SCENE CUT DETECTION IN SHORT VIDEOS USING PRE-PROCESSING TECHNIQUES AND CONVOLUTIONAL NEURAL NETWORKS

Nina LAWANKOWIT[1] and Sukree SINTHUPINYO[1]
1 Faculty of Engineering, Chulalongkorn University, Thailand;
nina_laws@hotmail.com (N. L.); sukree.s@chula.ac.th (S. S.)

ARTICLE HISTORY
Received: 3 February 2025 **Revised:** 24 February 2025 **Published:** 4 March 2025

ABSTRACT
This study presents a comprehensive approach to scene cut detection in videos using a combination of frame extraction and a Convolutional Neural Network (VGG16 model) with a Color Histogram method to prevent raw video file reconstruction. The main dataset contains 167,490 frames collected from various users on TikTok for the scene cut detection model development. However, since this research utilizes a modular design that consists of data preparation, data pre-processing, and model implementation stages, it promotes efficient use of resources and guarantees data safety. The results demonstrate that the model performs excellently in recognizing scene cuts. Specifically, video frames with scene cuts achieved a precision of 85.11%, a recall rate of 93.96%, and an F1-score of 0.8932. This indicates its ability to identify true transitions among the scenes while ensuring low false positive rates. In video frames without scene cuts, the model achieved low false positives, highlighting its ability to accurately distinguish between static and transitional scenes.
Keywords: Scene Cut Detection, Video Processing, Frame Extraction, Convolutional Neural Network

CITATION INFORMATION: Lawankowit, N., & Sinthupinyo, S. (2025). Scene Cut Detection in Short Videos Using Pre-Processing Techniques and Convolutional Neural Networks. *Procedia of Multidisciplinary Research, 3*(3), 39.

INTRODUCTION

According to HubSpot's 2024 State of Marketing & Trends Report, short-form video, data collected from 1,400+ global B2B and B2C marketers, is now the top content marketing format, offering the highest ROI at 31% compared to other formats, and it is the most widely used by marketers (Iskiev, 2024). The short videos offer a quick and engaging method for people to share content, manage marketing activities, and connect with others (Liu et al., 2019).

Consecutively, it is becoming harder to stand out and keep audiences interested. Social media influencers need to find creative and engaging ways to grab their attention as they scroll through their feeds (Ge et al., 2021). According to HubSpot's Short-Form Video Trends and Best Practices, one of the best practices for short-form video is using creative techniques like quick cuts. HubSpot's survey on short-form video trends emphasizes the importance of using fast transitions while avoiding long, static shots to create a fast-paced, engaging atmosphere that keeps viewers hooked (Bretous, 2024). However, manually identifying scene cuts and the focus point can be time-consuming and challenging (Yu & Srinath, 2001), especially for social media influencers who create a lot of content. Generating and disseminating content has also led to increasing issues of data privacy, especially with laws such as the Personal Data Protection Act (PDPA) in effect. This problem becomes more serious especially when scene cut detection is done using automated means as this often employs raw video data. Without the right measures in place, there is the possibility of the original material being accessed or reverse-engineered.

This research introduces proof of concept of a service capable of addressing the aforementioned challenges, specifically the privacy concerns. Compared to the traditional scene cut detection models, our concept takes significant consideration in emphasizing data protection to ensure privacy of the original media remains protected and compliant with PDPA guidelines. With the use of pre-processing techniques (such as Color Histogram) to convert raw video files into irreversible tensor files format and subsequently use them as an input to the scene cut detection model as a second-introduced part of the service. For the scene cut detection model, it utilizes Convolutional Neural Networks (CNN) in analyzing the frames of a video and identifying when a scene cut occurs (Qiao et al., 2020). This approach offers a different perspective from threshold-based scene cut detection techniques (Thakur, 2018), demonstrating a more flexible approach to handling complicated tasks, and enhancing the effectiveness of video analysis tasks such as video summarization and video indexing.

