Gender and Age Estimation from Facial Images Based on Multi-Task and Curriculum Learning

Toma Brezovan, Claudiu Ionuț Popîrlan
University of Craiova, Romania

***Abstract*—This study presents a multi-task deep learning approach for predicting age and gender attributes from facial images, with the aim of obtaining a robust dual classifier. The proposed system uses the pre-trained EfficientNet-B4 model as the feature extractor of the main model and incorporates a two-branch architecture, where the output of the gender classification branch informs the age prediction branch. This means a conditional feature learning with an explicit injection mechanism, by injecting gender information into the age field of the dual-task model, which is one of the novelties of our proposal. A curriculum learning strategy is applied during training to progressively improve the model's performance using various datasets, such as UTKFace, MORPH-II, and Adience. The proposed multi-phase curriculum learning strategy, which uses both multi-task learning and multi-dataset training, is another novelty of our proposal. Experimental results show that the model achieves high accuracy in both age and gender classification tasks while maintaining low inference latency. Furthermore, the experiments highlighted that the classification accuracy values of the proposed method, both for gender and age, as well as in all datasets used, are close to the best state-of-the-art results, which validates the robustness of the proposed classifier.**

Keywords*—*Age estimation; gender classification; multi-task learning; curriculum learning

I. Introduction

Human face analysis is an important area of research in computer vision, for performing and optimizing various facial perception tasks, such as face recognition, age estimation, gender determination, etc. In addition, biometric identification has been increasingly used recently in various fields, such as automatic detection of physical presence and confirmation of the identity of individuals through image analysis, and gender and age are two of the main biometric attributes [1].

Predicting age and gender from facial images is still an active task in the field of computer vision, having a wide range of applications in areas such as human-computer interaction, image retrieval, security, surveillance and web content filtering [2].

Recent advances in deep learning have significantly improved the performance of age and gender recognition systems, particularly through convolutional neural networks (CNNs) and multi-task learning (MTL) frameworks [3]. MTL approaches are especially attractive in this domain, as age and gender share underlying visual features that can be exploited to improve generalization.

However, existing models usually use either a specialized dataset of facial images using age and gender attributes for training, or other models trained on very large facial datasets (e.g., FaceNet [4], VGGFace [5], or ImageNet [6]), and then, through transfer learning, the training is refined also on a specialized dataset of facial images using age and gender attributes. In both cases, the training and then estimation process is limited to a not very large number of facial images, which depends on the size and diversity of the specific dataset used for training (or fine tuning).

One goal of our approach is to create a dual classifier for the age and gender of people in facial images, which is robust and can estimate the two attributes in various images. For this, we will use several public datasets for training, which contain images in various conditions, which will allow for more accurate estimation of both age and gender.

The tasks related to the prediction of age and gender of individuals are not completely independent, because there are studies that have demonstrated that incorporating gender information can improve the accuracy of age estimation from facial images [7]. Another goal of our approach is to exploit the dependency between the two tasks, which allows for increased accuracy for both age and gender estimation.

Although there are proposals for the correlated treatment of gender and age in facial images, these refer either to different ways of training the models or to modifying the structure of the models, but not to the direct dependence between the gender and age tasks. We use a simpler variant, regarding the modification of the model structure, resulting in a conditional or hierarchical model with dual tasks, as presented in Fig. 1.

In this study, we propose an efficient multitask deep learning model to predict age and gender in facial images. Although there are various proposals for this problem, the presented system incorporates efficient methods from the field of multitask learning, which allow increasing both the robustness of the system and the accuracy of the estimates.

The model employs *EfficientNet-B4* [8] as a pre-trained backbone, which is known for its balance between accuracy and computational cost. EfficientNet introduces a new strategy, known as Composite Scaling. It allows balancing and adjusting the depth, width, and resolution of the network, thus allowing the model to scale to arbitrary sizes. In addition, EfficientNet maintains a balance between performance and efficiency. EfficientNet (B0-B7) networks have been widely used recently in various fields of computer vision, especially in image classification [9]. They have also been used for age estimation in facial images [10].

A two-branch architecture is constructed, in which the age and gender branches are not totally independent. According to [7], gender classification output is fed into the age estimation branch to improve age prediction performance through feature sharing.

In the case of multi-task classification systems, there can be problems with the loss function, because the tasks usually differ in scale, complexity, or quality of the labels. Instead of using the weighted linear sum of losses for each individual task, we use a learnable task uncertainty by modeling homoscedastic uncertainty, as in [11].

Even if an efficient pre-trained model is used as feature extractor, the robustness of a dual-task model for simultaneous gender and age recognition depends largely on the dataset on which the dual model is trained. To increase the robustness of such a dual classifier, we used several established datasets for training and adapted the curriculum learning strategy [12].

Our work builds on these foundations, which are also the main contributions:

- Using the pre-trained *EfficientNet-B4* model as a backbone, we propose a dual-task classifier for gender and age in facial images, where the output of the gender branch is used in the input of the age branch, so that the gender prediction output is used to improve age estimation.
    - Unlike previous studies, such as [7], [13]–[15], where multiple subnetworks for the two tasks are created (and possibly trained separately) and interconnected, we propose a conditional feature learning with explicit injection mechanism, by injecting gender information into the age head of the dual-task model.
- Due to inconsistent age distributions and labeling schemes in the training datasets, we use a unified age classification strategy for all datasets, as well as a unified data loader for the datasets, which can be later used in the curriculum learning strategy. In addition to the unified age classification strategy, we propose an adaptive algorithm that learns age categories based on the age distribution in the datasets.
    - As far as we know, there are no proposals that automatically generate age groups based on information in datasets.
- We use a learnable task uncertainty so that the contribution of the task losses in the total loss is adaptive during training and the accuracy of the estimates is increased. This method is based on the method proposed in [11].
- We implement a multi-phase curriculum learning strategy for the training operation, based on several elements, including: a) age range coverage, b) age label granularity, and c) image quality. This strategy allows for increased prediction accuracy, as well as eliminating the catastrophic forgetting effect due to the sequential use of samples from the datasets.
    - Although it is not an adaptive method like in [16], unlike the previous methods [17], [18], the proposed method refers to both multi-task learning and multi-dataset training. To our knowledge, there have been no other proposals regarding multi-task learning and multi-dataset training

The remainder of the study is organized as follows: Section II provides a review of the related work. The proposed method is described in Section III which also includes the dual-task classifier and the curriculum learning strategy. The experimental results are presented in Section IV. The study is summarized, with some future directions, in Section V.

II. Related Work

Deep learning methods for age and gender recognition mainly use convolutional neural networks (CNN) trained from scratch or with minimal learning transfer. These models focused on learning hierarchical features directly from facial images. Levi and Hassner [19] introduced a CNN architecture for age and gender classification, demonstrating the effectiveness of CNNs in capturing age- and gender-relevant facial features. The paper [20] proposes an age and gender prediction method from face images using convolutional neural network (CNN). In this paper [21] authors utilized Haar Cascades for face detection and a CNN for classification, achieving real-time age and gender prediction.

A. Architectural Models

Recent advancements have incorporated sophisticated architectures, transfer learning, and attention mechanisms to enhance accuracy and robustness in age and gender recognition. The research [22] uses pre-trained models such as VGG19 and VGGFace, fine-tuned for age and gender prediction, using the MORPH2 dataset. The authors [23] compared custom CNN architectures with pre-trained models such as VGG16, ResNet50, and SE-ResNet50. The study found that transfer learning significantly improved performance over training from scratch. The paper [24] introduces an ensemble of attentional and residual CNNs, leveraging attention mechanisms to focus on informative facial regions, thereby enhancing prediction accuracy.

