



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 4, April 2024

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)

# Video Moderation for Inappropriate Content Detection Using Machine Learning

Siddharth H. Waikar<sup>1</sup>, Shivam Pichke<sup>2</sup>, Ganesh Khedker<sup>3</sup>, Shreyash Bhatkar<sup>4</sup>,  
Shrikant Thakre<sup>5</sup>, Prof. M.K. Nichat<sup>5</sup>

Department of Computer Science & Engineering, P. R. Pote (Patil) Collage of Engineering and Management,  
Amravati, Sant Gadge Baba Amravati University, Amravati, Maharashtra, India.

**ABSTRACT:** This paper presents a method for human detection and activity recognition (HDAR) in aerial video sequences captured by moving cameras on aerial platforms. Utilizing object detection models trained on the COCO dataset, specifically EfficientDetD7, the proposed approach achieves 92.9% accuracy in detecting humans across various challenging conditions. Furthermore, combining features extracted by pre-trained CNNs like ResNet and EfficientNet with LSTM for human activity recognition results in an average accuracy and F1 score of 80%, surpassing existing methods. The outcome is a robust HDAR system that integrates EfficientDetD7, EfficientNetB7, and LSTM for human detection and activity classification.

**KEYWORDS:** Video Content Moderation, Machine Learning, Image Processing, Convolutional Neural Network (CNN), Inappropriate Content detection

## I. INTRODUCTION

The prominence of online videos on media-sharing platforms has positioned them as significant sources of information. However, the proliferation of misleading videos across social media has raised concerns about the potential for these videos to deceive audiences and adversely impact society. The primary motive behind creating misleading videos is often driven by financial gain, such as luring users into clicking on phishing links or making purchases through false advertisements.

This profit-oriented incentive has led to a rapid surge in the dissemination of deceptive videos on social media. The sheer volume of misleading videos on media-sharing platforms poses a significant challenge to the automated detection of such content. The suggested framework capitalizes on a convolutional neural network (CNN) model pre-trained on relevant data, empowering it to effectively

## II. OBJECTIVES

- To develop a robust and automated system for moderating video content using a combination of machine learning and image processing techniques.
- Design and implement an advanced algorithm capable of identifying inappropriate content.
- Develop mechanisms to enable real-time processing of video content.
- Build a machine learning model that achieves high accuracy and reliability in inappropriate content.
- To develop a robust and automated system for moderating video content using a combination of machine learning and image processing techniques.
- Design and implement an advanced algorithm capable of identifying inappropriate content.
- Develop mechanisms to enable real-time processing of video content.
- Build a machine learning model that achieves high accuracy and reliability in inappropriate content.
- To provide an overview of the current state of video moderation and its relevance in the digital age.
- To delve into the role of machine learning in automating video content moderation and highlight the latest advancements in this field.
- To discuss the ethical considerations, potential biases, and societal implications associated with video moderation and machine learning.
- To assess the impact of video moderation on freedom of expression, user privacy, and the delicate balance between content control and open dialogue.

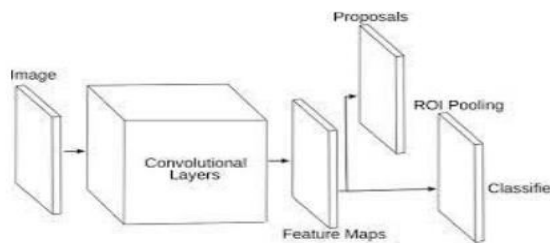


**Aim**

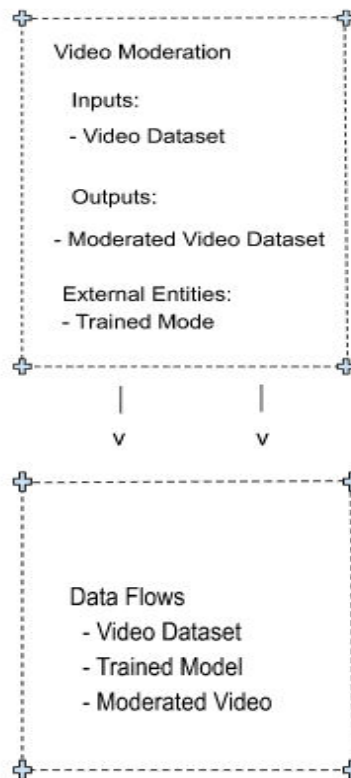
The primary aim of this project is to create a dependable and effective system employing CNN-based image processing and machine learning methods to autonomously identify and categorize inappropriate content within video media. By achieving this goal, the project aims to bolster online safety and instill greater confidence among users through continuous content monitoring across diverse video-sharing platforms.

