

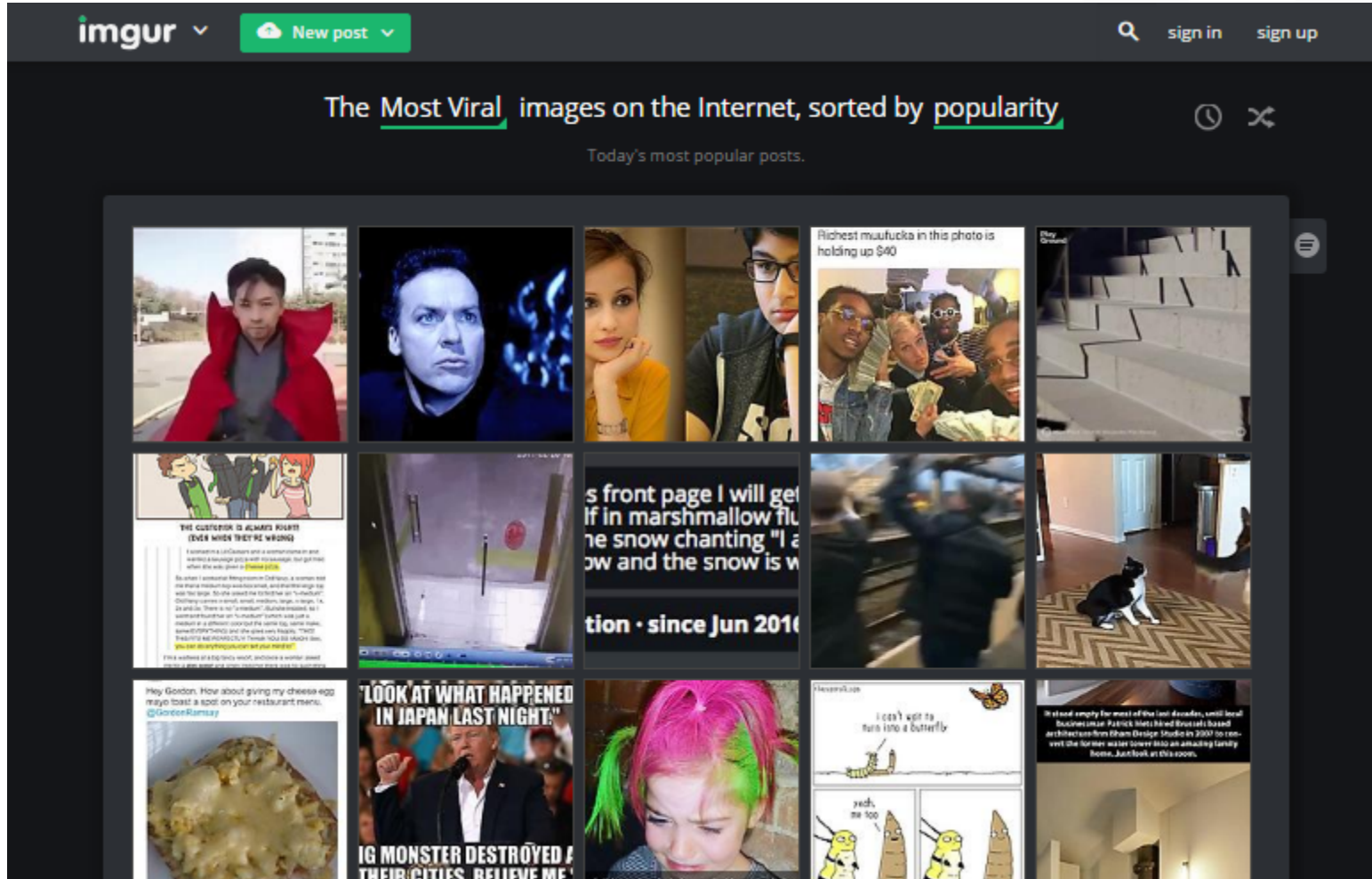


NATIONAL RESEARCH  
UNIVERSITY

# Разработка подхода для автоматической фильтрации нецензурных изображений из информационного потока

Дмитрий Желонкин

# Актуальность



# Постановка задачи

- Разработка подхода для автоматической фильтрации нецензурных изображений
- Нецензурными изображениями мог быть как фотографии так и рисунки
- Синонимы слова нецензурный: NSFW(Not Suitable For Work), nudity detection

