[[1]](#footnote-1)

Offensive Image Content Detection System through Text Extraction and Image Processing

*Aiman Fatima, Vaidehi Gupta, Shally Kumar, Kanishka Arora*

*Abstract*--

# Introduction

One of the main problems of the Internet is the presence of harmful (e.g. pornographic) or even illegal (e.g. pedophilic) contents which is not suitable for all age groups. The amount of this non-desired material is growing at an increasing rate. Nowadays, a lot of children under the age of 12 are on the internet for whom certain content is not suitable.

To design a system which would detect the offensive image and text content present in the form of an image against social media standards and block them from posting on any social media platform.

# Literature survey

In this section, literature review of papers published in the past 10 years in the field of image processing along with machine learning and its applications for recognition of offensive content in images like adult content, violence, etc. have been discussed. It also explores the relation between the current work and the already published work. A lot of researches have been made using artificial intelligence and its subset machine learning as well as data mining techniques to present many different systems. We have discussed a few papers that concerned our model and are related to our work below.

According to the paper [2], in the past 10 years there have already been many advances for detection of nudity and adult content using machine learning along with other technologies like deep learning.

It can be seen that Support Vector Machines (SVM) is one of the most commonly used technique in this domain [1][4][10]. It is also used along with other technique Bag-of-words method [9]. Other algorithms that is very popular is Convoluted Neural Network (CNN) wherein many different models for the same have been implemented. For instance, Fast R-CNN is used for the optimization of classification and bounding box regression tasks. Faster R-CNN is made use for additional subnetwork to generate region proposals and YOLO for object detection via a fixed-grid regression [2][3]. Techniques like FAISS or Elastic Search are used for identifying approximate nearest neighbor search and Logistic regr4ession and random forest for shallow classifier [3]. These are some of the popular techniques and algorithms used in image processing.

However, it was inferred from [1] that text extraction from images is still unexplored as compared to the latter. The methods and algorithms suggested for the same were Support Vector Machines (SVM) based method with the multilayer perceptron (MLP). It suggested text extraction from web images can also help in filtering of images with offensive language as future work. It is also helpful in conversion of web page to voice. Also, voice coding for blinds, intelligent transport system, Image tagging, robot vision and scene analysis, etc. are the possible works to be worked on [1]. For text analysis, the Support Vector Machine (SVM) algorithm and N-gram model based on Bayes’ theorem is applied and experimented to filter pornographic text for both Thai and English language web sites in the paper [13].

From the analysis of various previous research work, we realized that a lot of work can be done in the domain of text extraction and it can be a great future research work. There are tons of content on Social Media having textual content on Images, for instance memes shared these days are in form of text, which generally requires a good system for offensive content filtering. Hence, our work tries to explore this use case.

REFERENCES:

