Problem definition:

Given two images we want to assert whether the images were produced by the same user or not. Instead of tackling the problem as a regression model, we modeled it as a classification problem. The classification problem is posed as following:

1. The program in presented with two images.
2. The output from the program is ‘1’ if the images are from the user (similar images) and ‘0’ if the images are from different users (dissimilar images).

Data:

The data consists of 15500 images of different sizes most of them displaying the word “and” written by a user. For every user we have from 10 to 15 samples. There are 1500 users in total. We also had a few blank images which were discarded.