Age Estimation on Human-Face Images

Jinxi Xiao Yuecheng Xu Shangli Zhou
ShanghaiTech University
{xiaojx, xuych, zhoushl}@shanghaitech.edu.cn

Abstract

This project centers on solving the problem of human-face age estimation, capitalizing on advancements in machine learning and deep learning technologies. Our principal goal is to achieve precise age estimation of individuals by analyzing facial features extracted from images. To enrich the understanding of machine learning concepts acquired during this semester, a comprehensive set of traditional machine learning methods has been employed. Furthermore, novel internally-developed approaches are introduced to discern intricate patterns and correlations between facial characteristics and age labels. [1]

1 Introduction

Facial age estimation stands as a pivotal task with diverse applications, drawing considerable attention due to its historical significance and ongoing relevance. Despite the existence of well-established solutions, achieving flawless accuracy in age estimation remains an elusive goal, thus maintaining a high level of academic interest.

The central objective of age estimation involves determining the age of an individual depicted in a given facial image. This task can be approached from both classification and regression perspectives. In the classification paradigm, ages are treated as discrete values, reflecting the limited human lifespan spanning from 0 to 120 years. Consequently, this results in defining age groups as discrete classes, given the finite nature of age representation. Conversely, the regression viewpoint involves the mapping of facial images to their corresponding ages using a functional approach. To address this multifaceted challenge, our research explores several methodologies based on UTKFace dataset [1]:

1. Three distinct variations of K-Nearest Neighbors (KNN).
2. Utilization of Perceptron and Support Vector Machine algorithms for age classification.
3. Integration of ResNet [2] architecture for both age classification and regression tasks.
4. Implementation of a self-designed Coarse-to-Fine estimation process.
5. Exploration of the interconnections between age generative models and age estimation.

2 Related Work

2.1 Facial Age Estimation

Early approaches to facial age estimation predate the Deep Learning era. Young et al. [3] pioneered age classification, categorizing images into three age groups. Levi et al. [4] marked the onset of CNN utilization for age estimation. Advancements in neural networks led to various methods, including Cao et al.'s [5] ordinal regression using deep neural networks and Shin et al.'s [6] relative rank concept with a moving window mechanism. Currently, the state-of-the-art model on UTKFace is MiVOLO [7], a transformer-based feature fusion network with MAE 3.7.

[1] Codes for this report are stored at https://github.com/xiaojxkevin/age_estimation

Preprint. Under review.

2.2 Generative Models

The advent of GANs [8] expanded age estimation beyond prediction, delving into creative aspects. Works like [1] and [9] predict a person's appearance across different age periods based on their current facial image. The relevance of these generative models to our estimation approach is explored in Section 3.5.

2.3 Datasets and Benchmarks

IMDB-Wiki [10] stands as the largest dataset for age-related tasks, with MORPH [11] and UTKFace [1] also notable. UTKFace, chosen for its adequate data(over 23k images) and suitability for low-computation-resource conditions, covers ages from 1 to 116, with annotations limited to face crops.

3 Methods

3.1 K-Nearest Neighbors (KNN)

Age estimation can be likened to choosing the most similar samples from a pool, a process reminiscent of how children learn. In machine learning, this is encapsulated by the K-Nearest Neighbors (KNN) technique. The dataset is split into training and test sets. During testing, we identify the k samples with the smallest pixel differences for each test sample, employing either a voting mechanism (KNN classifier) or an averaging mechanism (KNN regression).

However, a critical consideration is whether comparing facial images based solely on pixel values is accurate. Defining the distance function is crucial in KNN. To explore this idea, we utilize PCA to compress images to lower dimensions and extract key features, and we will find closest neighbors based on these features.

3.2 Traditional Classification Methods

Perceptron and SVM have been chosen as the classification methods, extending them from the binary framework learned in class to a multi-class approach.

3.3 CNN Classifier and Regression

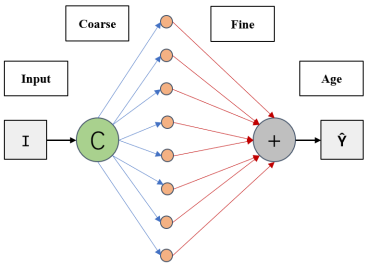
In pursuit of a more data-driven approach, we turn to deep learning, specifically utilizing Convolutional Neural Networks (CNNs) for age estimation, given that our inputs are images. ResNet [2], renowned for its outstanding performance on ImageNet datasets, emerges as our model of choice. To expedite our work, we leverage the pre-trained resnet18 model available in PyTorch. Our selection of resnet18 is motivated by considerations of parameter efficiency.

