Age Group and Gender Estimation

Team members:

Yale Li(A20502748)
Zhong Zheng(A20436842)

Contributions:

Zhong Zheng

- Data loading, cleaning, pre-processing
- Age group, gender model coding
- Train, validate, test age group, gender models.
- Record use of program video
- Write Solution, Implementation, Results and Discussion and Use of program parts of the report.

Yale Li

- Initialize the project and online docs, analyze the requirements of the project, organize meetings.
- Data presentation and analysis, model visualization.
- The abstract, problem statement, and visualization part writing, report formatting, and proofreading.

GitHub: https://github.com/Yale-Li/age_gender_estimation

Abstract

Automatically predicting age groups and gender from face images is an important and challenging task in many real-world applications. This difficulty is alleviated to some degree through convolutional neural networks (CNN) for their powerful feature representation. In this project, we constructed a CNN-based artificial neural network called residual networks of residual networks (RoR) which was proposed by Zhang et al.[1] in the paper Age Group and Gender Estimation in the Wild With Deep RoR Architecture to estimate the gender and age group. According to the paper, RoR achieves better performance for age group and gender classification than other CNN architectures. We trained the model on UTKFace[2], a large-scale face dataset that consists of over 20,000 face images. And finally, we got a superb result for the prediction of age group and gender, that both parts have an accuracy rate of more than 80%.

1.Introduction

Age and gender, two of the key facial attributes, play very fundamental roles in social interactions, making age and gender estimation from a single face image an important task in intelligent applications, such as access control, human-computer interaction, law enforcement, marketing intelligence, and visual surveillance, etc [1].

1.1 Previous works

In the past twenty years, human age and gender estimation from face images have benefited tremendously from the evolutionary development in facial analysis. Early methods for age estimation were based on geometric features calculating ratios between different measurements of facial features

[8]. After 2007, most existing methods used manually-designed features in this field, such as Gabor, LBP, SFP, and BIF [9]. Based on these manually-designed features, regression and classification methods are used to predict the age or gender of face images. SVM-based methods are used for age group and gender classification. For Regression, linear regression, SVR, PLS, and CCA are the most popular methods for accurate age prediction [10].

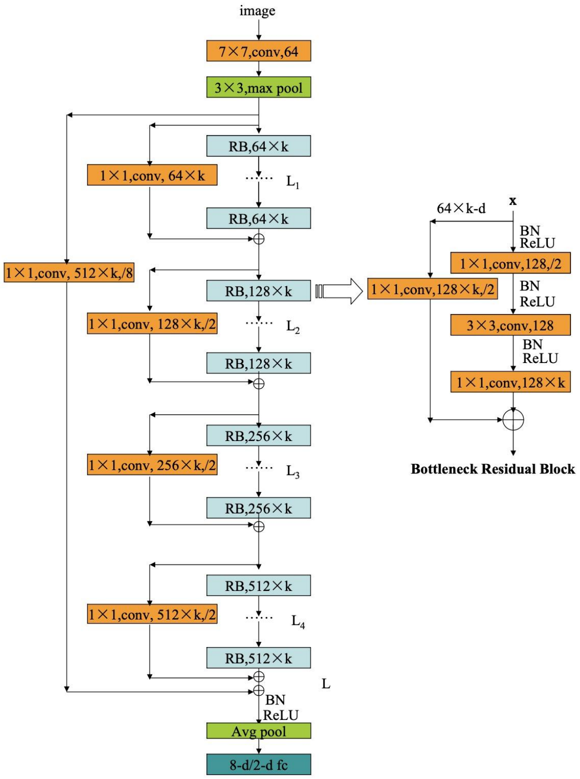
Deep learning, especially deep Convolutional Neural Networks (CNN), has proven itself to be a strong competitor to the more sophisticated and highly tuned methods [7]. Recent research on CNN showed that CNN models can learn a compact and discriminative feature representation when the size of training data is sufficiently large, so an increasing number of researchers start to use CNN for age and gender estimation [1].

The optimization ability of neural networks is critical to the performance of age and gender estimation while existing CNNs designed for age and gender estimation only have several layers, which severely limits the development of age and gender estimation [1].

