Training Effective Model for Real-Time Detection of NSFW Photos and Drawings

Dmirty Zhelonkin and Nikolay Karpov(✉)

National Research University Higher School of Economics,
25/12 Bolshaya Pechyorskaya str., 603155 Nizhny Novgorod, Russia
dmitryzhelonkin@gmail.com, nkarpov@hse.ru
https://www.hse.ru/en/staff/nkarpov

Abstract. Convolutional Neural Networks (CNN) show state of the art results on variety of tasks. The paper presents the scheme how to prepare highly accurate (97% on the test set) and fast CNN for detection not suitable or safe for work (NSFW) images. The present research focuses on investigating questions concerning identifying NSFW pictures with nudity by neural networks. One of the main features of the present work is considering the NSFW class of images not only in terms of natural human nudity but also include cartoons and other drawn pictures containing obscene images of the primary sexual characteristics. Another important considered issue is collecting representative dataset for the problem. The research includes the review of existing nudity detection methods, which are provided by traditional machine learning techniques and quite new neural networks based approaches. In addition, several important problems in NSFW pictures filtering are considered in the study.

Keywords: Image recognition · Pattern recognition · Not suitable or safe for work · Convolutional neural networks · Pornography detection

1 Introduction

The Internet as a quite free information dissemination platform provides access for numerous texts, images, and other types of uncensored information. Freedom of information on the Internet helps us make more thoughtful decisions based on a variety of sources. But sometimes Internet resources must follow the rules and engage censorship and deny public access to some of the data. Censorship can be determined by resource subjects or audience or the company, that must follow the law. One type of information which often must be censored is not suitable or safe for work (NSFW) data. NSFW is a class of information, that contains nudity, intense sexuality, pornography, obscene texts, and other disturbing content. In our paper, we limit NSFW content to pornographic pictures. In our work such type of content is named positive class, since this class is our aim to find. Hence, normal or suitable for work images are named as a positive class. Evidently, the

W. M. P. van der Aalst et al. (Eds.): AIST 2019, CCIS 1086, pp. 301–312, 2020.
https://doi.org/10.1007/978-3-030-39575-9_31

NSFW filtering images problem exists for the web resources, which are opened for uploading and publishing user's content with images. If we moderate such a resource we face three main tasks. First of all, a moderation process "by hand" takes too much time to analyze all the images. Big lag between uploading picture and publishing makes web resource less attractive for users.

Secondly, it should be noted that the research is provided for the Russian market and law. The investigation of law enforcement practice leads us to the fact that there is no single-valued definition of prohibited pornographic content in different countries. To determine what is the violation of the law, and what is not, usually uses expert assessments. That is why we should take into consideration what users determine as not suitable content. As a result, the definition of such content should be wide. In this work, we use the following definition: image is not suitable for work if it contains not covered full breast's areola and genitals or part of it. These things can be drawn or photographed. Thirdly, we must minimize unsuitable pictures as much as possible. Otherwise, Internet resources can break the law or lose the audience. In other words, the aim is to minimize false positive rate, however false negative rate is less important for us.

Our experiments indicate that modern pretrained neural networks show good results in our problem on many training set configurations. As a result the most challenging task is to collect representative and diverse dataset. The dataset that most appropriate to out NSFW class definition.

The main contribution of this research is the creation of a fast automatic image pre-filtering algorithm with a minimal false positive rate. In order to archive the goal we:

- review existing nudity detection methods;
- make a procedure to collect the data for training purpose;
- train chosen neural networks;
- compare precision, recall, and speed of our approaches with others.

The remainder of this paper is organized as follows: related works overview is discussed in Sect. 2. Section 3 provides information about the dataset formation procedure, while Sect. 4 presents a proposed methodology, experimental evaluation measures, and the baselines. In the end, we conclude and discuss our results on the test dataset.

2 Background

Filtering obscene images is not a novel task. In 1996 Fleck et al. [1] made one of the first attempts to create an automatic NSFW image detector. In 2012, Marie Short, together with co-authors, [2] conducted a study of works for 2002–2012 on filtering obscene images. Further, we observe several most important works from our point of view.

