Age estimation with pytorch

Amirmahdi Aramide

February 2024

1. Problem statement

Due to inherent challenges in facial image analysis, existing age estimation models based on PyTorch and convolutional layers often struggle with accuracy and generalizability on diverse datasets like UTKFace. This project aims to develop a ResNet-based architecture that leverages the power of residual connections and potentially incorporates advanced techniques like data augmentation to achieve significantly higher accuracy and robustness in age estimation across diverse demographics and image conditions. The goal is to create a more reliable and inclusive age estimation model suitable for real-world applications.

1.1. Challenges

- Genetics
- Life style
- Ethnicity
- Makeup

1.2. Goals

- Resistant to ethnicity
- Resistant to imaging location
- Accurate age calculation

2. Available methods

2.1. Regression

Propose a Convolutional Neural Network (CNN) model for estimating age from cropped face images. Instead of using the usual classifier layer at the end, this model features a regression layer.

2.2. Classification

Propose a Convolutional Neural Network (CNN) with fully connected with n neurons model for estimating age from cropped face images.

2.3. MIVOLO

Multi-input Transformer for Age and Gender Estimation The paper introduces MiVOLO, a method for age and gender estimation that utilizes a vision transformer and integrates both tasks into a unified dual input/output model. By incorporating person image data in addition to facial information, the model demonstrates improved generalization and the ability to estimate age and gender even when the face is occluded. Experimental results on multiple benchmarks show state-of-the-art performance and real-time processing capabilities. The model's age recognition performance surpasses human-level accuracy across various age ranges. The code, models, and additional dataset annotations are publicly available for validation and inference.

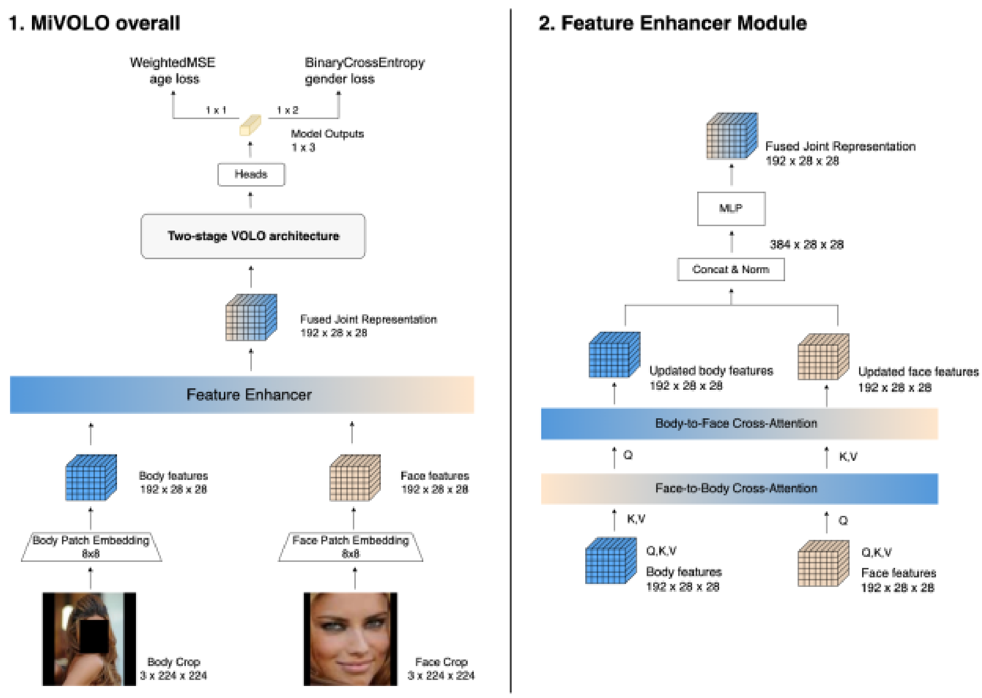
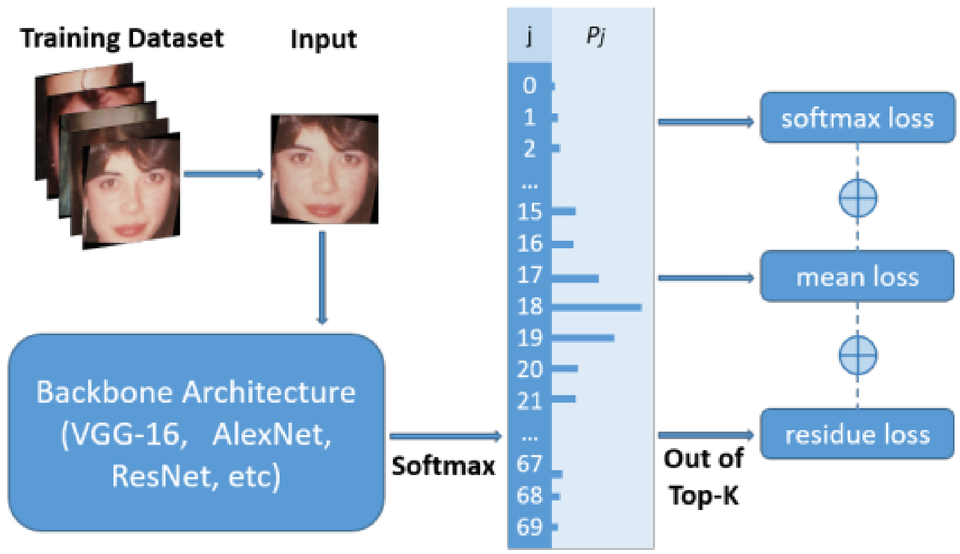


