Residual Blocks-Based Convolutional Neural Network for Age, Gender, and Race Classification

[1]Khasanova Nodira Gayrat Kizi, [2]Bong-Kee Sin

1,2Dept. of Artificial Intelligence Convergence, Graduate School Pukyong National University

[1]khasanovanodira55@gmail.com, [2]bkshin@pknu.ac.kr

연령, 성별, 인종 구분을 위한 잔차블록 기반 컨볼루션 신경망

하사노바 노디라[1], 신봉기[2]
부경대학교 인공지능융합학과 석사과정[1]
부경대학교 인공지능융합학과 교수[2]

Abstract

The problem of classifying of age, gender, and race images still poses challenges. Despite deep and machine learning strides, convolutional neural networks (CNNs) remain pivotal in addressing these issues. This paper introduces a novel CNN-based approach for accurate and efficient age, gender, and race classification. Leveraging CNNs with residual blocks, our method enhances learning while minimizing computational complexity. The model effectively captures low-level and high-level features, yielding improved classification accuracy. Evaluation of the diverse 'fair face' dataset shows our model achieving 56.3%, 94.6%, and 58.4% accuracy for age, gender, and race, respectively.

1. Introduction

Age, gender, and race classification models are increasingly vital across various domains, addressing social and technological challenges, enhancing inclusivity, and improving decision-making processes. They offer valuable insights for healthcare, marketing, human-computer interaction, and security. Responsible development and ethical considerations are paramount. These models find applications in individual identification, targeted advertising, and behavior predictions. In tourism, they help categorize visitors for historical sites. In marketing, demographic profiling guides customized advertisements. The proposal of CNN architecture that combines convolutional layers with residual blocks, that offers efficient feature extraction and outperforming established models. We employ the Fairface dataset and conduct a comparative analysis. This paper outlines related works, datasets, model architecture, experimental results, and future research directions."

2. Related works

In the domain of deep neural networks, multi-task learning is approached through two primary structures, soft parameter sharing, where separate networks for each task share a common structure and utilize a similarity function for regularization, and hard parameter sharing, where a single shared network is employed alongside task-specific output layers. Soft sharing, while effective, increases runtime space complexity with the number of tasks. Hard sharing is often preferred to mitigate this, reducing space complexity and preventing overfitting. Our research explores age, gender, and race classification within computer vision and machine learning, leveraging Convolutional Neural Networks (CNNs) like AlexNet, VGGNet, ResNet, and DenseNet, which have shown significant success in these areas. Our study used a dataset of 108,501 facial images from sources like YFCC-100M Flickr, Twitter, and online newspapers. The FairFace dataset includes seven racial groups (White, Black, Indian, East Asian, Southeast Asian, Middle Eastern, and Latino), and it was well-balanced, with nearly equal samples for all groups, as shown in Figure 1 [5].

3. Proposed model architecture.

In this paper, we proposed a novel approach to multi-label classification, distinct from conventional CNN architectures. The model is specifically designed to simultaneously predict age, gender, and race. It creatively integrates residual blocks and dedicated fully connected layers, which permits the network to effectively grasp distinct features associated with each label. This strategy leads to improved predictive accuracy across various demographic attributes, all within a single comprehensive model. Figure 4 shows a block diagram of a CNN algorithm that uses residual blocks to classify images based on age, gender, and race. The algorithm first preprocesses the input image, and then feeds it into a series of convolutional layers. The convolutional layers extract features from the image, which are then passed to a series of residual blocks. The residual blocks help to improve the performance of the algorithm by allowing it to learn residual information, which is the difference between the input image and the desired output. Finally, the output of the residual blocks is passed to a series of fully connected layers, which classify the image into one of the three categories (age, gender, or race).

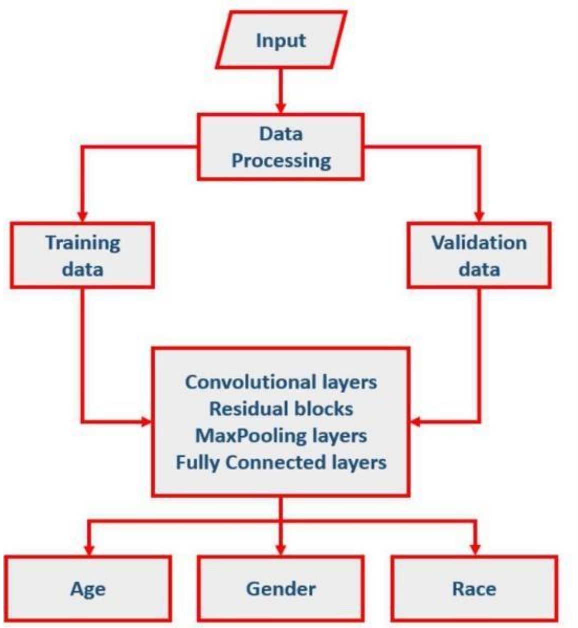

Figure 2: Proposed model architecture flowchart.

The model includes convolutional and pooling layers organized into three blocks for hierarchical feature learning (see Figure 3).