**III. METHODOLOGY**

The aim is to develop a comprehensive system that utilizes CNN-based image processing to detect obscene content within videos and applies blurring techniques to ensure a safer and more appropriate viewing experience. Below are the steps that followed during building the CNN model.



**Basic diagram of CNN**



**DFD Diagram**



#### IV. IMPLEMENTATION & RESULT

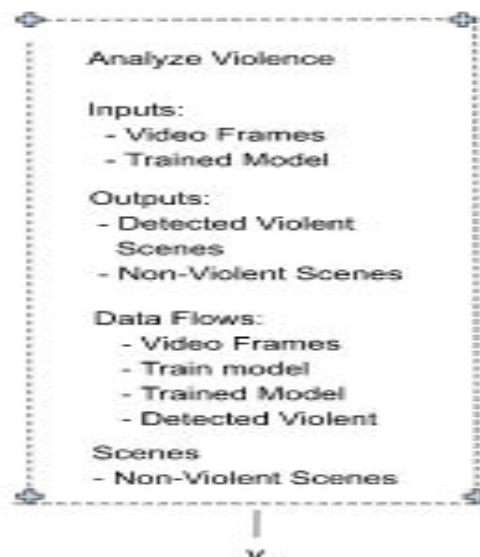
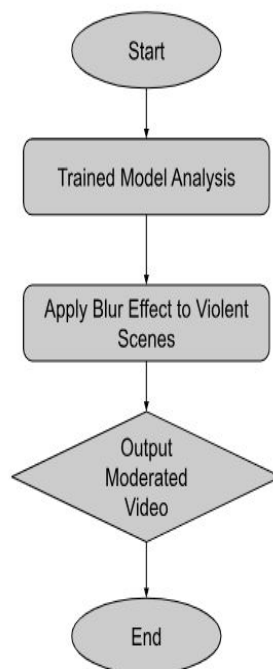
##### Implementation Steps

CNN Architecture: Design a CNN architecture optimized for image classification and object detection tasks.

Input Format: Configure the CNN to accept video frames as input.

Training: Train the CNN using the labeled dataset to learn to recognize and classify obscene content.

