



NETSPARK

Pornography Algorithms Comparison

For more than a decade Netspark Ltd. Has been developing sophisticated algorithms to identify harmful content in texts, images, and videos and embedded them into its products – parental control applications (Canopy, Netspark mobile), ISP solutions (now expanding outside Israel), and solutions for the B2B market (UGC companies, moderation companies and more).

We have decided to put our algorithms to the test and compared its accuracy with the accuracy of the algorithms of the tech giants – AWS, Google, and Microsoft.

Sure we were skeptical at first and were told that we have no chance against such powerful companies, with access to the brightest engineers on earth and with unlimited access to funds. Nevertheless, we decided to take a good look at ourselves and find out our real value by putting our pornography models to the test.

The methodology

If we are looking to classify pornographic content and block 100% of pornography, we can simply classify all content as containing pornography. This way, all content containing pornography shall actually be classified as such, but we will also classify all non-pornographic content as containing pornography. i.e. we would receive a lot of false-positive results.

This is exactly where the challenge lies, we need to be very accurate if we want to detect and classify only content containing pornography as such and ignore non-pornographic content.

We first need to understand that the population is not proportional. For example, if we take 1,000 images that only 10 of them contain pornography and use a specific model that identifies 90% of the pornography as such (10% mistakes) and another model that identifies non-porn images in 90% of the times (10% mistakes) - statistically speaking, the classification will be correct in 90% of the cases and incorrect in 10% of the cases.

So according to the example above, on every pornographic image, we classify correctly, we will misclassify 10 images. This means that out of 1,000 images, we will classify 108 as containing pornography - 99 mistakenly and 9 correctly.

The above example reveals the simple truth - image analysis is all about compromise. Either we choose to be more accurate and in this case, all classifications that the algorithm makes are correct, but the algorithm misses quite a few images, or we choose to be less accurate and some of the classifications shall be wrong.

Hence, in order to be able to compare the accuracy of the different algorithms, we needed to select the same accuracy level for all of them. Therefore we tried to set the accuracy level so that the algorithms will identify as many pornographic images/videos as possible. We were surprised to find out that our models managed to be 99% accurate, whereas the closest competitor managed to be only 96.5% accurate. Therefore, we restricted our models to the same accuracy, 96.5%.

Database

Running such a comparison is not only about setting the same accuracy levels. In order to make the comparison unbiased, we contacted Prof. Dr. Sandra Avila from the Institute of Computing at the University of Campinas (Unicamp) in Brazil.

Prof. Avila has put together a pornography and nudity database in the framework of writing a computer vision article[1]

The database contains two categories, each category contains 400 videos and also on average ~30 frames from each video (total 16,727 frames):

1. Pornography – the DB consists of several genres of pornography and depicts actors of many ethnicities, including multi-ethnic ones.
2. Non-pornography –
 - A. Easy – 200 videos were chosen at random from general public purpose video networks.
 - B. Difficult – 200 videos selected from textual search queries like “beach”, “wrestling”, “swimming”, which would be practically challenging for a detector – [example](#).

[Here](#) is a link to more information about the database and the way it was gathered.

After obtaining access to a comprehensive pornography database that was compiled by a third party whom we had no prior contact with, we went straight to work and started our comparison.

Comparison

Models we compared between

In order to conduct the comparison we used the following APIs:

1. **Netspark** - <https://rapidapi.com/user/Netspark>
2. **Microsoft Azure** - <https://docs.microsoft.com/en-us/azure/cognitive-services/content-moderator/image-moderation-api>
3. **AWS** - <https://docs.aws.amazon.com/rekognition/latest/dg/moderation.html>
4. **Google** - <https://cloud.google.com/video-intelligence/docs/analyze-safesearch>

Comparison Method:

- **Netspark** - Video was classified as pornographic only if models returned results of high probability - "Very likely" and "Likely".
- **Azure** - Here we had some issues conducting a non-biased comparison since we could not find Azure's video engine (we found only an engine that runs on frames), so we tested the frames model on the frames taken from the same video and decided that if more than one (1) frame is classified as pornography, then the video is marked as such - we should mention that this was the method that benefited Azure most (other methods brought their accuracy even lower).
- **AWS** - AWS's models count the number of nude frames in a video, so we decided that if more than one percent (1%) of the frames were classified as porn we classified the entire video as containing pornography.
- **Google** - Video was classified as pornographic only if models returned results of high probability - "Very likely" and "Likely".

