International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 9 Issue 12 ǁ 2021 ǁ PP. 79-84

Face Analysis in Image Collection for Graphical representation

C.DillyRani
JJ.college of engineering and technology (sowdambikaa group of institution)
Ammapettai, tiruchirappallai
Assistant ProfessorDepartment of Mathematics

Abstract:*In face analysis the concept of image collection in Social Medias with huge number of pictures [1]. It is used to analyze how the individuals that are present in the collections with interact each others. In this paper we use the word "connected" with class room students for each classes score that represent. It is estimated based on co-occurrence, closeness, facialexpressions. The nodes represent the students of the collection and the edges correspond represent the connectivities. Finally we represent the image of the students in which collection of the nodes present. We prove relevant result by Indian wedding celebration (IWC) cricket game, funds and social medias in an image collection.*
Keywords:-*IWC, Cricket game, friends, social medias.*
Index:-*Face recognition, expression, graph analysis and images, gender.*

Date of Submission: 15-12-2021 Date of acceptance: 30-12-2021

I. INTRODUCTION :

It has been a common practice to have collections of a huge number of images [1]. For eg. A social event – like a IWC – can have thousand of pictures. In practices, it is not so simple browse all images, and it can be very helpful to analyze the faces and build a graph with extracted from all faces that are present in the images in fig [1]. Where the nodes represent the student and the edges are connection between them. In addition there are very impressive advance application [2] and in the recognition of age [3] gender [4] and facial expression [5]. In [6] social networks are built by detecting and tracing faces in new sides, the idea in to establish with whom appearance on TV news. In [7] a social relation (defined as the association like worms, between two or more persons) are detected in face images in the wild like face clustering [8][9], eye[10], head [11], and facial land marks [12].

Our contribution is collection of a set of an image is present in a collection (i) connectivity measure every pair of students of the set that give a score that represent how connected they are (ii) estimation of a optimal graph in which the nodes represent the subject of the set, and the length of the edges that connect the nodes depends on the connectivity subjects by fig(1)

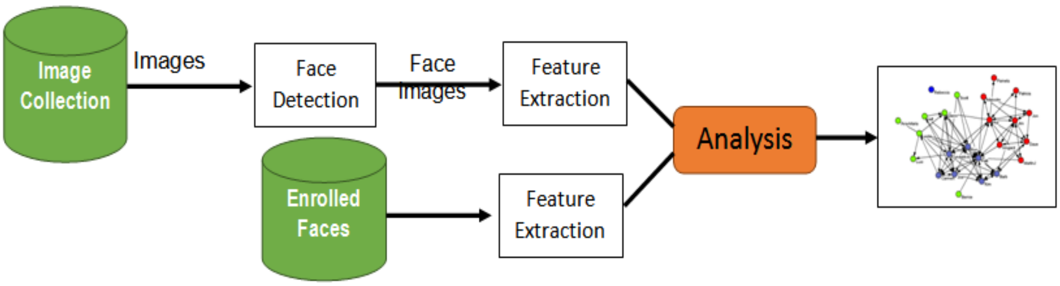


Fig.1. Block diagram of our proposed method.

1.Methodology

The method follows fig(1), it is necessary to do some feature extraction. After that, we define some matrices that can be used to measure the "connectivity" between the subjects that are present in the collection.

1.1 Feature Extraction

The idea of our approach is to analyze a set I of n images $\{I_t\}$, for t = 1 … n. we detect all faces of I using Multi-task Cascaded Convolutional Networks (MTCNN) [13] that has been demonstrated to be very robust in unconstrained environments. All detected faces are stored as set F of m face images $\{F_k\}$, for k = 1 … m. In addition, we store in vector z of m elements the image index of the detected face image, i.e., z_k = t, if face image F_k was detected in image I_t. Furthermore, the m bounding boxes of the detected faces are stored in matrix