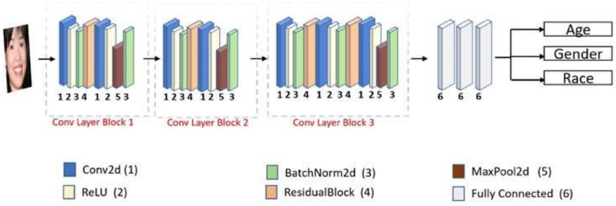

Figure 3: General view of the proposed model.

Convolutional layers employ ReLU activations and batch normalization extract features effectively. Residual blocks aid gradient flow and information preservation, while batch normalization and dropout layers prevent overfitting. Residual blocks tackle deep network training challenges by addressing vanishing gradient issues and capturing residual features. After convolutional layers, feature maps are flattened and input into fully connected layers for age, gender, and race prediction.

Fully connected layers consist of three branches, each responsible for predicting a specific label (age, gender, or race). Each branch begins with a linear layer, followed by ReLU activation and dropout. The third branch uses Alpha Dropout. The final fully connected layer in each branch produces the corresponding number of classes (9 for age, 2 for gender, and 7 for race). This model excels in accurately capturing and classifying diverse facial attributes, contributing to computer vision and socio-demographic analysis advancements.

4. Dataset

Our study used a dataset of 108,501 facial images from sources like YFCC-100M Flickr, Twitter, and online newspapers. The FairFace dataset includes seven racial groups (White, Black, Indian, East Asian, Southeast Asian, Middle Eastern, and Latino), and it was well-balanced, with nearly equal samples for all groups, as shown in Figure 1 [5].

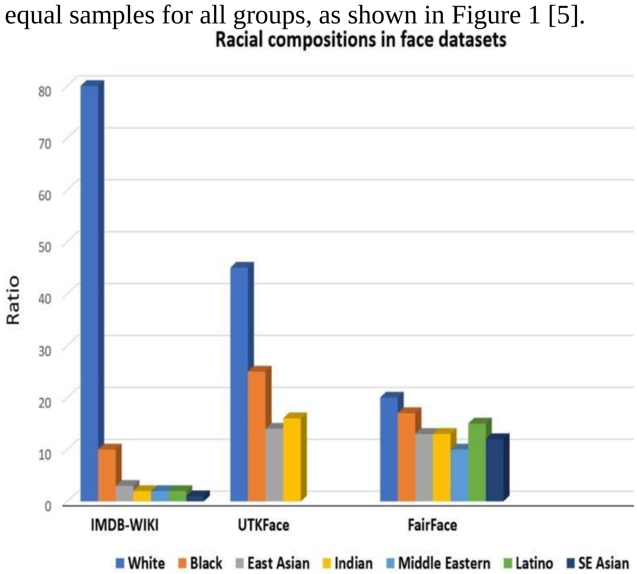

Figure 1: Racial compositions in face datasets.

5. Experimental Results

The experiments include various key decisions, such as setting the initial learning rate to 0.0001 and utilizing the Adam optimizer for network weight updates. We conducted training with batch sizes of 64 for both training and validation, over a maximum of 100 epochs.

Table 1 offers a comprehensive model comparisonin in terms of accuracy on fairface, revealing our model's superiority in age, gender, and race classification despite its lower complexity. Our model achieved an impressive 56.34% accuracy for age, 94.67% for gender, and 58.46% for race prediction.

Table 1: Comparison of the proposed model against existing models.

Model	**Accuracy**			**Number of parameters**
	Age	**Gender**	**Race**	
	%	%	%	Million
Resnet34	53.2	90.9	63.5	21.8
Densenet61	53.5	90.8	63.6	26.6
AlexNet	44.5	83.3	54.3	61.1
MobileNet_v2	41.9	80.3	47.5	3.5
Our model	**56.3**	**94.6**	**58.4**	**35.3**

In summary, the results validate the effectiveness of the proposal model in accurately predicting age, gender, and race, positioning it as a high-performing solution in these classification tasks.

6. Conclusion

The proposed CNN design excels in age, gender, and race classification, particularly in age and gender prediction, achieving accuracies of 56.3%, and 94.6%, respectively. Key components like ResidualBlock structures and multi-label classification empower the model to effectively capture complex features while predicting multiple attributes. This research highlights the potential of innovative CNN architectures in addressing intricate attribute recognition challenges.

Reference

[1] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," IEEE Conf. on Computer Vision and Pattern Recognition Workshops, 2015.

[2] Y. Sun, X. Wang, and X. Tang. Deep learning face representation from predicting 10,000 classes. In Proc. Conf. Comput. Vision Pattern Recognition, IEEE, 2014.

[3] Shafiq, Muhammad, and Zhaoquan Gu. "Deep residual learning for image recognition: a survey." *Applied Sciences* 12, no. 18 (2022).

[4] Galassi, Andrea, Marco Lippi, and Paolo Torroni. "Multi-task attentive residual networks for argument mining." *arXiv preprint arXiv:2102.12227* (2021).

[5] Karkkainen, Kimmo, and Jungseock Joo. "Fairface: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation." In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021.

[6] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Communications of the ACM* 60, no. 6 (2017).

[7] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).

[8] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.Hassler, "Age and gender classification using convolutional neural networks," IEEE Conf. on Computer Vision and Pattern Recognition Workshops, 2015.