International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Real-Time Age, Gender and Emotion Detection using Caffe Models

Prathamesh Kirpal, Nihal Kuthe, Prof. M.M. Gudadhe, Snehal Gajbhiye, Achal Tumsare, Anoop Chahande

Dept. of Information Technology, Priyadarshini College of Engineering

Nagpur, India

ABSTRACT

Age and gender classification has become applicable to an extending measure of applications, particularly resulting in the ascent of social platforms and social media. Regardless, execution of existing strategies on real-world images is still fundamentally missing, especially when considering the immense bounce in execution starting late reported for the related task of face acknowledgment. In this paper we exhibit that by learning representations through the use of significant Convolutional Neural Network (CNN) and Extreme Learning Machine (ELM). CNN is used to extract the features from the input images while ELM defines the intermediate results. We experiment our architecture on the recent Audience benchmark for age and gender estimation and demonstrate it to radically outflank current state-of-the-art methods. Experimental results show that our architecture outperforms other studies by exhibiting significant performance improvement in terms of accuracy and efficiency.

FACE DETECTION

Haar cascade XML file which is a classifier used to identify a specific object from the webcam. The haarcascasde_frontalface_default.xml provided by OpenCV used to recognize frontal face. OpenCV connects to the webcam which user can use to scan their faces for classification of age, gender and emotion

Face detection has a long research history. Yang et al compared some prominent face detection algorithms in year 2002, But they did not use any prominent algorithms such as Haar Classifiers in their studies. Haar Classifier is one of the most prominent and accurate object detection approach described by P. Viola and M. Jones. A thorough survey can be found in There are several natural

(lighting, pose angle, face marks) as well as digital (noise, glitches) variation imposed while detecting a face in a frame.DNN model was selected according to its superiority over other algorithms like (Haar Cascade, HoG, and CNN) in accuracy, speed, running at real-time on CPU, and detect faces in various scales and orientations.

Face Detection is a technology that used in different applications that detect human faces in digital images. Face detection also used for the psychological process by which humans locate and attend to faces in a visual scene We used OpenCV to catch the live image. Here, for detection the human faces, we used the HaarCascade image processing method. We saw that there was a situation where it didn't detect the human faces in the live images for the lack of contrast. So, we used histogram equalization to improve detection by increasing contrast. Haar-cascade: Face detection using Haar-cascade is based upon the training of a Binary classifier system using the number of positive images that represent the object to be recognized (such as faces of different peoples at the different scene) and even large number of negative images that indicate objects or feature not to be detected (images that are not human faces but can be anything else like a table, chair, wall, etc.) Actual Image Extracted human face.

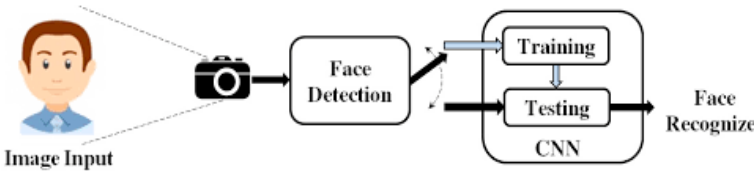


Fig (1).Face detection using cnn

2.Gender prediction

Images may not be perfect. There are many noises which are redundant. This can decrease system performance. To increase accuracy rate we have to make proper and effective feature extraction. This can be global or local which depends on shape, color, orientation.

1) Edge detection:Edge feature is mostly used for detecting the object. It finds the discontinuities in gray level. We can say that edge is the boundary between the regions.

2) Haar– like features: Viola and Jones proposed an algorithm which is called Haar-Classifiers for rapid object detection and pedestrian detection is applied. it is done with the haar like features which can be calculated efficiently by using Adaboost classifier and integral images in cascade classifier. Haar-like features can have high accuracy and in low

cost. Haar cascade is mostly used for face detection because of its easy calculation.

3) Detector using haar -Like features: In face detection, the image is first scanned, looking for patterns with indicate the presence of a face in the image. This is done by using haarlike features. The haar like features are created by two or three adjacent rectangles with different contrast values

A detailed survey of gender classification methods can be found in and more recently in.