Figure 4: MiVOLO. We present the overall model and a feature enhancer module in block 1 and block 2 respectively.

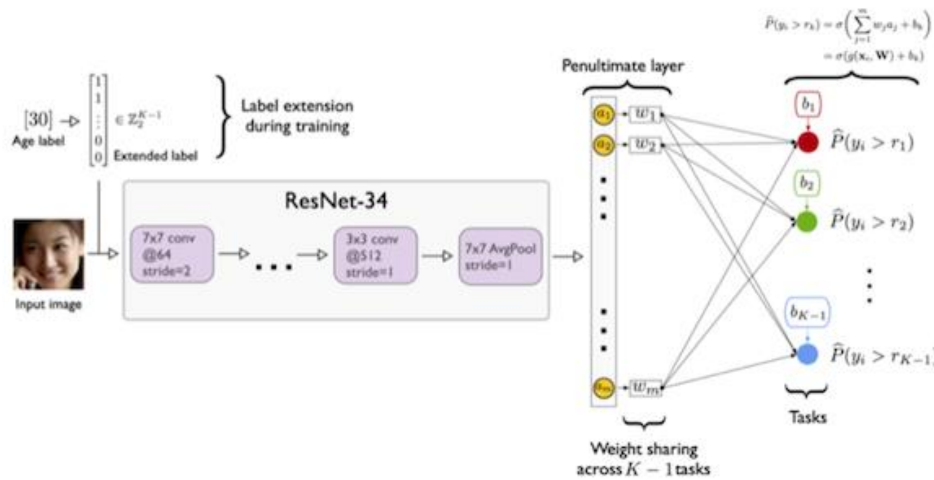
2.4. Adaptive mean residue loss

Adaptive Mean-Residue Loss for Robust Facial Age Estimation Automated facial age estimation has diverse real-world applications in multimedia analysis, e.g., video surveillance, and human-computer interaction. However, due to the randomness and ambiguity of the aging process, age assessment is challenging. Most research work over the topic regards the task as one of age regression, classification, and ranking problems, and cannot well leverage age distribution in representing labels with age ambiguity. In this work, we propose a simple yet effective loss function for robust facial age estimation via distribution learning, i.e., adaptive mean-residue loss, in which, the mean loss penalizes the difference between the estimated age distribution's mean and the ground-truth age, whereas the residue loss penalizes the entropy of age probability out of dynamic top-K in the distribution. Experimental results in the datasets FG-NET and CLAP2016 have validated the effectiveness of the proposed loss

