

UE20CS390A - Capstone Project Review #3

Quick Glance

Project Title: Sports Video Summarization

Project ID: 43

Project Guide: Prof Kokila P

Katta Lohith Krishna Kumar-PES2UG20CS158

Kanikaram Roopasreesai-PES2UG20CS154

Koyya Deekshitha-PES2UG20CS167

Maneesh Nand Reddy Kondaiahgari-PES2UG20CS186

Our Team



Roopasreesai K



Lohith K



Maneesh K



Deekshitha K

Outline

- Problem Statement
- Summary of Literature Survey
- Suggestions from Review-2
- Design Approach
- Proposed Architecture
- Design Description
- Technologies Used
- Project Progress
- References

Problem Statement

Sports video summarization is an emerging field that aims to condense long and complex sports videos into a short, concise summary. This technology can be useful for users who want to quickly catch up on the key moments from a game or event and for coaches and analysts who need to review footage to identify areas for improvement.

Summary of Literature Survey in Review 2

- The paper's main objective is to identify and profile different types of actions in sports videos. The proposed method consists of LSTM and the attention layer takes the input of events of action and predicts the likelihood of this action to be selected in the summary.
- This paper aims to develop an approach to build a highlight recognition framework, The proposed method comprises two models. A 3D CNN for feature extraction and an LSTM network to model temporal evolution generates a set of highlights at different timestamps and predicts the rank to highlights All high-ranked highlights were concatenated to form a summarized video.
- In this paper the authors proposed an approach for summarizing videos based on a combination of convolutional neural networks (CNNs), sparse autoencoders, and random forest classifiers. The proposed approach involves extracting key frames from the input video using multiple CNNs, followed by feature extraction using sparse autoencoders. The resulting features are then fed into a random forest classifier to identify the most representative frames for the video summary. Using several datasets and showing that it outperforms several state-of-the-art methods in terms of accuracy and efficiency.

Suggestions from Review 2

- To design the outline of the architecture.
- To explore better Feature Extractor.
- To research on datasets.

Design Approach

Existing Approach and results obtained:

- In the Existing System highlights are generated by a pre-trained AI-ML model which generates highlights. It used visual information and audio information and calculates a threshold and summarizes highlights based on that Information.
- This model fails as the crowd shouts even when the entry of a player and some defense sequence crowd observes silently. So this model won't recognize the second event as a highlight.

Proposed Methodology / Approach

Is there a need for changing the approach?

- Yes, there is a need to change this approach as when the highlights are required in a short duration, the existing system will fail to give importance to particular frame.