B. Multi-Task Learning

Multi-Task Learning (MTL), or *multi-attribute learning* (MAL) aims to improve the generalization of models by leveraging domain-specific information contained in the training signals of related tasks. In computer vision, this often involves simultaneously learning tasks such as object detection, semantic segmentation, and depth estimation.

There are two main categories for the MTL, related to the structure of these neural networks [25]:

1) Hard parameter sharing: where the initial layers of the network are shared among all tasks, while the task-specific layers are kept separate.

2) Soft parameter sharing: where each task has its own model with its own parameters, but regularization techniques are used to keep the parameters of different models similar. This approach can lead to high inference cost [26].

MTL networks were used for face attribute estimation [7], as well for other tasks, such as human pose estimation [27], or face alignment [28].

In MTLs, in addition to network structure, recent advances also relate to loss weighting. Several loss weighting methods are proposed to automatically combine the losses, such as: *task prioritization* [29], which weights task losses according to more difficult tasks during training, *uncertainty weighting* [11], which models loss weights as task-dependent but data-agnostic uncertainty, or *gradient normalization* [30], which learns loss weights to enforce that the gradient norms for each task are close

C. Curriculum Learning

Curriculum learning (CL) was introduced by Bengio et al. [31] and involves training models by gradually increasing the difficulty of training data. This method contrasts with traditional training, in which data is presented in random order. Easy-to-difficult ordering can also be used in multi-task learning, by determining a learning order of tasks to maximize the final result [17], [18]. There are several categories of curriculum learning [12], such as: *Teacher-student CL* [32], or *Implicit CL* [33]. In [34], a curriculum learning approach for classification tasks on small to medium-sized datasets is proposed, based on the order of samples.

III. Proposed Method

A. Dual-Task Architecture

Since hard parameter sharing allows multiple tasks to share low-level parameters [26], this structure has advantages such as lower storage costs. For this reason, the proposed deep multi-task network is based on *hard parameter sharing*, as illustrated in Fig. 1, where X represents the input tensor (images), while X_1 and X_2 represent the outputs of the model (predicted age and predicted gender, respectively). In addition, for a faster training and improved accuracy, the proposed dual-task classifier uses *EfficientNet-B4* as a shared backbone for feature extraction. The shared layer before the branches forces the model to compress shared task-based information into a lower-dimensional vector, encouraging learning of shared features and acting as an inductive bias.

Several studies across biomedical research and computer vision confirm that aging patterns differ significantly between men and women, which supports the integration of gender cues in age estimation models. Biological factors such as hormonal differences lead to gender-specific aging trajectories in skin elasticity, bone structure, and facial fat distribution [35]. In computer vision, deep learning-based age estimation models consistently achieve higher accuracy when gender information is included as input, indicating that aging cues vary in distribution across genders [36].

There are not many proposals for the correlated treatment of gender and age in facial images. In the study [13], a fusion layer between gender-based output and age-based output is proposed, which generates age prediction. In the study [14] a cascaded framework containing a parent network and multiple subnetworks is presented. For example, when using a gender classifier trained on gender information, the other two subnetworks are trained on male and female samples, respectively. In the study [15], age grouping for male and female faces is processed separately: one model is trained to group male faces into different age groups, and another model is learned to group female faces. In the study [36], a hybrid deep learning architecture is proposed for facial gender and age classification: a deep random forest is used to estimate facial gender, and then age is recognized under the conditional probability of gender alignment. In the study [7], it is proposed four two-level CNN models: the first level for gender classification, which contains one CNN, while the second level contains three CNNs, one for each age category (children, middle-aged people, adults, and the elderly).

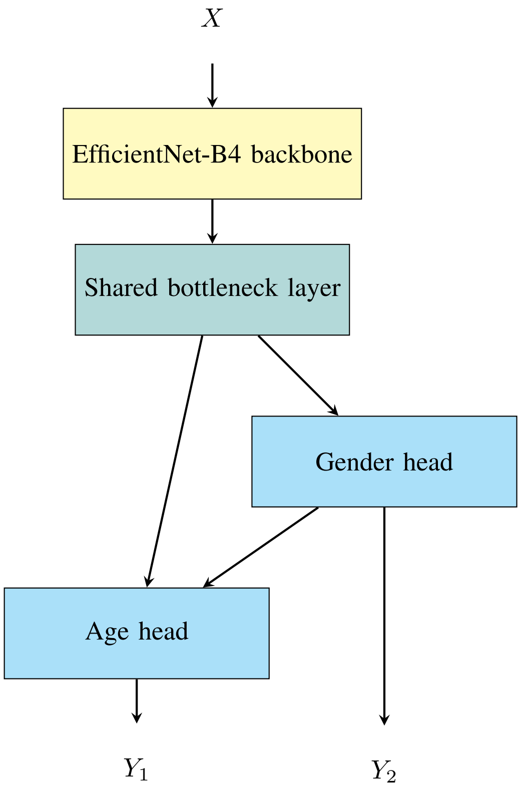

Fig. 1. Architecture of the dual-task classifier

In this proposal, a different approach has been chosen. We designed a *conditional* or *hierarchical model* with dual task, in which the gender prediction comes first, and the age classifier benefits from this prediction, as presented in Fig. 1.

More exactly, to inject gender information into the age head, we use a *hard conditioning* (post-prediction) method:

- Run gender head first.
- Concatenate predicted gender with shared features.
- Feed them into the age head.

The structure of the proposed architecture, as shown in Fig. 1, has several advantages:

- The model can learn gender-specific age patterns.
- It also encourages the shared feature extractor to learn disentangled but informative representations.

- Conditioning can reduce the variance in age prediction.

B. Training Phase

In the training process, we use two main improvements to increase the accuracy of the estimation operations:

- An adaptive method to determine task-specific weights when determining the total loss.
- A multi-dataset curriculum strategy by designing a training schedule, which is similar to domain adaptation and multi-source learning.

In multi-task learning models, despite the advantages offered by the hard parameter sharing structure, it presents a major drawback: for a global optimization, all task-specific loss objectives must be combined, requiring task-specific weights. Selecting these weights can be difficult and expensive [25]. Even though the simplest way is to sum partial task losses, this often leads to unbalanced training.

For this reason, in this proposal we will use an adaptive method to determine task-specific weights, based on the *Multi-Task Uncertainty Weighting* method [11]. This *first improvement* of the training process allows automatic learning of task weighting by modeling homoscedastic uncertainty.

The *second improvement* in the training process is related to how a deep learning model learns from its input data. Bengio [31] introduced the term *curriculum learning*, which is an optimization strategy for handling a minimization problem. When training a deep neural network, instead of using random samples, it is better to organize these samples so that the less complex examples are presented first.

Since a curriculum is implemented by ordering training data according to difficulty, we will need to generalize this to be able to handle training with multiple datasets, taking into account the following information:

- The quality of the images in the datasets,
- Age range coverage,
- Label granularity.

1) Dynamic learning of task weighting: Multi-task learning involves optimizing a model based on multiple objectives (determining a total loss based on the partial task losses) [see Eq. (1)]:

$$\mathcal{L} = \sum_i w_i \mathcal{L}_i, \tag{1}$$

where, task weights w_i are difficult to estimate.

In the proposal [11], the estimation problem of these weights is viewed as an aleatoric uncertainty, which is task-dependent (called *homoscedastic uncertainty*). In this case, the total loss is derived from maximum likelihood estimation, and for classification tasks, the minimisation objective (loss) has the following form (only two taks are considered here):

$$\mathcal{L} = \frac{1}{\sigma_1^2}\mathcal{L}_1 + \log(\sigma_1) + \frac{1}{\sigma_2^2}\mathcal{L}_2 + \log(\sigma_2), \tag{2}$$

where, σ_1 and σ_2 are observation noise scalars, representing the noise parameters for the two tasks. Using classical notation, σ_i represents the standard deviation, while σ_i^2 represents the variance.

