
TEMPE
abc 15
ARIZONA
DEADLY CRASH WITH SELF-DRIVING UBER
abc 15
ARIZON

What was their algorithm trying to do?

- Primary goal: identify object
- Once object is identified, figure out its path.

Where did their algorithm fail? Why did the error happen?

Bias in Machine Learning

Sauman Das

Why is this topic important? Why must this problem be addressed NOW?

Importance of Addressing Bias

- Machine Learning research is growing rapidly → starting to be applied in more situations
- If we do not address this now, it will spread too rapidly before we can fix it.

What are specific examples of where machine learning can show signs of bias? What is the impact of the bias when deployed?

Some Cases where Bias Impacts the Model

- Computer Vision
  - Face Recognition
  - Uber Self-Driving Car Example
  - Medical Datasets
- Natural Language Processing
  - Gender/Racial Bias in word vectors

Nurse is Closer to Woman than Surgeon? Mitigating Gender-Biased Proximities in Word Embeddings

Vaibhav Kumar[1*] Tenzin Singhay Bhotia[1*] Vaibhav Kumar[1*] Tanmoy Chakraborty[2]

[1]Delhi Technological University, New Delhi, India
[2]IIIT-Delhi, India
[1]{kumar.vaibhav1o1, tenzinbhotia0, vaibhavk992}@gmail.com
[2]tanmoy@iiitd.ac.in

Uncovering and Mitigating Algorithmic Bias through Learned Latent Structure

Alexander Amini*†
Massachusetts Institute of Technology
Cambridge, MA

Ava P. Soleimany†
Harvard University
Boston, MA

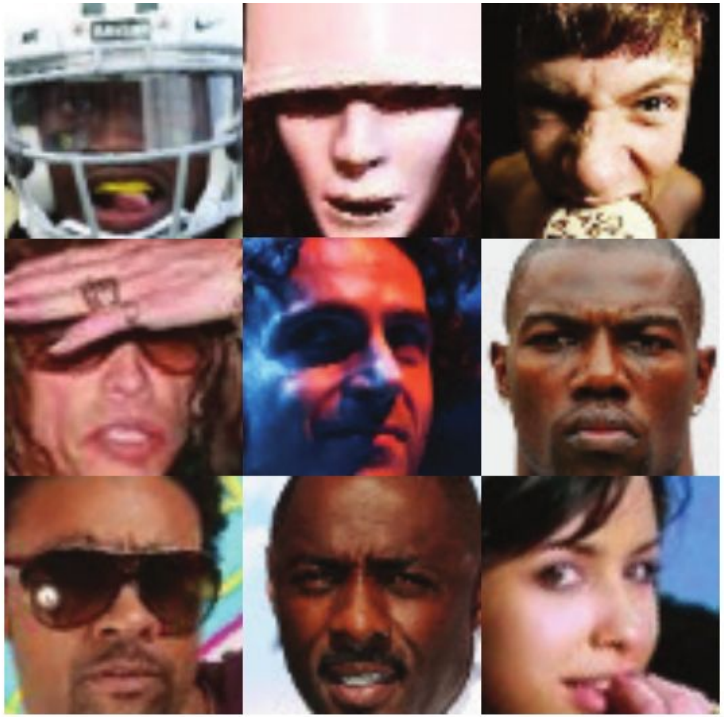
Wilko Schwarting
Massachusetts Institute of Technology
Cambridge, MA

Sangeeta N. Bhatia
Massachusetts Institute of Technology
Cambridge, MA

Daniela Rus
Massachusetts Institute of Technology
Cambridge, MA

What is wrong with this dataset?

It will fail on many images!



Current Research: How can we mitigate the bias in these applications?

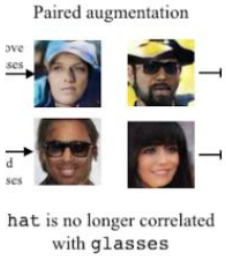
It is not easy to mitigate bias!

