Multi-Task Learning Based Age Detection for Image

Xinli Gu
NYU CDS
xg588
xg588@nyu.edu

Xi Yang
NYU CDS
xy2122
xy2122@nyu.edu

Abstract

In this work, we explore ways to accurately predict the age of a person from their portraits. We regard it as a classification problem first since the target age is divided into several groups. We also explore the possibility of regarding this problem as regression as we increase the granularity. After building the age detection model with just one label, two more labels (gender, race) are added to build the multi-task learning model. Multitask learning is believed to provide more information for the target task than the single-task leaning model. We learn all three tasks jointly to improve prediction accuracy. We used both single source and multiple source dataset to examine the possibility of multi-source multi-task leaning.

1 Project Description

1.1 Introduction

In recent years, artificial intelligence (AI) has been exploding thanks to breakthroughs in the field of machine learning and data science. In AI field, natural language processing and computer vision have breakthrough result in most tasks after the creation of BERT and CNN. Along with the explosion of computer vision, automatic human facial recognition has become an active research area that plays a key role in analyzing facial characteristics and human behaviors. Facial recognition and face detection is currently an important part in people's daily life, therefore, the need for accurately capturing facial features and expressions accelerates the development of facial recognition and detection.

Several papers have addressed the age classification problem. For example, Chen et al. (2014) and Eidinger et al. (2014) explored biological or real age estimation for single facial image. However, accurately predicting age is still a hard problem since increasing granularity of the age classes intuitively will make the classification harder and reduce the prediction accuracy. If we split the age range into even more groups, one can imagine how a classification problem will become a regression problem. When the number of groups is equal to the number of the data, the model is thought as a regression prediction. Our goal is to improve the prediction accuracy with multi-task learning for more age groups.

In this work, despite the target task age estimation, we jointly study different human facial analysis tasks including race prediction and gender recognition. All of three tasks use facial images as input. In race prediction task, we must detect if the people in a given image are while, black, asian, or indian. We will also indicate who are males and who are females among them in gender classification task. Instead of regarding these three tasks as seperating problems, we try to utilize multitask learning technic to jointly learn the tasks simultaneously in order to boost the performance of each individual task, especially the age classification task.

1.2 Multitask Learning

If training dataset is not large enough, treating the three tasks as seperate problems may lead to poor accuracy and difficulties in training models. In addition, as is shown in past researches, facial analytics tasks often share similar features or characteristics of human faces, so it's very natural to build a shared network with different branches to jointly learn the representations for facial tasks.

The concept of Multi-task learning was first mentioned in 1998, which is an extension of transfer learning, transfering the knowledge of other related tasks to the target task and learning the common features. Recently, Kaiser et al. (2017) proposed a huge model to simultaneously learn different tasks in NLP and computer vision and achieve promising results. In face detection area, Rothe et al. (2015)

proposed a multitask learning model to jointly learn apparent age classification and gender recognition from images. Therefore, multitask learning is an active researching area and by examine the results from multitask learning, we can have deeper understanding of human facial features and possibly reduce the computing resources.

1.3 Technical Challenge

Besides the single source dataset training, we also like to explore the possibility of using multiple data sources. Using multi-source datasets is big challenge for us. When people do multi-task learning for age detection, gender is usually added as another task. Age-Gender Multi-task learning is built on a single source dataset, which means the datasets have two labels: age and gender. However, in this work, we add race classification as the third task and we will test the difference between using race in the same source and using race information in a different source. The result is not clear before the experiment, and the technical setup is also a challenge for us.

2 Approach

In this work, We experiment whether multi-task learning will improve the age classification problem with increasing age group granularity.

2.1 Data Preprocessing

In order to test the model capacity of handling various age groups, we create three different settings for age groups. First we separate the age range 1-80 into three groups, 1-26, 27-52 and 53-80. Then, we create five groups, 1-11, 12-23, 24-39, 40-55 and 56-80. In the last, we run the model on ten age groups, 1-8, 9-16, 17-24, 25-32, 33-40, 41-48, 49-56, 57-64, 65-72 and 73-80. For race dataset, the labels are white, black, asian, indian and others, and for gender dataset, the labels are male and female.

