How to quantify gender bias in MORPH2, and which techniques can be used to mitigate potential bias generated by data

1st Kaan Gögcay
Master of Applied Intelligence
Hogeschool van Amsterdam
Amsterdam, Netherlands

I. Introduction

Facial age estimation is an important task in computer vision, with applications ranging from biometric security to age verification systems. Deep learning methods, especially Convolutional Neural Networks (CNNs), have achieved high accuracy in predicting age from facial images. However, the quality and balance of the datasets used for training these models strongly influence their performance and fairness.

The MORPH II dataset is widely used for facial age estimation research, but it contains a significantly higher number of male images compared to female images. This gender imbalance can lead to biased models that perform better for men while under performing for women. Despite the frequent use of MORPH II in research, there is limited work quantifying gender bias or evaluating methods to mitigate it.

This research approach aims to investigate the extent of gender bias in the MORPH II dataset and explore techniques to reduce it. By analyzing dataset characteristics, training models on both imbalanced and balanced data, and testing bias mitigation methods, the study seeks to provide practical insights for developing models that try to reduce the bias between genders and try to get the results as equal as possible.

II. Research Problem

Facial age estimation using deep learning is an important task in computer vision. The MORPH II dataset, which is widely used for training and testing models, has many more male images than female images. This imbalance can cause gender bias, meaning that models may perform better for men and worse for women. So far, few studies have measured how much bias exists or how it affects model performance. Understanding and measuring this bias is important because it helps make AI systems more fair and guides better practices for preparing datasets and evaluating models.

In this study, the goal is to identify and test possible techniques to reduce gender bias in data sets. This will involve analyzing the data set to measure the imbalance, training models to compare performance between men and women, and evaluating different methods to mitigate bias. Ultimately, it will provide a reference resource for researchers and developers working with biased data sets.

III. Research Purpose

The purpose of this study is to investigate and reduce gender bias in the MORPH II data set for the estimation of facial age using deep learning. Specifically, the study will focus on quantifying differences in model performance between male and female subjects as a measurable variable.

The research will involve training Convolutional Neural Networks (CNNs) on the data set and evaluating model performance separately for male and female subjects. A set of appropriate quantitative metrics, such as Mean Absolute Error (MAE) or other error measures, will be used to assess differences in performance. Comparisons will focus on identifying patterns of gender bias and determining whether certain model configurations or mitigation techniques reduce these differences.

In addition, the study aims to evaluate techniques to mitigate bias, such as data resampling, augmentation, and fairness-based training methods. In doing so, the study intends to provide practical insights and recommendations for researchers and developers working with imbalanced data sets. This research is feasible given the availability of the MORPH II dataset and established CNN frameworks, and it contributes to the field by addressing fairness in AI and improving the reliability of facial age estimation models.

IV. Research Questions

Based on the research purpose, the following questions will guide this study:

1) What is the distribution of male and female images in the MORPH II dataset, and how is age represented within each gender?
2) How does gender imbalance in MORPH II influence the performance of age estimation models for male and female subjects?
3) What techniques are most effective in reducing gender bias in age estimation models?

V. Research Methods

This study uses a combination of quantitative research methods to answer the research questions, build models, and validate the findings. Each research question is approached with multiple techniques to ensure comprehensive results.

A. RQ1: What is the distribution of male and female images in the MORPH II dataset, and how is age represented within each gender?

To answer this question:

- **Exploratory Data Analysis (EDA):** Analyze the dataset to quantify gender and age distributions using summary statistics, percentages, and visualizations such as charts and histograms.

B. RQ2: How does gender imbalance in MORPH II influence the performance of age estimation models for male and female subjects?

To answer this question, the study will use the following methods:

- **Modeling with an imbalanced dataset:** Train a Convolutional Neural Network (CNN) on the imbalanced MORPH II dataset.
- **Modeling with a balanced dataset:** Train the same CNN architecture on a balanced version of the dataset, created using resampling or augmentation.
- **Comparison of results:** Evaluate and compare the performance of both models using metrics such as Mean Absolute Error (MAE) for male and female subjects to assess the effect of dataset imbalance on gender bias.

C. RQ3: Which techniques are most effective at reducing gender bias in age estimation models?

To answer this question, the study will use the following methods:

- **Identify bias mitigation techniques:** Review and select methods such as data resampling, augmentation, and fairness-aware training approaches.
- **Apply techniques:** Implement the selected methods on the dataset and train models accordingly.
- **Compare effectiveness:** Evaluate the models using metrics like Mean Absolute Error (MAE) for male and female subjects, and determine which techniques reduce gender bias most effectively.

VI. Research Planning

References

The following papers are relevant to this research and might be consulted throughout the study. These references are listed in the bibliography below. Note that my research won't be limited to these papers.

Dates	Tasks / Deliverables
26 Sept	Start on Research Approach and submit it
26–30 Sept	Collect and prepare MORPH II dataset; initial EDA (RQ1)
1–6 Oct	EDA on a balanced dataset
7-10 Oct	Prepare and present symposium presentation (block 1)
11-12 Oct	Train CNN on imbalanced dataset; evaluate by gender (RQ2)
13–14 Oct	Train CNN on balanced dataset; compare performance (RQ2)
15–19 Oct	Identify and implement bias mitigation techniques (RQ3)
20–24 Oct	Train models with mitigation; evaluate effectiveness (RQ3)
25–29 Oct	Analyze results; draft findings and recommendations
30 Oct–6 Nov	Draft research approach paper, peer or teacher review
7 Nov	Submission of research approach paper

TABLE I
Updated research planning and timeline for gender bias study in MORPH II.

References

[1] A. Puc, V. ˇStruc, and K. Grm. Analysis of race and gender bias in deep age estimation models. In 202028th European Signal Processing Conference (EU-SIPCO), 2021

[2] T. Feldman and A. Peake, "End-to-end bias mitigation: Removing gender bias in deep learning," 2021.

[3] Y. Keller, "Hierarchical Attention-based Age Estimation and Bias Estimation," Bar-Ilan University, 2021.

[4] T. Wongvorachan, O. Bulut, J. X. Liu, and E. Mazzullo, "A comparison of bias mitigation techniques for educational classification tasks using supervised machine learning, 2024.