Psychiatr. Pol. 2022; 56(4): 877–888
PL ISSN 0033-2674 (PRINT), ISSN 2391-5854 (ONLINE)
www.psychiatriapolska.pl
DOI: https://doi.org/ 10.12740/PP/138243

Classification of nine types of pornographic materials using the sAI 0.3 model. Continuation of the pilot study

Wojciech Oronowicz-Jaśkowiak[1], Adam Siwiak[2], Krzysztof Róg[3], Agnieszka Oronowicz-Jaśkowiak[4]

[1] Polish-Japanese Academy of Information Technology, Department of Informatics
[2] Institute of Psychology, Jagiellonian University
[3] Faculty of Psychology, SWPS University of Social Sciences and Humanities
[4] First Radiology Department, Maria Skłodowska-Curie National Research Institute of Oncology

Summary

Aim. Legal pornographic materials are a heterogenous group of audiovisual materials that depict one or more person over the age of eighteen engaging in sexual activities. The aim of this study was to train a model that could classify given types of pornographic materials.

Method. Materials included in the training set (3,600 materials) and the validation set (900 materials) were manually classified and tagged by psychologists-sexologists. Then, a deep neural network was trained on the dataset. Six models based on different architectures of convolutional neural networks were included in the study (ResNet152, ResNet101, VGG19, VGG16, Squeezenet 1.1, Squeezenet 1.0). Each model was trained on the same group of photographs, and fast.ai library was used for the training process.

Results. The final model allows for the classification of more types of pornographic materials with greater efficiency than the pilot model, and thanks to the manual labelling of individual photographs, the limitations of the classification are known.

Conclusions. The possible applications of the model in clinical sexology and psychiatry are discussed. The application of deep neural networks in sexology seems to be particularly promising for at least two reasons. Firstly, a tool for automated detection of pornographic materials involving minors can be developed and used during criminal proceeding. Secondly, after retraining the presented model on photographs of men and women not engaging in sexual activity the model could be used to filter content that is inappropriate for minors.

Key words: pornographic materials, neural networks

Introduction

The detection and classification of pornographic materials using machine learning models, including deep neural networks, has become a new subject discussed by scientists. Short et al. [1] reviewed the publications related to the characteristics of the internet pornography that had appeared over the previous ten years. The analysis was based on 46 publications, and half of the studies were conducted online, 43% of them were conducted using paper-and-pencil surveys, a few used both of these methods, and one study was conducted by phone. Many different definitions of pornography are adopted in the publications, but in as many as 84% of them, this definition was not disclosed to the participants. For this reason, Short et al. [1] believe it is necessary to precisely define the concept of pornography in every future paper, since the issue of definition has a direct impact on the results of the studies and their comparison between each other. The failure to disclose the definition leads to the fact that some authors believe that visible genitals are a sufficient criterium for classifying a material as pornographic, while others consider a material pornographic only when it depicts a full sexual intercourse. Short et al. [1] note that a well-formalized definition should contain information about the type of pornography and how a particular material is to affect the person who sees it. In their view, a material should be considered pornographic when its character is unambiguously sexual and the material is viewed in order to induce sexual arousal or sexual fantasies.

Over two decades ago, Scott [2] distinguished only two types of pornography:

(1) soft-core – no deviant behaviors are presented, sexual activity depicted was agreed upon by all the parties involved (this category also included "neutral" materials used for sex education);

(2) hard-core – deviant behaviors and sexual violence are depicted; examples include pedophile and zoophile activities.

Several years later, Hald [3] presented his subjects with 31 different categories of pornographic materials, and identified eight that were most often chosen by the respondents. Hald's [3] publication illustrates how significantly the development of new technologies can affect the diversity of pornographic materials. Nowadays, due to the universal access to the Internet, pornographic materials are widely available.

At the same time, the number of persons seeking help due to the problematic use of pornography has been increasing. Data from support groups in Poland show that between 2009 and 2012, the number of the sexual dependency group members increased by 340% [4]. One of the common problems is the development of masturbatory conditioning in response to a specific type of sexual act that may not be achievable in reality. Therapeutic interventions in patients with pornography addiction are difficult, given that users of pornography pay very little attention to the negative aspects of frequent viewing of pornographic content [5]. Attention is also drawn to the fact that pornographic materials can negatively affect the development of children and adolescents [6]. The threats include, among others, the creation of unrealistic expectations regarding sexual intercourse (including the frequent appearance of fetish elements), and sexualization of children's behavior.