It is widely acknowledged that the performance of CNN-based age and gender estimation relies heavily on the optimization ability of the CNN architecture, where deeper and deeper CNNs have been constructed. From 5-conv+3-fc AlexNet to the 16-conv+3-fc VGG networks and 21-conv+1-fc GoogleNet, then to thousand-layer ResNets, both the accuracy and depth of CNNs were promptly increased. In order to dig into the optimization ability of the residual network's family, Zhang et al. [11] proposed Residual Networks of Residual Networks architecture (RoR), which added shortcuts level by level based on residual networks, and achieved the state-of-the-art results on low-resolution image data sets such as CIFAR-10, CIFAR-100 and SVHN at that time [1].

1.2 Methods

The paper proposed a new CNN-based method for age group and gender estimation leveraging Residual Networks of Residual Networks (RoR), which exhibits better optimization ability for age group and gender classification than other CNN architectures[1]. Predicting the exact age of an image would be very challenging. The paper proposed to predict the age group instead of predicting the exact age. Age group selection is described in the next section.



Pre-RoR with Bottleneck Blocks

We follow the general idea of this proposed solution to construct deep neural networks and train our age group and gender prediction models using cropped UTKFace dataset[2].

1.3 Results

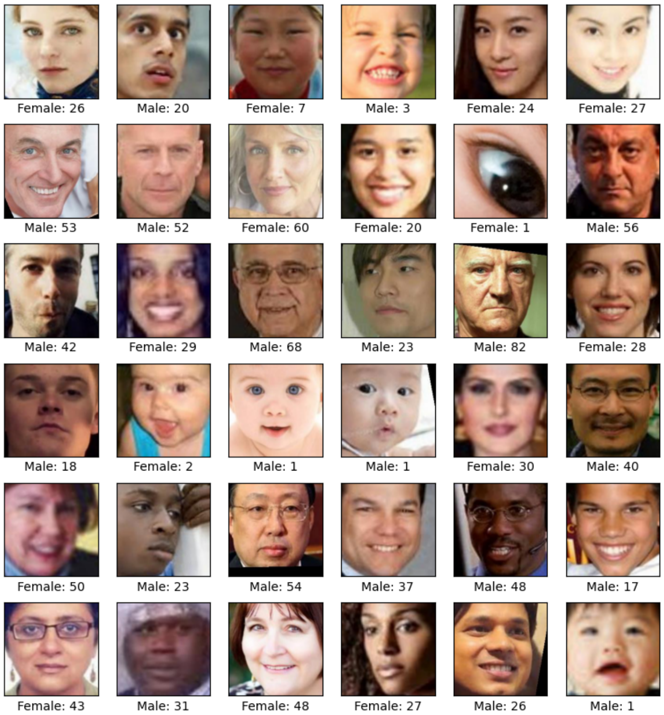
we got a superb result for the prediction of age group and gender, that both parts have an accuracy rate of more than 80%.

2. Experiments

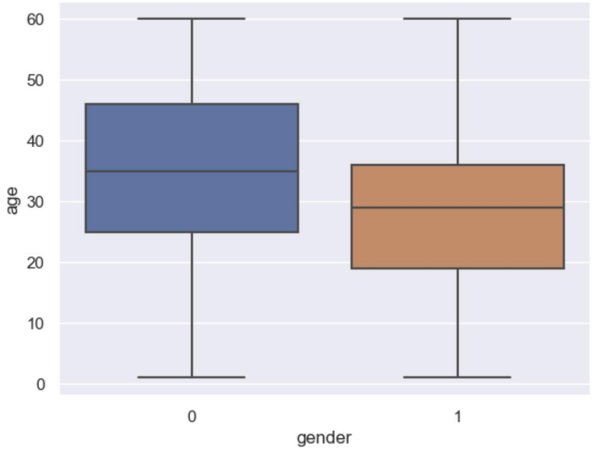
2.1 Data pre-processing

The file name of each image in the UTKFace dataset contains the following information: age, gender, and race. In our case, we only need to parse the age and gender information of each image.

