Title: Recognizing pornographic visual content with object detection

The Internet allows people to upload any digital data into world-wide-web, including unsafety and harmful content, such as pornography. Censoring pornography visual content is one of the most challenging problems in Computer Vision.

In recent years, many efforts have been made to distinguish pornographic visual content from normal ones. One of the earliest approaches involved recognizing the exposed skin ratio to decide if there is a nude person in that image [1,2]. Sometimes, identifying facial and posing can improve the model performance [3,4]. However, this simple and uncostliness approach came with disadvantages, which are depending heavily on the quality and resolution of input images; misclassifying safety images that contain a skin-like object or vast amount of exposed body skin; or suffering from high false-positive prediction [5]. Later, another approach to identifying sensitive visual content is extracting visual features from images from the same label and mapping them into a dictionary codebook. Then, a machine learning classifier is adapted to identify the pornographic elements based on that codebook's feature[6,7,8]. While this method came with better performance than the skin detection-based approach, the difficulty of selecting appropriate features to represent pornographic images——as well as the diversity of this type of visual content——makes it hard to deploy a good predicting model.

With the robustness of deep learning, there are many approaches recently using deep neural networks architectures——mainly Convolutional Neural Network (CNN)——to train better models to tackle this problem. Deep learning method comes with state-of-the-art performances on different experiments [9, 10]. However, the biggest limitation of this approach is data, as neural network often demands a large-scale dataset for training and validating the model. Of course, we can augment training data (such as flip, zoom, or change some inner colors and dimensions) but it also comes with noise or bias. Also, modifying hyperparameters for a robust neural networks model with confidence performance is a conundrum, too.

One key element to identify pornography are sexual organs and sensitive objects portray blatantly within that content, regardless of visual or textual [11]. Previous study [12] shows that 98% of pornographic visual content exposes sensitive sexual objects such as nake breast, genital, anus, dildo, etc. which raise the erotic level on the viewer's eye. Thus, recognizing these organs and objects means identifying pornography. Therefore, in this paper, we proposed a novel approach that detecting sexual objects as a key factor to identify pornographic visual content. Our main proposed approach is adapting few short learning Aid CNN model——which is one of the best performance object detection model in recent years that only need a small number of data for training——to identify if there are any sexual object on the input image, thus conclude if the image is safety or not. To ensure the performance of our approach, facial recognition and skin extraction are also deployed as the side factor of judging images. A supervised discrimination model Support Vector Machine is adapted to receive output features from these modules, and determine if the input visual content is safe or not. Which the combination models from the previous approaches, this proposed hybrid can overcome previous approaches's limitations.

In the experiment, our model achieved the performance of 99.98% Accuracy on the open datasets APorn28, BPorn55, and XFDB, overcomes the current state-of-the-art. Deploy in internet supervision system, we hope our method could ensure the safety of the user, especially children when surfing the internet without worrying about inappropriate content.