International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/IJSRCSEIT

A Hybrid Deep Learning Approach for Explicit Content Detection in Images on Social Media and Internet

Pradeep N. Fale[1]*, Dr. Krishan Kumar Goyal[2], Dr. Shivani[3]

[*1]Research Scholar, Department of Computer Science and Engineering, Bhagwant University, Ajmer, Rajasthan, India

[2]Dean, Faculty of Computer Application, RBSMTC, Agra, India

[3]Bhagwant University, Ajmer, Rajasthan, India

Article Info
Volume 8, Issue 3
Page Number : 126-134

Publication Issue :
May-June-2022

Article History
Accepted: 10 May 2022
Published: 22 May 2022

ABSTRACT

The exponential increase in the amount of explicit content has presented numerous obstacles to the current way of life. This is especially true in situations wherein children and minors have unrestricted access to the internet. This process of screening the image features of all the TV channels in Malaysia imposes a huge censorship cost on the service providers like Unifi TV because all films, both local and foreign, are required to obtain the suitability approval in Malaysia before they can be distributed or shown to the public. This paper proposes the use of a hybrid model of Deep Learning (DL) approaches, specifically CNN+SVM and CNN+XGBOOST, in order to further improve the process of explicit images recognition in visual contents. The goal of this paper is to use this issue to our advantage. Transfer learning was performed using the previously trained model in order to find a solution to a new binary classification problem including explicit and non-explicit images. The effectiveness of the model that has been developed is examined using a dataset that has recently been compiled and contains more than number of examples of explicit non-explicit images photographs. The CNN+XGBOOST approach was able to acquire the best performance in terms of accuracy of 99.81 % after tests were run on the dataset. This was in comparison to the 76.49% accuracy attained by the CNN+SVM model. For better evaluation we also present the comparison of the proposed systems with standard state of art mechanisms viz. NB, LDA, SVM, RF, KNN, DT, and LR.

Keywords : Abusive Content Detection, Natural language processing, Social Media, Deep Learning, Machine Learning

I. INTRODUCTION

The advent of social media had a profound impact on the practices and goals that are associated with the use of communications technology in the current day. When mass communication media were originally brought into existence, they were utilised in connection with the moral and ethical responsibilities that were determined by society standards. Aside from that, various forms of mass communication have been effectively utilised for the purposes of educational and vocational training. The advent of social media has made it possible for everyone with an internet connection to express their opinions on any topic they want. This may be done through the use of social networking sites such as Facebook, YouTube, Snapchat, and Twitter, amongst others. A recent study conducted on social media indicated that persons have a low tolerance for others, which presents itself as aggressiveness, in which they use phrases that may be hurtful to the sensibilities of others. This study also revealed that individuals have a lack of empathy for others. On the other hand, the overwhelming majority of online platforms contain guidelines concerning the uploading of material, as well as penalties for violating such regulations.

Filtering out inappropriate visual information from a variety of sources is a significant challenge in many settings, including schools, homes, and places of employment, amongst others (internet TV, online pages, and so on). However, despite the fact that numerous attempts have been made within this body of literature to find a solution to the problem of detecting explicit content, there is a distinction between the meaning of the word "Explicit images" in our research and that of the previous academic publications [3, [7], and [11]. This is because our research focuses on the detection of explicit content rather than on the detection of explicit images. Even if a woman wearing only a bikini is considered acceptable for viewing in the United States or Europe, she is still considered to be adult material in India and even in other countries such as Malaysia, Indonesia, or Brunei, where she is still considered to be adult material.

The primary objective of this study is to automate the difficult and time-consuming task of explicit images identification by utilising the capabilities of deep learning (DL) techniques, as outlined in [1]. To be more specific, it is proposed to use a particular DL model called Convolutional Neural Networks (CNN) [2], which have recently achieved the best performances in all visual recognition tasks (classification, segmentation, detection, localization, etc.), including the recognition of pornography and other forms of adult content, as reported in [3]–[6].

When we compared the performance of five conventional machine learning techniques to that of five novel hybrid deep learning approaches, we found that deep learning techniques outperformed machine learning algorithms by a significant margin. This led us to conclude that deep learning techniques are superior.

The organization of this paper is as described in the following: In the second section, a review of the relevant literature is presented. In the third section, datasets and models are discussed. In the fourth section, findings are evaluated, and the last section presents a conclusion.