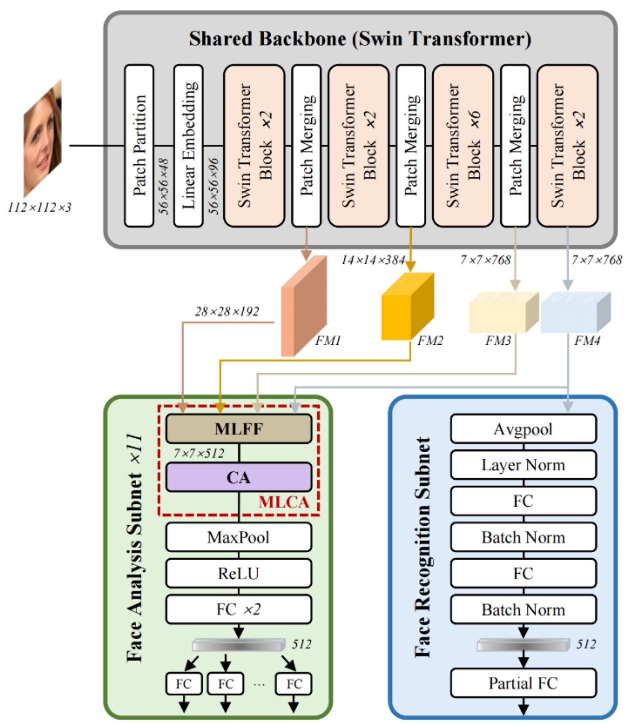


2.5. Coral

Rank consistent ordinal regression for neural networks with application to age estimation The paper addresses the issue of capturing relative ordering information in class labels for tasks like age estimation. It introduces the COnsistent RAnk Logits (CORAL) framework, which transforms ordinal targets into binary classification subtasks to resolve inconsistencies among binary classifiers. The proposed method, applicable to various deep neural network architectures, demonstrates strong theoretical guarantees for rank-monotonicity and consistent confidence scores. Empirical evaluation on face-image datasets for age prediction shows a significant reduction in prediction error compared to reference ordinal regression networks.



2.6. A Multi-task Transformer for Face Recognition, Expression Recognition, Age Estimation and Attribute Estimation The paper introduces SwinFace, a multi-task algorithm for face recognition, facial expression recognition, age estimation, and face attribute estimation. It utilizes a single Swin Transformer with task-specific subnets and a Multi-Level Channel Attention module to address conflicts and adaptively select optimal features. Experimental results demonstrate superior performance, achieving state-of-the-art accuracy of 90.97% on facial expression recognition (RAF-DB) and 0.22 error on age estimation (CLAP2015)



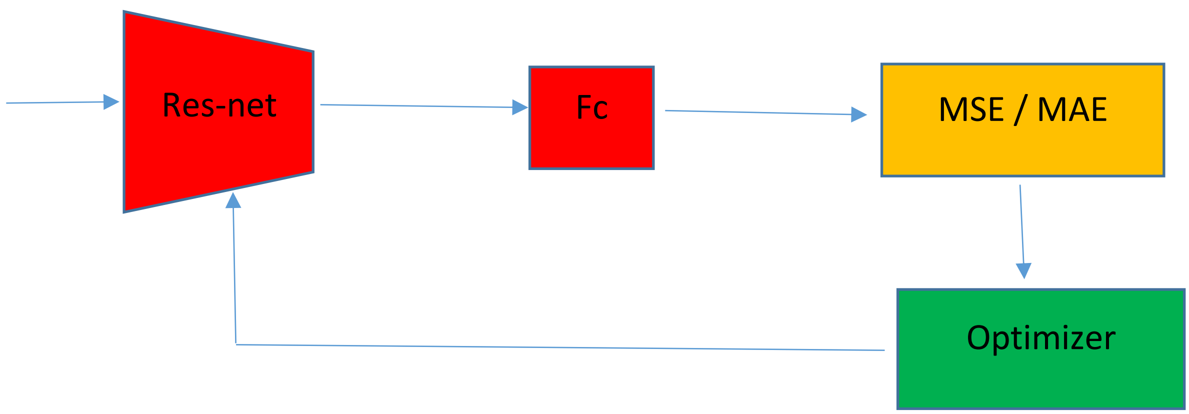
2.7. Deep Regression Forests for Age Estimation

The paper introduces Deep Regression Forests (DRFs) as an end-to-end model for age estimation from facial images. DRFs address the challenge of heterogeneous facial feature space by jointly learning input-dependent data partitions and data abstractions. The proposed method achieves state-of-the-art results on three standard age estimation benchmarks, demonstrating its effectiveness in capturing the nonlinearity and variation in facial appearance across different ages.

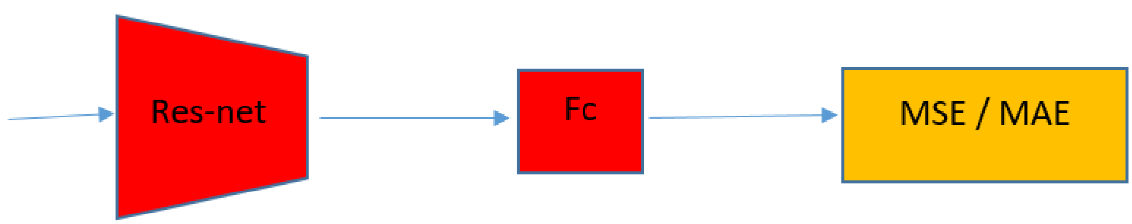
	Method	Summary	link
1	Regression	Using cnn and fully connected as regression layer	code
2	Classification	Using cnn and classifier layer	code
3	MIVOLO	Mivolo's architecture shines through its dual-input processing, combining facial image information with full-body images for superior age and gender estimation. It leverages a dual-stream CNN with residual connections, ultimately employing a regression layer for continuous age prediction, exceeding the accuracy of traditional classification methods.	code
4	Adaptive Mean Residue Loss	This loss function tackles facial age estimation by combining two components: 1) **Mean Loss:** penalizes the difference between the estimated and ground-truth age distributions' means, ensuring overall accuracy. 2) **Residue Loss:** focuses on the "tail" of the estimated distribution, using entropy to prioritize penalizing predictions deviating significantly from the true age group, leading to improved robustness and generalizability across diverse datasets.	code
5	Coral	The paper proposes CORAL, a framework that leverages binary classification to capture ordinal relationships in labels like age, leading to more	code