Videos comparison results

The table below shows the percentage of videos that are classified as pornography by the compared algorithms:

Category	Explanation	Netspark	Azure	AWS	Google
Non-porn - Easy	False-positive - lower percentage is better (0%)	0.50%	6.47%	0.50%	1.00%
Non-porn - Difficult	False-positive - lower percentage is better (0%)	10.00%	34.00%	31.00%	16.50%
Porn	True-positive - higher is better (100%)	96.50%	94.26%	96.50%	94.75%

We were very surprised but also very happy to see the results specified in the table above. **It is evident that Netspark's models are significantly more accurate than all the tech giants' models.**

If we examine the results of the **Non-porn - Easy category**, we can see that Netspark and AWS showed similar results when mistakenly identifying only 0.5% as pornography, whereas Microsoft identified 6.47% of the content as containing pornography and Google 1%.

If we examine the results of the **Non-porn - Difficult category**, we can see that Netspark mistakenly identified only 10% of the content as pornography, while Google mistakenly identified 16.5% of the content as pornography (65% more errors than Netspark) and Microsoft and Amazon mistakenly identified more than 30% of the content as pornography (More than 3 times as many errors than Netspark).

If we examine the results of the **Porn category**, we can see that only Netspark and Amazon identified 96.5% of the content as pornography.

As specified above, Netspark is able to detect 99% of the pornography when we take into account the results that our engine classifies as possible pornography, Microsoft and Google are able to identify less than 95%.

Frames comparison results

The pornographic database we obtained included a mix of pornographic frames and non-pornographic frames under the Porn category. We could have classified the frames according to our understanding but this would mean that we temper with the database's authenticity and questions about the integrity of the comparison could have risen. This is why we decided to run the comparison only on the Non-porn category.

Since non-porn videos do not contain pornographic frames and porn videos contain a mix of pornographic and non-pornographic frames, we can examine the false positive rate - i.e. how many times did the models mistakenly classify a frame as pornography, while it is not.

The table below illustrates the false positive rate, i.e. the percentage of non-porn videos that according to the algorithms contained pornographic frames:

Category	Explanation	Netspark	Azure	AWS	Google
Non-porn - Easy	False-positive - less is better (0%)	0.00%	6.47%	3.98%	10.45%
Non-porn - Difficult	False-positive - less is better (0%)	3.50%	34.00%	22.00%	37.50%

We were totally astounded when we examined the above results. **We never imagined that we would be 6 times more accurate than the closest competitor (!)**

With Google's algorithm, when we do the same and take into account also the results classified as possible pornography, Google's algorithm would result in 95.5% accuracy of porn detection. In **Non-porn Easy** category the mistake rate would be 1.5% and in **Non-porn Difficult** the mistake rate would be 33%

Conclusion

Our models are far more accurate than our competitors.

Not only that our models produce the highest levels of true-positive results, but our false-positive classifications are also significantly lower than the competition – **in difficult frames, our models are more than 6 times more accurate than the nearest competitor.**

For more information please contact us at:

sales@netspark.com

[1] Sandra Avila, Nicolas Thome, Matthieu Cord, Eduardo Valle, Arnaldo de A. Araújo. "Pooling in Image Representation: the Visual Codeword Point of View." Computer Vision and Image Understanding, volume 117, issue 5, p. 453-465, 2013. DOI: <http://dx.doi.org/10.1016/j.cviu.2012.09.007>