Empowering Video Summarization Through NLP Algorithms

Ramya R[1], Saravana Balaji K[2], Ragul A S[3], Kannan T[4]
[1, 2, 3, 4] Kamaraj College of Engineering and Technology

Abstract- *Video summary is the process of reducing long videos into shorter versions while maintaining essential details and characteristics, so that viewers can more easily and quickly understand the main points of the video. Video summarization presents several obstacles, including processing videos without subtitles and summarizing sound-only videos. The system addresses these issues by utilizing algorithms such as Edmundson, LexRank, Latent Semantic Analysis (LSA), and TextRank. The system allows summaries to be produced even when textual annotations or subtitles are missing. This is made possible by audio recognition methods that produce captions from videos that only have sound. This method allows users to choose preferred summary lengths, which improves user happiness and usability. It also provides thorough coverage of video content and gives a great degree of flexibility. In addition, the technique offers a thorough explanation of the steps taken to produce the final summary film, including video clip along with retrieved subtitles, guaranteeing coherence and relevancy in the condensed material. The method demonstrates the revolutionary potential of natural language processing (NLP) in video summarizing, opening new avenues for multimedia content analysis research and enhancing the productivity and efficacy of video summarization assignments. LexRank performed the best out of all the algorithms used, exhibiting unmatched accuracy and productivity in producing precise video summaries. Its success highlights how important NLP approaches were in changing the video summarizing environment and opening doors for more creative applications and cutting-edge methods.*

Keywords- Video Summarization, Natural Language Processing, Speech Recognition, Subtitle Generation, Automatic Summarization, Duration Fitting.

I. INTRODUCTION

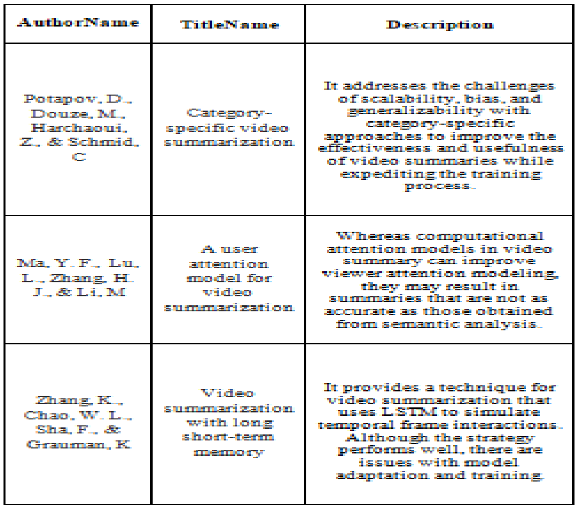
The need for effective video summarizing approaches has been highlighted by the increase in internet video material. The computational complexity of traditional methods makes creative solutions necessary. In order to address the problems associated with video summarization, this research presents a novel method that combines deep learning and Natural Language Processing (NLP).

Video summarization has benefited from deep learning, especially from 2D Convolutional Neural Networks (CNNs), which automatically extract important information from video sequences. Pattern recognition and feature extraction continue to provide difficulties, nevertheless. This research provides a comprehensive framework to improve the effectiveness and efficiency of video summarization by utilizing NLP in conjunction with deep learning.

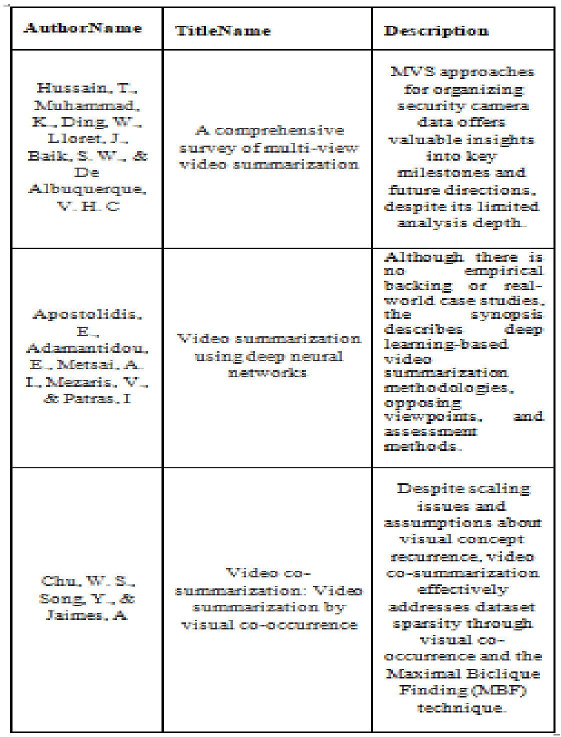
Setting the stage for further exploration of the combination of NLP and deep learning approaches, this introduction promises to transform video summarization in the digital age. The task of creating subtitles for videos that don't have them is solved by using Facebook's speech recognition technology, WIT.AI. With the use of audio input, this utility effectively makes subtitles. Benefits of using speech recognition technology for subtitle production are explored, along with insights into the deployment process.

