Facial Age Estimation by Curriculum Learning

Wei Wang
Graduate School of Fundamental Science and Engineering
Waseda University
Tokyo, Japan
wangwei@toki.waseda.jp

Takaaki Ishikawa
Global Information and Telecommunication Institute
Waseda University
Tokyo, Japan
takaxp@ieee.org

Hiroshi Watanabe
Graduate School of Fundamental Science and Engineering
Waseda University
Tokyo, Japan
hiroshi.watanabe@waseda.jp

***Abstract*—Curriculum learning has been widely used in training neural networks because of its significant improvements in generalization capability. However, it has not been applied to age estimation tasks. In this paper, we incorporate curriculum learning into age estimation. Experimental result of the proposed method on AFAD database for age prediction shows a substantial reduction of the prediction error compared to the traditional training strategy.**

Keywords—age estimation, convolutional neural network, curriculum learning, feature extraction

I. INTRODUCTION

Age estimation is to infer a person's age based on distinct patterns derived from the facial appearance[1]. As real-world applications emerge explosively, such as age invariant person identification, human machine interaction and access control, automatic facial age estimation has become more and more prevalent recently[2]. Although the problem has attracted much attention among the research community, challenges of age estimation are still alive, such as uncontrolled illumination, variant posting, expression and personal life styles[3].

Past works focus on hand-crafted features by applying decision trees and utilizing a heuristic training frame, while current works adopt the architecture of convolutional neural networks, which has proved to be effective on large-scale image classification tasks[4]. Among them the ResNet is a great milestone, which contains novel residual learning framework that can simplify the training of networks that are deeper than previous[5].

In order to gain better performance, we apply curriculum learning strategy when training, which holds that the model can learn better when the dataset is not randomly given but organized in a meaningful order[6]. Compared with the traditional training strategy, curriculum learning adapts the concept of guiding the optimization process, not only contributing faster convergence but also leading the model towards better local minima. The process of curriculum learning, in brief, is distributing the training data into several complex levels, allowing the model to learn simpler data at first, which can avoid introducing noisy early[6].

II. PROPOSED METHOD

The core ideal of curriculum learning is splitting the training data into a number of complex levels, allowing the model to learn simpler data in the beginning, and then increasing the difficulty gradually. Although the specific method can be different, it contains two steps: curriculum design and curriculum training.

A. Curriculum Design

In age estimation tasks, two assumptions are proposed: (a) The complexity of samples within one category is different. (b) Within one feature space that is composed of feature vectors, the closer the feature is to the clustering center, the less difficult the sample is.

In the process of curriculum design, the main mission is to calculate the clustering center by a density-distance clustering method. After that, samples in each category can be split into three subsets, based on the Euclidean distance between the clustering center and each sample.

To calculate the feature vectors, the InceptionResNet-V2 model will be applied. After pre-processing, the sample will be delivered to InceptionResNet-V2 model to obtain the fc_256 layer feature vectors. Then, for each input image, the corresponding feature vector can be calculated. After acquiring all of the feature vectors, the clustering center of each category is accessed by easily computing the average value of feature vectors for each sample. Lastly, by counting the Euclidean distance between clustering center and each sample, three subsets can be attained.

B. Curriculum Training

Curriculum training is the procedure of model optimization, which involves combining current training set with harder samples. Training details are shown in Fig. 1. In the beginning, only the easiest data will be used for the training model, so that the model can learn the basic features within that category. Then, the next difficulty subset will be merged into the current subset. Importantly, in order to decrease the impact of difficulty subset, we need to lower down the learning rate. After that, if the result on validation data does not improve in *m* epochs, the model will start learning the next harder curriculum.

It can be observed that the accuracy of the model does not decrease. On the contrary, the generalization ability of the model is improved and the network can avoid over-fitting. Lastly, three subsets are merged together and the learning rate is reduced in order to achieve fine-tuning.

III. EXPERIMENT

A. Datasets and Preprocessing

The database used in our experiments is AFAD[7], which is the biggest Asian face dataset for age estimation, containing more than 160 thousand facial images and corresponding age labels in the range of 15-40 years. It can be downloaded from *https://github.com/afad- dataset/tarball.*

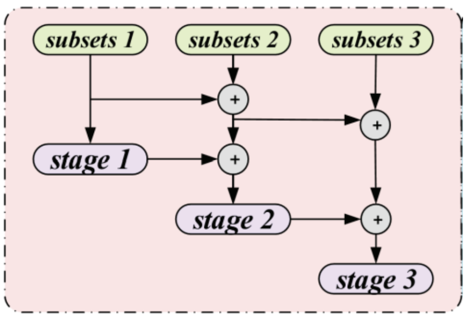

Fig. 1. Training process with designed curriculum.

(1)Input Image (2)Face Detection (3)Face Alignment

Fig. 2. The face image pre-processing pipeline.

