



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)

# Survey on Video Moderation for Inappropriate Content Detection Using Machine Learning

Prof. M.K.Nichat, Siddharth waikar, Shivam Pichke, Ganesh Khedker, Shreyash Bhatkar,  
Shrikant Thakre

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

**ABSTRACT:** The remarkable surge in video content on YouTube has captivated billions of viewers predominantly from the younger generation. Regrettably, this surge has also provided a platform or malicious individuals to disseminate disturbing visual material. Instances include the misuse of animated cartoons to share inappropriate content with children. To address this concern, the integration of a real-time automatic video content filtering mechanism within social media platforms is strongly recommended. This research introduces an innovative Machine-learning architecture designed to identify and categorize unsuitable content in videos. The proposed framework leverages a convolutional neural network (CNN) model enabling it to accurately detect inappropriate content in videos

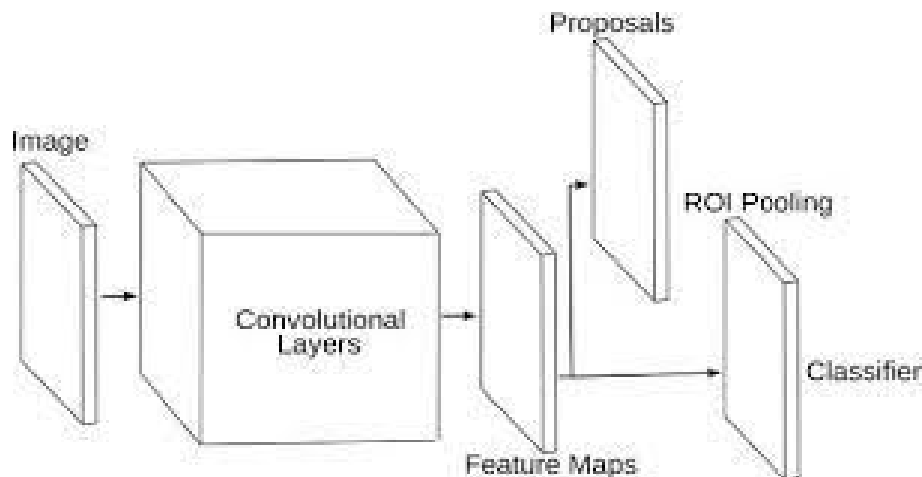
**KEYWORDS:** video moderation for inappropriate content, technology, optimization, emerging contaminants, digital platform, smart systems.

## 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. 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



**Step:-**

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.

