International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Special Issue: 1 Volume No.7 April – 2023
DOI:10.46647/ijetms.2023.v07si01.045 ISSN: 2581-4621

A Deep Learning Based Approach for Inappropriate content Detection and Classification of YouTube Videos

[1]Mr. B.Ramana Reddy, [2]N.Harshitha Chowdary, [3]K.Bhumika, [4]S.Ajay Kumar, [5]A.Chenchu Lakshmi
[1]Assistant Professor, Department of CSE, AITS, Tirupati
[2,3,4,5]Student, Department of CSE, AITS, Tirupati

Abstract
The exponential growth of videos on YouTube has attracted billions of viewers among which the majority belongs to a young demographic. Malicious uploaders also find this platform as an opportunity to spread upsetting visual content, such as using animated cartoon videos to share inappropriate content with children. Therefore, an automatic real-time video content filtering mechanism is highly suggested to be integrated into social media platforms. In this proposed system, a novel deep learning-based architecture is proposed for the detection and classification of inappropriate content in videos. For this, the proposed framework employs an ImageNet pre-trained convolutional neural network (CNN) model known as EfficientNet-B7 to extract video descriptors, which are then fed to bidirectional long short-term memory (BiLSTM) network to learn effective video representations and perform multiclass video classification. An attention mechanism is also integrated after BiLSTM to apply attention probability distribution in the network. These models are evaluated on a manually annotated dataset of cartoon clips collected from YouTube videos. Experimental results demonstrated that EfficientNet-BiLSTM (accuracy D 95.66%) performs better than attention mechanism based EfficientNet-BiLSTM (accuracy D 95.30%) framework. Secondly, the traditional machine learning classifiers perform relatively poor than deep learning classifiers. Hence the CNN with EficientNet-BiLSTM achieved better results in child inappropriate video content detection and classification.
Keywords: Deep learning, social media analysis, video classification, bidirectional LSTM, CNN, EfficientNet

Introduction

The creation and consumption of videos on social media platforms have grown drastically over the past few years. This creates opportunities for malicious users to indulge in spamming activities by misleading the audiences with falsely advertised content. In an attempt to provide a safe online platform, laws like the children's online privacy protection act (COPPA) imposes certain requirements on websites to adopt safety mechanisms for children under the age of 13. YouTube has also included a ''safety mode'' option to filter out unsafe content. Apart from that, YouTube developed the YouTube Kids application to allow parental control over videos that are approved as safe for a certain age group of childrenControlling the unsafe content phenomena, disturbing videos still appear even in YouTube Kids due to difficulty in identifying such content. An explanation for this may be that the rate at which videos are uploaded every minute makes YouTube vulnerable to unwanted content. Besides, the decision-making algorithms of YouTube rely heavily on the metadata of video (i.e., video title, video description, view count, rating, tags, comments, and community flags). Hence, filtering videos based on the metadata and community flagging is not sufficient to assure the safety of children. Prior techniques addressed the challenge of identifying disturbing content (i.e., violence, pornography, etc.) from videos by using traditional hand-crafted features on frame-level data. In recent years, the state-of-the-art performance of deep learning has motivated researchers to employ it in image and video processing. The most frequent applications of image/video classification employed the convolutional neural networks. Apart from that, the long-short term memory (LSTM), a special type of recurrent neural network (RNN) architecture, has

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Special Issue: 1 Volume No.7 April – 2023
DOI:10.46647/ijetms.2023.v07si01.045 ISSN: 2581-4621

proven to be an effective deep learning model in time-series data analysis. Hence, this study targets the YouTube multiclass video classification problem by leveraging CNN (EfficientNet-B7) and LSTM to learn video effective representations for detection and classification of inappropriate content.

Related Work

1. Material and Methods:

1.1 Deep Learning Method:

In contrast to machine learning algorithms, there is a growing trend of using deep learning architectures to learn the video-based feature representations in video classification. The proposed methodology provides a system to resolve theproblem of disturbing content in videos. This work employs deep learning architecture which has already been applied successfully in several applications for video classification problems. As shown in Fig. 2, the proposed system is dividedinto three main classifications namely (1) video preprocessing, (2) deep feature extraction, and (3) video representation andclassification. In the video preprocessing stage, the collected YouTube videos are preprocessed to remove all irrelevant or missing video information. It also rescales the extracted frames of each video clip into fixed dimensions. The preprocessed video frames of each video clip are for- warded as an input to an ImageNet pre-trained EfficientNet-B7 model for feature extraction. The extracted features are experimented with the BiLSTM network to learn effective video representations, which subsequently passed to the fullyconnected and softmax layers for final video classification.

2. BiLSTM NETWORK

The recurrent neural networks produce good network per- formance in modeling the hidden sequential patterns of time-series data. However, the vanishing gradient problem hampers an update of network parameters during the back- propagation process. It is usually resolved by using two variations of RNNs, which are: LSTM and gated recurrent unit (GRU). Conceptually, the network structure of LSTM is as same as RNN, but a special unit ‘‘memory cell’’ is introduced in LSTM to replace the update process of RNN. The memorycell of LSTM maintains information for a longer duration of time.

3. SOFTMAX CLASSIFIER:

The softmax activation function in an output layer of any deep learning model is considered as a softmax classifier.To classify each video clip into one of three classes (i.e., fantasy violence, sexual-nudity and safe), the proposed modelintegrated a softmax activation function in the last fully connected layer to determine the relative probability of three output units.