Remark. There are some small differences from the relation in the work [11], because we are performing classification tasks, not regression tasks:

- $\mathcal{L}_1$ and $\mathcal{L}_2$ are cross-entropy losses, not Euclidean losses.
- Coefficients $\frac{1}{2\sigma_i^2}$ were replaced by $\frac{1}{\sigma_i^2}$.

As in [11], in practice the dual-task model should be trained to predict the log variance,

$$s = \log(\sigma^2), \tag{3}$$

Because it is more numerically stable.

To achieve this, two trainable scalar parameters, p_1 and p_2, will be defined as Eq. (4),

$$p_i = \log(\sigma_i^2),\ i \in \{1, 2\}, \tag{4}$$

which represents the log of the variance (uncertainty) of each task. These trainable parameters:

- Will be initialized to 0 and updated in the training phase through gradient descent, just like any other weight.
- They will be used as uncertainties associated with the tasks when calculating the total loss:

$$\mathcal{L} = \mathrm{e}^{-p_1}\mathcal{L}_1 + p_1 + \mathrm{e}^{-p_2}\mathcal{L}_2 + p_2 \tag{5}$$

Remark. Using Eq. (3), the Eq. (2) and Eq. (5) are the same.

2) Multi-dataset curriculum strategy: Typically, deep classification networks are trained using a specific dataset, which is relevant to the respective domain, and the method used for training is to randomly select samples at each iteration step.

This proposal attempts to extend classical training methods in the following way:

- Instead of using a single dataset to train the dual-task classifier, we use multiple datasets that are common in the field of estimating the age and gender of people in images (UTKFace, MORPH-II, and Adience)
- We use a learning strategy based on the *curriculum learning* method [31], which involves training models by gradually increasing the difficulty of the training data.

When a neural network is trained sequentially on multiple datasets, a phenomenon called *catastrophic forgetting* can occur, in which the neural network, after being trained on one dataset and reaching a certain level of performance, loses that knowledge when trained on a new dataset. There are several proposals to reduce forgetting in curriculum learning, such as:

- Regularization-based methods [37], where methods are applied that penalize changes in important weights for previously learned data
- Progressive training with overlap [38], which involves gradually increasing the proportion of more difficult samples, while the model continues to train on the previous ones

We will define a multi-phase curriculum training strategy that allows training the dual-task model using multiple datasets, according to the *progressive training with overlap* method. In other words, an ordered sequence of examples, from simple to complex, will need to be determined, which will be exposed in the training process.

To be consistent with [31], the following questions must be answered:

- What does a sample mean, given that multiple datasets are used?
- How can the difficulty of a sample be estimated?
- How can the training schedule be specified so that the model is exposed from easier to more complex samples?

Of the two classification tasks, the one associated with age is the most difficult, because the facial aging process is random, both for each individual person and for different categories of people. For this reason, age-related characteristics contribute to determining the difficulty of a sample. To be able to efficiently and robustly learn the age classifier:

- A uniform age category scheme will need to be created, depending on the structure of the datasets used for training.
- A generic, instantiable data loader will need to be created that allows the return of samples from a specified dataset, which refers to a specified list of age categories (age bins).

The answer to the first question is the following: a sample from a multi-dataset curriculum strategy is a sequence of samples from a list of specified datasets, associated with a specified list of age bins. This can be formalized as follows.

Let N_b be the number of age categories, and

$$\mathcal{B}_a = \{0, 1, \ldots, N_b - 1\},$$

the set of age bins. Let $\mathcal{D} = \{\mathcal{D}_i\}_{i=1}^{N_d}$ be the set of training datatsets, where each dataset $\mathcal{D}_i$, $1 \leq i \leq N_b$, is specified as:

$$\mathcal{D}_i = (Name_i, \mathcal{S}_i),$$

where, $Name_i$ is the name of the dataset, and $\mathcal{S}_i \subseteq \mathcal{B}_a$ is the set of age bins the dataset $Name_i$ contains.

Denoting by $\mathcal{DL}$ the generic data loader, a k^{th}-*sample* from a multi-dataset curriculum strategy is associated to a sequence of datasets, $DSeq_k \subseteq \mathcal{D}$, and a set of age bins, $BSet_k \subseteq \mathcal{B}_a$, and it can be defined as a sequence (a list) of *elementary samples*:

$$\begin{aligned} \mathcal{X}_k &= \mathcal{X}(DSeq_k, BSet_k) = [\mathcal{X}_k^1, \ldots, \mathcal{X}_k^p], \\ \mathcal{X}_k^j &= \mathcal{DL}(Name_{k_j}, BSet_k),\ 1 \leq j \leq p, \end{aligned} \tag{6}$$

where, $Name_{k_j} \in DSeq_k(\mathbb{N})$ for $1 \leq j \leq p$.

Remark. The order of the datasets in Eq. (6) matters because they form a sequence (a list) and not a set.

The difficulty of the samples depends on the datasets used for training, and the calculation of this score considers elements such as :

- Image quality in datasets,
- Age range coverage,
- Number of samples per age range.

We will denote such a scoring function with $score()$.

A *schedule training* is a list of ordered samples [see Eq. (7)],

$$\begin{aligned} \mathcal{ST} = [\mathcal{X}(DSeq_1, BSet_1), \ldots, &\mathcal{X}(DSeq_t, BSet_t)] = \\ &= [\mathcal{X}_1, \ldots, \mathcal{X}_t], \end{aligned} \tag{7}$$

which meet the following conditions:

- $score(\mathcal{X}_i) < score(\mathcal{X}_{i+1}),\ \forall i \in \{1, \ldots, t-1\}$,
- $DSeq_i$ is a subsequence of $DSeq_{i+1}$,
- $BSet_i \subseteq BSet_{i+1}$,
- $DSeq_t = \mathcal{D}$,
- $BSet_t = \mathcal{B}_a$.

The first condition has its source in the curriculum learning method, while the others have their source in the progressive training with overlap method.

Remark. Some examples of curricula (list of multi-dataset samples) are presented in Section IV-D.

3) The training algorithm: The function that describes the training method is described in Algorithm 1, where $model$ is the dual-task classifier, while $scheduleList$ is the list of training samples.

In Algorithm 1:

1) The parameter $model$ represents an instance of the class $DualClassifier$, which implements the dual task classifier, as presented in Fig. 1. The class constructor uses only two parameters: the number of classes for the age attribute, and for the gender attribute:

   $model \leftarrow DualClassifier(ageClasses, genderClasses)$

2) The parameter $scheduleList$ is a dictionary implementing the multi-dataset curriculum strategy. The

Algorithm 1 Training with Multi-Dataset Curriculum Strategy

```
Require: $model$, $scheduleList$
function TRAINCURRICULUM($model, scheduleList$)
    for $schedule \in scheduleList$ do
        for $(datasets, bins, epochs) \in schedule$ do
            for $dataset \in datasets$ do
                for $epoch \in epochs$ do
                    TRAINONEEPOCH(
                        $model, trainLoader$)
                end for
            end for
        end for
    end for
end function
```

algorithm that generates the dictionary is adaptive and does not require parameters (the datasets *MORPH-II*, *UTKFace* and *Adience* are predefined).

3) Each element in a training list has an additional element, $epochs$, which specifies the number of training epochs for that sample.

4) The parameter $trainLoader$ has the following meaning:

$$filteredDataset \leftarrow FilteredDataset(dataset, bins)$$
$$trainLoader \leftarrow DataLoader(filteredDataset)$$

The function TRAINONEEPOCH is briefly described in Algorithm 2.