2.1 Skin-Based and Bag-of-Visual-Word Approaches

Approaches which are based on the search and analysis of regions with human skin are widely presented in the literature [1,3,4]. For instance, Margaret Fleck and co-authors in 1996 [1] proposed a two-step algorithm. The first stage is the collection of images with a large proportion of pixels with a color close to the body. The second stage is a geometric analysis of “skin” areas.

Another widely used approach in the image classification task is the BoVW (Bag-of-Visual-Words). A dictionary of such features or words can be prepared in several ways. The most popular methods are SIFT (Scale-invariant feature transform) [5] and SURF (Speed Up Robust Features) [6].

Some of the first people who applied BoVW in order to solve the problem of filtering obscene images were Deselaers with co-authors [7]. They showed results superior to all previous algorithms based on the search and analysis of regions with human skin. Sandra Avila et al. in 2009 [8] developed the BoVW based approach and suggested to add color information to SIFT features (Hue-SIFT).

2.2 Convolutional Neural Networks

Today convolutional neural networks (CNN) show the best results in classification and other image tasks. The most well-known example of CNN application in the classification problem, which would significantly exceed all other algorithms, is the victory of AlexNet [9] in ILSVRC 2012 (ImageNet Large Scale Visual Recognition Challenge) [10].

The results shown on ImageNet not only allowed us to increase quality of the classification on a particular task but also allowed us to significantly improve the classification results for many other problems. It turned out that the model trained on ImageNet can be retrained (fine-tuned) for other tasks, even not only classification ones. If we consider the classical architecture of CNN, it turns out that the majority of layers in the neural network generate a feature vector. Then, machine learning model [9], fully connected layers [11] or convolutional layers [12] solve classification task based on the generated features. As a result, the network trained on ImageNet generates features that can effectively represent a variety of images and objects. Some of the most popular models are (ResNet [13], Inception [14], VGG [11] and their modified versions. In this paper, we will use one of the modifications of the network Inception-Xception [15], rethinking and optimization of several known CNN architectures – MobileNet [12] and one of the variants of the architecture ResNet – ResNet-50. All these types of CNN showed good results on the sample ImageNet, but at the same time, learns quickly because of compact and effective architectures.

In the problem of filtering pornographic images, neural networks are also used and show better results compared to other methods. One of the first successful applications of CNN for this task is the work of Muhammad Mustafa in 2015 [16]. He retrained two well-known architectures: AlexNet [9] and GoogleNet [17] on a sample of NPDI [18]. Sample NPDI is one of the few public samples for comparison of models filtering obscene images. Among newer works trained on the NPDI sample, one can distinguish the work of 2017 by Maurizio Perez and co-authors [19]. They experimented with filtering pornographic videos using the same network architectures and slightly improved neural network training techniques, showing better results for this dataset.

Today at least two free models are available for comparison. The first model[1] is the ResNet-50 network with a reduced number of folds on each layer, pretrained on ImageNet and trained by specialists from Yahoo on its own proprietary sample. The second model[2] was trained to filter pornographic photos, not drawings. The model has the same architecture as the Yahoo network, also trained on a closed sample, but its main purpose is the classification of poses during sexual intercourse. Objective of the project is significantly different from ours, but the output of the network can be reduced to our formulation of the problem.

3 Dataset Formation

Generally, CNN demonstrates satisfactory results only after training on a huge dataset, but gathering sample from scratch is a challenging task since collectors should strongly identify criteria for every class, accurately choose data sources, elaborate gathering procedure and finally collect the data. All these points require time, resources and certain competencies. As a result, several carefully prepared datasets were considered.

3.1 Existing Datasets

There are several datasets for nudity detection task [8,20]. But they are not applicable to our problem as nudity and our NSFW class definition are not the same. In other words, picture with open genitals or with quite small not covered skin area is NSFW content, but it cannot be uniquely defined as nudity. Moreover usually nudity class does not contain drawn pictures. Also, there is Pornography database [18]. In contrast to the previous two, this dataset contains photos and drawn pictures and determine two classes very close to ours. The dataset is a collection of films and selected frames from the films. The benign images also divided into two subclasses: “Easy” and “Difficult”. We use only the frames for our task. Close investigation ofPornography database shows us several problems. The first problem is label correctness. NSFW class consists of frames from films with NSFW content, but it seems every n-th screenshots was placed into sample without label checking. But pornographic videos from the database have benign scenes at the beginning, also pornographic scenes have frames which contain only faces, interior and etc. According to our negative class criteria, 67 images from 200 randomly taken pictures from pornographic (only 33,5%) can be named as NSFW. In the same time we did not find any wrong labeled images in the positive class as a result a benign class is correct.