Furthermore, recognizing the similarities between classification and regression tasks, we apply resnet18 to both, with variations in the last fully connected layer and loss function to accommodate the specific requirements of each task.

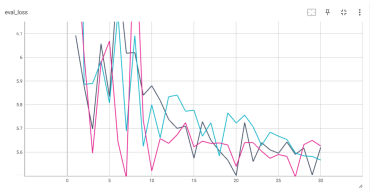
3.4 Coarse-to-Fine Estimation Process

To further enhance result accuracy, we introduce a coarse-to-fine process, enhancing the model's sensitivity to specific age periods and overall performance.

We first train an overall coarse classification model, which divides the face images into 8 age groups for prediction. Subsequently, we train 8 expert models, each specializing in its assigned age group. These expert models exhibit high accuracy in predicting the age group they are designed for. After the completion of age group classification by the overall model, the images are then assigned to the corresponding expert models for further detailed regression. This coarse-to-fine estimation process allows for a more refined and accurate age estimation for each individual. Model architecture is shown in Figure 1a



(a) Architecture of Coarse-to-fine model



(b) Validation results for CNN classification: Pink for original; black for weights 1.25 and blue for weights 1.5

3.5 Connections with Generative Models

We closely examine the work of [1], which predicts a person's face at different age periods given their current age. The key proposition is the existence of a manifold $\mathcal{M}$ representing the distribution of face images. The authors leverage age as a conditioning factor for training an adversarial network. Recognizing the potential applicability to age estimation, we conceptualize this task as the inverse of age estimation. Strengthening the generative model's ability to predict current age images could further imply its capacity in capturing age-related features.

However, for simplicity and efficacy, the authors categorize images into 10 groups, akin to our approach in Section 4.1. This categorization is facilitated by the *one-hot encoding* mechanism. The generator requires age information, leading to the one-hot encoding of age into a 10-dimensional vector. Crucially, age estimation involves fine-grained classification, necessitating a compact label class space. Clustering ages into groups is a coarse-grained procedure. Labeling each age as a class would result in a sparse vector of approximately length 100, posing challenges for both computation and model performance.

Due to time constraints, a solution for this challenge has not been proposed in this report. However, we think that future advancements in generative models may offer resolution to this issue.

4 Experiments

4.1 Dataset Preparations

The dataset is divided into training, validation, and test sets with a ratio of [0.8, 0.1, 0.1] to facilitate network training. To validate the appropriateness of our split, we examine the class distribution(refer to Section 4.1) of the train set: [0.080, 0.025, 0.039, 0.076, 0.450, 0.259, 0.053, 0.018], which closely mirrors the distribution in the overall dataset. Furthermore, during data loading, we implemented data augmentation, primarily involving pixel value modifications.

In order to build a coarse-grained label class space and simplify the problem for traditional methods, we clustering images labelled with ages ranged from 1 to 116 into 8 groups shown in the table below:

class	0	1	2	3	4	5	6	7
ages	1-3	4-6	7-12	13-21	22-37	38-65	66-84	85 and older

4.2 Implementation Details and Results

4.2.1 KNN, Perceptron and SVM

KNN: For two images, we defined the distance function to be $d(A, B) = \sqrt{\sum_{i=1}^{n}(A_i - B_i)^2}$ where A_i and B_i represent the pixel values of image A and B at position i, respectively. For KNN classifier, we calculate the mode of k samples; while for KNN regression, we calculated the average of k samples. Also, we set k in $[1, 3, 5, 8, 10, 12, 15, 20, 50, 100, 150, 200]$. In addition, as for KNN-PCA, dimensions after compression were set to 1, 10, 100, and 200. Best results are shown in Table 2.
Perceptron: We set learning rate to 0.03 and performed 10 epochs. Despite these settings, the samples were still classified into the aforementioned 8 age groups. The best results are shown in Table 2.

SVM: We set learning rate to 10^{-7} and added a *L2* regularization into the loss. The following regularization coefficients were tested: $[10^{-7}, 2.5\times10^{-6}, 5\times10^{-6}, 7.5\times10^{-6}, 10^{-5}, 2.5\times10^{-5}, 5\times10^{-5}, 7.5\times10^{-5}]$. The best results are shown in Table 2.

