Investigation into Ethnic Bias in Face Recognition Models

Motivation

With the rise of Artificial Intelligence, its usage in various fields has highlighted racial discrimination concerns that have not yet been addressed. It turns out that machine learning models fail to generalise their algorithms to data of non-white population. The literature suggests that it's due to the unbalanced datasets on which the mentioned models train. The discriminatory outcomes can be found practically everywhere - healthcare, surveillance systems, law enforcement etc.[1]. Cases of most interest to us include those regarding facial recognition. Those provoke a strong sense of injustice as they end in wrongful arrests, restricted access to services, and violations of privacy.

The aim of this project is to explore the underlying factors that contribute to the ethnic disparity caused by the machine learning models in face recognition.

Dataset Description

We will be using the VGG-Face2 Mivia Ethnicity Recognition (VMER) dataset[3] that includes over 3 million `jpg` face images and their corresponding ethnicity labels.
The images vary in pose, age, illumination and ethnicity. The dataset is approximately gender-balanced but in terms of race, it's highly unbalanced (look Figure 1).

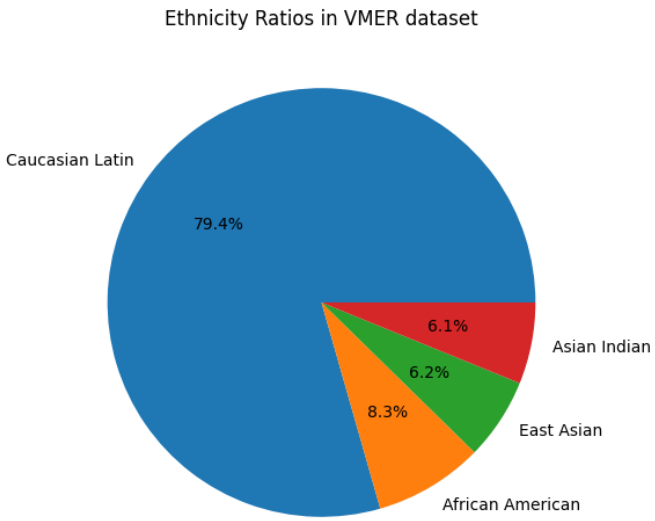

Figure 1: Ethnicity Ratios in VMER Dataset

The dataset has been downloaded from Kaggle and the labels in the XML format from MIVIA Laboratory Website.

Method

1. **Data Preprocessing**
   We transform the raw image data into algorithm-suitable format by normalising pixel values. We use min-max normalization to scale pixel values between 0 and 1 for consistency across images. Moreover, we adjust image dimensions to a uniform size.

2. **Data Exploration**
   We check for check for class imbalance, skewed distribution of images and the distribution of image sizes.

3. **Model Selection**
   To do the face recognition task, that is identify individuals from images, we use Convolutional Neural Networks (CNN).

4. **Model Training**
   The dataset is already split into training and validation data.

5. **Evaluation**
   We evaluate our model based on Receiver Operating Characteristic (ROC) Curve.

Our research will focus on analyzing subsets of the larger VGGFace2 dataset. In these subsets, we will have control over the proportions of the ethnicities, allowing us to systematically adjust the representation of each group. By applying our face recognition methods to datasets with varying ethnic compositions, we aim to compare the performance of the models across these different groups. The goal is to explore the trade off between representation and performance balance, determining how much representation of each ethnicity is required to prevent significant performance imbalances. This will help us better understand the threshold at which under-representation begins to affect the accuracy and fairness of the face recognition model for specific ethnic groups.

Related Research

The paper "VGGFace2: A dataset for recognising faces across pose and age"[2] analyzes the same dataset used in our project. The authors trained ResNet-50 and Squeeze-and-Excitation (SE) blocks, which are powerful convolutional neural network architectures, to improve face recognition performance, focusing on variations in pose and age. They used the dataset's diversity in image attributes, such as pose, lighting, and age, to train models that performed better than previous benchmarks on face recognition tasks. Their approach included both training from scratch and pre-training on MS-Celeb-1M before fine-tuning on VGGFace2. While their study primarily addressed technical aspects of face recognition, it provides a foundation for our project, which will specifically investigate ethnic bias in models trained on VGGFace2. Our work will focus on how these models handle faces of different ethnicities and explore methods to mitigate any discovered biases.

References

[1] J. Buolamwini and T. Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In S. A. Friedler and C. Wilson, editors, *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, volume 81 of *Proceedings of Machine Learning Research*, pages 77–91. PMLR, 23–24 Feb 2018.

[2] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, pages 67–74, 2018.

[3] A. Greco, G. Percannella, M. Vento, and V. Vigilante. Benchmarking deep network architectures for ethnicity recognition using a new large face dataset. *Machine Vision and Applications*, 2020.