II. Related Work

In the past, it was common practice to use conventional feature descriptors such as LBP (local binary patterns) [12], SIFT (scale invariant feature transform) [13], or HOG (histogram of oriented gradients) [14] to obtain internal or external develop a greater sense and then use these descriptors to differentiate between images with sensitive content

and images with normal contents. However, this practice has become less common in recent years.

For example, the research presented in [11] made an effort to recognize pictures that showed pornographic or nudist scenes by proposing the use of a new variant of SIFT known as Hue-SIFT to extract global image features. This was done in conjunction with a Bag of Feature (BoF) model, which was used to acquire a global representation of the image. Later on, the authors demonstrated that the same recognition rate as that which was reported by those skin-based algorithms could be achieved without the detection of skin or forms in the explicit images pictures.

In a different piece of research [15], the authors handled the issue of detecting pornography by suggesting the application of high-level semantic characteristics. They optimized the BoF model in order to bridge the gap between the high-level and low-level scene properties. To do this, they combined the contextual information included in the pornographic images' visual language with the spatial characteristics of the pornographic images themselves. In a later section of [16], it was proposed that the regions of interest (RoI) could be used to solve the problem of inaccurate pornography detection. This was done under the assumption that the task at hand was comparable to object detection and that the human visual system uses the visual attention model to solve problems of this nature. Their proposed framework included four stages: the first was the detection of skin regions; the second was the construction of visual saliency maps; the third was the detection of pornographic regions based on threshold segmentation; and the fourth stage was the extraction of features including colour, texture, intensity, and skin.

The research presented in [17] was the first attempt at using mid-level picture descriptors for the purpose of resolving the pornography detection job in movies. To be more specific, the authors proposed using a novel video frame descriptor that makes use of local binary patterns in conjunction with BossaNova, which is a powerful mid-level picture representation that is detailed in [18]. The identical task was accomplished with BossaNova once more in [19]. The BoF model was utilised in [20], in which the authors reported the use of a multi-instance modelling approach based on spatial pyramid partitions (SPP) to shift the target problem (pornography detection) into a MIL problem. This was accomplished with the help of the BoF model.

The study that was done in [7] was the first time that Temporal Robust Features (TRoF) were used in the task of detecting pornography in films. It was suggested that the TRoF features should be aggregated into mid-level features by utilising FV (Fisher vector), a new form of BoVW (bag of visual words) model.

Since the development of convolutional neural networks, beginning with the ground-breaking work presented in [2], all previous performances of hand-crafted feature descriptors have been significantly enhanced, and these low-level descriptors are no longer the focus of the attention of researchers and practitioners. [CNNs] have allowed for a significant leap in the performance of feature descriptors. The research presented in [21] provides an overview of the efforts that have been done in the field of adult content recognition, the majority of which make use of hand-crafted features. The reader who is further interested might discover the survey in [22] to be quite helpful. In this survey, the authors compared and reviewed the various local feature extraction methods that are used in the field of online pornography detection. Recently, with the advent of Deep Learning techniques, all areas of visual recognition, such as detection, localization, classification, and so on, have seen considerable improvements in all fields, including biometric [23], bioinformatics [24], and so on.

The field of pornography content recognition is not an exception, and many attempts have been made to improve upon the earlier results with the assistance of DL, such as the ones that were published in [3, [4], [6], [10], and [25].In [6], a new CNN architecture was proposed to firstly quickly identify the coarse images with no or fewer skin/facial sensitive contents. Next, a fine detection was performed to identify the target pornography contents in a selected subset of all video frames. Finally, the proposed architecture was used to identify coarse images with no or fewer skin/facial sensitive contents. In a later section of [25], a technique based on deep learning was suggested as a decision support tool that could assist with the evaluation of sexual assaults. The authors of [4] developed a Weighted Multiple Instance Learning (WMIL) strategy that may be combined with a CNN model in order to identify regions that contain pornographic content.

In another piece of research, referred to as [10], it was recommended that an ensemble of CNN-based classifiers be used in conjunction with a probability model that was based on uncertain inferencing in order to perform the task of recognizing adult content in still photos. In conclusion, writers in [3] advocated the utilization of a pre-trained CNN model referred to as ResNet-50 [26] in order to identify potentially sensitive pornographic content within photographs.