The gender prediction is assumed as a classification problem and the output layer of this network is a SoftMax with 2 nodes, which represents Male and Female classes.this model into memory and passing the output of the face detection process (detected face) through the gender prediction network, then we will have the predicted values for both classes as an output from the network. now we can take the maximum value of the output and use it as a predicted gender.

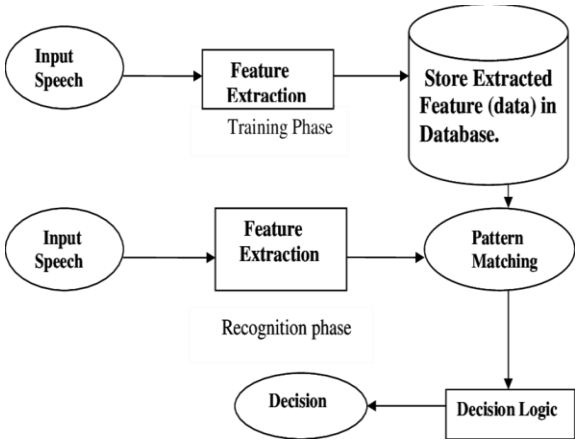


Fig (2).Gender detection flowchart

3.Age prediction

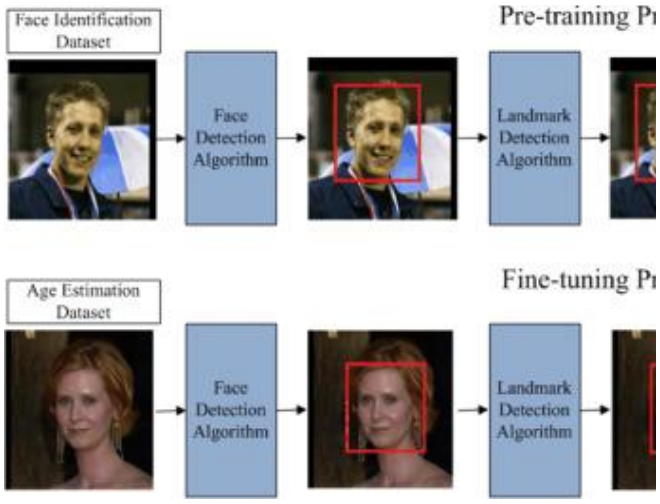
The ages grouped as a following (0–2), (4–6), (8–12), (15–20), (25–32), (38–43), (48–53). Like gender prediction Using the model in as a network that uses 9 layers, eight of them are fully connected layers and the last one is an output layer. By loading this model into memory and passing the output of the face detection process (detected face) through the age prediction network, then we will have the predicted values for all classes as an output from the network. now we can take the maximum value of the output and use it as a predicted age group.

This is practically like the sexual orientation identification part aside from that the relating

prototxt record and the caffe model document are "deploy_agenet.prototxt" and "age_net.caffemodel". Moreover, the CNN's yield layer (likelihood layer) in this CNN comprises of 8 qualities for 8 age classes ("0–2", "4–6", "8–13", "15–20", "25–32", "38– 43", "48–53" and "60-") Preferably, Age Prediction ought to be drawn closer as a Regression issue since we are anticipating a genuine number as the yield. Notwithstanding, assessing age precisely utilizing relapse is testing. Indeed, even people can't precisely foresee the age dependent on taking a gander at an individual. Be that as it may, we have a thought of whether they are in their 20s or in their 30s. In view of this explanation, it is savvy to outline this issue as an order issue where we attempt to gauge the age bunch the individual is in. For instance, age in the scope of 0-2 is a solitary class, 4-6 is another class, etc. The Adience dataset has 8 classes separated into the accompanying age bunches '(0-2)', '(4-6)', '(8-12)', '(15-20)', '(25-32)', '(38-43)', '(48-53)', '(60-100)'. In this way, the age expectation network has 8 hubs in the last softmax layer demonstrating the referenced age ranges.It ought to be remembered that Age forecast from a solitary picture is definitely not an exceptionally simple issue to explain as the apparent age relies upon a ton of elements and individuals of a similar age may appear to be truly unique in different pieces of the world.

