Deep Residual Neural Network for Age Classification with Face Image

1st Raya Rahadian
School of Computing
Telkom University
Bandung, Indonesia
raya.rahadian322@gmail.com

2nd Suyanto Suyanto
School of Computing
Telkom University
Bandung, Indonesia
suyanto@telkomuniversity.ac.id

***Abstract*—One of the challenges in computer vision is age classification. There have been many methods used to classify someone age from the image of their faces. Convolutional neural network (CNN) gives a high accuracy but it cannot be used on many layers. Therefore, a residual technique is applied on convolutional neural network then named residual neural network. In this paper, some Residual Networks are applied to develop an age classification with face image using the Adience dataset that has 19,370 face images from 2,284 individuals grouped into eight categories: 0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, and 60-100 years. Three techniques: cyclical learning rate, data augmentation, and transfer learning are observed. Six training scenarios are performed to select the best model. Experimental results show that Resnet34 is the best model with an average F1 score of 0.792 that is achieved by data augmentation, transfer learning, and trained on the image with size 224 x 224 pixels.**

Keywords*—*age classification, cyclical learning rate, face image, multi-class, residual neural network

I. Introduction

One challenge in computer vision is age classification. Age tells us about how long they have lived and effecting various aspect of our life such as social interaction, health, and others. Age classification on face image intends to recognize someone's age from their face image. The information of someone's age is helpful in many things. For example, company needs information about their consumer's ages to help them determine their strategy to sell their products. Various methods are used to classify someone's age but the ability to classify someone's age from face image is still lacking [1, 2]. One of the methods is calculating the distance between features in the face image but that method doesn't work for non-rigid face image likes photos in social media [2].

Convolutional Neural Network (CNN) is one of deep neural network. Convolutional neural network consists of many layers, the example of layers used in convolutional neural network are convolution, pooling, and fully connected layer. One of convolutional neural network model is residual neural network (ResNet). Residual neural network is used in this research because of residual neural network abilities to solve the vanishing gradient problem on convolutional neural network that has many layers. The residual neural network model used in this research is Resnet34.

The proposed method by Gil Levi and Tal Hassner (2015) [2] is convolutional neural network with three layers and two fully connected layers. The convolutional neural network is trained from scratch with 1e-3 learning rate then 1e-4 learning rate after 10,000 iterations. Two methods are used to train the convolutional neural network it is center crop and over-sampling. On center crop, face image that has been cropped with size 227 x 227 on the center of the image is inputted into the convolutional neural network. On over-sampling, the image with size 227 x 227 is cropped five times, one in the middle of the image meanwhile the rest is on around the face. Accuracy is used to evaluate the model and the model got 50.7% accuracy [2].

The purpose of this research is to implement and evaluate Resnet34 performances on classify age from face image on Adience dataset using cyclical learning rate, data augmentation, and transfer learning.

II. Developed System

The block diagram of the developed model is illustrated by Fig. 1. There are three processes in the model.

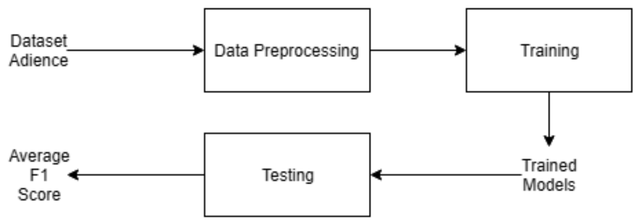

Fig. 1. Block diagram of the developed system

First, a data preprocessing is used to remove the unlabeled data and put the data in their respective class which are 0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, and 60-100 age group. The data is divided into three part which are train, valid, and test data as can be seen on figure 2. Data augmentation is performed on train data as can be seen on Fig. 3. Mean and standard deviation is used to normalize the data, ImageNet mean and standard deviation is used for data that is going to be trained on the model that use transfer learning method. Meanwhile, mean and standard deviation from augmented train data, valid data, and test data is used for data that is going to be trained on the model that doesn't use transfer learning method.

Second, six models are proposed which are three models trained on face image with size 128 x 128 and the last three models on 224 x 224. The transfer learning method is used on one model, data augmentation used on one model, and transfer learning and data augmentation used on one model. Meanwhile, cyclical learning rate is used on all models. The training is performed until the model is overfitting.

