International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 20 (2018) pp. 14595-14599
© Research India Publications. http://www.ripublication.com

Detection of Illegitimate Divulgation of Obscene Contents Using Tensorflow

S.Anoosha Devi[1], V.Arvind[2]
[1]School Of Computing (IT), SASTRA Deemed University, Thanjavur, India.
[2]School of Computing (CSE), SASTRA Deemed University, Thanjavur, India.

Abstract

Agile development of technology, with a camera in every pocket has created and entitled a global audience, for every social media post, non-consensual porn has become increasingly prevalent. Illicit activities such as nonconsensual porn or revenge porn, phishing scam, digital sexual assault, morphed pornography are proliferating threats. The detection of such offensive contents is a challenge of growing importance in multimedia forensics and security. Multiple facial recognition systems and porn detection algorithms can be combined with the emerging concepts of machine learning and cloud computing to formulate an feasible method to resolve this issue. This paper mainly focuses on identification of raunchy contents on the electronic devices using an open source machine learning framework- tensorflow and verification of the authenticity with the registered users.The source URL of the identified dissentious contents along with the victim's authentication ID are reported to the respective search engines with the request of removal.

Keywords: NSFW, Machine Learning, Cloud Server, Tensorflow, python

INTRODUCTION

The advancements in the digital technologies has drastically reformed the aspects of human civilisation in recent decades. With the tremendous evolution of internet applications, the emergence of digital sexual assault has become a casual custom. This pervasive problem poses alarming threats for security and has cultivated complexities in legal and emotional aspects.

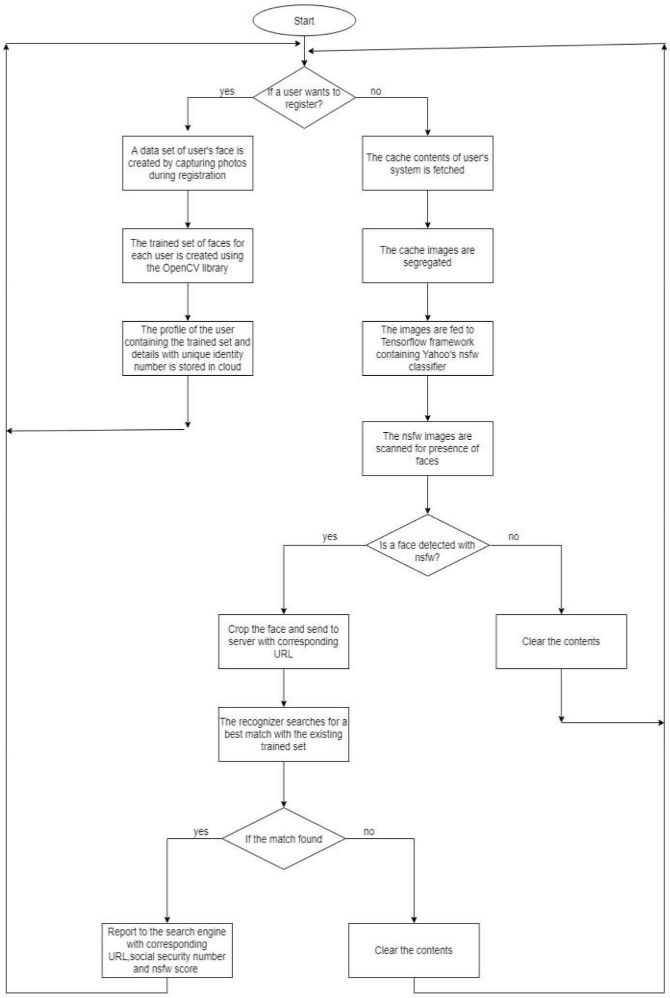
Though there are initiatives taken by groups like the Cyber Civil Rights Initiative (CCRI) regarding this crucial issue, there is no provision in IPC that is particularly steered towards chastening the cyber harassment/cyberstalking until recently. Moreover 93 percent of non-consensual porn victims endure emotional disorders, 82 percent are alleged of deterioration in social and occupational behaviour. 51 percent have admitted on suicidal attempts. 42 percent are reported of seeking psychological treatments. Mental health therapists also face complexity in comprehending the dimensions of the admitted issue in forensic view. This technology could contrive a greater social impact by eradicating unsought X-rated contents to a greater extent, thus retarding the social and clinical work loads in the glimpse of forthcoming digital era[4[11]].