[1] https://github.com/yahoo/open_nsfw.

[2] https://github.com/ryanjay0/miles-deep.

The second problem is frame quality. A big part of the dynamic frames are blurred and several NSFW pictures cannot be absolutely classified by human.

The third is a size of the dataset. Pornography database has 6387 and 10340 images for negative and positive classes correspondingly. Since CNN requires huge dataset for training and clear labels we suppose that such number is not enough for training effective classifier from scratch or fine-tuning. Moreover, our NSFW definition mentioned above is wider than the negative subset of this dataset. As a result, the Pornography database is not perfectly suitable for our tasks, but it was used in several experiments. The results are presented in the Experimental Results section.

3.2 Dataset Gathering Procedure

Consequently, we have to collect our own dataset. The main issue for collecting dataset is how to make it representative and able to be a good source for training classifier. In order to resolve this problem we apply an iterative procedure:

- Gather initial sample;
- Conduct a validation experiment;
- Investigate misclassified items.

If there is obvious misclassified subclass then add more images from the most problem subclass and repeat the second stage. Otherwise end procedure.

Validation experiment in the second stage is testing of fine-tuned model pretrained on ImageNet. As testing model architectures were considered some widely used neural network such as AlexNet, ResNet and Inception. Searching the best model type for validation experiment was not provided. Number of epochs determined by the moment when accuracy improvement on validation set was less than 0.005. Usually, this figure was between 3 and 10 epochs. Certainly, we can get better results if we train network longer, but we believe that such a criterion makes it possible to maintain a balance between time and result. We fine-tune only fully connected and the last convolutional layers. Intermediate dataset was split into training (80%), validation (10%) and testing (10%) sets. We explore the test set in order to observe errors and find the most problematic picture cluster. Subsets using strategy is to train network on the training data, choose the best model according to error on validation images and investigate quality on the test set.

The testing experiment strategy is not optimal, but it gives us information which helps us to improve our sample.

3.3 Application of the Iterative Procedure

At the first few iterations we exclude drawn pictures from NSFW class to simplify the searching misclassified subclasses process.

First few iterations show us that we should add more safety images with big flesh color areas and with human faces. These results are quite obvious since a huge part of NSFW images contain faces. It should be noted that every iteration network achieves more than 90% accuracy on the validation set. It can testify to the fact, that the architectures can be well trained on a specific dataset (20000 of NSFW images and 20000 of suitable data), but new examples show poor models generalizing power even with extensive use of augmentations. Since chosen architectures show great generalization on a more difficult challenge of classification of 1000 classes, we decide that the problem is in the dataset and we need to collect more data. Our aim was to collect around half a million pictures for both classes. While expanding dataset especially NSFW class we faced a problem. Whereas proven benign images are placed openly for research purposes datasets such as VOC2012, ImageNet, INRIA person and others, gathering tremendous number of pornographic images is a challenge. First of all, as it was already mentioned, there is no universal NSFW class definition. As a result, we observe that many images which are placed in pornographic resources are safe according to our definition of negative class, but at the same time, there is quite a small number of resources with similar NSFW marks as ours. As the main source of such images we used Danbooru sites and microblogs, for instance tumblr.com. Such microblogs allow users to post any images. Some users publish NSFW images in their blogs and tag images or whole blogs. In order to collect the positive class we rely on description of those images from source, because we cannot check every image manually. We did not use prepared crawlers with collected NSFW URLs pictures[3,4] since they did not exist at the time of dataset gathering formation.

From the web sites, which support picture tagging, we use images which are marked by "NSFW", "Not safe" and other similar tags or at least one tag which describes content as not suitable in our definition. From the microblogs without picture tagging we use images which are placed in microblog with the name containing NSFW keywords or phrases for example sex, naked, porno and etc.

Drawn pictures are grouped in accordance with discovered description. In other words, we firstly collect not suitable pictures according to the description. Then we gather similar in style and content drawn images, for example benign anime and hentai. And finally, we add this drawn set to the dataset and improve it by the iterative procedure. As a result, we have a structure of dataset as it is described below.