Algorithm 2 Training One Epoch

```
Require: $model$, $trainLoader$
function TRAINONEEPOC($model, trainLoader$)
    for $(images, labels) \in trainLoader$ do
        $out \leftarrow model(images)$
        $\mathcal{L}_{age} \leftarrow$ ENTROPYLOSS($out_{age}, labels_{age}$)
        $\mathcal{L}_{gender} \leftarrow$ ENTROPYLOSS($out_{gender}, labels_{gender}$)
        $\mathcal{L} \leftarrow \mathrm{e}^{-p_{age}}\mathcal{L}_{age} + p_{age} + \mathrm{e}^{-p_{gender}}\mathcal{L}_{gender} + p_{gender}$
        BACKWARD($\mathcal{L}$)
    end for
end function
```

In Algorithm 2:

1) Loss values for age and gender are calculated using a *cross entropy loss* function to be consistent with classification tasks.

2) The total loss value is calculated using Eq. (4) and Eq. (5).

3) p_{age} and p_{gender} are two trainable scalar parameters, which are initialized to 0 and updated in the training phase at each epoch.

IV. Experiments and Results

A. Datasets

We evaluated several publicly available facial datasets, both controlled and uncontrolled, that allow for the estimation of both gender and age of individuals based on face images, in the field of multi-task learning. A controlled dataset is generated in a controlled environment with some limited variability during image capture, while an uncontrolled dataset involves high variability in real-life image capture [3].

As mentioned in Section III, we chose to use three widely used public domain facial datasets, *MORPH-II* [39], *UTKFace* [40] and *Adience* [41]:

- *MORPH-II* is the *most commonly used* dataset in literature. It is a controlled dataset with some environmental variability (the images are captured in a constrained environment).
- *UTKFace* is a controlled dataset with the *widest age range* (0 to 116 years). Similar to MORPH, the images in UTKFace are captured in a constrained environment.
- *Adience* is the *most challenging* (and uncontrolled) dataset [3].

A part of these datasets are used for training and validation, while the other part is used for testing.

1) UTKFace [40]: UTKFace is a large-scale facial dataset containing over 20,000 images, spanning a long age range from 0 to 116 years. In addition to age labels (which represent the exact age), the images in the dataset also contain binary gender labels (0 - Male, 1 - Female), as well as ethnicity labels. The images are 200×200 pixels in size and are aligned and cropped.

Some qualitative elements related to UTKFace:

- Images are generally front-facing.
- Lighting and background are generally consistent, with low variation.
- Label quality: because they are derived from filename, not manually verified, can contain noise, especially in age.
- UTKFace is more balanced than other datasets in terms of age range but it underrepresents elderly and infants

2) MORPH-II [39]: MORPH-II is a very large public dataset of non-celebrity individuals, containing approximately 55,000 facial images from over 13,000 subjects, with an age range of 16 to 77 years. The images have a variable size, often over 200×240, and they have three labels: exact age, gender (binary label), and race (black, white, others).

Qualitative elements related to MORPH-II:

- Images are frontal well-aligned.
- Lighting and background is controlled (studio-like) and uniform.
- Label quality is high, with metadata.
- MORPH-II has a strong gender imbalance ($\approx 85\%$ Male) and racial imbalance ($\approx 77\%$ African-American).
- Temporal data: It contains multiple images per person over time.

3) Adience [41]: Adience is a challenging dataset containing approximately 26,000 facial images of 2,284 subjects, collected from difficult real-world scenarios. Each image from the dataset is labelled with a gender class label (binary), and an age group class label. There are eight age groups (0–2, 4–6, 8–13, 15–20, 25–32, 38–43, 48–53, and 60+), and the distribution of face images across age group and gender class labels in Adience is shown in Table I.

TABLE I. THE AGE AND GENDER DISTRIBUTION IN ADIENCE

	0-2	4-6	8-13	15-20	25-32	38-43	48-53	60+	Total
Male	4.69%	5.60%	5.72%	4.52%	14.04%	7.65%	2.49%	2.75%	47.18%
Female	4.20%	7.58%	7.32%	5.59%	16.47%	6.48%	2.59%	2.59%	52.82%
Total	8.89%	13.18%	13.05%	10.11	30.52%	14.14%	5.08%	5.34%	100.00%

Qualitative elements related to Adience:

- Images unconstrained, non-frontal, real-world poses.
- High variation of lighting and background (wild dataset).
- Label quality images are human-labeled, but age groups, not exact ages. In addition, there are gaps between age groups.
- The distribution of labels is more balanced across genders and ages than MORPH-II, but still uneven between some age groups (as in Table I).

B. Preprocessing Operations

To achieve the research goals, several preprocessing steps were performed on the datasets.

1) Preprocessing file names in UTKFace dataset: : A step specific only to the UTKFace dataset, consisted of uniformly processing the image file names, to eliminate labeling noise (since labels are extracted from file names):

- Incorrect age labels: Some file names contain unrealistic ages due to estimation or annotation errors, and there is also a mismatch between the apparent age and the actual age written in the file name.
- Ambiguous gender labels: The labels are 0 for male and 1 for female, but some are mislabeled (especially for children or when gender is difficult to determine visually).
- Non-standard or corrupted filenames: Some images are mislabeled or don't follow the standard file naming pattern, and a few are even missing tags completely.

2) Face detection and cropping: To obtain a robust classifier, but also because there are large differences between datasets in terms of framing faces in images, before the training, evaluation and testing operations, the two operations must be performed.

3) Face alignment and resizing: The face alignment operation is useful for improving classification accuracy in face-based tasks, such as age and gender classification, because:

- It reduces variability, as faces in datasets may be skewed, rotated, or improperly centered.
- Improves feature learning, as aligned faces lead to better feature consistency across different images, facilitating convergence and generalization.

Image resizing is important because CNN-based backbones (such as EfficientNet-B4) accept images of predefined sizes (380 × 380 in the case of EfficientNet-B4).

Remark. Face detection, cropping, face alignment are operations included in the generic class used for loading custom datasets in the training, validation and testing operations.

C. Age Labeling Strategy and Data Augmentation

- Label binning for age
- Handling class imbalance

Since the training operation is performed using three datasets, a unified age labeling scheme will need to be designed. To achieve this, the following elements were taken into account:

- Adience is the only dataset that has age groups, the others use exact age.
- The age range for the three datasets differs: 0 to 116 for UTKFace, 16 to 77 for MORPH-II and 0 to 100 for Adience.
- The age ranges at Adience are not consecutive (for example, the range for the second group is 4 to 6, and for the third group is 8 to 13).
- The distribution of images for age is generally imbalanced for all datasets.

This scheme will have eight consecutive age groups, which will cover the entire age range between 0 and 100 years (in the UTKFace dataset there are very few images of people over 100 years old, so they can be included in the last group). We will denote the age groups as follows:

$$\begin{aligned} &g_1 = [0, c_1], \\ &g_i = [c_{i-1} + 1, c_i],\ i \in \{2, 3, \ldots, 7\}, \\ &g_8 = [c_7 + 1, 100]. \end{aligned} \tag{8}$$

Denoting the Adience age groups as,

$$[a_1, b_1], [a_2, b_2], \ldots, [a_8, b_8], \tag{9}$$

we will need to determine the values for the variables $c_1, \ldots, c_7$, so that:

1) Age ranges from Adience groups to be included in the new ranges.
2) The number of images in all datasets, for the eight age groups, should be as balanced as possible.

Because the values for the boundaries $c_1, \ldots, c_7$ are integers, this is an integer optimization problem. We use an age histogram function:

$$f(x, y) = \sum_{i=x}^{y} n(i),$$

where, $n(i)$ is a function that counts in all three datasets images with age i. In addition, we will denote by $f_i = f(x_i, y_i)$ the number of images in group g_i [Eq. (8)].