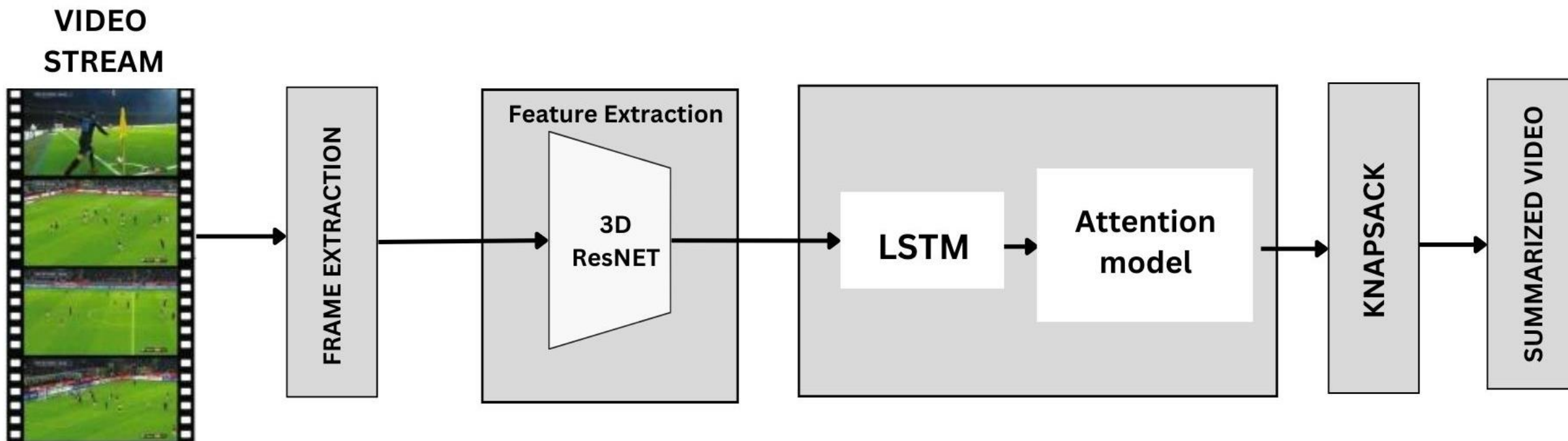
Design Details

- **Novelty:** The Novelty of this project is to automate the process of summarization and to generate highlights on basis of how important a particular highlight is. The model gives a weight value to the frames from which we can say how much is a particular highlight important. The weight values help in selecting the duration of the output summary as required. The Novelty makes this model unique from the other existing model approaches.
- **Reusability:** With some tweaks to the training data and code, this model can be used again to produce highlights for other games or sports.
- **Performance:** The Deep learning technique considers both spatial and temporal features and the machine learning technology gives weight to each frame efficiently. So, the model is expected to achieve good performance.
- **Legacy to Modernization:** This model generates automated highlights whereas the old systems are done manually.