Early methods for age estimation are based on calculating ratios between different measurements of facial features. Once facial features (e.g. eyes, nose, mouth, chin, etc.) are localized and their sizes and distances measured, ratios between them are calculated and used for classifying the face into different age categories according to hand-crafted rules. More recently, [41] uses a similar approach to model age progression in subjects under 18 years old. As those methods require accurate localization of facial features, a challenging problem by itself, they are unsuitable for in-the-wild images which one may expect to find on social platforms.

Caffe is a CNN framework which allows researchers and other practitioners to build a complex neural network and train it without need to write much code. For estimation of age using the convolution neural network, gathering a large dataset for training the algorithm is a tedious and time consuming job. The dataset needs to be well labeled and from a social image database which has the private information of the subjects i.e. age.



Fig(3).Age detection flowchart

4.EMOTION DETECTION

The case of speech, facial expressions reflect the emotions that a person can be feeling. Eyebrows, lips, nose, mouth, muscles of the face: they all reveal the emotions we are feeling. Even when a person tries to fake some emotion, still their own face is telling the truth. The technologies used in this field of emotion detection work in an analogous way to the ones used with speech: detecting a face, identifying the crucial points in the face which reveal the emotion expressed and processing their positions to decide what emotion is being detected.

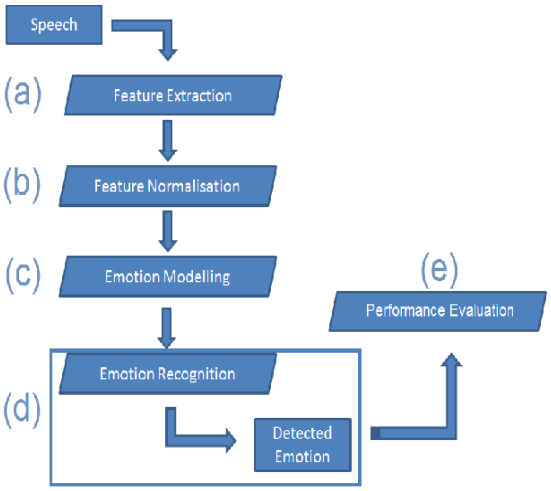
FER2013 is a Kaggle dataset that contains labeled 3589 test images, 28709 train images. We don't have to do data augmentation because the dataset has been built with wide range of images. The database holds grayscale pictures of human faces. We don't use transfer learning because our dataset contains grayscale images and doesn't fit in 3 channels pretrained Models. We use 3 convolutional layers. Input [48x48x1] carries the pixel values of given image. Hence images have width equal to 48, height equal to 48, and with one color channel.

· Step 1: Normalizing the data between 0 and 1.

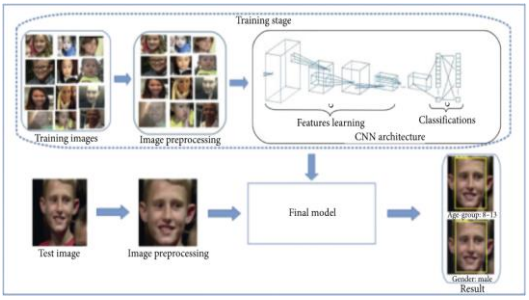
· Step 2: We use 3 convolutional layers. In each layer, we do Batch Normalization, RELU activation function and useMaxPooling. In a fully connected layer we use the RELU activation function and SOFTMAX function.

· Step 3: Calculate the loss function using Adam optimizer ·

Step 4: to use the trained model later, save the weights in fer.h5.



Fig(4).Emotion detection flowchart



fig(5). Final output of age,gender and emotion

5.MERGING OF ALL THE MODULES

As per the above modules, age has been detected separately as one module, gender as another module and emotion has also another module. This project is made to merge all the three modules using algorithms and work as a one module to give (age,gender and emotion) as one output. It gives the transparency and is used to detect all the three (age,gender and emotion) at once.

To work this three modules as one, Many algorithms with the knowledge of Artificial Intelligence, Machine Learning and many other basic technologies are used.

