International Journal of Reconfigurable and Embedded Systems (IJRES)
Vol. 99, No. 1, Month 2099, pp. 1∼1x
ISSN: 2089-4864, DOI: 10.11591/ijres.v99.i1.pp1-1x

Real-Time Obscene Scene Nudity Detection and Blurring in a Video Clip

Jesima Rizwana[1], Md. Fahim Hasan[2], Motakabbir Hossain[3], Kaniz Ferdous Binte Zahangir[4], Humaira Mir Prothoma[5], Mr Md. Tanzim Reza[6], Dr Jia Uddin[7]
[1-5]Department of Computer Science and Engineering, BRAC University, Dhaka, Bangladesh

Keywords:

Video
Obscenity
Neural Network
Nudity Detection
YOLOv5
YOLOv6
Faster R-CNN
OpenCV
PIL
Image Processing
Concealment
Blur

ABSTRACT

Videos are widely consumed by people of all ages as a form of entertainment, information and education. However, not all videos are made for everyone. Many videos contain obscenities such as nudity, violence, blood, and gore which should not be watched by children or people who feel repulsed by these obscenities. Obscene content can negatively affect a child's mindset, and it can even traumatize people with weak mental constitutions. The real problem begins when these obscene videos are publicly available on the Internet, and anyone can watch them easily by downloading or streaming them online without getting any kind of warning. Moreover, people can even encounter these obscenities on live video streams or video calls. In our research, we have worked to detect and blur nude and obscene sexual content from videos in real-time. In that respect, this paper proposes a Neural Network-based approach. We have detected whether sexually explicit content is present in a video or not and blurred only the detected contents from the video frames. To detect nude and obscene contents, we have used different object detection algorithms such as Faster R-CNN, YOLOv5 and YOLOv6. These three respectively gave us mean average precision values of 0.382, 0.663 and 0.508 at 0.5 IOU threshold. Although with an mAP value less than YOLOv5, we chose YOLOv6 as it has proved to be the most optimal for our solution in regards of both accuracy and speed. And to blur, we have tried a total of five methods provided by two image processing libraries, OpenCV and PIL. Among those, we have selected the averaging method of OpenCV since it has best suited our needs. Additionally, we have attempted to reduce the rate of false positives so that any decent content does not get incorrectly labelled as obscene. This detection and blurring of obscene contents will contribute to ensuring safety in internet browsing for everyone.

Corresponding Author:

Jia Uddin, Ph. D.

Assistant Professor
AI and Big data Department, Woosong University, Daejeon, South Korea
Associate Professor(On Leave)
Department of Computer Science and Engineering, BRAC University, Dhaka, Bangladesh
Email: jia.uddin@bracu.ac.bd, jia.uddin@wsu.ac.kr

1. INTRODUCTION

Due to technological advancement, increased the affordability of internet-enabled devices. Among all other things such as communication, education and entertainment, people spend a significant amount of time every day on the Internet and other platforms to watch video content [1]. According to Statista, in 2020, an adult person in the United States spent a staggering 312 minutes per day watching digital video on devices [2]. The usage of videos is diverse. Be it for the sake of education, work, gathering knowledge, or entertainment. Nevertheless, in the entertainment media nowadays, many videos created to entertain people contain sensitive and obscene content such as nudity and sex [6]. These contents, in recent days, have penetrated the mainstream media deeply. Many mainstream movies and web series contain these, which are inappropriate to watch for children and people with strict moral values. Moreover, most sexual video contents objectify women, which can adversely affect the minds of our younger generation.

According to Ward (2016), watching nudity and sexual content can result in ”higher levels of body dissatisfaction, greater self-objectification, greater support of sexist beliefs and adversarial sexual beliefs, and greater tolerance of sexual violence toward women” (p.1) [3]. Viewers of such media can also experience changes in their sexual beliefs [4] and preferences, for example, they might get influenced to desire a particular body shape or practice sexual behaviours shown in these videos [5].