It separates the client from the server and thus brings scalability and future proofing in such strict measures for data privacy. This research also describes how this methodology and design considerations were adopted in this service application to real-world cases of social media influencers and video content creators. With privacy-preserving pre-processing techniques and flexible server-side architecture, these components are architected entirely from the start.

LITERATURE REVIEWS

Detecting Object-Level Scene Changes in Images with Viewpoint Differences Using Graph Matching

Their research proposed a method for detecting changes in object levels to distinguish scene changes in images based on viewpoint differences, utilizing objection detection and a graph-matching network. The network consists of two techniques: an object detection module and a graph matching module. The object detection module detects object boundaries and creates object features. The graph matching module, subsequently, feeds the object features and constructs assignments using a graph neural network and an optimal matching layer. Subsequently, the boundaries of unmatched objects are distinguished as object-level changes. In summary, this technique involves separating objects in each image using object detection and utilizing object pairing to create objects' similarities, allowing the network to detect objects that remain or disappear based on the similarities. In comparison to this research, our scene cut detection model focuses on a similar theme but with a different approach. While the referenced study targets object-level changes in images, emphasizing scene change detection based on viewpoint differences, our model explores scene cut detection in video sequences. Despite differences, both research addresses challenges in scene change/cut detection and offer potential applications. Comparing the methodology with the referenced research, it provides a comprehensive understanding of scene cut detection techniques and alternatives.

Change Detection and Feature Extraction Using High-Resolution Remote Sensing Images

Their research presented a method for detecting changes from high-resolution satellite images with the purpose of identifying changes on the Earth's surface. This method addresses the time-consuming process issues. For instance, examining the vacant areas and physically surveying regions. The analysis of the data for change detection depends on image interpretation skills and the resolutions of the input data. Thus, computer vision techniques, such as color histograms, were employed to display successful change detection and categorize various areas, including water, bodies, roads, vacant land, and more. With this technique, it has been evaluated with an overall accuracy of 88.2%. The work will significantly contribute to urban planning. In comparison, both their and our research share a similar goal of utilizing images for change/cut detection, but with differences in methodology and focus. While the referenced study detects the changes on the Earth’s surface using high-resolution satellite images, our scene cut detection model focuses on a different domain including the techniques employed.

Detection of Cut Transition in Videos Using Optical Flow and Clustering

Their research presented a scene detection technique by segmenting frames from videos. They introduced a method using clustering algorithms to identify the sudden changes in the video by leveraging light flow and then analyze the number of signification scene changes. In Figure 2, it illustrates the flow of retrieving the scene from the video sequence by employing clustering methodologies with optical flow corner detector techniques. For the optical flow technique, two different approaches were experimented with: Features from Accelerated Segemetned Test (FAST), which utilized corner detection method by selecting pixels that would be the point of interest; Oriented FAST and Rotated BRIEF (ORB), where it considered features as binary strings and handle in-plane rotations and enhance feature detection performance. This research concluded that the ORB method is more effective than the FAST method because the ORB method detects prominent corner points better. In contrast, their research utilizes optical flow and clustering algorithms techniques for scene cut detection, but our scene cut detection model utilizes a pre-trained CNNs model as a feature extractor. With the capabilities of video preprocessing and CNNs, we extract and calculate differences from features of the video sequences to detect scene cuts. The absence of optical flow and Canny edge detection in our methodology distinguishes our approach from the aforementioned study.

Video Objects Detection Using Deep Convolutional Neural Networks

Their research presented a study model of CNNs detailing its main features and applications. With the ImageAI technique, they utilized a CNNs model integrated with the YOLOv3 model which allows map evaluation of the trained model. The process classifies and predicts objects present in each frame of the video dataset, resulting in object detection in videos. According to this research and experiment, the model demonstrated high accuracy in object detection and tracking. However, due to factors such as size and lighting, the model is not able to detect all objects. In comparison, our approach shares a similar focus on utilizing CNNs for object detection in video frames. While the referenced study’s CNN model integrates YOLOv3 model for object classification and prediction in video datasets, our research might employ a different CNN architecture. However, since the foundation of both research theses were based on the CNNs model, the referenced study could provide a comprehensive understanding and technique that would be valuable for our scene cut detection model development.