**III. LITERATURE REVIEW**

Sr. No.	Author	Title	Journal	Findings	Advantages	Disadvantages
1.	Q.-F. Zheng, W. Zeng, et al.	“Shape-based adult images detection,”	in Proceedings of the 3rd International Conference on Image and Graphics, pp. 150–153, December 2020.	They propose a skin detection method using multi-Bayes classifiers in the paper.	<b>Efficient Skin Region Detection:</b> Utilizing shape features of skin regions can lead to efficient and accurate skin detection. Shape-based approaches can be robust in identifying skin areas within images.	<b>Limited to Shape Features:</b> Relying solely on shape features may limit the method's ability to detect subtle variations in skin tones or other distinguishing characteristics, potentially leading to false negatives or misclassification.
2.	Gonzalo Molpeceres Barrientos, Rocio Alaiz-Rodríguez, et al.	“Learning Techniques for the Detection of Inappropriate Erotic Content in Text”	International Journal of Computational Intelligence Systems Vol. 13(1), 2020, pp. 591–603	They present the methodology used to build the proposed sexual/erotic text classifier.	<b>Methodological Clarity:</b> The presentation of the methodology for building a sexual/erotic text classifier offers transparency and clarity in the process, which can be crucial for both understanding and replicating the approach.	<b>Limited to Binary Classification:</b> The method described appears to be designed for binary classification (sexual/erotic or not), which may not address the nuances of content moderation where a broader range of categories might be required.
3.	Artha Agastya1, Arief Setyanto, et al.	“Convolutional Neural Network for Pornographic Images Classification”	School of Computing and IT, Taylor’s University, Subang Jaya, Malaysia. (2019)	They had shown CNN, especially the VGG16 network. It is supported by the fact that the testing accuracy reached 0.938 on average. Compared to the state-of-the-art method, this proposed method has better precision, with a 2% difference.	<b>Feature Learning:</b> Convolutional Neural Networks (CNNs) excel at automatically learning relevant features from image data, making them well-suited for tasks like pornographic image classification where visual patterns are crucial.	<b>Data Bias:</b> CNN models can inherit biases present in the training data, potentially leading to unfair or inaccurate predictions, especially in sensitive tasks like classifying pornographic content. Mitigating bias requires careful dataset curation and model evaluation.
4.	S.Anoosha Devi, et al.	“Detection of Illegitimate Divulgence of Obscene Contents Using Tensorflow”	International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 20 (2019) pp. 14595-14599.	This paper mainly focuses on the identification of raunchy contents on electronic devices using an open-source machine learning framework-tensorflow and verification of the authenticity with the registered users.	<b>Open-Source Machine Learning Framework:</b> Utilizing an open-source machine learning framework like TensorFlow can make the approach accessible to a wide range of developers and researchers, promoting transparency and collaboration.	<b>Privacy Concerns:</b> Reporting URLs and victim authentication IDs to search engines could raise privacy concerns, as it involves sharing potentially sensitive information. Careful handling of user data and privacy implications is necessary.
5.	Kamrun Nahar Tofa, et al.	Inappropriate scene detection in a video stream.	PhD thesis, BRAC University, 2019.	They nude detection is implemented with the use of two algorithms. The first algorithm detects humans from the processed frames and then crops them out. The second algorithm is the nude detection algorithm used to determine if the images are nudes or not.	<b>Two-Step Approach:</b> The use of a two-step approach, with the first algorithm identifying humans in processed frames and cropping them out, followed by the second algorithm for nude detection, allows for a more targeted and efficient detection process.	<b>Algorithm Dependency:</b> The effectiveness of this method relies heavily on the accuracy and performance of the two algorithms chosen. If either algorithm underperforms, it may lead to false positives or false negatives, affecting the overall system's reliability.
6.	Aaron Halfaker and R. Stuart Geiger.	ORES: Lowering Barriers Participatory Machine Learning in Wikipedia.	<i>Proceedings of the ACM on Human-Computer Interaction</i> 4, CSCW2(2020).	Collaborative efforts between human editors and machine learning systems contributed to a more robust and sustainable model for maintaining large-scale knowledge platforms like Wikipedia.	<b>Quality Assessment:</b> ORES helps assess the quality of Wikipedia edits by providing predictions on whether an edit is likely to be damaging or beneficial. This assists in maintaining the overall quality of Wikipedia content.	<b>Bias and Fairness Concerns:</b> Machine learning models like ORES may inherit and potentially exacerbate biases present in the training data. If the training data is not representative, it can lead to biased assessments of articles, impacting marginalized or underrepresented topics or communities on Wikipedia.



