Captivating Video Editing Secrets: Unsupervised Deep Learning and PySceneDetect Analysis of Creative Strategies

Richard Anthony Álvarez

IPHS 484 Senior Seminar (Spring 2023) Prof Elkins and Chun, Kenyon College

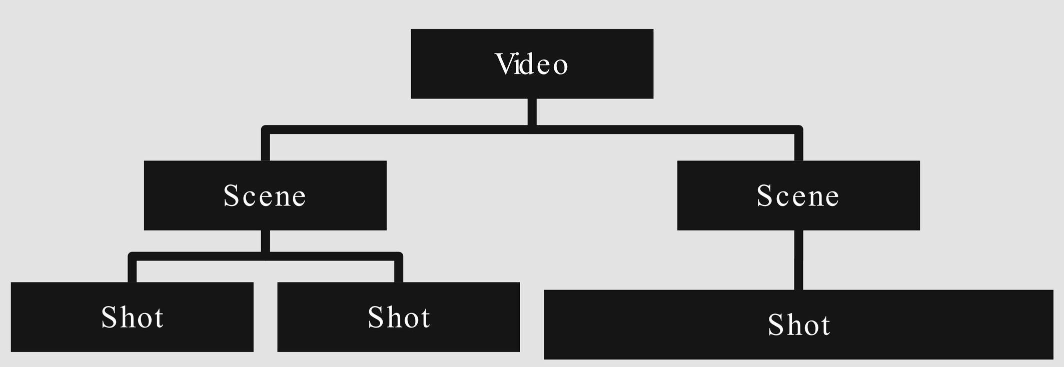
Github Repository

Abstract

What factors explain why some short-format videos go viral and others don't? This research creates an original dataset of YouTube shots across five genres, extracts both textual and video editing features, and uses AutoML in order to identify optimal models to explain which features explain the variance in virality metrics. This project found that short videos have a wide range of editing styles that achieve virality.

Shot Boundary Detection

Shot boundary detection is a crucial process in video analysis, as it identifies changes in shots and allows for efficient video structuring and analysis. PySceneDetect, an open-source tool for detecting scene changes in videos, was used in this project to detect shot boundaries and extract various video statistics for analysis.

