



Description of the neural network based on AB/DL pictures. Possible implications for forensic sexology

Correspondence to:

Wojciech Oronowicz-Jaśkowiak
Faculty of Information Technology
Polish-Japanese Academy of Information
Technology
86 Koszykowa St.
02-008 Warsaw, Poland
e-mail: oronowiczjaskowiak@pjwstk.edu.pl

Wojciech Oronowicz-Jaśkowiak¹ , Piotr Wasilewski² 

¹Faculty of Information Technology, Polish-Japanese Academy of Information Technology, Warsaw, Poland

²Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Poland

Submitted: 09.07.2021

Accepted: 23.11.2022

Abstract

Purpose: Neural networks might be an appropriate solution for the categorization of child sexual abuse materials (CSAM) in forensic cases. The aim of this study was to present a neural network model that may be able to categorize objects and behaviors, which are visible in CSAM, using pictures visually similar to CSAM (AB/DL), involving persons who have paraphilic preferences for watching adult women or men dressed like children or involved in activities typical for children, such as playing.

Methods: The dataset consisted of 2251 photos divided into five classes. 1914 photos were randomly used for the training of the neural network, while 337 photos were used for its later validation. The Fast.ai and PyTorch libraries were used for the training of the neural network using the ResNet152 model. We used five classes, two of which were imported from the sexACT dataset, and three of which that were collected for this study.

Results: The model was able to classify selected classes with a relatively high accuracy (95%); on the other hand, further improvement of the network is needed, considering the fact that the final validation loss was moderate (0.17).

Conclusions: The model presented might be effective in the classification of several objects and behaviors presented in a number of pornography categories (“paraphilic infantilism”, “sexual activity”, “nude women”, “dressed women”, “sexual activity – spanking”). As the results are promising, further research on real CSAM is justified.

Key words: pornography, neural networks, CSAM.

INTRODUCTION

The Internet is an environment vulnerable to the commission of sexual crimes [1]. Individuals with sexual interest in children may contact other individuals with similar interests, engage in different forms of sexual communication with children (including sending threats or sexually explicit material), promote sexual tourism or even target children as potential victims for abuse. This paper focuses only on child sexual abuse materials (CSAM).

Given the highly complex nature of the motives for accessing child sexual abuse materials [1, 2], the aim of the expert witnesses in child abuse cases is to evaluate child sexual abuse materials using their psychological and sexological knowledge. The use of anthropological knowledge to assess child sexual abuse materials allows for the estimation of the biological (developmental) age-range of people in the given material. Expert witnesses evaluate child sexual abuse materials using special rating systems to estimate the sexual aggression that is shown in

the pictures and videos involving children (see COPINE Scale [3]). Expert witnesses in child abuse cases play an important role in justice systems; however, a number of major methodological problems have yet to be resolved. Using neural networks and deep learning might be an appropriate way of categorizing child sexual abuse materials.

Due to the importance of the conclusions drawn by the expert witnesses in child abuse cases, it would be helpful to present the exact quantity of data in specific categories. However, due to the large number of photos and videos involved (child sexual abuse materials collections might include from several to over a thousand photos and videos), such an evaluation may not be feasible. Neural networks allow for the examination of amounts of data that could not be manually evaluated by expert witnesses in child abuse cases. Such witnesses could benefit from machine model evaluation when working with material that is too categorically blurred for evaluation. Neural networks might also be applicable in the work of police

officers pre-evaluating child sexual abuse materials. In that case, machine learning could suggest what materials should be sent for further analysis conducted by expert witnesses. Another example of possible application is in the education of future sexologists and anthropologists. Neural networks could help teach students to determine which materials could be classified as CSAM.

