Regular paper

Journal of the KIECS. pp. 99-104, vol. 17, no. 1, Feb. 28. 2022, t. 111, pISSN 1975-8170 | eISSN 2288-2189
http://dx.doi.org/10.13067/JKIECS.2022.17.1.99

소규모 합성곱 신경망을 사용한 연령 및 성별 분류

Uraimov Jamoliddin* · 류재홍**

Age and Gender Classification with Small Scale CNN

Uraimov Jamoliddin* · Jae Hung Yoo**

요 약

인공지능은 놀라운 이점으로 우리 삶의 중요한 부분을 차지하고 있다. 기계는 이미지에서 물체를 인식하는 것, 특히 사람들을 정확한 나이와 성별 그룹으로 분류하는 것에 있어서 인간을 능가하고 있다. 이러한 측면에서 나이와 성별 분류는 최근 수십 년 동안 컴퓨터 비전 연구자들 사이에서 뜨거운 주제 중 하나였다. 심층 합성곱 신경망(CNN) 모델의 배포는 최첨단 성능을 달성했다. 그러나 대부분의 CNN 기반 아키텍처는 수십 개의 훈련 매개 변수로 매우 복잡하기 때문에 많은 계산 시간과 자원이 필요하다. 이러한 이유로 기존 방법에 비해 훈련 매개 변수와 훈련 시간이 현저히 적은 새로운 CNN기반 분류 알고리즘을 제안한다. 덜 복잡함에도 불구하고 우리 모델은 UTKFace 데이터 세트에서 연령 및 성별 분류의 더 나은 정확도를 보여준다.

ABSTRACT

Artificial intelligence is getting a crucial part of our lives with its incredible benefits. Machines outperform humans in recognizing objects in images, particularly in classifying people into correct age and gender groups. In this respect, age and gender classification has been one of the hot topics among computer vision researchers in recent decades. Deployment of deep Convolutional Neural Network(: CNN) models achieved state-of-the-art performance. However, the most of CNN based architectures are very complex with several dozens of training parameters so they require much computation time and resources. For this reason, we propose a new CNN-based classification algorithm with significantly fewer training parameters and training time compared to the existing methods. Despite its less complexity, our model shows better accuracy of age and gender classification on the UTKFace dataset.

키워드

Age Classification, Gender classification, Unfiltered Image Recognition, Imbalanced Classification Problems
나이 분류, 성별 분류, 필터링 되지 않은 영상 인식, 불균형 분류 문제

Ⅰ. Introduction

Community interactions are strongly rely on members' age and gender. These interactions varies in salutations and communication approaches for males towards females and youngs towards elders [1]. Recent development of technologies has allowed application of age, gender attributes in multiple

* 전남대학교 컴퓨터공학과 석사과정 재학
(uraimov-net@mail.ru)
** 교신저자: 전남대학교 컴퓨터공학과
・접 수 일 : 2021. 12. 31
・수정완료일 : 2022. 01. 24
・게재확정일 : 2022. 02. 17

・Received : Dec. 31, 2021, Revised : Jan. 24, 2022, Accepted : Feb. 17, 2022
・Corresponding Author : Jae Hung Yoo
Dept. of Computer Engineering, Chonnam Nat. Univ
Email : jhy@jnu.ac.kr

domains, including security and video surveillance, customer behavior analyses and prediction, biometrics, electronic vending machines, smart stores, human-computer interaction [2]. This led research communities to develop more sophisticated algorithms with more accurate classification capabilities

Early studies suggested classifying age and gender attributes in facial photographs based on differences in facial feature sizes[3]. These features like nasal, oral, eyes sizes and distance and their ratio calculated for classification. However, the machine learning techniques used in these early systems did not fully utilize the huge number of face images and public datasets to improve their classification performance.

More recently, the emerging and state-of-the-art methods are neural network based techniques providing highest accuracy. However, these approaches rely on the very complex computation architecture, thus requires much resources and time for model training. The goal of our work is to build method with reduced architecture but better performance. Our proposed deep learning architecture consists of six convolution layers followed by two fully connected layers pretrained on UTKFace dataset [4] for age and gender classification.

In this paper, a small scale CNN networks are investigated for the image classification problems. In section II, typical CNN are introduced. In section III, image datasets are reviewed. In section IV, proposed architecture is explained. In section V, experimental results show the effectiveness of the proposed method followed by the conclusion and reference sections [5-17].

Ⅱ. Related works

In the last decades, we are witnessing that deep learning techniques have shown excellent performance. Generally, most of the works rely on the regression, classification, ranking and label distribution based methods [5].