		accurate age estimation with consistent and reliable confidence scores.	
6	SwinFace	SwinFace, a single Swin Transformer with specialized subnets and adaptive feature selection, excels in multitasking, showcasing state-of-the-art results in face recognition, expression recognition, age estimation, and attribute estimation.	code
7	Deep Regression Forests for Age Estimation	This paper proposes Deep Regression Forests, a novel model combining data partitions and abstractions to effectively estimate age from facial images, achieving state-of-the-art results on standard benchmarks.	paper

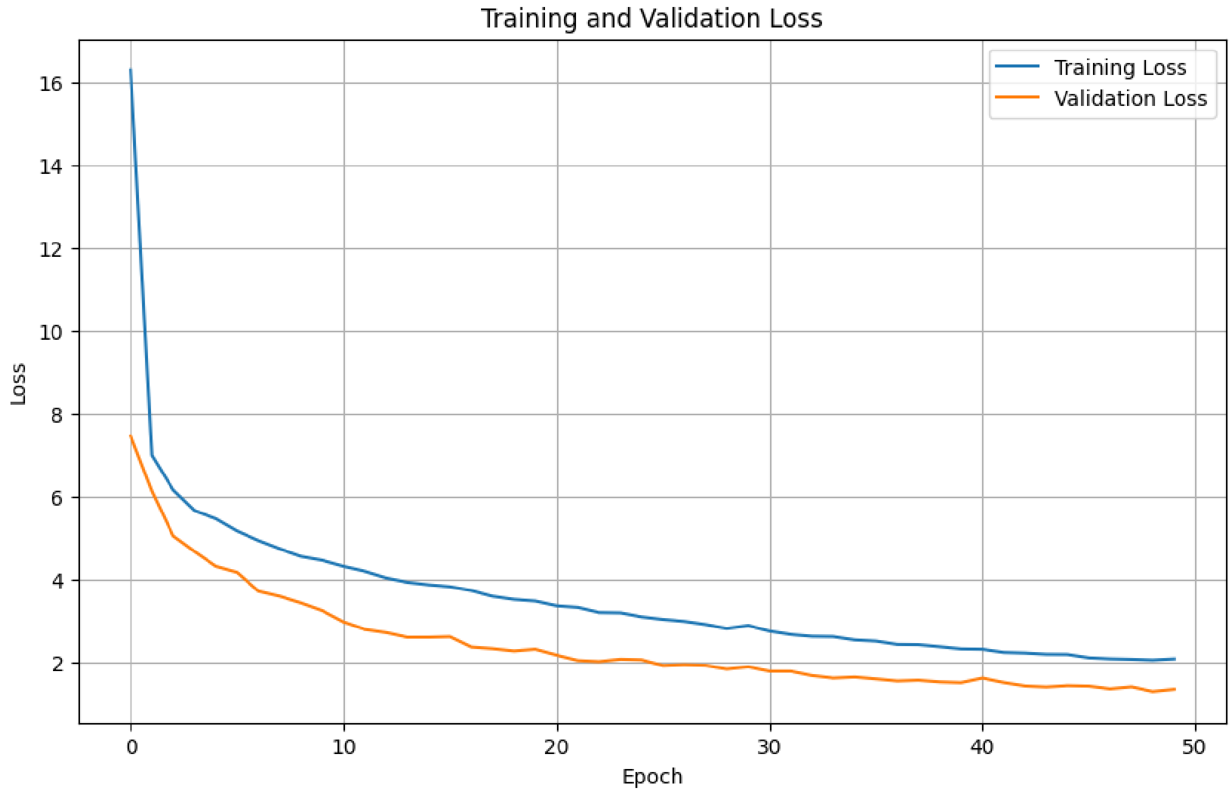
3. Diagram
   - Train



   - Evaluate



4. Results



Training and Validation Loss
Training Loss
Validation Loss
Loss
Epoch
16
14
12
10
8
6
4
2
0
10
20
30
40
50