International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET2310265

NLP based Video Summarisation using Machine Learning

Prof. Kapil Hande[*1], Hrushi Karlekar[*2], Pranit Yeole[*3], Aditya Likhar[*4], Himanshu Rangari[*5]
[*1] Guide, Department of Computer Science & Engineering, Priyadarshini Bhagwati College Of Engineering,Nagpur, Nagpur, Maharashtra, India
[*2,3,4,5] Student, Department of Computer Science & Engineering, Priyadarshini Bhagwati College Of Engineering,Nagpur, Nagpur, Maharashtra, India

ARTICLEINFO

Article History:

Accepted: 01 April 2023
Published: 14 April 2023

Publication Issue
Volume 10, Issue 2
March-April-2023

Page Number
456-461

ABSTRACT

More people are capturing their daily lives with video data, because to the wide availability of recording equipment. However, the sheer volume of video content makes it more challenging to manage, especially for lengthy movies like security or CCTV footage. For a richer and more succinct condensing of the film will result from automatically detecting the key sections and frames of larger videos and providing them with captions. Users still need to spend time searching or navigating through a summarized video. To extract a shortened version of the footage's information into text form, automatic video summarizing has been proposed. With simply a text summary, the suggested system provides a quick semantic understanding of a lengthy film using LSTM model and the summary can be taken in 3 major different languages (English, Hindi, & Marathi).

Keywords : Video Summarizer, Text Summarizer, Long Short Term Memory, Machine Learning, Summarizer, Abstractive, Extractive

I. INTRODUCTION

One of the most popular ways to access visual information is now video. It would take almost 85 years just to view every video that is published to YouTube every day due to the massive volume of video data! Therefore, it is crucial to have automated methods for video content analysis and comprehension. Automatic video summarization in particular is a crucial tool for assisting human users in browsing video material. An effective video summary would condense the main points of the original video into a concise, viewable overview.

There are many ways that video summaries can cut down on the length. In this study, we concentrate on the two most typical ones: keyframe selection, in which the system identifies a sequence of defining frames [1] [2] [3] [4] [5] , and key sub-shot selection, in which the system identifies a sequence of defining techniques each of which is a set of frames that is temporally contiguous and spans a brief time interval.

The interdependency among video frames is intricate and incredibly heterogeneous when it comes to a video summary. This is not wholly unexpected because human viewers evaluate whether the structure would prove useful to maintain for a summary based on their high-level semantic grasp of the video's contents (and how the narratives are developing). Temporally close video frames, for instance, are frequently visually identical and communicate redundant information, therefore they should be condensed when determining what the keyframes are. The opposite, however, is untrue. Thus, visually comparable frames do not always need to be contemporaneous. Consider summarising the video as "leave home in the morning, come back for lunch at home, leave again and come home at night," for instance.

While the frames related to the “at home” scene can be visually similar, the semantic flow of the video dictates none of them should be eliminated. Thus, a summarization algorithm that relies on examining visual cues only but fails to take into consideration the high-level semantic understanding about the video over a long-range temporal span will erroneously eliminate important frames. Essentially, the nature of making those decisions is largely sequential – any decision including or excluding frames is dependent on other decisions made on a temporal line.

II. LITERATURE SURVEY

B. Mahasseni et al. [6] in order to select a sparse collection of video frames that best represent the input video, the topic of unsupervised summarization of video is addressed in this study. Our main concept is too unsupervised develop a deep summarizer networks to reduce the distance between training films and the dissemination of their summaries.

M. Z. Khan et al. [7] evaluation of four comparison datasets made up of films depicting various events from both first- and third-person perspectives shows that our performance is competitive with that of fully supervised state-of-the-art methods.