B of m x 4 elements with coordinates $b_k = (x_1,y_1,x_2,y_2)_k$ for face image k. After face detection is performed, for each face image k, we compute the width and height (w_k,h_k) of the bounding box. Afterwards, we compute: i) the age (a_k), the gender (g_k) using the library py-agender [14] that offers very good results. ii) The facial expressions (e_k) are defined as a vector of seven probabilities [15] for: angry, disgust, scared, happy, sad, surprised, and neutral. iii) The 68 facial landmarks (l_k) give the coordinates (x,y) of the eyebrows (left and right), eyes (left and right), nose, mouth and jawline. For this end, we use the library Dlib [15], iv) the pose vector (v_k) defines the direction where the face is looking to, for this end we use the projection of the roll vector obtained in [11], and v) the face descriptor of elements (x_k), in this case we use descriptor with uniform. In our experiments, we use ArcFace [16], that computes an embedding of d = 512 elements with outstanding results. The descriptors of m detected faces are stored in matrix X of m x 512 elements. In addition, we have a list of n_e enrolled subjects that we want to analyze. It can be defined manually or using a face clustering algorithm. The descriptors of the enrolled faces have uni-norms and they are stored in a matrix X_e of n_e x 512 elements. It is very simple to detect if the enrolled subjects are present in the image collection: we compute the mxn_e –element matrix $Y = XX_e^T \geq \theta$, where θ is a threshold that we set to 0.4. If Y (k,i) = 1, that means that subject S_i, was detected in image z_k.

1.2 Analysis

The following matrices have n_e x n_e elements. The indices of the elements of the matrices are (i,j). Element (i,j) gives a 'connectivity measure' between subjects S_i and S_j. The matrices are symmetric, that means for matrix X, $X_{ij} = X_{ji}$. In addition, $X_{ii} = 0$. In all connectivity matrices of our approach, a high / low value of X_{ij}means that the connectivity between S_i and S_j is high / low.

1. Co-occurrence Matrix (C): Element C_{ij} is defined as the number of images in the collection in which subjects S_i and S_j are present. This matrix is easily computed by defining a 'presence matrix' P of n_e x n elements in which element P_{it} is 1/0 if subject S_i is present / absent in image I_t. S_i, S_j is the no.of image collection in the students.
2. Closeness Matrix (D): Element D_{ij} is defined as the sum of the 'closeness factors' of subject S_i and S_j in those images of the collection in which S_i and S_j are present. The factor is computed as follows: For subjects S_i and S_j, we measure the size in pixels of the bounding box of the faces as $a_i = \sqrt{w_i h_i}$ and $a_j = \sqrt{w_j h_j}$, the average $(a_i + a_j)/2$, and the distance d in pixels between the centers of the bounding boxes in the image. The 'closeness factor' is computed for those images that have similar face sizes (min $(a_i/a_j,a_j/a_i)<0.7$) and for the faces that are close enough $(d/a<n_f)$. In our experiments, we set $n_f$to 4. With the first criterium, we can avoid faces that are close in the image but because of the perspective they are far away in 3D space. With the second criterium, we consider only those pairs of images that are close enough. In our case, the faces must be closer than 4a (4 'faces'). Thus, the 'closeness factor' is defined as $(n_f – d/a)/n_f$. In this case, the factors is close to one, if the faces are very close; it is for example 0.25, if the distance of the faces is 3a, and it is zero if the distance is greater than 4a.
3. Connection Matrix (Z): Element $Z_{ij}$is defined as the sum of the 'connection factors' of subjects S_i and S_j in those images of the collection in which S_i and S_j are present. The factor is computed as follows: For subject S_i and S_j, we compute the intersection of vectors v_i and v_j defined as the projected roll vectors (computed from the head pose) that start at the point that is in the middle of both eyes (compute as the center of the landmarks corresponding to the eyes of the face). If the intersection of the vectors are in front of the faces, we compute both distance from intersection point to the middle of both eyes, and select the minimum (d). If d is very small, it means that one subject is seeing close to the other one. The 'connection factor' is defined (like the 'closeness factor') as $(n_f – d/a)/n_f$ for $d/a <n_f$.
4. Empathy Matrix (E): Element E_{ij} is defined as the sum of the 'empathy factors' of subjects S_i and S_j in those images of the collection in which S_i and S_j are present. The factor is computed as follows: For subject S_i and S_j, we measure the first six elements of the expression vector (we avoid the neutral expression) as vecotrse'_i and e'_j. The idea of the 'empathy factor' is to be close to one (or zero), if the expressions of subjects S_i and S_j are similar (or different). For this end, we use the cosine similarity,that means, we normalize the vectors as $e'_i/||e'_i||$ and $e'_i/||e'_j||$ and define the 'empathy factor' as the dot product of them.
5. Happiness Matrix (H): Element H_{ij} is defined as the sum of the 'happiness factors' of subjects S_i and S_j in those images of the collection in which S_i and S_j are present. The factor is computed as follows: For subjects S_i and S_j, we extract the fourth element of the expression vector (that corresponds to the happiness) $h_i = e_i (4)$ and $h_j = e_j(4)$, and define the 'happiness factor' as the average $(h_i + h_j)/2$.

