Diverse Demographic Data Generation

Increases Zero-Shot Accuracy for Image Classification

Javon Hickmon - University of Washington

INTRODUCTION

- Multimodal models require more high-quality data with rich cross-modal representations to perform well compared to their unimodal counterparts.
- This is compounded by the fact that existing public datasets are strongly biased towards Caucasian faces, while other races are typically underrepresented.
- By generating multimodal data for a broad range of demographics, I aim to increase the accuracy and equity of such models.

OBJECTIVE

1. Increase the top-1 accuracy of CLIP ViT-L/14 on the IdenProf dataset.
2. Improve the per-class accuracy of CLIP ViT-L/14 on the IdenProf dataset, specifically for classes containing underrepresented demographics.
3. Understand which demographics help to improve classification accuracy.

Ex. PROFESSION TEXTS

Profession: A photo of a doctor

Race 7: A photo of a white doctor

Race 4: A photo of a white doctor

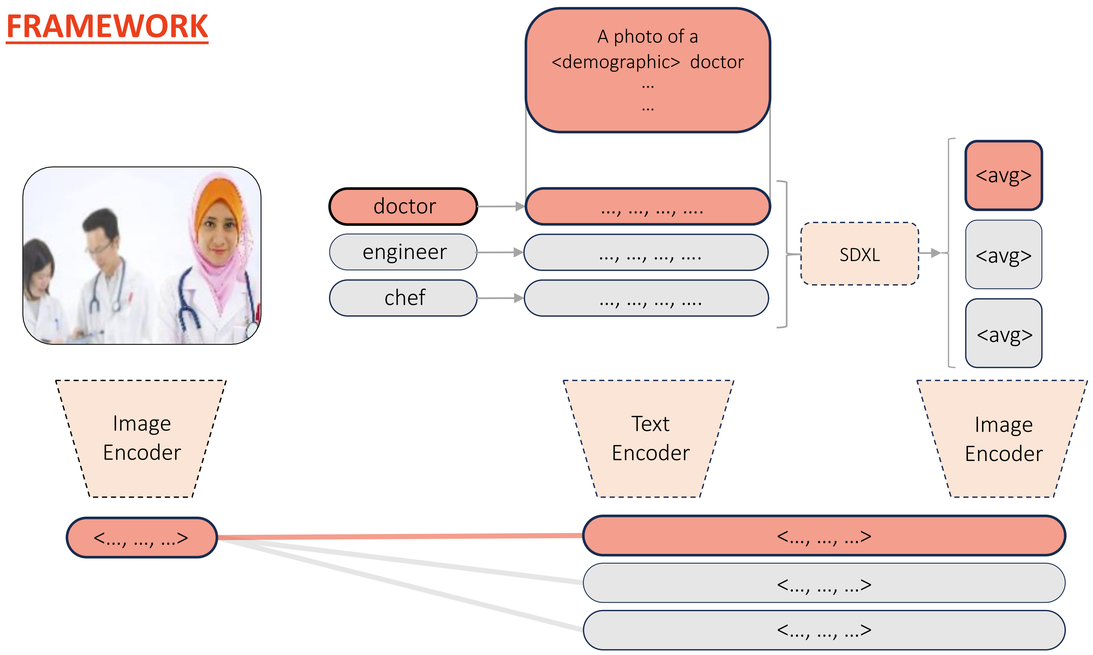
Gender: A photo of a male doctor

Age: A photo of a 0-2 year old doctor

PER-CLASS RESULTS – Avg D3G Race7

	white	black	latino/hispanic	east asian	southeast asian	indian	middle eastern
"A photo of a <race> person"	9.40	61.94	15.38	66.24	5.88	60.61	26.48
<race> + Profession	68.92	70.90	15.38	43.46	20.59	57.58	13.24
<race> + Gender	20.72	67.91	15.38	44.73	11.76	51.52	29.86
<race> + Age	11.57	65.67	11.54	59.92	14.71	69.70	25.07

FRAMEWORK



ANALYSIS

For each of these tests, the aim is to classify the profession of the person within the query image. Generating images helped most when the text utilized race as the demographic. It helped the least when gender was the demographic.

	"A photo of a"	Profession	Race 7	Race 4	Gender	Age
CLIP, Classify Profession	95.14		94.73	95.22	96.52	94.81
D3G, Classify Profession	95.54		95.22	95.30	96.52	95.06
Avg D3G, Classify Profession	95.87		95.62	95.38	96.76	95.54
CLIP, Classify Race7	28.20	44.65			28.61	25.69
D3G, Classify Race7	31.85	45.38			32.90	30.96
Avg D3G, Classify Race7	32.33	45.46			33.55	32.25

CONCLUSION & FUTURE WORK

For future directions, I aim to:

- Generate images based on demographics within the dataset (i.e. , generate more images of Hispanic people if they are underrepresented)
- Try modifying demographics of the person within the existing image

ETHICS STATEMENT

- This technique does not remove demographic biases but rather modifies the direction of the model bias.
- The images generated by the model can often reinforce certain demographic biases, and this method should only be used within appropriate contexts.