Privacy-Preserving Action Recognition Using Coded Aperture Videos
The authors propose a method of action recognition, tapping into the capabilities of delayed coded aperture imaging techniques to protect the identities of the subjects. Their research examines the applicability of phase correlation in the Fourier space for direct motion estimation between consecutive video frames and proposes it as a solution to the problem of privacy preservation without the use of raw video footage. The transformation of features in the frequency space ensures, in a way, the confidentiality of the information contained in the original video while facilitating the detection of a given activity or scene. On the contrary, our research mainly focuses on scene cut detection in TikTok videos and more on how tensor representations with added Color Histogram technique reinforce privacy considerations. This is crucial in the case of video content created and shared on social networking sites, where due to the drastic requirements of privacy and high performance, it is easier to employ irreversible transformations of the data, alongside application-specific tailoring of the VGG16 network, thus resolving the chapter on video motion analysis in a far more productive way.

RESEARCH METHODOLOGY

In response to the need for an efficient and privacy-preserving scene cut detection framework for short videos, this study adopts a structured methodological approach that integrates pre-processing techniques and convolutional neural networks. The methodology comprises data collection, data pre-processing, model implementation, and performance evaluation, ensuring accuracy and computational efficiency while safeguarding data privacy.

Data Collection
This research utilizes a TikTok video dataset as the primary source for scene cut detection. Through the utilization of TikTok videos, they provide a variety of topics, visual content, and user-generated videos. A significant number of TikTok videos, totaling 167,490 frames datasets, were collected to ensure coverage of scene variations and transitions. The quality of the videos should be defined by criteria which are 1) resolution is at least 720p 2) an aspect ratio is 9:16 3) frame rate is 30 frames per second 4) video length is within 10-95 seconds.

Data Pre-Processing
In the present module of data preprocessing, the aspect of privacy is focused on primarily. Traditional unedited videos or video clips normally have certain visual aspects that are sensitive and can be easily exposed when the videos are accessed in the public domain. As a solution to this problem, video frames are transformed into tensors using pre-processing techniques such as Gaussian blur, Adaptive Thresholding, Canny Edge Detection, and Color Histogram. This research investigated with its privacy-preserving advantages to ensure that the tensors used in the scene cut detection model cannot be reverse-engineered to recreate the original video frames. In addition, to prove that original frame pictures cannot be extracted from the tensors, a reverse transformation will be attempted for the eventual generation of the original frame pictures from the Tensors.

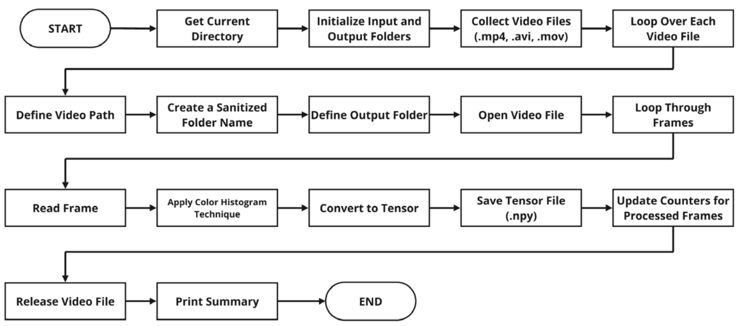

Figure 1 The process of converting video data into a sequence of tensors for the data pre-processing stage.

Model Implementation

The implementation of the scene cut detection model uses pre-trained deep learning models and similarity metrics for achieving frame transition classification to differentiate consecutive video scenes. This process involves taking high-level visual features of video frames or tensors with the help of a VGG16 model, which captures complex patterns and details. By comparing these features through cosine similarity, we assess how visually distinct consecutive frames are. This metric is rather useful when we are interested in scene cuts as this would contribute to quantifying the degree of difference between feature vectors and therefore serves as a good summary statistic on all our dimensional information. This approach enables us to exploit sophisticated image processing methods in video content scene cut detection for boosting specificity and sensitivity.

