Song et al., 2015 [23] present TvSum an unsupervised summarization framework. Introduce a new benchmarks dataset. This approach produces superior quality summaries compare with other approaches. Yuan et al.,2017[22] present a novel Deep Side Semantic Embedding (DSSE) model to generate video summaries. In semantic relevance can be more effectively measured. Fu et al., 2019[4] proposed a GAN-based training framework through an unsupervised and supervised video summarization approach. The generator is focused on Ptr-Net that generates the cutting points of summarized fragments. Where “SumMe, TVSum, YouTube, and LoL datasets with remarkable improvements”. Rochan et al., 2019[11] present a method that learns to generate optimal video summaries. This r model goal is to learn a mapping function F: V → S.

Chu et al., 2015[8] developed a Maximal Biclique Finding (MBF) algorithm that is optimized to find sparsely co-occurring

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 01, January-2020

patterns. The results suggest that summaries generated by visual co-occurrence tend to match more closely with human-generated summaries. Agyeman et al.,2019[7] present a deep learning approach to summarizing long soccer videos which are three-dimensional Convolutional Neural Network (3D-CNN) and Long Short-term Memory (LSTM) – Recurrent Neural Network (RNN). Fajtl et al. 2019[6] propose a novel method for supervised bi-directional recurrent networks such as BiLSTM combined with attention. Elfeki et al. 2019[4] conduct extensive experiments on the compiled dataset in addition to three other standard benchmarks. Vasudevan et al., 2017[3] Introduce a new dataset, annotated with diversity and query-specific relevance labels. In the video, summarization can be a single-view or multiview. In single-view video Summarization proposed for summarizing a single-view using videos supervised approaches usually stood out with best performances. On the other hand, multi-view Video Summarization proposed a method that tends to rely on feature selection in using an unsupervised optimization paradigm.

Table 1 Summary of related works

SN	Articles on topics	Objective	Methods
1	Single-view video Summarization [4]	▪ summarizing single-view videos ▪ Determinantal point processes (DPP)	▪ RNN, LSTM, (Bi-LSTM) and DPP
2	Multiview video Summarization [4]	▪ summarization methods tend to rely on feature selection ▪ an unsupervised optimization paradigm ▪ multi-view video summarization	▪ Graph-based approach ▪ 3D structure view axis
3	Supervised Methods Summarization [6]	▪ A target label annotation which is a user-created	▪ TVSUM, DPP SQDPP, andBiLSTM
4	Unsupervised Methods Summarization [23]	▪ use hand-crafted heuristics to satisfy ▪ Diversity, representativeness.	ILSUM, GAN, and VAE
5	Reinforcement approach	Reinforcement deep learning approach works in the sequential process	DSN

6. GAN BASED VIDEO SUMMARIZATION

GAN-based training framework is a neural network that consists two adversarial Networks called generator and discriminator. This framework which combines the merits of unsupervised and supervised video summarization approaches [23]. The generator network is an attention-aware Ptr-Net that generates the cutting points of summarized fragments whereas the discriminator is a 3D CNN classifier to judge whether a fragment is from a ground-truth or a generated summarization. Therefore, GAN is a better one than that of others in different metri

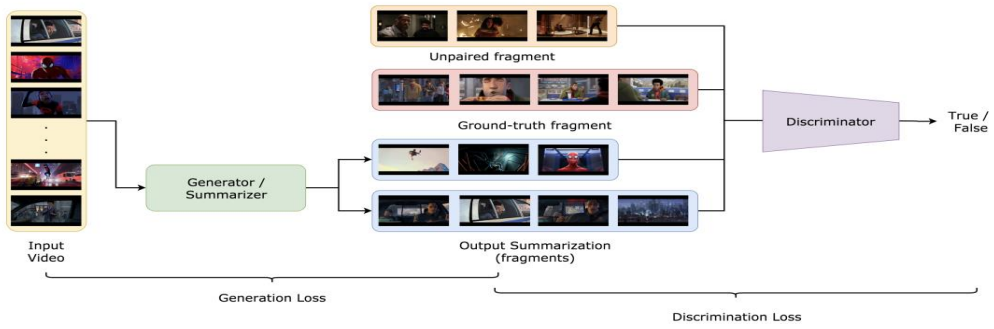

Figure 4 an overview of our method. Present a GAN-based approach to video summarization

