Additional learning on object detection: a novel approach in pornography classification*

Hoang-Loc Tran(✉)[1,2], Quang-Huy Nguyen[1,2], Dinh-Duy Phan[1,2]
, Thanh-Thien Nguyen[1,2], Khac-Ngoc-Khoi Nguyen[1,2], and Duc-Lung Vu(✉)[1,2]

[1] University of Information Technology
[2] Vietnam National University Ho Chi Minh City
Thu Duc Dist. Ho Chi Minh City, Vietnam
locth@uit.edu.vn, 15520306@gm.uit.edu.vn,
{duypd,thiennt}@uit.edu.vn, 15520386@gm.uit.edu.vn,
lungvd@uit.edu.vn

Abstract. In this paper, we proposed a new approach for pornographic classification by recognizing sensitive objects on images. To handle the misdetection and wrong judgment, a novel training strategy named additional learning was developed to help object detection model learns from mistakes, therefore increasing the method performance. Furthermore, a separate SVM classifier was trained to classify pornography and benign images from sexual object detected using Mask R-CNN model. Benchmarked by the NPDI-800 dataset, our proposed method achieved an accuracy of 84.625% and 90.125%, before and after applying additional learning strategy respectively. Besides, our proposed model also improves the false positive rate from 22.16% to 3.56% in our manually collected dataset.

Keywords: Computer vision · Image processing · Object detection algorithms · Pornographic content recognition and classification.

1 Introduction

Pornography has been a long-lasting problem in society nowadays. Not only it brings a huge negative impact toward the young generation but also provokes the ratio of commit crime, especially sexual crime. Although extensively studied and researched, recognizing pornographic images and videos remains a challenging problem in computer vision. The difficulties of addressing pornographic content often due to various factors, such as people's definition of pornography, the variation exhibited in scenario, person pose, background, lightning, scales, especially the highly similar between some certain pornographic images and normal ones.

To address the problem, various methods have been proposed based on different definition toward pornography. With the base hypothesis that pornographic image often portray a large ratio of explicit body, skin based approaches focus on determining the ratio of skin and face of the image, therefore conclude the safety of that image based on a custom threshold. Low level-based approaches develop description model to specify

* Supported by Vietnam National University Ho Chi Minh City

a visual codebook represent hand-crafted features extracted from pornographic images, then using a classifier model such as SVM to classify pornographic image under the robustness of machine learning algorithms. One of the finest approaches in pornographic visual content identification problem belongs to deep learning approaches, modifying pre-train neural networks models to learn the global features from obscene images automatically. However, these approaches comes with some certain weakness. Determining explicit pornographic images from skin and face ratio have been known for the large ratio of misdetection and wrong judgement [1], while the various features that describe pornographic images make it quite a challenging problem to determine appropriate features presenting visual low-level codebooks [2]. While deep learning method is the state-of-the-art, the lack of public pornographic dataset and the rich diversity and complexity of pornographic content that make the creation of a ”standard” dataset with quantity and quality for training pornographic model become difficult. By the lack of data for training, many deep learning method often suffer from high rate false-positive classification [3].

As we recognized, there is a significant part of pornographic content describe sexual body parts, textually and visually. Extremely rarely it has a pornographic content without sexual objects description, as these parts play a significant role to cause people’s euphoria. Hence, recognizing sexual objects in content means identifying pornography. In other words, detecting sexual objects is one of the root of the pornographic content classification problem. In the light of that conclude, our study proposed a method that focused on recognizing pornographic image and video-based on detecting sensitive objects such as anus, breast, or genitals. Our proposed method included all the advantages of the existing methods, which are modifying the neural network model to study low-level features from images and identifying obscene objects on images automatically. After that, to transfer from pornographic object detection to pornographic classification problem, a separated discriminative model SVM was trained based on the predicted results from object detection model to identify sensitive images.

One of the most crucial problem of object detection is wrong detection i.e. the false-positive rate. To deal with wrong detection problem as well as to improve the performance of our detection model, we proposed a binary-training-phase strategy called additional learning. The initial training phase worked as a classic strategy which learned to recognize sexual objects, and the additional training phase, which is our novel approach, served to learn again with the combination between the original training set and false recognized objects, thus improved the total accuracy and strengthened the model’s performance.