##### Flowchart

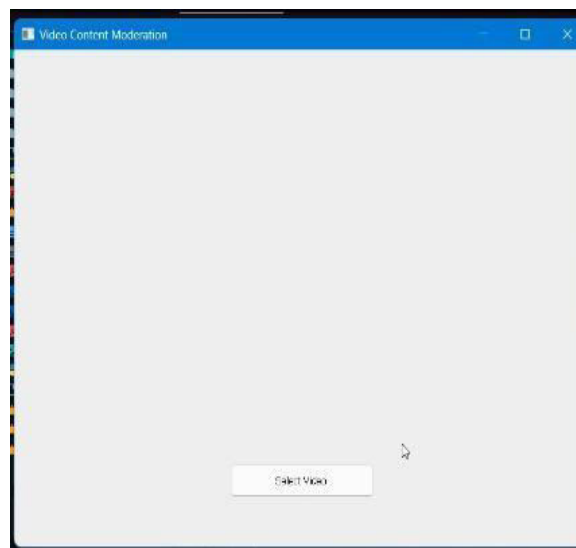




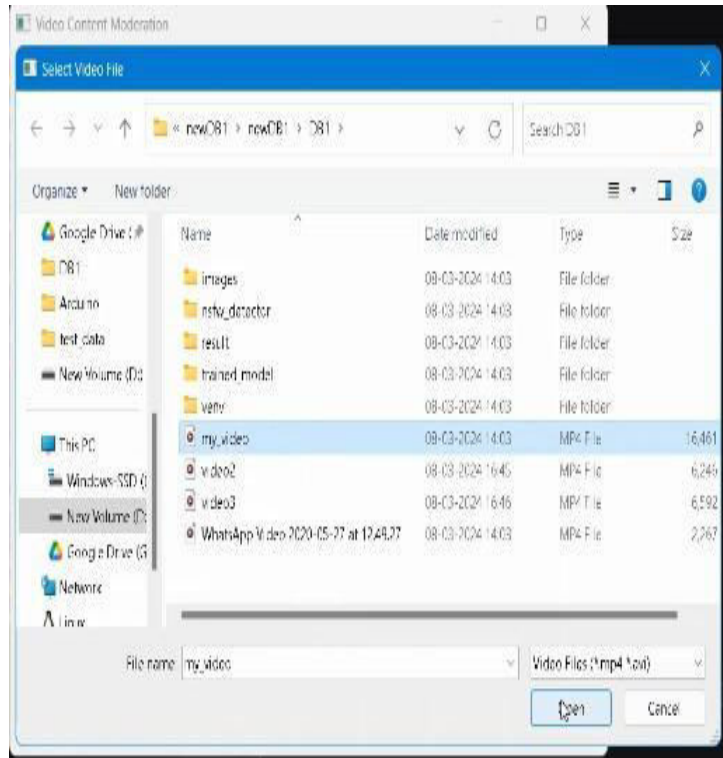
**Step 1: Login Using Credentials**



**Step 2 : Select Video**





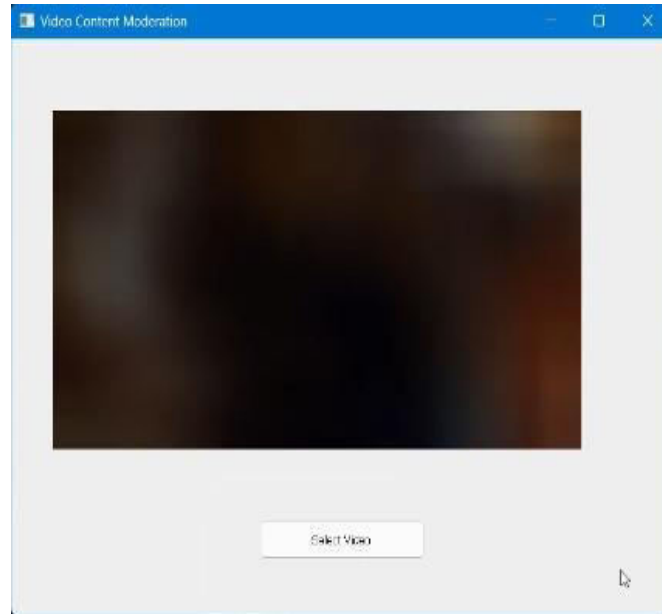


**Results**

**Input image**



### Output image



The above image illustrates the outcome of the project. It displays an initial image containing inappropriate content that is effectively blurred out. The project successfully implemented a video content moderation system for detecting and blurring inappropriate content using machine learning techniques.

By employing Convolutional Neural Networks (CNN) for image detection and tracking within the video, the system effectively identified obscene scenes. Upon detection, the system applied blurring techniques to censor the inappropriate content, ensuring a safer viewing experience.

This solution provides a robust method for automated content moderation in video formats, contributing to the maintenance of online decency standards.

### V. FUTURE SCOPE

The future scope for video moderation concerning inappropriate content detection based on machine learning is vast and evolving rapidly. As technology advances, machine learning techniques are playing a crucial role in automating the process of content moderation, especially for videos. Here are some key aspects and future possibilities in this space:

AI-Powered Video Moderation:

1. **Improved Accuracy:** Future developments in machine learning models, especially deep learning algorithms like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), will continue to enhance the accuracy of video content moderation systems in detecting inappropriate content, such as violence, hate speech, nudity, or graphic imagery.

2. **Real-Time Detection:** Advancements in real-time video analysis using techniques like object detection, action recognition, and frame-level analysis will enable platforms to promptly detect and moderate inappropriate content as videos are being uploaded or streamed.

3. **Multimodal Analysis:** Integrating multiple AI techniques for analyzing audio, video, and text content together to understand context and detect subtle cues that may indicate inappropriate content, leading to more comprehensive moderation capabilities.

4. **Contextual Understanding:** Future systems will focus on understanding the cultural context, intent, and nuances of content to avoid misinterpretations and false positives in moderation.



**5. Semi-Supervised and Self-Learning Models:** Implementation of semi-supervised learning techniques and self-learning models will enable video moderation systems to continuously improve by learning from new data and user feedback.