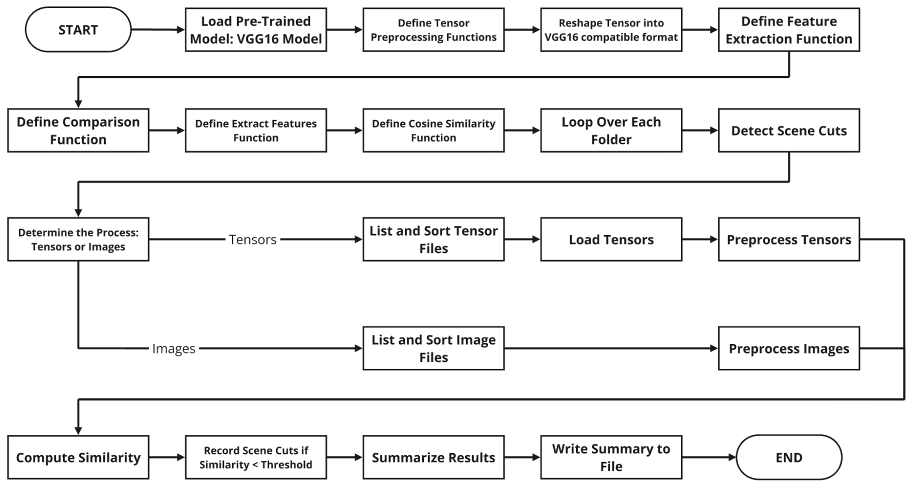

Figure 2 The workflow of the scene cut detection framework, depicting the step-by-step process from initialization to scene cut detection.

Performance Evaluation

All aspects of the model are considered, various performance measures are introduced and including True Positives (TP), True Negatives (TN), False Positives (FP), False Negatives (FN), Precision, Recall, F1-Score, and mainly Turn Around Time. These measures will evaluate how well the model performs the task of detecting the scene cuts, more specifically, how many cuts were recognized correctly and misclassified. Two types of videos are considered in this evaluation, videos with cuts and without cuts.

RESEARCH RESULTS

System Architecture Analysis

The system proposed is developed on two components to maintain user privacy and efficiency with scalability. The pre-processing and analyzing tasks are hence separated into user-side and server-side components with privacy-preserving seamless operations. 1) User-Side Processing: On the user end, there is a preprocessing service that transforms video files into tensors using the Color Histogram technique. This method abstracts out the visual content of the video into its statistical representation without regard to any spatial or structural details. Thus, these tensors are not reversible to the original video format and are secure from user privacy breaches. This service makes it easy for several users to process their videos at the same time without the need to queue, which enhances the customer experience. 2) Server-Side Processing: Once the video is converted into tensors, the user uploads them to the server, through which they will be passed for scene cut detection. The server uses a CNNs-based model to process these tensors and provide the output back to the user in the form of a detailed scene cut analysis.

This model will house pre-processing on the user side and analysis on the server, thereby reducing the server load and ensuring that sensitive video content never leaves the user's device.

Data Size Comparisons

The data size comparison between the original video files and the tensors generated using pre-processing techniques was conducted on 10 random samples (14,430 frames). The size of the original video files varies between 1.48 MB for shorter durations (e.g., 13 seconds) to 18.6 MB for longer durations (e.g., 96 seconds). The tensor file size difference between pre-processing techniques 1) Canny Edge Detection's tensor file sizes are significantly larger, ranging from 2.37 GB to 30.9 GB, depending on the video duration. 2) Adaptive Thresholding's tensor file size is also substantial, ranging from 2.37 GB to 30.9 GB. This is due to the binary transformation applied to each frame, which results in dense pixel-level information being preserved. This is due to the dense edge data retained by the Canny Edge Detection method. 3) Gaussian Blur's tensor file sizes are the largest among all techniques, ranging from 7.12 GB to 92.8 GB. The increase in size is caused by the smoothing process, which retains a broader range of pixel information while suppressing noise, making it less optimal for data size reduction. 4) Color Histogram's tensor file sizes are much smaller, ranging from 0.8 MB to 3.36 MB. This drastic reduction highlights the efficiency of the Color Histogram technique in reducing data size while preserving the necessary information for analysis. Among the different methods evaluated, Color Histogram appears to be the most effective method, generating tensor files that contain information useful for analysis while being significantly smaller than the original video sizes.