Regression method models facial ageing as a regression issue and outputs predicted age values from facial images. Thus, they mostly deploy Euclidean loss as optimization function. Yi et al. [6] proposed a multi-scale CNN for age regression and deployed mean squared loss for model training. Whereas, classification approaches performed the age prediction as a multi-class categorizing issue and separate each age into independent classes[1,7]. Classification approach eases the CNN network training but does not handle interclass relations [5].

Ranking CNN [8] built a series of CNN-based binary classifiers and summed up their predictions to obtain the estimated age.

Another approach that has showed the state-of-the-art performance on different age prediction benchmarks is Label Distribution Learning (LDL) [11], however, it solves the age prediction as a probability distribution over all potential age vectors. In [9,10], the authors proposed to take the expectation value of output distribution as the predicted age.

These methods have been verified effectively on constrained imaging conditions, however, they require long training times because the model architecture is very complex and the number of parameters to be trained is large.

Ⅲ. Datasets

The UTKFace dataset contains over 20,000 face color images for different age ranges (from infants to 116 years old). The data set can be used for multiple purposes, such as landmark localization, classification, regression, etc. because the images have age, gender, and ethnicity labels. It consists of images in various poses, resolutions, and lighting (Figure 1). Figure 2 shows the age distribution of images in the data set. To train the proposed model, we pre-processed the input image and scaled it to 256 x 256 pixels. Table 1 shows the distribution of the UTKFace dataset. The test data consists of 20% of the full set. The validation data has 15% of the remaining 80% of the dataset using the stratified random sampling for the 8 predefined age groups.

Fig. 1 Sample facial images from UTKFace dataset.

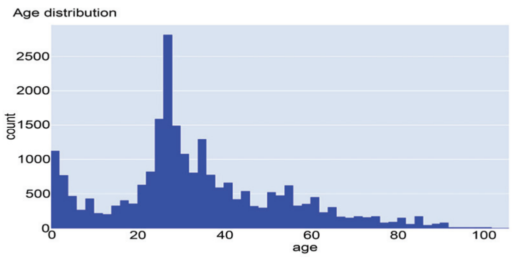

Fig. 2 Age distribution of images in the dataset.

Table 1. Cardinality of UTKFace dataset.

Training	Validation	Testing	Total
16391	2892	4821	24104

Ⅳ. Proposed Method

In this section we present our proposed CNN model of age and gender classification. To simplify network architecture and decrease number of parameters we applied simple convolutional neural network design. Figure 3 shows the general view of the proposed model.

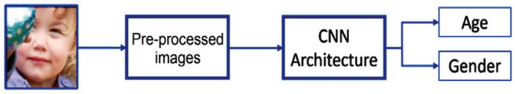

Fig. 3 General view of our model.

Model architecture is given in figure 4. The CNN design is an end-to-end sequential deep learning architecture for both age and gender classification jointly, including feature extraction and classification phases.

The feature extraction phase has six convolutional layers, with the corresponding parameters, including the number of filters, the kernel size of each filter, and the stride. It contains the convolutional layer, activation layer (rectified linear unit (ReLU)), batch normalization (instead of the conventional Local Response Normalization), max-pooling layer.

The classification stage, on the other hand, contains two fully connected layers, that handle the classification phase of the model. The both fully connected layer contain 512 neurons, followed by a ReLU, and a dropout layer at a dropout ratio of 0.5. Finally, we applied two output neurons separately, 8 for age group classes and 2 for gender classes respectively. Figure 4 shows the numerical view of the model.

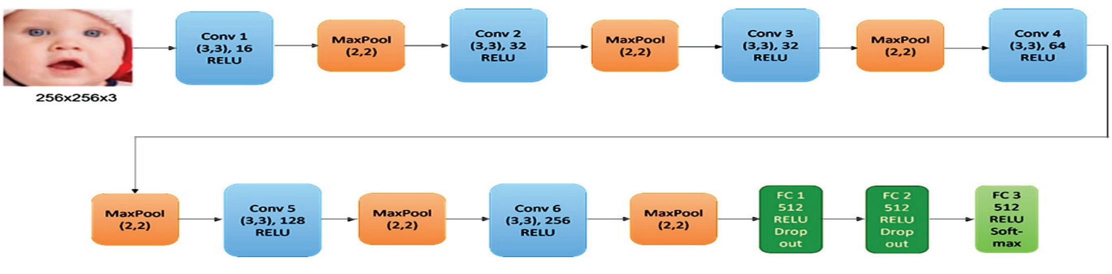

Fig. 4 Numeric view of the proposed model

Ⅴ. Experimental Results

To train the age and gender classification CNN, after series of empirical experiments, we set the initial learning rate to be 0.0001 to allow model train longer. To make our model able to generalize and extract features correctly, we apply the Adam optimizer to update network weights during training. We set training and validation batch size to 32 and performed training with max epoch=30 [12].