All in all, not only is categorizing what is and what is not pornography difficult, but also classifying pornographic materials into types is an increasingly difficult task as the Internet develops.

For the purpose of this study, it has been assumed that legal pornographic materials are a heterogeneous group of audiovisual materials that depict one or more persons over the age of eighteen engaging in sexual activities.

The use of neural networks in sexology

Deep neural networks are a functional equivalent of a human nervous system using mathematical and statistical structures that can be viewed as artificial neurons. The network consists of a network of artificial neurons and their connections, which can be connections with feedback, in which case the response is delayed, or without feedback, in which case the response is immediate [7]. The models using neural networks perform complex mathematical tasks and base their predictions or classification of the database on their calculations. As early as in 2007, Posner and Rothbart [8] demonstrated that neural networks can be used in psychological sciences to predict and understand various forms of human behavior. Although first studies using neural networks in medical and social sciences have already been carried out several years ago – for example, to analyze the relationship between certain behaviors and specific personality types [9] – it seems that in the last few years, these applications have been gaining popularity.

It seems that neural networks may prove useful in a clinical setting, particularly in sexology. An example of a diagnosis that may in the future be supported by machine learning models is the diagnosis of paraphilic disorders in cases related to viewing pornography involving minors.

The diagnosis of sexual preference disorders in a perpetrator accused of accessing pornography involving minors is a demanding task and often requires to analyze all the materials collected by the offender [10]. The possible diagnosis of the pedophilic disorder in a sexual offender has legal consequences for them, including the conditions and course of their sentence, as well as therapeutic consequences, i.e., undergoing specialist therapy. For these reasons, persons with sexual preference disorders often hide certain sexual fantasies[1], and an objective assessment of their pornographic material is gaining in importance. However, it is difficult to conduct a thorough analysis of databases of several thousand pornographic materials.

As shown earlier, there are many types of pornographic material and their classification is difficult. Therefore, more and more accurate image classification techniques,

[1] What is more, as a consequence, paraphilias and their diagnostic criteria are based mainly on information on persons who came into conflict with the law or sought help themselves [11].

including neural networks, are employed. The study of Wang and Kosiński [12] compared the effectiveness of recognition of sexual orientation by a human being and by a model. A human assessor was able to correctly distinguish heterosexual men from homosexual men in 61% of cases. The neural network was able to correctly label 81% of cases, and after presenting the model with five images of the same face, the accuracy of the model increased to 91% [12][2].

An interesting approach to the combined analysis of image and motion has been suggested by Perez et al. [13]. Nudity presented on images on the Internet does not always have a sexual context. To improve the accuracy of the classification, the authors suggested to analyze motion. By combining a static image (a photograph) with motion (a video), they obtained a model characterized by classification accuracy of 97.9% and they reduced the error rate by 65.4%.

Aim and justification of the study

The sAI 0.1 model for classification of pornographic materials has been recently proposed by Oronowicz-Jaśkowiak [10]. The pilot version of the model has been trained to differentiate between seven types of pornography. The model's accuracy is 70%. The methodology used for selection of the training material was not systematic (see below); in addition, the model's accuracy for classification of images outside the training set was low. For this reason, a new neural network model has been trained based on the sAI 0.1 model [10] and sexACT database [14]. New research in this direction is important for several reasons.

Firstly, due to a similar visual content of pornographic materials involving adults and pornographic materials involving minors, new models based on this one may be created and used for forensic-sexological opinions [15].

Secondly, the model as it is can be used for research, for example, for automatic classification of pornographic materials depicting adults and minors and assigning photographs depending on study conditions, for example, in the emotional Stroop test [16].

Methodology

The study material comprised 4,500 photographs in total, 3,600 of which were used as a training set, and the remaining 900 were used as a validation set. The process of creating a model differentiating pornographic materials involving minors consisted of several steps.

[2] The ethical aspect of practical applications of the model requires further analysis.

Step 1. Preparation of the training and validation set

In order to train the model for automatic classification of different types of pornographic materials, photographs depicting adult pornography were downloaded and labeled, assigning it into appropriate categories of pornographic materials.

There are hundreds of categories of adult pornography. These categories may concern the type of sexual activity, the context presented on a given material, or the presence of a specific object (a fetish). For the purpose of this study, we distinguished several groups of pornographic materials, following literature reviews in this field [1, 3].