All the images in three datasets are portraits, but we would like to only work on the face area of the portraits. Therefore, for the training data, we use the cropped dataset. For testing, we apply Multi-Task Cascaded Convolutional Neural Network (MTCNN) (Zhang et al., 2016) on the images to detect faces. MTCNN is a state-of-the-art face detection and alignment model, this framework adopts a cascaded structure with three stages of carefully designed deep convolutional networks that predict face and landmark location. Due to the relatively small data size, we also adopt the standard data augmentation technic.

2.2 Modeling Pipeline

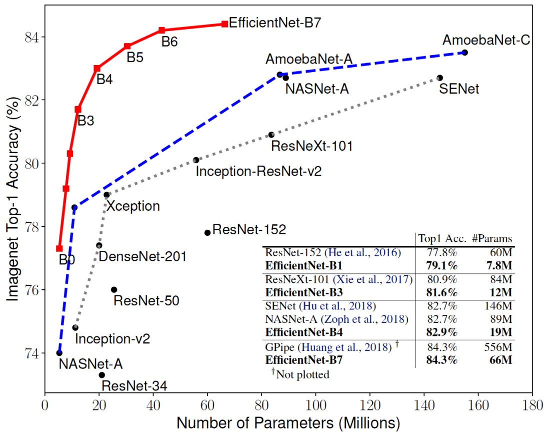
We build a framework of multitask learning based on CNN to simultaneously learn common features of age classification, gender recognition and race prediction. As inspired by Sang et al. (2017), in our model, each task takes in the input from either the same data source or different data sources. Then the inputs are passed into a shared CNN block in which the model learns shared features. At the end of the network, these three tasks are separated into three fully connected branches with task specific losses. Then we combine the losses together to form an overall loss and use the back propagation algorithm to minimize the total loss.

We propose models for three different problems, only predicting age, simultaneously predicting age and gender and simultaneously predicting age, gender and race.

Figure 1 shows the solution diagram for this framework.

2.3 Efficient Net

The main part for the multi-task learning framework is the shared CNN block, which is used to learn the common features from the three tasks. The CNN based model we adopt is EfficientNet which was proposed by Tan and Le (2019).



	Top1 Acc.	#Params
ResNet-152 (He et al., 2016)	77.8%	60M
EfficientNet-B1	**79.1%**	**7.8M**
ResNeXt-101 (Xie et al., 2017)	80.9%	84M
EfficientNet-B3	**81.6%**	**12M**
SENet (Hu et al., 2018)	82.7%	146M
NASNet-A (Zoph et al., 2018)	82.7%	89M
EfficientNet-B4	**82.9%**	**19M**
GPipe (Huang et al., 2018) †	84.3%	556M
EfficientNet-B7	**84.3%**	**66M**

†Not plotted

Figure 2: Efficient Net

As is indicated by the name of the network, EfficientNet can scale up very fast and retain a high accuracy. To train the EfficientNet, we need to optimize both the accuracy and FLOPS. From figure 2 we can see that EfficientNet-B7 achieves 84.3%