Table 1 Data Size Comparison of Original Videos and Pre-Processed Tensors

Video Duration (seconds)	**Original Video File Size (MB)**	**Tensors File Size After Applying Pre-Processing Techniques**			
		Canny Edge Detection (GB)	**Adaptive Thresholding (GB)**	**Gaussian Blur (GB)**	**Color Histogram (MB)**
44	5.58	10.2	10.2	30.8	2.44
41	7.63	9.74	9.74	29.2	2.31
49	5.2	11.5	11.5	34.5	2.74
30	5.08	7.05	7.05	21.1	1.67
13	1.48	2.37	2.37	7.12	0.8
53	8.6	12.6	12.6	37.9	3.01
43	5.95	10.1	10.1	30.4	2.41
63	7.35	14.6	14.6	43.8	3.47
96	18.6	30.9	30.9	92.8	7.36
49	7.6	11.5	11.5	34.7	2.75

Effectiveness of Pre-Processing Techniques as a Privacy Measure

Within the scope of our concerns pertaining to privacy preservation during the conduct of video analysis, the pre-processing techniques are vital as far as the concealment of visual content from exposure while still allowing accurate analysis. The chosen pre-processing techniques are 1) Canny Edge Detection highlights the edges present in an image concentrating on the outlines and edges while eliminating the internal details. It helps to retain some structural details that are needed to perform certain functions like detecting transition cuts in a scene, however, makes it difficult to recreate the unprocessed image. 2) Adaptive Thresholding highlights some areas of the image with strong contrast while binarizing the image in contrast. It tones down certain areas, making it easier to discern certain elements while accomplishing what it seeks to accomplish but tends to lose the finer aspects. 3) Gaussian Blur seeks to smooth the image by taking the average of all the pixel values in the local neighborhood, normally blurring the fine details as well as high-frequency components. The outcome in turn is a smudged image which tends to hide the visual content of the image but might even be able to keep structural information to a certain extent and hence it can be analyzed for a few types of analysis. 4) Color Histogram reduces an image to calculate the distribution of intensities per channel: the red, green, and blue channels. With no area information, it easily obscures information that may be sensitive but keeps enough information on color variations for applications like detecting changes in the scene. On the downside, some assessment methods that need intricate details may not be possible due to the limitations of the approach considering the absence of spatial information.

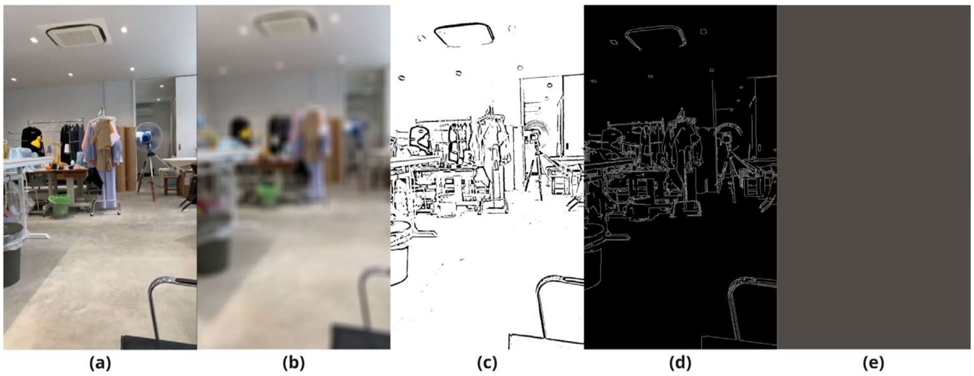

Figure 3 Visual comparison of different image processing techniques.
Note: (a) is an original frame image. (b) is a reverted image from the Gaussian Blur technique, demonstrating a blurred effect. (c) is a result of Adaptive Thresholding, highlighting regions of high contrast in a binary format. (d) is an image after applying Canny Edge Detection, showcasing the edges and boundaries in the scene. (e) is an abstract image constructed based on histogram tensor color distribution.