The proposed model uses a numerous machine learning and deep learning algorithms chained together to train and create datasets, spot the NSFW(Not Suitable For Work) classifications, recognise images and identify facial features[7]. A set of image files opened on Chrome browser on Windows platform is gathered from the cache and fed as input to a python code that uses tensorflow(a framework for open source machine learning) adopting yahoo's NSFW classifiers[3]. The caffe model, originally used to detect NSFW content, is converted to tensorflow model by using github (https://github.com/mdietrichstein/tensorflow-open_nsfw) contribution[2]. Though caffe is a fast and open framework, tensorflow proves to be advantageous because of the fact that it is easily deployable and provides better compatibility for GPUs and multi-machine configurations.

The code returns a NSFW score that measures the nudity content in the file in terms of percentage. If this parameter exceeds the threshold value, the image file is scanned for facial recognition-a biometric application that analyzes facial feature patterns and contours to uniquely identify a person's face[13]. There are enormous and efficient facial recognition techniques contemporarily which includes generalized face detection matching method and the adaptive regional blend method[8]. These facial recognition systems compute based on the various human face nodal points. The values calculated for the variable corresponding to, points of a person’s face leads in distinctively recognising the target person. testing the model with new set of inputs. The submitted model uses OpenCV(open source computer vision library) that has python interface and supports Windows to detect and recognize human faces.

The non-consensual activity detected is directed to the search engine with the URL of the link containing the NSFW content, the NSFW score and the person's authentication id and details with the license agreements.

PROSPECTIVE PARADIGM


Start
If a user wants to register?
yes
no
A data set of user's face is created by capturing photos during registration
The trained set of faces for each user is created using the OpenCV library
The profile of the user containing the trained set and details with unique identity number is stored in cloud
The cache contents of user's system is fetched
The cache images are segregated
The images are fed to Tensorflow framework containing Yahoo's nsfw classifier
The nsfw images are scanned for presence of faces
Is a face detected with nsfw?
yes
no
Crop the face and send to server with corresponding URL
Clear the contents
The recognizer searches for a best match with the existing trained set
If the match found
yes
no
Report to the search engine with corresponding URL,social security number and nsfw score
Clear the contents

In order to mitigate the victimisation to revenge porn, all computers could be embedded with system programs to discern the prevalence of NSFW content. The users can sign up with the server program to avert the threat of digital sexual assault. The application would acquire the user's authentication id/social security number and necessary details of the user. A dataset of user's images is created by capturing photos using the client's webcam during sign up process. The data set of each registered user is schematically stored with unique keys.

An inbuilt system program/client program is scheduled to run automatically each time when the an image file is loaded into the cache. This client program runs an analysis test on the fetched cache contents to identify files with NSFW presence. These detected files are analysed for the faces[9]. The faces associated with the obscene themes are pruned and sent to the server with relative source URL of the file. The server programs is initiated. The input images are correlated with the trained image set stored in the server's cloud. If any match cases are found the server examines the seriousness of the problem stated and reports it to the analogous search engine with necessary elementary facts such as social security number/unique identification id, URL of the designated link, NSFW score.

CLIENT ARCHITECTURE

The cache is a high speed memory of limited storage capacity. Every file from the browser's data centres are viewed by the end machine by loading it in the cache. Therefore, a regular monitoring process on the contents of the cache can be made to trace any form of illicit activities on the internet. A check-up program for detection of NSFW materials can be run every time when the browser's cache is loaded with a new file. Care should be taken to reject the rescanning of previously processed contents as this would shoot up the processing/service time. To avoid this complexity, the client program is initiated on the occasion of deletion or updation of cache contents. This technique works because for faster processing the cache data should be in sync with the data warehouse. This urges the need for regular and frequent updation or discardation of the cache data. There are efficient scheduling operations to delete the cache contents. For instance, when a matching cache is invalidated, it is cleared. When all the history list messages that references the cache is deleted, the history caches are deleted. Similarly, when corresponding matching cache is deleted, an XML cache is cleared. But this process actually decelerates the circulation of porn in the internet as the response time of detection of the crime depends on the frequency of clearance of cache contents. i.e. The process becomes ineffective when the cache is not deleted frequently. This can be overcomed with an efficient algorithm to retrieve cache contents without rescanning the repeated contents.