Machine learning algorithms are believed to be the best solution for automatic pornography detection and classification. Research in this area has been conducted for several years. Before the now widely-used convolutional neural networks an attempt was made to create a pornography detection system [4]. The authors involved achieved 75% of accuracy, which is a rather insufficient result, in order to use similar a system in clinical practice. The authors used a Support Vector Machine (SVM) – a type of deep learning algorithm that is able to perform supervised learning in the classification of data. Further studies aimed to create pornography detection system using deep neural networks [5]. Particularly good results have been achieved with the use of convolutional neural networks. For example, Castrillón-Santana *et al.* [6] presented a model that is over 93% accurate in terms of child detection. Vitorino *et al.* [7] used deep convolutional neural networks to detect the presence of CSAM content in an image. These authors have used standard neural network architectures and proposed a new one.

It is also worth noting that the detection of child sexual abuse materials might also include not only images, but also videos [8] or audio [9]; videos containing CSAM could be used as training material for model training. It is worth noting that some approaches used to detect pornography are not related to computer vision. For example, Wazir *et al.* [9] presented model that is able to detect pornography using audio features. These authors presented the finding that acoustic pornography recognition might be as accurate as the “traditional” approach (using image processing) with an accuracy of 86.50%. It has also been shown that pornography might be successfully classified as using bag-of-words associated with certain files [10]. Recently, an interesting solution was proposed by Pereira *et al.* [11], who state that machine learning could help law enforcement identify CSAM. Consequently there is a problem with collecting CSAM imagery to train models, due to ethical and legal constraints. To overcome this issue, the authors trained convolutional neural networks on metadata of every CSAM image in their database and achieved a degree of accuracy of 97%. This approach shows that it is not always necessary to have CSAM images in order to conduct research. The conclusions of a recent literature review are also important [12]. The authors of this concluded that all kind of approaches (deep-learning algorithms with multi-modal image or video descriptors merged together) are important in CSAM detection, and that the best results could be obtained if multiple approaches are used in com-

ination. It is worth adding that European Commission guidelines concerning the application of AI in pornography detection have recently been introduced [13], so such models can be used in the EU (for example Google is now able to detect inappropriate content, such as CSAM, by providing Google SafeSearch [14]).

The aim of this study was to present a neural network model that may be able to categorize objects and behaviors which are visible in child sexual abuse materials using pictures visually similar to CSAM. For legal reasons the training dataset did not include actual CSAM; pictures visually similar to CSAM were used instead. Previous studies have shown that it is possible to determine if the material contains pornographic or non-pornographic content; however, the categorization of specific types of child sexual abuse materials has not been attempted.

METHODS

In order to train the neural network to categorize several types of pornography, we used “AB/DL” photos. “AB/DL” (Adult Baby/Diaper Lover) is the name of an Internet community [15] involving persons who have paraphilic preferences for watching adult women or men dressed like children or involved in activities typical for children, such as playing (paraphilic infantilism). “AB/DL” might be considered as a legally obtainable substitute for child sexual abuse materials, given that the pictures show “AB/DL” images are visually similar to CSAM (see also Table 5). However, it should be noted that there is no link between paraphilic infantilism and pedophilic preferences [16].

We had used five classes, two of which were imported from the sexACT dataset [17], and three collected for this study. A detailed description of the preparation of the materials included in the first two classes had been already described; therefore, in this paper we describe only the process of creating classes that have not been presented before.

The model was trained on a computer with the following specifications: Linux Ubuntu 18.04.2 LTS, 32 GB RAM, Intel Core i5-9400F 2.90 GHz, GF RTX 2070. The Fast.ai [18] and PyTorch [19] libraries were used for the training of the neural network using the ResNet152 model [20]. The architecture of ResNet152 consisted of 152 layers including convolutional layers and neural activation functions. It is used as a residual learning framework in the detection of classes of images. As already noted, the model used in the current study uses a convolutional neural network for image recognition [21].

The dataset consisted of 2251 photos divided into five classes (“paraphilic infantilism”, “sexual activity”, “nude women”, “dressed women”, “sexual activity – spanking”). 1914 photos were randomly used for the training

of the neural network, while 337 were used for its later validation.

The process of creating the dataset consisted of three stages.