Negative class:

- 250000 – photos of the real world;
- 120000 – drawn pictures;
- 35000 – photos of people with big open skin areas;
- 15000 – face images;
- 25000 – photos of people in the crowd and alone;
- 5000 – desert images;
- 420000 – total positive sample size.

[3] https://github.com/GodelBose/NSFW_Detection.
[4] https://github.com/EBazarov/nsfw_data_source_urls.

Positive class:

- 300000 – images with natural human nudity;
- 120000 – drawn NSFW pictures;
- 420000 – total negative sample size.

3.4 Upgrading of Dataset A

We observe that neural networks after training on dataset A become too “strict” in terms of negative class. In other words people in bikini and other types of very open suits are constantly qualified as NSFW images by many network architectures and learning strategies. Quality metrics are presented in the Experimental Results section.

More precise investigation of dataset A shows us that there are quite a lot of examples of wrongly gathered images from negative class. It was discovered that some blogs without drawn pictures with NSFW keywords in name contain mostly erotic images with covered genitals and breast. Moreover, sometimes prohibited class definition in microblog changes over time. For instance, sometimes the author places in his blog mostly erotic images and in some time posts only images, which correspond to our NSFW class specification. Another problem appears when people mark benign pictures with few NSFW tags. We suppose that it can be explained by different understanding of what pornographic image is and a human factor. Benign images are selected on the desirable level. We haven’t discovered any wrongly collected pictures for the negative class. Furthermore, networks trained on dataset A make very few mistakes in classification photos of the world. Most parts of errors are pictures that contain people in revealing clothes.

Drawn pictures have few wrongly marked instances (less than 5%), but classification quality of such images is quite poor (80% in contrast to 95% for the photos). We suppose that the reason for it is imbalanced subclasses of drawn pictures in negative class. The largest part of collected benign drawn content is anime, but positive class contains not only NSFW anime images or hentai.

As a result, we decide to upgrade dataset A by the next improvements:

- Use stronger NSFW photos gathering conditions;
- Increase benign images proportion in positive class;
- Balance drawn pictures distribution in both classes and increase its proportion in the sample.

Also, we modified conditions of NSFW content:

- From Danbooru websites, we collect images marked by “NSFW”, “not safe” and other at least three similar tags which describe content as not suitable in our definition.
- From microblogs, we use image placed in microblog with a name containing NSFW keywords for example sex, porno and 98 of 100 images from this blog must be NSFW (images must be checked by the experts).

Since new conditions require manual checking we were forced to reduce new sample size. Also, as we have not saved pairs “images - tags” and “image - blog name” we had to recollect positive class. The final distribution of the new sample of dataset B described below.

Negative sample:

- 30000 – photos of the world;
- 24000 – photos of people with big open skin areas;
- 23000 – drawn pictures;
- 77000 – total positive samples size.

Positive sample:

- 44000 – images with natural human nudity;
- 33000 – drawn NSFW pictures;
- 77000 – total negative sample size.

Models trained on dataset B show more appropriate results and the last sample was chosen as the main.

4 Experimental Results

In this section we describe conducted experiments with the collected data and some other datasets.

4.1 Proposed Methodology

As it was already discussed convolutional neural networks are state of the art in various tasks. So CNN were chosen as the main classification method. For training networks fine-tuning was applied, since using pretrained networks significantly reduces training time and allows to achieve better results. Xception, MobileNet and ResNet-50 architectures were used for training, since these networks show competitive results in ILSVRC and other competitions, they are quite fast, compact and they have pretrained weights on ImageNet in Keras. We did not provide experiments with architecture modifications since vanilla networks demonstrate desirable results. Also we have limited computational budget and we cannot provide many experiments.

Most networks were trained in Keras framework with Tensorflow backend using Adam optimizer with learning rate 1e–4. We reduce learning rate by a factor 2 when validation accuracy did not improve for a three epochs. In all experiments we train networks for 100 epochs and choose the best by validation accuracy. Random flip, rotation, shear, brightness and gamma jittering augmentation techniques were used in order to improve generalization power. We convert Yahoo model in Keras for fine-tuning. Also Caffe framework was applied for testing vanilla Yahoo and Miles-Deep pretrained networks.