In summary, this paper proposed a method to detect pornographic content based on human sensitive objects. It took images as input and gave classification results as output which let us know if an image is benign or pornographic. To discriminate pornographic content from the normal one, the proposed method focused on detecting human sensitive objects. If the model found anus, women nude breast, or genitals in images, it identified that image as pornography and vice versa. This led to the root problem which we were dealing with turned into the sensitive object detection. To reduce the false-positive rate, this paper proposed the train-booster strategy which contain two phases of training. Finally, an SVM is applied to determine whether the content is pornography or not. In the experiment, our additional training strategy helped boost the total accuracy of Mask R-CNN model from 84.625% to 90.125% on the open NPDI-800 dataset [4], proving the effectiveness of our method. Our main contribution lies in the new strategy to reduce the false-positive rate hence boosting the performance of the model.

The rest paper is organized as follows. Section 2 describes in details some relevant approaches in pornographic detection. After that, the detail of our proposed method is presented in Section 3. Section 4 provides the details of our experiments and the results, while in section 5, we discuss our study and our future work in details.

2 Related Works

One of the earliest approaches for recognizing pornography image is calculated the ratio of human body skin exposure on that image, hence conclude the safety of the image. Since skin cell identification and segmentation involve color and texture information, color-spaces are often applied to identify whether the pixel is skin or not, therefore segment the areas of human skin. Balamurali et al. [5] proposed a binary model consist of skin extraction using YCbCr color space and face identification with Viola-Jones algorithms to calculate the ratio of face and skin areas on the image. If the percentage between human face and skin body lesser than 30% or the skin areas detected is half an image, this picture is declared as pornographic explicit and vice versa. Although Zaidan et al. [1] praised the skin-based approaches for its effectiveness in classifying between obscene and benign images, there are still some certain problems, mainly because of the quality and resolution that affect the performance of images. Moreover, wrong decisions can be made when it comes to images with skin-like objects or athletic images, which have a vast amount of exposed skin.

Another approach for recognizing pornographic images is using feature descriptors to extract hand-crafted features from images, then develop a classifier to determine whether the image is pornography or not based on extracted features. Avila et al. [4] presented a descriptor named BossaNova to represent the conceptual point of view in visual content. Extended from Bag-of-Visual-Work model, BossaNova computing histogram of distances between descriptors found in image and the visual dictionary in order to preserve essential information about the distribution of Hue-SIFT descriptor. Developing an open pornographic video dataset named NPDI-800, the research team extracted key-frames from every video and applied a binary model consist of BossaNova and SVM to detect labels for each video via a major voting scheme. Moreira et al. [6] introduced a space-temporal detector and descriptor called Temporal Robust Features which was a custom-tailor for effective and efficient description to extract local information on the image. Then, these features are aggregate into a mid-level representation using a state-of-the-art Bag-of-Visual-Words model named Fisher Vectors. To benchmark the method, an extension of NPDI-800 named NPDI-2k was developed.

Taking into account of the remarkable results achieved by deep learning architectures on various computer vision tasks, recent advances in pornography detection applied deep learning convolutional neural networks (ConvNet) to learn pornographic features from images automatically. Moustafa et al. [7] remodeled two pre-trained neural networks architectures AlexNet (ANet) and GoogLeNet (GNet) for pornographic images detection. By feeding a two-way Softmax with the output from the last layer of two architectures respectively, the models yielded a probability that pointed out whether the input image is pornographic or normal. Eventually, the two new architectures are combined together to developed new classifying models called AGNet and AGbNet with distinct metrics. In AGNet, the final score was determined by calculating the average between two probabilities, while in AGbNet, the result was choosing max. Mahadeokar et al. [8] developed an NSFW classification model by replicated ResNet 50-layer networks using half number of filters in each layer. For training the pre-trained residual network on ImageNet 1000 class dataset, the authors applied scaled augmentation to avoid over fitting. The last layer of every networks was replaced with two fully-connected layers and fine-tuned with an NSFW dataset. Furthermore, CaffeOnSpark, an open source deep learning framework developed by Yahoo, was applied for training NSFW model on Hadoop and Spark cluster. Vitorino et al. [9] adopted a transfer learning approach, training GoogLeNet model to classify a large scale everyday object dataset ImageNet. Then, the model was fine-tuned to learn pornographic features from NPDI-2k dataset. Finally, the last layer of the model architecture was replaced with an SVM with RBF kernel as a discriminative model to identify pornographic images. Wehrmann et al. [3] proposed ACORDE, a deep learning architecture comprised of separate ConvNet for image feature extraction and LSTM recurrent networks for sequence learning on video, thus optimise the performance of ACORDE on adult video.