The criterion for distinguishing the categories of pornographic materials does not reflect the complexity of the problem. It was decided to include the most common categories of pornographic materials; however, not all of the included groups are among the most common ones, as the clinical significance of selected, less common groups of materials, was taken into account as well, as was the case of "AB/DL" ("Adult Baby / Diaper Lover") category. It was also determined to distinguish materials depicting a person having a tattoo[3], which was justified by the possibility of a relatively easy differentiation between these materials and other categories by the neural network.

First, the photographs were semi-automatically downloaded from websites containing pornographic materials of various types. Next, a team of three sexologists-psychologists assessed the quality of each photograph and assigned it to an appropriate category. Table 1 presents the inclusion criteria concerning the quality of a photograph.

Table 1. **Criteria concerning the quality of the study material**

Inclusion criteria	Exclusion criteria
Good visibility of the whole body or a fragment relevant for a given type of pornographic material.	Blurred image making it impossible to identify the types of objects depicted.
Photo resolution at least 800 x 600 px.	Photo resolution less than 800 x 600 px.
File format .jpeg or .jpg.	Other file formats, in particular vector graphics.

Each pornographic material was labeled by sexologists-psychologists in terms of the following objects depicted on each photograph (see Table 2). In addition, each photograph was given a unique identification number.

[3] This category was interesting from the point of view of image processing by the neural network; however, it does not necessarily have a great clinical significance, as the fetish concerning tattoos is rarely a subject of clinical interventions.

Table 2. **Objects depicted on the photographs labeled by sexologists-psychologists4**

Category	Identifying factors
Common for all categories	a) ID of the photograph; b) number of persons; c) number of women; d) number of men; e) no data on gender; f) vagina depicted; g) penis depicted; h) buttocks depicted; i) anus depicted; j) breasts depicted; k) penetration; l) sexual activity; m) clothing items
“AB/DL” category	a) children’s attire; b) bottle; c) pacifier; d) plush toys; e) other toys; f) diaper; g) children’s hairstyle
“Acrotomophilia” category	a) missing upper limb; b) both upper limbs missing; c) missing lower limb; d) both lower limbs missing; e) crutches; f) wheelchair; g) lower limb prosthesis; h) bilateral lower limb prosthesis; i) deformities
“BDSM” category	a) machinery; b) rope; c) chain; d) handcuffs; e) collar; f) whip; g) blindfold; h) mask; i) swatter; j) riding whip; k) spur; l) nipple clamps; m) ball stretcher; n) chastity belt; o) gag
“Feet fetishism” category	a) one foot; b) two feet; c) foot or feet involved in sexual activity
“High heels fetishism” category	a) high heels
“Latex fetishism” category	a) latex apparel
“Tattoo fetishism” category	a) tattoos or a tattoo
“Knee socks fetishism” category	a) knee socks; b) over the knee socks
“Graviditiophilia fetishism” category	a) visible signs of pregnancy

The collected database comprised 4,500 materials, each of which was thoroughly labeled and the quality of which was verified.

In order to analyze the characteristics of objects and situations depicted on photographs, selected descriptive statistics for individual types of pornographic materials were presented. Some characteristics were common for all the types (see Table 2), and some were unique for the classes. A detailed analysis of information and objects depicted on the materials is beyond the scope of this article. Table 3 presents basic descriptive statistics common for all the classes.

Table 3. **Basic descriptive statistics of the number of persons depicted on photographs**

	n	Minimum	Maximum	Mean	Standard deviation
Number of persons	4,500	1	10	1.1831	0.48105
Number of women	4,500	0	10	0.9671	0.41293
Number of men	4,500	0	2	0.1089	0.31929

[4] The basic characteristics of pornographic materials will be presented in a separate article, due to the amount of data and their cognitive significance for understanding the key features of objects depicted on pornographic materials.

Step 2. Selection of an appropriate pre-trained model

In order to select a pre-trained model with most fitting parameters, a series of studies was conducted in which selected models were tested.

Six models based on different architectures of convolutional neural networks were included in the study. Each model was trained on the same group of photographs, and fast.ai library [17] was used for the training process. The photographs underwent a standard augmentation implemented in the fast.ai library, and none of the default values for hyperparameters of augmentation were modified. Each series (for a given pre-trained model) consisted of 50 training cycles. Table 4 presents the final training loss, validation loss, and classification accuracy[5].

