

Udacity ML NanoDegree Project Proposal

Nudity / NSFW Detection In Images Using Deep Learning

Deepanshu Yadav

August 2019

CHAPTER 1

Domain Background

Internet is flooded with images. Many images posted are uploaded without any censorship. The images may belong to NSFW (Not suitable For Work). The use of censorship on images especially belonging to Nudity is still debatable as it may depend on culture to culture. But the following repercussions can still happen if such content is not actively monitored :

- The usage of nude content for prostitution and human trafficking. Many web pages are full of with ads using this content to promote such activities.
- Censoring such content can prevent underage kids to view such content. Such content is also actively promoting child pornography.

Therefore it is the job of content hosting sites to prevent such misuse by censoring.

CHAPTER 2

Problem Statement

The goal of this project to detect and tell severity namely explicit nude , animation , semi nude etc.Also it should be independent of gender.

That is if the given image contains nudity then it should identify of what severity other wise show non nude.

CHAPTER 3

Datasets

A few of the sources where dataset is available.

- CTCI community

<http://www.cti-community.net/>

: This discord community hosts nsfw content particularly animated images. The data can be scraped from here.

- B praneeth's dataset

https://archive.org/details/NudeNet_classifier_dataset_v1

This dataset consists 1,78,601 images mostly thumbnails crawled from Pornhub, 1,21,644 images from Reddit (mostly nsfw subreddits) and 1,30,266 from GantMans dataset. The GantMan's model is <https://github.com/infinitered/nsfwjs> his training data is https://github.com/alex000kim/nsfw_data_scraper

The author author image augmentation using <https://github.com/mdbloice/Augmentor> and the end result is 256X256 image. Also for sfw 68,948 from Facebook, 98,359 from GantMans dataset and 55,137 from Reddit(sfw sub reddit). So it's about 18 GB dataset.

We will classify dataset into 4 categories which are mentioned in next chapter . Also only thing missing is the semi - nude(refer chapter 4). which can be collected from instagram and semi-nude subreddit which will be done by me.

Note : The reason we have taken sfw images size (size means total number of images) smaller than nsfw images because many images are of human. Therefore if increase the size of images of nsfw category the model will become more of human detector than nsfw detector.

CHAPTER 4

Solution

We will use the pretrained model to detect from here.

<https://github.com/bedapudi6788/NudeNet-models/tree/master/v1>

The pre trained model tell us whether is non-nude / nude. We will further develop model to categorize into 4 categories.

- Animated
- Explicit Nude
- Semi Nude
- Porn

CHAPTER 5

Existing Benchmarks and Evaluation

<https://dzone.com/articles/nudity-detection-and-abusive-content-classifiers-r>

. This article discusses various issues with nsfw censoring.

<https://dzone.com/articles/nudity-detection-and-abusive-content-classifiers-r>

. This article compares existing benchmarks for nsfw detection.

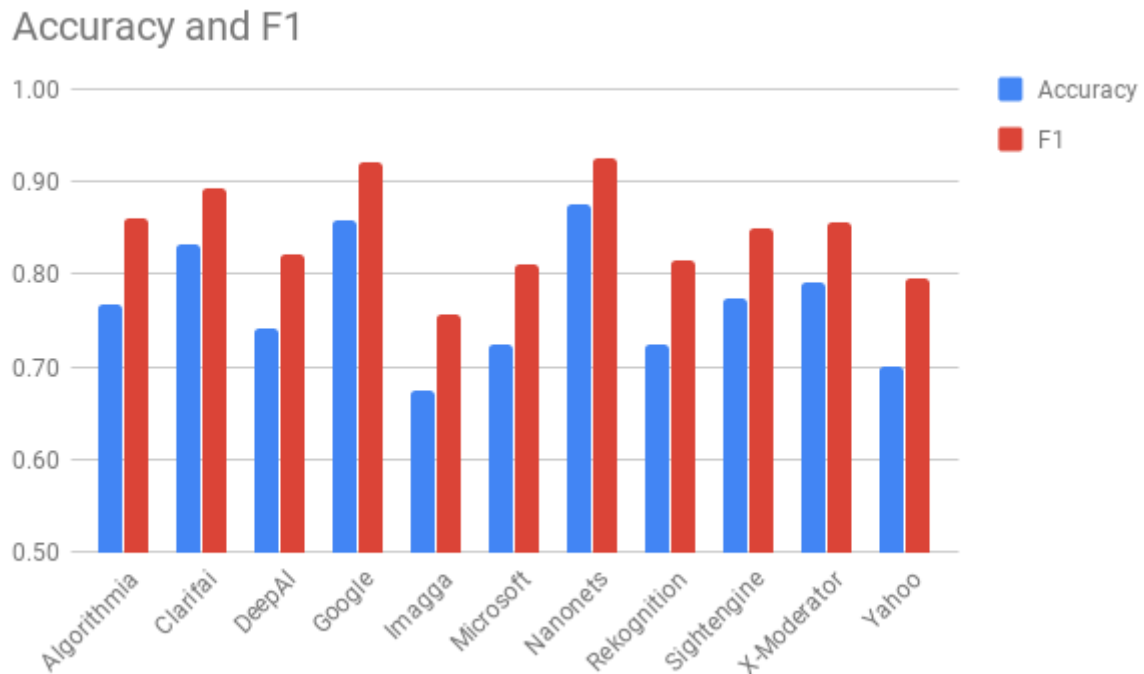


Figure 5.1: Compares the F1 score and accuracy of various API available for NSFW detection.