Videos containing sexual content and nudity are more widespread on the Internet and easily accessible, which puts children at great risk [7]. So, to protect the children from encountering these potentially harmful contents and ensure safety it is imperative that these contents are filtered and flagged [10]. Furthermore, if this can be done in real-time, it will help ensure protection against these obscene sexual contents even in live video streams and video calls [9]. So, our target is to develop an algorithm that can detect sexually explicit content from videos as well as blurring in real time, keeping the best possible accuracy [12].

The article from [8] investigated the shortcomings of existing convolutional neural network methodologies by focusing on the attentional control of CNN on the projected nude regions within the frames in order to reduce the FNR (False Negative Result). Furthermore, in the availability of diverse backgrounds, CNN's previous techniques ignored small-scale pornographic content. The ”You Only Look Once” (YOLO) algorithm was used in the pornography and nudity detection algorithms to identify individuals as small regions of interest (ROIs), which were then classified using methodologies [17]. Furthermore, it demonstrated that the ”You Only Look Once” (YOLO) object analyzer outperforms the YOLO approach significantly. Many evaluations were performed on the previously mentioned dataset to compare the contributions of various YOLO algorithms [16]. A resection survey was also carried out to demonstrate the impact of YOLO inclusion. YOLO outperformed the network in terms of accuracy by only 85.6% to 89.5%. Furthermore, when tested on the data set and model, the motion-based color filter achieves an overall accuracy of 93% [18]. Each user anticipates receiving three consecutive screenshots in 879 nanoseconds.

According to the report in [14], a system for detecting inappropriate scenes in video content, such as nudity, drugs, gore, and so on, is proposed. The process was completed in three steps, the first of which was to convert the videos into multiple frames [19]. Then, three different algorithms were used to detect inappropriate scenes, and the percentage of inappropriate scenes was calculated and displayed as a result. To detect objects and scenes, YOLO's Object Detection algorithm was used. Nudity detection was assisted by Python's nudepy library, which works by detecting skin-colored pixels and identifying nudity based on pixel count and region [15]. Furthermore, with the model's assistance, any video that exceeds the model's base scale can be detected and classified as pornographic [13]. The devices used to develop the framework demonstrated an estimated 90% accuracy in determining both the nudity and violent content of the video for nudity detection.

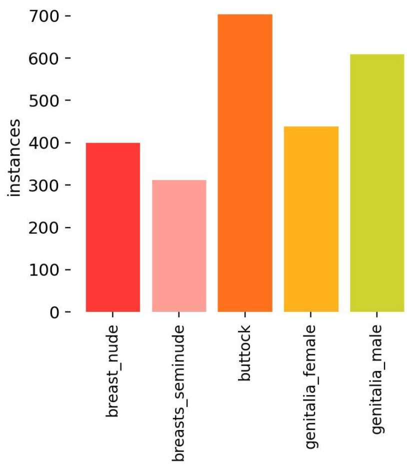
A study from [21] was conducted to evaluate the performance of various image blurring methodologies in decreasing the risk of recognition system. The personal data blurring procedure was discovered to be more useful than minimizing facial classification when compared to the Gaussian blurring method, the Box Blurring procedure, and a various private information color fringing method.

Another method suggested in [22] is a simple method for simply acknowledging and characterizing blurred zones in images. This applies to a variety of audio and visual designing approaches, such as feature extraction, depth information appraisal, and pattern classification. This method has a 73.6% accuracy [20].

2. METHOD

We were usually required to label the dataset in order to train and analyze the models [23]. Initially we have collected 3500 images from various platforms. To ensure adequate comparability with the base study, we kept the research's training, validation, and testing sets. For training, validation, and testing, the dataset was divided into three equal segments. As a result, 2450 photos were allotted to the training segment, 700 to the validation segment, and 350 to testing. We utilized many photos to be labeled using bounding boxes, resulting in the most widely used training set. In Figure 1(a), it is a demo for labelling and annotation. In Figure 1(b), there is a bar chart for classes of the objects that we used for detection and blurring.