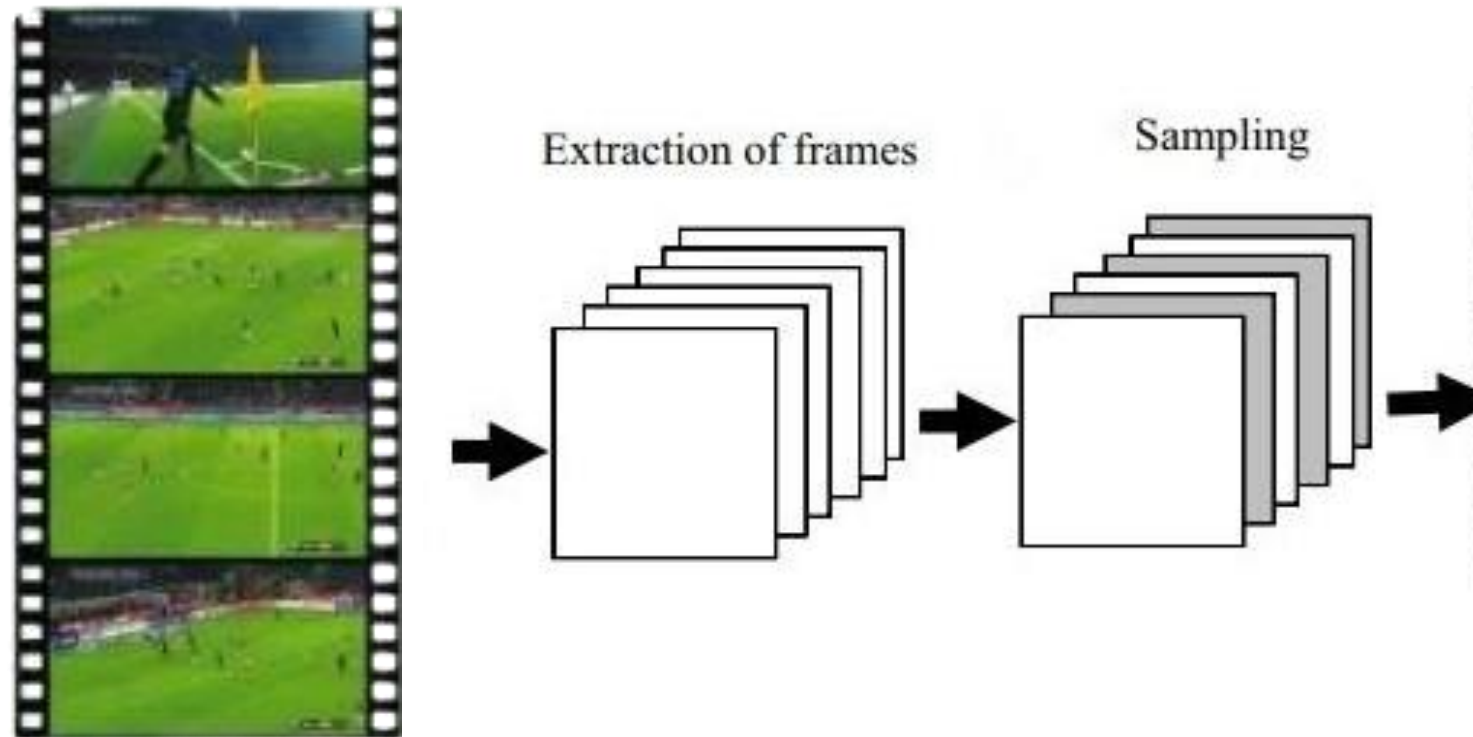
Design Details

- **Interoperability:** As the Model is being executed on Google Collab. The Model can run on other Operating Systems if Google Collab is available to be used on that particular Operating System.
- **Reliability:** This Model will be able to produce the expected result.