The client program uses glob API (nirsoft) to collect all the cache contents from a specified cache path of the browser (Google Chrome) and save it in a directory. (https://www.nirsoft.net - Chrome cache view)[1]. It then segregates the image files separately for the detection of porn. The vital classifiers needed for the NSFW detection combines several Convolutional Neural Networks(CNN) and deep learning algorithms[12]. Several such classifiers are open-sourced by many contributors. The Yahoo open NSFW classifier has been used here for NSFW detection using the tensorflow deep learning framework (developed by google brain team). Tensorflow is opted for the fact that each node in this implementation is regarded as a tensor operation thus providing modularity. Further, tensorflow uses object-oriented paradigm and it can be operated on multiple GPUs. A github repository by Marc Dietrichstein consists of the caffe to tensorflow converted implementation of the yahoo open NSFW classifiers. This model retains the actual caffe weights. Currently, jpeg images alone are accepted. The model can be operated on two input types- encoded string of base64 or float tensor. If the chosen input type is float tensor, either of the two types of image loading mechanism - yahoo image loader and tensorflow image loader can be selected. These image loaders convert the images to black and white format, crop, resize to uniform dimensions, and convert it into arrays for computation. Yahoo image loader uses skimage module and PIL(python image library) packages. Whereas, the tensorflow image loader doesn't bank on these packages. Instead it imitates the caffe's image loading mechanism. A minor discrepancy can occur on comparison of the results by both mechanism because of diversity in methods used for resizing the picture and encoders and decoders used for jpeg images. Depending upon the input type preferred the image is processed, layers are created and weights for each neuron in the hidden layer are assigned. This activation process continues and weights are assigned for last layer. Thus, if an input image is fed to the code, a value between 0 to 1 analogous to the extent of nudity in the image and the corresponding percentage value is returned. If this percentage value is greater than a threshold value of 70%, the image is considered to be NSFW. These images are segregated for visage detection.

The Haar Cascade classifiers in cv2 package of python is used to detect the presence of face. The faces are cropped and converted to gray scales images in jpg format and stored in file. Connection is established with the server using port number and the encoded URL is sent. Later the image file is encoded and sent for identification the faces. The server uses the trained set of faces of registered users to identify the person. If a match is found in the trained set, person is stated to be victimised. The URL along with the unique id of the person's file is printed as output. Further emendation can be made by reporting to the crime to respective search engine with amendments such as URL of the relevant source, the person's details collected during sign up process and the salient license agreements for further investigations.

Algorithm Client

```
1.while(true)
    1.1.if cache is updated or deleted
        1.1.1.Retrieve cache contents
        1.1.2.Segregate the images
        1.1.3.Check for presence of nsfw contents
              using yahoo nsfw classifiers in
              Tensorflow
        1.1.4.if nsfw contents are present
            1.1.4.1.Check for faces with the nsfw
                    contents
            1.1.4.2.if faces are present
                1.1.4.2.1.Crop the faces
                1.1.4.2.2.Convert to grayscale
                1.1.4.2.3.Establish connection with
                          Server
                1.1.4.2.4.Send request along with
                          respective URL and
                          identified encoded face
                          image
            1.1.4.3.else
                1.1.4.3.1.Clear contents
        1.1.5.else
                1.1.5.1.Clear retrieved contents
2.End
```

SERVER ARCHITECTURE

The end users, who deprecate the presence of their lewd images on the internet shall register in this application. The python code used in the front end uses the openCV library for face detection using deep learning concept. The access to webcam is gained using cv2.videocapture() function in openCV module and several snapshots of the registered users is captured to create a data set. The webcam has an inbuilt face detection mechanism to focus on faces. The face should not be too close or too far with respect to the camera as the scaling is not done dynamically. The images are captured continuously and swiftly in a short span of 20 ms interval for each image. A lumpful amount of images (almost twenty sample snapshots) of the individual is captured to avoid human errors. To complete the sign up process, the user is requested to provide their unique identity number(social security number,aadhar number,etc.) and other needed details.