(a)



(b)

Figure 1. Labelling Images for Datasets (a) and After Giving The Datasets To Train The Codes (b)

Also, after initially researching and trying detection with YOLOv3, YOLOv4 and Faster R-CNN (Detectron2), we stepped into working on YOLOv5 and YOLOv6 as well. And for having the best fitted accuracy and speed in comparison to the previously tested models and algorithms, we finally did the whole work on YOLOv5 and YOLOv6.

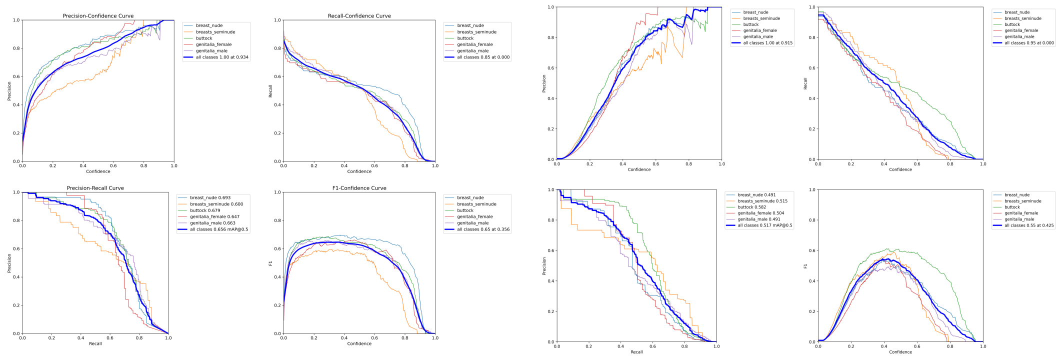
3. RESULTS AND DISCUSSION

The proposed model for our work based on detection and blurring nudity in a video clip has been described in this section. The accuracy and the loss were calculated while training our models on the dataset over 100 epochs. In the dataset, all of our pictures were resized to 640×640 pixels. While training on the dataset, the mAP@0.5 values reached 0.6626 and 0.5083 respectively for YOLOv5 and YOLOv6 (Figure 3(a), 3(b), 4(a) and 4(b). In this section the blurring result of the proposed model has been described. Different blurring methods were used to find out the fastest method. The models were then implemented in entirely new video and images for testing purposes.

In this section, results of the detection and results of the blurring times of the proposed models have been described. Different blurring methods were used to find out the fastest method. The models were then implemented in entirely new video and images for testing purposes.

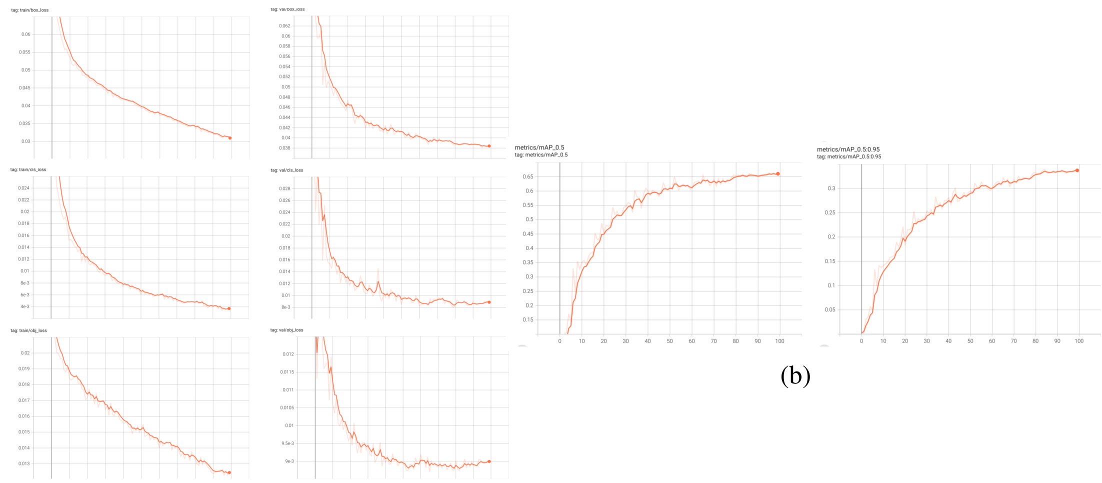
3.1. YOLOv5 and YOLOv6 - Detection and Results