D. Sahrawat et al. [8] with Kernel Temporal Segmentation (KTS) for shot segments and a global attention-based customized memory network coupled with LSTM for shot score learning, we provide a straightforward method for summarising videos. The model's learning capacity is increased by the updated memory network known as the Global Attention Memory Module (GAMM), and with its incorporation of LSTM, it is further able to pick up better contextual characteristics. Research on data sets such as TVSum and SumMe reveals that our technique performs roughly 15% better than the state-of-the-art.

R. Agyeman et al. [9] by utilising the spatiotemporal learning capabilities of three-dimensional convolutional neural networks, also known as (3D-CNN) and long short-term memory - recurrent neural networks, this research provides a deep learning strategy to summarise lengthy football movies.

T. Hussain et al. [10] there are a number of proposed low-level features and technique-based soft computing strategies that fall short of completely utilizing MVS. In this paper, author integrate soft computing methods based on deep neural networks into a two-tier system to accomplish MVS. Target-appearance-based shot segmentation is carried out by the first online layer, which then saves the results in a lookup table before sending them to the cloud for additional processing. In order to obtain probabilities of in formativeness and provide a summary, the second tier collects specific characteristics from each frame of the order in the lookup list and passes them to deep bidirectional short-term memory.

G. Yalınız et al. [11] for the unsupervised video summarization problem, an approach combining deep reinforcement learning and autonomously recurrent neural networks is put forth. In this method, the algorithm can be developed with more layers and steps without encountering gradient-related issues.

III. METHODS AND MODULES

We outline our procedures for summarising videos in this section. We begin by clearly stating the problem including the markings before quickly going over LSTM [12] [13], the foundation of our strategy. Then, we present our first summarising model, called LSTM. Then, we go over how we may improve LSTM by combining it with the determinantal point process (DPP), which further takes into account the summarization structure (such as the diversity among chosen frames).

Figure 1. Proposed System Flow Diagram

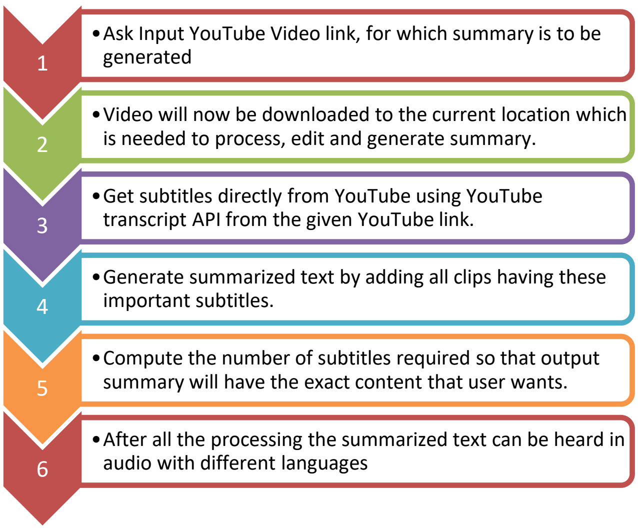

Figure 2. Proposed System Working Flow

Our suggested technique uses consumer choice and image quality to distinguish intriguing sections from lengthy films. In order to produce audio and written summaries, deep visual captioning systems extract key frames from intriguing segments. To create sentence summaries from the entire video, extractive techniques are fed captions from interesting segments. The section summary is useful for organizing and searching through films, and each of the segment captions is useful for quickly locating the right temporal offset in lengthy recordings.

A. Requirements

i) LSTM:

Long short-term memory networks, or LSTMs, are employed in deep learning. Many recurrent neural networks are able to learn long-term dependencies, particularly in tasks involving sequence prediction. The complete input sequence is read by an encoder LSTM, with one word being sent into the encoder at each timestamp. The information is then processed at each timestamp, and the contextual data from the input sequence is captured.

