ADScAI
Applied Data Science & Artificial Intelligence Symposium

Automated Censoring of Cigarettes and Liquor Drinking in Videos using Deep Learning Techniques

Jayapradha Perinparajah
Department of ICT
University of Vavuniya
Vavuniya, Sri Lanka
pradhajp2408@gmail.com

Malshi Imasha
Department of ICT
University of Vavuniya
Vavuniya, Sri Lanka
imashamalshi7@gmail.com

Dineskaran Navaratnarajah
Department of ICT
University of Vavuniya
Vavuniya, Sri Lanka
Dineskaran58@gmail.com

Thanushigan Panchalingam
Department of ICT
University of Vavuniya
Vavuniya, Sri Lanka
thanuthanush0325@gmail.com

Tharindu Shihan Honnanthara
Department of ICT
University of Vavuniya
Vavuniya, Sri Lanka
tharinduh67@gmail.com

Lahiru Lakmina
Department of ICT
University of Vavuniya
Vavuniya, Sri Lanka
lahirulakmina1999@gmail.com

Satkunarajah Suthaharan
Department of Physical Science
University of Vavuniya
Vavuniya, Sri Lanka
suthaharan@vau.ac.lk

Rukshani Puvanendran
Department of ICT
University of Vavuniya
Vavuniya, Sri Lanka
rukupuvan@vau.ac.lk

Exposure to content depicting smoking, alcohol consumption, and other addictive behaviors on social media platforms has been linked to an increase in youth engagement with these substances. This trend is concerning, as early exposure to such content can normalize substance use and lead to initiation among impressionable youth. Therefore, automatic censoring of such content is essential to ensure alignment with community standards and legal regulations, protecting users from exposure to inappropriate material. This research introduces A-Censor, an advanced deep learning system designed to automatically detect and censor smoking and alcohol consumption in videos, addressing critical public health concerns. The development process encompassed data collection, model training, and evaluation. In the data collection phase, a dataset of 3,000 images was assembled, comprising neutral, alcohol consumption, and smoking instances. Feature extraction was performed using MobileNetV2, while classification was conducted using algorithms such as Faster R-CNN, RNN, Gradient Boost, and SVM. The optimal model, Faster R-CNN with a MobileNet backbone, achieved a superior accuracy of 93.53%, outperforming other models. Following detection of smoking and alcohol consumption instances, a blurring technique is applied to obscure harmful content while preserving video quality. A-Censor offers an efficient, automated solution for content moderation, promoting a healthier digital environment.

***Keywords*—Automated Censorship, Deep Learning, Faster R-CNN, Mobile Net, Object Detection, Smoking Detection**

I. Introduction

The prevalence of smoking and alcohol consumption in media, particularly in videos, poses significant public health risks, especially for impressionable audiences such as adolescents. Research has shown that exposure to smoking in movies is linked to increased smoking initiation among young viewers [1][2]. Manual content moderation of such harmful depictions is often inefficient, requiring substantial human effort and resources. This highlights the need for automated solutions that can effectively detect and censor smoking and drinking scenes in videos [3][4].

This study aims to develop an automated censorship system using deep learning techniques to detect and obscure smoking and drinking scenes in videos, thereby protecting viewers from harmful content while maintaining video quality.

II. Related Works

Existing research in content moderation has increasingly focused on employing machine learning and deep learning techniques to identify harmful content in various media forms. Early approaches primarily utilized convolutional neural networks (CNNs) for detecting explicit content, including smoking and alcohol usage [1][2]. However, current systems face significant challenges related to scalability, precision, and real-time implementation. to overfitting, reducing the model's adaptability to unseen content variations [5]. However, CNN, R-CNN are known for its efficiency in object detection, allowing for the generation of region proposals and classification in a single framework [6]. By incorporating MobileNet, the model benefits from reduced computational complexity while maintaining high accuracy, making it suitable for real-time applications. Furthermore, this study explores various data augmentation methods to enhance the robustness of the training dataset, aiming to create a scalable and efficient system capable of accurately detecting and moderating smoking and

https://doi.org/10.31705/ADScAI.2025.08

alcohol consumption in diverse video contexts. Furthermore, the models struggle to maintain high accuracy across diverse video sources, often leading to false positives or negatives in content detection [3]. Recent studies have explored multi-task learning methods to improve content detection performance, but these approaches still encounter limitations in generalizing across different media types [4].

Our research addresses these gaps by leveraging with MobileNet as a backbone for feature extraction and classification algrithms suh as Faster R-CNN, SVM, RNN and gradient boost.