Table 4. **Review of the efficacy parameters of the pre-trained models**

Architecture	Training loss	Validation loss	Classification accuracy
ResNet152 [18]	0.044511	0.390854	89.33%
ResNet101 [19]	0.058082	0.478085	87.33%
VGG19 [20]	0.180486	0.476862	85.22%
VGG16 [20]	0.167459	0.502113	85.00%
Squeezenet 1.1 [21]	0.406253	0.840177	73.22%
Squeezenet 1.0 [21]	0.398491	0.799043	74.11%

Based on the results presented in Table 4, ResNet152 was considered the best model and therefore, it was used for the further stages of the study.

Step 3. Excluding individual categories of the study material

In order to determine whether individual categories have a negative impact on the model's accuracy, a series of trainings was conducted, during which individual categories of pornographic materials were excluded. Table 5 presents the end results concerning this experiment.

Table 5. **The process of training of the ResNet152 model excluding individual categories**

Category excluded from the training	Training loss	Validation loss	Classification accuracy
"AB/DL"	0.04375	0.444737	88.22%
"Acrotomophilia"	0.03379	0.357016	91.10%
"BDSM"	0.04756	0.329997	90.72%
"Feet fetishism"	0.03771	0.445882	89.09%
"Graviditiophilia fetishism"	0.03210	0.433339	89.47%

[5] When a model showed signs of overtraining, results of the last training cycle preceding the overtraining were presented.

table continued on the next page

"High heels fetishism"	0.03824	0.376584	90.10%
"Knee socks fetishism"	0.03755	0.342921	90.97%
"Latex fetishism"	0.04520	0.535365	88.00%
"Tattoo fetishism"	0.03572	0.504898	88.47%

As a consequence of the results presented in Table 5 above, it can be generally stated that excluding a category does not significantly influence the performance during classification of individual types of pornographic materials. Therefore, there is no reason to remove any category from the data set, as none of the categories reduce the accuracy of the model.

Step 4. Adjusting the hyperparameters of the final model

In order to select the best hyperparameters for the network, calculations were made to determine the optimal learning rate of the model [22]. The procedure was repeated after each series of five training cycles. Thus, a total of four stages of training were conducted, and each stage consisted of five training cycles.

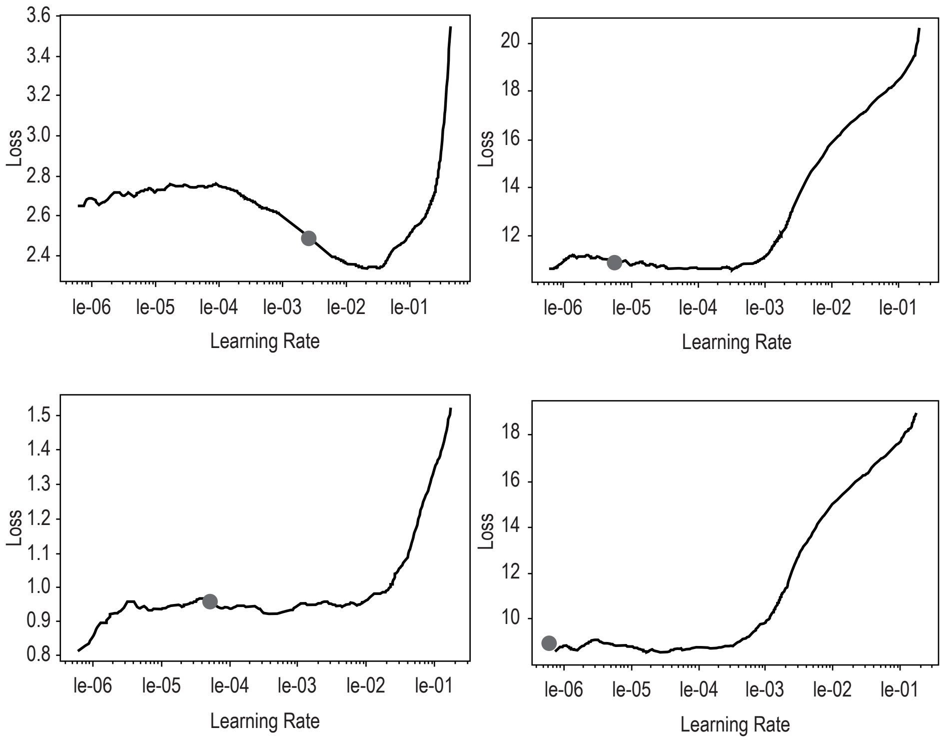

Figure 1. **Chart showing the optimal learning rate**

Figure 1 presents the optimal learning rate. Optimal learning rate is determined in the point of steepest gradient (indicated with a dot), respectively: 1e-03, 1e-05, 1e-04, 1e-06. The training was interrupted after Stage 4., as signs of overtraining were observed.