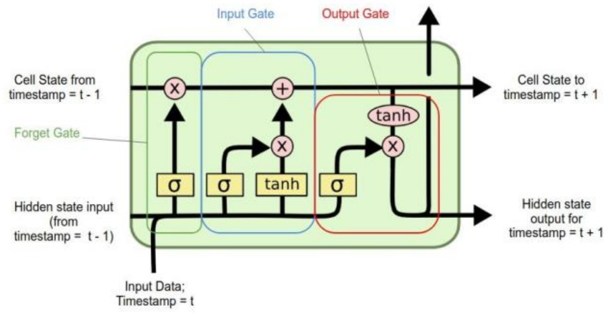

Figure 3. LSTM Working Module

ii) Natural Language Processing (NLP):

The practise of condensing long texts into manageable paragraphs or sentences is called NLP text summarization. This technique preserves the text's meaning while still extracting important information. This shortens the amount of time needed to comprehend lengthy items like articles without omitting important details. [14]

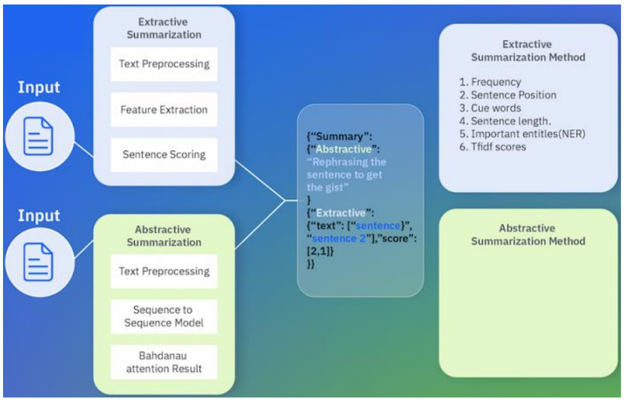

Figure 3. NLP Working Module [14]

iii) Convolutional Neural Network (CNN):

In Automatic Text Summarization (ATS), CNN plays an important role as CNN is meant to work on images and videos. With a text matrix representation, CNN-ATS, a convolutional neural network, provides automatic text summarization. In order to identify the best CNN configurations, analyse the phrases, and choose the most informative one, the CNN-ATS deep learning system was utilised to examine the improvements brought about by the increase in depth. For document summaries, important sentences are extracted. [15]

IV. RESULTS

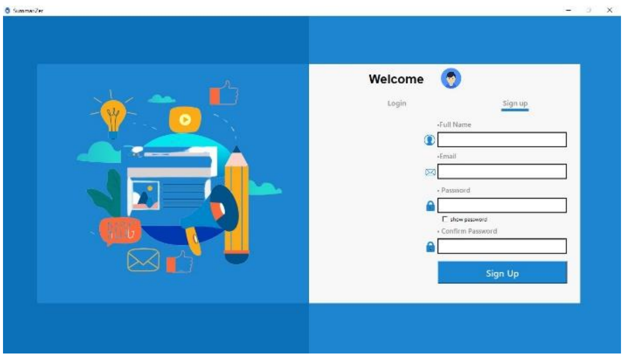

Figure 5. Signup Page

As shown in Figure 5 above a signup form is created for the new user where user can register themselves by providing some basic details.

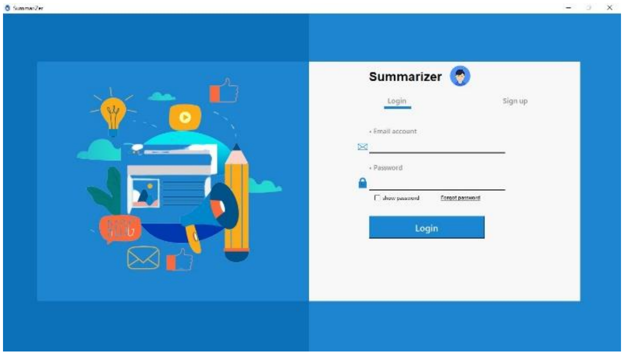

Figure 6. Login Page

As shown in figure 6 above if the user is new to the panel and registered themselves can now login by entering username and password in the given fields.