III. Proposed Methodology

In this section we will discuss about the proposed Abusive and Explicit Content Detection model. We have built a Python based Hybrid DL/ML model for the detection of explicit content from the text as well as Image datasets. The Hybrid model for Image dataset is based on CNN-SVM and CNN-XGBOOST. For evaluation of the obtained results we use Accuracy, MAPE and RMSE as the performance parameters.

3.1 Hybrid CNN-SVM and CNN-XGBOOST Model for Explicit/Non-Explicit Image Detection

According to the previous debates, it is critical to design an explicit content detection (ECD) technology which not only identifies the data as explicit, non-explicit, or suspicious, but also has to be scalable, quick, and accurate in order to be effective.

Following Figure 1 depicts our proposed ECD system, which accepts as input a list of files and checks for the content kinds (i.e., image, video, or non-image file). Using a predetermined threshold probability, each recognized image/video file will be submitted to the CNN-SVM and CNN-XGBOOST models for categorization of content as explicit or non-explicit depending on the content's explicit or non-explicit classification.

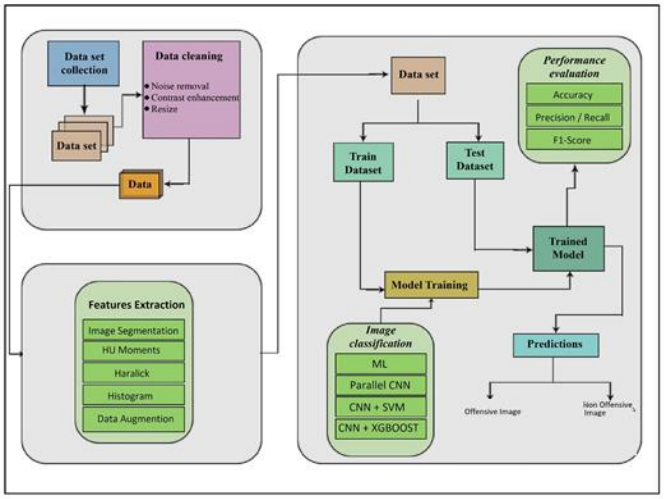

Figure 1. Architecture Model for Proposed Explicit Content Detection Model

3.2.1 CNN-SVM Model

The suggested method combines the best features of both SVM and CNN classifiers into a single system. A convolutional neural network (CNN) is a type of neural network that consists of many fully connected layers and has a learning process that is supervised. CNN operates in a similar manner to how we humans do, and it is capable of learning invariant local characteristics quite effectively. When dealing with

raw digit pictures, it is capable of extracting the most discriminating information. When a 5x5 kernel/filter has been used to extract the most recognizable characteristics from the raw input pictures, the proposed system is called a 5x5 system. nxn input neurons from the input layer are convoluted with a mm filter in the convolutional layer, which results in an output with the size (n-m+1) (n-m+1) of the convolutional layer. Each layer's output is used as the input for the layer above it in the hierarchy. When calculating effective sub-regions from a raw digit image, the receptive field feature of CNN is used to aid in the computation. An SVM attempts to represent a multi-dimensional dataset in space in which data items belonging to different classes have been divided by a hyperplane in order to describe the dataset; this is known as the Support Vector Machine (SVM). The Classification algorithm has the capacity to reduce the classification error on data that has not yet been viewed. The separating hyperplane is sometimes referred to as an optimum hyperplane in some instances. However, it is discovered that SVM is unsuccessful when dealing with noisy data. SVM is effective when dealing with binary classification. Because of the shallow design of SVM, there are certain difficulties in learning deep features when the architecture is shallow.

As a result of the current research, it is proposed a hybrid CNN-SVM model, in which SVM is used as a classification technique and the softmax layer of CNN is substituted by SVM. During the course of this application, CNN is employed as a feature extractor, while SVM is used as a binary classifier. The suggested hybrid CNN-SVM model's architecture is depicted in Figure 2, which describes the model's overall design. Figure 2 Represents the Layer architecture model for CNN-SVM.