The optimization problem can be formulated as following:

Problem 1: Given the Adience age groups as in Eq. (9), and the new age groups as in Eq. (8), find all integer values for the variables $c_1, \ldots, c_8$ that minimize the mean squared deviation of group sizes:

$$\frac{1}{8}\sum_{i=1}^{8}(f_i - \overline{f})^2,$$

subject to the following constraints:

$$c_i \in [b_i, a_{i+1} - 1], \ \forall i \in \{1, \ldots, 7\}.$$

This discrete optimization problem does not always have a unique solution, due in particular to the age distribution of the datasets. This is not a problem, because there is no need for just one canonical solution. In conclusion, we will choose the first solution, if there are multiple possible solutions.

In this case we use the following solution:

$$(c_1, \ldots, c_7) = (2, 6, 14, 23, 34, 45, 59),$$

noindent which means that the age groups are as follows:

$$0-2, 3-6, 7-14, 15-23, 24-34, 35-45, 46-59, 60-100.$$

Even after finding the solution, the age groups still remain unbalanced, as seen in Fig. 2, due to age group constraints in the Adience dataset.

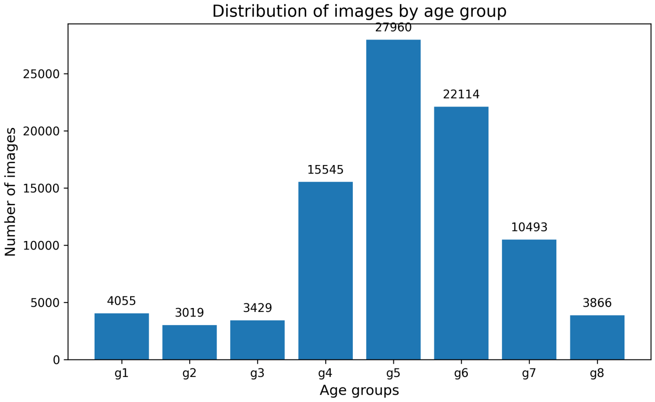

Fig. 2. Distribution of images in datasets by age groups.

To further reduce the imbalance, two operations will be performed:

- Randomly removing images from those age groups that have too many images.
- Generating new images in age groups that have too few images.

The generation of new images is based on augmentation operations, such as: rotation, skewing, and random distortion. In addition, these methods can enhance the dataset without introducing distortions that degrade classification accuracy.

Above the median value, drop-out operations will be performed, and below this value, augmentation operations. After performing these operations, the histogram of images by age group is shown in Fig. 3.

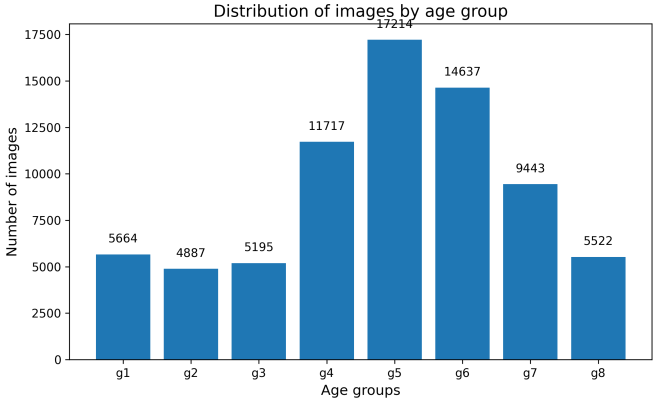

Fig. 3. Distribution of images in datasets by age groups after augmentation2.

The $\frac{1}{9}$ ratio has been reduced to $\frac{1}{3.5}$. To avoid overfitting, we do not believe this ratio should be reduced further. When necessary, a weighted sampling method will be used.

D. Implementing a Multi-Dataset Curriculum Strategy

To implement a *multi-dataset curriculum strategy*, we use the following elements:

- A multi-dataset sample is represented by a list of samples from a list of datasets, related to a list of age bins, as described in Eq. (6).
- A score related to a a multi-dataset sample specifies the training difficulty of that sample, which comes from:
  - The quality of images from each dataset
  - The degree of imbalance of age bins
- A training schedule is a list of multi-dataset samples ordered by their score, as in Eq. (7).

Since we are not yet ready to present an adaptive algorithm that automatically learns a multi-dataset curriculum, regardless of the list of datasets used for training, we will describe the static creation of such a strategy for the three datasets used in this research. Such an algorithm will be the subject of a future study.

According to the quantitative elements of the datasets presented in Section IV-A, the following three difficulty criteria

TABLE II. Datasets Difficulty Criteria

Image quality	Age bins imbalance	Gender imbalance
UTKFace	Adience	UTKFace, Adience
MORPH-II	UTKFace	MORPH-II
Adience	MORPH-II	-

are presented in Table II, in increasing order of difficulty for each dataset:

Remark. The most important difficulty criterion is *Gender imbalance*, because the dual-task model learns gender first, and then age. The difficulty order is as follows:

- *Gender imbalance*: important since the dual-task model depends on gender output for age prediction.
- *Image quality*: impacts feature extraction and prediction robustness.
- *Age imbalance*: makes learning certain age bins harder.

Considering the above remarks and the difficulty criteria, as shown in Table II, below is an ordered list of multi-dataset samples:

1) Multi-dataset sample 1 (Gender-balanced, medium-age, decent quality):
`{'datasets': ['utkface', 'adience'], 'bins': [3, 4, 5]}`
2) Multi-dataset sample 2 (Gradual age complexity increase):
`{'datasets': ['utkface', 'adience'], 'bins': [3, 4, 5, 6, 7]}`
3) Multi-dataset sample 3 (Full age distribution):
`{'datasets': ['utkface', 'adience'], 'bins': [0, 1, 2, 3, 4, 5, 6, 7]}`
4) Multi-dataset sample 4 (Incorporate harder dataset gradually):
`{'datasets': ['utkface', 'adience', 'morph2'], 'bins': [3, 4, 5]}`
5) Multi-dataset sample 5 (Adds aging extremity):
`{'datasets': ['utkface', 'adience', 'morph2'], 'bins': [3, 4, 5, 6, 7]}`
6) Multi-dataset sample 6 (Full complexity and diversity):
`{'datasets': ['utkface', 'adience', 'morph2'], 'bins': [0, 1, 2, 3, 4, 5, 6, 7]}`

Remarks:

- The previous samples were written as dictionaries in the *Python* language.
- Instead of the age groups $g_1, \ldots, g_8$, were used age bins: $[0, 1, \ldots, 7]$.

E. Results

In this section, the results of experiments performed on the test parts of the UTKFace, MORPH-II, and Adience datasets are presented.

The proposed method is implemented in PyTorch. The training was performed on the Google Colab framework, with an NVIDIA Tesla T4 graphics card with 16 GB of GPU memory. The maximum number of epochs was set to 45, but an early stopping mechanism was used to avoid overfitting. A dynamic learning rate was used via the ReduceLROnPlateau learning rate scheduler in PyTorch. The total training time was approximately 8 hours.

The tests were performed on a Windows PC with an Intel Core Ultra 7, 16-core, 5 GHz processor and 32 GB of CPU memory. The total number of test images in the 3 datasets is approximately 9000. Using a batch size of 64, the total testing time was approximately 5 minutes.

The *accuracy* is used to measure the performance of the model, both for age and gender estimation. The accuracy metric is calculated based on the true positive (TP), false positive (FP), true negative (VN), and false negative (FN) values, as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}.$$

Table III presents the estimation performance results for gender and age for all three datasets.

TABLE III. Gender and Age Accuracy for All Datasets

	Gender	Age
Adience	92.25	73.12
MORPH-II	97.92	77.17
UTKFace	97.74	69.83

It is observed that, although the MORPH-II dataset was introduced last in the training process and has a strong gender imbalance, the classification performances are the best in this case (both in gender and age). Although it is the most gender-unbalanced, the image quality is better than other datasets, which makes the learning process for images in MORPH-II better.