In the figures, we can see the Precision-Confidence Curve, Recall-Confidence Curve, Precision-Recall curve, F1-Confidence Curve. For precision, recall and F1 score, greater values indicate that the model can detect more accurately(Figure 2(a) and Figure2(b)). For the Precision-Recall Curve, Area Under Curve (AUC) is considered. A better model has a larger AUC. From the curves shown in figure, it is evident that YOLOv5 clearly outperforms YOLOv6 in terms of detection. However, YOLOv6 proved itself to be significantly faster than YOLOv5.



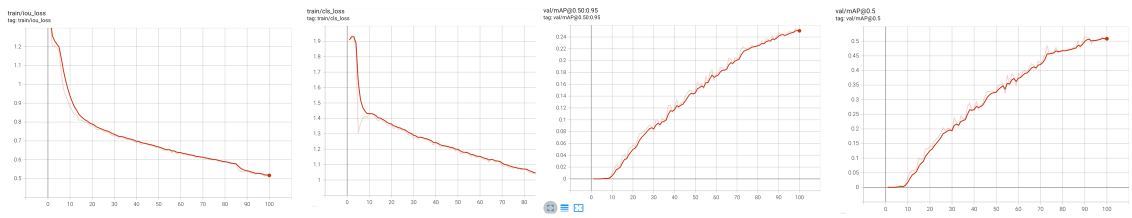
(a) (b)

Figure 2. YOLOv5 P, R and PR Curve (a) and YOLOv6 - P, R, PR and F1 (b)



(a) (b)

Figure 3. YOLOv5 Train Loss and Validation Loss (a) and YOLOv5 mAP (b)



(a) (b)

Figure 4. YOLOv6 Training Loss and Validation Loss (a) and YOLOv6 mAP(b)

3.2. Blur Times and Results - YOLOv5 and YOLOv6

As per our concern on the work to be done in real-time, faster detection and faster blurring was encouraged. Even if the detection was faster in YOLOv6, the blurring (with all necessary methods, e.g. Gaussian, Median) created a bit of buffering issues. To find the fastness, we tried using 5 different blurring methods of OpenCV and Python Imaging Library(PIL):

Table 1. Blur Timings - OpenCV/YOLOv5

OpenCV		
Median Blur	Gaussian Blur	Average Blur
14.09757137	21.69613719	12.7930088
15.43405342	21.67938781	13.20445609
15.29584622	22.4433012	13.84958696
14.82680988	23.06503415	13.21205378
14.71921921	22.1556921	13.54963064
14.87470002	22.20791049	13.32174726

Table 2. Blur Timings - PIL/YOLOv5

PIL	
Gaussian Blur	Box Blur
13.463434934616	14.8195593357086
14.3288419246673	13.9096252918243
14.7272689342498	13.9061765670776
15.4157016277313	14.6888945102691
14.1470916271209	12.6811256408691
14.41646781	14.00107627

Table 3. Blur Timings - OpenCV/YOLOv6

OpenCV		
Median Blur	Gaussian Blur	Average Blur
12.49306393	30.66199422	9.415367842
12.4664228	31.84477472	8.666805267
12.51437354	29.78105116	9.191741467
12.05743599	30.22601199	8.482309103
11.64797616	28.93356848	8.605831385
12.23585448	30.28948011	8.872411013

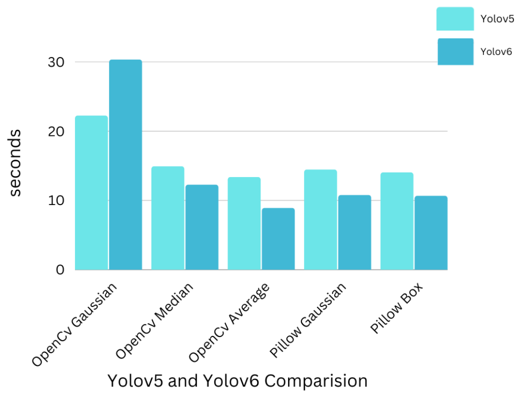
Table 4. Blur Timings - PIL/YOLOv6

PIL	
Gaussian Blur	Box Blur
10.64559317	10.69285965
10.29630065	11.28603435
10.91101742	9.702563047
11.33054686	9.906082153
10.54887247	11.5955081
10.74646611	10.63660946