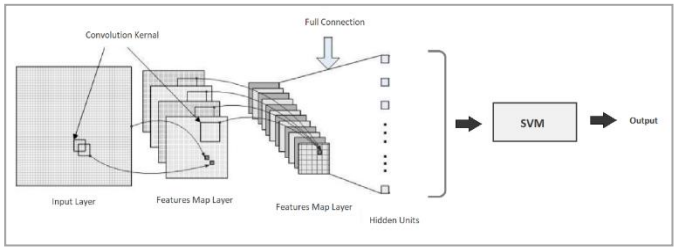

Figure 2. Architecture Diagram of CNN-SVM Model

3.2.2 CNN-XGBOOST Model

Figure 3 depicts the specific design of the Classification algorithm for image classification, which is used for image classification. First, the input picture data is normalised and sent to the CNN's input layer, where it is further processed. After training CNN using the BP method for numerous epochs in order to acquire an appropriate structure and image analysis, XGBoost replaces the output layer of CNN with a soft-max classifier and uses the trainable characteristics from CNN for the training phase of the algorithm. The CNN-XGBoost model then receives the updated classification results from the testing pictures. Combining the two exceptional classifiers, our CNN-XGBoost model can automatically extract features from input and produce more precise classification results than each classifier alone.

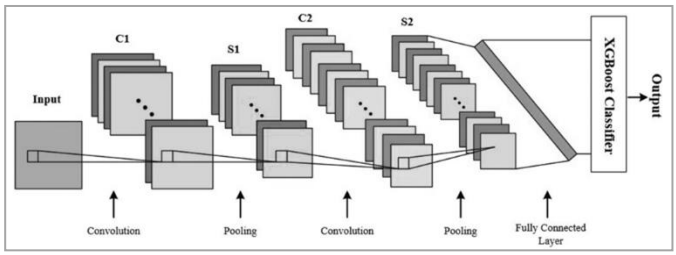

Figure 3. Architecture Diagram of CNN-XGBOOST Model

3.3 Dataset

It is required to balance the dataset in order to choose a higher number of images that have been classified as inappropriate as would otherwise be the case. For Explicit/Explicit images Image detection we used a NSFW Dataset obtained from Kaggle.

3.4 Experimental Setup

This study describes the implementation of the prediction models experimented on the Windows 10 Professional platform. It will, however, be implemented across a number of different platforms. It is performed on a machine with 8 GB of RAM and a 256 GB solid-state drive (SSD). MS Excel 2010 was used to prepare the data. For implementation, Python 3.7 and above was used to write the code for the models.

3.5 Performance Parameter

RMSE (Root Mean Square Error): To calculate the RMSE, the following equation is used

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - o_i)^2}$$

Where,
n: number of samples
f: forecasts
o: observed values

MAPE (mean absolute percentage error): To calculate the MAPE, the following equation is used

$$\text{MAPE} = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

Where A_t is the actual value and F_t is the forecast value.
n be the number of fitted points

Accuracy: The accuracy of the algorithms is obtained by following equation

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n}$$

IV. Experimental Results

For Image Dataset

In this section we present the result analysis of both CNN-SVM and CNN-XGBOOST systems. For better evaluation we also present the comparison of the proposed systems with standard state of art mechanisms viz NB, LDA, SVM, RF, KNN, DT, LR. The performance parameter comparison for all the algorithms is presented in table 1.

TABLE 1. Performance Parameters Comparison of Machine Learning Algorithms for Image Dataset

Algorithms	Accuracy
NB	60
LDA	59
DT	65
SVM	70
LR	76
KNN	77
RF	80.14

We obtained the results for CNN+SVM and CNN+XGBOOST algorithms and the results for MAPE, RMSE and Accuracy for both the algorithms is presented in table 2.

Table 2. Comparison of Performance Parameters of Hybrid Deep Learning Models

Algorithm	Accuracy	MAPE	RMSE
CNN+SVM	76.49	13.99	23.23
CNN+XGBOOST	99.81	5.99	10

The final comparison of our hybrid CNN-SVM and CNN-XGBOOST model with all the implemented ML model is shown in figure 4.