4.2 Evaluation Measure

In order to evaluate proposed approach we run several experiments and write the results of the experiments to the Table 1. Columns "Precision (neg/pos)" and "Recall (neg/pos)" show precision and recall of testing negative and positive subsets correspondingly. For other two columns "Recall for NSFW photos" and "Recall for NSFW drawings" we split the testing positive NSFW subset into photos and drawn pictures respectively.

Table 1. Evaluation on dataset B

Architecture	Procedure	Frame-work	Precision (neg/pos)	Recall (neg/pos)	Recall for NSFW photos	Recall for NSFW drawings
Yahoo_ResNet50_1by2	–	Caffe	0.81/0.98	0.99/0.76	0.87	0.64
Miles-Deep_ResNet50_1by2	–	Caffe	0.56/0.98	1.0/0.19	0.26	0.09
Yahoo_ResNet50_1by2	Fine-tuning	Keras	0.97/0.95	0.95/0.96	0.97	0.93
ResNet50	Fine-tuning	Keras	0.95/0.96	0.96/0.95	0.97	0.91
ResNet50 (grayscale)	Fine-tuning	Keras	0.95/0.95	0.95/0.95	0.96	0.92
Xception	Fine-tuning	Keras	0.96/0.98	0.98/0.96	0.97	0.94
Xception	Re-training	Keras	0.96/0.75	0.86/0.97	0.98	0.96
MobileNet	Fine-tuning	Keras	0.95/0.94	0.95/0.96	0.97	0.93

In Table 1 the first column notes a network structure like ResNet50, Xception or MobileNet. Such a name like Yahoo_ResNet50_1by2 means a particular implementation of the corresponding architecture. Column "Procedure" demonstrates whether we use fine-tuning procedure, pretraining on additional data or use pretrained model as-is. All models were fine-tuned only on the dataset B and pretrained on ImageNet. In retraining procedure we pretrain model using dataset A and then fine-tuned on dataset B. Next we specify a framework which was used: Caffe or Keras. As we can see, the best precisions shown by Xception network both fine-tuned and retrained. Also, we can see that fully trained Xception network works slightly better on drawn pictures than fine-tuned one. The downside is that a fully retrained network has much less precision of NSFW class. It should be noted that MobileNet architecture shows competitive precisions. One more interesting experiment with grayscale images has been done. We evaluated, how useful color information for the networks in this task. Since we have only pretrained weights for color images, standard ResNet50 architecture was fine-tuned on 3 channel gray pictures. Surprisingly results are very close to the color modification. This fact gives us possible way to speed up inference.

Also several experiments were conducted on the NPDI dataset with Xception architecture pretrained on ImageNet. Training procedure was the same as in experiments with dataset B. We can see in Table 1 that pretraining on the dataset B is better than other options. As it was already mentioned the definition of NSFW content is different from ours. Consequently using of NPDI train sample leads to metrics improvements. Also since NPDI dataset has noisy labels we did

Table 2. Evaluation on dataset NPDI

Architecture	Dataset	Precision (neg/pos)	Recall (neg/pos)
Xception	NPDI	0.93/0.96	0.94/0.95
Xception	A	0.63/0.88	0.85/0.69
Xception	B	0.92/0.76	0.50/0.97
Xception	A; NPDI	0.91/0.91	0.85/0.95
Xception	B; NPDI	0.95/0.98	0.97/0.97

not try to improve these achievements or try to use NPDI as pretraining sample and only convinced of the superiority of dataset B in close task (Table 2).

We conduct several speed tests except quality evaluation. Speed is important characteristic of algorithms especially in commercial usage. We compare several architectures on machine with NVIDIA GTX 1080, CUDA 9.0, cuDNN 7.3.1, Keras 2.2.4 and TensorFlow 1.12. It was done with 64 batch size and constant image in RAM (Random Access Memory).

As we can see in the Table 3, the slowest result shown by ResNet50. Xception architecture performs just a little bit faster. Yahoo_ResNet50_1by2 runs more than 2 times faster on Caffe than vanilla ResNet50 on Keras. The fastest architecture in this table is MobileNet.

Table 3. Images per second performance.