The first stage involved extracting photos indexed on Google. We searched for relevant images using the following keywords: “spanking”, “spanking sex”, “nude women”, “dressed women”, “T-shirt women”, “dressed women”. Photos were automatically downloaded via a download manager. Between 435 to 856 photos were downloaded for each keyword. All in all, 5763 files were downloaded. The number of photos varied across the different classes: “sexual activity – spanking” consisted of 608 photos, “nude women” consisted of 831 photos and “dressed women” consisted of 491 photos.

The second stage involved the automatic deletion of photos which could not be read by the computer (for example, they might be in a format other than .jpg or .jpeg). This was performed by running a command from the fast.ai library. The number of files remaining after this operation for the “sexual activity – spanking” was 558, for “nude women” 810, and for “dressed women” 471.

The third stage involved the manual selection of photos conducted by three psychologists-sexologists. Some of the photos were deleted due to fact that they were not representative of the category, the picture was actually a drawing, or the photo was blurred or duplicated. After

this stage, we assessed the correctness of the classification of 56 images. Lawshe’s content validity ratio was 1.00 for 11 images and 0.3 for one image, and overall CVI was 0,944, which can be considered as high classification accuracy. The final number of photos after the four stages is shown in Table 1. The characteristics of the materials included in the dataset and they hypothetical equivalent in CSAM are shown in Table 2.

RESULTS

Two stages of neural network training using a one cycle approach [22] were carried out. The results of the first phase along with some basic parameters are presented in Table 3 and Figure I. Before the second training phase, the optimal learning range was computed (see Figure II), using the function from the fast.ai library [18, 23]. The results of the second phase, along with the basic parameters, are presented in Table 4 and Figure III.

Validation loss is the degree of error present after running the validation data, and training loss is the degree of error present after running the training data. Accuracy is defined as metrics for evaluating models; it is the fraction of the predictions made by the machine learning model properly classified.

As shown in the Tables 3 and 4, as the number of epochs increased, the training loss decreased. On the other

Table 1. Characteristics of the materials included in the dataset

Class	Dataset	Modification	Number of photos
“Paraphilic infantilism” (AB/DL)	sexACT 0.1 dataset	We added 75 photos (original dataset consists of 437 photos)	512
“Sexual activity”	sexACT 0.1 dataset	We removed 23 photos (original dataset consists of 370 photos)	347
“Nude women”	Created for the purpose of this study	N/a	433
“Dressed women”	Created for the purpose of this study	N/a	696
“Sexual activity – spanking”	Created for the purpose of this study	N/a	263
Total			2251

Table 2. Characteristics of the materials included in the dataset and they hypothetical equivalent in child sexual abuse materials (CSAM)

Class	Characteristics of photos included in dataset	Hypothetical equivalent in CSAM
“Paraphilic infantilism” (AB/DL)	Photos of adult women dressed like children. The following elements appeared: diapers, pacifiers, bottles, toys, sleepers.	Photos of dressed children.
“Sexual activity”	Photos of adult women and men during sexual activity, including vaginal, anal and oral sex. The following elements appeared: penises, anus, vagina, buttocks, breasts.	Photos of children during sexual activity, including vaginal, anal and oral sex.
“Nude women”	Photos of nude adult women not engaging in sexual activity.	Photos of nude children not engaging in sexual activity.
“Dressed women”	Photos of dressed adult women not engaging in sexual activity.	Photos of dressed children not engaging in sexual activity.
“Sexual activity – spanking”	Photos of adult women and men engaging in sexual activity, including spanking.	Photos of children as victims of sexual violence.