Although the advantage of deep learning helps neural network models learn the common global features from explicit sexual images, these approaches do not identify sexual body parts specifically on images as a major part of pornographic photos often include sensitive organs. Fortunately, with the robust development of object detection algorithms, this is no longer a difficult problem. Shen et al. [2] developed a combine model called EFUI (Ensemble Framework using Uncertain Inference) consist of SSD [20] to identify sensitive semantic components included genital, breast, ass, nude body and sexual action. While SSD model has a slight possibility of false detection in practical, the training dataset is prepared with slight noisy data to incorporate the prior global confidence. Wang et al. [10] located female breast and sex organs on image using Multiple Instance Learning model, applying a generic pornographic content detector to recognise a pornographic image if it contains at least one exposed sexual organs. Comparing with the transitional pornographic classification methods such as image retrieval or bag-of-features, multi-instance generic detector achieved more accurate results. Notably, in Tabone et al. study [11], a multi-class ConvNet was applied to divide sensitive organs into five main classes: buttock, female breast, female genital, male genital and sex toy, which has a significant similar to the genital to be ignored. As female genital appearance was recognized quite different between obscene posing and sexual activities, this class was divided into two sub-classes: female genital posing and female genital active. Moreover, the authors proposed an extra class named benign to define the concept of normal body parts, thus strengthen the recognition performance.

3 Proposed Method

When examined the nature of the problem, we found that if an image is called as pornography, it must ”describe or show naked people and sexual acts in order to make people feel sexually excited”[3]. In other words, pornography must show off the sensitive objects of human. These objects usually are male or female genitals, breast, and anus. From this observation, sensitive body parts detection could be used as a potential approach for the pornography recognition problem. In general, the proposed method is presented in Figure 3. Different from the existing methods which usually trained the model only one time, the proposed method trained the model up to two times in binary-training phase. To obtain the sensitive object detector, Mask R-CNN was selected.

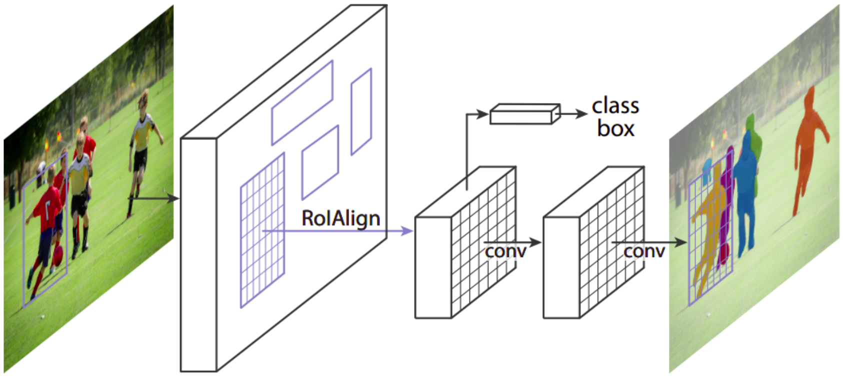
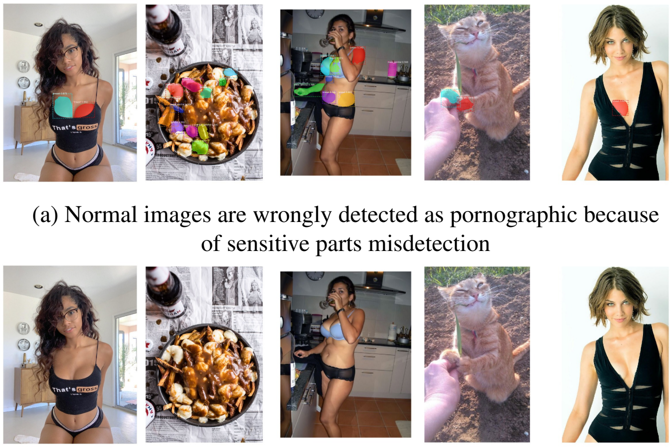