The process of automatically extracting relevant information from the generated subtitles through NLP-based summarization algorithms is described. In particular, a variety of summary methods are used, including Text Rank, Lex Rank, Luhn's, Edmundson, and Latent Semantic Analysis. The advantages and disadvantages of each technique for summarizing video content are examined.

II. DENTIFY, RESEARCH AND COLLECT IDEA



AuthorName	TitleName	Description
Potapov, D., Douze, M., Harchaoui, Z., & Schmid, C	Category-specific video summarization	It addresses the challenges of scalability, bias, and generalizability with category-specific approaches to improve the effectiveness and usefulness of video summaries while expediting the training process.
Ma, Y. F., Lu, L., Zhang, H. J., & Li, M	A user attention model for video summarization	Whereas computational attention models in video summary can improve viewer attention modeling, they may result in summaries that are not as accurate as those obtained from semantic analysis.
Zhang, K., Chao, W. L., Sha, F., & Grauman, K	Video summarization with long short-term memory	It provides a technique for video summarization that uses LSTM to simulate temporal frame interactions. Although the strategy performs well, there are issues with model adaptation and training.



AuthorName	TitleName	Description
Hussain, T., Muhammad, K., Ding, W., Lloret, J., Baik, S. W., & De Albuquerque, V. H. C	A comprehensive survey of multi-view video summarization	MVS approaches for organizing security camera data offers valuable insights into key milestones and future directions, despite its limited analysis depth.
Apostolidis, E., Adamantidou, E., Metsai, A. I., Mezaris, V., & Patras, I	Video summarization using deep neural networks	Although there is no empirical backing or real-world case studies, the synopsis describes deep learning-based video summarization methodologies, opposing viewpoints, and assessment methods.
Chu, W. S., Song, Y., & Jaimes, A	Video co-summarization: Video summarization by visual co-occurrence	Despite scaling issues and assumptions about visual concept recurrence, video co-summarization effectively addresses dataset sparsity through visual co-occurrence and the Maximal Biclique Finding (MBF) technique.

III. STUDIES AND FINDINGS

The research started with a thorough investigation of the difficulties associated with video summary, identifying major obstacles like the requirement to process videos without subtitles and the work of summarizing content from sound-only videos. A variety of strategies were used to overcome these challenges, including a group of algorithms that included Edmundson, LexRank, Latent Semantic Analysis (LSA), and TextRank. One noteworthy accomplishment was the creation of methods for overcoming the lack of textual annotations or subtitles by using audio recognition techniques to produce captions for films that don't have any textual clues.

Additionally, in order to maximize usability and pleasure, the project concentrated on improving the user experience by putting in place a flexible system that enables users to customize summary lengths to their preferences. A crucial component is providing thorough coverage of the video content while retaining a high level of adaptability, with an emphasis on producing summaries that are both logical and pertinent. The initiative aims to strengthen the coherence and usefulness of the created summaries by offering comprehensive explanations of the process, along with video segments accompanied by retrieved subtitles.

The project also shown the revolutionary potential of natural language processing (NLP) to transform the field of video summarization and pave the way for future research in multimedia content analysis. As a result of a thorough performance review, LexRank was determined to be the best algorithm. Its unmatched productivity and accuracy in creating accurate video summaries highlighted the project's major contribution to the field's advancement.

By giving users, the option to alter summary lengths, the project improved the usability of video summarization. Accuracy was also raised by audio recognition techniques, which extract extra information from sound-only videos. These results highlight the significance of user-driven customization and multimodal integration for efficient summarization.

IV. PROPOSED SOLUTIONS

Creating subtitles for videos that don't have any is a task that Facebook's speech recognition tool, WIT.AI, efficiently handles. This tool effectively creates subtitles by using audio input, which enhances accessibility and comprehension of video content. Along with details on the deployment procedure, the advantages of using speech recognition technology for subtitle production are examined. Additionally, the study explores the automatic extraction of pertinent data using NLP-based summarizing techniques from the generated subtitles. A variety of summary techniques, including as Text Rank, Lex Rank, Luhn's, Edmundson, and Latent Semantic Analysis, are used and assessed for how well they summarize video information, taking into account both the benefits and drawbacks of each method.