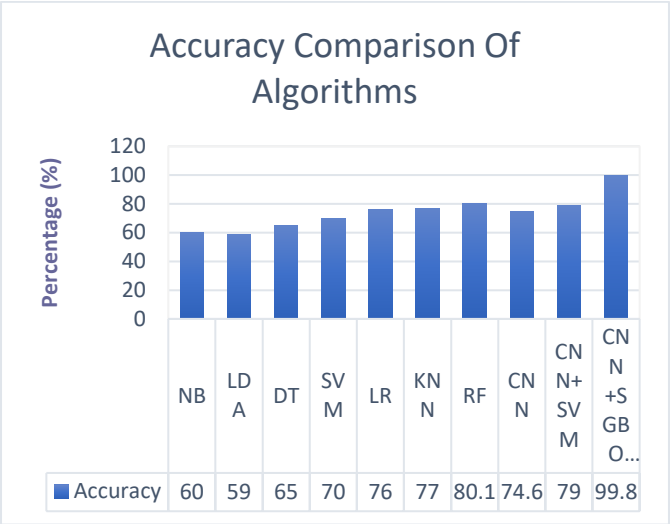

Figure 4. Accuracy Comparison of All Algorithm with Proposed Hybrid Models for Image Dataset

V. Conclusion

Within the scope of this investigation, our proposed solution to the problem of binary classification of explicit image contents is to make use of ResNet, a relatively new and highly effective architecture of deep neural networks (explicit vs. non-explicit images). We created a new dataset of explicit images that includes over 4,000 instances of explicit and non-explicit images. This definition is referred to during the process of visual content regulation of TV stations. Following the completion of a series of studies, it was discovered that the CNN+SVM and CNN XGBOOST Hybrid Models led to a significant improvement in the explicit images classifier's overall performance, as well as state-of-the-art outcomes in terms of accuracy. In upcoming work, we plan to expand our dataset to include additional classes of sensitive content, such as various kinds of pornographic acts, as well as additional facets of explicit image instances, such as complicated explicit images and partial explicit images, amongst other things. This will be done in addition to the expansion of our explicit images dataset. In addition, future efforts may benefit from CUDA-enabled implementation in order to enable deployment of deep learning models on embedded platforms for high-speed utilization of explicit images detection in high-resolution (SD/HD/FHD) video frames. This would be done in order to enable high-definition (FHD), standard definition (SD), and ultra-high definition (FHD) video frame detection.

VI. REFERENCES

[1]. Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[2]. A. Krizhevsky, "ImageNet Classification with Deep Convolutional Neural Networks," NIPS, vol. 4, no. 4, pp. 253–262, 2012.

[3]. A. Nurhadiyatna, S. Cahyadi, F. Damatraseta, and Y. Rianto, "Adult content classification through deep convolution neural network," Proc. - 2017 Int. Conf. Comput. Control. Informatics its Appl. Emerg. Trends Comput. Sci. Eng. IC3INA 2017, vol. 2018-Janua, pp. 106–110, 2018.

[4]. X. Jin, Y. Wang, and X. Tan, "Pornographic Image Recognition via Weighted Multiple Instance Learning," IEEE Trans. Cybern., vol. PP, pp. 1–9, 2018.

[5]. F. Nian, T. Li, Y. Wang, M. Xu, and J. Wu, "Pornographic image detection utilizing deep convolutional neural networks," Neurocomputing, vol. 210, pp. 283–293, 2016.

[6]. K. Zhou, L. Zhuo, Z. Geng, J. Zhang, and X. G. Li, "Convolutional neural networks based pornographic image classification," Proc. - 2016 IEEE 2nd Int. Conf. Multimed. Big Data, BigMM 2016, pp. 206–209, 2016.

[7]. D. Moreira et al., "Pornography classification: The hidden clues in video space–time," Forensic Sci. Int., vol. 268, pp. 46–61, 2016.

[8]. M. D. More, D. M. Souza, and R. C. Barros, "Seamless Explicit images Censorship : an Image-to-Image Translation Approach based on Adversarial Training," IEEE Int. Jt. Conf. Neural Networks, 2018.

[9]. A. P. B. Lopes, S. E. F. De Avila, A. N. A. Peixoto, R. S. Oliveira, M. D. M. Coelho, and A. D. A. Araújo, "Nude detection in video using bag-of-visual-features," Proc. SIBGRAPI 2009 - 22nd Brazilian Symp. Comput. Graph. Image Process., pp. 224–231, 2009.

[10]. R. Shen, F. Zou, J. Song, K. Yan, and K. Zhou, "EFUI: An ensemble framework using uncertain inference for pornographic image recognition," Neurocomputing, vol. 322, pp. 166–176, 2018.