Third, testing is performed on test data to get the F1 score from each class to get the average F1 score. Average F1 score is used to evaluate the performance of the models.

A. Dataset

This research uses Adience dataset. Adience dataset created to facilitate age and gender recognition study. The source of Adience dataset is photo albums from Flickr and have various pose, noise, lighting, and other features. Photo albums from Flickr are used to get non-rigid face image. Total face image from Adience dataset is 19,370 face images from 2,284 subjects and labeled on eight class which are age group of 0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, and 60-100 [2], [11].

B. Residual Neural Network

Residual neural network (ResNet) is a neural network that can solve the vanishing gradient problem on deep convolutional neural network. Kaiming He and others create a new concept of neural layer called residual block [3]. Residual block has a main connection and short connection [3]. The gradient can go through short connection so the gradient doesn't have to go through many activation layers, therefore, the gradient value doesn't vanish or zero [3].

C. Transfer learning

Transfer learning is a method to create a neural network model from a model that has been trained on the dataset by transferring the knowledge, therefore, the model isn't trained from scratch. The knowledge on the model that has been trained on dataset is transferred to the new model by training the model fully connected layer on the model using new dataset then fine-tuning is used on the model. In this research, transfer learning is used on the model that has been trained on dataset ImageNet. The model that has been trained on dataset ImageNet can be used in various case [8].

D. Cyclical learning rate

Learning rate is one of hyper-parameter that is used in training convolutional neural network model. Learning rate updates the weights on the model when training the model. To get optimal learning rate training model on many learning rate values is performed but that consume lots of time even using early stopping method, therefore, cyclical learning rate is used. The cyclical learning rate is a method to train a neural network using various learning rate between minimal learning rate (base learning rate) and maximum learning rate (max learning rate) on a cycle [6, 7]. A cycle consists of two step size, the first step size increases the learning rate from base learning rate to max learning rate then on the second step size decreases the learning rate from max learning rate to base learning rate [7]. In this research, only one cycle is used.

E. Data Augmentation

Data augmentation is one of the regulation methods to train a neural network model. Data augmentation manipulates the data by increasing or modify the data in various ways, therefore, lots of data can be created from one data. Data augmentation is used to reduce the overfit on neural network model [4]. There are various ways to do data augmentation one of the popular ways of doing data augmentation is the traditional ways which are using flipping, rotation, cropping, increase or decrease the brightness, and others [4, 5]. In this research data augmentation is done by horizontal flipping, rotation with maximum 45 degrees, zoom in or out 10% from the original image, height and width shift with maximum 10% from the original image, and increase or decrease the brightness.

III. Results and Discussion

On Resnet34 Model that has been trained using transfer learning, eight epochs are performed on fully connected layers with base learning rate 4e-4 and max learning rate 1e-2 then five epochs are performed to train on the whole neural network with base learning rate 4e-7 and max learning rate 1e-5. The training is stopped on five epochs because the model hit overfit on the six epochs. Meanwhile, on Resnet34 that has been trained without using transfer learning, five epochs are performed with base learning rate 4e-5 and max learning rate 1e-3. The training is stopped on five epochs because the model hit overfit on the six epochs.

Table I and Fig. 4 show that the model Resnet34 trained on size 224 x 224 using cyclical learning rate, transfer learning, and data augmentation is the best model with average F1 score of 0.792. The training is performed on PC using Ryzen 5 3600, GTX 1070 8 GB, 16 GB RAM, and SSD. The model with fastest training time with training time 346 seconds is Resnet34 model that has been trained on transfer learning with image size 128 x 128 while model with slowest training time with training time 4149 seconds is Resnet34 model that has been trained on transfer learning and data augmentation with image size 224 x 224.

TABLE I. Average F1 Score and Training Time on the Six Resnet34 Models

Model	Size	Training Time (Seconds)	Average F1 Score
Transfer Learning	128x128	346	0.686
Data Augmentation		772	0.681
Transfer Learning and Data Augmentation		1645	0.744
Transfer Learning	224x224	676	0.709
Data Augmentation		1962	0.720
Transfer Learning and Data Augmentation		4149	0.792

IV. Conclusion

Resnet34 model that has been trained on the image with size 224 x 224 using cyclical learning rate, data augmentation, and transfer learning has the best average F1 score of 0.792. The result shows that cyclical learning rate make choosing learning rate easier than before, speed up training process, and still achieve average F1 score above 0.6 for all models. In the future, expected to involve more hyper-parameter and use bigger architecture. Using data augmentation and transfer learning is better than just using one technique. A higher image size used on training can make model learns better because more features are learned. In the future, more hyper-parameters and use bigger architecture can be investigated.