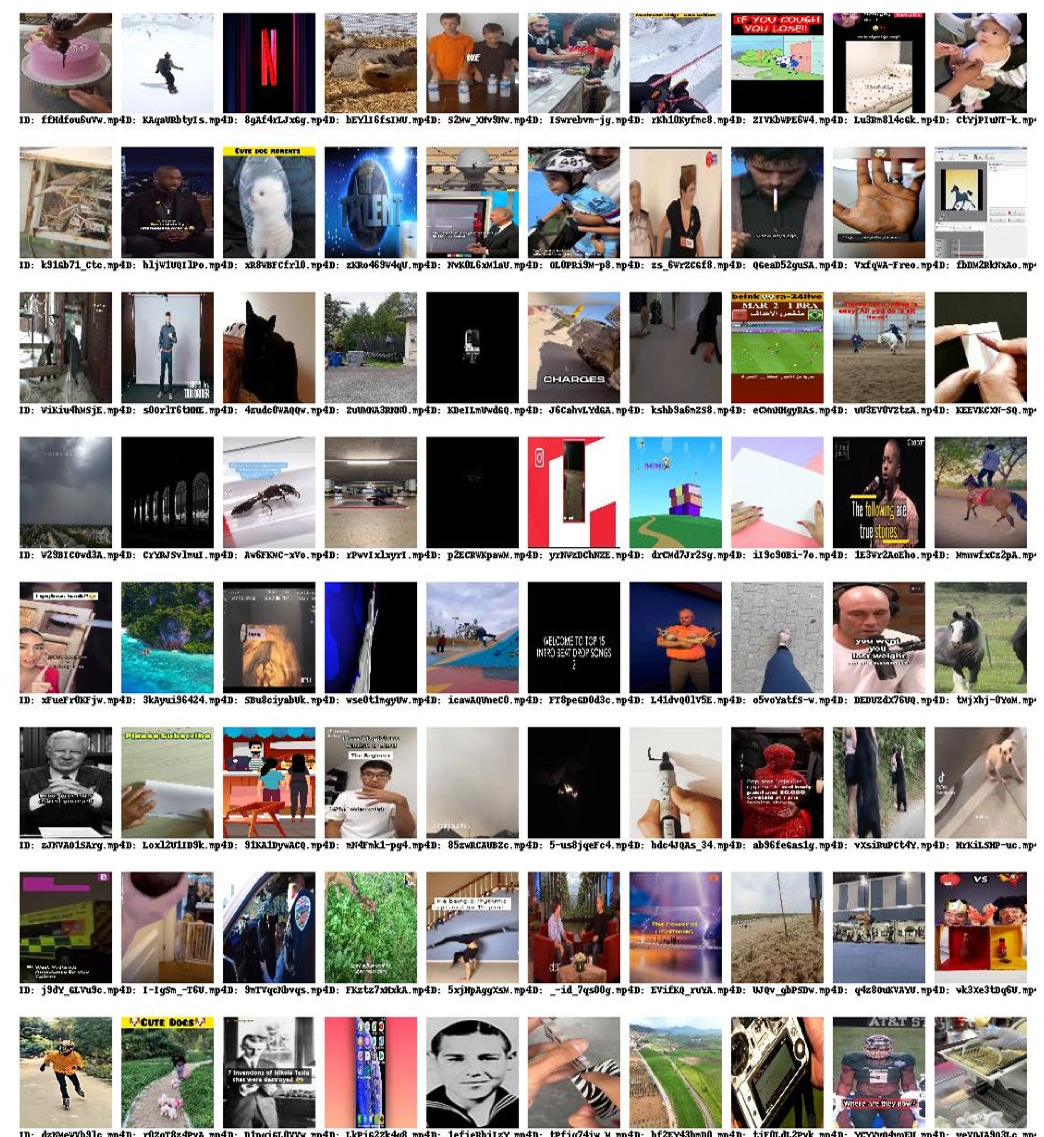


(*Figure 1.* Hierarchy of video)

By analyzing the shot duration statistics, it is possible to identify the most common editing styles employed in short videos and compare them to traditional editing styles. Shot boundary detection and the extracted video statistics can help in identifying the dominant themes and trends prevalent in short videos. By comparing the video statistics of different categories, it is possible to identify patterns and trends in editing style that are unique to specific categories such as music, comedy, education, and travel. Overall, the extracted video statistics and shot boundary detection can provide valuable information for content creators and video producers, informing the creation and optimization of short videos for social media platforms.

Data Collection

In this project, we collected over 6,000 short-format videos from YouTube using the YouTube v3 API and yt-dlp. However, we faced several challenges, including rate limits and file storage issues. To overcome these issues, we implemented a script that tracked the number of requests made and paused when the limit was reached, and used an SSD with fast read and write times for efficient file storage.



(*Figure 2.* Thumbnails from scraped videos)

Training the Model

In this project, we used PyCaret, an open-source machine learning library in Python, to train an unsupervised clustering model on the extracted video features. PyCaret is designed to streamline the machine learning workflow, enabling us to preprocess the data, select the most appropriate clustering algorithm, and tune the hyperparameters of the algorithm for optimal performance.

We experimented with several clustering algorithms, including K-Means, Hierarchical Clustering, and Gaussian Mixture Models. After evaluating the performance of each algorithm, we selected K-Means clustering as the most suitable algorithm for this project.

We further optimized the hyperparameters of the K-Means clustering algorithm, including the number of clusters, the initialization method, and the maximum number of iterations. We used the elbow method and silhouette score analysis to determine the optimal number of clusters, which we found to be 5.

Finally, we used the trained model to assign each video in the dataset to a specific cluster based on its extracted features. The resulting clusters provided insights into the different editing styles and techniques used for successful short-format videos on YouTube.

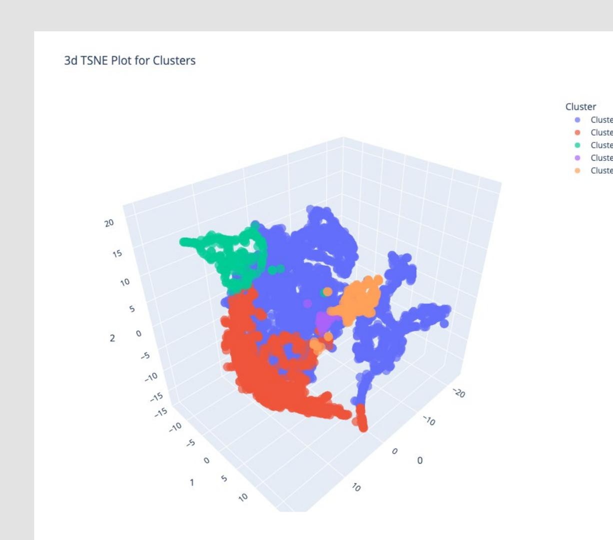
Code available at:
https://github.com/raulduk3/pySceneDetect-video-editing-trends