Fig. 1: Mask R-CNN main architecture [12]

Mask R-CNN [12] has been well known as one of the most effective and efficient method to deal with object detection and instance segmentation problem. Created as the extension of the famous architecture Faster R-CNN, Mask R-CNN can detect multiple objects with high performance. With the combination of Residual Network (ResNet) [13] and Feature Pyramid Network [14] architecture as the main backbone as well as the Fully Convolutional Network as an extended mask branch, Mask R-CNN is able to describe segmentation mask for object instances in the pixel-level. For the core operation of local features extraction, an unique pooling algorithms called RoIAlign [12] is applied to help Mask R-CNN to achieve the pixel-to-pixel alignment, therefore improves the bounding box localization and instance mask segmentation to predict object in high detail. The main architecture of Mask R-CNN can be observed in Figure 1.

The reason for choosing Mask R-CNN models for this paper mainly because it is a noticeable algorithm for both object detection and instance segmentation tasks. Our

[3] https://www.oxfordlearnersdictionaries.com/definition/english/pornography



(a) Normal images are wrongly detected as pornographic because of sensitive parts misdetection

(b) No sensitive object found in normal images

Fig. 2: (a) Results from initial training phase with no augmentation and (b) results after augmenting Mask R-CNN model with additional learning strategy

target of recognizing private body parts, not only because it could identify whether the image is safe or not, but also it could lead to censoring inappropriate visual content automatically in the future. Because of that, Mask R-CNN is the most suitable choice as it is considered to be one of the most significant models which have the ability to recognize objects in the pixel-level. However, due to the lack of data as well as the complexity of training parameters, Mask R-CNN model is quite easy to suffer from over-fitting, thus leads to wrong judgment and misdetection. To improve the prediction's performance, the proposed two-phase training strategy is developed alongside with model's parameters improvement. While the labeling process in pixel level for Mask R-CNN training is quite a tedious task, this is a crucial step for developing the automatic censoring pornographic visual content that we are currently working on.

The sensitive object detector, which had been trained by the Mask R-CNN model, was used to detect if human expose their sensitive part in an image, thus gave the conclusion whether that image is pornography or not. Not like other methods that usually stop training after achieving the model, our proposed method had an extra step to validate and re-train the model which called the additional learning phase.

The purpose of this additional training was that we wanted to improve the false-positive rate. Since Mask R-CNN had a great advantage in detecting sensitive objects, it was also too sensitive with normal images. This leads to the high rate of false positive when benign objects were wrongly detected as sensitive parts of human.

To deal with this problem, an additional verification set, which contained all normal images, was used to improve the performance of the sensitive object detector. As shown in Figure 3, if any normal images in the verification set were classified as pornography, which meant that normal objects have been detected wrongly in these images, it will be re-annotated with pseudo-negative-labels included wrong-male-genitals, wrong-