Regarding the other two datasets, Adience and UTKFace, the results are different for gender and age: the classification accuracy for age is better in Adience, while the accuracy for gender is better in UTKFace.

This can be explained for the following reasons:

- Although the images in Adience are collected from difficult real-world scenarios, the image quality is better than in the case of UTKFace, and the image sizes are quite large, over 200×200 (most are around 600×600).
- In UTKFace, all images are 200×200 in size and are aligned and cropped. But because they need to be scaled to the size of 380×380 (which is the input size for EfficientNet-B4) the scaling operation introduces noise.

In conclusion, age estimation for the UTKFace dataset is more difficult than for the other datasets.

In the following, we will compare the results of the proposed method with state-of-the-art works on facial gender and age group estimation.

Table IV presents the results of estimating the performance of the proposed method on the Adience dataset and compares this with other proposed approaches based on deep learning.

TABLE IV. Gender and Age Accuracy for Adience Dataset

Method	Gender accuracy	Age accuracy
Zhang et al., 2017 [42]	93.24	66.74
Duan et al., 2018 [43]	88.20	52.30
Gurnani et al., 2019 [44]	91.80	62.11
Khan et al., 2020 [45]	**93.60**	69.40
Garain et al., 2021 [46]	81.80	66.10
Saha et al. 2023 [47]	84.94	**75.10**
Proposed method	92.25	73.12

The methods specified in Table IV use the 8 age groups as defined in the Adience dataset (0–2, 4–6, 8–13, 15–20, 25–32, 38–43, 48–53, and 60+).

Table V presents comparisons of performance estimation on the MORPH-II dataset, while Table VI presents comparisons of performance estimation on the UTKFace dataset.

TABLE V. Gender and Age Accuracy for MORPH-II Dataset

Method	Gender accuracy	Age accuracy
Guo et al., 2014, [48]	**98.50**	70.00
Han et al., 2015 [49]	97.60	77.40
Wang et al., 2017 [50]	98.00	**85.30**
Khan et al., 2020 [45]	96.70	75.00
Han et al., 2018 [51]	81.80	66.10
Proposed method	97.92	77.17

TABLE VI. Gender and Age Accuracy for UTKFace Dataset

Method	Gender accuracy	Age accuracy
Krizhevsky et al., 2017, [52]	85.10	55.96
Huang et al., 2017 [53]	87.28	59.22
Das et al., 2018 [54]	**98.23**	**70.10**
Yuan et al., 2024 [55]	90.91	64.74
Proposed method	97.74	69.83

The vast majority of the methods specified in Table VI use 7 age groups: a) baby: 0 to 3 years, b) child: 4 to 12 years, c) teenagers: 13 to 19 years, d) young: 20 to 30 years, e) adult: 31 to 45 years, f) middle aged: 46 to 60 years, and g) senior: 61 and above years.

It is observed that, in all three tables, the classification results of our method do not exceed the best results of the other proposals, but the following elements should be noted:

- The classification accuracy values for ours method are close to the best results of the other methods, both for gender and age.
- Related the classification accuracy for age groups, the values are not always conclusive, because each proposal had its own age grouping scheme. With the exception of the Adience dataset, many other methods use between 3 and 6 age groups, while our method used 8 age groups, which led to a decrease in accuracy.
- Each of the other proposals used a single dataset for training and testing, while our proposal used all 3 datasets.

One of the goals of our method is to create a robust classifier for the gender and age of individuals in facial images, which means that testing the classifier on unknown images should have high accuracy, both for estimating age and gender.

F. Ablation Study

There are two important factors in our proposal. One is dynamic learning of task weighting when calculating the total loss. The other is the use of multi-dataset curriculum strategy in the training phase. We investigate the effects of the two factors on the accuracy of classification by gender and age.

1) Dynamic learning of task weighting: In multi-task learning, the total loss must be determined based on the partial losses in the task, as presented in Eq. (1). Usually the total loss is simply the sum of partial task losses:

$$\mathcal{L} = \sum_i \mathcal{L}_i.$$

Another approach is to manually determine normalized weights w_i, depending on the difficulty of the tasks:

$$\mathcal{L} = \sum_{i=1}^{n} w_i \mathcal{L}_i,$$

such that $\sum_{i=1}^{n} w_i = 1$.

We used a modified verison of the proposal from [11],

$$\mathcal{L} = \sum_{i=1}^{n} \mathrm{e}^{-\log(\sigma_i^2)} \mathcal{L}_i + \log(\sigma_i^2) \tag{10}$$

As described in Eq. (3), Eq. (4) and Eq. (5), which learns the log of the variance, $\log(\sigma_i^2)$, for each classification task i.

The results shown in Tables IV, V and VI use the Eq. (10), for $n = 2$ [or similarly, Eq. (5)].

To see the importance of using dynamic learning of task weighting, we then retrained of the dual classifier. We used the multi-dataset curriculum strategy in the training part, but in Algorithm 2, we used the following two variants for calculating the total loss value:

$$\mathcal{L} = \mathcal{L}_{age} + \mathcal{L}_{gender}, \tag{11}$$

and:

$$\mathcal{L} = 0.6 \cdot \mathcal{L}_{age} + 0.4 \cdot \mathcal{L}_{gender}. \qquad (12)$$

Remark. In Eq. (12), the weight for age classification is higher than for gender classification, because this task is more difficult.

In testing, the values obtained for accuracy were quite close when Eq. (11) and Eq. (12) were used. The differences were, however, greater between using Eq. (11) [or Eq. (12)] and Eq. (5).

Table VII presents the classification accuracy results for gender and age when using both the original Eq. (5), and Eq. (11) to calculate the total loss. An increase of approximately 2 per cent in accuracy is observed comparing the use of Eq. (11) with the use of Eq. (5).

TABLE VII. Comparison of Model Performance Depending on the Calculation Method of Total Loss

Dataset	Total loss determination	Gender	Age
Adience	Equation 5	**92.25**	**73.12**
	Equation 11	91.27	71.65
MORPH-II	Equation 5	**97.92**	**77.17**
	Equation 11	97.23	75.31
UTKFace	Equation 5	**97.74**	**69.83**
	Equation 11	94.68	66.25

2) Multi-dataset curriculum strategy: To determine the contribution of the training method, we will perform several experiments in which the dynamic learning of task weighting is used as in Eq. (5). Since there are many possible combinations, we will only test two combinations, which seem to be more important for the study.

- Since the accuracy values for the MORPH-II dataset are the highest, and it uses only age groups $g4$, $g5$ and $g6$ (bins $[3-5]$), the samples will be ordered according to the age group balance and the results obtained in the main ordering scheme. We will use the following ordered list of multi-dataset samples (will be noted as **list 1**):
  - (1.1) Core mid-age learning:
    ```
    {'datasets': ['morph2'], 'bins':
    [3, 4, 5]}
    ```
  - (1.2) Extended core mid-age:
    ```
    {'datasets': ['morph2', 'utkface',
     'adience'], 'bins': [3, 4, 5]}
    ```
  - (1.3) Young age focus:
    ```
    {'datasets': ['utkface', 'adience'
    ], 'bins': [0, 1, 2]}
    ```
  - (1.4) Elderly focus:
    ```
    {'datasets': ['utkface', 'adience'
    ], 'bins': [6, 7]}
    ```
  - (1.5) Mid-age re-focus:
    ```
    {'datasets': ['morph2', 'utkface',
     'adience'], 'bins': [3, 4, 5]}
    ```
  - (1.6) All datasets:
    ```
    {'datasets': ['morph2', 'utkface',
     'adience'], 'bins': [0, 1, 2, 3,
    4, 5, 6, 7]}
    ```
- Another variation of the curriculum is based on the uneven age distribution and imbalances of the Adience, UTKFace, and MORPH2 datasets. This time MORPH-II is no longer considered an advantageous dataset. We will denote this curriculum as **list 2**:
  - (2.1) Core mid-age learning:
    ```
    {'datasets': ['adience', 'utkface'
    , 'morph2'], 'bins': [3, 4, 5]}
    ```
  - (2.2) Young age specialization:
    ```
    {'datasets': ['adience', 'utkface'
    ], 'bins': [0, 1, 2]}
    ```
  - (2.3) Elderly age focus:
    ```
    {'datasets': ['adience', 'utkface'
    ], 'bins': [6, 7]}
    ```
  - (2.4) Balanced reintroduction:
    ```
    {'datasets': ['adience', 'utkface'
    ], 'bins': [0, 1, 2, 3, 4, 5, 6,
    7]}
    ```
  - (2.5) Fine tuning (all datasets):
    ```
    {'datasets': ['adience', 'utkface'
    , 'morph2'], 'bins': [3, 4, 5]}
    ```

Table VIII presents the classification accuracy results for gender and age when using three lists of multi-dataset samples in training phase: the original list and the two lists described above (**list 1** and **list 2**).