Table 2 shows comparison of our model against other state-of-the-art methods. Despite of having less complexity our model provided highest accuracy on both age and gender classification. We achieved 61.34% and 88.67% correct classification of age and gender respectively. This gives the improvement compare to the 50.7% and 86.8% accuracy of the previous system for unfiltered data reported in [1].

Table 2. Comparison of the proposed method against existing methods.

Model	Accuracy Age %	Accuracy Gender %	Number of Parameters M	Training Time min
AlexNet	55.96	85.1	61	398
VGG11	57.06	83.63	133	90
ResNet18	58.6	87.4	11	120
DenseNet161	59.22	87.28	26.6	150
Our model	61.34	88.67	11.5	75

Performance of our proposed model can be concluded from the confusion matrix in the figure 5 and figure 6. It is obvious from the confusion matrix that our model learned age classes of 0-2, 25-32 and 60 +. However, there are some cases that it confused to correctly distinguish neighbor classes. Figure 6 proves that our model in very rare cases misclassifies gender from the images.

Although our system can be accepted as the competitive candidate method for the multi-class age categorization problem, it can be denied as the solution of underage prevention system for buying age-restricted goods from the store. Confusion matrix in the figure 5 shows that system predict over 52% of age group 15-20 into older age groups. This is due to the imbalanced distribution of class members. Data augmentation does not help at all. We solve this problem by introducing the class weight that can be applied to the weighted loss calculation.

$$w_i = E(m_i)/m_i \,. \quad (1)$$

Here, m_i is the cardinality of class i in the training set. E is the expectation operator. Corresponding confusion matrix is showing in the figure 7. Across each row the diagonal entry dominates the off diagonal ones. The balanced system with classwise weighted loss can be accepted as the solution of underage prevention. But trade off using the balanced system is that the accuracy will be down to 52.85%. That still be comparable to 50.7% in the previous system for unfiltered data reported in [1].

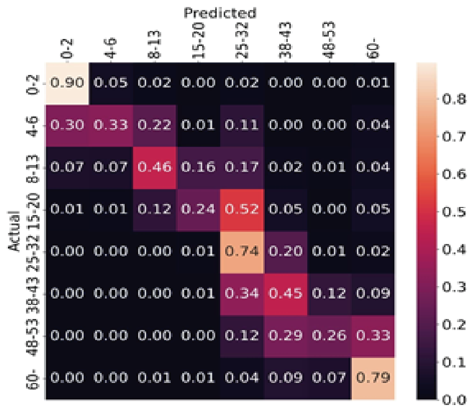

Fig. 5 Confusion matrix of age classification

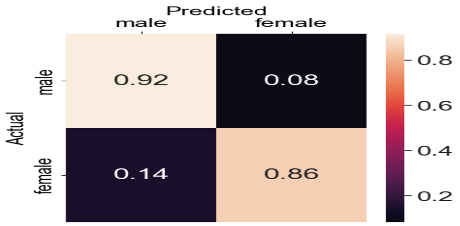

Fig. 6 Confusion matrix of gender classification

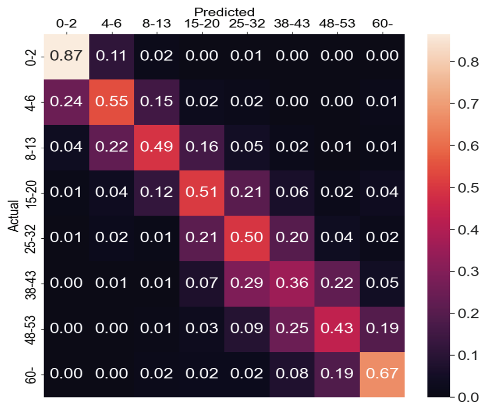

Fig. 7 Confusion matrix of age classification with classwise weighted loss.

Ⅵ. Conclusions

In this paper, a new convolutional neural networks based model is proposed to classify age and gender from facial images. Proposed model is trained on UTKFace dataset. Experimental results showed that our model achieved increased accuracy in both age and gender classification outperforming well-known deep CNN architectures. Complement system with the classwise weighted loss overcomes the difficulty in pursuing the underage prevention system. Our future work is going to be exploring the proposed model to other facial related tasks.

