IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁙ Impact Factor 8.066 ⁙ Peer-reviewed / Refereed journal ⁙ Vol. 10, Issue 7, July 2023

DOI: 10.17148/IARJSET.2023.107125

IDENTIFICATION OF AGE USING MTCNN ALGORITHM

Roshansingh Rajput[1], Samarth Handral[2], Abhishek Nazare[3]

Department of MCA, KLS Gogte Institute of Technology/VTU, India[1]

Department of MCA, KLS Gogte Institute of Technology/VTU, India[2]

Department of MCA, KLS Gogte Institute of Technology/VTU, India[3]

Abstract: The Convolutional Neural Network (CNN) has shown remarkable versatility across various applications. One such emerging area of research is age range estimation using CNN, which finds applications in numerous domains and continues to be a state-of-the-art field for investigation, aiming to enhance estimation accuracy.

In our proposed work, we leverage a deep CNN model to identify people's age ranges. Initially, we employ the MTCNN technique to extract only face images from the dataset, discarding unnecessary features unrelated to the face. To improve the model's performance, we utilize data augmentation through random crop techniques.

Our approach also incorporates transfer learning, where a pretrained face recognition model, VGG-Face, serves as the foundation for building our age range identification model. The effectiveness of our work is evaluated on the Adience Benchmark, where the performance on the test set significantly surpasses existing state-of-the-art methods.

Overall, our research demonstrates the prowess of CNN-based age range estimation, showcasing promising results and contributing to the advancement of this domain.

Keywords: Computer Vision, Image Detection, Feature Matching, MTCNN, NumPy, OpenCV, PyTorch , FaceNet , P-Net, O-Net , R-Net , NMS , Face-Alignment.

I. INTRODUCTION

Recently, face images have found diverse applications in face recognition[1], surveillance systems[15], emotion identification[16], and smart attendance systems, serving multifarious purposes.

However, age range identification using CNN has emerged as a challenging task, especially among adults where the age ranges appear similar compared to children and the elderly[8]. The distinct facial features of children and the elderly make their age identification relatively easier[7], but for adults, the presence of almost identical facial features complicates the process.

Among various neural network architectures, CNN has proven exceptionally effective in processing image data for tasks like object detection, face recognition, and image classification. This superiority stems from its ability to efficiently handle the vast number of parameters in image data, leveraging different convolutional layers to extract essential features and simplifying the image data processing through convolution with filters.

In our research, we have employed a novel approach to identify age ranges across different age groups. Instead of building a custom CNN model with limited datasets, we opted for transfer learning, which proves more effective. By utilizing a pre-trained model designed for image classification and face recognition[3], we achieved accurate age range estimation in our work. To enhance the performance of our face recognition system, we adopt a two-step approach, utilizing both MTCNN and FaceNet. First, we employ MTCNN, a widely used target detection network known for its high accuracy, lightweight nature, and real-time capabilities, to perform face detection. This provides us with precise face coordinates as the initial input.

The face recognition process is then divided into two main steps: face detection and face recognition. MTCNN performs face detection to obtain accurate face coordinates. Based on these results, FaceNet comes into play for face recognition.

DOI: 10.17148/IARJSET.2023.107125

MTCNN operates as follows: the test image is continuously resized to create an image pyramid. This pyramid is then fed into the P-Net, which generates numerous candidate face images. These candidates are further refined using the R-Net, and after removing many redundant candidates, the remaining images are processed by the O-Net. Finally, MTCNN outputs the precise bounding box coordinates for each face.

In comparison to DeepFace, FaceNet retains the crucial face alignment step but abandons the feature extraction process. Instead, it directly employs a CNN to train the model end-to-end after face alignment, which simplifies the overall procedure and streamlines the recognition process.