The Performance Contributed to Scene Cut Detection Model Comparisons
The performance difference between the four techniques was conducted on 10 random samples (14,430 frames) under the same threshold set for scene cut detection. Based on the result, the Color Histogram technique produces the best F1-score, which is 0.9111, while the rest produces fairly similar results.

Table 2 Scene Cut Detection Performance Comparison of Pre-Process Techniques

Actual Scene Cuts	**Detected Scene Cuts**			
	Canny Edge Detection	**Adaptive Thresholding**	**Gaussian Blur**	**Color Histogram**
5	6	6	6	5
17	9	17	17	21
14	38	61	27	18
4	15	9	5	4
14	16	15	15	16
8	13	14	19	6
14	15	15	18	13
16	16	20	25	22
15	9	16	26	16
19	10	16	39	23
Precision	70.70%	65.08%	63.96%	85.42%
Recall	88.10%	97.62%	100.00%	97.62%
F1-Score	0.7845	0.7810	0.7802	0.9111

Scene Cut Detection Model Performance

A total of 116,640 frames from video with scene cuts were processed. These video frames ranged from 6 seconds to 96 seconds and featured a wide range of content genres and various settings. The following metrics were used to evaluate the performance of the scene cut detection system for video frames that contain scene cuts.

1) True Positives (TP): The number of scene cuts correctly identified by the detection system.
2) True Negatives (TN): The number of non-scene cuts correctly identified by the detection system.
3) False Positives (FP): The number of incorrect detections where the system identified a scene cut that is not actually present.
4) False Negatives (FN): The number of actual scene cuts that were not detected by the system.
5) Precision: To measure the precision of the scene cut detection, we used the following equation:

$$Precision = \frac{TP}{TP + FP} \times 100$$

6) Recall (Sensitivity): To measure the recall of the scene cut detection system, we used the following equation:

$$Recall = \frac{TP}{TP + FN} \times 100$$

7) F1-Score: Balances precision and recall into one score, we used the following equation:

$$F1 - Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Table 3 Performance Result of Scene Cut Detection Model

Video Type	TP	FP	FN	TN	Precision (%)	Recall (%)	F1-Score
Scene Cuts	1,898	332	122	114,188	85.11%	93.96%	0.8932
Without Scene Cuts	N/A	31	N/A	50,719	N/A	N/A	N/A

The model displays a strong performance in detecting scene cuts, especially for video frames with evident and clear transitions. For video frames with scene cuts, the model has a precision of 85.11%, recall of 93.96%, and F1-score of 0.8932. The F1-score's value near 1 suggests that the model's predictions are both accurate and comprehensive. This means that this model is pretty good at identifying True Positives while minimizing False Positives. What this precision score means in real terms is that nearly all detected scene cuts were real scene transitions leading to fewer unnecessary alerts given by it. Accuracy evaluation is ignored due to it is heavily impacted by class imbalance; this can mislead the interpretation of model quality, as it does not account for the difficulty of detecting scene cuts. Therefore, this research only concerns precision, recall, and F1-score as the performance evaluation.

However, the recall score paints a different picture as there were 122 missed scene cuts (False Negatives). This implies that while the model has some degree of high performance, it may occasionally fail to pick up on more subtle or less obvious things such as those that have only slight or minimal variations. In this case, we could be talking about situations where there are not very blatant differences in scenes or almost imperceptible transitions. This aligns with the observation of extremely high False Negative rates on video frames with low mobility or little change suggesting the need for more work to find more opaque transition detections. For video frames without scene cuts, 50,719 were identified as True Negatives while only 31 instances were identified as False Positives by the model. True Positives (TP) and False Negatives (FN) are not considered as they are not relevant since there are no actual scene cuts to detect.