Table 3. Training – stage 1

Epoch	Validation loss	Accuracy
0	1.055738	0.623145
1	0.666002	0.762611
2	0.518696	0.824926
3	0.428738	0.863501

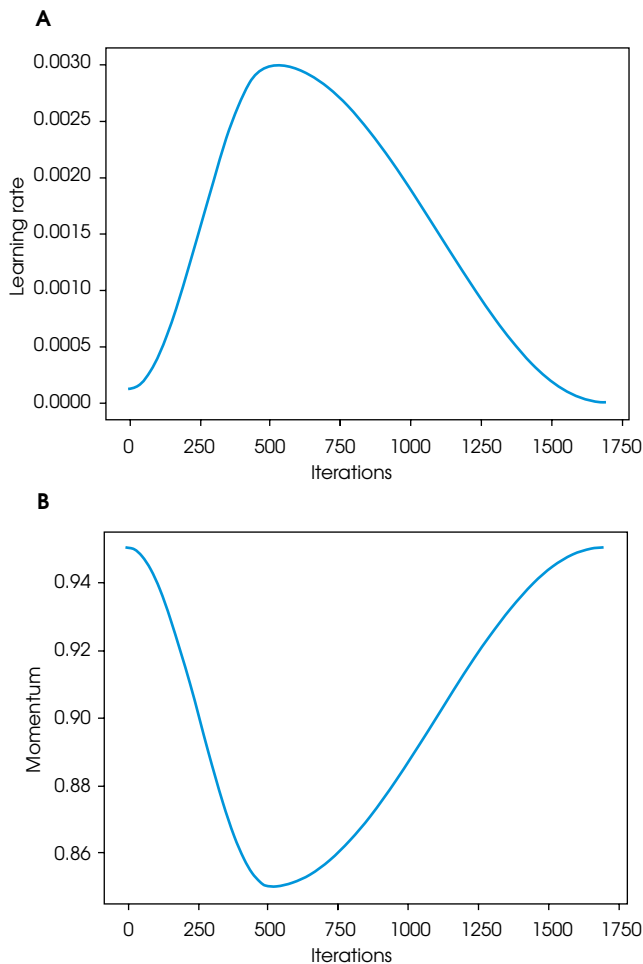


Figure I. Function used for calibration of training – stage 1

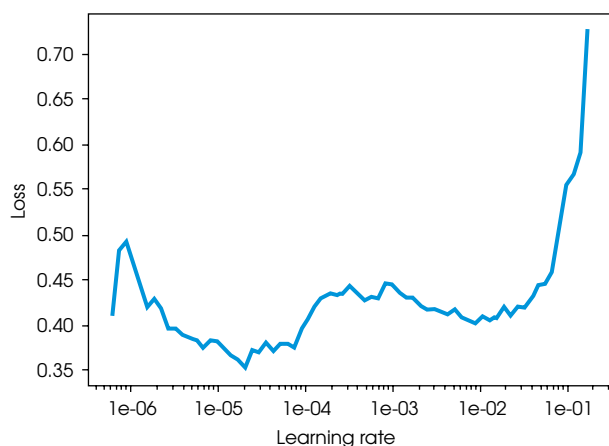


Figure II. Learning rate – stage 1

Table 4. Training – stage 2

Epoch	Validation loss	Accuracy
0	0.310678	0.901176
1	0.261772	0.917647
2	0.266130	0.910588
3	0.269927	0.915294
...
13	0.170804	0.950588
14	0.167400	0.955294

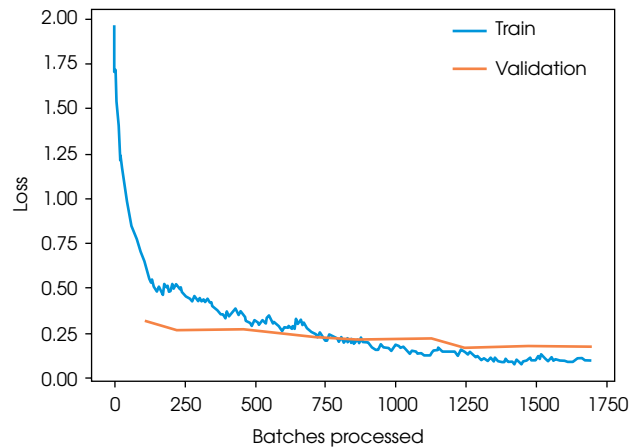


Figure III. Train and validation loss – stage 2

hand, validation loss was constant. The second learning phase resulted in higher accuracy, lower training loss, and a slightly lower validation loss.