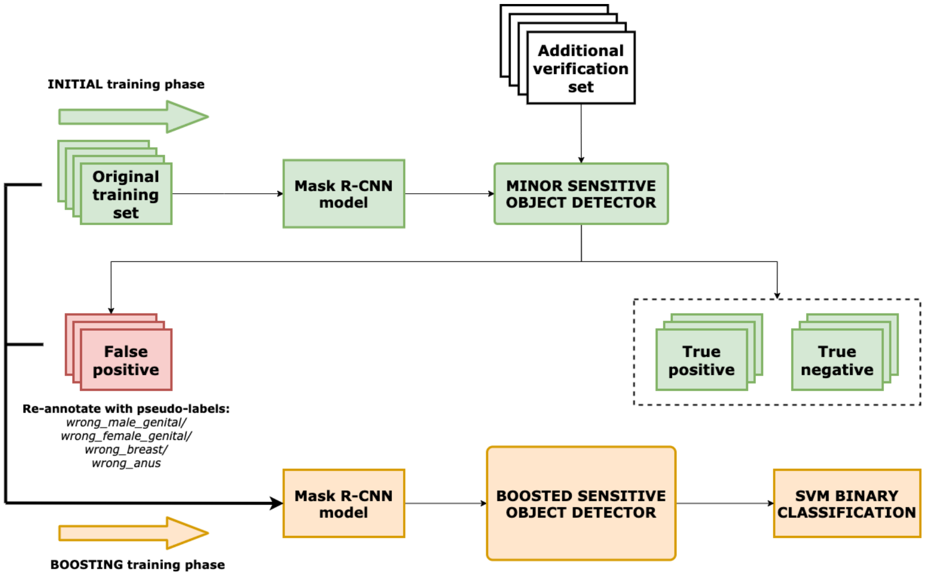


Fig. 3: Proposed method's training procedure

female-genitals, wrong-breast, and wrong-anus. Based on this proposal, there were a total of 8 labels in the additional training phase: four real labels to detect sensitive objects and the four corresponded pseudo-labels to avoid misdetection. Because the number of training class was different from the initial phase, the model had to be re-trained completely. The training set of the additional training phase is the combination of original training set and false-positive images from the initial validation set. This combination came with pseudo labels that helped the proposed model to discriminate features from pornographic images and normal ones hence reducing the false-positive rate and increasing the total accuracy.

The final step of the training phase has been done with the help of the Support Vector Machine [15] (SVM) binary classification model. Using the results from the sensitive object detector, which included the number of each kind of sensitive object and its highest confidence, the SVM finally trains a classifier that has the ability to distinguish pornography from normal images.

4 Experimental Results

4.1 Experiment Environment

To conduct experiments, the Google Colab Pro had been used to train and test the proposed model. In specific, the computation's hardware included the Intel(R) Xeon (R) CPU @ 2.30GHz came with 25GB RAM and NVIDIA Tesla P40 with 16GB memory. Python version 3.7 and TensorFlow version 1.15.2 had been used for the development environment.

Objects Name	Training	Validating	**Total**
male_genital	2055	799	**2854**
female_genital	2079	298	**2377**
breast	7740	949	**8689**
anus	1046	110	**1156**
pseudo_male_genital	2367	329	**2696**
pseudo_female_genital	715	111	**826**
pseudo_breast	6342	1280	**7622**
pseudo_anus	224	34	**258**

Table 1: Distribution of annotated objects of two training phases in total

4.2 Dataset processing

Initially, in the first training phase, a small annotated dataset was created by collecting nearly 8.000 explicit pornographic images from the internet. These images was then annotated in polygon mask for four sexual body parts, including female breast, male/female genitals, anus and split into two sub-sets namely Original Train and Original Val for Mask R-CNN training.

In the additional training phase, the Additional Verification set had been built with 40.000 images. This set included all non-porn images and was also divided into two sub-directories namely: (i) normal subset which contained 30.000 images from usual social life, (ii) sexy subset which contained 10.000 images that show human with sexy pose and exposed skin. Based on the false-positive prediction on that verification set, we extracted an pseudo set with over 7.000 images which are annotated with pseudo-class corresponded with the four original class from the first training phase. From the segmentation masks and labels predicted by the minor sensitive object detection, we converted them into polygon coordinate information as well as renamed these labels into pseudo respective with the original ones for the additional training phase. Then, the pseudo set was combined with the original training set to create the Combination Set for the next training stage, which included 12.850 images for training and 2.165 images for validating. The Combination Set consisted of 8 annotated class including 4 original explicit sexual classes and their corresponding pseudo class, which distribution are described in detail in Table 1. Eventually, a 40.000-testing set included 15.000 normal, 20.000 pornography and 5.000 sexy images was developed as the final measurement to evaluate the object detection model after the binary-phase training. The quantity of each image-set is described in Table 2.

4.3 Experiment Detail

In the proposed method, the training phase has been split into two phases which used two different image sets. The first training phase use the Original Set to train and validate. These images helped the Mask R-CNN model learn features to detect sensitive objects of humans includes male and female genital, breast, and anus. To implement the Mask R-CNN model, ResNet101 + FPN backbone had been used while the learning rate had been set to 0.001. We trained our model with 100 epochs which included

Name	Porn	Normal	Sexy	Pseudo	**Total**
Original Train	6.850	N/A	N/A	N/A	6.850
Original Val	1.107	N/A	N/A	N/A	1.107
Additional Verification	N/A	30.000	10.000	N/A	40.000
Combination Train	6.850	N/A	N/A	6.000	12.850
Combination Val	1.107	N/A	N/A	1.058	2.165
Final Test	20.000	15.000	5.000	N/A	40.000

