Project Plan

Xinyue Hu, Zhuoan Zhou

Project idea

We want to do a program that can find a good photo in a set of photos. The set will be photos of the same person or the same location. To achieve this goal, measures of the blur, exposure, orientation, facial expression, and composition need to be involved. The most challenge one will be the facial expression because there can be one or more people in a photo and can be no one there. In the circumstance that a person existing, we also need to consider how big his or her face is. And then recognize his or her facial expression.

Data set description

For facial expression part, we will use the dataset from kaggle[1]. This dataset contains seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). 500 happy images and 500 random images from others will be used in this project. We will also download 500 images without human.

We will download 300 images from Burst library (a free image source) as the good photo set, then use photoshop to modify their exposure, sharpness, contrast, and orientations in order to get 600 “bad” photos. This will take several hours.

Test set will be 50 sets of photos. Each set will have four photos for our program to pick the best one among them. Total number of photos will be 200.

Code and libraries

Blur: This function will evaluate the sharpness of the image by simply measuring the average square gradient. Then return the value as part of the feature of the image.

Exposure: This function will evaluate the exposure of the image by making a histogram of pixel values. The histogram contains 8 columns, each column represents 8 values. This histogram will be returned as part of the image features.

Orientation: First, we do Canny edge detection, Then do Hough transform to get straight lines. After that, we get a bunch of coordinates of end points of each detected lines. For each coordinate, calculate the difference of x and y between two points. And then we get delta x list and delta y list. Sort these two list and transform delta x and delta y from pixel value difference into proportion of corresponding axes. Count the number of values between 0 to 0.05 and others. Return these two numbers of x and y. Then we get a 1x4 feature vector for orientation.

Facial expression: First, represent each image with the bag of SIFT feature. Secondly, train the svm detects three categories (happy, unhappy, and no people in the image). The dataset will be 1000 images in the facial expressions dataset from kaggle plus 500 images without human. We will use the confidence of happy category to be one of the features of the image.

Object detection: In this part, ImageAI library will supply function that can help us detect objects in images. After objects detection, we compute the centers of them. Next, we split the image into 4 equal parts horizontally and vertically, respectively. Set each intersection corner as a cluster center. Then, cluster every object into 25 cluster centers. Calculate the number of objects in each cluster center and transform the results into a 1x25 histogram vector as our feature vector.

Back Propagation Algorithm: At last, we combine every feature together into a big feature, feed it into back propagation algorithm for classification. BP is a kind of algorithm that very good at solving nonlinear problems, which is very suitable for our application.

Library: Pandas, sklearn, TensorFlow, openCV

Evaluation criteria and baselines for comparison

There are 4 photos in one set in test dataset, only one of which is a good photo. We input data into our algorithm by group. Each photo will receive a score. The one got the highest score will be classified as a good photo. Do the same thing to all 50 groups. Last, The final accuracy is obtained by dividing the number of correct groups by the total group number.

The first baseline we choose is randomness. The probability of choosing a good photo randomly is 25%. Therefore, if the accuracy of our algorithm higher than 25%, that shows our algorithm performs well.

The second baseline will be the performance of baseline back propagation on our test sets. Using bag of SIFT features to get features from the 300 “good photos” and 600 “bad” photos. Then use these features to train our back propagation.

Collaboration plan

Xinyue will be working on facial expressions, object detection, and back propagation.
Zhuoan will be working on blur, exposure, orientation, and facial expressions.

Reference:

[1] "Challenges in Representation Learning: A report on three machine learning contests." I Goodfellow, D Erhan, PL Carrier, A Courville, M Mirza, B Hamner, W Cukierski, Y Tang, DH Lee, Y Zhou, C Ramaiah, F Feng, R Li, X Wang, D Athanasakis, J Shawe-Taylor, M Milakov, J Park, R Ionescu, M Popescu, C Grozea, J Bergstra, J Xie, L Romaszko, B Xu, Z Chuang, and Y. Bengio. arXiv 2013.