In our experiment, the 165501 images are split into 99300 training (TRAIN), 33100 validation (VAL) and 33100 (TEST) images, and all images were resized to 128×128×3 pixels. For the training images, we split them into three subsets with different complexity, and the number of each subsets is the same.

The preprocessing pipeline is conducted for all images including face detection and face alignment, which is shown in Fig. 2. We employ the DLIB model to detect the face based on histogram of oriented gradients feature, and then the detected face is fed into Haar feature-based cascade classifiers to get the 68 facial landmarks, which can get location of the facial regions including mouth, right/left eyebrow, right/left eye, nose and jaw.

B. *Convolutional Neural Network Architectures*

In order to test the ability of curriculum learning for age estimation form AFAD dataset, the modern ResNet-34 CNN architecture is chosen in our experiment, proven to be powerful in a variety of image classification tasks because of the residual learning framework that can simplify the training of networks that are deeper than previous networks[5].

C. *Hardware and Software*

Our experiments are carried out in pytorch 1.5, and the ResNet-34 models are implemented on NVIDIA K40 GPU with 12GB onboard memory.

D. *Evaluation Metrics*

In age estimation, Mean Absolute Error (MAE) is the most widely used error metric. We also use the Root Mean Squared Error (RMSE) to evaluate the performance of our model.

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - x_i)^2}{n}} \quad (2)$$

where y_i is the predicted rank of the i-th test example and x_i is the corresponding ground truth. We train the each model for 250 epochs, and choose the best model through MAE performance base on validation set. Then, the best model is assessed in test set, and the RESE and MAE performances are attained correspondingly. At last, the worst, median and best performances are reported from all reported model result within 250 training epochs.

IV. RESULT AND DISCUSSION

We conducted our experiments on the AFAD dataset for age estimation to observe the performance of the model after utilizing the curriculum learning method. All the implements are based on the ResNet-34 architecture. As is shown in Table I, the performance of the model has improved significantly after applying the curriculum learning strategy, which achieves 3.72 on the metric of MAE compared with the model without curriculum learning(which is 7.65 on average). Meanwhile, the state-of-art method which has applied the COnsistent RAnk Logits (CORAL) framework can reach 3.47 on the same metric[1].

TABLE I. AGE PREDICTION ERRORS ON THE TEST SETS.

Method	**AFAD**		
	Ten experiments	***MAE***	***RMSE***
With Curriculum Learning	Best Result Median Result Worst Result	3.64 3.72 3.75	5.01 5.01 5.06
	AVG ± SD	**3.72 ± 0.02**	**5.03 ± 0.03**
Without Curriculum Learning	Best Result Median Result Worst Result	7.53 7.68 7.74	9.92 10.09 10.20
	AVG ± SD	7.65 ± 0.11	10.07± 0.14
SOTA Model	AVG ± SD	3.47 ± 0.05	4.71 ± 0.06

In terms of the metric of RMSE, the figure reveals that the best performance of the model can hit 5.01 after applying the curriculum learning method, whereas the figure increases to 9.92 when we train the model directly. Meanwhile, the SOTA model can provide 4.71[1], indicating the remaining space that can be improved.

V. CONCLUSIONS

In this paper, we employ the curriculum method to age estimation tasks. We split the dataset into three subsets with different complexity by a density-distance clustering method. The training starts with the easiest subset and harder subset will be merged into the current subset when result on validation data meets the requirement. The experiment result the on AFAD dataset shows significant improvement can be achieved by adapting curriculum learning, indicating the potential of curriculum learning in training neural networks. The statistic also shows that there is still some room for improvement in contrast to SOTA model performance.

REFERENCES

[1] W. Cao, V. Mirjalili, and S. Raschka, “Rank-consistent ordinal regression for neural networks,” arXiv preprint arXiv:1901.07884, Jan. 2019.

[2] T. Dhimar and K. Mistree, "Feature extraction for facial age estimation: A survey," 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, 2016, pp. 2243-2248, Mar.2016.

[3] B. Chen, C. Chen, and W. H. Hsu, “Cross-age reference coding for age invariant face recognition and retrieval,” Proceedings of the European Conference on Computer Vision, Springer. pp. 768-783, 2014.

[4] Y. Fu, G. Guo and T. S. Huang, “Age Synthesis and Estimation via Faces: A Survey,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.32, No.11, pp.1955-1976, Nov. 2010.

[5] S. Chen, C. Zhang, M. Dong, J. Le, and M. Rao, “Using Ranking-CNN for age estimation,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR2017), pp. 5183–5192, Jul. 2017.

[6] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, “Curriculum learning,” Proceedings of the 26th Annual International Conference on Machine Learning (ICML2009), Montreal, pp. 41-48, June 2009.

[7] Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua, “Ordinal regression with multiple output CNN for age estimation,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR2016), pp. 4920–4928, June 2016.

978-1-7281-9802-6/20/$31.00 ©2020 IEEE