- A lot of new research is being developed in this area
- Princeton Visual AI Lab focuses mainly on representation and fairness!
- The impact of this research is really crucial.
  - Fixing this problem now will make problems much easier in the future

Princeton Visual AI Lab

Home People Research Teaching Awards Outreach

All Representative Annotation Diversity Fairness ImageNet Interaction Interpretability Language Objects Video



Fair Attribute Classification through Latent Space De-biasing

Vikram V. Ramaswamy, Sunnie S. Y. Kim and Olga Russakovsky.

Computer Vision and Pattern Recognition (CVPR), 2021.

[paper] [website] [code] [bibtex]

Evolving Graphical Planner: Contextual Global Planning for Vision-and-Language Navigation

Zhiwei Deng, Karthik Narasimhan and Olga Russakovsky.

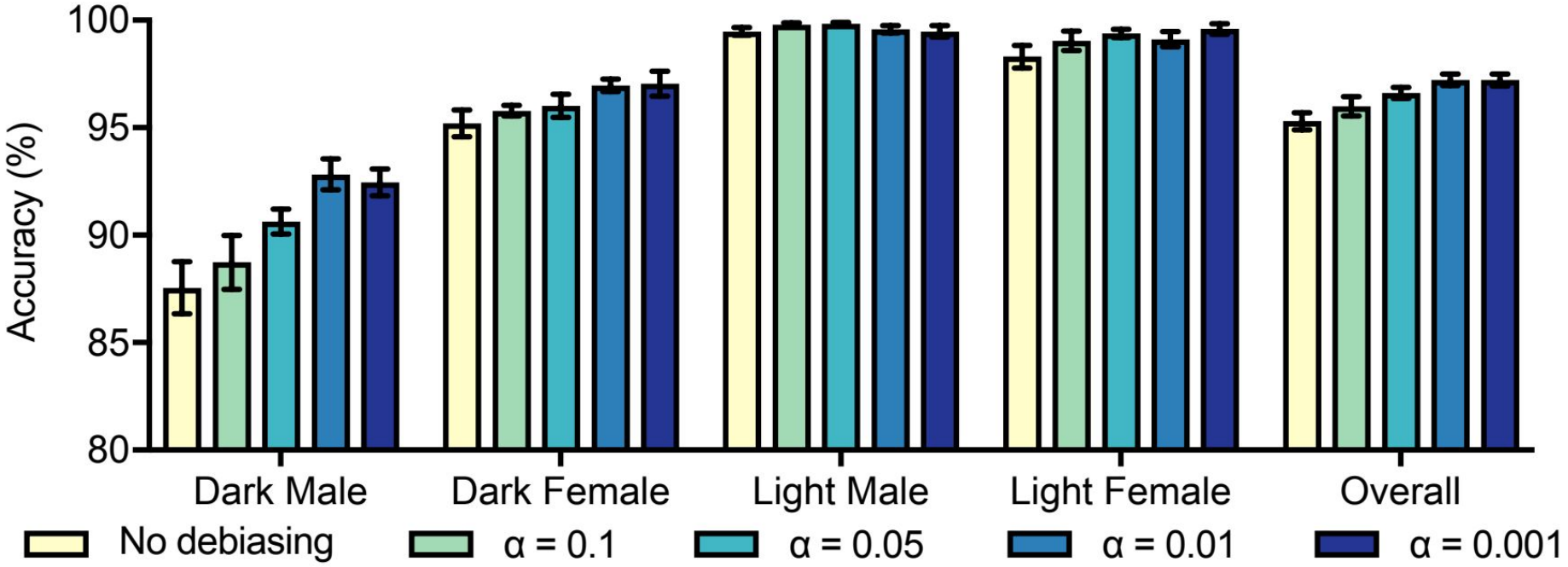
Neural Information Processing Systems (NeurIPS), 2020.

[paper] [bibtex]

REVISE: A Tool for Measuring and

Crowdsourcing in Computer Vision

Results of their approach → intelligent batch sampling



GANs for Debiased Data: Latent Space Perturbing

- CVPR 2021
- Datasets tend to show certain correlations
  - Ex) People wearing hats also wear glasses