Architecture Diagram



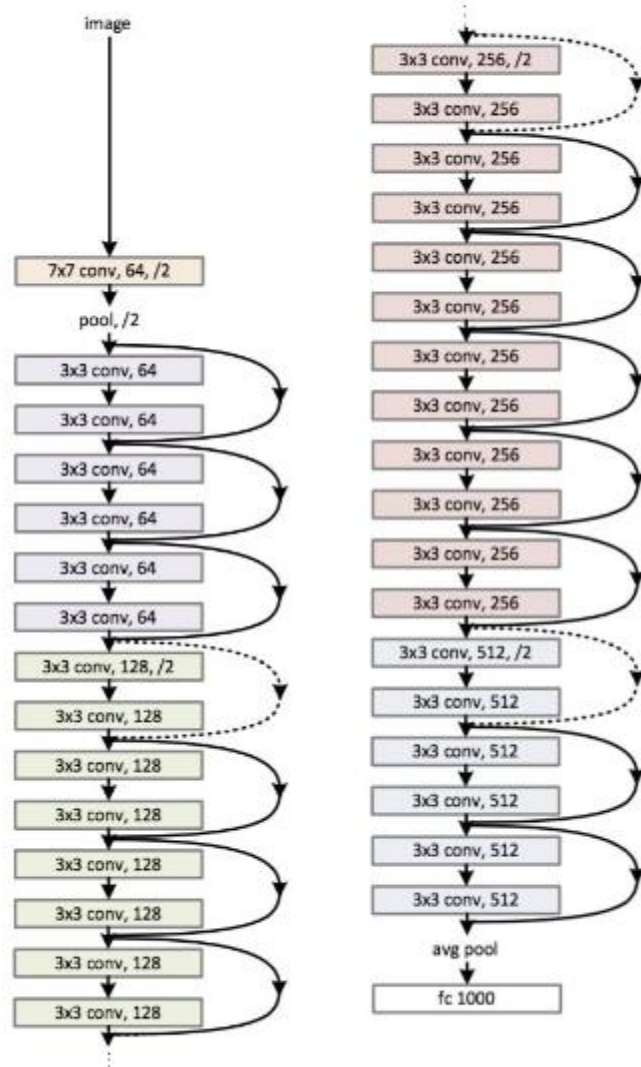
Video Segmentation



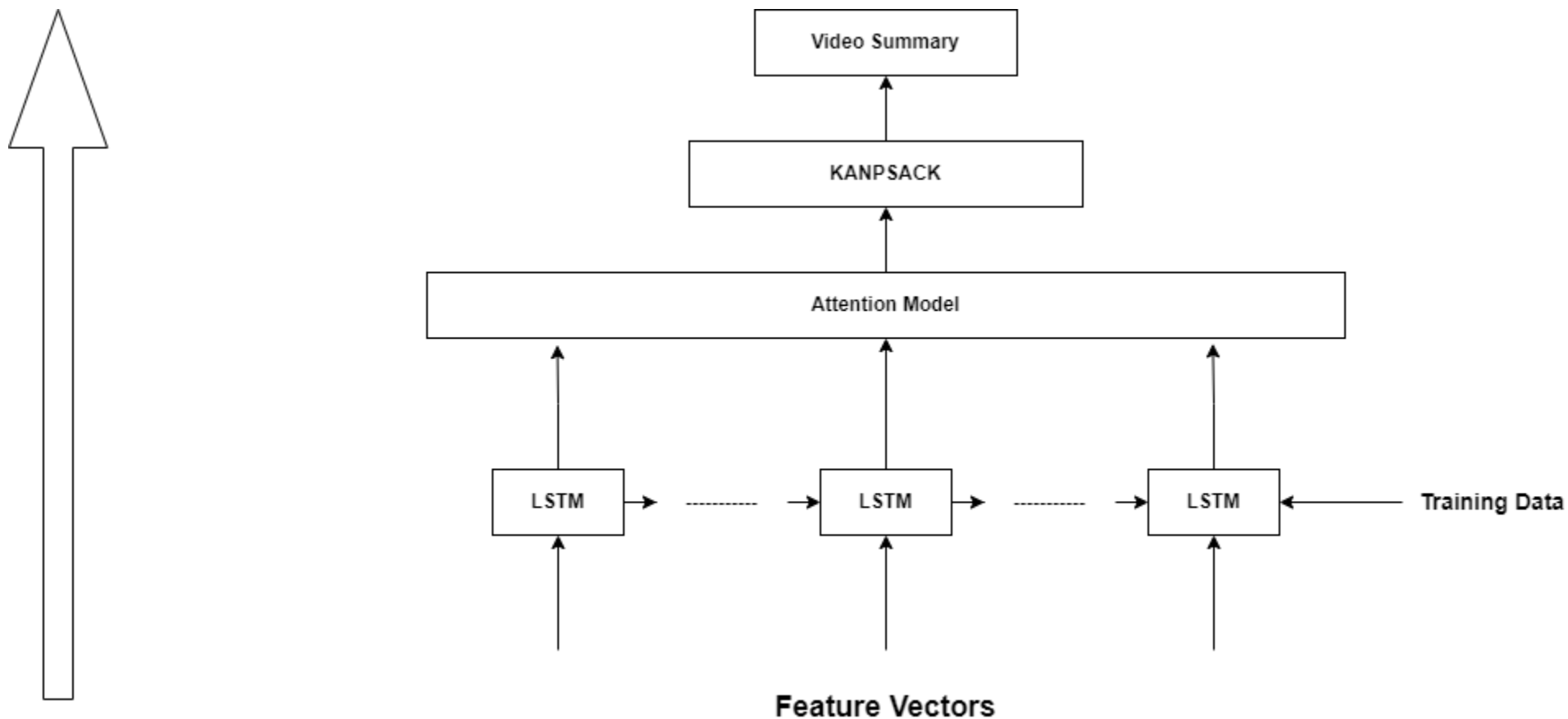
3D ResNet

- 3D ResNet is a CNN model used to extract features from frames(images).
- It gives the output as a feature vector.
- Skip Connections are used to overcome the vanishing gradient problem.
- The Output of every convolution layer is typically a feature map that is flattened(compressed) to 1D space which is named a Feature Vectors.