Finally, the total connectivity matrix T is defined as the average: T = (C + D + Z + E + H)/5. We tested different weighted sums with similar results.

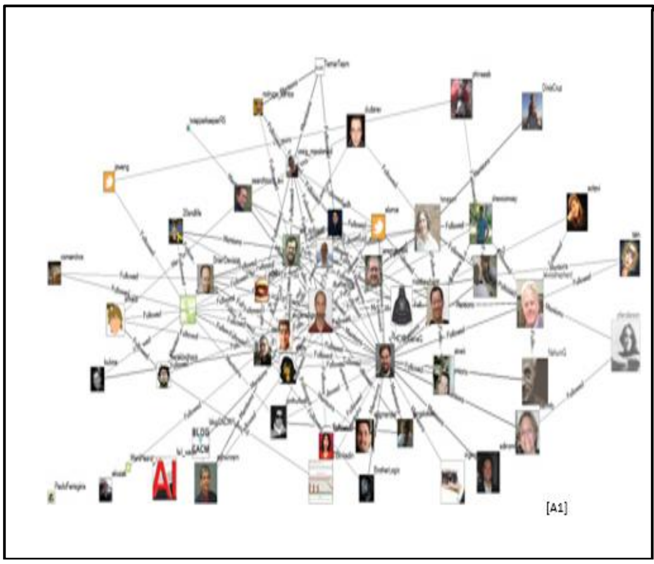
1.3 Graph Construction

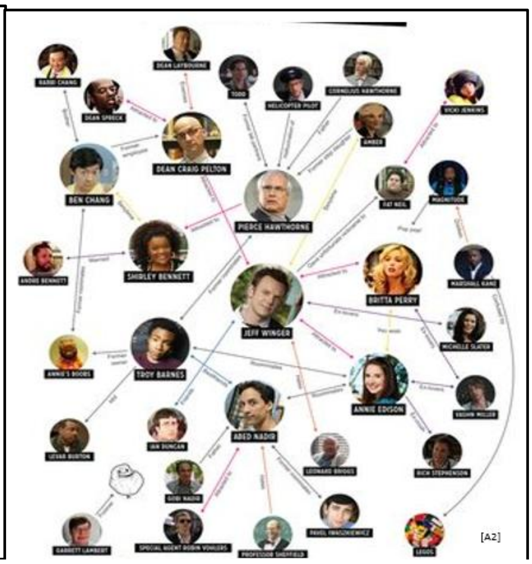
In order to build a graph that represent the connectivity of the subjects to be analyzed, we define a graph of n_e nodes (one for each subject) and locate them in a 2D space in position (x_i,y_i) for i = 1 … n_e. The key idea of the graph, is that the closer are the nodes i and j, the higher the connectivity between them. That means, the distance of both nodes $\Delta_{ij} = \|(x_i,y_i) - (x_j,y_j)\|$ should be a value that represents the 'no-connectivity' between subjects S_i and S_j. In our approach, we use the 'connectivity' defined in previous section (see matrix T). We tested several definitions for 'no-connectivity' and the best one –in terms of the visualization of the graph – was given by $W_{ij} = 1/\sqrt{Q_{ij}} + 1$, where $Q_{ij} = 99 \times T_{ij}/\max(T)$, that means the 'no-connectivity' W_{ij} for subjects S_i and S_j gives values between 0.1 (when the connectivity is maximal) and 1.0 (when there is no connectivity). In addition, we set W(i,j) = 0. The idea of our method is to find the coordinates of the nodes (x_i,y_j), so that the distance of the nodes Δ_{ij}, should be similar to the 'no-connectivity' W_{ij}. This is an optimization problem that can be represented by minimizing the Frobenius norm: error = $\| \Delta - W \|^2_F \rightarrow$ min. we solve this problem using the simplex search method [16] with a starting point given by the solution of a graph drawing by force-directed placement[17], in which the length of each edge in the graph is proportional to its weight W_{ij}. In our experiments, we report the Mean Absolute Error (MAE), as a metric of the performance of the graph, defined as the mean difference $|\Delta_{ij} - W_{ij}|$ in all pairs of nodes. Obviously, MAE cannot be zero in many cases where (eg., A is friend of B and C, but B and C are enemies).