6.CONCLUSION

The present study on facial analysis, particularly emotion recognition, age estimation and gender recognition, proposed developing a Multi-task learning framework capable of solving the three tasks simultaneously in a fast and efficient way. Various models were developed using a chosen CNN as the baseline architecture, derived from the known Exception, altering and adapting the network to various Multi-task structures. The difference between the various developed models relies on the number of shared residual modules. Having two datasets, one with emotion labels and one with gender and age labels, the networks were trained using two multi-task learning methods: Task-based Regularization and Domain-based

Regularization. The Task-based Regularization approach uses an aggregated loss function as the objective function to train the CNN. This approach was used to train the gender and age tasks. The Domain-based Regularization approach implies that sharing the parameters between tasks trained with task-related datasets makes the shared parameters adapt to the complete set of tasks, instead of fitting to a task-specific domain, inducing a better generalization capacity to the network. The method to train in such a way, batch by batch training, was developed specifically for this dissertation. The models' results are successful, supporting the developed multi-task learning method success. Furthermore, the results state comparable results between the multi-task learning models and the single-task ones, providing evidence that a model with shared layers – more efficient, faster at inference times and lighter – is a better choice if one seeks solving different tasks simultaneously. The present research is nonetheless affected by faults and limitations. One of the limitations concerns the used gender and age dataset, which is biased towards dark-skinned individuals, males, and the age interval of 25 to 55 years old, approximately. Also, the Disgust class of the emotion dataset has quite few samples. Future work would benefit from testing the framework with different baseline architectures and different datasets. Additionally, the development of an end-to-end system would be appropriate, as in this way it would be possible to deploy the model in a live (or just video footage) setting. The model is very capable of such, as it is small, fast and efficient. There are various face detection and alignment CNNs available, making the development of an end-to-end system quite simple: by implementing an available system, i.e. the MTCNN (K. Zhang et al., 2016), and feeding the outputted aligned face image to the developed model Finally, it would be interesting to implement this model on a live business service. There are various situations in which it could be implemented, such as crowd analytics, intelligent marketing, and HCI (human-computer interaction). This is the current tendency: one small step in the Computer Vision field, a giant leap for the fusion between humankind and technology.

7.REFERENCES

Abousaleh, F. S., Lim, T., Cheng, W.-H., Yu, N.-H., Hossain, M. A., & Alhamid, M. F. (2016). A novel comparative deep learning framework for facial age estimation. EURASIP

Journal on Image and Video Processing, 2016(1), 47. https://doi.org/10.1186/s13640-016- 0151-4 Arriaga, O., Valdenegro-Toro, M., & Plöger, P. (2017). Real-time Convolutional Neural Networks for Emotion and Gender Classification. Retrieved from http://arxiv.org/abs/1710.07557 Baluja, S., & Rowley, H. A. (2007). Boosting sex identification performance. International Journal of Computer Vision. https://doi.org/10.1007/s11263-006-8910-9

Bordes, A., Glorot, X., Weston, J., & Bengio, Y. (2012). Joint learning of words and meaning representations for open-text semantic parsing. Artificial Intelligence and Statistics, 127– 135. https://doi.org/10.1109/CVPR.2015.7298594

Bruna, J., Szlam, A., & LeCun, Y. (2014). Signal Recovery from Pooling Representations. ArXiv Preprint. Retrieved from http://arxiv.org/abs/1311.4025 Cadène, R., Thome, N., & Cord, M. (2016). Master's Thesis : Deep Learning for Visual Recognition. Retrieved from http://arxiv.org/abs/1610.05567 Canziani, A., Paszke, A., & Culurciello, E. (2017). An Analysis of Deep Neural Network Models for Practical Applications, 1–7. Retrieved from http://arxiv.org/abs/1605.07678 Caruana, R. (1997). Multitask learning. In Machine learning (pp. 41–75). https://doi.org/10.1109/TCBB.2010.22 Chen, D., Ren, S., Wei, Y., Cao, X., & Sun, J. (2014). Joint cascade face detection and alignment. In European Conference on Computer Vision (pp. 109–122). https://doi.org/10.1007/978-3-319-10599-4_8 Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017–Janua, 1800–1807. https://doi.org/10.1109/CVPR.2017.195

Ciresan, D., Meier, U., Masci, J., Gambardella, L. M., & Schmidhuber, J. (2011). HighPerformance Neural Networks for Visual Object Classification. Advances In Neural Information Processing Systems, 12. https://doi.org/http://arxiv.org/abs/1102.0183

Ciresan, D., Meier, U., & Schmidhuber, J. (2012a). Multi-column Deep Neural Networks for 84 Image Classification. International Conference of Pattern Recognition, (February), 3642–3649. https://doi.org/10.1109/CVPR.2012.6248110