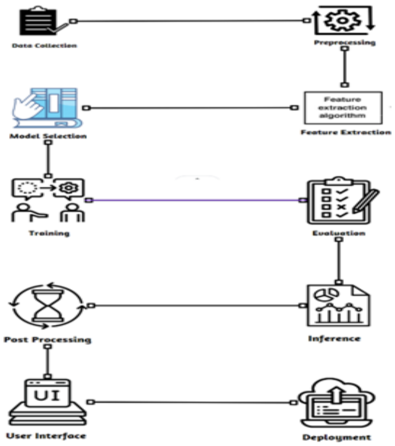


Figure 1: System Methodology

The process of video summarizing entails gathering several video clips and adding subtitles to improve the visual content. Preprocessing operations are carried out, such as text normalization and format harmonization. Important segments of video footage are identified using feature extraction techniques, such as Natural Language Processing (NLP) algorithms like LexRank and Luhn's algorithm. These segments are then transformed into feature vectors for analysis. Using user-specified settings and extracted information, a summary method based on natural language processing generates succinct video summaries. Incorporating subtitles facilitates comprehension, while user customisation choices improve the summary experience. For uniformity and relevancy, video segments are carefully combined and edited using extracted subtitles to produce the final summary film. The system has an easy-to-use user interface that lets users enter movies, adjust preferences, and check summary results. The process of deployment entails assessment and validation in order to determine effectiveness, utility, and dependability. This validates the practical application and promotes future advancements in multimedia content analysis.

TextRank Algorithm:

TextRank is a video summarizing algorithm that works similarly to Google's PageRank for web pages. Sentences from video subtitles are given relevance rankings, resulting in a graph with nodes representing sentences and edges reflecting semantic similarity. Key sentences for the summary are identified by TextRank using iterative computations. It uses timestamps to make sure the summary makes sense chronologically by matching specific lines to relevant parts of the video. In the end, TextRank automates the crucial information extraction process from subtitles, making it possible to automatically create succinct and perceptive video summaries without the need for human interaction.

LexRank Algorithm:

LexRank algorithm is used to summarize videos. In order for LexRank to function, sentences are represented as vectors, and then nodes are sentences, and edges are sentences' similarity. LexRank uses a version of Google's PageRank algorithm to order sentences according to how important they are in the document. To ensure that the summary accurately conveys the major ideas and important points mentioned in the subtitles, the phrases with the highest ranking are chosen to provide a brief synopsis of the video material. All things considered, LexRank makes it easier to automatically extract important information from video subtitles, which helps to produce logical and educational video summaries.

Luhn'sRank Algorithm:

For video summarizing, Luhn's algorithm is used, with an emphasis on finding and picking sentences from the subtitles that include important keywords or phrases. Sentences are rated according to their keyword density, and the highest-scoring ones are chosen to be included in the summary. This technique makes it easier for significant information to be automatically extracted, which helps to produce succinct and educational video summaries.

Edmundson Algorithm:

For video summarizing, the Edmundson algorithm is applied, with a focus on selecting sentences that contain salient features such as keywords and phrases from the subtitles. After classifying sentences according to cue frequency, presence, and perhaps other criteria, the highest-scoring sentences are chosen for the summary. This technique makes it easier for important information to be automatically extracted, which helps create clear and educational video summaries.

Latent Semantic Analysis:

For video summarizing, the Latent Semantic Analysis (LSA) technique is used. LSA use dimensionality reduction techniques such as Singular Value Decomposition (SVD) to transform video subtitles into a conceptual space, from which it extracts phrases and analyzes their semantic links. Based on the vectors in this space, it determines how similar sentences are to one another and ranks them accordingly. The most important details from the video content are condensed into a brief summary using the sentences that have the highest similarity scores. In general, LSA makes it easier for significant content to be automatically extracted from subtitles, which improves the effectiveness of video summarization.