III. Methodology

A. Dataset Preparation

A dataset of 3,000 instances was categorized into three classes: Smoking, Drinking, and Neutral. The dataset was split into 70% for training, 15% for validation,and 15% for testing using train_test_split for balanced representation.

B. Augmentation

Data augmentation techniques such as rotation, zooming, flipping, and contrast adjustments were applied using Keras' ImageDataGenerator to enhance the diversity of the training dataset. These transformations help prevent overfitting by introducing variability into the training process, ensuring that the model generalizes well to unseen data.

C. Video Frame Extraction

Frames were extracted and resized to 224x224 pixels for accurate localization while preserving context.

D. Model Development

The MobileNetV2 was used as the backbone for feature exraction. Four models were evaluated:

- Faster R-CNN: Utilized a Region Proposal Network (RPN), trained with the Adam optimizer and fine-tuned anchor box sizes.
- Gradient Boosting Classifier: Trained on MobileNetV2-extracted features.
- Support Vector Machine (SVM): Applied MobileNetV2 feature extraction before training. Neural
- Recurrent Network (RNN): Implemented with LSTM layers for sequence classification.

Hyperparameters were optimized to improve detection performance and reduce overfitting. Adjusting the batch size influences memory utilization and learning stability, while fine-tuning the learning rate helps in achieving faster convergence. Finally, Gaussian blur was applied to detected smoking and drinking objects for content obscuration.

IV. Results and Discussions

The TABLE I depicts the model performance between four classification algorithms based on the metric such as accuracy, precision, recall and F1-score. The Faster R-CNN achieved the highest accuracy at 93.53%, along with strong precision, recall, and F1-score values, indicating its effectiveness in detecting smoking and alcohol scenes. Gradient Boost and SVM classifiers also performed well, with accuracies of 92.39%. Their precision, recall, and F1-scores were slightly lower than those of Faster R-CNN, suggesting they are competent alternatives but may not capture complex patterns as effectively. The RNN model exhibited the lowest performance among the evaluated models, with an accuracy of 86.79%. In summary, Faster R-CNN demonstrated superior performance in detecting smoking and alcohol scenes, making it the most suitable model among those evaluated. However, Gradient Boost and SVM also showed promise and could be considered depending on specific application requirements.

Model	**Acc. (%)**	**Preci.(%)**	**Rec.(%)**	**F1(%)**
Faster R-CNN	93.53	92.80	93.10	92.95
Gradient Boost	92.39	91.50	92.10	91.80
RNN	86.79	85.90	86.30	86.10
SVM	92.39	91.20	92.00	91.60

TABLE I
Model Performance Metrics

The data collection phase is a time-intensive process that often fails to encompass the full spectrum of real-world variations. This limitation can hinder the performance and generalizability of detection models. To address these challenges, We have planned to expand the dataset including a broader range of instances, thereby enhancing the model's ability to generalize across different contexts. Additionally, optimizing model performance through advanced algorithms and fine-tuning techniques can further improve detection accuracy.

V. Conclusion

This research developed an automated censorship system for detecting and moderating smoking and alcohol scenes in videos. The Faster R-CNN with MobileNet backbone achieved 93.53% accuracy, surpassing the other models. Real-time blurring effectively obscured harmful content without significantly affecting video quality. Further, the current system focuses on visual detection, and hence, integrating audio or subtitle analysis through multimodal approach could reduce misclassifications.

References

[1] Sukthankar, "Violence detection in video using computer vision techniques," Lecture Notes in Computer Science, vol. 6855, no. 2, pp. 332–339, Springer Berlin Heidelberg, 2011. doi: 10.1007/978-3 642-23678-5_39.

[2] P. Zhou, Q. Ding, H. Luo, and X. Hou, "Violent interaction detection in video based on deep learning," Journal of Physics: Conference Series, vol. 844, no. 1, Jun. 2017. doi: 10.1088/1742-6596/844/1/012044.

[3] N. F. Khan, A. Hussain, A. Khan, and I. U. Din, "Categorized violent contents detection in cartoon movies using deep learning model: mobile net," in Proc. 5th Int. Conf. on Next Generation Computing, 2019.

[4] A. Salekin, M. Rashid, M. M. Rahman, and J. Almhana, "Deep Puff: Censoring via Machine Learning Cigarette Use," in Proc. 12th Int. Conf. on Developments in eSystems Engineering (DeSE), Kazan, Russia, Oct. 2019, 10.1109/DeSE.2019.00056. pp. 556-561.

[5] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv preprint arXiv:1804.02767, 2018.