TABLE VIII. Comparing Model Performance based on the List of Multi-Dataset Samples

List of multi-dataset samples	Dataset	Gender	Age
Original list	Adience	92.25	73.12
	MORPH-II	97.92	77.14
	UTKFace	97.74	69.32
List 1	Adience	81.25	60.37
	MORPH-II	99.08	87.25
	UTKFace	85.32	58.45
List 2	Adience	92.75	73.82
	MORPH-II	97.12	76.87
	UTKFace	96.21	67.24

Based on the values in this table, the following remarks can be drawn:

- In the case of **list 1**, the accuracy values are extremely high for the MORPH-II dataset, but extremely low for the other two datasets. The model learns classifications for MORPH-II very well because this dataset is the first in the list of datasets. And because of this, the learnings for the other two datasets are at a disadvantage.
- The original list and **list 2** generate predictions with similar accuracy: the accuracy is slightly better in the case of the original list for the MORPH-II and

UTKFace datasets, and in the case of **list 2** the accuracy is slightly better for Adience.

In conclusion, evaluating the difficulty of multi-dataset samples is a complex and difficult operation. It is expected that in the future the manual design of a curriculum will be replaced with an adaptive algorithm that automatically learns a multi-dataset curriculum.

V. Conclusion

This study presents a deep multi-task learning approach for facial gender and age recognition. Facial features are extracted using a pre-trained model, EfficientNet-B4, which has a balanced ratio between accuracy and computational cost. A two-branch architecture is proposed, where the output of the gender branch is fed into the age estimation branch to improve age prediction.

Besides the mechanism for injecting gender information into the age branch of the dual-task model, which is in the architectural part of the proposal, the other contributions of the proposal reside mainly in the training part of the dual classifier.

For robust classification, instead of training the model on a single dataset, three of the most well-known datasets were used simultaneously and a unified strategy for determining age groups was proposed for all datasets. To improve training efficiency and final model performance, a multi-dataset curriculum strategy was proposed, which is based on the curriculum learning method for single datasets, which it extends to multiple datasets.

Another improvement to increase the accuracy of the estimation operations consists of implementing an adaptive method for determining the task-specific weights when determining the total loss. This uses two trainable scalar parameters, representing the logarithm of the variance of each task, which are used to calculate the total loss.

The experimental study conducted on three datasets used (Adience, MORPH-II and UTKFace) validated our contributions. Comparing the results of the proposed method with the state-of-the-art works on gender and age estimation, the classification accuracy values of our method are close to the best results of the other methods, both for gender and age. These results are valid for all datasets used, which demonstrates the robustness of the proposed dual classifier.

In its current form, the proposed method has some limitations and constraints, which could be remedied in the future:

- The multi-dataset curriculum method, although theoretically described, relies on manual curriculum creation.
- Optimal generation of age groups is done automatically, but it is based on predefined age groups in the Adience dataset.

Future research are in several directions:

- Developing an adaptive algorithm that automatically learns a a multi-dataset curriculum, based on loss values, instead of a static and manually developed one.
- Development of an adaptive algorithm that automatically determines age groups for the datasets used, based only on the annotation .csv files.
- Developing a similar lightweight model (e.g. based on a pre-trained model such as EfficientNet-B0), which can be transformed into a model for iOS applications.

References

[1] P. Terhorst, N. Damer, and A. K. F. Kirchbuchner, “Suppressing Gender and Age in Face Templates Using Incremental Variable Elimination,” in *2019 International Conference on Biometrics (ICB)*, Crete, Greece, 2019, pp. 1–8.

[2] M. Benkaddour, “CNN Based Features Extraction for Age Estimation and Gender Classification,” *Informatica*, vol. 45, no. 5, pp. 697–703, 2021.

[3] S. Gupta and N. Nain, “Review: Single attribute and multi attribute facial gender and age estimation,” *Multimedia Tools and Applications*, vol. 82, pp. 1289–1311, 2023.

[4] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A unified embedding for face recognition and clustering,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, X. Xie, M. Jones, and G. Tam, Eds., Boston, MA, USA, 2015, pp. 815–823.

[5] O. Parkhi, A. Vedaldi, and A. Zisserman, “Deep Face Recognition,” in *Proceedings of the British Machine Vision Conference (BMVC)*, X. Xie, M. Jones, and G. Tam, Eds., Swansea, UK, 2015, pp. 41.1–41.12.

[6] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. Berg, and L. Fei-Fei, “ImageNet Large Scale Visual Recognition Challeng,” *International Journal of Computer Vission*, vol. 115, pp. 211—252, 2015.

[7] H. Liao, L. Yuan, M. Wu, L. Zhong, G. Jin, and N. Xiong, “Face Gender and Age Classification Based on Multi-Task, Multi-Instance and Multi-Scale Learning,” *Applied Sciences*, vol. 12, no. 23, p. 12432, 2023.

[8] M. Tan and Q. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” in *International Conference on Machine Learning (ICML)*, Long Beach, California, USA, 2019, pp. 6105–6114.

[9] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, “A review of convolutional neural networks in computer vision,” *Artificial Intelligence Review*, vol. 57, p. 99, 2024.

[10] I. Aruleba and S. Viriri, “Deep Learning for Age Estimation Using EfficientNet,” in *Advances in Computational Intelligence. IWANN 2021*, ser. Lecture Notes in Computer Science, I. Rojas, G. Joya, and A. Catal, Eds. Springer, 2021, vol. 12861, pp. 407–419.

[11] R. Cipolla, Y. Gal, and A. Kendall, “Multi-task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 7482–7491.

[12] P. Soviany, R. Ionescu, P. Rota, and N. Sebe, “Curriculum Learning: A Survey,” *International Journal of Computer Vision*, vol. 130, pp. 1526–1565, 2022.

[13] J. Xing, K. Li, W. Hu, C. Yuan, and H. Ling, “Diagnosing deep learning models for high accuracy age estimation from a single image,” *Pattern Recognition*, vol. 66, pp. 106–116, 2017.

[14] J. Wan, Z. Tan, G. Guo, S. Li, and Z. Lei, “Auxiliary demographic information assisted age estimation with cascaded structure,” *IEEE Transactions on Cybernetics*, vol. 48, no. 9, pp. 2531–2541, 2018.

[15] K.-H. Liu and T.-J. Liu, “A Structure-Based Human Facial Age Estimation Framework Under a Constrained Condition,” *IEEE Transactions on Image Processing*, vol. 28, no. 10, pp. 5187–5200, 2019.

[16] Y. Kong, L. Liu, J. Wang, and D. Tao, “Adaptive Curriculum Learning,” in *EIEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, 2021, pp. 5047–5056.

[17] A. Pentina, V. Sharmanska, and C. Lampert, “Curriculum learning of multiple tasks,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 5492–5500.