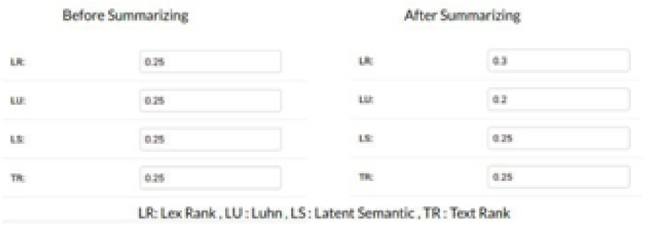
Fitting Summarization to User-Defined Duration:

Each sentence in the video subtitles is given a ranking. In order to match the user-provided duration, the total length of the video and the number of subtitles are used to calculate the average duration per subtitle. The summarizing approach, which modifies the amount of phrases added or subtracted to satisfy the time constraint, is used to select the top-ranked subtitles. With this adaptive method, the summary video adjusts to the user's desired watching time, preserving the most relevant content.

V. RESULT ANALYSIS & CONCLUSION

The video summary system's analysis shows how well it works to solve important issues like processing videos without subtitles and summarizing videos with just sound. The system's adaptability and versatility are demonstrated by the way it uses algorithms like as Edmundson, LexRank, Latent Semantic Analysis (LSA), and TextRank to provide summaries even in the absence of textual annotations or subtitles. The incorporation of audio recognition techniques augments usability by producing captions for audio-only films, permitting users to designate custom summary durations for a customized encounter. Furthermore, the technology guarantees thorough coverage of video content while preserving flexibility, giving consumers accurate and pertinent summaries.

Lastly, we evaluated how well Edmundson, Luhn, LexRank, and LSA performed as video summarizing techniques. While all the methods showed promise for reducing video material, our strategy based on natural language processing (NLP) and specifically utilizing LexRank outperformed the others. Using NLP approaches for thorough content analysis in multimedia applications resulted in considerable improvements over traditional methods for content condensing and information retrieval. LexRank stood out as the most accurate and effective approach, proving that it might revolutionize the video summarizing market.



Before Summarizing		After Summarizing	
LR:	0.25	LR:	0.3
LU:	0.25	LU:	0.2
LS:	0.25	LS:	0.25
TR:	0.25	TR:	0.25

LR: Lex Rank , LU : Luhn , LS : Latent Semantic , TR : Text Rank

Figure 2: Weights before and after summarization of a video.

It is clear from the figure that Luhn's weight was reduced while Lex's weight grew. Thus, Lex did better and Luhn did the worst in this video. Therefore, there is a one unit increase in Lex Rank's weight and a one unit decrease in Luhn's.

The graph of subtitle text with it being relevant in the summarized video for the algorithms are described in Fig 3, Fig 4, Fig 5, Fig 6, Fig 7.