Table 2: Quantity of each sub-set distributed on their labels

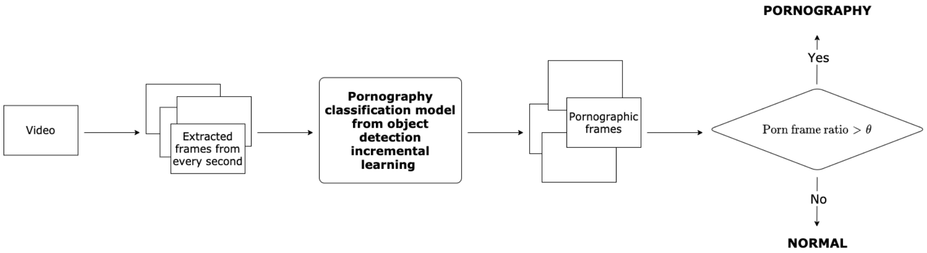

PORNOGRAPHY
Yes
Video
Extracted frames from every second
Pornography classification model from object detection incremental learning
Pornographic frames
Porn frame ratio > θ
No
NORMAL

Fig. 4: Proposed method's training procedure

500 training steps and 200 evaluation steps for every epoch. All the training images were resized into 512 x 512 and augmentation strategy was randomly flip. The initial step of our strategy was training data with the first 60 epochs with the head layer of Mask R-CNN network. After that, we fine-tuned the model with the 40 later epochs, using the whole network architectures. The result from the initial training phase was the sensitive object detector that could detect sensitive objects in images.

In the additional training phase, an Additional Verification set had been built with 40.000 non-porn images. All images in this set were not pornographic content because this verification set aimed to improve the sensitivity of the model. If any image in this set was detected as positive result, it was considered as false positive and then would be annotated with corresponding pseudo labels and used in the additional training phase. This step helped model discriminate between pornographic content and normal content which look like pornographic. The predicted and segmented results of Mask R-CNN can be seen in Figure 5.

After the two phases of training, the model was completed with the help of an SVM binary classification. Through experiments with various settings, the linear kernel has been chosen since it had given the significant results from others.

4.4 Result and Evaluation

To evaluate the effectiveness of the proposed method, 2 testing sets had been used included a manually collected testing set and the NPDI-800 benchmark dataset. In the manually collected testing set, 40.000 images was gathered from the internet included 20.000 pornography, 15.000 normal and 5.000 sexy images. The evaluation results in

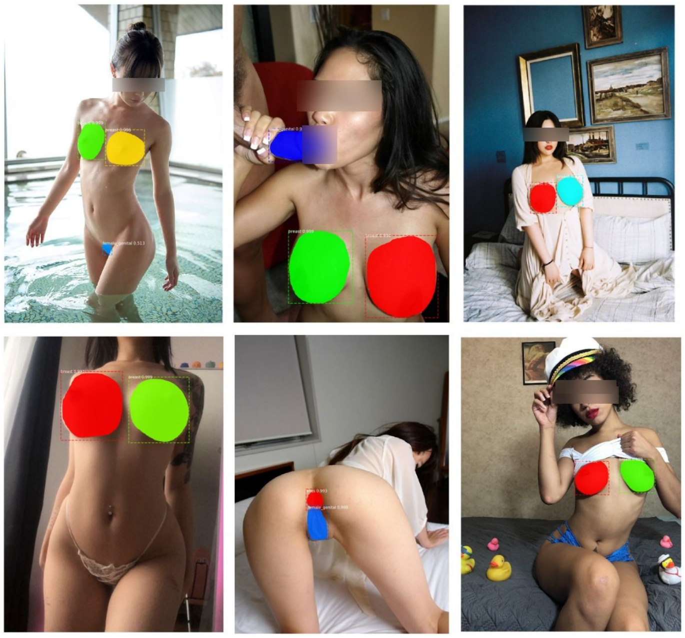


Fig. 5: Mask R-CNN performance

detail are presented in Table 3. In the pornographic class, the accuracy slightly decreases from 89.03% to 84.43% but there is a significant improvement on normal and sexy classes. In specific, the accuracy has increased from 86.45% to 97.48% on normal class. Furthermore, the proposed model also boosts the accuracy on sexy class from 52.02% to 93.32%. In total, the proposed model achieved 90.43% accuracy compares to 83.44% from classic training method. The effectiveness of the proposed additional training on sensitivity and specificity is shown in Table 4. The proposed method has improved the total accuracy by reducing the false positive rate from 22.16% in classic training strategy to 3.56% in additional training strategy. In the other hand, the true positive rate is also affected by the novel training strategy which reduces from 89.03% to 84.43%. However, the true nature of this problem is not detecting pornographic content as much as possible but trying to discriminate the pornography and normal content. Being too sensitive may incur user experience in unpleasant way. Because of that, we focus on improving the total accuracy by reducing the false positive rate and accepting a little bit decrease in the true positive rate.