The BGR images are converted to grayscale images using cv2.cvtcolor(img,cv2.COLOR_BGR2GRAY) function to reduce noise interventions, gain faster results and enhance accuracy[5]. The collected images are scanned for detecting faces using the Haar Cascade classifiers. OpenCV uses the concept of cascading classifiers to create a repository of pre-trained classifiers for facial features and gestures as XML files in opencv/data/haarcascades folder. The valid facial images are cropped to uniform dimensions using cv2.rectangle() function to prune out the unnecessary background details. This is appended with the respective id to frame a dataset. cv2.imwrite() is used here to store the images in a file.

The datasets are retrieved from the path and processed using image module in python image library(PIL) . PIL lends functions to generate instances by fetching images from files, creating a new image or modifying the existing image and saving it in various image formats. The images are converted into shades of gray and converted in the format of array of 8-bit unsigned integers using array() function in the numpy module of python libraries.

The array of images now trained using the train() function of Local Binary Patterns Histograms(LBPH) face recogniser in the cv2 module of python[10]. In the LBPH method each pixels is compared with the corresponding adjacent pixels to create histogram for every image. The trained images are now saved as YML files using save() function of LBPH recognier.

The client programs checks the systems for the prevalence of the images containing NSFW contents and sends the corresponding cropped face image and URL associated with it. The server is in an optimal running state at all time, waiting for requests from the client. When the server is fed with URL and the new image as input, the recogniser creates histogram for the image and hunts for the best match among the existing histograms of the registered users to identify the person. If the person is identified by the recogniser the confidentiality is compared the threshold value of seventy percent to verify the victims. The respective identity number, NSFW score, and the URL is printed. With legitimate authorisation, these information can be sent to corresponding search engines for further investigations and revisions.

Algorithm Server

```
1.if user wants to register
    1.1.Get the user details
    1.2.Get the access to web camera
    1.3.Capture ample amount of images to create
      dataset
    1.4.Convert images to grayscale
    1.5.Use Haar Cascade classifiers in the openCV
      package to identify faces
    1.6.Crop faces to uniform dimensions
    1.7.Train the data set
    1.8.Save the user details and trained data set with
      unique id
2.while(true)
    2.1.if client request is received
        2.1.1.Initiate recognizer with the received face
          image
        2.1.2.if match is found in the trained set
            2.1.2.1.Report to search engine
        2.1.3.else
            2.1.3.1.Clear contents
3.End
```

ADVANTAGES

In the present scenario, the internet has been amalgamated with our routines. The digitisation of data has paved a broader path for privacy policies. Invasion of privacy is the major intimidation of every individual. Reinforcement of justifiable legislation for the social issues on the online platform regarding the intrusion of privacy is still in an hazy state of judiciousness. Many prodigious software companies are working on this issue to subjugate the seriousness. Facebook is experimenting on tools on removing posts related to morphed porn,revenge or non-consensual porn contents. But such technique failed because it couldn't differentiate the consensual and non-consensual posts. Google and other immense search engines, on other hand, has made a bid to limit the prevailing defective contents by circulating e-forms and identifying socially disturbing contents. Google has addressed and removed 43 percent of such requests by manually reviewing it. But this task would non-productive and time-consuming for the search engine companies as each submission is subjected to various kinds of legal proceedings. Adding up to the trouble, many notorious users can misuse the provided facility for personal vengeance thus piling up undesirable requests. Another major point of consideration is there are conflicts between the agreements proposed by the developers on freedom of expression and rights to privacy. To balance and bridge gaps between these rights and to cut down the costs, some online platforms have come up with models to reduce the occurence of the NSFW contents in the search results than to remove it entirely. However, this can only be a partial solution to the massive problem.

This theme proposed can be worthwhile on a larger scale as it is generalised and automated. There would be minimal interference of the legislation on embarking upon the problem. The model precisely identifies the porn contents that are on circulation without the consent of the individual. It also mitigates the hazards on threats of intrusion of privacy in internet usage. With a betterment in algorithms to retrieve cache contents and highly scalable faster support support systems the model could respond faster, thus alleviating the impacts. i.e. the system could be designed to delete the contents before one could share thus eradicating the replication and proliferation of non-consensual contents. Also, this a back end process and doesn't completely involve the interruption of service providers or the search engines. Thus, the suggested scheme, with a betterment in technology and efficacious alternatives, can be a solution to a wide variety of internet related immoralities with respect to revenge porn and morphed pornography.

