Understanding the Bias in Deep Learning Based Gender Classification Systems

Anoop Krishnan,* Ali Almadan
Faculty: Ajita Rattani
Department of Electrical Engineering and Computer Science, College of Engineering

Automated gender classification has important applications in many domains such as demographic research, law enforcement, online advertisements, as well as human computer interface. Research search questioned the fairness of this technology across gender and race. Specifically majority of the studies raised the concern of higher error rates of the race based gender classification system for dark skinned people like Afro American and for women. However to date, the majority of the studies were limited to African American and Caucasian only. There can be two reasons for this, since these classification systems are data driven, either these systems lack uniform data distribution of individual races (which leads to data imbalance), or the algorithm (here, we talk about machine learning algorithms) which is a black box. We tackled the data imbalance using a demographically balanced dataset (so far it is the best), Fairface for the study. It contains data (images) of male and female (so, our research is constrained to binary form), belongs to seven races, and are evenly distributed. The second problem, is what we are trying to tackle it down. What classification algorithms can be used or how the different classifications algorithms perform (architectural differences), when they are trained to classify images based on their gender. Here, we used deep learning based solution which involves state of the art technology for classification, estimation so on. Finally, even in this era, social justice, human rights, equality are violated. Since most of the technology in this period involves deep learning machine solutions, these machines should not violate equality, but that is not the case. That is where the relevance of this study stands.