Machine Learning Application On Detecting Nudity In Images

Yong Lin, Yujun Wu

1. Introduction

We built a mobile app that help people get opinions and recommendations from their social network. The app enables people to create picture poll and have friends to vote on it. Since it contains user-generated content, it is essential to censor the content to ensure that there is no pornographic content in the picture. Monitoring all the content created in the app merely by human is labor cost. Therefore, we wanted to apply the knowledge learned from machine learning class to help detect the pornographic pictures with some learning algorithms. Our algorithm should be able to detect the suspected pornographic picture, which could be censored by app admin in the end.

We collected the preliminary training examples(labeled images) from online adult websites. The nudity detection includes skin region recognition, feature extraction and image classification. Two classification methods are applied, which are classification by skin area and classification by regionization. Comparing the results, classification by regioniaztion is significantly better.

2.Dataset

For the training data set, we collected 6026 normal pictures, marked as negative samples and 8917 pornographic pictures, marked as positive samples. The normal pictures consist of the categories of Man's Fashion, Women's Fashion, Design, Food, Travel. The size of the picture varies from 1k to 100k. 991 positive images and 670 negative images are used as the testing set.

3. Classification by skin area

3.1 Skin Detection

The most important feature that provides clues to image content is color. Color is a low level feature, which makes it computationally inexpensive and therefore suitable for real- time object characterization, detection and localization (Martinkauppi, 2002).

Nudity often consists of showing naked persons, special shots of sexual organs, or a picture of sexual intercourse (Lin et al., 2003). The presence of large skin area is a main trait of these pictures. Therefore, skin color is a basic feature used in nudity detection. A disadvantage of systems using color as a primary feature is that the systems will not work with black and white images. However, nude images are rarely in black and white.

The simplest methods in skin detection define or assume skin color to have a certain range or values in some coordinates of a color space. The RGB color space is one of the most widely used color spaces for storing and processing digital image. However, the RGB color space alone is not reliable for identifying skin-colored pixels since it represents not only color but also luminance. Skin luminance may vary within and across persons due to ambient lighting so it is not dependable for segmenting skin and non-skin regions. Chromatic colors are more reliable and these are obtained by eliminating luminance through some form of transformation. Some studies show that HSV is invariant to highlights at white light sources, to matte surfaces, and ambient lighting. Thus, the color space we used in this article is HSV.

By referring to prior work from others and trial-errors, a set of threshold values in HSV color space is set up. Hue(H) =100 ~ 130, Saturation(S) =14 ~ 140, Value(V) = 89 ~ 255. This threshold works very well. It has 90% accuracy rate with 15% false positive rate.

With this skin pixel recognition mechanism as a building block, some high level features can be extracted in this experiment.

3.2 Features

3.2.1. Normalized skin size

Normalized skin size is defined by the percentage of skin pixels relative to the image size. It is one of the main features that affect the probability of an image being nude or not. The larger the percentage of skin is, the higher the probability that an image will be classified nude.

3.2.2. Normalized n-th largest connected skin area size

Normalized n-th largest connected skin area size is defined as the ratio between the number of pixels in the n-th largest skin area to the total skin pixel count in the image. We take 1, 2 and 3 as n in this study. These features mainly characterize the concentration of skin area because nudity pictures always show grouped skin pixels rather than scattered skin pixels. All the 4 features were normalized so they are not affected by their scale factor.

3.2.3. Normalized average distance among the three largest connected skin areas

The relative distance among connected skin areas can characterize the typical pattern of a pornographic picture. One feature was the normalized average distance among the three largest connected skin areas. The steps to extract this feature was as follows: 1) find the geometric centers of the three largest skin areas in an image; 2) calculate the distances between any two of these three centers; 3) average the three distances into a value 4) divide this value by the diagonal length of the image.

3.3 Classificaiton

3.3.1 Logistic Regression

The Hessian H of the log-likelihood function for the logistic regression was calculated. H was used in the Newton Method to update theta.Consider it as converged when the difference between two theta was less than 0.0000001.

3.3.2 Gaussian Discriminative Analysis

Since the input feature x are continuous-valued random variables, we applied the GDA model to train the data. The parameters,0,1 were calculated and the training error obtained was 0.4372, which was unacceptable high.

3.3.3 Bayes Naive

In this study, Bayes Naive algorithm was applied to train our model with the 4 above-mentioned features extracted. Since Naive Bayes algorithm is only applicable to discrete features while the above-mentioned features are continuous, discretization must be applied to preprocess our features. In this study, the continuous features were evenly discretized into a scale of 0 to 10. Concretely, the normalized values were scaled up by a factor of 10 and then rounded out to the nearest integer.

3.3.4 SVM

Support Vector Machine algorithm with a gaussian kernal (sigma = 1) was applied to train our model.