Ciresan, D., Meier, U., & Schmidhuber, J. (2012b). Transfer Learning for Latin and Chinese Characters with Deep Neural Networks. Neural Networks (IJCNN), The 2012 International Joint Conference On, 1–6.

https://doi.org/10.1109/IJCNN.2012.6252544 Clevert, D.-A., Unterthiner, T., & Hochreiter, S. (2016). Fast and Accurate Deep Network Learning By Exponential Linear Units (Elus). Iclr 2016, 1–14. Cook, A. (2017). Batch Normalization. Retrieved August 10, 2018, from https://alexisbcook.github.io/2017/global-average-pooling-layers-for-objectlocalization/ Cope, G. (2017). Kernels in Image Processing. Retrieved November 30, 2017, from http://www.naturefocused.com/articles/photography-image-processing-kernel.html Dauphin, Y., Pascanu, R., Gulcehre, C., Cho, K., Ganguli, S., & Bengio, Y. (2014). Identifying and attacking the saddle point problem in high-dimensional non-convex optimization, 1–14. Retrieved from http://arxiv.org/abs/1406.2572 Dehghan, A., Ortiz, E. G., Shu, G., & Masood, S. Z. (2017). DAGER: Deep Age, Gender and Emotion Recognition Using Convolutional Neural Network. Retrieved from http://arxiv.org/abs/1702.04280 Ekmekji, A. (2016). Convolutional Neural Networks for Age and Gender Classificatio. Technical Report. Eldan, R., & Shamir, O. (2016). The Power of Depth for Feedforward Neural Networks, 1– 33. Retrieved from http://arxiv.org/abs/1512.03965 FER-2013. (2018). Retrieved December 12, 2017, from https://www.kaggle.com/c/challenges-in-representation-learning-facial-expressionrecognition-challenge/data Gao, F., & Ai, H. (2009). Face age classification on consumer images with gabor feature and fuzzy LDA method. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (Vol. 5558 LNCS, pp. 132–141). https://doi.org/10.1007/978-3-642-01793-3_14 Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. 580–587). https://doi.org/10.1109/CVPR.2014.81 85 Golomb, B. A., Lawrence, D. T., & Sejnowski, T. J. (1991). Sexnet: A neural network identifies sex from human faces. Advances in Neural Information Processing Systems 3, (July), 572–7. Retrieved from http://dl.acm.org/citation.cfm?id=118953 Gomez, V., Cortes, A., & Noguer, F. (2015). Object Detection for Autonomous Driving Using Deep Learning. Meeting of the Universitat Politecnica de Catalunya, Spain, (December). Gong, Y., Wang, L., Guo, R., &

Lazebnik, S. (2014). Multi-scale orderless pooling of deep convolutional activation features. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (Vol. 8695 LNCS, pp. 392–407). https://doi.org/10.1007/978-3-319-10584-0_26 Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., … Wang, G. (2017). Recent Advances in Convolutional Neural Networks. ArXiv, 1–14. https://doi.org/10.3389/fpsyg.2013.00124 Guo, G., & Mu, G. (2014). A framework for joint estimation of age, gender and ethnicity on a large database. Image and Vision Computing, 32(10), 761–770. https://doi.org/10.1016/j.imavis.2014.04.011 Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. Neurocomputing, 187, 27–48. https://doi.org/10.1016/j.neucom.2015.09.116 He, K., Zhang, X., Ren, S., & Sun, J. (2015a). Delving deep into rectifiers: Surpassing humanlevel performance on imagenet classification. In Proceedings of the IEEE International Conference on Computer Vision (Vol. 2015 Inter, pp. 1026–1034). https://doi.org/10.1109/ICCV.2015.123 He, K., Zhang, X., Ren, S., & Sun, J. (2015b). Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(9), 1904–1916. https://doi.org/10.1109/TPAMI.2015.2389824 He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90 Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. Retrieved from http://arxiv.org/abs/1502.03167 Karpathy, A. (2018). Convolutional Neural Networks for Visual Recognition. Retrieved June 19, 2018, from http://cs231n.github.io/optimization-1/#optimization Keras. (2018). Retrieved January 15, 2018, from https://keras.io/