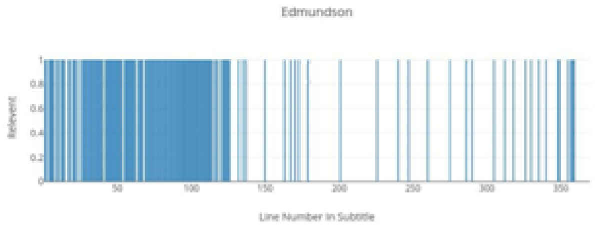


Fig 3: Edumundson Algorithm

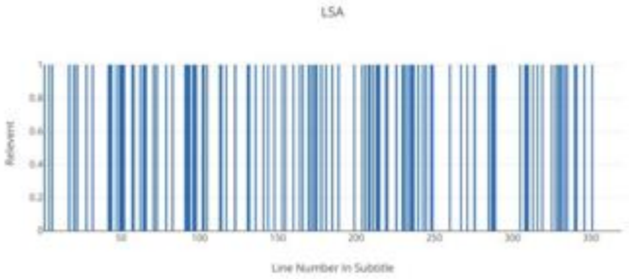


Fig 4: LSA Algorithm

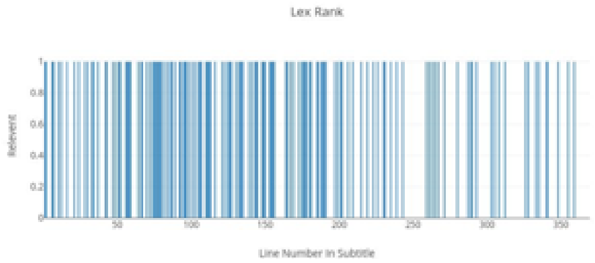


Fig 5: Lex Rank Algorithm

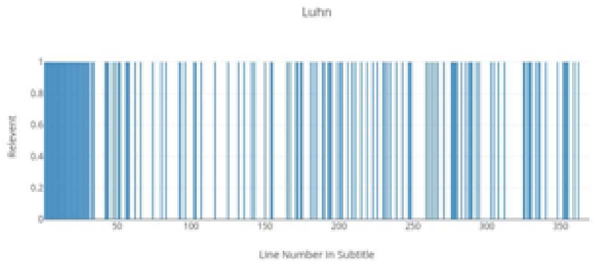


Fig 6: Luhn Algorithm

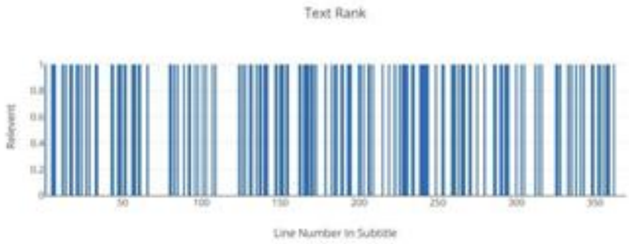


Fig 7: Text Rank Algorithm

REFERENCES

[1] Potapov, D., Douze, M., Harchaoui, Z., & Schmid, C. (2014). Category-specific video summarization. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VI 13 (pp. 540-555). Springer International Publishing.

[2] Ma, Y. F., Lu, L., Zhang, H. J., & Li, M. (2002, December). A user attention model for video summarization. In Proceedings of the tenth ACM international conference on Multimedia (pp. 533-542).

[3] Zhang, K., Chao, W. L., Sha, F., & Grauman, K. (2016). Video summarization with long short-term memory. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII 14 (pp. 766-782). Springer International Publishing.

[4] Hussain, T., Muhammad, K., Ding, W., Lloret, J., Baik, S. W., & De Albuquerque, V. H. C. (2021). A comprehensive survey of multi-view video summarization. Pattern Recognition, 109, 107567.

[5] Apostolidis, E., Adamantidou, E., Metsai, A. I., Mezaris, V., & Patras, I. (2021). Video summarization using deep neural networks: A survey. Proceedings of the IEEE, 109(11), 1838-1863.

[6] 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).

[7] Gong, Y., & Liu, X. (2000, June). Video summarization using singular value decomposition. In Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662) (Vol. 2, pp. 174-180). IEEE.

[8] Meena, P., Kumar, H., & Yadav, S. K. (2023). A review on video summarization techniques. Engineering Applications of Artificial Intelligence, 118, 105667.

[9] Li, X., Zhao, B., & Lu, X. (2017). A general framework for edited video and raw video summarization. IEEE Transactions on Image Processing, 26(8), 3652-3664.

[10] DeMenthon, D., Kobla, V., & Doermann, D. (1998, September). Video summarization by curve simplification. In Proceedings of the sixth ACM international conference on Multimedia (pp. 211-218).

[11] Ma, Y. F., Hua, X. S., Lu, L., & Zhang, H. J. (2005). A generic framework of user attention model and its application in video summarization. IEEE transactions on multimedia, 7(5), 907-919.

[12] Gong, B., Chao, W. L., Grauman, K., & Sha, F. (2014). Diverse sequential subset selection for supervised video summarization. Advances in neural information processing systems, 27.

[13] Otani, M., Nakashima, Y., Rahtu, E., Heikkilä, J., & Yokoya, N. (2017). Video summarization using deep semantic features. In Computer Vision–ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part V 13 (pp. 361-377). Springer International Publishing.

[14] Ngo, C. W., Ma, Y. F., & Zhang, H. J. (2003, October). Automatic video summarization by graph modeling. In Proceedings Ninth IEEE International Conference on Computer Vision (pp. 104-109). IEEE.

[15] Rochan, M., Ye, L., & Wang, Y. (2018). Video summarization using fully convolutional sequence networks. In Proceedings of the European conference on computer vision (ECCV) (pp. 347-363).

[16] Zhao, B., Li, X., & Lu, X. (2017, October). Hierarchical recurrent neural network for video summarization. In Proceedings of the 25th ACM international conference on Multimedia (pp. 863-871).

[17] Mei, S., Guan, G., Wang, Z., Wan, S., He, M., & Feng, D. D. (2015). Video summarization via minimum sparse reconstruction. Pattern Recognition, 48(2), 522-533.

[18] Wei, H., Ni, B., Yan, Y., Yu, H., Yang, X., & Yao, C. (2018, April). Video summarization via semantic attended networks. In Proceedings of the AAAI conference on artificial intelligence (Vol. 32, No. 1).

[19] Mahasseni, B., Lam, M., & Todorovic, S. (2017). Unsupervised video summarization with adversarial lstm networks. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (pp. 202-211).