Architecture	Framework	Images per second
MobileNet	Keras	596
MobileNet v2 [21]	Keras	548
Yahoo_ResNet50_1by2	Keras	423
VGG 16	Keras	230
Xception	Keras	206
ResNet50	Keras	203

If we sum up our results, Tables 1 and 3 demonstrate that MobileNet shows the best tradeoff between speed and quality. That is why it is chosen to run in production system. You can check the inference of our neural network on the demonstration version, which is available at fapme.me[5] with Russian user interface.

5 Conclusion

This research is dedicated to the topic of creation of highly accurate convolutional neural network for filtering not suitable or safe for work images. As a

[5] https://fapme.me.

NSFW class of images we consider not only photos in terms of natural human nudity, but also include cartoons and other drawn pictures containing obscene images of the primary sexual characteristics.

In order to achieve our goal, we review existing nudity detection methods which are provided by traditional machine learning techniques and quite new neural networks based approaches. In addition, several important problems in NSFW pictures filtering are considered in the study.

We have found that existing neural networks trained on NSFW the dataset achieve more than 90% accuracy on validation. Consequently the most challenging task is to collect representative and diverse dataset.

The results of the work are:

- dataset iterative collecting procedure;
- artificial neural networks trained on the dataset;
- comparison with other existing approaches.

The best tradeoff between precision, recall and speed has been achieved using pre-trained MobileNet v1 after fine-tuning procedure on our dataset.

References

1. Fleck, M.M., Forsyth, D.A., Bregler, C.: Finding naked people. In: Buxton, B., Cipolla, R. (eds.) ECCV 1996. LNCS, vol. 1065, pp. 593–602. Springer, Heidelberg (1996). https://doi.org/10.1007/3-540-61123-1_173
2. Short, M.B., Black, L., Smith, A.H., Wetterneck, C.T., Wells, D.E.: A review of Internet pornography use research: methodology and content from the past 10 years. Cyberpsychol. Behav. Soc. Netw. **15**(1), 13–23 (2012)
3. Flores, P.I.T., Guillén, L.E.C., Prieto, O.A.N.: Approach of RSOR algorithm using HSV color model for nude detection in digital images. Comput. Inf. Sci. **4**(4), 29 (2011)
4. Platzer, C., Stuetz, M., Lindorfer, M.: Skin sheriff: a machine learning solution for detecting explicit images. In: Proceedings of the 2nd International Workshop on Security and Forensics in Communication Systems, pp. 45–56. ACM (2014)
5. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision **60**(2), 91–110 (2004)
6. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006). https://doi.org/10.1007/11744023_32
7. Deselaers, T., Pimenidis, L., Ney, H.: Bag-of-visual-words models for adult image classification and filtering. In: 19th International Conference on Pattern Recognition, ICPR 2008, pp. 1–4. IEEE (2008)
8. Lopes, A.P.B., Avila, S.E.F.d., Peixoto, A.N.A., Oliveira, R.S., Araújo, A.d.A.: A bag-of-features approach based on hue-sift descriptor for nude detection. In: Proceedings of the XVII European Signal Processing Conference (EUSIPCO), Glasgow, Scotland (2009)
9. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)

10. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 248–255. IEEE (2009)
11. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
12. Howard, A.G., et al.: MobileNets: efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017)
13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
14. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. CoRR abs/1512.00567 (2015). http://arxiv.org/abs/1512.00567
15. Chollet, F.: Xception: deep learning with depthwise separable convolutions. arXiv preprint pp. 1610–02357 (2017)
16. Moustafa, M.: Applying deep learning to classify pornographic images and videos. arXiv preprint arXiv:1511.08899 (2015)
17. Szegedy, C., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
18. Avila, S., Thome, N., Cord, M., Valle, E., Araújo, A.D.A.: Pooling in image representation: the visual codeword point of view. Comput. Vis. Image Underst. **117**(5), 453–465 (2013)
19. Perez, M., et al.: Video pornography detection through deep learning techniques and motion information. Neurocomputing **230**, 279–293 (2017)
20. Jones, M.J., Rehg, J.M.: Statistical color models with application to skin detection. Int. J. Comput. Vis. **46**(1), 81–96 (2002)
21. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: MobileNetV2: inverted residuals and linear bottlenecks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4510–4520. IEEE (2018)