4.2.2 ResNet18

We initially employ ResNet18 for classification by adjusting the output of the last fully connected layer to 117 and using the *CrossEntropy* loss. To further analyze test MAE shown in Table 2, we examine MAE in coarse age groups from Section 4.1. Specifically, MAE for classes 5-7 is 8, 13 and 13 respectively, indicating suboptimal estimation for older age groups. To address this, we adjust the weights in loss for ages in class 5 to 1.25 and 1.5 subsequently, while maintaining weights at 1.0 for all other ages. Evaluation losses are visualized in Figure 1b, demonstrating that setting weights to 1.25 generally improves both overall MAE and the MAE of class 5.

Subsequently, we implement ResNet18 for regression, configuring the output of the last fully connected layer to be 1. Utilizing the *L1* loss for supervision surprisingly yields superior performance, achieving an MAE of 5.12. This improvement may stem from the inadequacy of using only CrossEntropy loss to supervise a fine-grained problem with over 100 classes, suggesting the potential benefit of introducing more specialized loss functions, such as the triplet loss.

4.2.3 Coarse-to-Fine Process

In this part, we continued with the approach of ResNet18. For the overall model, we modified the output layer to have 8 unit s, corresponding to the 8 age groups for classification. We set the initial learning rate to 1×10^{-3} and applied a discount factor of 0.9 every 2 epochs. The overall model was trained for a total of 30 epochs with a batch size of 256. For the 8 expert models, we only used the data corresponding to their respective age groups for training and testing. We set the initial learning rate to 5×10^{-4} and applied a discount factor of 0.9 every 3 epochs. The expert models were trained for a total of 20 epochs with a batch size of 64. The output layer of the expert models was set to 1 for regression tasks. During the integration testing phase, we input the test images into the overall model to predict the age stage to which each image belongs. Then, we pass each image to the corresponding expert model for accurate regression prediction within that age stage.

Model	Accuracy/MAE in expertise area
overall	75.6%
expert0	0.45
expert1	0.78
expert2	1.15
expert3	1.62
expert4	2.99
expert5	5.11
expert6	3.56
expert7	4.62

Table 1: Test values for coarse-to-fine model on each stage of test dat.a

Method	MAE	Accuracy
SVM	-	52.1%
Perceptron	-	47.8%
KNN	16.43	51.0%
KNN-PCA	16.61	11.0%
CNN Classifier	5.90	-
CNN-C-w-1.25	5.72	-
CNN Regression	5.12	-
Coarse-to-Fine	**4.48**	**75.6%**

Table 2: General results of all methods. For accuracy, we consider classifying into 8 age group instead of the numerical values.

5 Conclusions

In traditional machine learning methods, SVM has shown the best performance.
KNN-PCA showed a faster speed than traditional KNN, however, some information about age was lost during the phase of dimension reduction. We think its demerit outweigh its merit.
Among all methods, our self-designed Coarse-to-Fine model exhibits the best performance , which performs one year below the SOTA model methoned in Section 2.1. Notably, it successfully reduces the loss on classes 5-7, representing the most challenging age groups for estimation. This substantiates the effectiveness of the coarse-to-fine approach. While we did not provide a solution for integrating generative models, it remains a promising avenue for further exploration after this class.

References

[1] Z. Zhang, Y. Song, and H. Qi, "Age progression/regression by conditional adversarial autoencoder," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, July 2017.

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, June 2016.

[3] Y. H. Kwon and N. da Vitoria Lobo, "Age classification from facial images," *Computer Vision and Image Understanding*, vol. 74, no. 1, pp. 1–21, 1999.

[4] 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)*, pp. 34–42, 2015.

[5] W. Cao, V. Mirjalili, and S. Raschka, "Rank consistent ordinal regression for neural networks with application to age estimation," *Pattern Recognition Letters*, vol. 140, pp. 325–331, 2020.

[6] N.-H. Shin, S.-H. Lee, and C.-S. Kim, "Moving window regression: A novel approach to ordinal regression," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, June 2022.

[7] M. Kuprashevich and I. Tolstykh, "Mivolo: Multi-input transformer for age and gender estimation," *arXiv preprint arXiv:2307.04616*, 2023.

[8] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014.

[9] X. Tang, Z. Wang, W. Luo, and S. Gao, "Face aging with identity-preserved conditional generative adversarial networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7939–7947, 2018.

[10] R. Rothe, R. Timofte, and L. Van Gool, "Dex: Deep expectation of apparent age from a single image," in *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pp. 252–257, 2015.

[11] K. Ricanek and T. Tesafaye, "Morph: a longitudinal image database of normal adult age-progression," in *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pp. 341–345, 2006.