From Table 1, 2, 3 and 4, we see that the blurring times consisting of two sections for OpenCV and PIL. We have trained these methods for both YOLOv5 and YOLOv6 respectively.

3.3. Blurring Results and Comparison Between YOLOv5 and YOLOv6

The tables given (Table 1, 2, 3 and 4) illustrate the performance of each model on a same 7seconds sample video. By comparing the average of the results our goal was to find out the fastest blurring method possible.



(a)

(b)

Figure 5. Speed Comparison Bar Chart (a) and Detection and Blur Comparison Between YOLOv5 and YOLOv6 Model Training (b)

Combination of YOLOv6 and OpenCV average blur worked faster in comparison to both models. In the bar chart above (Figure 5), the lowest bar stands for the faster speed in case of blurring. Our limitation was not having access to a better GPU, due to which we were facing some timing issues in blurring speeds. Lastly, from (Figure 5(b)), we get a comparison between the visualisation of blurred images after running those through our models.

4. CONCLUSION

There are countless videos on the Internet, and to know whether a video is safe to watch or not, we need to determine if the video in question contains obscene content. On top of that, real-time detection and blurring will be beneficial to safeguard the users from obscenities displayed during live video streams and video calls. Therefore, we have made improvements on the existing research that has been previously done on detecting obscene content from video clips and develop an algorithm using YOLOv6 and OpenCV to achieve real-time detection and blurring of obscene content without giving up on the accuracy. With a more specific approach, this can be utilized to censor unnecessary vulgarities in mainstream media so that more people can experience them without encountering unwanted obscene content. This can make the Internet a safer environment for everyone, especially children. With a larger and more specific dataset, we plan to refine and improve this research even further Our goal from this work is to develop our research skill to a greater extent and to put more contribution to the related field.

ACKNOWLEDGEMENTS

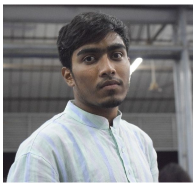
By the Grace of the Almighty, by the inspiration, help, support and motivation by our respected Supervisor and Co Supervisor, we were blessed to complete this conference paper with fruitful research and thorough dedication. We also look forward to work more and take our research to the next steps with more elaborating knowledge and studies. Our goal and motto is to put further contribution to the relevant field and increase our datasets to work on new models for upgrading. Also, we shall look for funding to take it to the professional fields in future.