References

[1] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," *IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 34-42.

[2] O. Agbo-Ajala and S. Viriri, "Deeply Learned Classifiers for Age and Gender Predictions of Unfiltered Faces," *Hindawi The Scientific World Journal*, vol. 2020, AID 1289408, Apr. 2020, pp. 1-12.

[3] Y. Kwon and N. da V. Lobo, "Age classification from facial images," *Computer vision and image understanding*, vol. 74, 1999, pp. 1-21.

[4] Z. Zhang, Y. Song, and H. Qi, "Age Progression/Regression by Conditional Adversarial Autoencoder," *IEEE Conf. on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 2017, pp. 4352-4360.

[5] Y. Lin, J. Shen, Y. Wang, and M. Pantic, "FP-Age: Leveraging Face Parsing Attention for Facial Age Estimation in the Wild," *arXiv preprint* arXiv:2106.11145, 2021, pp. 1-10.

[6] D. Yi, Z. Lei, and S. Z. Li, "Age estimation by multi-scale convolutional network," in *Computer Vision – ACCV 2014, Workshops:* Singapore, Singapore, Nov. 2014, pp. 144–158.

[7] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," *IEEE Trans. on Information Forensics and Security*, vol. 9, no. 12, 2014, pp. 2170–2179.

[8] S. Chen, C. Zhang, M. Dong, J. Le, and M. Rao, "Using Ranking-CNN for Age Estimation," in *2017 IEEE Conf. on Computer Vision and Pattern Recognition*, July 2017, pp. 742–751

[9] R. Rothe, R. Timofte, and L. V. Gool, "DEX: Deep Expectation of Apparent Age from a Single Image," in *IEEE Int. Conf. on Computer Vision Workshop*, Dec. 2015, pp. 252–257.

[10] R. Rothe, R. Timofte, and L. V. Gool, "Deep Expectation of Real and Apparent Age from a Single Image Without Facial Landmarks," *Int. J.*

of Computer Vision, vol. 126, no. 2, Apr. 2018, pp. 144 - 157.
[11] B. B. Gao, H. Y. Zhou, J. Wu, and X. Geng, "Age estimation using expectation of label distribution learning," in *Proc. of the 27th Int. Joint Conf. on Artificial Intelligence,* Stockholm, Sweden, July 2018, pp. 712 - 718.
[12] D. P. Kingma and L. J. Ba, "Adam, A method for stochastic optimization," In *Proceedings of the Int. Conf. for Learning Representations (ICLR),* San Diego, CA, USA, May 2015, pp. 1 - 15.
[13] S. Salimov and J. Yoo, "A Design of Small Scale Deep CNN Model for Facial Expression Recognition using the Low Resolution Image Datasets," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 16, no. 1 , 2021, pp. 75-80.
[14] J. Yoo, "An Extension of Unified Bayesian Tikhonov Regularization Method and Application to Image Restoration," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 15, no. 01, 2020, pp. 161-166.
[15] J. Yoo, "A Unified Bayesian Tikhonov Regularization Method for Image Restoration," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 11, no. 11, 2016, pp. 1129-1134.
[16] J. Yoo, "Self-Regularization Method for Image Restoration," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 11, no. 1, 2016, pp. 45-52.
[17] S. Park, "Optimal QP Determination Method for Adaptive Intra Frame Encoding," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 10, no. 9, 2015, pp. 1009-1018.

저자 소개(Authors)

우라이모프 자멀리딘 (Uraimov Jamoliddin Kuldashali ugli)

2018년 Tashkent University of Information Technologies, Software Engineering and Information security 공학과 졸업(공학사)
2021년 현재 전남대학교 대학원 컴퓨터공학과 재학
※ 관심분야 : 심층 신경망, 영상처리, 컴퓨터 비전

류재홍(Jae-Hung Yoo)

1981년 한양대학교 기계공학과 졸업(공학사) (BE in Mechanical Engineering from Hanyang Univ. in 1981)
MA in Computer Science from Univ. of Detroit in 1986
PhD in Computer Science from Wayne State Univ. in 1993
Joined as a faculty member in the Dept. of Computer Engineering, Yosu Nat. Univ. in 1994
Became a faculty member in the Dept. of Computer Engineering, Chonnam Nat. Univ. in 2006
※(Main research areas: Artificial Neural Networks, Pattern Recognition, Machine Learning, Image Processing and Computer Vision)