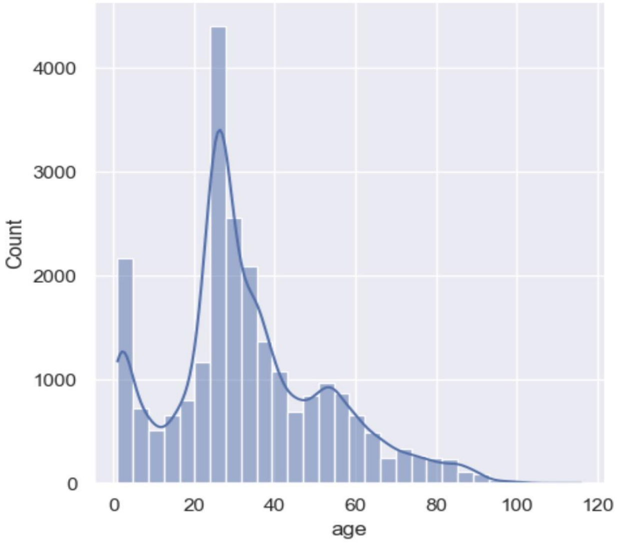
We use the parse-fileName function provided by this article[6]. There are a total of 23708 images in the dataset. We use 14000 images to train, 6000 images to validate, and the rest for testing.



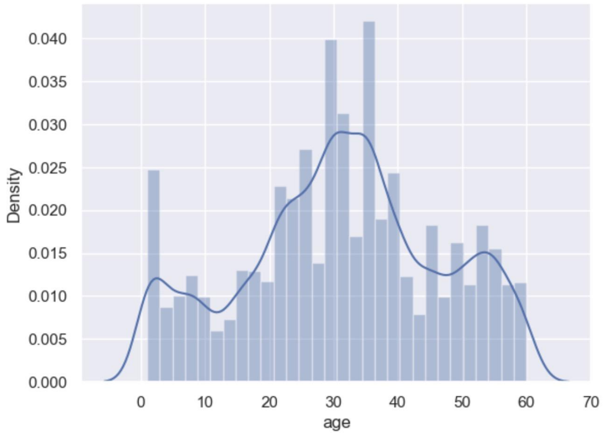
We observed that each gender is almost equally distributed, so it is ok to use the original dataset for training.



However, age distribution has a very high variance. That will cause the model to perform well at some particular age groups, in our case, the model might be over-performing in classifying the age group of 0-4 and 24-28. Also, there is a few age data distributed over age 60.



To tackle this problem, we decided to randomly select part of these two age groups: 0-4 and 24-28 instead of using all of them for training. Also, we choose not to use the age over 60(Below the right graph shows the final distribution). After data selection, there are 16500 images remaining for age group classification. We use 9000 images for training, 3000 images for validation, and the rest for testing.



2.2 Image modification

We resize the image to (32, 32). The reason we resize the image to a smaller size is that a larger image size would consume training time exponentially, and we don't have either enough time budget or a powerful GPU to achieve it.

2.3 Choosing age group

Predicting the exact age from an image is very challenging, instead, the paper chooses to predict the age group. For example, predicting that one's age is from 20-25. In our case, we choose the age group gap as 5. For instance, 0-5, 6-10, so on and so forth. Therefore, we convert the regression problem into a classification problem. Since we only train the data that ranges from 0 to 60 ages, there are a total of 12 age groups.

2.4 Model design

2.4.1 RoR (Residual Networks of Residual Networks)

We follow the model architecture design that the paper proposes[1] shown in the previous section. First, we have one 7x7 convolution layer with 64 filters followed by a 3x3 max pooling. Then we start with a nested residual module. The outer residual block is to copy the identical output by the first 7x7 convolution layer to add it to the output of the multiple inner residual blocks. The proposed model has 4 inner residual blocks. For convolution layers in each residual block, we use 64, 128, 256, and 512 filters sequentially. Inside each bottleneck residual block, we have 3 convolution layers, and the filter size is 1x1, 3x3, and 1x1 respectively. Before each convolution layer, we have a batch normalization layer, and we use the RELU activation for the convolution layers. The nested residual block is followed by an average pooling layer and one dense layer with 256 output units. For gender classification, the output layer uses sigmoid activation and the output unit is 1; for the age group classification, the output layer uses softmax activation and the output unit is the number of age groups that can be altered, in our case, it is 12. We choose binary_crossentropy loss for gender classification, and categorical_crossentropy loss for age group classification, and choose accuracy metric for both classification problems.