DISCUSSION

The aim of this study was to present a neural network model that may be able to categorize objects and behaviors which are visible in child sexual abuse materials, using pictures visually similar to CSAM. This was achieved by training a neural network that can identify several pornographic categories that are visually similar to CSAM with high efficiency.

The model presented might be effective in the classification of several objects and behaviors presented in range of pornography categories (“paraphilic infantilism”, “sexual activity”, “nude women”, “dressed women”, “sexual activity – spanking”). As shown above, the best-performing model based on convolutional neural networks achieves a degree of accuracy of 0.95 in the validation set. The results obtained are similar to other research [accuracy 95% vs. 93%; 6]. The model presented, unlike similar research [5], is able to conduct multiclass classification. On the other hand, further improvement of the network is needed, considering the fact that the final validation loss was moderate (0.17). It should be

also noted that the pre-trained model used (ResNet152) and fast.ai library were designed for classifying pornography and are rather simple. Nevertheless, the fast.ai model seems to be the best starting point for developing a specialized tool. The model could also be improved by adding new types of information, for example keywords associated with pornographic materials [10] or acoustic information [9].

The main limitation of this study was its dataset, as it did not include actual CSAM. As a professional model should be trained on CSAM, the neural network developed in this study may be regarded as a demonstration of the possibilities of this approach. Considering the results of this study, further research in this field may result in the development of a valuable tool for expert witnesses in child abuse cases. Such experts could apply the neural network approach in many complex cases. In particular, neural networks might be used for assessing pornographic materials, estimating the degree of the sexualization of children or even predicting the age of children presented in the pictures. However, even considering the significant technological advancement we have seen, it seems that the expert witnesses in child abuse cases will not be replaced by artificial intelligence.

Firstly, expert witnesses in child abuse cases are personally responsible for their forensic opinions and conclusions. It cannot be expected that the authors of the software will take legal responsibility in cases of misclassification of pornographic materials.

Secondly, even if neural networks were able to classify pornographic material with excellent accuracy, it is still

almost certain that the network will make errors. Even if the error rate of the neural network were to be much lower than that of forensic experts, verification by another expert enables the high degree of accuracy required in the judicial system.

Thirdly, the raw results obtained by processing a dataset with a neural network are not sufficient to be a proof in legal cases. Expert witnesses in child abuse cases are not only there to give their opinion, but also to present the thought process behind it. The opinion of the expert witness can determine the fate of a case. Because of its key importance, a justification of the expert witness' conclusion must be provided so that the conclusions are logical. As of this moment, neural networks cannot provide an answer to the question of how the results were computed.

CONCLUSIONS

A neural network was trained in order to classify pornographic materials in selected categories. The results of this study seem to be promising for future research aimed at training neural networks on real child sexual abuse materials. In the future, tools based on a similar model might facilitate the work of expert witnesses in child abuse cases [24]. Due to the fact that the model presented was not trained on real child sexual abuse materials, but only photos visually similar to CSAM, it is not possible to use this model in forensic practice. However, as the results are promising, further research on real CSAM is justified.

Conflict of interest

Absent.

Financial support

Absent.