1. Tridib Chakraborty et al, Text Recognition using Image Processing. International Journal of Advanced Research in Computer Science, 8 (5), May-June 2017,765-768
2. Zhao Z., Zheng P., et al, Object Detection with Deep Learning: A Review. IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, April 2019
3. Gandhi S, Kokkula S, et al, Image Matters: Scalable Detection of Offensive and Non-compliant Content / Logo in Product Images. arXiv preprint arXiv:1905.02234. August 2019
4. Platzer, Christian & Stuetz, Martin & Lindorfer, Martina. (2014). Skin Sheriff: A Machine Learning Solution for Detecting Explicit Images. SFCS 2014 - Proceedings of the 2nd International Workshop on Security and Forensics in Communication Systems. 10.1145/2598918.2598920.
5. T. Deselaers, L. Pimenidis and H. Ney, "Bag-of-visual-words models for adult image classification and filtering," 2008 19th International Conference on Pattern Recognition, Tampa, FL, 2008, pp. 1-4. doi: 10.1109/ICPR.2008.4761366
6. ] Arentz, Will Archer and Bjørn Olstad. “Classifying offensive sites based on image content.” Computer Vision and Image Understanding 94 (2004): 295-310.
7. Giuseppe A et al (2009). Detection of images with adult content for parental control on mobile devices?. Proceedings of the 6th International Conference on Mobile Technology, Application and Systems, Mobility '09. 10.1145/1710035.1710070.
8. Prashant Tomer et al., 2013. ON LINE SOCIAL NETWORK CONTENT AND IMAGE FILTERING, CLASSIFICATIONS, Int. J. Engg. Res. & Sci. & Tech. 2013
9. Kansara, Krishna B. and Narendra M. Shekokar. “A Framework for Cyberbullying Detection in Social Network.” (2015).
10. Hosseinmardi, Homa & Arredondo Mattson, Sabrina & Rafiq, Rahat Ibn & Han, Richard & Lv, Qin & Mishra, Shivakant. (2015). Detection of Cyberbullying Incidents on the Instagram Social Network.
11. H. Hosseinmardi, R. I. Rafiq, R. Han, Q. Lv and S. Mishra, "Prediction of cyberbullying incidents in a media-based social network," *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, San Francisco, CA, 2016, pp. 186-192.
12. Haoti Zhong, Hao Li, Anna Squicciarini, Sarah Rajtmajer, Christopher Grin, David Miller, and Cornelia Caragea. 2016. Content-driven Detection of Cyberbullying on the Instagram Social Network. In Proceedings of the Twenty-Fi‡h International Joint Conference on Articial Intelligence (IJCAI’16). 3952–3958.
13. Polpinij, Jantima & Sibunruang, Chumsak & Puangpronpitag, Somnuk & Chamchong, Rapeeporn & Chotthanom, Anirut. (2008). A Web Pornography Patrol System by Content-based Analysis: In Particular Text and Image. 500 - 505. 10.1109/ICSMC.2008.4811326.
14. Mofaddel, Mahmoud & Bakheet, Samy. (2010). Adult image content filtering: A statistical method based on Multi-Color Skin Modeling. 2010 IEEE International Symposium on Signal Processing and Information Technology, ISSPIT 2010. 366-370. 10.1109/ISSPIT.2010.5711812.
15. Shih, Jau-Ling & Lee, Chang-Hsing & Yang, Chang-Shen. (2007). An adult image identification system employing image retrieval technique. Pattern Recognition Letters. 28. 2367-2374. 10.1016/j.patrec.2007.08.002.
16. Zheng, Huicheng & Liu, Hongmei & Daoudi, Mohamed. (2004). Blocking objectionable images: Adult images and harmful symbols. 1223 - 1226 Vol.2. 10.1109/ICME.2004.1394442.
17. –
18. J. Ruiz-del-Solar et al, Characterizing Objectionable Image Content (Pornography and Nude Images) of specific Web Segments: Chile as a case study, IEEE,  January 2005
19. Kim, Wonil & Lee, Han-Ku & Yoo, Seong-Joon & Baik, Sung. (2005). Neural Network Based Adult Image Classification. 481-486. 10.1007/11550822\_75.
20. Schmidt, Anna & Wiegand, Michael. (2017). A Survey on Hate Speech Detection using Natural Language Processing. 1-10. 10.18653/v1/W17-1101.
21. Nobata, C et al. (2016). Abusive Language Detection in Online User Content. 145-153. 10.1145/2872427.2883062.
22. –
23. Chen, Ying & Zhou, et al. (2012). Detecting Offensive Language in Social Media to Protect Adolescent Online Safety. 71-80. 10.1109/SocialCom-PASSAT.2012.55.
24. Pitsilis, Georgios & Ramampiaro, Heri & Langseth, Helge. (2018). Detecting Offensive Language in Tweets Using Deep Learning.
25. Xiang, Guang & Fan, et al. (2012). Detecting offensive tweets via topical feature discovery over a large scale twitter corpus. 1980-1984. 10.1145/2396761.2398556.
26. Zampieri, M et al. (2019). SemEval-2019 Task 6: Identifying and Categorizing Offensive Language in Social Media (OffensEval).
27. ] Dadvar, M et al. (2013). Improving Cyberbullying Detection with User Context. In Proceedings of 35th European Conference on IR Research, ECIR 2013, Advances in Information Retrieval. pp 693-696. 10.1007/978-3-642-36973-5\_62.
28. Razavi, Amir & Inkpen, Diana & Uritsky, Sasha & Matwin, Stan. (2010). Offensive Language Detection Using Multi-level Classification. 16-27. 10.1007/978-3-642-13059-5\_5.
29. Zampieri, M, et al. (2019). Predicting the Type and Target of Offensive Posts in Social Media. 1415-1420. 10.18653/v1/N19-1144.
30. Van Hee C, Jacobs G, Emmery C, Desmet B, Lefever E, Verhoeven B, et al. (2018) Automatic detection of cyberbullying in social media text. PLoS ONE 13(10): e0203794. <https://doi.org/10.1371/journal.pone.0203794>
31. Basilio, J et al. (2011). Explicit Content Image Detection. Signal & Image Processing. 1. 10.5121/sipij.2010.1205.
32. –
33. Choudhary, Kaustubh. (2012). Image Steganography and Global Terrorism. IOSR Journal of Computer Engineering. 1.10.9790/0661-0123448.
34. Akusok, A, Yoan M. et al (2015). Arbitrary Category Classification of Websites Based on Image Content. Computational Intelligence Magazine, IEEE. 10. 30-41. 10.1109/MCI.2015.2405317.
35. Adebayo J, Onifade, et al, (2011). Combating Terrorism with Biometric Authentication Using Face Recognition. 10.13140/RG.2.1.2381.5206
36. ] K.N. Natei, et al., Extracting Text from Image Document and Displaying Its Related Information. Journal of Engineering Research and Application www.ijera.com ISSN : 2248-9622, Vol. 8, Issue5 (Part -V) May 2018, pp 27-33
37. ] N.K. Gundu, et al., Text Extraction from Image and Displaying its Related Information. International Journal of Scientific and research Publications, Volume 4, Issue 12, December 2014

1. [↑](#footnote-ref-1)