[11]. A. P. B. Lopes, S. E. F. De Avila, A. N. A. Peixoto, R. S. Oliveira, and A. De A. Araújo, "A bag-of-features approach based on Hue-SIFT descriptor for nude detection," Eur. Signal Process. Conf., no. Eusipco, pp. 1552–1556, 2009.

[12]. W. Zhou, A. Ahrary, and S. I. Kamata, "Image description with local patterns: An application to face recognition," IEICE Trans. Inf. Syst., vol. E95-D, no. 5, pp. 1494–1505, 2012.

[13]. D. G. Lowe, "Object recognition from local scale-invariant features," Proc. Seventh IEEE Int. Conf. Comput. Vis., pp. 1150–1157 vol.2, 1999.

[14]. N. Dalal, B. Triggs, and D. Europe, "Histograms of Oriented Gradients for Human Detection," 2005.

[15]. L. Lv, C. Zhao, H. Lv, J. Shang, Y. Yang, and J. Wang, "Pornographic images detection using high-level semantic features," Proc. - 2011 7th Int. Conf. Nat. Comput. ICNC 2011, vol. 2, pp. 1015–1018, 2011.

[16]. J. Zhang, L. Sui, L. Zhuo, Z. Li, and Y. Yang, "An approach of bag-of-words based on visual attention model for pornographic images recognition in compressed domain," Neurocomputing, vol. 110, no. July 2012, pp. 145–152, 2013.

[17]. C. Caetano, S. Avila, S. Guimar, and A. D. A. Ara, "Pornography Detection using BOSSANOVA Video Descripor," pp. 2–6, 2014.

[18]. S. Avila, N. Thome, M. Cord, E. Valle, and A. De A. Araújo, "Pooling in image representation: The visual codeword point of view," Comput. Vis. Image Underst., vol. 117, no. 5, pp. 453–465, 2013.

[19]. C. Caetano, S. Avila, W. R. Schwartz, S. J. F. Guimarães, and A. de A. Araújo, "A mid-level video representation based on binary descriptors: A case study for pornography detection," Neurocomputing, vol. 213, pp. 102–114, 2016.

[20]. D. Li, N. Li, J. Wang, and T. Zhu, "Pornographic images recognition based on spatial pyramid partition and multi-instance ensemble learning," Knowledge-Based Syst., vol. 84, pp. 214–223, 2015.

[21]. C. X. Ries and R. Lienhart, "A survey on visual adult image recognition," Multimed. Tools Appl., vol. 69, no. 3, pp. 661–688, 2014.

[22]. Z. Geng, L. Zhuo, J. Zhang, and X. Li, "A comparative study of local feature extraction algorithms for Web pornographic image recognition," Proc. 2015 IEEE Int. Conf. Prog. Informatics Comput. PIC 2015, pp. 87–92, 2016.

[23]. R. Nejad, Elaheh Mahraban and Affendey, Lilly Suriani and Latip, Rohaya Binti and Ishak, Iskandar Bin and Banaeeyan, "Transferred Semantic Scores for Scalable Retrieval of Histopathological Breast Cancer Images," pp. 1–8, 2018.

[24]. R. Banaeeyan, H. Lye, M. F. Ahmad Fauzi, H. Abdul Karim, and J. See, "Semantic facial scores and compact deep transferred descriptors for scalable face image retrieval," Neurocomputing, 2018.

[25]. K. Fernandes, J. S. Cardoso, and B. S. Astrup, "A deep learning approach for the forensic evaluation of sexual assault," Pattern Anal. Appl., vol. 21, no. 3, pp. 629–640, 2018.

[26]. R. G. Crane, "Deep Residual Learning for Image Recognition 2015," no. (ed.), Oxford, U.K.,

Pergamon Press PLC, 1989, Section 3, pp.111-120. (ISBN 0-08-036148-X), pp. 1–9, 1989.

Cite this article as :

Pradeep N. Fale, Dr. Krishan Kumar Goyal, Dr. Shivani, "A Hybrid Deep Learning Approach for Explicit Content Detection in Images on Social Media and Internet", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 3, pp. 126-134, May-June 2022.
Journal URL : https://ijsrcseit.com/CSEIT228345