DISCUSSION & CONCLUSION

For scene cut detection, a TikTok video dataset serves as the major source of the study, which includes a wide range of subjects, visual content, and even user-generated videos. The aim of the research was to develop a reliable model able to accurately identify scene cuts with 167,490 collected frames from 200 TikTok videos. The system architecture is divided into three modules: data preparation, preprocessing, and the implementation of the model. This modular prototype is significantly beneficial from the resource utilization point of view and has the added benefit of managing sensitive data separately to allow for greater security levels. Furthering this architecture, the system separates the workflow into two sides of the model, the user side and the server side. With this division, privacy preservation can be blended effortlessly with effective data use and scalability in computation, which solves the challenges posed by scene cut detection.

The model's performance varies depending on scene complexity and video length. The model was highly effective in detecting clear transitions but showed limitations in recognizing subtle scene changes, leading to a moderate recall rate. This suggests that while the VGG16 model effectively captures prominent features, its fixed input size and reliance on color-based features may restrict its ability to detect more nuanced transitions. In other words, the model performs exceptionally well for typical scene cuts but may require enhancements for complex scenes with minimal visual differences.

According to the study results, it can be discussed that the integration of pre-processing techniques, particularly Color Histogram, not only enhances scene cut detection accuracy but also serves as a privacy-preserving measure by transforming raw video frames into irreversible tensor representations. This approach ensures that visual content is anonymized while preserving essential transition features, complying with data privacy regulations such as the Personal Data Protection Act (PDPA). Nevertheless, it can be observed that although Color Histogram effectively abstracts visual content, challenges remain in fully safeguarding against advanced reverse-engineering techniques. Consequently, the study demonstrates the feasibility of achieving high-accuracy scene cut detection while prioritizing data privacy, paving the way for secure video processing in social media applications.

The discussion portrayed in this study aligns with the findings of Qiao et al. (2020), who demonstrated the effectiveness of CNNs in scene change detection and extends their work by introducing privacy-preserving pre-processing techniques. Additionally, the results are consistent with Thakur (2018), who highlighted the limitations of threshold-based techniques, further supporting the superiority of deep learning-based approaches in handling complex scene transitions. This study also corroborates the work of Liu et al. (2019), emphasizing the importance of privacy considerations in automated video processing.

This study provides two main contributions: practical and theoretical insights. For practical contributions, the proposed model enables social media influencers and video content creators to efficiently detect scene cuts while ensuring data privacy, enhancing content management and compliance with privacy laws. Furthermore, the modular design of the system architecture supports scalability, allowing future extensions to accommodate additional video formats or advanced detection techniques. For theoretical contributions, the study confirms that privacy-preserving pre-processing techniques can effectively anonymize visual content while maintaining high scene cut detection accuracy. This adds to the existing literature on secure video processing by validating the use of irreversible tensor transformations combined with CNNs.

However, this study has some limitations that must be considered to enhance future research. One limitation involves the reliance on the VGG16 model, which, despite its effectiveness, may struggle with detecting subtle transitions or highly complex scenes due to its fixed input size and feature extraction approach. Consequently, future research may explore more