1.4 Graph Representation

For the node representation: a node is typically represented as a point or a circle. Alternatives in this case can be the size of the circle, color of the circle and color of the boundary of the circle. In our case for node i, we put the face of subject S_i in a circle (the radius is related to the number of images in which S_i is present), and the color of the boundary is related to the predominant facial expression of S_i. We use the colors based on [18]: scare (dark gray), angry (red), disgusted (green), sad (blue), surprised (white), happy (yellow), neutral (gray). For the edge representation: an edge is typically represented as a line, or (bi)directional arrow. Alternatives in this case can be the color and the width of the line. In our case for edge (i,j), we decided to use lines (and not arrows) to make the graph simpler. In addition, our connectivity matrix is symmetric and it has no sense to use arrows[1]. The color of the lines, in our case, is set to the most common expression between subjects S_i and S_j, and the width of the line corresponds to the number of images in which both subjects are present. If S_i and S_j do not have pictures together ($C_{ij} = 0$), there will be no edge between them.

(1. The definition of connection given by matrix Z can be changed, if we consider the case when subject S_i is looking to subject S_j not vice-versa).





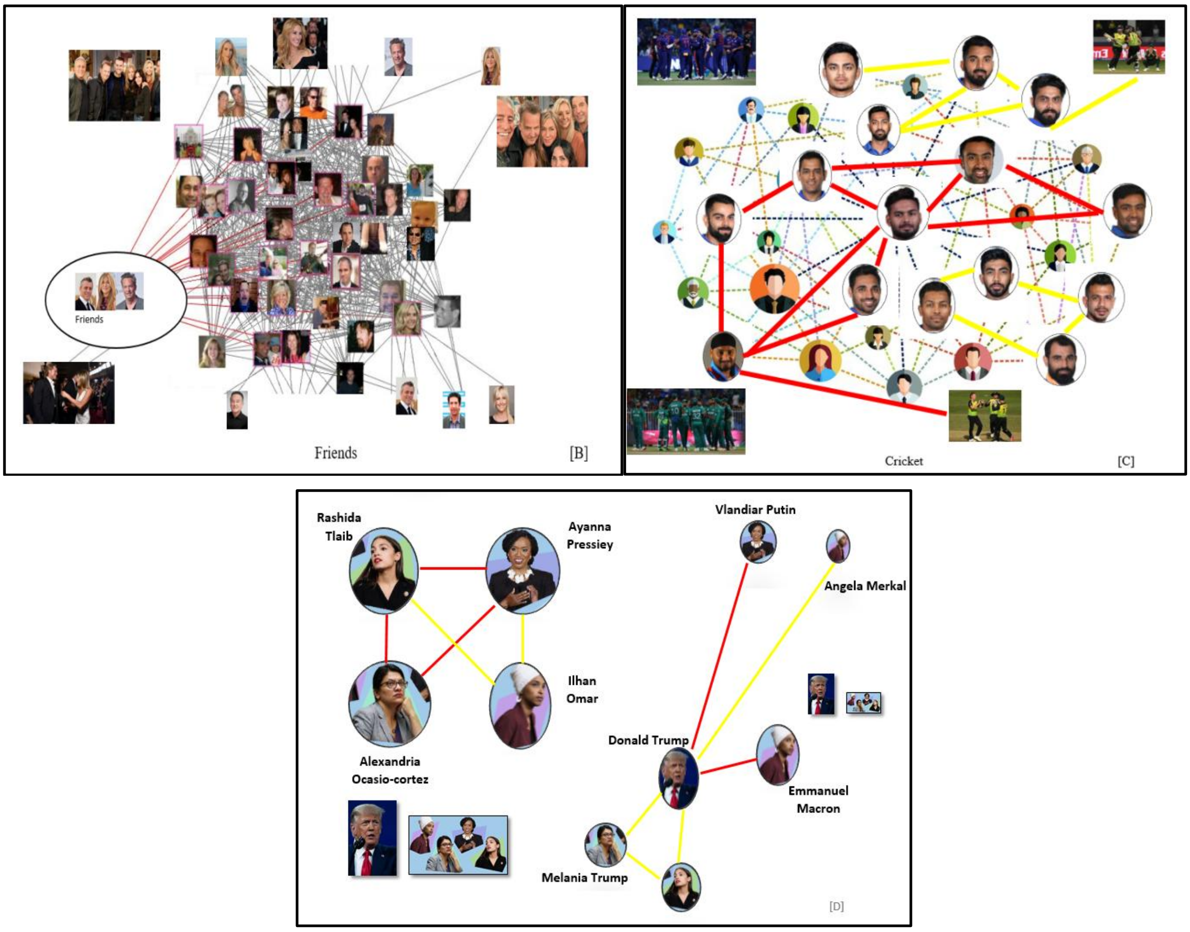


Fig.2. Graph Representations. A)IWC – A1: Proposed graph representation. A2: Nodes of Proposed graph with edges given manually (blue lines: couples, green arrows: parents to sons and daughters, orange lines: Siblings, magenta lines: friends. B) Friends (ellipse is plotted manually), C) Cricket: two teams (red and white stars drawn manually). D) Trump – The Squad. In Addition, some examples of the GUI.