Results

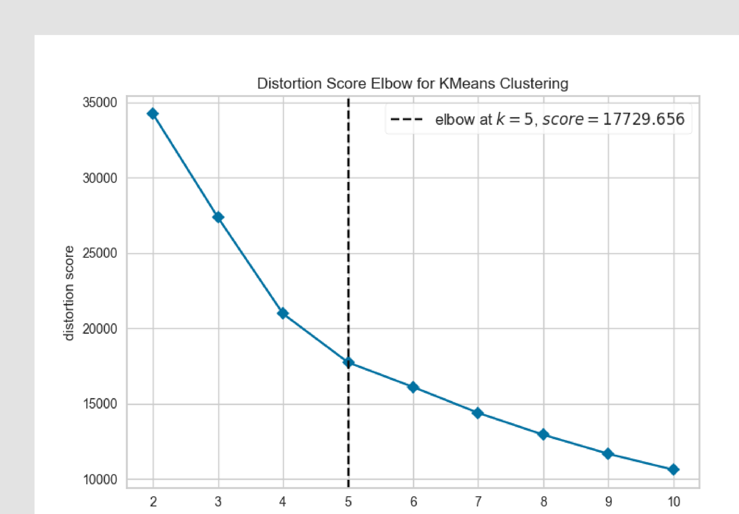
Our analysis of short-format videos using unsupervised clustering techniques has led us to identify five distinct clusters of videos with similar characteristics. We believe that this number of clusters is optimal based on the clustering metrics, which show a peak in the Calinski-Harabasz score at five clusters. This indicates that the clustering algorithm is able to differentiate between the clusters effectively, with high inter-cluster variance and low intra-cluster variance.

Our analysis of the five clusters revealed distinct editing styles and themes that were prevalent in the videos. For example, the first cluster was characterized by fast-paced, energetic videos with high shot variance, while the second cluster featured slower-paced videos with longer shots and more static camera movements. The other clusters exhibited similar differences in shot duration, camera movement, and overall pacing, suggesting that these are key factors in differentiating between different editing styles in short-format videos.

However, it is important to note that our analysis is not exhaustive and that there is more work to be done in analyzing the clusters of data. For example, we can conduct a qualitative analysis of the videos in each cluster to gain a better understanding of the editing styles and themes represented. We can also explore the relationship between the clusters and other metadata such as video category, upload date, and creator demographics.



(*Figure 3.* 3D TSNE plot for clusters)



(*Figure 4.* Elbow plot)

Conclusion

In conclusion, this project offers valuable insights into the creative strategies and techniques used in short-format video editing. By utilizing PySceneDetect and unsupervised deep learning techniques, we were able to extract various video statistics and train a clustering model to identify the different editing styles and themes in our dataset.

Our findings can be used by content creators and video producers to create and optimize short videos for social media platforms. However, it is important to note the limitations of our dataset and analysis, such as potential biases in our selection of search queries and the subjective nature of clustering.

Despite these limitations, we were able to achieve promising results, as demonstrated by the clustering metrics such as silhouette, Calinski-Harabasz, and Davies-Bouldin. Our recommendations for future work include expanding the dataset to include videos from other platforms and exploring supervised learning techniques for more accurate classification.

Overall, this project demonstrates the power of unsupervised deep learning techniques in analyzing trends in short-format video editing and provides a foundation for future research in this area.

References

Abdulhussain, Sadiq H., et al. "Methods and Challenges in Shot Boundary Detection: A Review." *Entropy*, vol. 20, no. 4, 4, Apr. 2018, p. 214. *www.mdpi.com*, https://doi.org/10.3390/e20040214.

Pycaret/Pycaret: An Open-Source, Low-Code Machine Learning Library in Python. https://github.com/pycaret/pycaret. Accessed 1 May 2023.

Soe, Than Htut. *AI Video Editing Tools. What Editors Want and How Far Is AI from Delivering?* arXiv:2109.07809, arXiv, 16 Sept. 2021. *arXiv.org*, https://doi.org/10.48550/arXiv.2109.07809.

Xie, Haolin, and Yu Sun. "An Intelligent Video Editing Automate Framework Using AI and Computer Vision." *Artificial Intelligence and Machine Learning*, Academy and Industry Research Collaboration Center (AIRCC), 2022, pp. 201–09. *DOI.org (Crossref)*, https://doi.org/10.5121/csit.2022.121216.

Acknowledgements

I am deeply grateful to my family, Professors Chun and Elkins, as well as LLMs, for their unwavering support and encouragement throughout this project.