7. CHALLENGES IN VIDEO SUMMARIZATION

In video summarization, the process is trivial in skimming important sections of the original content. The main challenge is the 1) training and preprocessing unbalanced training-test length.2) complexity in application and development.3) The temporal relationship between video frames in information like video tags, captions, comments and so on will need to be investigated in the future[22].4) inexpensiveness of training video mostly the annotated dataset.

8. APPLICATION OF VIDEO SUMMARIZATION

In video summarization, Keyframe extraction is an important part of many video applications, like video indexing, browsing, and video retrieval. Many professional and educational applications that involve generating or using large volumes of video and multimedia data are prime candidates for taking advantage of video content analysis techniques [33]

- movie trailer (film industry)
- Advert creation (Advertisement)
- football highlights (Recreation means)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 01, January-2020

- Anomaly detection from video surveillance(security)
- Remove redundancy
- Reduce computational time, storage requirements
- Data visualization, Labeling
- Search, Retrieval, Recommendation

9. CONCLUSION

This paper presents video summarization techniques, applications, and challenges. the architecture of video summarization focusses on a chunk of video summarize into short skim of potential information. Recently classical computer vision techniques for video summarization methods are dynamically shifting to deep learning especially deep generative model, Recurrent neural network, variational auto encoders. Video summarization may handle in supervised (TVSUM, RNN and DPP SQDPP and BiLSTM), unsupervised
(ILSUM, GAN and VAE) and even deep reinforcement learning approach (DSN).
GAN-based training framework as a powerful means of image and video generation. Video summarization is challenged different factor these ranges from dataset until computational device, especially in new deep learning models. The application video summarization can be used in a different scenario for different reasons these can be recreation, film industry, and security and reduce computation power. In general, a deep generative model and variational auto encoder is a relatively good way of video summarization techniques in both static and dynamic summarization approaches

REFERENCES

[1] Otani, M., Nakashima, Y., Rahtu, E., & Heikkila, J. (2019). Rethinking the Evaluation of Video Summaries. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7596-7604).

[2] Luthra, V., Basak, J., Chaudhury, S., & Jyothi, K. A. N. (2008). A Machine Learning based Approach to Video Summarization.

[3] Vasudevan, A. B., Gygli, M., Volokitin, A., & Van Gool, L. (2017, October). Query-adaptive video summarization via quality-aware relevance estimation. In *Proceedings of the 25th ACM international conference on Multimedia* (pp. 582-590). ACM.

[4] Fu, T. J., Tai, S. H., & Chen, H. T. (2019, January). Attentive and Adversarial Learning for Video Summarization. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 1579-1587). IEEE.

[5] Zhang, S., Zhu, Y., & Roy-Chowdhury, A. K. (2016). Context-aware surveillance video summarization. *IEEE Transactions on Image Processing*, *25*(11), 5469-5478.

[6] Fajtl, J., Sokeh, H. S., Argyriou, V., Monekosso, D., & Remagnino, P. (2018, December). Summarizing Videos with Attention. In *Asian Conference on Computer Vision* (pp. 39-54). Springer, Cham.

[7] Agyeman, R., Muhammad, R., & Choi, G. S. (2019, March). Soccer Video Summarization Using Deep Learning. In *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)* (pp. 270-273). IEEE.

[8] Chu, W. S., Song, Y., & Jaimes, A. (2015). Video co-summarization: Video summarization by visual co-occurrence. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3584-3592).

[9] Otani, M., Nakashima, Y., Rahtu, E., Heikkilä, J., & Yokoya, N. (2016, November). Video summarization using deep semantic features. In *Asian Conference on Computer Vision* (pp. 361-377). Springer, Cham.