Some of the benchmark are :

- <https://github.com/deppomf/DeepCreamPy> . A very good model to classify nsfw animated content.
- <https://github.com/bedapudi6788/NudeNet> A very good model for detection and classification.
- <http://pury.fi/> Their API can be used to censor nsfw content.

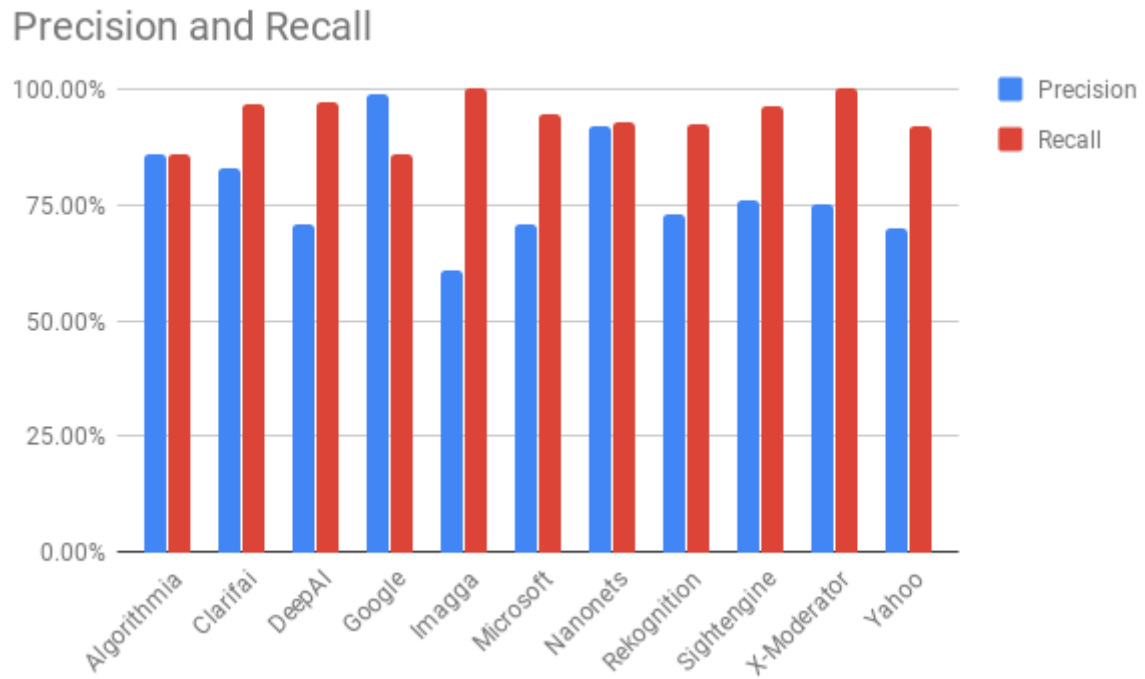


Figure 5.2: Compares the Precision and Recall of various API available for NSFW detection.

The following evaluation metrics will be used :

- Precision
- Recall
- F1 -Score
- Accuracy

CHAPTER 6

Design

- Data Collection and Augmentation As mentioned <https://github.com/mdbloice/Augmentor> will be used for augmenting. Here is an example which the original author <https://medium.com/@praneethbedapudi/nudenet-an-ensemble-of-neural-nets-for-nudity-detection-and-censor> has used.

```
# Random rotation, flips, zoom, distortion, contrast, skew and brightness
pipeline.rotate(probability=0.2, max_left_rotation=20,
                max_right_rotation=20)
pipeline.flip_left_right(probability=0.4)
pipeline.flip_top_bottom(probability=0.8)
pipeline.zoom(probability=0.2, min_factor=1.1, max_factor=1.5)
pipeline.random_distortion(probability=0.2, grid_width=4, grid_height=4,
                           magnitude=8)
pipeline.random_brightness(probability=0.2, min_factor=0.5, max_factor=3)
pipeline.random_color(probability=0.2, min_factor=0.5, max_factor=3)
pipeline.random_contrast(probability=0.2, min_factor=0.5, max_factor=3)
pipeline.skew(probability=0.2, magnitude=0.4)
```

I am also thinking of my own preprocessing steps.

- Training

Initial suggestion would be to use resnet 50 or Xception (used by the above mentioned author).

- Evaluation

We will use the metrics mentioned previously and can also deploy the model for testing.

Bibliography

- [1] https://archive.org/details/NudeNet_classifier_dataset_v1
- [2] <http://www.cti-community.net/>
- [3] <http://pury.fi/>
- [4] <https://github.com/bedapudi6788/NudeNet>
- [5] <https://github.com/bedapudi6788/NudeNet-models/>
- [6] <https://dzone.com/articles/nudity-detection-and-abusive-content-classifiers-r>
- [7] <https://github.com/deeppomf/DeepCreamPy>
- [8] <https://medium.com/@praneethbedapudi/nudenet-an-ensemble-of-neural-nets-for-nudity-detection-and-censoring-c8fcef6cc92>
- [9] <https://medium.com/@praneethbedapudi/nudenet-an-ensemble-of-neural-nets-for-nudity-detection-and-censoring-d9f3da721e3>
- [10] <https://github.com/bedapudi6788/NudeNet-models/tree/master/v1>
- [11] <https://docs.aws.amazon.com/sagemaker/latest/dg/image-classification.html>
- [12] Deep residual learning for image recognition Kaiming He, et al., 2016 IEEE Conference on Computer Vision and Pattern Recognition
- [13] <https://towardsdatascience.com/comparison-of-the-best-nsfw-image-moderation-apis-2018-84be8da65303>