Beside the manually collected dataset, the NPDI-800 dataset was also used to evaluate the proposed model. This dataset consisted of 400 videos which are divided into 2 groups namely pornographic and normal. Because the testing data were videos, it was necessary to have a strategy to conduct the testing since the proposed model only works on image. Therefore, every second in each video of NPDI-800 was extracted as key-frames. For each extracted key-frame, the final model, which achieved from the SVM

Class	Classic Training	additional Training	Quantity
Normal	86.45%	**97.48%**	15.000
Sexy	52.02%	**93.32%**	5.000
Porn	**89.03%**	84.43%	20.000
Total Accuracy	83.44%	**90.43%**	40.000

Table 3: Custom dataset prediction results

	TP	TN	FP	FN	Total Accuracy
Classic training	**89.03%**	77.85%	22.16%	**10.97%**	83.44%
Train-boosting	84.43%	**96.44%**	**3.56%**	15.58%	**90.43%**

Table 4: Rate results

binary classification, made consideration and then gave the prediction. This strategy is graphically presented in Figure 4, where

$$\text{porn frame ratio} = \frac{\text{porn frame}}{\text{total frame}}$$

In order to avoid wrong detection, a video is considered as pornography if only it contains at least θ percentage of porn frames. The selected value of θ will decide how sensitive the model be. If θ is too low i.e. a video will be treated as pornographic when it only have few predictably pornographic frames, the model becomes too sensitive

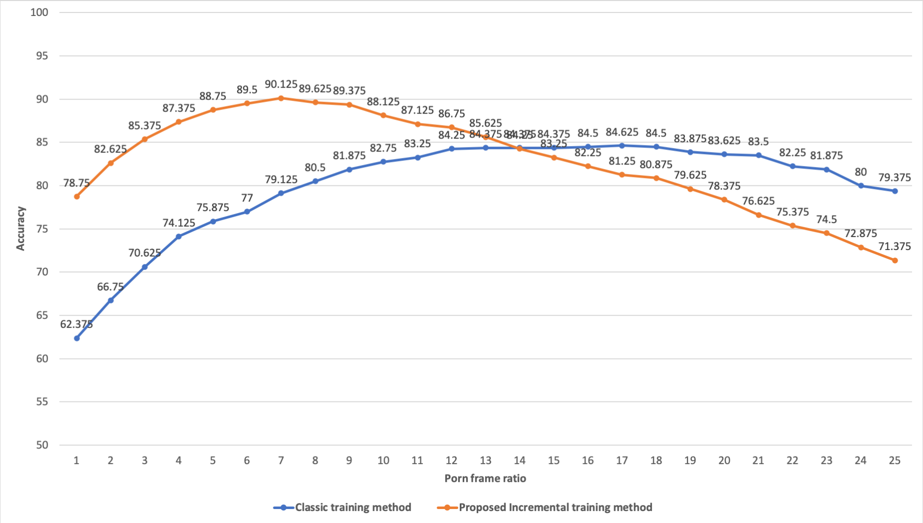


Fig. 6: NPDI-800 result

since not all predictions are correct. This leads to high false-positive rate. In the other hand, if θ is too high, the threshold to identify if a video is pornographic become too high and easily lead to miss detection, in other words, high false-negative rate. In this paper, we choose θ by experimental results from Figure 6. In this figure, the horizontal axis represents for the value of θ while the vertical axis represents for the accuracy, the blue line shows the result from classic training strategy and the orange line shows the proposed method's result. To reach the peak performance, this value may be varied in two approaches. Based on experiments, $\theta = 7$ is chosen which means a video should have at least 7% porn frames over the total number of frames to be determined as pornographic video. Compare to the pure classification from object detection approach, the proposed additional learning gives a significant improvement whose result is **90.125%** compare to **84.625%** from original method.