[10] Khan, S., & Pawar, S. (2015). Video summarization: survey on event detection and summarization in soccer videos. *International Journal of Advanced Computer Science and Applications*, *6*(11).

[11] Rochan, M., & Wang, Y. (2019). Video Summarization by Learning from Unpaired Data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7902-7911).

[12] Ghafoor, H. A., Javed, A., Irtaza, A., Dawood, H., Dawood, H., & Banjar, A. (2018). Egocentric Video Summarization Based on People Interaction Using Deep Learning. *Mathematical Problems in Engineering*, *2018*.

[13] Taskiran, C. M. (2006, January). Evaluation of automatic video summarization systems. In *Multimedia Content Analysis, Management, and Retrieval 2006* (Vol. 6073, p. 60730K). International Society for Optics and Photonics.

[14] Gong, B., Chao, W. L., Grauman, K., & Sha, F. (2014). Diverse sequential subset selection for supervised video summarization. In *Advances in Neural Information Processing Systems* (pp. 2069-2077).

[15] Wei, H., Ni, B., Yan, Y., Yu, H., Yang, X., & Yao, C. (2018, April). Video summarization via semantic attended networks. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

[16] Fu, T. J., Tai, S. H., & Chen, H. T. (2019, January). Attentive and Adversarial Learning for Video Summarization. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 1579-1587). IEEE.

[17] Elkhattabi, Z., Tabii, Y., & Benkaddour, A. (2015). Video summarization: Techniques and applications. *International Journal of Computer and Information Engineering*, *9*(4), 928-933.

[18] Jadon, S., & Jasim, M. (2019). Video Summarization using Keyframe Extraction and Video Skimming. *arXiv preprint arXiv:1910.04792*.

[19] Divakaran, A., Peker, K. A., & Sun, H. (2001, January). Video summarization using motion descriptors. In *Storage and Retrieval for Media Databases 2001* (Vol. 4315, pp. 517-522). International Society for Optics and Photonics.

[20] Cai, S., Zuo, W., Davis, L. S., & Zhang, L. (2018). Weakly-supervised video summarization using variational encoder-decoder and web prior. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 184-200).

[21] Yuan, Y., Mei, T., Cui, P., & Zhu, W. (2017). Video summarization by learning deep side semantic embedding. *IEEE Transactions on Circuits and Systems for Video Technology*, *29*(1), 226-237.

[22] Song, Y., Vallmitjana, J., Stent, A., & Jaimes, A. (2015). Tvsum: Summarizing web videos using titles. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5179-5187).

[23] Cai, S., Zuo, W., Davis, L. S., & Zhang, L. (2018). Weakly-supervised video summarization using variational encoder-decoder and web prior. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 184-200).

[24] Sebastian, T., & Puthiyidam, J. J. (2015). A survey on video summarization techniques. *Int. J. Comput. Appl*, *132*(13), 30-32.

[25] Mundur, P., Rao, Y., & Yesha, Y. (2006). Keyframe-based video summarization using Delaunay clustering. *International Journal on Digital Libraries*, *6*(2), 219-232.

[26] Potapov, D., Douze, M., Harchaoui, Z., & Schmid, C. (2014, September). Category-specific video summarization. In *European conference on computer vision* (pp. 540-555). Springer, Cham

[27] Zhou, K., Qiao, Y., & Xiang, T. (2018, April). Deep reinforcement learning for unsupervised video summarization with diversity-representativeness reward. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

[28] Shemer, Y., Rotman, D., & Shimkin, N. (2019). ILS-SUMM: Iterated Local Search for Unsupervised Video Summarization. *arXiv preprint arXiv:1912.03650*.

[29] Srinivas, M., Pai, M. M., & Pai, R. M. (2016). An Improved Algorithm for Video Summarization–A Rank Based Approach. *Procedia Computer Science*, *89*, 812-819.

[30] Jiang, R. M., Sadka, A. H., & Crookes, D. (2009). Advances in video summarization and skimming. In *Recent Advances in Multimedia Signal Processing and Communications* (pp. 27-50). Springer, Berlin, Heidelber.