7.	J. Mallmann, A. O. Santin, E. K. Viegas, R. R. dos Santos, J. Geremias	PPCensor: Architecture for real-time pornography detection in video streaming	<i>Future Generation Computer Systems</i> Volume 112, November 2020, Pages 945-955	The classification procedure starts with an input layer, which receives the raw image as the input, i.e., a pixel matrix. The resulting data is then surpassed for each posterior layer until the output layer is reached. Finally, the output layer generates a probability vector for each considered class label as the output. For instance, the class probabilities for normal or pornographic frames are generated.	<b>Real-time Detection:</b> PPCensor is designed to detect explicit content in video streams in real-time, which is essential for ensuring immediate content filtering and moderation on video streaming platforms. This can help prevent explicit content from being disseminated to viewers.	<b>False Positives and False Negatives:</b> Real-time content filtering systems, including pornography detection, are prone to both false positives and false negatives. False positives can lead to legitimate content being blocked or flagged incorrectly, while false negatives allow inappropriate content to slip through.
8.	Fudong Nian, Teng Li, Yan Wang, Mingliang Xu, Jun Wu c	Pornographic image detection utilizing deep convolutional neural networks	<i>Neurocomputing</i> Volume 210, 19 October 2016, Pages 283-293	In this paper, we proposed a pornographic image detection method based on deep convolutional neural network. The proposed deep network has minimal complexity and is very efficient for our task. And we designed a very appropriate pornographic data acquisition and augmentation method that is improved based on the experience and observation by human.	<b>High Accuracy:</b> Deep Convolutional Neural Networks (CNNs) have demonstrated exceptional performance in image classification tasks, including pornographic content detection, achieving high accuracy rates even with large and diverse datasets.	<b>Interpretability:</b> Deep CNNs are often considered black-box models, making it challenging to interpret and understand the decision-making process behind predictions. This lack of interpretability can hinder trust and transparency in content moderation systems.
9.	Jiann-Shu Lee, Yung-Ming Kuo, Pau-Choo Chung, E-Liang Chen	Naked image detection based on adaptive and extensible skin color model	<i>Pattern Recognition</i> Volume 40, Issue 8, August 2007, Pages 2261-2270	Effectively blocking or filtering out pornography has become a critical issue. In this paper, a new naked image detection algorithm is proposed.	<b>Accuracy:</b> Skin color models can be highly effective in identifying naked or partially clothed human bodies, particularly when combined with other image processing techniques such as edge detection or texture analysis.	<b>False Positives:</b> Skin color alone may not always accurately indicate nudity, leading to false positives where non-nude images are incorrectly flagged as containing nudity. This can be particularly problematic in cases where clothing or background colors resemble skin tones.

#### IV. 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.

#### ACKNOWLEDGMENTS

We acknowledge the valuable contributions of researchers, institutions, and organizations whose work has informed this survey. Their impactful research forms the basis for our comprehensive review of waste management practices. We express our sincere gratitude to the authors and researchers whose pioneering contributions form the backbone of this survey paper.

#### 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] 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>.
- [11] Xuanjing Shen, Wei Wei, and Qingji Qian. The filtering of internet images based on detecting erotogenic-part. In *Proc. 3rd International Conference on Natural Computation (ICNC)*, 2007.
- [12] T. Deselaers, L. Pimenidis, and H. Ney. Bag-of-visualwords models for adult image classification and filtering. In *Proc. 19th International Conference on Pattern Recognition(ICPR)*, pages 1 –4, 2008.
- [13] J. Wang, J. Li, G. Wiederhold, and O. Firschein. System for screening objectionable images. Technical Report 1998-5, Stanford InfoLab, 1998.
- [14] H.A. Rowley, Y. Jing, and S. Baluja. Large scale imagebased adult-content filtering. In *Proc. International Conference on Computer Vision Theory and Applications*, 2006.
- [15] Xiaoyin Wang, Changzhen Hu, and Shuping Yao. A breast detecting algorithm for adult image recognition. In *Proc. International Conference on Information Management, Innovation Management and Industrial Engineering.*, 2009.
- [16] Hogyun Lee, Seungmin Lee, and Taekyong Nam. Implementation of high performance objectionable video classification system. In *Proc. 8th International Conference in Advanced Communication Technology (ICACT)*, 2006.
- [17] A.P.B. Lopes, S.E.F. de Avila, A.N.A. Peixoto, R.S. Oliveira, M. de M. Coelho, and A. de A. Araujo. Nude detection in video using bag-of-visual-features. In *Proc. XXII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI)*, pages 224 –231, 2009.



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