REFERENCES

[1] Turner, A. (2022, January 3). How Many People Have Smartphones Worldwide (Jan 2022).
[2] Stoll, J. (2021, July 14). U.S. video consumption by device 2023. Statista.
[3] Ward, L. M. (2016). Media and Sexualization: State of Empirical Research, 1995–2015. The Journal of Sex Research, 53(4–5), 560–577.doi:10.1080/00224499.2016.1142496.
[4] Collins, R., Strasburger, V., Brown, J., Donnerstein, E., Lenhart, A., & Ward, L. (2017). Sexual Media and Childhood Well-being and Health. Pediatrics, 140 (Supplement2), S162-S166. doi:10.1542/peds.2016-1758x.
[5] Flood, M. (2009). The harms of pornography exposure among children and young people. Child Abuse Review, 18(6), 384–400. doi:10.1002/car.1092.
[6] Karavarsamis, S., Ntarmos, N., Blekas, K., and Pitas, I. (2013). Detecting Pornographic Images by Localizing Skin ROIs.
[7] Wehrmann, J., Simões, G., Barros, R., Cavalcante, V. (2018). Adult content detection in videos with convolutional and recurrent neural networks.
[8] AlDahoul, N.; Abdul Karim, H.; Lye Abdullah, M.H.; Ahmad Fauzi, M.F.; Ba Wazir, A.S.; Mansor, S.; See, J. Transfer Detection of YOLO to Focus CNN's Attention on Nude Regions for Adult Content Detection. Symmetry 2021, 13, 26. https://doi.org/10.3390/sym13010026.
[9] Meek, Allen. (2009). Trauma and Media : Theories, Histories and Images. 172-173.
[10] Sriram, R. (2020, June 24). Why ages 2-7 matter so much for brain development. Edutopia.
[11] Geethapriya. S, N. Duraimurugan, S.P. Chokkalingam. (2019). Real time object detection with YOLO.
[12] Coulson, J. (2015). The Problem with Exposing Kids to Sexual and Violent Content [Blog].
[13] Ross, C. (2012). Overexposed and Under-Prepared: The Effects of Early Exposure to Sexual Content [Blog].
[14] Tofa, K. N., Ahmed, F., Shakil, A. (2017). Inappropriate scene detection in a video stream.
[15] Ap-apid, R. (2005). An Algorithm for Nudity Detection.
[16] Xing, X., Liang, Y., Dang, J., Huang, S., Han, R., & Liu, X. et al. (2011). Safe Vchat: Detecting Obscene Content and Misbehaving Users in Online Video Chat Services. doi:http://dx.doi.org/10.1145/1963405.1963501.
[17] Behrad, A., Salehpour, M., Ghaderian, M., Saiedi, M., & Barati, M. (2012). Content-based obscene video recognition by combining 3D spatiotemporal and motion-based features. EURASIP Journal On Image And Video Processing, 2012(1). doi:10.1186/1687-5281-2012-23.
[18] Lee, J. S., Kuo, Y. M., Chung, P. C., & Chen, E. L. (2007). Naked image detection based on adaptive and extensible skin color model. Pattern recognition.
[19] Islam, R. S., Siddiqui, R., Roy, D. (2019). Blurring of inappropriate scenes in a video using image processing.
[20] More, M. D., Souza, D. M., Wehrmann, J., & Barros, R. C. (2018, July). Seamless nudity censorship: an image-to-image translation approach based on adversarial training. In 2018 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.
[21] Pulfer, E. (2019). Scholarworks@UARK.
[22] Su, B., Lu, S., Tan, c. (2011). Blurred Image Region Detection and Classification. Research Gate. doi: 10.1145/2072298.2072024
[23] Yuan, K., Tang, D., Liao, X., Wang, X., Feng, X., Chen, Y., ... Zhang, K. (2019, May). Stealthy porn: Understanding real-world adversarial images for illicit online promotion. In 2019 IEEE Symposium on Security and Privacy (SP) (pp. 952-966). IEEE.

BIOGRAPHIES OF AUTHORS

Jesima Rizwana
Student of BRAC University, Department of Computer Science and Engineering. This was her final semester and by submitting her undergraduate thesis, she shall be graduated by this Summer 2022. She can be contacted at email: jesima.rizwana@g.bracu.ac.bd, jesimarizwana1999@gmail.com



Md. Fahim Hasan
Student of BRAC University, Department of Computer Science and Engineering. He can be contacted at email: md.fahim.hasan@g.bracu.ac.bd, fhsourav@gmail.com

Motakabbir Hossain
Student of BRAC University, Department of Computer Science and Engineering. This was his final semester and by submitting his undergraduate thesis, he shall be graduated by this Summer 2022. He can be contacted at email: motakabbir.hossain@g.bracu.ac.bd, rohithossain008@gmail.com

Kaniz Ferdous Binte Zahangir
Student of BRAC University, Department of Computer Science and Engineering. She can be contacted at email: kaniz.ferdous.binte.zahangir@g.bracu.ac.bd, kaniz.ferdous20@gmail.com

Humaira Mir Prothoma
Student of BRAC University, Department of Computer Science and Engineering. She can be contacted at email: humaira.mir.prothoma@g.bracu.ac.bd, humairamirprothoma@gmail.com