Table 6. **End results of the model after the adjustment of hyperparameters**

Stage	Training loss	Validation loss	Classification accuracy
Stage 1	1.084223	0.448046	85.89%
Stage 2	0.895491	0.357969	90.11%
Stage 3	0.865549	0.322195	91.00%
Stage 4	0.848440	0.304644	91.88%

Discussion

The aim of this study was to train a model that could classify given types of pornographic materials. The process of material selection was presented, as well as the empirical justification for the selection of the ResNet152 architecture. It was also indicated that the use of all nine categories of pornographic materials is justified. In the last stage, hyperparemeters of the model were adjusted. The final model was able to classify nine types of pornographic materials with nearly 92% accuracy, which may be considered a satisfying score.

In the pilot model [10], the photographic material was obtained through the Google search engine, using appropriate keywords related to the pre-defined types of pornographic materials. From these materials, photographs used for the training of the model were manually selected. The model was able to classify seven types of pornography with nearly 70% accuracy. A significant number of errors resulted from an incorrect assignment of the photo to a category [10].

To minimize the number of errors and increase the accuracy of the model in this study, the photographs were assigned to individual categories by experts who used restrictive inclusion and exclusion criteria while assessing the research material. Each photograph was assessed in terms of presence or absence of the classified objects. Then, a pretrained model was selected experimentally to maximize the accuracy of the classification.

The applications of deep neural networks in sexology seems to be particularly promising for at least two reasons. Firstly, a tool for automated detection of pornographic materials involving minors can be developed and used during criminal proceeding. The model presented in this paper cannot be used (without modification) for the differentiation between pornographic materials involving minors and adults; however, in the future (using different training material), the classification is possible. The model could also be used to estimate the age of depicted persons in the materials [23] and creating expert opinions for the court [15]. As a consequence, it will be possible to improve models differentiating between child pornography and other categories of materials, which may be applicable in criminal cases related to accessing pornography involving children.

Secondly, after retraining the presented model on photographs of men and women not engaging in sexual activity or on photographs of objects not associated with people, the model could be used to filter content that is inappropriate for minors. It is important, because access to pornography is not restricted on the Internet [24]; moreover, it is often not wanted [25]. In secure browsers for children[6], there are already systems that allow filtering pornographic materials involving adults, but this model would add the value of differentiating between different types of fetishistic material. Simple models of neural networks that allow differentiation between sexual activity, nudity, and the absence of thereof, could be supplemented with a more subtle analysis, involving materials that may be inappropriate for minors, but do not depict nudity or sexual activity (fetishistic materials).

Thirdly, not only does the model allow to perform a simple classification of a photograph into an appropriate category, but also it enables to search for images that do not match any of the known categories. It is possible that thanks to this approach, it will be possible to search for pornographic materials that have not yet been described, for example materials associated with rare paraphilias. This approach would require a modification of the neural network architecture for a different task – not classification, but finding similar patterns in a given type of images. Nonetheless, it seems an interesting direction of research.

Fourth, as mentioned earlier, this kind of model may allow in the future to generate better stimulus material for the emotional Stroop test [16]. In this test, a patient is asked to react as quickly as possible (usually by pressing an appropriate key) to visual stimuli that appear on the screen. Generating different experimental conditions requires a lengthy process of selecting multiple photographs, which could be automated. As a consequence, conditions could be created that would allow to achieve better clinical significance of the test.

Note

The sAI 0.3 model is registered in the OSF database and is available under DOI number: 10.17605/OSF.IO/M2X6P.

[6] Internet browsers that scan the content of a website for content inappropriate for children before displaying the website.