advanced CNN architectures or hybrid models that can better capture nuanced scene changes. Another limitation is related to the dataset, which consists of TikTok videos. Although diverse, the dataset may not generalize well to other platforms or video types. Therefore, future research should expand the dataset to include videos from different social media platforms or genres and conduct comparative analyses to enhance model robustness. Lastly, although Color Histogram proved effective for privacy preservation, further investigations are needed to assess its resilience against advanced reverse-engineering techniques. Future studies could explore additional privacy-preserving methods, such as differential privacy or homomorphic encryption, to strengthen data security. These considerations provide a foundation for enhancing the scene cut detection model's accuracy, generalizability, and privacy preservation, thereby contributing to the advancement of secure video processing technologies.
The findings of this study can be applied in many ways where privacy preservation and video analysis play a major role. For instance, the scene cut detection capability would come in handy to media companies and content developers so that they can edit their footage faster without risking the privacy of any person. In surveillance and other security operations, this can assist in pinpointing the specific scenes in footage that are necessary and at the same time preserving the identity of the individuals. Video-sharing websites would also take advantage of them, as a more risk-free way of allowing users to post videos is to do so under legislations mandating user control over the videos. This can even be applied in schools and by researchers for the analysis of video content without revealing the identities of the people appearing in the videos. To sum up, this model enables various sectors of the economy to utilize video content more responsibly.

REFERENCES

Voorveld, H. A., Van Noort, G., Muntinga, D. G., & Bronner, F. (2018). Engagement with social media and social media advertising: The differentiating role of platform type. *Journal of advertising, 47*(1), 38-54.

Iskiev, M. (2024). *The 2024 State of Marketing & Trends Report: Data from 1400+ Global Marketers*. Retrieved from https://blog.hubspot.com/marketing/hubspot-blog-marketing-industry-trends-report?hubs_content=blog.hubspot.com/marketing&hubs_content-cta=The%202024%20State%20of%20Marketing%20%26%20Trends%20Report%3A%20Data%20from%201400%2B.

Liu, G. F., Gao, P. C., Li, Y. C., & Zhang, Z. P. (2019, August). Research on the influence of social media short video marketing on consumer brand attitude. In *2019 5th International Conference on Social Science and Higher Education (ICSSHE 2019)* (pp. 433-438). Atlantis Press.

Ge, J., Sui, Y., Zhou, X., & Li, G. (2021). Effect of short video ads on sales through social media: the role of advertisement content generators. *International Journal of Advertising, 40*(6), 870-896.

Bretous, M. (2024). *6 Short-Form Video Trends Marketers Should Watch in 2024 [New Data]*. Retrieved from https://blog.hubspot.com/marketing/short-form-video-trends?hubs_content=blog.hubspot.com/marketing/hubspot-blog-marketing-industry-trends-report&hubs_content-cta=short-form video renaissance.

Yu, J., & Srinath, M. D. (2001). An efficient method for scene cut detection. *Pattern Recognition Letters, 22*(13), 1379-1391.

Qiao, H., Wan, X., Wan, Y., Li, S., & Zhang, W. (2020). A novel change detection method for natural disaster detection and segmentation from video sequence. *Sensors, 20*(18), 5076.

Thakur, M. K. (2018). Dynamic dual threshold based schemes for abrupt scene change detection in videos. *Journal of Theoretical and Applied Information Technology, 96*(14).

Doi, K., Hamaguchi, R., Iwasawa, Y., Onishi, M., Matsuo, Y., & Sakurada, K. (2022). Detecting object-level scene changes in images with viewpoint differences using graph matching. *Remote Sensing, 14*(17), 4225.

Sharma, V. K., Luthra, D., Mann, E., Chaudhary, P., Chowdary, V. M., & Jha, C. S. (2022). Change detection and feature extraction using high-resolution remote sensing images. *Remote Sensing in Earth Systems Sciences, 5*(3), 154-164.

Dash, N., Nayak, S. K., & Majumdhar, J. (2021, August). Detection of Cut Transition in Videos Using Optical Flow and Clustering. In *2021 Asian Conference on Innovation in Technology (ASIANCON)* (pp. 1-7). IEEE.

Nife, N. I., & Chtourou, M. (2022, May). Video Objects Detection Using Deep Convolutional Neural Networks. In *2022 6th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)* (pp. 1-6). IEEE.

Wang, Z. W., Vineet, V., Pittaluga, F., Sinha, S. N., Cossairt, O., & Bing Kang, S. (2019). Privacy-preserving action recognition using coded aperture videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 0-0).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.