

PROBLEM STATEMENT

Engaging trailers are crucial for capturing audience attention and building anticipation for upcoming movies, music albums, and other events. Traditionally, creating these trailers involves manually selecting key moments—a process that is both time-consuming and labour-intensive. To address this, an automated **multimodal** approach to trailer generation integrates **audio and video features** to select key moments efficiently. By leveraging audio cues to guide video frame selection, this method streamlines trailer creation, **saving time** while **improving efficiency**. This innovative approach redefines trailer generation by ensuring a balance between creativity and automation.

BACKGROUND

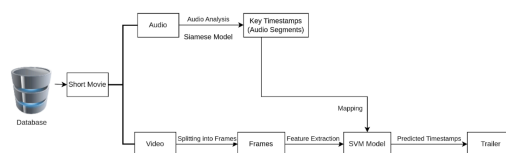
Manual trailer creation is a time-intensive process that requires selecting key scenes to captivate audiences. Automating this task using multimodal data analysis can streamline the process and enhance creativity in film making

DATASET AND FEATURES

The dataset contains 311 short films, and their trailers sourced from platforms like **YouTube**, **Vimeo**, and **ShortOfTheWeek**, known for promoting short films and Director's works. Movies have an average duration of **12 minutes** while the trailer's average duration is **50 seconds**. The dataset includes five columns : movie name, video link, duration, trailer link and trailer duration, with most films in the horror thriller genre. All content is publicly available, of high auditory and visual quality, and age appropriate.

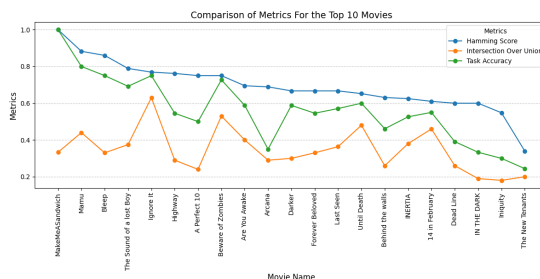
DESIGN APPROACH / METHODS

The short film and trailer are converted to mp4, split into audio and video, and analyzed. Audio Features form triplet data for a **Siamese LSTM** model to predict trailer worthy segments. Video Features from the CLIP model train a **One Class SVM** to refine timestamps obtained from the audio analysis. The final trailer uses only the films original content.



RESULTS AND DISCUSSION

The audio-video processing method highlighted key features in horror films such as screams and unsettling imagery. Three evaluation metrics were used to assess the performance of machine learning models. The **Hamming Score** was **0.6930** indicating over 60% of the predicted segments were trailer worthy. The **Intersection over union** score was **0.3455** showing 30% overlap with ground truth, with room for more relevant segments to be included. The **Task Accuracy** was **0.5625** meaning over 50% of the predicted segments were correctly predicted as a part of the trailer. These metrics demonstrate the effectiveness in key-moment extraction for trailer. The results obtained for the top 10 movies analyzed are shown in the graph below



SUMMARY OF PROJECT OUTCOME

This research presents a novel approach to trailer generation, emphasizing **audio and video modalities** to extract impactful segments, with a focus on horror films. Key auditory features like screams and unsettling atmospheres, along with visual elements such as darkness and fear-filled imagery, are prioritized through an **audio-guided visual model**. Computationally efficient, the method achieves an average Hamming Score of 0.6930, showing that **over 60%** of extracted segments are **highly trailer worthy**. This approach aims to assist short filmmakers in identifying key moments and generating impactful trailers, a goal that has been successfully achieved through this method.

CONCLUSION AND FUTURE WORK

The model demonstrates strong performance, with over 60% of predicted segments being trailer worthy and significant overlap with ground truth segments. While IOU scores were relatively lower due to the variability in trailer content, the results validate the model's effectiveness. Future work includes efforts to enhance the robustness by incorporating advanced features and diversifying the dataset to include a broader range of sub-genres.

REFERENCES

- [1] P. Papalampidi, "Structure-aware narrative summarization from multiple views," era.ed.ac.uk, Jan. 2023, doi: <https://doi.org/10.7488/era/2946>.
- [2] Gan, Bei & Shu, Xiujun & Qiao, Ruizhi & Wu, Haoqian & Chen, Keyu & Li, Hanjun & Ren, Bo. (2023). Collaborative Noisy Label Cleaner: Learning Scene-aware Trailers for Multi-modal Highlight Detection in Movies.
- [3] C. Bretti, P. Mettes, Hendrik Vincent Koops, Daan Odijk, and Nanne van Noord, "Find the Cliffhanger: Multi-modal Trailerness in Soap Operas," Lecture Notes in Computer Science, pp. 199–212, Jan. 2024, doi: https://doi.org/10.1007/978-3-031-53308-2_15.
- [4] P. Mishra, C. Diwan, S. Srinivasa, and G. Srinivasaraghavan, "AI based approach to trailer generation for online educational courses," CSI Transactions on ICT, vol. 11, no. 4, pp. 193–201, Nov. 2023, doi: <https://doi.org/10.1007/s40012-023-00390-1>.

Authors



Nikhil Giridhar
PESUG21CS384



Prajna R
PESUG21CS417



Pranathi Praveen
PESUG21CS428



Shreeja Rajesh
PESUG21CS564



Dr. Surabhi Narayan
Dept of C.S.E