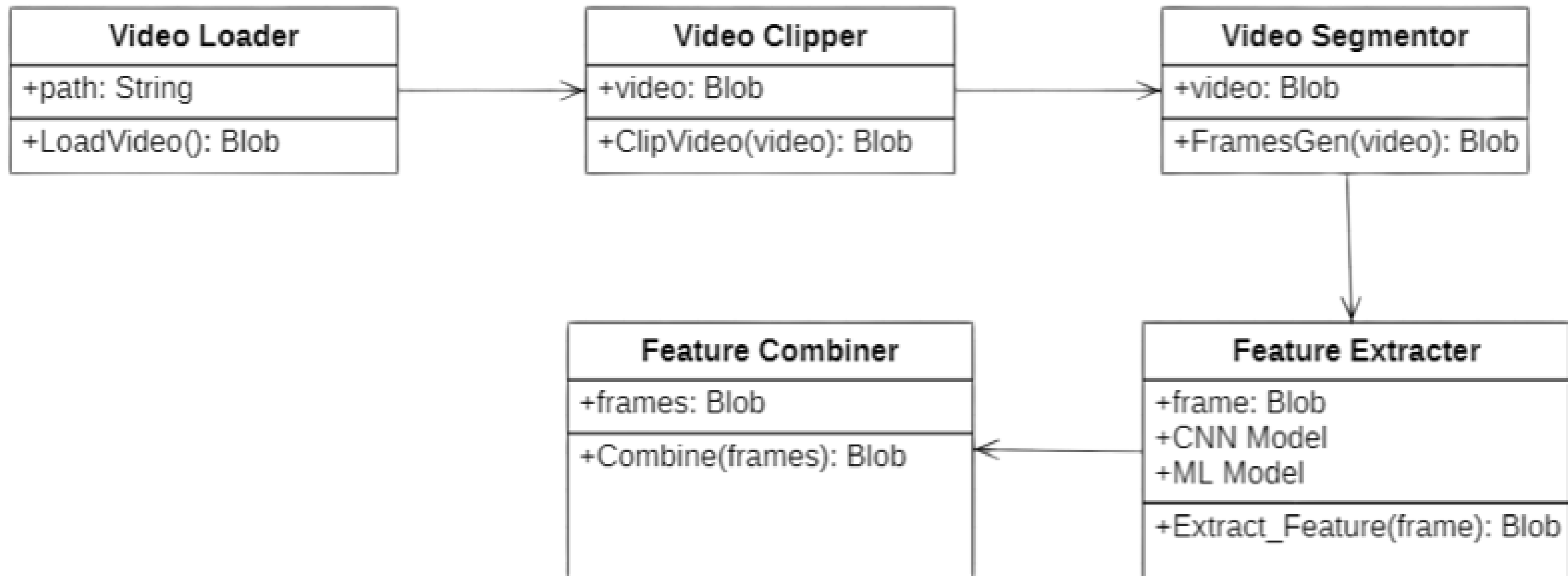
34-layer residual



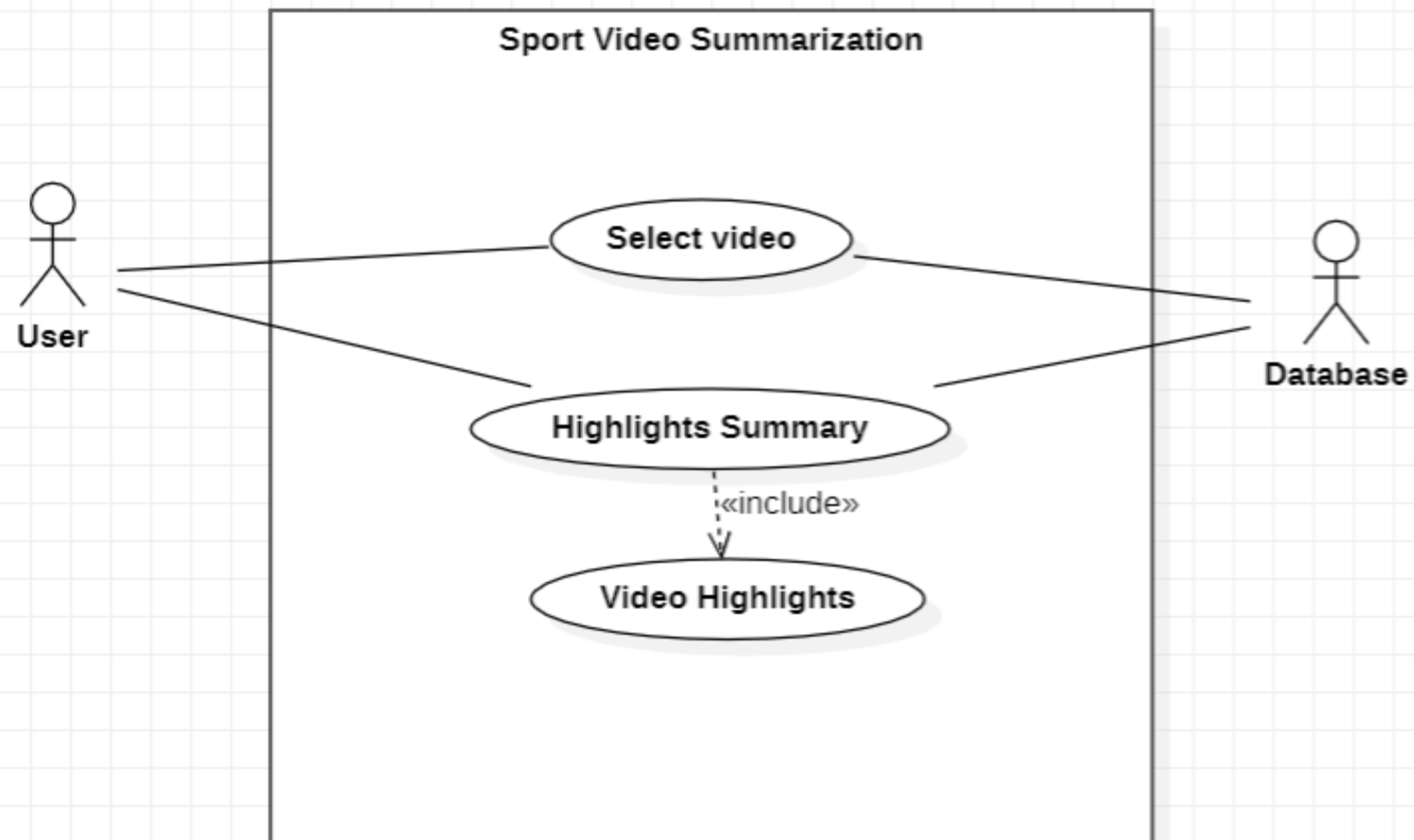
LSTM based Attention Layer



Master Class Diagram



Use-Case Diagram



Technologies Used

- Key Frame extraction is done using Multi-CNN Models.
- Programming Languages like Python.
- Image feature extraction is done through pre-trained models like 3D ResNet.
- Google Collab is used for execution purposes.

Project Progress

- Literature Survey.
- We have acquired a dataset and got access to use that dataset. Now we are analyzing the dataset if it works for our Model or not.

GANTT CHART

JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC

PROBLE STATEMENT

1 JAN - 25 MAR

LITERATURE
SURVEY

26 JAN - 30 NOV

HIGH LEVEL
ARCHITECTURE

26 JAN - 20 APRIL

HIGH LEVEL
DESIGN

18/03-30/04

DRAFT
IMPLEMENTATION PLAN

1 MAR - 25 MAY

CAPSTONE PHASE 2

16 APR - 30 SEP

DATA
PREPROCESSING

01/08 - 15/09

IMPLEMENTATI
ON OF MODELS

01/09 - 01/11

TESTING

1 SEPT - 15 NOV

TESTING AND
DEPLOYMENT

10/1 - 15/12

Summarization

- This Model will generate video highlights while considering the importance of each frame i.e., by giving it a bias value to each frame and extracting the highlights based on those bias values.
- This Model is expected to return concise highlights which are much related to the key moments in the game.

References

- [1] Nair, M.S., Mohan, J. Static video summarization using multi-CNN with sparse autoencoder and random forest classifier. *SIViP* **15**, 735–742 (2021). <https://doi.org/10.1007/s11760-020-01791-4>