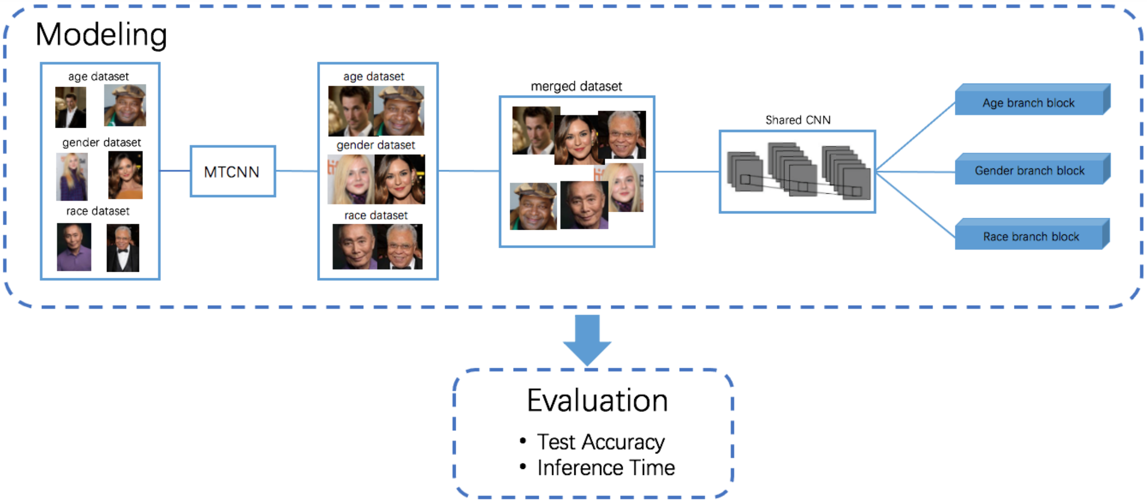


Figure 1: Solution Diagram

top-1 accuracy but is 8.4x smaller and 6.1x faster than GPipe. EfficientNet-B1 is 7.6x smaller and 5.7x faster than ResNet-152.

In our experiment, we tried EfficientNet-B4 to B7 to test the model capabilities and the scaling effect.

3 Implementation Details

Models are tranined in the Google Colab with P100. The framework of the model is Tensorflow2.4.0. There are two datasets for training. One is IMDB-WIKI Dataset. Another one is UTKFace Dataset. Each model is trained 40 epochs, with batch size of 256. There are three main steps in this experiment. First, different models are trained on IMDB-WIKI dataset to compare the model performance between single-task and two-task learning. Second, different models are trained on UTKFaca dataset to test whether adding more tasks will improve target task learning performance. Third, models are trained on multi-source datasets to test the feasiability of multi-source datasets.

4 Experimental Results

The final model helps to predict age, gender and race for people in images. Three demos are shown below, with results for multi-task learning based on 3 age classes, 5 age classes and 10 age classes, respectively.

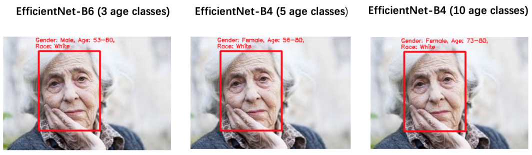

EfficientNet-B6 (3 age classes)
EfficientNet-B4 (5 age classes)
EfficientNet-B4 (10 age classes)
Gender: Male, Age: 53-80, Race: White
Gender: Female, Age: 56-80, Race: White
Gender: Female, Age: 73-80, Race: White

Figure 3: experiment demo

5 Experimental Evaluation

5.1 Evaluation for IMDB-WIKI Dataset

First, all the models are trained on IMDB-WIKI dataset. There are total 20,000 images, and each image has age and gender labels. After conducting train-validation split, 16,000 images are used for training and 4,000 images are used for validation.

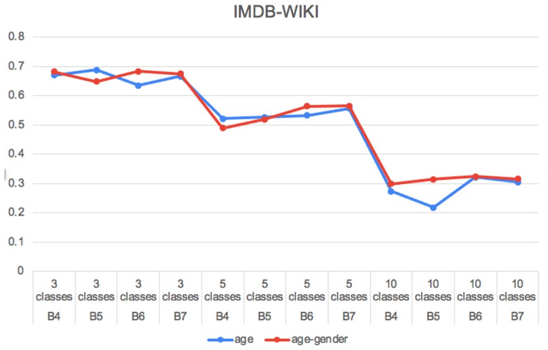


Figure 4: validation accuracy for age prediction with different models and different age classes based on IMDB-WIKI dataset

Above figure shows the performance of models. EfficientNet-B4, EfficientNet-B5, EfficientNet-B6 and EfficientNet-B7 have similar performance in age prediction. EfficientNet-B7 is more complex than other models, however, the model performance is similar to others. It is possible that the age prediction task is easy, complicated models are not needed for this task. Also, the training data is limited, which cannot provide enough information for complicated models to learn. Thus, no matter how complicated the model is, it will have the same performance as basic models.