II. RELATED WORKS

A dataset with one million celebrity faces was created, which ultimately increased the accuracy of face recognition[1]. Face recognition with fresh data from numerous identities that minimize label noise[2], either utilizing a single It is done to analyze and survey an image or group of faces[3]. The VGG-Face model[4], the ranking SVM[5], and determining enough embedding space by applying a variety of learning techniques that effectively model data with multiple linear regression function[6] are used for age estimation. A craniofacial growth model is put out to simulate growth-related changes in the structure of human faces[7]. Different faces are separated into a number of tiny regions for the extraction of Local Binary Pattern(LBP), which is then concatenated into feature vectors and used as a face descriptor[8]. For determining age from various angles, two new approaches are proposed: Ranking-CNN[9] based on rank relationship and deep CNN architecture for low quality face images[10]. Identification of gender and age is carried out using unfiltered faces on Adience Benchmark[11] and deep CNN[12] in the presence of little learning data. There are also numerous research challenges.

The studies mentioned above demonstrated that a few research projects have already been completed using various measurements and techniques. Despite the vast amount of work that has been done, a substantial advancement in the recognition of age range in the adult age group due to similar facial features has not been made. In this experiment, we first isolated just the facial features from the joint face alignment using the MTCNN network, and then we determined age range using the VGG-Face model, which greatly reduced overfitting and enhanced identification among the adult age group in comparison to earlier research. In order to assess the effectiveness of our work utilizing unconstrained face photos, a comparison analysis of several face models has also been conducted.

III. DATA PREPARATION

In our suggested study, the facial features are extracted using a network called MTCNN (Multi-task Cascaded Convolutional Networks)[17]. MTCNN consists of three CNN phases, Proposal Network (P-net), Refine Network (R-net), and Output Network (O-net), respectively. Proposal Network (P-Net), a CNN model, is fed all of the input images. As a result of P-net, the candidate window inside the image and its bounding box regression vector are received. In order to recognize faces of all sizes, as shown in Fig. 2, we build an image pyramid.

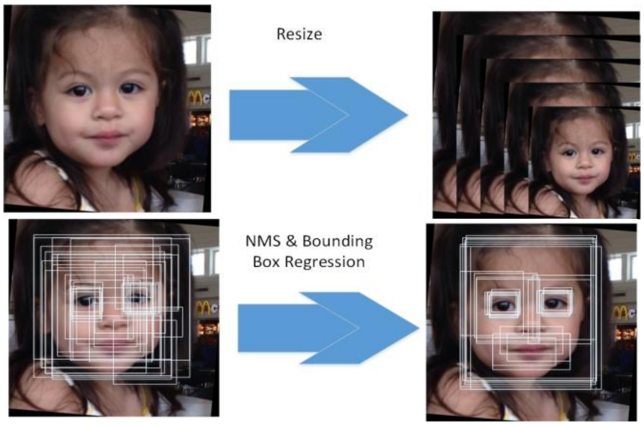

Fig. 2 demonstrates the use of image pyramids, NMS, and bounding box

Following the feeding of those candidates to the refined network (R-Net), we identified the bounding box regression vector and employed Non-Max Suppression (NMS) to locate and integrate highly overlapped candidates, as shown in Fig. 3. Results from this network are sent to O-Net.

DOI: 10.17148/IARJSET.2023.107125

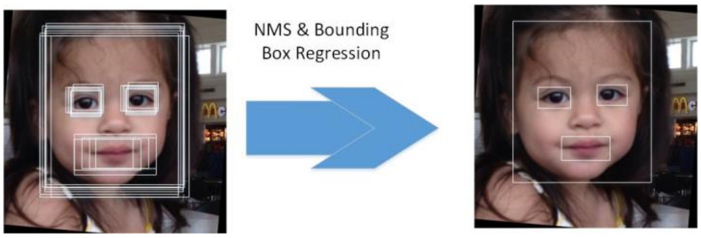

Fig. 3. NMS and bounding box regression (R-Net)

In Fig. 4, the Output Network (O-Net) produces three essential outputs: the bounding box coordinates, the coordinates of five facial landmarks, and the confidence level for each detected box. The resulting bounding box image is saved as a new image and subsequently fed into the proposed model for further processing.