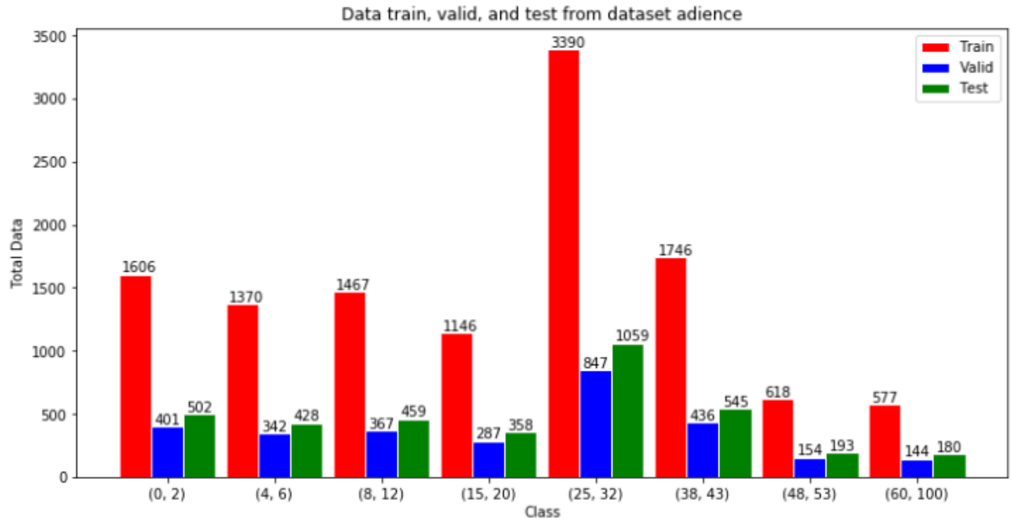

Fig. 2. Distribution of the training set, valididation set, and testing set from the Adience dataset