At the same time, models with age single-task learning have similar performance with age-gender two-task learning . It is possible that gender prediction task is not related to age prediction task. Moreover, it is also possible that the gender prediction task is too easy, thus it cannot provide age prediction task with some useful information. Thus gender prediction task not benefit age prediction task a lot.

It is obvious that increasing age granularity will decrease model performance. All models perform best in 3 age classes, and perform worst in 10 age classes. The data for training is limited. As increasing age granularity, the model becomes more and more complex. In order to maintain the model accuracy, more data should be provided for training. At the same time, with limited data, the number of images in each class decreases as we increase the number of classes. As shown in the figure below, there are less than 1000 images in age classes 1-8, 9-16 and 65-72, which leads to the model learning little from them. Also, when increasing age granularity, the data is more and more imbalanced, which also exacerbates the deterioration of model performance.

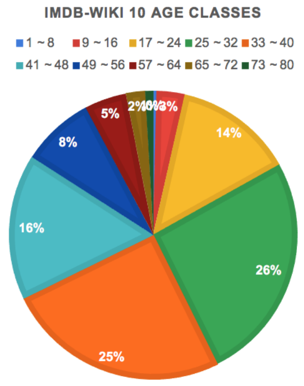


Figure 5: data distribution for 10 age classes based on IMDB-WIKI dataset

5.2 Evaluation for UKTFace Dataset

Second, all the models are trained on UTkFace dataset. There are total 20,000 images, and each image has age, gender and race labels. After conducting train-validation split, 16,000 images are used for training and 4,000 images are used for validation.

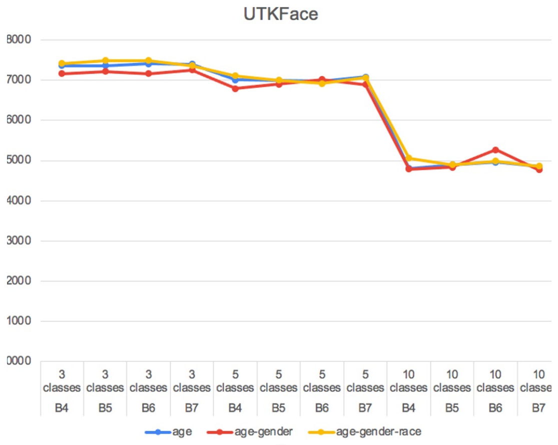


Figure 6: validation accuracy for age prediction with different models and different age classes based on UTKFace dataset

It is seen that age-gender-race multi-task learning performs better than age single-task learning, and age-gender two-task learning is the worst. Adding a race prediction task helps improve the target task learning. It is possible that the race prediction task is quite related to the age prediction task. If two tasks are conducted at the same time, those tasks can share information with each other, thus improving model generalization, as well as the model performance. The reason why age-gender multi-task learning is the worst is the same as before.

5.3 Evaluation for Comparison Between UKTFace Dataset And IMDB-WIKI Dataset

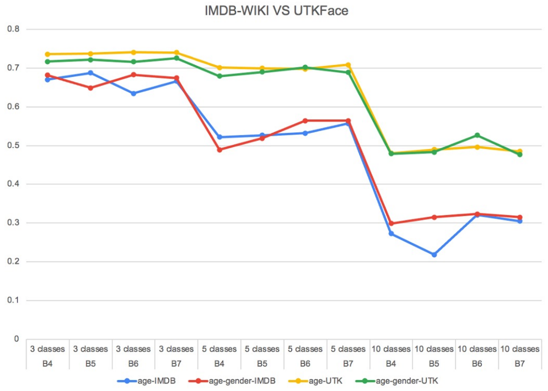


Figure 7: validation accuracy for age prediction with different models and different age classes based on two datasets

It is clear that models perform better in UTKFace dataset no matter how many age classes. From the below data distribution, It is seen that data is more balanced in UTKFace dataset. In IMDB-WIKI dataset, half of data are in the 24-39 age class and only 1% data are in the 1-13 age class. Without doubt, imbalanced data result in bad performance.