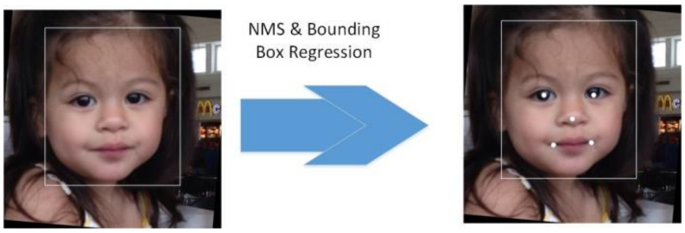

Fig. 4. Output-network

The dataset's images are resized to 256x256 pixels. Using a random crop, these photos were recovered to be 224x224 pixels in size[18]. To enhance the data, the retrieved photos are rotated and flipped horizontally. Eight classes were created, one for each age group. Images were divided 80/20 for training and validation purposes for each class. Finally, the network receives this dataset.

IV. WORKING OF MTCNN

MTCNN (Multi-Task Cascaded Convolutional Networks) is a popular deep learning-based face detection algorithm. It was proposed in the paper "Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks" by Zhang et al. in 2016.

MTCNN is designed to perform three tasks simultaneously: face detection, facial landmark localization, and face alignment. Let's go through the key steps of how MTCNN works:

1.Image Pyramid Generation:
MTCNN operates on an image pyramid, which is a set of scaled-down versions of the input image. By using an image pyramid, MTCNN can detect faces at different scales, allowing it to handle faces of various sizes in the input image.

2.First Stage - Proposal Network (P-Net):
The first stage, or the Proposal Network (P-Net), is responsible for generating candidate face regions (bounding boxes) that potentially contain faces. P-Net is a fully convolutional network that takes the input image and scans it with a small window to classify each location as face or non-face. Additionally, it regresses the bounding box coordinates for potential face regions.

3.Non-Maximum Suppression (NMS):
After obtaining the candidate face regions from P-Net, non-maximum suppression is applied to filter out highly overlapping bounding boxes, keeping only the ones with the highest confidence score. This step helps to reduce redundant detections.

4.Second Stage - Refine Network (R-Net):
The second stage, or the Refine Network (R-Net), further refines the candidate face regions generated by P-Net. R-Net takes the regions proposed by P-Net and refines them by classifying the regions as face or non-face more accurately and adjusting the bounding box coordinates accordingly.

5.NMS and Face Alignment:
Similar to the first stage, non-maximum suppression is applied again to filter out redundant detections after R-Net. Additionally, MTCNN performs facial landmark localization during this stage, identifying the facial landmarks (such as eyes, nose, and mouth) within each detected face region. These facial landmarks are then used for face alignment, which helps normalize the faces' orientation.

6.Third Stage - Output Network (O-Net):
The third and final stage, or the Output Network (O-Net), performs a more precise face classification and bounding box regression. O-Net takes the refined face regions from the second stage and further improves the face classification and localization accuracy.

V. MULTI-SCALE DETECTION CASCADE

The multiscale detection cascade of CNN consists of a series of face detector. Each handles only one face of her. Relatively small range. Each face detector consists of: from 2 stages. In the first phase, multiscale proposals are generated. From a full convolutional network. Then the second stage Face and non-face predictions for candidate windows generated in the first stage. If the candidate window is classified

As for the face, we will continue to adjust the position of the candidate window.