2.4.1 CNN (Convolutional Neural Networks)

The CNN architecture is similar to RoR architecture, only without all the shortcuts and has fewer convolutional layers.

2.5 Training, validation, and testing

For training the gender prediction model, we use 14000 images and 3500 validation images for 20 episodes, and the rest for testing. For training the age group prediction model, we use 9000 images for training, with 3000 images for validation for 50 episodes, and the rest for testing. We also perform parameters searching, and found that when we have a batch size of 64 the models take fewer epochs to converge, and perform better. For the optimizer, we use Adam with a learning rate of 0.0001 for both models.

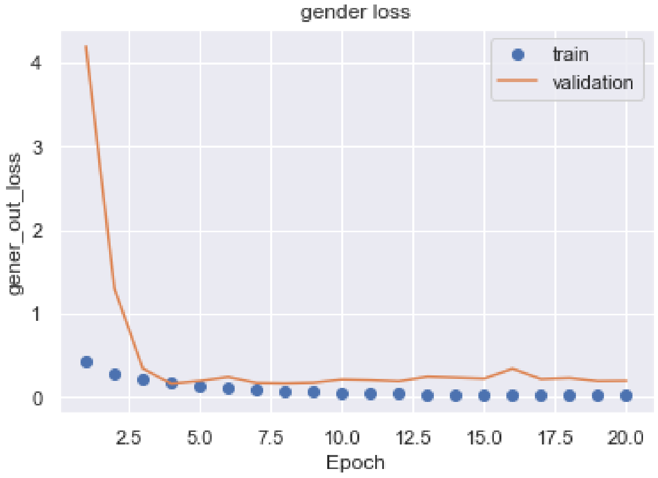
3. Results

Summary of the two models' performance.

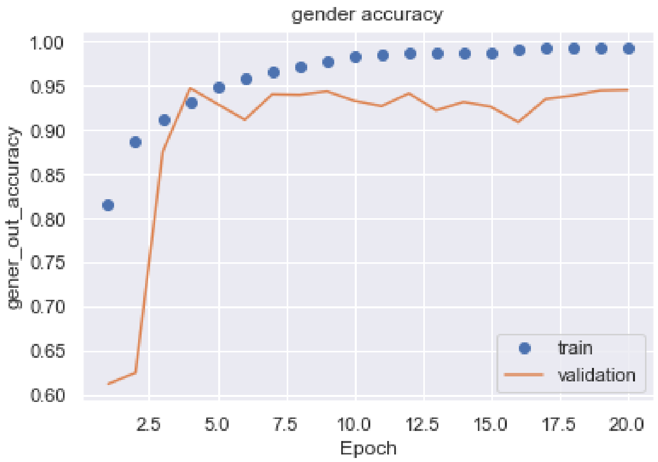
Accuracies	Gender	Age
CNN	85%	21%
RoR	84.9%	85.86%

