

Computer Vision and Robotic Learning

MAE 340 Junior Independent Work Proposal

Matthew Coleman

February 2022

1 Introduction

In computer vision, it is common for researchers to provide data to a model in the form of labeled or annotated images or videos. While this works well in practice for accomplishing tasks such as image classification, detection, segmentation, etc., such datasets may contain human bias in image selection or annotation, leading to models that perpetuate that bias in their real-world implementations.[1] In order to combat this drawback, some computer vision tasks focus instead on learning directly from the world, rather than from humans. An example could be a model which is meant to predict the next frame of a video, being trained by the actual next frame of the video as the ground truth. While this setting represents an improvement over feeding in labeled data, such models also suffer in that they are unable to compute predictions in real-time or predict images too far into the future.

2 Abstract

3 Procedure

References

- [1] D. Zhao, A. Wang, and O. Russakovsky. Understanding and Evaluating Racial Biases in Image Captioning. *arXiv e-prints*, art. arXiv:2106.08503, June 2021.