AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild

Ali Mollahosseini, Behzad Hasani and Mohammad H. Mahoor. *IEEE Transactions on affective computing, Vol 10. No. 1, January-March 2019*

Affect from InterNet is the larges database of the categorical and dimensional models of affect in the wild. The database is created by querying emotion related keywords from three search engines and annotated by expert human labelers. In this section, the process of querying the Internet, processing facial images and extracting facial landmarks, and annotating facial expression, valence, and arousal of affect are discussed.

Emotion-related keywords were combined with words related to gender, age, or ethnicity, to obtain nearly 362 strings in the English language such as "joyful girl", "blissful Spanish man", "furious young lady", "astonished senior". These keywords are then translated into five other languages: Spanish, Portuguese, German, Arabic and Farsi. The direct translation of queries in English to other languages did not accurately result in the intended emotions since each language and culture has differing words and expressions for different emotions. Therefore, the list of English queries was provided to native non-English speakers who were proficient in English, and they created a list of queries for each emotion in their native language and inspected the quality of the results visually. The criteria for high-quality queries were those that returned a high percentage of human faces showing the intended queried emotions rather than drawings, graphics, or non-human objects. A total of 1,250 search queries were compiled and used to crawl the search engines in out database. Since a high percentage of results returned by our query terms already contained neutral facial images, no individual query was performed to obtain additional neutral face.

Three search engines (Google, Bing, and Yahoo) were queried with these 1,250 emotion related tags. Other search engines such as Baidu and Yandex were considered. However, they either did not produce a large number of facial images with intended expressions or the did not have available APIs for automatically querying and pulling image URLs into the database. Additionally, queries were combined with negative terms (e.g., "drawing", "cartoon", "animation", "birthday", etc.) to avoid non-human objects as much as possible. Furthermore, since the images of stock photo websites are posed unnaturally and contain watermarks mostly, a list of popular stock photo websites was compiled and the results returned from the stock photo websites were filtered out.

A total of 1,800,000 URLs returned for each query were stored in the database. The OpenCV face recognition was used to obtain bounding boxes around each face. A face alignment algorithm via regression local binary features was used to extract 66 facial landmark points. The facial landmark localization technique was trained using the annotations provided from the 300 W competition. More than 1M images containing at least one face with extracted facial landmark points were kept for further processing.

The average image resolution of faces in AffectNet are 425×425 with STD of 349×349 pixels. We used Microsoft cognitive face API to extract these facial attributes on 50,000 randomly selected images from the database. According to MS face API, 49 percent of the faces are men. The average estimated age of the faces is 33.01 years with the standard deviation of 16.96 years. In particular, 10.85, 2.9, 30.19, 26.86, 14,46 and 13.75 percent of the faces are in age ranges $[0, 10), [20, 40), [30, 40), [40, 50)$ and $[50, -)$, respectively. MS face API detected forehead, mouth, and eye occlusions in 4.5, 1.08 and 0.49 percent of the images, respectively. Also, 9.63 percent of the faces wear glasses, 51.07 and 41.4 percent of the faces have eye and lip make-ups, respectively. In terms of head pose, the average estimated pitch, yaw, roll are 0.0, -0.7, and -1.19 degrees, respectively.

1. (1 point) Which two search engines were mentioned, but not queried?

    1. _____

    2. _____

2. (1 point) What is the average image resolution of faces in AffectNet?

    ○ 255×255

    ○ 349×349

    ○ 425×425

    ○ 400×400

3. (5 points) Mark sentences as True or False

Sentence	T	F
One of the combined keywords mentioned in second paragraph was "furious girl"		
There was a total of over 1,200 search queries compiled		
Less than 70 facial landmark points were extracted using DFS algorithm		
Bounding boxes for faces were obtained using OpenCL		
Less than 10 percent of faces wear glasses		

4. (3 points) List three non-English languages used for querying the search engines

    1. _____

    2. _____

    3. _____

Question:	1	2	3	4	Total
Points:	1	1	5	3	10
Score:					