Since the information of the co-occurrence matrix and presence matrix is stored, we propose to display the graph in a Graphical User Interface (GUI) in which the user can interact with the nodes and edged. By clicking onto a node, the user can obtain information related to the subject of the node: e.g., gender, age and facial expressions. In addition, the images of the collection in which the subject is present can be displayed according to the facial expression (the images can be stored using the expression 'happiness' for example). On the other hand, by clicking onto an edge (that connect two nodes, in our case two subjects), the user can display the corresponding images of the connected subjects (images of the collection in which both subjects are present).

II. RESULTS AND DISCUSSION

In this section, we report the results obtained in four experiments that we used to validate the proposed approach. The implementation was done in Python 3.6.4 (for face detection and feature extraction).

2.1 Indian Wedding Celebration images

In 15th Jun'2021 we took a private family album of a wedding celebration in which we participated. That means, we know exactly the subjects that are present and the relations that they have. In order to protect the privacy of the participants, we public only the attributes and descriptors of the faces, and for the graph representation we use synthetic faces generated by a GAN model [20]. This dataset has 639 images and 2280 faces. The obtained graph is illustrated in Fig.2-A1. In Fig.2-A2, we provide an additional graph with the same nodes, in which the edges are drawn manually according to the existing relation between the participants (couples, parents, sons, daughters, siblings and friends).

Discussion: (MAE = 0.1944) In the graph, we observe that main participants are the bride and the groom. In the GUI, we show the histogram of the expression of the bride (she was happy in 78% of the pictures). The man facial expression of the wedding is 'happiness' (yellow), however, there are some people that is angry (the parents of the bride). It is very impressive, how close are the nodes in the graph for many of the strongest relationships of the participants.