5 Conclusion

In this paper, we proposed an approach for the pornography identification which applied a new training strategy call additional training. Our new strategy contained two phases of training named initial training phase which was the classic approach, and additional training phase which re-trained the model with pseudo labels from false-positive images. The proposed model had proven its performance by reducing the false positive rate hence increasing the total accuracy. The final result, which achieved from Mask R-CNN model combined with an SVM binary classification, had shown the better performance which is 90.125% compared to 84.625% from pure object detection approach on NPDI-800 dataset. On the manually collected dataset, the proposed method also showed the significant improvement in both accuracy and false positive rate which are 90.43% and 3.56% respectively.

Acknowledgments

This research is funded by Vietnam National University Ho Chi Minh City (VNU-HCM) under grant number B2019-26-02.

References

1. Zaidan, A. A., Karim, H.A., Ahmad, N.N., Bahaa, B. and Sali, A. (2013). An automated anti-pornography system using a skin detector based on artificial intelligence: A Review. International Journal of Pattern Recognition and Artificial Intelligence, vol. 27.
2. Shen, R., Zou, F., Song, J., Yan, K., Zhou, K. (2018). EFUI: an Ensemble Framework using Uncertain Inference for pornographic image recognition. Neurocomputing. vol. 322. 10.1016/j.neucom.2018.08.080.
3. Wehrmann, J., Simões, G. S., Barros, R. C., Cavalcante, V. F. (2018). Adult content detection in videos with convolutional and recurrent neural networks. Neurocomputing, vol. 272. 10.1016/j.neucom.2017.07.012.
4. de Avila, S., Thome, N., Cord, M., Valle, E. and Araújo A. (2013). Pooling in Image Representation: the Visual Codeword Point of View. Computer Vision and Image Understanding (CVIU), vol. 117, pp. 453–465.

5. Balamurali, R. and Chandrasekar, A. (2019). Multiple parameter algorithm approach for adult image identification. Cluster Computing, vol. 22.
6. Moreira, D., de Avila, S., Perez, M., Moraes, D., Testoni, V., Valle, E., Goldenstein, S. and Rocha, A. (2016). Pornography classification: the hidden clues in video space-time. Forensic Science International (Forensic Sci. Int.), vol. 268, pp. 46–61.
7. Moustafa, M. N. (2015). Applying deep learning to classify pornographic images and videos. In 7th Pacific-Rim Symposium on Image and Video Technology (PSIVT).
8. Mahadeokar J., Pesavento G. (2016) Open Sourcing a Deep Learning Solution for Detecting NSFW Images, Available online: `https://yahooeng.tumblr.com/post/151148689421/open-sourcing-a-deep-learning-solution-for`.
9. Vitorino, P., Avila, S., Perez, M., and Rocha, A. (2018). Leveraging deep neural networks to fight child pornography in the age of social media. Journal of Visual Communication and Image Representation, vol. 50, pp. 303 – 313.
10. Wang, Y., Jin, X., and Tan, X. (2016). Pornographic image recognition by strongly-supervised deep multiple instance learning. International Conference on Image Processing (ICIP), pp. 4418–4422.
11. Tabone, A., Bonnici, A., Cristina, S., Farrugia, R. and Camilleri, K. (2020). Private Body Part Detection using Deep Learning. Proceedings of the 9th International Conference on Pattern Recognition Applications and Methods (ICPRAM), pp. 205–211.
12. He, K., Gkioxari, G., Dollár, P., Girshick, R.B. (2017). Mask R-CNN. IEEE International Conference on Computer Vision (ICCV), pp. 2980–298.
13. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778.
14. Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2117-2125.
15. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, vol. 20, no. 3, pp. 273–297.
16. Abdulla W. (2017) Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. GitHub repository, GitHub, Available online: `https://github.com/matterport/Mask_RCNN`.
17. Redmon, R., Divvala, S., Girshick, R.B., Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788.
18. Redmon, J., Farhadi, A. (2017). YOLO9000:Better, Faster, Stronger. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7263–7271.
19. Redmon J., Farhadi A. (2018) YOLOv3: An Incremental Improvement, Arxiv, arXiv:1804.02767.
20. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C. (2016), SSD: Single Shot MultiBox Detector. European Conference on Computer Vision (ECCV), pp. 21–37.