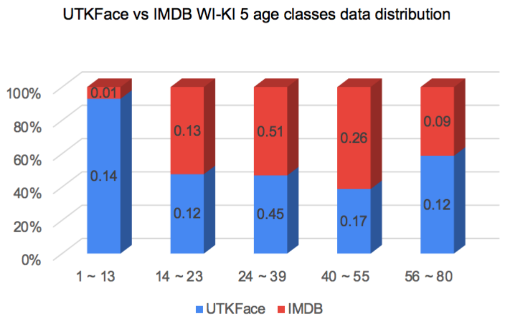


Figure 8: 5 age classes data distribution based on UTK-Faca dataset and IMDB-WIKI dataset

5.4 Evaluation for Multi-source Datasets

In practice it is difficult to find a large dataset with multiple labels. multi-source datasets are a good way to solve this problem. In the project, experiments are also conducted to test the model performance on multi-source datasets. Data from IMDB-WIKI dataset and UTKFace dataset are combined together, which includes 20,000 images from IMDB WI-KI with age and gender labels, and 20,000 images from UTKFace with only race label. Models performance is shown below. The model trained on multi-source performs the worst.

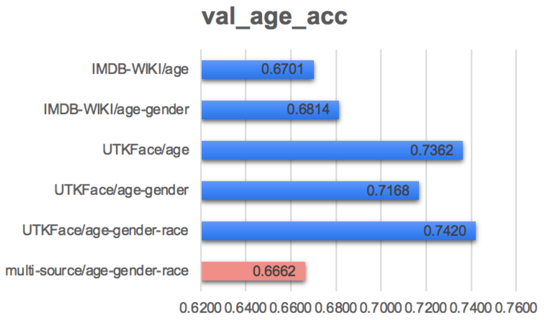


Figure 9: age prediction validation accuracy for different datasets and different tasks with 3 age classes

6 Conclusion

Multi-task learning has many advantages over single-task learning, such as reducing memory cost, increasing learning efficiency and improving model generalization. However, it is not suitable for all cases. When tasks are not related enough, the multi-task learning may not help target task learning, even worse, it may decrease target task performance. In conclusion, adding related tasks benefits target task learning. On the contrary, the performance of multi-task learning with irreverent tasks is no better than single-task learning.

Data with multiple labels plays a key role in multi-task learning. In reality, datasets can be put together to provide multi-labels. However, compared with single-source dataset with multiple labels, models perform worse on multi-source datasets. At the same time, it is essential to make sure each data is balanced on each task when conducting multi-task learning. If the data is balanced on one task and imbalanced on others, the performance of model is still not good enough.

References

Bor-Chun Chen, Chu-Song Chen, and Winston H. Hsu. 2014. Cross-age reference coding for age-invariant face recognition and retrieval. In *Proceedings of the European Conference on Computer Vision (ECCV)*.

Eran Eidinger, Roee Enbar, and Tal Hassner. 2014. Age and gender estimation of unfiltered faces. *IEEE Transactions on Information Forensics and Security*, 9(12):2170–2179.

Lukasz Kaiser, Aidan N. Gomez, Noam Shazeer, Ashish Vaswani, Niki Parmar, Llion Jones, and Jakob Uszkoreit. 2017. One model to learn them all. *CoRR*, abs/1706.05137.

Rasmus Rothe, Radu Timofte, and Luc Van Gool. 2015. Dex: Deep expectation of apparent age from a single image. In *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pages 252–257.

Dinh Viet Sang, Le Tran Bao Cuong, and Vu Van Thieu. 2017. Multi-task learning for smile detection, emotion recognition and gender classification. In *Proceedings of the Eighth International Symposium on Information and Communication Technology*, SoICT 2017, page 340–347, New York, NY, USA. Association for Computing Machinery.

Mingxing Tan and Quoc V. Le. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. *CoRR*, abs/1905.11946.

Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. 2016. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503.