**6. Privacy Protection:** Developments in privacy-preserving AI techniques such as federated learning and differential privacy will allow for effective video moderation without compromising users' privacy.

## VI. CONCLUSION

Through the attainment of these objectives, this research seeks to contribute to the advancement of video content moderation practices by leveraging the power of machine learning. Particularly CNN-based image analysis, to create a dependable and high-performance system for the accurate detection of inappropriate content within videos. This system holds the potential to significantly enhance online safety and user trust across various video-sharing platforms.

## VII. ACKNOWLEDGMENT

We gratefully acknowledge the collaborative efforts of all team members for their invaluable support throughout the research process. Each member contributed significantly to the research, demonstrating dedication and expertise in their respective roles. Our collective effort and teamwork were essential in completing this project successfully.

We would also like to express our appreciation to the International Journal for Science and Advance Research in Technology for considering our work and providing the opportunity to publish our findings

## REFERENCES

- [1] Q.-F. Zheng, W. Zeng, G. Wen, and W.-Q. Wang, "Shape-based adult images detection," in Proceedings of the 3rd International Conference on Image and Graphics, pp. 150–153, December 2020.
- [2] Gonzalo Molpeceres Barrientos, Rocío Alaiz-Rodríguez, Víctor González-Castr, , Andrew C. Parnell, "Learning Techniques for the Detection of Inappropriate Erotic Content in Text" International Journal of Computational Intelligence Systems Vol. 13(1), 2020, pp. 591–603
- [3] S.Anoosha Devi, V.Arvind "Detection of Illegitimate Divulgarion of Obscene Contents Using Tensorflow" International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 20 (2019) pp. 14595-14599.
- [4] Artha Agastya1, Arief Setyanto, Kusriani, Dini Oktarina Dwi Handayani "Convolutional Neural Network for Pornographic Images Classification" School of Computing and IT, Taylor's University, Subang Jaya, Malaysia.(2019)
- [5] Kamrun Nahar Tofa, Farhana Ahmed, Arif Shakil, et al. Inappropriate scene detection in a video stream. PhD thesis, BRAC University, 2019.
- [6] Chang Xu, Dacheng Tao, and Chao Xu. A survey on multi-view learning. arXiv preprint arXiv:1304.5634, 2020.
- [7] Jing Zhao, Xijiong Xie, Xin Xu, and Shiliang Sun. Multi-view learning overview: Recent progress and new challenges. Information Fusion, 38:4354, 2021.
- [8] Mary B Short, Lora Black, Angela H Smith, Chad T Wetterneck, and Daryl E Wells. A review of internet pornography use research: Methodology and content from the past 10 years.
- [9] Nitish Srivastava, Elman Mansimov, and Ruslan Salakhudinov. Unsupervised learning of video representations using lstms. In International conference on machine learning, pages 843852, 2021.
- [10] S.Anoosha Devi, V.Arvind "Detection of Illegitimate Divulgarion of Obscene Contents Using Tensorflow" International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 20 (2019) pp. 14595-14599.
- [11] Artha Agastya1, Arief Setyanto, Kusriani, Dini Oktarina Dwi Handayani "Convolutional Neural Network for Pornographic Images Classification" School of Computing and IT, Taylor's University, Subang Jaya, Malaysia.(2019)
- [12] Kamrun Nahar Tofa, Farhana Ahmed, Arif Shakil, et al. Inappropriate scene detection in a video stream. PhD thesis, BRAC University, 2019.
- [13] Chang Xu, Dacheng Tao, and Chao Xu. A survey on multi-view learning. arXiv preprint arXiv:1304.5634, 2020.
- [14] Jing Zhao, Xijiong Xie, Xin Xu, and Shiliang Sun. Multi-view learning overview: Recent progress and new challenges. Information Fusion, 38:4354, 2021.
- [15] Mary B Short, Lora Black, Angela H Smith, Chad T Wetterneck, and Daryl E Wells. A review of internet pornography use research: Methodology and content from the past 10 years. <https://www.sciencedirect.com/science/article/abs/pii/S0959652604000174>
- [16] Wisakha Phoochinda and Saraporn Kriyapak, "Electronic Waste Recycling Business: Solution, Choice, Survival", International Journal of Sustainable Development and Planning [Vol. 16, No. 4, August, 2021, pp. 693-700] <https://www.iieta.org/journals/ijstdp/paper/10.18280/ijstdp.160409>





INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details