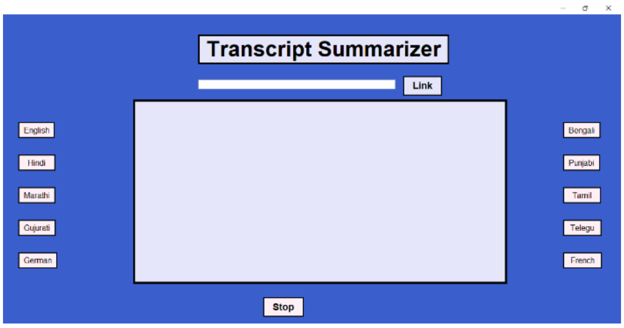

Figure 7. Main Dashboard Module of Video Summarizer

The above image shows the dashboard where an input area is given for the video link and there are multiple buttons displaying different languages in it. User can upload and YouTube video URL in the input field and then select any of the given language button and the summary will be generated after a few second. There is an additional feature of playing the generated summary in the selected language. This feature enhance the creativity of the project.

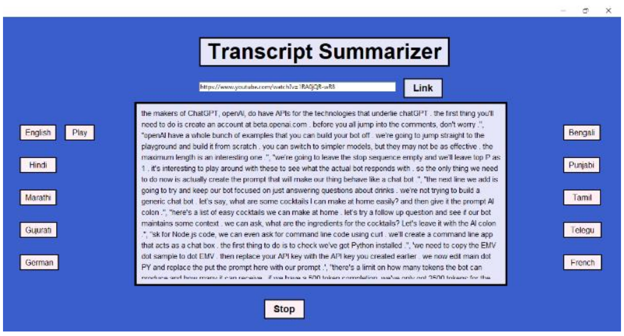

Figure 8. Displaying Summary in English Language

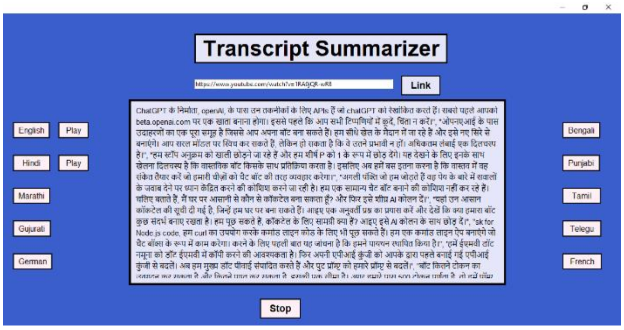

Figure 8. Displaying Summary in Hindi Language

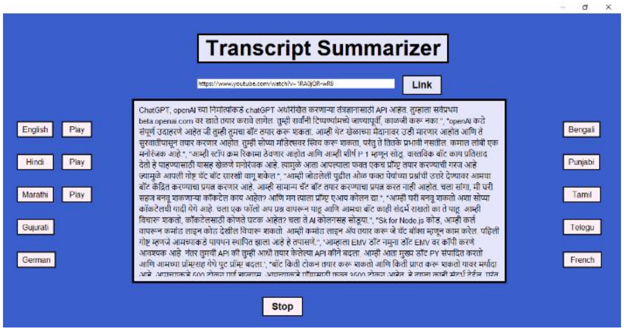

Figure 9. Displaying Summary in Marathi Language

V. CONCLUSION

In order to create new supervised learning techniques for automatic video summarization, our work investigates long short-term memory. On two difficult benchmarks, our LSTM-based models perform better than competing techniques. The ability of LSTMs to model variable range inter-dependencies as well as our suggestion to combine LSTMs' strength with DPP to explicitly represent inter-frame repulsiveness to promote a variety of selected frames are two of the major contributing elements. We demonstrate how, despite the variability in style and content of other annotated video datasets, it is possible to mitigate the demand for LSTMs' huge number of annotated samples. The preliminary findings are quite encouraging and point to possible future research avenues for creating more complex methods that can combine a large number of existing video datasets for video summarising.