- [2] R. Agyeman, R. Muhammad and G. S. Choi, "Soccer Video Summarization Using Deep Learning," 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), San Jose, CA, USA, 2019, pp. 270-273, doi: 10.1109/MIPR.2019.00055.
- [3] M. Sanabria, F. Precioso and T. Menguy, "Profiling Actions for Sport Video Summarization: An attention signal analysis," 2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP), Tampere, Finland, 2020, pp. 1-6, doi: 10.1109/MMSP48831.2020.9287062.
- [4] M. E. Anjum, S. F. Ali, M. T. Hassan and M. Adnan, "Video summarization: Sports highlights generation," INMIC, Lahore, Pakistan, 2013, pp. 142-147, doi:10.1109/INMIC.2013.6731340.
- [5] J. Han, Y. Zhang and Z. Sun, "Key Frame Extraction Based on Sports Statistical Characteristics," 2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), Dalian, China, 2022, pp. 908-911, doi: 10.1109/IPEC54454.2022.9777583.
- [6] K. Davila, F. Xu, S. Setlur and V. Govindaraju, "FCN-LectureNet: Extractive Summarization of Whiteboard and Chalkboard Lecture Videos," in *IEEE Access*, vol. 9, pp. 104469-104484, 2021, doi: 10.1109/ACCESS.2021.3099427.
- [7] H. Boukadida, S. -A. Berrani and P. Gros, "Automatically Creating Adaptive Video Summaries Using Constraint Satisfaction Programming: Application to Sport Content," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 4, pp. 920-934, April 2017, doi: 10.1109/TCSVT.2015.2513678.