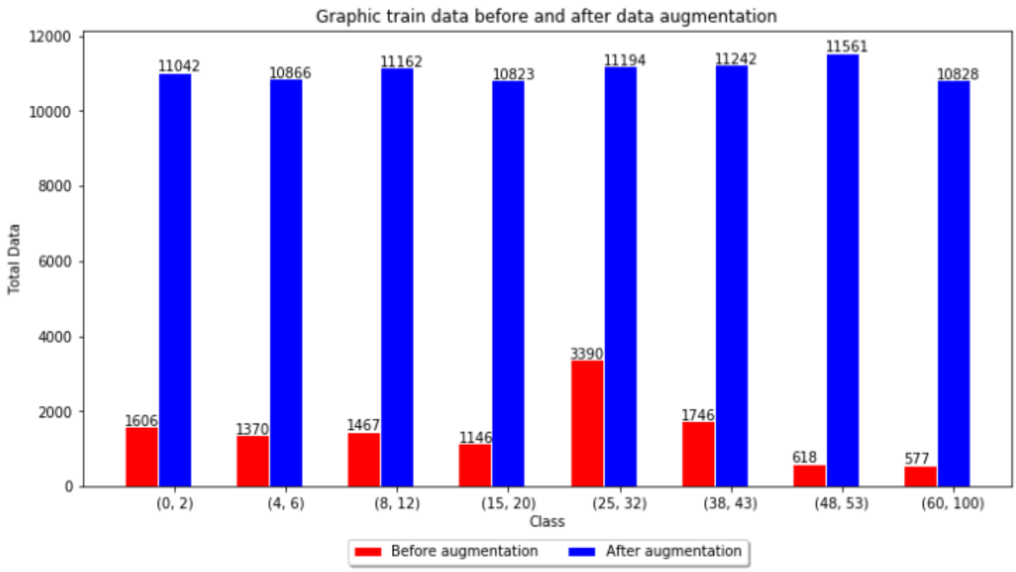

Fig. 3. Training set before and after data augmentation

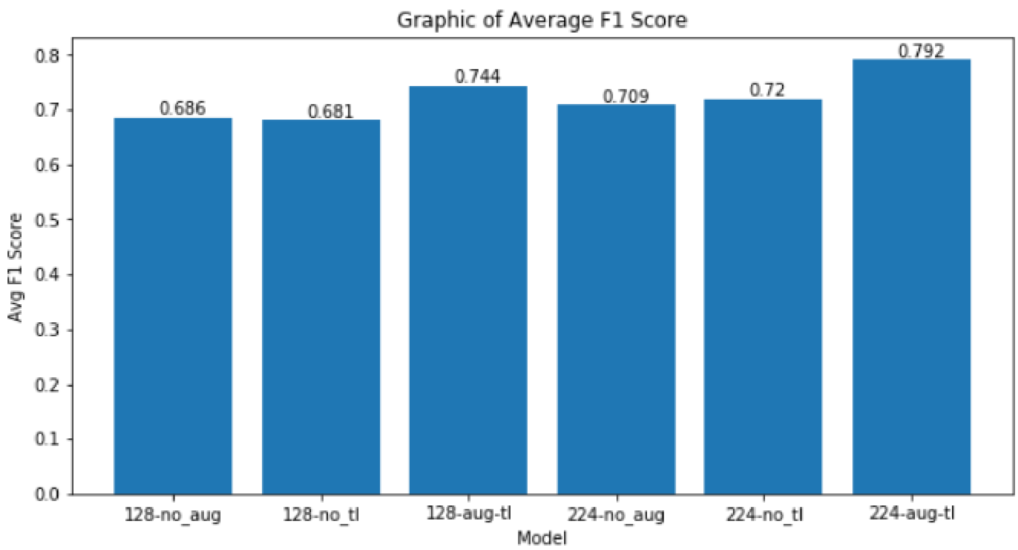

Fig. 4. Average F1 Score on six Resnet34 models

REFERENCES

[1] S. H. Lee, S. Hosseini, H. J. Kwon, J. Moon, H. I. Koo, and N. I. Cho, “Age and gender estimation using deep residual learning network,” in 2018 International Workshop on Advanced Image Technology (IWAIT), Chiang Mai, 2018, pp. 1–3.

[2] G. Levi and T. Hassncer, “Age and gender classification using convolutional neural networks,” in 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Boston, MA, USA, 2015, pp. 34–42.

[3] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” arXiv:1512.03385 [cs], Dec. 2015.

[4] J. Wang, S. Mall, and L. Perez, “The Effectiveness of Data Augmentation in Image Classification using Deep Learning,” Stanford University research report, Dec. 2017.

[5] A. Mikolajczyk and M. Grochowski, “Data augmentation for improving deep learning in image classification problem,” in 2018 International Interdisciplinary PhD Workshop (IIPhDW), Swinoujście, 2018, pp. 117–122.

[6] L. N. Smith, “A disciplined approach to neural network hyper-parameters: Part 1 -- learning rate, batch size, momentum, and weight decay,” US Naval Research Laboratory Technical Report 5510-026, Mar. 2018.

[7] L. N. Smith, "Cyclical Learning Rates for Training Neural Networks," 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, 2017, pp. 464-472.

[8] S. Kornblith, J. Shlens, and Q. V. Le, “Do Better ImageNet Models Transfer Better?,” CVPR 2019 Oral, Jan. 2019.

[9] L. Siebert, “Pre-trained Deep Convolutional Neural Networks for Face Recognition,” University of Groningen, The Netherlands, 2018.

[10] R. Rothe, R. Timofte, and L. V. Gool, “DEX: Deep EXpectation of Apparent Age from a Single Image,” in 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), Santiago, Chile, 2015, pp. 252–257.

[11] E. Eidinger, R. Enbar, and T. Hassner, “Age and Gender Estimation of Unfiltered Faces,” IEEE Trans.Inform.Forensic Secur., vol. 9, no. 12, pp. 2170–2179, Dec. 2014.

[12] J. Howards and others, “Fastai Library Documentation,” 2018. [Online]. Available: https://docs.fast.ai/. [Accessed: 20-Jan-2019].