PRESENTER
Sepehr Dehdashtian

Utility-Fairness Trade-Offs and How to Find Them

Background

When building classification systems with demographic fairness considerations, there are **two objectives to satisfy**:
1) maximizing utility for the specific task and **2) ensuring fairness** w.r.t. a known demographic attribute. These objectives often compete, so optimizing both can lead to a trade-off between utility and fairness.

Questions We Answer

1) What are the optimal trade-offs between utility and fairness?

2) How can we numerically quantify these trade-offs from data for a desired prediction task and demographic attribute of interest?

Trade-Offs Definitions

Definition 1. *Data Space Trade-Off* (DST)

$$f_{\lambda}^{DST} := \underset{f \in \mathcal{H}_X}{\arg\inf} \Big\{ (1-\lambda) \inf_{g_Y \in \mathcal{H}_Y} \mathbb{E}_{X,Y} \left[\mathcal{L}_Y \left(g_Y \left(\boldsymbol{f}(X) \right), Y \right) \right] + \lambda \, \mathrm{Dep}\big(\boldsymbol{f}(X), S | Y = y\big) \Big\}, \quad 0 \leq \lambda < 1$$

Definition 2. *Label Space Trade-Off* (LST)

$$Z_{\lambda}^{LST} := \underset{Z \in L^2}{\arg\inf} \Big\{ (1-\lambda) \inf_{g_Y \in \mathcal{H}_Y} \mathbb{E}_Y \Big[\mathcal{L}_Y \big(g_Y(Z), Y \big) \Big] + \lambda \, \mathrm{Dep}\big(Z, S | Y = y\big) \Big\}, \quad 0 \leq \lambda < 1$$

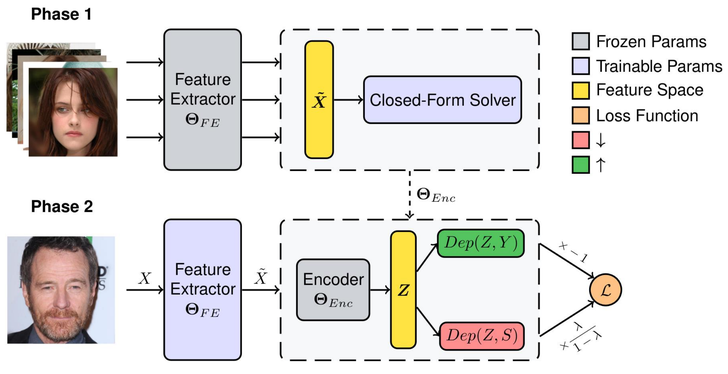
Fairness Criteria

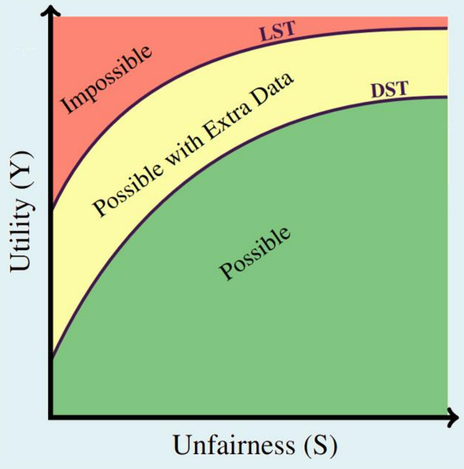
1) Demographic Parity (DP)

2) Equalized Opportunity (EO)

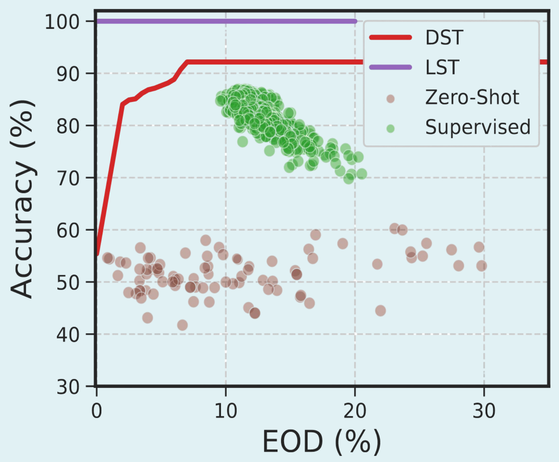
3) Equality Of Odds (EOO)

U-FaTE





Data Space Trade-Off (**DST**) and Label Space Trade-Off (**LST**) divide the utility (e.g., accuracy) versus fairness space into **three regions**.



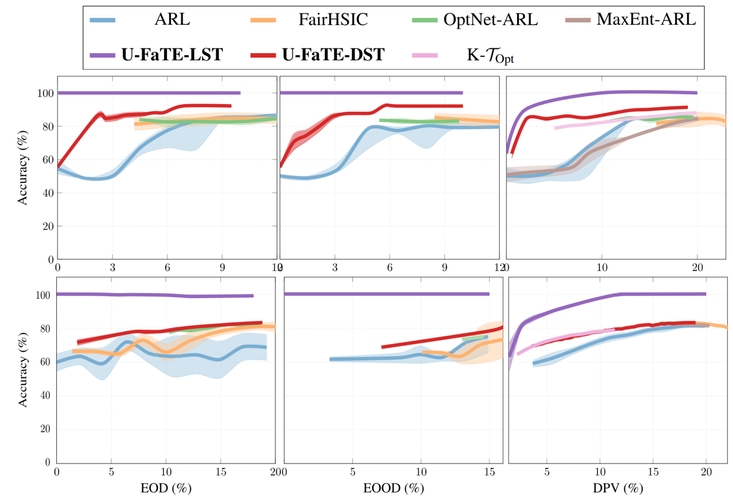
We empirically estimate **DST** and **LST** on **CelebA** and evaluate the utility (high cheekbones) and fairness (**gender & age**) of over **100 zero-shot** and **900 supervised models**.

Paper

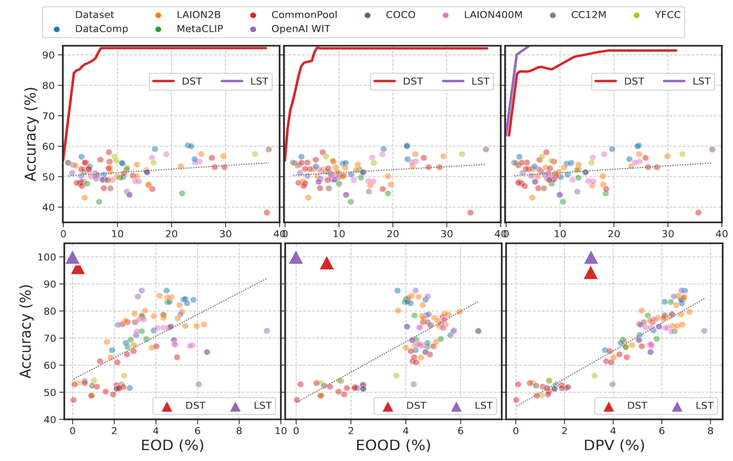
Sepehr Dehdashtian,
Bashir Sadeghi,
Vishnu Naresh Boddeti

CVPR
JUNE 17-21, 2024
SEATTLE, WA

How do FRL Methods Compare?



How fair are CLIP Models?



How fair are pre-trained image models?