References

- [8] M. Merler et al., "Automatic Curation of Sports Highlights Using Multimodal Excitement Features," in IEEE Transactions on Multimedia, vol. 21, no. 5, pp. 1147-1160, May 2019, doi: 10.1109/TMM.2018.2876046.
- [9] R. Agyeman, R. Muhammad and G. S. Choi, "Soccer Video Summarization Using Deep Learning," 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), San Jose, CA, USA, 2019, pp. 270-273, doi: 10.1109/MIPR.2019.00055.
- [10] G. Mujtaba, A. Malik and E. -S. Ryu, "LTC-SUM: Lightweight Client-Driven Personalized Video Summarization Framework Using 2D CNN," in IEEE Access, vol. 10, pp. 103041-103055, 2022, doi: 10.1109/ACCESS.2022.3209275.
- [11] D. Xu, H. Fu, L. Wu, M. Jian, D. Wang and X. Liu, "Group Activity Recognition by Using Effective Multiple Modality Relation Representation With Temporal-Spatial Attention," in IEEE Access, vol. 8, pp. 65689-65698, 2020, doi: 10.1109/ACCESS.2020.2979742.
- [12] C. Lin and Y. Chen, "Sports Video Summarization with Limited Labeling Datasets Based on 3D Neural Networks," 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Taipei, Taiwan, 2019, pp. 1-6, doi: 10.1109/AVSS.2019.8909872.
- [13] M. Sanabria, F. Precioso and T. Menguy, "Profiling Actions for Sport Video Summarization: An attention signal analysis," 2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP), Tampere, Finland, 2020, pp. 1-6, doi: 10.1109/MMSP48831.2020.9287062.
- [14] M. Tavassolipour, M. Karimian and S. Kasaei, "Event Detection and Summarization in Soccer Videos Using Bayesian Network and Copula," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 24, no. 2, pp. 291-304, Feb. 2014, doi: 10.1109/TCSVT.2013.2243640

References

- [15] A. Tejero-de-Pablos, Y. Nakashima, T. Sato, N. Yokoya, M. Linna and E. Rahtu, "Summarization of User-Generated Sports Video by Using Deep Action Recognition Features," in IEEE Transactions on Multimedia, vol. 20, no. 8, pp. 2000-2011, Aug. 2018, doi: 10.1109/TMM.2018.2794265
- [16] Godi, M., Rota, P., Setti, F. (2017). Indirect Match Highlights Detection with Deep Convolutional Neural Networks. In: Battiato, S., Farinella, G., Leo, M., Gallo, G. (eds) New Trends in Image Analysis and Processing – ICIAP 2017. ICIAP 2017. Lecture Notes in Computer Science(), vol 10590. Springer, Cham
- [17] C. Yan, X. Li and G. Li, "A New Action Recognition Framework for Video Highlights Summarization in Sporting Events," 2021 16th International Conference on Computer Science & Education (ICCSE), Lancaster, United Kingdom, 2021, pp. 653-666, doi: 10.1109/ICCSE51940.2021.9569708
- [18] M. Z. Khan, S. Jabeen, S. ul Hassan, M. A. Hassan and M. U. G. Khan, "Video Summarization using CNN and Bidirectional LSTM by Utilizing Scene Boundary Detection," 2019 International Conference on Applied and Engineering Mathematics (ICAEM), Taxila, Pakistan, 2019, pp. 197-202, doi: 10.1109/ICAEM.2019.8853663
- [19] 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)
- [20] Chianese, A., Miscioscia, R., Moscato, V., Parlato, S., & Picariello, A. (2004, August). A fuzzy approach to video scene detection and its application for soccer matches. In Proceedings of the 4th International Conference on Intelligent Systems Design and Application

References

- [21] F. Chen and C. De Vleeschouwer, "Formulating Team-Sport Video Summarization as a Resource Allocation Problem," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 2, pp. 193-205, Feb. 2011, doi: 10.1109/TCSVT.2011.2106271.
- [22] C. Yan, X. Li and G. Li, "A New Action Recognition Framework for Video Highlights Summarization in Sporting Events," 2021 16th International Conference on Computer Science & Education (ICCSE), Lancaster, United Kingdom, 2021, pp. 653-666, doi: 10.1109/ICCSE51940.2021.9569708.
- [23] P. Kadam et al., "Recent Challenges and Opportunities in Video Summarization With Machine Learning Algorithms," in *IEEE Access*, vol. 10, pp. 122762-122785, 2022, doi: 10.1109/ACCESS.2022.3223379.

Thank You