[18] N. Sarafianos, T. Giannakopoulos, C. Nikou, and I. A. Kakadiaris, “Curriculum Learning for Multi-task Classification of Visual Attributes,” in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, Austin, Texas, SUA, 2017, pp. 2608–2615.

[19] 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.

[20] K. Ito, H. Kawai, T. Okano, and T. Aoki, "Age and Gender Prediction from Face Images Using Convolutional Neural Network," in *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Honolulu, HI, USA, 2018, pp. 7–11.

[21] B. Abirami, T. Subashini, and V. Mahavaishnavi, "Gender and age prediction from real time facial images using CNN," *Materials Today: Proceedings*, vol. 33, no. 7, pp. 4708–4712, 2020.

[22] P. Smith and C. Che, "Transfer Learning with Deep CNNs for Gender Recognition and Age Estimation," in *2018 IEEE International Conference on Big Data*, Seattle, WA, USA, 2018, pp. 2564–2571.

[23] V. Sheoran, S. Joshi, and T. Bhayani, "Age and Gender Prediction Using Deep CNNs and Transfer Learning," in *Computer Vision and Image Processing. CVIP 2020*, ser. Communications in Computer and Information Science, S. Singh, P. Roy, B. Raman, and P. Nagabhushan, Eds. Springer, 2021, vol. 1377, pp. 293–304.

[24] A. Abdolrashidi, M. Minaei, E. Azimi, and S. Minaee, "Age and Gender Prediction From Face Images Using Attentional Convolutional Network," 2020. [Online]. Available: https://arxiv.org/abs/2010.03791

[25] S. Ruder, "An overview of multi-task learning in deep neural networks," 2017. [Online]. Available: https://arxiv.org/abs/1706.05098

[26] L. Liu, Y. Li, Z. Kuang, J.-H. Xue, Y. Chen, W. Yang, Q. Liao, and W. Zhang, "Towards Impartial Multi-task Learning," in *9th International Conference on Learning Representations, ICLR 2021*, Austria, Virtual Event, 2021.

[27] S. Li, Z.-Q. Liu, and A. B. Chan, "Heterogeneous multi-task learning for human pose estimation with deep convolutional neural network," *International Journal of Computer Vision*, vol. 113, no. 1, pp. 19–36, 2015.

[28] Z. Zhang, P. Luo, C. Loy, and X. Tang, "Learning Deep Representation for Face Alignment with Auxiliary Attributes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 113, no. 1, pp. 918–930, 2016.

[29] M. Guo, A. Haque, D. Huang, S. Yeung, and L. Fei-Fei, "Dynamic Task Prioritization for Multitask Learning," in *CComputer Vision. ECCV 2018*, ser. Lecture Notes in Computer Science, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Springer, 2018, vol. 1122, pp. 282–299.

[30] Z. Chen, V. Badrinarayanan, C.-Y. Lee, and A. Rabinovich, "Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks," in *International Conference on Machine Learning, ICML 2018*, Stockholm, Swede, 2018, pp. 794–803.

[31] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proceedings of the 26th Annual International Conference on Machine Learning (ICML)*, Montreal, Quebec, Canada, 2009, pp. 41–48.

[32] G. Hacohen and D. Weinshall, "On the power of curriculum learning in training deep networks," in *Proceedings of the 36th International Conference on Machine Learning (ICML)*, vol. 97, 2019, pp. 2535–2544.

[33] S. Sinha, A. Garg, and H. Larochelle, "Curriculum by smoothing," in *Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS)*, vol. 97, Vancouver, BC. Canada, 2020, pp. 21 653–21 664.

[34] S. Chaudhry and A. Sharma, "Data Distribution-Based Curriculum Learning," *IEEE Access*, vol. 12, pp. 138 429–138 440, 2024.

[35] S. Fitzmaurice and H. Maibach, "Gender Differences in Skin," in *Textbook of Aging Skin*, M. Farage, K. Miller, and H. Maibach, Eds. Springer, Berlin, Heidelberg, 2010, pp. 282–299.

[36] S. Dammak, H. Mliki, and E. Fendri, "Gender effect on age classification in an unconstrained environment," *Multimedia Tools and Applications*, vol. 80, pp. 28 001–28 014, 2021.

[37] F. Benzing, "Unifying Importance Based Regularisation Methods for Continual Learning," in *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics (PMLR)*, 2022, pp. 2372–2396.

[38] H. Fayek, L. Cavedon, and H. Wu, "Progressive learning: A deep learning framework for continual learning," *Neural Networks*, vol. 128, pp. 345–357, 2020.

[39] 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)*, Southampton, UK, 2006, pp. 341–345.

[40] 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)*, Honolulu, HI, USA, 2017, pp. 4352–4360.

[41] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2170–2179, 2014.

[42] K. Zhang, C. Gao, L. Guo, M. Sun, X. Yuan, T. Han, Z. Zhao, and B. Li, "Age Group and Gender Estimation in the Wild With Deep RoR Architecture," *IEEE Accesss*, vol. 5, pp. 22 492–22 503, 2017.

[43] M. Duan, K. Li, C. Yang, and K. Li, "A hybrid deep learning CNN–ELM for age and gender classification," *Neurocomputing*, vol. 275, pp. 448–461, 2018.

[44] A. Gurnani, K. Shah, V. Gajjar, V. Mavani, and Y. Khandhediyae, "SAF-BAGE: Salient Approach for Facial Soft-Biometric Classification - Age, Gender, and Facial Expression," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, HI, USA, 2019, pp. 839–847.

[45] K. Khan, M. Attique, R. Khan, I. Syed, and T.-S. Chung, "A Multi-Task Framework for Facial Attributes Classification through End-to-End Face Parsing and Deep Convolutional Neural Networks," *Sensors*, vol. 20, no. 2, p. 328, 2020.

[46] A. Garain, B. Ray, P. Singh, A. Ahmadian, N. Senu, and R. Sarkar, "GRA_Net: A Deep Learning Model for Classification of Age and Gender From Facial Images," *IEEE Access*, vol. 9, pp. 85 672–85 689, 2021.

[47] A. Saha, S. Kumar, and P. Nithyakani, "Age and Gender Prediction using Adaptive Gamma Correction and Convolutional Neural Network," in *International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, 2023, pp. 1–5.

[48] G. Guo and G. Mu, "A framework for joint estimation of age, gender and ethnicity on a large database," *Image and Vision Computing*, vol. 32, no. 10, pp. 761–770, 2014.

[49] H. Han, C. Otto, X. Liu, and A. Jain, "Demographic Estimation from Face Images: Human vs. Machine Performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 6, pp. 1148–1161, 2015.

[50] F. Wang, H. Han, S. Shan, and X. Chen, "Deep Multi-Task Learning for Joint Prediction of Heterogeneous Face Attributes," in *12th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2017)*, Washington, DC, USA, 2017, pp. 173–179.

[51] H. Han, A. Jain, F. Wang, S. Shan, and X. Chen, "Heterogeneous Face Attribute Estimation: A Deep Multi-Task Learning Approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 11, pp. 2597–2609, 2018.

[52] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 284–90, 2017.

[53] G. Huang, Z. Liu, L. V. D. Maaten, and K. Weinberger, "Densely Connected Convolutional Networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 2261–2269.

[54] A. Das, A. Dantcheva, and F. Bremond, "Mitigating Bias in Gender, Age and Ethnicity Classification: A Multi-task Convolution Neural Network Approach," in *European Conference on Computer Vision (ECCV)*, Munich, Germany, 2018, pp. 573–585.

[55] H. Yuan, Y. He, P. Du, and L. Song, "Multi-task learning using uncertainty to weigh losses for heterogeneous face attribute estimation," 2024. [Online]. Available: https://arxiv.org/abs/2403.00561