Sample Dataset:

{'box': [1942, 716, 334, 415], 'confidence': 0.9999997615814209, 'keypoints': {'left_eye': (2053, 901), 'right_eye': (2205, 897), 'nose': (2139, 976), 'mouth_left': (2058, 1029), 'mouth_right': (2206, 1023)}}
{'box': [2084, 396, 37, 46], 'confidence': 0.9999206066131592, 'keypoints': {'left_eye': (2094, 414), 'right_eye': (2112, 414), 'nose': (2102, 426), 'mouth_left': (2095, 432), 'mouth_right': (2112, 431)}}
{'box': [1980, 381, 44, 59], 'confidence': 0.9998701810836792, 'keypoints': {'left_eye': (1997, 404), 'right_eye': (2019, 407), 'nose': (2010, 417), 'mouth_left': (1995, 425), 'mouth_right': (2015, 427)}}
{'box': [2039, 395, 39, 46], 'confidence': 0.9993435740470886, 'keypoints': {'left_eye': (2054, 409), 'right_eye': (2071, 415), 'nose': (2058, 422), 'mouth_left': (2048, 425), 'mouth_right': (2065, 431)}}

VI. REQUIREMENTS

1.MTCNN Library: You need to have the MTCNN library installed. MTCNN is a popular face detection and alignment algorithm that can be used for age estimation as well.

2.Python and Libraries: Ensure you have Python installed on your system. Additionally, you will need the following Python libraries:

- NumPy: For numerical computing.
- OpenCV: For image processing and computer vision tasks.
- TensorFlow or PyTorch: These deep learning frameworks are used to implement and run the MTCNN model.

3.Age Labeled Dataset: You will require a dataset with images labeled with corresponding ages. The dataset should be diverse and representative of the age groups you wish to detect.

4.Pre-trained Model: Obtain a pre-trained MTCNN model that includes age estimation capabilities. Several versions of MTCNN exist with various pre-trained models available.

5.Hardware Resources: Age detection using MTCNN can be computationally intensive, especially if you are processing a large number of images in real-time. Ensure that you have adequate CPU/GPU resources to run the model efficiently.

6. Data Preprocessing: You will need to preprocess the input images to fit the requirements of the MTCNN model. Common preprocessing steps include resizing, normalization, and face alignment.

7. Evaluation Metrics: Depending on the use case, you may need evaluation metrics to measure the accuracy of the age estimation. Common metrics include Mean Absolute Error (MAE) or Mean Squared Error (MSE) between the predicted ages and ground truth ages.

8. Deployment Environment: Determine the environment where you will deploy the age detection system. It could be a local machine, server, or cloud-based infrastructure.

DOI: 10.17148/IARJSET.2023.107125

VII. CAN BE IMPLEMENT

The MTCNN (Multi-Task Cascaded Convolutional Networks) algorithm is primarily designed for face detection, facial landmark localization, and alignment. While it is a powerful tool for locating faces in images and extracting facial landmarks, it is not specifically intended for age estimation.

Age estimation is a more complex task that involves predicting the age of an individual based on their facial features and appearance. It requires specialized models and datasets designed explicitly for age prediction.

However, MTCNN can be used as a pre-processing step for age estimation. By using MTCNN to detect faces and localize facial landmarks, you can extract relevant regions of interest (ROI) for age estimation models. These ROIs can then be fed into age estimation algorithms like deep learning-based convolutional neural networks (CNNs) to predict the age of the individuals.

CONCLUSION

In conclusion, Age Detection using the MTCNN algorithm presents a robust and efficient solution for estimating the age of individuals from images. By leveraging deep learning techniques and facial feature extraction, MTCNN enables accurate age estimation, facilitating numerous practical applications across industries. As the field of computer vision and deep learning continues to advance, further improvements in age estimation accuracy and the development of more comprehensive datasets will enhance the reliability and applicability of this technology.

REFERENCES

[1] B. Yang, J. Yan, Z. Lei, and S. Z. Li, “Aggregate channel eatures for multi-view face detection,” in IEEE International Joint Conference on Biometrics, 2014, pp. 1-8.

[2] Guo Y., Zhang L., Hu Y., He X., Gao J. (2016) MS-Celeb-1M: A Dataset and Benchmark for Large-Scale Face Recognition. In: Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision – ECCV 2016.