2.2 Friends

We downloaded videos of the sitcom friends with six main character (Matt Liblanc, Matthew Perry, Jennifer Aniston, David Schwimmer Lisa kudrow with guest stars like (Julia Roberts, Brad pitt, Robin Williams, Erin Brockovich (2000), Emma Roberts.,)[2]. This dataset has 16420 images and 27038 faces. For this we select the six main characters and 18 guest stars is illustrated in fig 2-B.

->Discussion: (MAE = 0.2312) from the graph, we observe that the main participants coincide with the six main characters of friends. The main facial expression in the graph for this (‘happiness’ (yellow). There is for example a connection with Brad Bit (he was present in 363 images, and from them, Lisa kudrow was present in 67 as well).

2.3 Cricket

On Nov11,2010 we downloaded the ESPNCricinfo the album “VBARC 2010” of pictures taken by Bruee Taylor. In there picture we observe the different cricket games played on June 20, 1973. This dataset has 1131 images and 4550 faces. The graph is illustrated in Fig 2-C. In this set, there are two teams of players (see red and white stars included manually in the graph). Then we select the players of both teams.

->Discussion: (MAE = 0.0679). 16% and 6% of the detected faces belong to the red and white teamsresp, we observe in the graph, that both teams are red team. The facial expression are ‘scared’ (gray) and ‘happiness’ (yellow).

2.4 Donald Trump – The Squad

On Aug 16,2019 (4.32pm) we downloaded images from twitter given the hashtags#. Trump and # Donald Trump. In those days, there was a problems an informal clique of four freshman congress women who have become an avatar of the progressive resistance in Washington (the squad). The dataset has 494 images 124 were automatically removed because they were duplicated and 677 faces the graph is illustrated on Fig 2-D.

->Discussion: (MAE = 0.1251). in graph, we obtained that main participant is Trump, has a ‘happy’ relation, to his wife and daughter, and a ‘sad’ and ‘angry’ relation to putin and markel respectively.

III. CONCLUSION

In this paper, we presented a graphical tool that can be used to detect and analyze social relations in an image collection. We proposed an optimal graph representation that is based on the ‘connectivity’ of the subjects. We based our measurement on co-occurrence, closeness, facial expressions, theorientation of the head. The nodes represent the subjects of the collection, and the edgesare their connectivities. In our solution, the closer the nodes, the more connected are the subjects. We propose a representation for the nodes and edges (colors are related to facial expressions and size are related to presence). The graph can be used in a graphical user interface (GUI) in which we can display the original images that shows the connection of the people. Finally, we present relevant results by analyzing a wedding celebration, a sitcom video, a cricket game and images extracted from twitter given a hashtag. We believe that this tool can be very helpful to detect the existing social relations in an image collection. For future work, we would like to expand to three or more people by analyzing the connectivity matrix.

REFERENCES

[1]. D.Mery, ‘Face analysis: State of the art and ethical chal-lenges,” in Proceeding of the Pacific Rim Symposium on Image and Video Technology (PSIVT 2019): Workshop Vision-Tech, 2019.

[2]. D.Mery, I.Mackenney, and E. Villalobos, “Student attendance system in crowded classrooms using a smartphone camera,” in 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2019, pp. 857-866.

[3]. R.Rothe, R.Timofte, and L. Van Gool, “Deep expectation of real and apparent age from a single image without facial landmarks,” International Journal of Computer Vision, vol.126, no.2-4, pp. 144-157, 2018.

[4]. F.H.d.B.Zavan, O.R.Bellon, L.Silva, and G.G.Medioni, “Benchmarking parts based face processing in the wild for gender recognition and head pose estimation,” Pattern Recognition Letters, vol.123, pp.104-110, 2019.

[5]. O.Arriaga, M.Valdenegro-Toro, and P.Ploge, “Real-time convolutional neural networks for emotion and gender classification,” arXiv preprint arXiv:1710.07557, 2017.