# Существующие подходы

- Methods based on skin-region analysis
- Methods based on object detection
- ...
- Deep Convolutional Neural Networks

# Особенности задачи

- Отсутствие общедоступной выборки для оценки моделей
- Сложно дать универсальное определение нецензурного изображения
- Не большое число доступных обученных моделей для сравнения

# Негативный класс наблюдений

- Было собрано 430000 негативных наблюдений:
  1. 380000 фотографий окружающего мира
  2. 25000 изображений людей с большими открытыми участками кожи
  3. 15000 нарисованные изображения
  4. 10000 фотографий лиц

- Было собрано 2 выборки с нецензурными картинками из Danbooru сайтов (1) или служб микроблогов (2)
  1. Нецензурное изображение – изображение помеченное хотя бы одним тэгом, относящимся к группе 18+, или изображение размещено в блоге с тематикой 18+ (420000 изображений)
  2. Нецензурное изображение – изображение помеченное более чем тремя тэгами, относящимся к группе 18+, или изображение относится к блогу, в котором случайные 98 из 100 изображений нецензурные (77000 изображений)

1 - <https://danbooru.donmai.us>

2 - <https://www.tumblr.com/>

# Выборка

- 77000 положительных наблюдений:
  - 44000 реальных нецензурных изображений
  - 33000 нарисованных нецензурных изображений
  
- 77000 негативных наблюдений:
  - 21000 нарисованных изображений
  - 24000 изображений лиц и людей с большими открытыми участками кожи
  - 32000 фотографий окружающего мира
  
- Обучающая / Тестовая: 70000/7000



# Frameworks

## GoogLeNet v1 – Input 128x3x224x224

| Library                | Class                                    | Time (ms) | forward (ms) | backward (ms) |
|------------------------|--|-----------|--------------|---------------|
| Nervana-neon-fp16      | <a href="#">ConvLayer</a>                | 230       | 72           | 157           |
| Nervana-neon-fp32      | <a href="#">ConvLayer</a>                | 270       | 84           | 186           |
| TensorFlow             | <a href="#">conv2d</a>                   | 445       | 135          | 310           |
| CuDNN[R4]-fp16 (Torch) | <a href="#">cudnn.SpatialConvolution</a> | 462       | 112          | 349           |
| CuDNN[R4]-fp32 (Torch) | <a href="#">cudnn.SpatialConvolution</a> | 470       | 130          | 340           |
| Chainer                | <a href="#">Convolution2D</a>            | 687       | 189          | 497           |
| Caffe                  | <a href="#">ConvolutionLayer</a>         | 1935      | 786          | 1148          |
| CL-nn (Torch)          | <a href="#">SpatialConvolutionMM</a>     | 7016      | 3027         | 3988          |
| Caffe-CLGreenTea       | <a href="#">ConvolutionLayer</a>         | 9462      | 746          | 8716          |

<https://github.com/soumith/convnet-benchmarks>

# Сравнение моделей

|                                       | Precision<br>(neg/pos) | Recall (neg/pos) | recall_pos_photo | recall_pos_drawing |
|---------------------------------------|------------------------|------------------|------------------|--------------------|
| caffe_yahoo_resnet50_1by2 (1)         | 0.81/0.98              | 0.99/0.76        | 0.87             | 0.64               |
| caffe_miles-deep_resnet50_1by2<br>(2) | 0.56/0.98              | 1.0/0.19         | 0.26             | 0.09               |
| caffe_yahoo_ft                        | 0.97/0.95              | 0.95/0.96        | 0.97             | 0.93               |
| keras_resnet50_ft                     | 0.95/0.96              | 0.96/0.95        | 0.97             | 0.91               |
| keras_resnet50_grayscale_ft           | 0.95/0.96              | 0.96/0.95        | 0.96             | 0.92               |
| keras_xception_ft                     | <b>0.96/0.98</b>       | <b>0.98/0.96</b> | 0.97             | 0.94               |
| keras_xception_big_dataset            | 0.96/0.75              | <b>0.86/0.97</b> | <b>0.98</b>      | <b>0.96</b>        |
| keras_xception_2xft                   | 0.93/0.92              | 0.92/0.93        | 0.96             | 0.90               |

ft – finetuning, pos- positive(nsfw) class

(1) - [https://github.com/yahoo/open\\_nsfw](https://github.com/yahoo/open_nsfw)

(2) - <https://github.com/ryanjay0/miles-deep/tree/master/model>



NATIONAL RESEARCH  
UNIVERSITY

Thank you  
for your attention!