[3] https://www.youtube.com/watch?v=ZjbWF9f3VD4

[4] K. Liu, H. Liu, S. Pei, T. Liu and C. Chang, ”Age Estimation on Low Quality Face Images,” 2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS), Hsinchu, Taiwan, 2019, pp. 295-296, doi: 10.1109/AICAS.2019.8771612.

[5] Y. Fu, Y. Xu and T. S. Huang, ”Estimating Human Age by Manifold Analysis of Face Pictures and Regression on Aging Features,” 2007 IEEE International Conference on Multimedia and Expo, Beijing, 2007, pp. 1383-1386, doi: 10.1109/ICME.2007.4284917.

[6] Taigman, Y.; Yang, M.; Ranzato, M.; and Wolf, L. 2014. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[7] R. Takahashi, T. Matsubara and K. Uehara, ”Data Augmentation using Random Image Cropping and Patching for Deep CNNs,” in IEEE Transactions on Circuits and Systems for Video Technology, doi: 10.1109/TCSVT.2019.2935128.

[8] K. Liu, H. Liu, S. Pei, T. Liu and C. Chang, ”Age Estimation on Low Quality Face Images,” 2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS), Hsinchu, Taiwan, 2019, pp. 295-296, doi: 10.1109/AICAS.2019.8771612.

[9] S. Kumar, S. Singh and J. Kumar, ”A study on face recognition techniques with age and gender classification,” 2017 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, 2017, pp. 1001-1006, doi: 10.1109/CCAA.2017.8229960.

[10] S. H. Nam, Y. H. Kim, N. Q. Truong, J. Choi and K. R. Park, ”Age Estimation by Super-Resolution Reconstruction Based on AdversarialNetworks,” in IEEE Access, vol. 8, pp. 17103-17120, 2020, doi:10.1109/ACCESS.2020.2967800.

[11] Zhu, H., Zhang, Y., Li, G. et al. Ordinal distribution regression for gaitbased age estimation. Sci. China Inf. Sci. 63, 120102 (2020).

[12] R. Takahashi, T. Matsubara and K. Uehara, ”Data Augmentation usingRandom Image Cropping and Patching for Deep CNNs,” in IEEE Transactions on Circuits and Systems for Video Technology, doi: 10.1109/TCSVT.2019.2935128.

[13] Sawant, M.M., Bhurchandi, K.M. Age invariant face recognition: asurvey on facial aging databases, techniques and effect of aging. ArtifIntell Rev 52, 981–1008 (2019).

DOI: 10.17148/IARJSET.2023.107125

[14] Cao D., Lei Z., Zhang Z., Feng J., Li S.Z. (2012) Human Age Estimation using Ranking SVM. In: Zheng WS., Sun Z., Wang Y., Chen X., YuenP.C., Lai J. (eds) Biometric Recognition. CCBR 2012. Springer, Berlin, Heidelberrg.
[15] L. O. Chua and T. Roska, “Reprogrammable CNN and supercomputer,”U.S. Patent 5 355 528, Oct. 11, 1994.
[16] X. Cao, Y. Wei, F. Wen, and J. Sun, “Face alignment by explicit shape regression,” International Journal of Computer Vision, vol 107, no. 2, pp. 177-190, 2012.
[17] S. Yang, P. Luo, C. C. Loy, and X. Tang, “From facial parts responses to face detection: A deep learning approach,” in IEEE International Confer-ence on Computer Vision, 2015, pp. 3676-3684.
[18] Q. Zhu, M. C. Yeh, K. T. Cheng, and S. Avidan, “Fast human detection using a cascade of histograms of oriented gradients,” in IEEE Computer Conference on Computer Vision and Pattern Recognition, 2006, pp. 1491-1498.
[19] C. Zhang, and Z. Zhang, “Improving multiview face detection with mul-ti-task deep convolutional neural networks,” IEEE Winter Conference on Applications of Computer Vision, 2014, pp. 1036-1041.
[20] T. F. Cootes, G. J. Edwards, and C. J. Taylor, “Active appearance models,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pp. 681-685, 2001.