[6]. J.Deng, J.Guo, N.Xue, and S.Zafeiriou, “Arcface: Additive angular margin loss for deep face recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp.4690-4699.

[7]. T.Baltrusaitis, A.Zadeh, Y.C.Lim, and L.P.Morency, “Open-face 2.0: Facial behavior analysis toolkit,” in 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). IEEE, 2018, pp.59-66.

[8]. C.Otto, D.Wang, and A.K.Jain, “Clustering millions of faces by identity,” IEEE transactions on pattern analysis and machine intelligence, vol.40, no.2, pp.289-303, 2017.

[9]. Y.Shi, C.Otto, and A.K.Jain, “Face clustering: representation and pairwise constraints,” IEEE Transactions on Information Forensics and Security, vol.13, no.7, pp. 1626-1640, 2018.

[10]. O.Lorenz and U.Thomas, “Real time eye gaze tracking system using cnn-based facial features for human attention measurement,” in proceedings o the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications – Volume 5 : VISAPP, INSTICC. SciTePress, 2019, pp. 598-606.

[11]. N.Ruiz, E.Chong, and J.M.Rehg, "Fine-grained head pose estimation without keypoints," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2018.
[12]. M.Zhu, D.Shi, M.zheng, and M.Sadiz, "Robust facial landmark detection via occlusion adaptive deep networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 3489-3496.
[13]. K.Zhang, Z.Zhang, Z.Li, and Y.Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," IEEE Signal Proceeding Letters, vol.23, no.10, pp. 1499-1503, 2016.
[14]. Y.Uchida, "Py-Agender: Keras implementation of a cnn network for age and gender estimation (http://pypi.org/project/py-agender/)," 2019.
[15]. D.E.King, "Dlib-ml: A machine learning toolkit," Journal of Machine Learning Research, vol.10, no.Jul, pp. 1755-2009.
[16]. J.C.Lagarias, J.A.Reeds, M.H.Wright, and P.E.Wright, "Converage properties of the nelder-mead simplex method in low dimension," SIAM Journal on optimization, vol.9, no.1, pp. 112-147, 1998.
[17]. T.M.Fruchterman and E.M.Reingold, "Graph drawing by force-directed placement," Software: Practice and experience, vol.21, no.11, pp. 1129-1164, 1991.
[18]. T.M.Sutton and J.Altarriba, "Color associations to emotion and emotion laden words: A collection of norms for stimulus construction and selection," Behaviour research methods, vol.48, no.2, pp. 686-728, 2016.
[19]. Z.Zhang, P.Luo, C.C.Loy, and X.Tang, "Learning social relation traits from face images," in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 3631-3639.
[20]. T.Karras, S.Laine, and T.Aila, "A style based generator architecture for generative adversarial network," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp.4401-4410.
[21]. O.M.Parkhi, A.Vedaldi, and A.Zisserman, "Deep face recognition," in British Machine Vision Conference (BMVC2015), vol.1, 2015, p.6.
[22]. P.Li, L.Prieto, D.Mery, and P.J. Flynn, "On low-resolution face recognition in the wild: Comparisons and new techniques," IEEE Transactions on Iformation Forensics and Security, vol.14, no.8, pp.2000-2012, 2019.
[23]. X.Sun, P.Wu, and S.C.Hoi, "Face detection using deep learning: An improved faster rcnn approach," Neurocomputing, vol.299, pp.42-50, 2018.
[24]. B.Renoust, T.Kobayashi, T.D.Ngo, D.D. Le, and S. Satoh, "When face-tracking meets social networks: a story of politics in news video," Applied network science, vol. 1, no.1, pp.4, 2016.
[25]. X.Huang, A.Dhall, R.Goecke, M.Pietikainen, and G.Zhao, "Multimodel framework for analyzing the affect of a group of people," IEEE Transactions on Multimedia, vol.20, no.10, pp. 2706-2721, 2018.
[26]. M.Hoai and A.Zisserman, "Talking heads: Detecting humans and recognition their interactions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 875-882.