3.1 CNN (Convolutional Neural Networks)

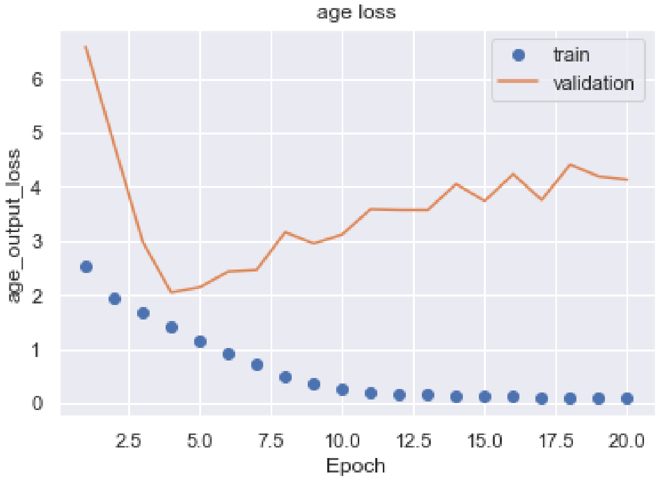
3.1.1 Gender loss



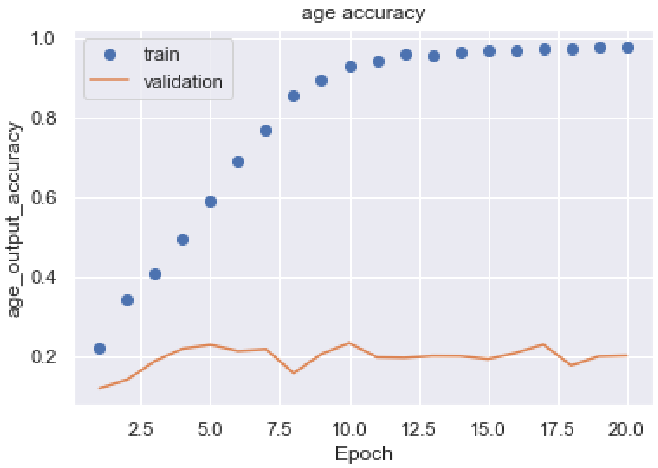
3.1.2 Gender accuracy



3.1.3 Age loss

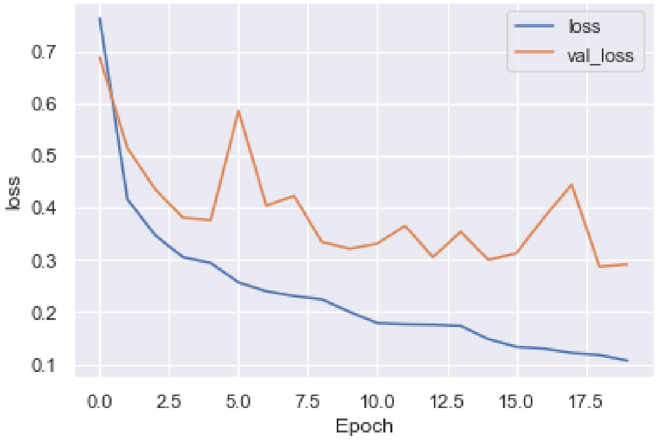


3.1.4 Age accuracy

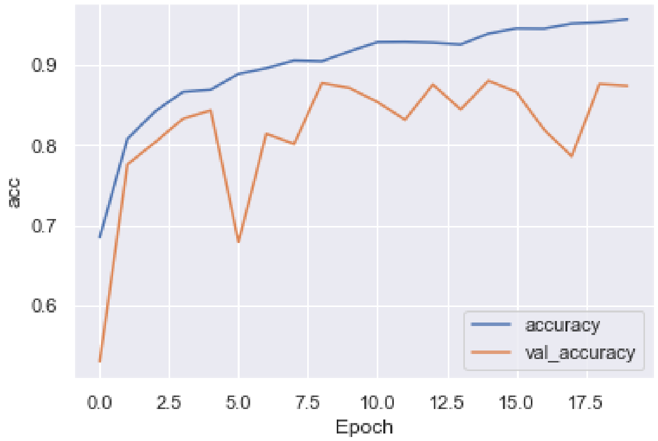


3.2 RoR (Residual Networks of Residual Networks)

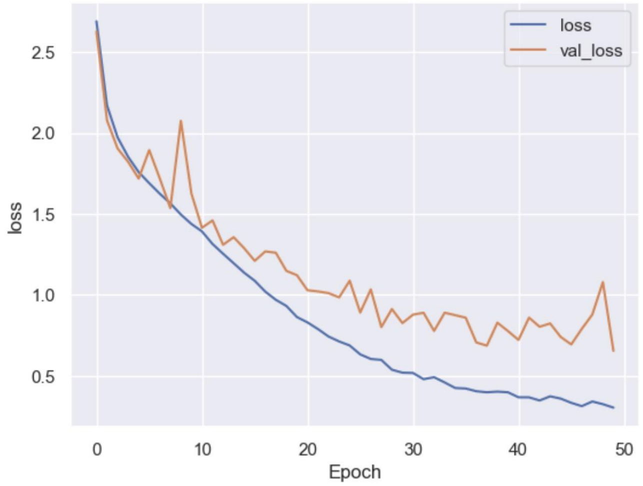
3.2.1 Gender loss



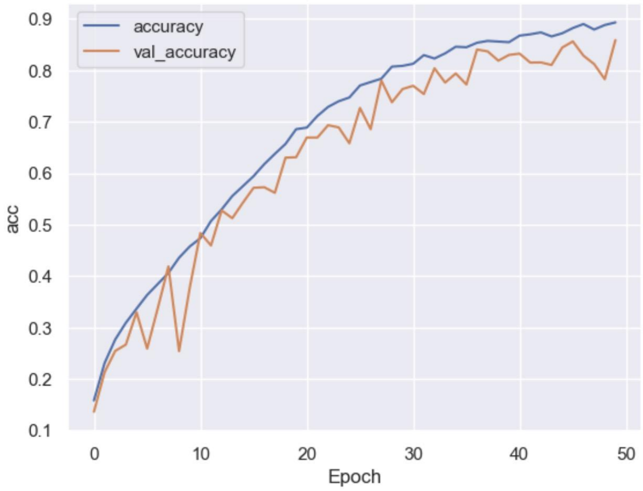
3.2.2 Gender accuracy



3.2.3 Age loss

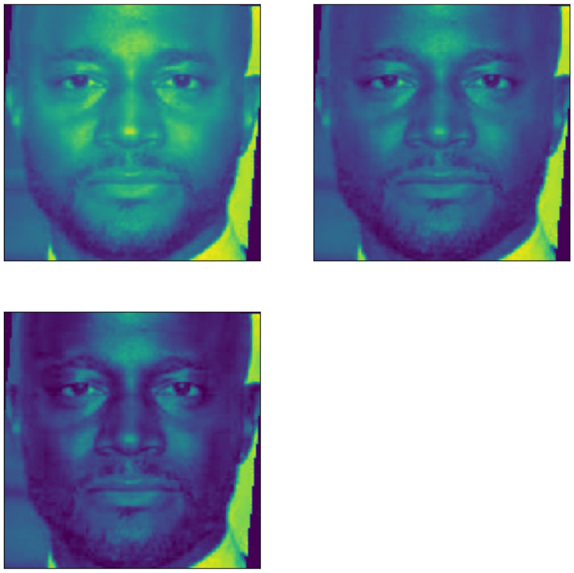


3.2.4 Age accuracy

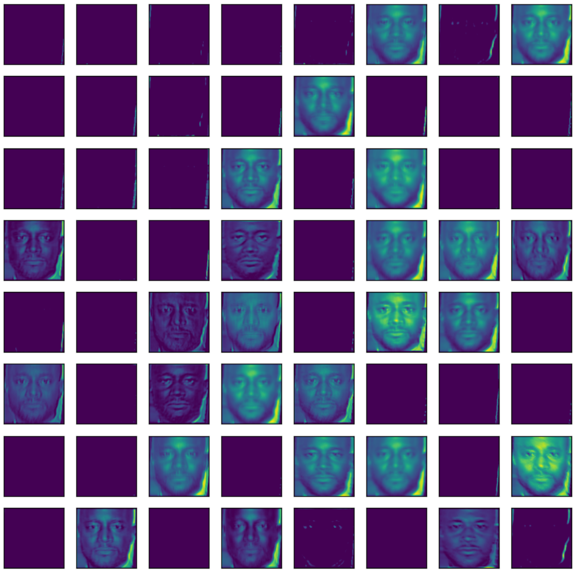


3.3 Visualize RoR model

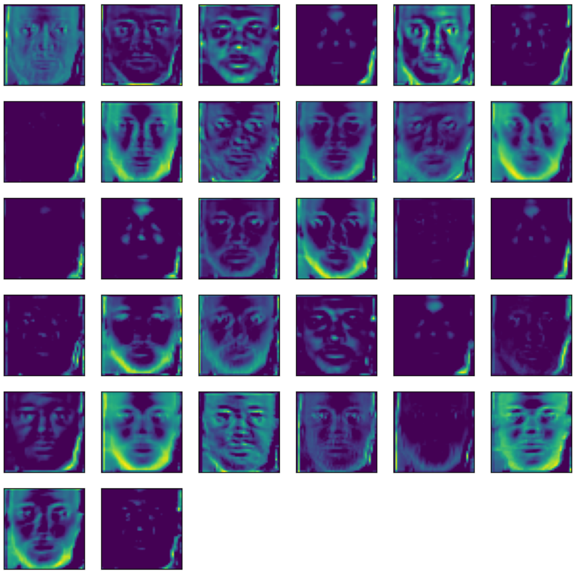
Visualization1: input layer



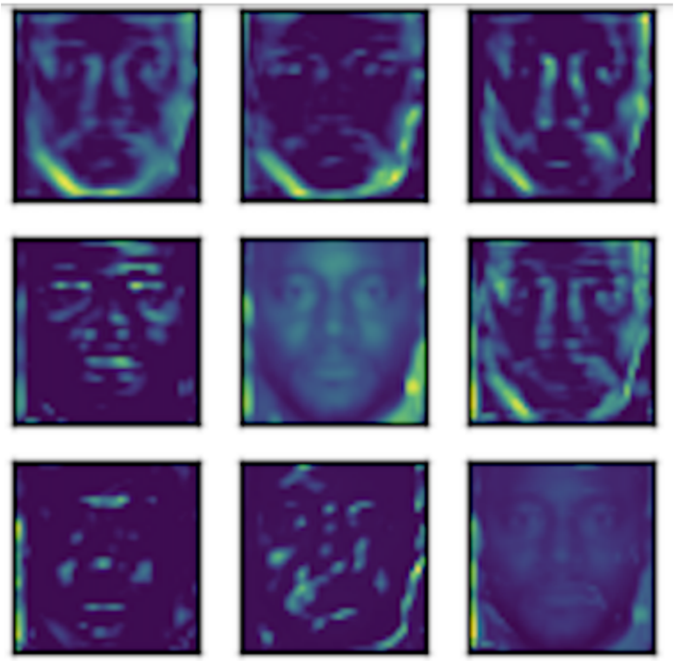
Visualization 2: layer 3



Visualization 3: layer 10



Visualization 4: part of layer 33



4. Conclusions

For gender prediction, CNN and RoR Architecture have similar results with an accuracy of around 85%. But for age prediction, the CNN model overfits quickly after several epochs of training and got a poor accuracy of around 20%. RoR has an edge that can continuously reduce the loss and increase the accuracy.

