Fairness-aware Dataset generation using Crowd-sourcing platform for Face Detection applications.

Dushyant Singh Udawat
NetID: ds35

1. Scope of the Project

Fairness in AI is an advancing field that is dedicated towards making AI/ML solutions unbiased with respect to attributes of individuals that they cannot control. This is specifically important in the case of Face detection algorithm because their large-scale deployment can put minority groups at a disadvantage [5]. While there have been several proposed solutions – Custom dataset creation for training [1], GAN-based attribute removal, usage of custom loss functions [6], and attribute-based threshold tuning [3]. We believe that creation of a fairness aware dataset will help advance the field because we will be tackling the problem at its root. Previously, Fairness based dataset has been prepared with respect to race, age, and gender [1], but further detailing is rare. In this work, we define more attributes for every face and attempt to create a more detailed dataset. We also use crowd-sourcing platform as we want to test their feasibility for fairness-based application and want to gauge the cost associated with data collection of such scale.

2. Methods

To understand the methods to be used during the independent study, it is necessary to understand the work already accomplished by research laboratory.

Completed Phases of the project –

1. Dataset Selection for Attribute generation – CelebA.
2. Crowdsourcing platform selection – Amazon Mechanical Turk
3. Definition of Framework for Data Collection.
4. Deployment of Assignment on Mechanical Turk
5. Data Cleaning
6. Training models to predict attributes with high probability for unseen images.

Phases to be completed and Techniques used (written in sub-bullets)–

1. Training models to predict attributes with high probability for unseen images.
    a. Object Localization, and
    b. Classification using CNN.
2. Using Trained models to predict attributes for all the images in CelebA dataset.
    a. Celeb Identity matching
    b. Prediction using the trained models.
3. Definition of sampling techniques for this data for multiple applications.
    a. This is a rather open-ended question as to how we can use the data in various ways so as to bring better performance for face detection algorithms and better gauge their fairness.
4. Use the dataset to compare the face detection algorithms in use today. - Retina Face, and yolo-v5-face.

a. Here, I will prepare a script which will take inputs of face attributes (like Skin, lips etc.) and a face detection model (yolo-v5-face, RetinaFace).
b. We will use certain metrics to gauge fairness as outlines in [2], some of those are False Non-Match rate, false match rate, general accuracy of face detection segregated into different attributes.

3. Final Deliverables

The final deliverables of the project will be in the form of

A. a codebase comprising of the following –

- **a. Trained models with reasonable test accuracy (July 5th)**
  These models will take in images as input and return the class of the attribute as output. For each attribute, there will be a separate classifier. The number of classifier models will depend on the number of attributes that have sufficient trainable data. I think we will find at least 6 such attributes.
- **b. A dataset where each image of CelebA dataset has defined attributes – Eyes, Nose, Hairs, Ears, Eyebrows, lips, chin etc. (July 5th)**
  This dataset should have image path, id and primary key and attribute classes corresponding to each entry in the primary key. If enough information can't be found by us regarding that attribute. It will be omitted from the file as column.
- **c. A Script to use the dataset to compare the face detection algorithms for their bias given the attribute. (July 31st)**
  This Script should take three inputs – data, face detection model, and attribute. Then it should display the accuracy, False positive rate, False Negative rate for that attribute segmented by different classes of that attribute. For example, for skin, the script will be able to tell the bias of a particular face detection model w.r.t light and dark skin.
- **d. The codebase will be properly documented with Readme files to understand all the operations. (July 31st)**
  This requirement is not quantizable, but Readme files should contain details about what each file is doing and what its purpose in the whole codebase is. All the functions in the codebase should have a description that helps the future readers of the database understand it better.

B. A written report that contains the following –

- **a. Results of the trained models on various attributes (Corresponding to 3.A.a) (Aug 05)**
- **b. Discussion regarding sampling techniques (of Dataset) for various applications. (Aug 05)**
  This is an open-ended question and will be answered after discussion with other members of the lab involved in the project.
- **c. Results show at least two face detection algorithm's biases with respect to three attributes. (Corresponding to 3.A.c) (Aug 05)**

References

1. Yang, Yu, and Aayush Gupta. "Enhancing fairness in face detection in computer vision systems by demographic bias mitigation." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 36. No. 1. 2022.
2. H. F. Menezes, A. S. C. Ferreira, E. T. Pereira and H. M. Gomes, "Bias and Fairness in Face Detection," 2021 34th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Gramado, Rio Grande do Sul, Brazil, 2021, pp. 247-254, doi: 10.1109/SIBGRAPI54419.2021.00041.
3. Philipp Terhörst, Jan Niklas Kolf, Naser Damer, Florian Kirchbuchner, Arjan Kuijper, "Post-Comparison Mitigation of Demographic Bias in Face Recognition Using Fair Score Normalization ", arXiv:2002.03592
4. Shervin Minaee, Ping Luo, Zhe Lin, Kevin Bowyer, "Going Deeper into Face Detection: A Survey ", arXiv:2103.14983
5. Elijah Mayfield, Michael Madaio, Shrimai Prabhumoye, David Gerritsen, Brittany McLaughlin, Ezekiel Dixon-Román, and Alan W Black. 2019. Equity beyond bias in language technologies for education. In Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications. 444–460
6. Zeyu Wang, Klint Qinami, Ioannis Christos Karakozis, Kyle Genova, Prem Nair, Kenji Hata, and Olga Russakovsky. 2020. Towards fairness in visual recognition: Effective strategies for bias mitigation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 8919–8928.