References

1. Short MB, Black L, Smith AH, Wetterneck CT, Wells DE. *A review of internet pornography use research: Methodology and content from the past 10 years.* Cyberpsychol. Behav. Soc. Netw. 2012; 15(1): 13–23.
2. Scott AD. *Pornografia. Jej wpływ na rodzinę, społeczeństwo, kulturę.* Gdańsk: Human Life International; 1995.
3. Hald GM. *Gender differences in pornography consumption among young heterosexual Danish adults.* Arch. Sex. Behav. 2006; 35(5): 577–585.
4. Gola M. *Neuronalne mechanizmy nałogowych zachowań.* In: Habrat B, editor. *Zaburzenia uprawiania hazardu i tzw. nałogi behawioralne.* Warszawa; Wydawnictwo IPiN [in press].
5. Hald GM, Malamuth NM. *Self-perceived effects of pornography consumption.* Arch. Sex. Behav. 2008; 37(4): 614–625.
6. Quadara A, El-Murr A, Latham J. *The effects of pornography on children and young people.* Melbourne: Australian Institute of Family Studies; 2017.
7. Nawrocka M, Drozd M, Maszczyk A, Gołaś A. *Wprowadzenie do predykcji z wykorzystaniem sztucznych sieci neuronowych i metod statystycznych.* Ogrody Nauk i Sztuk 2015; 5: 203–211.
8. Posner MI, Rothbart MK. *Research on attention networks as a model for the integration of psychological science.* Ann. Rev. Psychol. 2007; 58(1): 1–23.
9. Read SJ, Monroe BM, Brownstein AL, Yang Y, Chopa G, Miller LC. *A neural network model of the structure and dynamics of human personality.* Psychol. Rev. 2010; 117(1): 61–92.
10. Oronowicz-Jaśkowiak W. *Pilotażowy model sieci neuronowej do zastosowań związanych z klasyfikacją siedmiu typów materiałów pornograficznych.* Przegląd Seksuologiczny 2019; 1(55): 21–31.
11. Lew-Starowicz Z. *Seksuologia sądowa.* Warszawa: Wydawnictwo Lekarskie PZWL; 2000.
12. Wang Y, Kosiński M. *Deep neural networks are more accurate than humans at detecting sexual orientation from facial images.* J. Pers. Soc. Psychol. 2018; 114(2): 246–257.
13. Perez M, Avila S, Moreira D, Moraes D, Testoni V, Valle E et al. *Video pornography detection through deep learning techniques and motion information.* Neurocomputing 2017; 230: 279–293.
14. Oronowicz-Jaśkowiak W. *SexACT database v. 0.1.* www.sexailab.pl (retrieved: 2 Jan 2020).
15. Oronowicz-Jaśkowiak W. *The application of neural networks in the work of forensic experts in child abuse cases.* Postępy Psychiatr. i Neurol. 2019; 4(28): 273–282.
16. Oronowicz-Jaśkowiak W, Lew-Starowicz M. *Polish adaptation of emotional Stroop test in assessment of pedophilia – A pilot study.* Psychiatr. Pol. 2021; 55(1): 85–100.
17. Jeremy H. *Fast.ai software library*. www.fast.ai (retrieved: 10 Jul 2019).
18. He K, Zhang X, Ren S, Sun J. *ResNet-152.* www.kaggle.com (retrieved: 10 Jul 2019).
19. He K, Zhang X, Ren S, Sun J. *Deep residual learning for image recognition.* Proceedings of the IEEE conference on computer vision and pattern recognition, 2016. pp. 770–778.
20. Simonyan K, Zisserman A. *Very deep convolutional networks for large-scale image recognition.* ArXiv preprint 2014. www.arxiv.org (retrieved: 10 Jul 2019).

21. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. *SqueezeNet: AlexNetlevel accuracy with 50x fewer parameters and< 0.5 MB model size.* ArXiv preprint 2016. www.arxiv.org (retrieved: 10 Jul 2019).

22. Smith LN. *Cyclical learning rates for training neural networks.* IEEE Winter Conference on Applications of Computer Vision, 2017. pp. 464–472.

23. Yang TY, Huang YH, Lin YY, Hsiu PC, Chuang YY. *SSR-Net: A Compact Soft Stagewise Regression Network for Age Estimation.* International Joint Conferences on Artificial Intelligence 2018; 5(6).

24. Wolak J, Mitchell K, Finkelhor D. *Unwanted and wanted exposure to online pornography in a national sample of youth Internet users.* Pediatrics 2007; 119(2): 247–257.

25. Makaruk K, Włodarczyk J, Michalski P. *Kontakt dzieci i młodzieży z pornografią.* Warszawa: Fundacja Dajemy Dzieciom Siłę; 2017.