4. EVALUATION METRICS:

The performances of multiclass video classification models are evaluated by calculating the accuracy, precision, recall and f1 score using confusion matrices. Accuracy is the ratio of number of correct predictions for each class to the total number of predictions of all classes, and is calculated as:

$$Accuracy = \frac{1}{N}\sum_{c=0}^{N-1}\frac{(T_p^c + T_N^c)}{(T_p^c + T_N^c + F_p^c + F_N^c)} \times 100\%$$

In equation, c represents a particular class index from N number of classes, Tp denotes the true positives, TN denotes true negatives, Fp denotes false positives and FN denotes false negatives. Precision is the ratio of total number of correct predictions of positive instances to the total number of predictions with positive instances. It is calculated as:

$$Precision = \frac{1}{N}\sum_{c=0}^{N-1}\frac{T_p^c}{(T_p^c + F_p^c)} \times 100\%$$

The recall (also known as sensitivity) is th he ratio of total number of correct predictions of positive instances to the total number of instances in an actual class. The recall and f1 score are calculated by

DOI:10.46647/ijetms.2023.v07si01.045 ISSN: 2581-4621

using equations

$$Recall = \frac{1}{N}\sum_{c=0}^{N-1}\frac{T_p^c}{(T_p^c + F_N^c)} \times 100\%$$

$$F1Score = 2 * (\frac{Precision * Recall}{Precision + Recall})$$

4. Result:

In this process, the results obtained through experimental evaluations of different machine learning and deep learning approaches for video classification are presented. At first, three pre-trained convolutional neural network models including EfficientNet-B7 are employed as video classifiers to determine the performances of these ImageNet pre-trained CNN models in our multiclass video classification problem. For each model, the last three layers of the pipeline are discarded and added with a fully connected layer of softmax activation function. The transfer learning approach is implemented in a manner where weights of all layers in the model are fixed except the last fully connected layer. After training each pre-trained convolutional neural network model using the transfer learning approach, as shown in Table 3, it is analysed that the EfficientNet-B7 model performs comparatively better than VGG-19 and Inception-V3 on the YouTube cartoon video dataset. It has achieved the highest recall score which means that the EfficientNet-B7 model retrieves morerelevant instances than the remaining two pre-trained CNN models. Hence, further experiments are carried out with theEfficientNet-B7 as a base classifier.

5. Conclusion:

In this paper, a new method was suggested to improve the detection of inappropriate content in youtube videos using deep learning. Transfer learning using EfficientNet-B7 architecture is employed to extract the features of videos. The extracted video features are processed through the BiLSTM network, where the model learns the effectivevideo representations and performs multiclass video classification. It can assist any video sharing platform to either remove the video containing unsafe clips or blur/hide any portion with unsettling frames. It may also help in the development of parental control solutions on the Internet through browser extensions where child unsafe content can be filtered automatically.

6. References:

[1] J. Marsh, L. Law, J. Lahmar, D. Yamada-Rice, B. Parry, and F. Scott, *Social Media, Television and Children*. Sheffield, U.K.: Univ. Sheffield, 2019. [Online]. Available: https://www.stac-study.org/downloads/ STAC_Full_Report.pdf

[2] L. Ceci. *YouTube—Statistics & Facts*. Accessed: Sep. 01, 2021. [Online]. Available: https://www.statista.com/topics/2019/youtube/

[3] M. M. Neumann and C. Herodotou, "Young children and YouTube: A global phenomenon," *Childhood Educ.*, vol. 96, no. 4, pp. 72–77, Jul. 2020, doi: 10.1080/00094056.2020.1796459.

[4] S. Livingstone, L. Haddon, A. Görzig, and K. Ólafsson, *Risks and Safety on the Internet: The Perspective of European Children: Full Findings and Policy Implications From the EU Kids Online Survey of 9-16 Year Olds and Their Parents in 25 Countries*. London, U.K.: EU Kids Online, 2011. [Online]. Available: http://eprints.lse.ac.U.K./id/eprint/33731

[5] B. J. Bushman and L. R. Huesmann, "Short-term and long-term effects of violent media on aggression in children and adults," *Arch. Pediatrics Adolescent Med.*, vol. 160, no. 4, pp. 348–352, 2006, doi: 10.1001/arch- pedi.160.4.348.

[6] S. Maheshwari. (2017). *On YouTube Kids, Startling Videos Slip Past Fil- ters*. The New York Times. [Online]. Available: https://www.nytimes.com/ 2017/11/04/business/media/youtube-kids-

DOI:10.46647/ijetms.2023.v07si01.045 ISSN: 2581-4621

paw-patrol.html
[7] C. Hou, X. Wu, and G. Wang, "End-to-end bloody video recognition by audio-visual feature fusion," in *Proc. Chin. Conf. Pattern Recognit. Comput. Vis. (PRCV)*, 2018, pp. 501–510, doi: 10.1007/978-3-030-03398-9_43.
[8] A. Ali and N. Senan, "Violence video classification performance using deep neural networks," in *Proc. Int. Conf. Soft Comput. Data Mining*, 2018, pp. 225–233, doi: 10.1007/978-3-319-72550-5_22.