References

1. Merdian HL, Curtis C, Thakker J, Wilson N, Boer DP. The three dimensions of online child pornography offending. *Journal of Sexual Aggression* 2013; 9: 121-132.
2. Babchishin KM, Hanson R, Hermann CA. The characteristics of online sex offenders: a meta-analysis. *Sex Abuse* 2011; 23: 92-123.
3. Quayle E. The COPINE project. *Irish Probation Journal* 2018; 5: 65-83.
4. Lin C, Tseng HW, Fuh CS. Pornography detection using support vector machine. In: 16th IPPR Conference on Computer Vision, Graphics and Image Processing; 2003, p. 123-130.
5. Nian F, Li T, Wang Y, Xu M, Wu J. Pornographic image detection utilizing deep convolutional neural networks. *Neurocomputing* 2016; 210: 283-293.
6. Castrillón-Santana M, Lorenzo-Navarro J, Travieso-González CM, Freire-Obregón D, Alonso-Hernandez JB. Evaluation of local descriptors and CNNs for non-adult detection in visual content. *Pattern Recognition Letters* 2018; 113: 10-18.

7. Vitorino P, Avila S, Perez M, Rocha A. Leveraging deep neural networks to fight child pornography in the age of social media. *Journal of Visual Communication and Image Representation* 2018; 50: 303-313.
8. Gangwar A, Fidalgo E, Alegre E, González-Castro V. Pornography and child sexual abuse detection in image and video: a comparative evaluation. In: 8th International Conference on Imaging for Crime Detection and Prevention (ICDP 2017); 2017. DOI: 10.1049/ic.2017.0046.
9. Wazir ASB, Karim HA, Abdullah MHL, Mansor S. Acoustic pornography recognition using recurrent neural network. In: 2019 IEEE International Conference on Signal and Image Processing Applications; 2019, p. 144-148.
10. Karamizadeh S, Arabsorkhi A. Methods of pornography detection: review. In: Proceedings of the 10th International Conference on Computer Modeling and Simulation; 2018, p. 33-38. DOI: <https://doi.org/10.1145/3177457.3177484>.
11. Pereira M, Dodhia R, Anderson H, Brown R. Metadata-based detection of child sexual abuse material. arXiv preprint 2020: arXiv:2010.02387. DOI: <https://doi.org/10.48550/arXiv.2010.02387>.
12. Lee HE, Ermakova T, Ververis V, Fabian B. Detecting child sexual abuse material: a comprehensive survey. *Forensic Science International: Digital Investigation* 2020; 34: 301022.
13. European Commission. White Paper on Artificial Intelligence – a European approach to excellence and trust. from www.eur-lex.europa.eu (Accessed: 26.03.2022).
14. Mulfari D, Celesti A, Fazio M, Villari M, Puliafito A. Using Google Cloud Vision in assistive technology scenarios. In: IEEE Symposium on Computers and Communication (ISCC); 2016, p. 214-219.
15. Hawkinson K, Zamboni BD. Adult baby/diaper lovers: an exploratory study of an online community sample. *Arch Sex Behav* 2014; 43: 863-877.
16. Doshi SM, Zanzrukiya K, Kumar L. Paraphilic infantilism, diaperism and pedophilia: a review. *J Forensic Leg Med* 2018; 56: 12-15.
17. sexAI lab dataset sexACT. www.sexailab.pl (Accessed: 05.07.2021).
18. Howard J. Fast.ai software library. www.fast.ai (Accessed: 05.07.2021).
19. Ketkar N. Introduction to pytorch. In: *Deep Learning with Python*. Berkeley: Apress; 2017.
20. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016, p. 27-30.
21. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems 25 (NIPS 2012)*; 2012, p. 1097-1105.
22. Smith LN. Cyclical learning rates for training neural networks. In: IEEE Winter Conference on Applications of Computer Vision (WACV); 2017. DOI: <https://doi.org/10.48550/arXiv.1506.01186>.
23. Smith LN. A disciplined approach to neural network hyper-parameters: Part 1-learning rate, batch size, momentum, and weight decay. Preprint arXiv: arXiv:1803.09820. DOI: <https://doi.org/10.48550/arXiv.1803.09820>.
24. Thurzo A, Kosnáčová HS, Kurilová V, Kosmel' S, Beňuš R, Moravský N, Varga I. Use of advanced artificial intelligence in forensic medicine, forensic anthropology and clinical anatomy. *Healthcare* 2021; 9: 1545.