DISADVANTAGES

Though this ideology seems to be worthwhile, there arises unfavourable circumstances in implementation of this model. If the application is inbuilt, only the newer advanced version of the devices could be benefitted. The former versions should be upgraded and patches must be created for reconciling the application with the systems. Additional compatibility issues can also commence as the model uses frameworks that does high-end computational formulations. For amelioration, GPUs with high processing power should be used. Low-configuration machines with lesser RAM would be discordant for the application because this would impede the processing capacity of the machine by occupying portion of main memory for the back-end process.

The execution time of the program is extensively large because of adoption of many deep learning algorithms and frequent exchange of information with the cloud system[6]. The cache is frequently modified/cleared and thus the program is frequently invoked. Humongous amount of the data is processed each time. Efficient algorithms for retrieving cache contents without reprocessing the same file has to be articulated.

This method is bounded to process only jpeg images. Code should be scaled to process different forms of data in different file formats. Also, for sending and receiving huge data files to the cloud system excess of internet is used up in this process. For optimised performance the machine needs to be connected with a strong internet service provider all-time. The prodigious amount of data generated during the creation and training of the data set for each individual, face recognition and the NSFW classifiers requires efficient data storage and retrieval system. Eminent compression techniques has to be enforced. Apt database management queries has to be exerted.

CONCLUSION

This model propounds an elucidation on the optimal solution to detect the illegitimate divulgation of obscene contents in the internet to mitigate the cyber threats concerned with revenge porn. An insight to this model is found in the github repository- https://github.com/arvindrvs/nsfw_face. This detection is made efficient by the deploying the yahoo's NSFW classifiers in tensorflow implementation. The storage optimization could be done by delegating the server processes to robust cloud entities. Coupling with an advanced accomplishments in the techniques for faster retrieval of the cache contents, the proposed model could provide an outspread solution. Thus, the automated inbuilt version of this ideology equipped with cutting-edge technologies could conduce a prominent solution.

ACKNOWLEDGEMENT

The publications would have been inconceivable without the guidance of Mrs.R.Ezhilarasie madam, SASTRA University. We are greatly gratified for the assistance provided by our faculty.

REFERENCES

[1] Nirsoft, Google chrome cache extraction, https://www.nirsoft.net/utils/chrome_cache_view.html

[2] Github, Yahoo NSFW Tensorflow implementation,

https://github.com/mdietrichstein/tensorflow-open_nsfw

[3] Tensorflow, Python platform for deep learning,

https://github.com/tensorflow/tensorflow

[4] Paarijaat Aditya, Rijurekha Sen, Bernt Schiele, Bobby Bhattacharjee, Tong Tong Wu, etc., I-Pic: A Platform for Privacy - Compliant Image Capture, In Proceedings of the ACM Annual International Conference on Mobile Systems, Applications, and Services (MobiSys - 16), 2016.

[5] OpenCV, Open source classifier for image processing, https://github.com/opencv/opencv

[6] Wikipedia, Running a web server in cloud, https://en.wikipedia.org/wiki/Cloud_computing

[7] Wikipedia, NSFW (Not Suitable For Work) ,https://en.wikipedia.org/wiki/Not_safe_for_work

[8] M. Mathias, R. Benenson, M. Pedersoli, etc, Face detection without bells and whistles. In proceedings of European Conference on Computer Vision (ECCV), 2014.

[9] S. Joon Oh, R. Benenson and M. Fritz, etc., Person detection and recognition in personal photo collections, In International Conference on Computer Vision (ICCV), 2015.

[10] Terence Sim and Li Zhang, Controllable face privacy, In International Conference and Workshops on Automatic Face and Gesture Recognition, 2015.

[11] Paarijaat Aditya, Technical Report: I-Pic: A Platform for Privacy-Compliant Image Capture.

http://www.mpisws.org/~paditya/papers/ipic-tr.pdf.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton., Imagenet classification with deep convolutional neural networks (CNN). In Conference on Neural Information Processing Systems (NIPS), 2012.

[13] J. Deng, W. Dong, R. Socher, etc .ImageNet: A Large-Scale Hierarchical Image Database, Computer Vision and Pattern Recognition (CVPR), 2009