4. Classification by regionization

4.1. Feature Extraction

4.1.1 Skin filtering

Since we only care about how many skin pixels an image contains and how they are distributed across the image, we process all the images by filtering the skin pixels. For each sample image, a matrix with the same size as the image was created. Each element in this matrix corresponds to a pixel in the image. After applying a skin-pixel filter mentioned in Section 3.1 on all the pixels in the image and storing the answers in this matrix, we can get a binary matrix of which each element stores 1 or 0 representing whether its corresponding pixel in the image is considered as a skin pixel or not. Figure 2 shows the visualization of this binary matrix converted from its original image (as shown in Figure 1). The white regions represent skin pixels and the black regions represent non-skin pixels.

Figure 1 . Original image

Figure 2. A binary(black & white) matrix visualization for skin pixels

4.1.2 Regionization

In order to represent the distribution of skin pixels across an image, recording whether a pixel is a skin pixel for all pixels in images is an exhaustive approach. However, images will have different size therefore different numbers of pixels. Moreover, taking information about all pixels individually as features will definitely make a very high VC dimension. This is a cause for overfitting. To reduce the number of features and VC dimension, we chose a 25-d vector to represent skin distribution.

The process of downsizing an image with any size to a 5 x 5 matrix(i.e. 25-d vector by concatenating rows or cols) is called regionization. Specifically, the way to regionize an image is simple. First thing first, partition the binary matrix obtained in Section 5.1 into 5x5 regions (as shown in Figure 3). Secondly, average the binary values within each region and put these average values into a 25-d vector in row-first order. Figure 4 gives a visualization of this averaging process. These are the two steps to do the regionization to an image. With regionization, features representing the distribution and concentration of skin pixels in an image can be extracted.

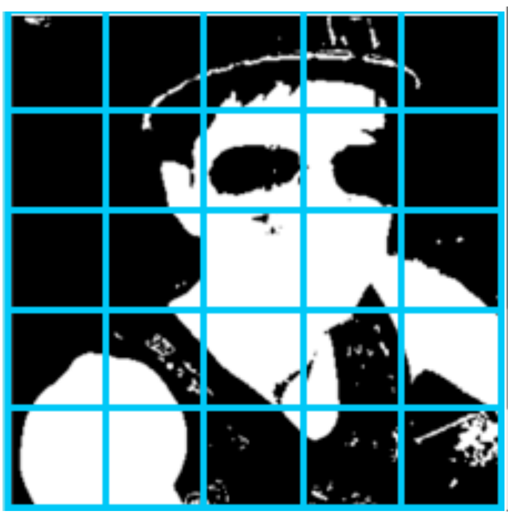

Figure 3. Partition a binary image into 5x5 blocks

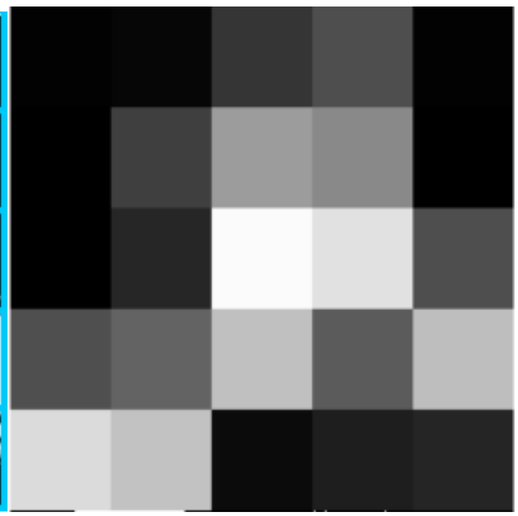

Figure 4. The visualization of the result of regionization

4.2 Classification

After downsizing all the images into a 5x5 matrix of skin proportions, all the data set were divided into 10 divisions. 9 of them were used as training set by the SVM with Guassian kernal (sigma = 1.75) and the rest 1 was used as a test group. The best hypothesis is obtained by 10-fold cross validation.

5. Result

5.1 Classification by skin area

With features characterizing skin areas and their relative distance, the best training error obtained among logistic regression, Gaussian discriminative analysis. Naïve Bayes and SVM was 0.2810, which was very high. But it led to the exploration of classification by regionization, which significantly improved the performance.

5.2 Classification by regionization

With the features extracted by 5 x 5 regionization, the optimal performance for the SVM model has the accuracy of 92.44% on training set and accuracy of 81.71% in testing. It has false positive rate of 23.82% and false negative rate of 14.96%.

6. Reference

Martinkauppi, B. 2002. Face Colour Under varying Illumination. [online]. Available: http://herkules.oulu.fi/isbn9514267885/isbn9514267885 .p df. (02 October 2003)

Lin, Y., Tseng, H. & Fuh, C. Pornography Detection Using Support Vector Machine. *16th IPPR Conference on Computer Vision, Graphics and Image Processing (CVGIP 2003)*.

Y. Xu, B. Li, X. Xue, and H. Lu. Region-based pornographic image detection. *IEEE 7th Workshop on Multimedia Signal Processing (MMSP)*, pages 1–4, November 2005.

J. Yang, Y.-G. Jiang, A. G. Hauptmann, and C.-W. Ngo. Evaluating bag-of-visual-words representations in scene classification. *In ACM Multimedia Information Retrieval (MIR)*, pages 197–206, New York, NY, USA, 2007. ACM.