From the intermediate images above, especially the last picture, we can observe that even in the last few layers there are some concrete faces around with other abstract faces. It's contributed by the short paths in the RoR Architecture. And thanks to this feature we obtain a superb performance much better than a simple CNN model.

5. References

1. K. Zhang et al., "Age Group and Gender Estimation in the Wild With Deep RoR Architecture," in IEEE Access, vol. 5, pp. 22492-22503, 2017, doi: 10.1109/ACCESS.2017.2761849.
2. “UTKFace | Large Scale Face Dataset.” UTKFace, https://susanqq.github.io/UTKFace/. Accessed 28 Oct. 2022.
3. “Face Image Project.” Adience, https://talhassner.github.io/home/projects/Adience/Adience-main.html. Accessed 31 Oct. 2022.
4. Tal Hassner, Shai Harel*, Eran Paz* and Roee Enbar, Effective Face Frontalization in Unconstrained Images, IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Boston, June 2015.
5. Keras Team. “About Keras.” Keras: The Python Deep Learning API, https://keras.io/about/. Accessed 27 Oct. 2022.
6. eward96. “Age and Gender Prediction on UTKFace | Kaggle.” Kaggle: Your Machine Learning and Data Science Community, Kaggle, 14 Oct. 2022, https://www.kaggle.com/code/eward96/age-and-gender-prediction-on-utkface.
7. Razavian, A.S., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN Features Off-the-Shelf: An Astounding Baseline for Recognition. 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, 512-519.
8. Kwon, Y. H., & Lobo, N. D. V. (1999). Age Classification from Facial Images. Computer Vision and Image Understanding, 74(1), 1–21. https://doi.org/10.1006/cviu.1997.0549.
9. Guo, G., Guowang Mu, Fu, Y., & Huang, T. S. (2009). Human age estimation using bio-inspired features. 2009 IEEE Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/cvpr.2009.5206681.
10. Guodong Guo, Yun Fu, Dyer, C., & Huang, T. (2008). Image-Based Human Age Estimation by Manifold Learning and Locally Adjusted Robust Regression. IEEE Transactions on Image Processing, 17(7), 1178–1188. https://doi.org/10.1109/tip.2008.924280.
11. Zhang, Ke, et al. "Residual networks of residual networks: Multilevel residual networks." IEEE Transactions on Circuits and Systems for Video Technology 28.6 (2017): 1303-1314.