VI. REFERENCES

[1]. H. W. J. Z. D. S. S. Zhang, "An integrated system for contentbased video retrieval and browsing," Pattern recognition, vol. 30(4), p. 643–658, 2018.

[2]. B. C. W. G. K. S. F. Gong, "Diverse sequential subset selection for supervised video summarization," In: NIPS, 2014.

[3]. P. R. Y. Y. Y. Mundur, " Keyframe-based video summarization using delaunay clustering," International Journal on Digital Libraries, vol. 6(2), p. 219–232, 2016.

[4]. D. H. G. C. T. Liu, "A hierarchical visual model for video object summarization," IEEE transactions on pattern analysis and machine intelligence , vol. 32(12), p. 2178–2190, 2020.

[5]. Y. G. J. G. K. Lee, "Discovering important people and objects for egocentric video summarization," In: CVPR, 2012.

[6]. M. L. a. S. T. B. Mahasseni, "Unsupervised Video Summarization with Adversarial LSTM Networks," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA,, no. doi: 10.1109/CVPR.2017.318., pp. pp. 2982-2991, 2017.

[7]. S. J. S. u. H. M. A. H. a. M. U. G. K. M. Z. Khan, "Video Summarization using CNN and Bidirectional LSTM by Utilizing Scene Boundary Detection,," International Conference on Applied and Engineering Mathematics (ICAEM), Taxila, Pakistan, no. doi: 10.1109/ICAEM.2019.8853663., pp. pp. 197-202, 2019.

[8]. D. Sahrawat et al., "Video Summarization using Global Attention with Memory Network and LSTM," 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM), Singapore,, no. doi:

10.1109/BigMM.2019.00-20., pp. pp. 231-236,, 2019.

[9]. R. M. a. G. S. C. R. Agyeman, "Soccer Video Summarization Using Deep Learning," 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), San Jose, CA, USA,, no. doi: 10.1109/MIPR.2019.00055., pp. pp. 270-273, 2019.

[10]. K. M. A. U. Z. C. S. W. B. a. V. H. C. d. A. T. Hussain, "Cloud-Assisted Multiview Video Summarization Using CNN and Bidirectional LSTM," in IEEE Transactions on Industrial Informatics, Vols. vol. 16, no. 1, no. doi: 10.1109/TII.2019.2929228., pp. pp. 77-86, Jan. 2020.

[11]. G. Y. a. N. Ikizler-Cinbis, "Unsupervised Video Summarization with Independently Recurrent Neural Networks," 2019 27th Signal Processing and Communications Applications Conference (SIU), Sivas, Turkey,, no. doi: 10.1109/SIU.2019.8806603, pp. pp. 1-4, 2019.

[12]. F. S. J. C. F. Gers, "Learning to forget: Continual prediction with lstm," Neural computation, vol. 12(10), p. 2451–2471, 2020.

[13]. S. S. J. Hochreiter, "Long short-term memory," Neural computation, vol. 9(8), p. 1735–1780, 2019.

[14]. A. N. NLP, "What is NLP Text Summarization: Benefits & Use Cases," Accern NoCode NLP, 28 September 2022. [Online]. Available: https://accern.com/blog/nlp-text-summarization/. [Accessed 13 April 2023].

[15]. N. B. A. G. Wajdi Homaid Alquliti, "Convolutional Neural Network based for Automatic Text Summarization," International Journal of Advanced Computer Science and Applications(IJACSA), vol. Volume 10, no. Issue 4, 2019., 2019.

Cite this article as :

Prof. Kapil Hande, Hrushi Karlekar, Pranit Yeole, Aditya Likhar, Himanshu Rangari, "NLP based Video Summarisation using Machine Learning", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 10 Issue 2, pp. 456-461, March-April 2023. Available at doi : https://doi.org/10.32628/IJSRSET2310265
Journal URL : https://ijsrset.com/IJSRSET2310265