Progress Report: CNN-ViT Showdown for Racial Equality in Upscaling and Facial Recognition

Team Members: Justin Tsai 3040802612 (justin_cy.tsai@berekely.edu), Shou-Jen Chen 3041983168 (lawrancechen@berkeley.edu), Derek Xu 3037640738 (derek_xu@berkeley.edu)
Course: CS 289A – Graduate Project
Date: November 5, 2025

Outline

Graduate Project | CS 189 - Introduction to Machine Learning
Abstract: We seek to quantify how CNNs and ViTs differ in "racial fairness" towards completing image tasks when trained on imbalanced datasets.
Introduction: CNNs and ViTs are common image-task models, but typically suffer from dataset imbalances. This is a commonly discussed issue about ML deployment in public, and thus far most work has been on mitigation rather than quantifying the impact and comparing model families.
Related Work: RFW, Fairface. Referenced papers are in this link.
Methods/Approach: We use ViT/ResNet50 to upscale images of faces of different races. We evaluate fairness by comparing how well the models perform for each underrepresented race against when balanced datasets are used. The primary goal of our evaluation is to determine if the ViT or ResNet50 architecture is more fair for image upscaling and facial recognition.
Evaluation/Analysis: The methods we used to evaluate the racial fairness of each model and how racial fairness varies between models and among races. Some per-race metrics such as PSNR, SSIM, for upscaling; TPR/FPR/accuracy for face recognition, and then deriving group-fairness measures such as max-mine performance gap and standard deviation between races.
Conclusions: For upscaling there seems to be little to no difference; hinting that we may be less different than we think we are.

Progress:

Specific tasks completed:

- Custom FairFace Dataset class, image augmentations: **Justin**
- Custom RFW Dataset class: **Derek**
- ResNet50-UNet upscaler & FastViT-UNet upscaler: **Justin**
- Race-based aggregation for parsing results: **Shou-Jen**
- Main training loop/staged unfreezing: **Shou-Jen**
- Perception-based loss (VGG-based feature extraction): **Shou-Jen**
- Microbatch training: **Justin**
- RFW dataset access obtained: **Justin**
- Additional ResNet-based upscalers using different convolution methods (eg. Pixel Shuffle, interpolation + convolutions, over-scaling then downscaling): **Justin**

Technical Setup and Resources:

Repository: https://github.com/Maltomatic/RaceComp || Colab Training/Results: Link

We have built the upscaling models using a CNN (ResNet50)-based backbone and ViT (FastViT, Apple 2023)-based backbone. These were modified into UNet structures with additional convolved skip connections for the upscaling task, taking a 3*112*112 input and upscaling it to 3*224*224 with room for training to even higher resolutions.
Custom dataset classes were designed to use the FairFace dataset, and work is underway to adapt that to use the Racial Faces in the Wild (RFW) dataset.

As far as training is concerned, we have completed training for all the racially-biased experiments. “Minority” races were sampled at 20% the possibility of normal races, and then the models were evaluated by how well they could recreate images of all races, even ones that were underrepresented in training. Training is set to unfreeze only the latter stages of the backbone networks that were used, and run for a total of three epochs at different stages of unfreezing.

Validation Set Results:

Minority (W = 0.2)	Epoch	mean_SSIM	min_SSIM	max_SSIM	delta_SSIM	mean_PSNR	min_PSNR	max_PSNR	delta_PSNR
All	3	0.9960	0.9960	0.9960	0.0	37.03	36.95	37.09	0.14
Black	3	0.993	0.993	0.993	0.0	34.56	34.54	34.58	0.039
Southeast-Asian	3	0.991	0.991	0.991	0.0	32.95	32.94	32.97	0.030
Latino_Hispanic	3	0.994	0.994	0.994	0.0	35.75	35.73	35.77	0.040
Middle Eastern	3	0.995	0.995	0.995	0.0	35.93	35.91	35.95	0.040
White	3	0.993	0.993	0.993	0.0	34.56	34.54	34.58	0.039
Indian	3	0.9943	0.9940	0.9940	0.0	35.74	35.65	35.79	0.14
East Asian	3	0.9866	0.9860	0.9870	0.001	31.83	31.73	31.90	0.17

Current issues and mitigation plans:

One unexpected issue we have run into was the over-fairness of both models. The initial intent was to observe how each model differed in their generalization performance; i.e. which type of model performed worse trying to upscale an underrepresented race relative to its own “fully balanced” training data. However, as observed, there is no discernible difference—both model structures recreate images quite well.

At this time, we are pursuing two avenues in parallel. Firstly, we are continuously lowering the represented races in the dataset—we intend to run a batch of tests on 0.05 weights for the underrepresented minority and see how the models continue to generalize. We will also further reduce the input image size to 56x56 and see if this further reduced input information affects the model performance noticeably.

Goals and Updated Timeline:

The goals remain unchanged: we will continue the training and bias experiments and seek to observe how underrepresentation really affects a model performance. In addition we will test the new parameters and try to identify

1) what model structure produces the best image upscaling results, and
2) how little information is needed to finally noticeably impact model performance.

Schedule

Week	Goal	Manager	Subtasks
W9 (10/22~25)	“Fairness” scoring method; literature review	Shou-Jen	- Obtain RFW access(licensing) - Review SOTA methods; identify repositories with public code we can reference - Define accuracy quantification for image upscaling - Define “fairness” quantification - Setup Colab+Github Repo env
W10 (10/26~11/1)	Complete ResNet and ViT architecture modifications and initial training for **upscaling task**; establish baseline	Justin	- Replicate testing environment for ResNet and FastViT - Unfreeze rear 8 layers of each model to do transfer-learning - Evaluate baseline trained on full FairFace dataset
	Implement racially biased sampling method, training loop	Shou-Jen	- Modify dataloader/training script to under-sample minorities - Adjust loss calculations to consider perception loss and race buckets for future analysis
W11 (11/2~11/8)	Run experiments with re-training on skewed datasets	Shou-Jen	- Generate skewed train splits for each experiment with a different overrepresented race - Upscale images using models - Calculate upscaling accuracy/error for images - Compare error for each racial group
W12 (11/9~11/15)	Complete ResNet and ViT architecture modifications and initial training for **facial recognition**; establish baseline	Derek	- Replicate testing environment for ResNet and FastViT - Unfreeze rear 2 layers of trained upscale model to do transfer-learning for facial recognition - Evaluate baseline trained on full RFW dataset
	Complete modifications and train with new upscaling model parameters—smaller input size, reduced minority representation	Justin	- Modify ResNet and ViT UNets to take further reduced image sizes - Run experiments with reduced
W13 (11/16~11/22)	Run experiments with re-training on skewed datasets	Justin	- Generate skewed train splits for each experiment with a different overrepresented race - Test facial recognition using models - Calculate recognition accuracy/error for images - Compare error for each racial group
W14 (11/23~11/29)	Organize results and identify model traits; paper first draft & edits	Derek	- Analyze model performance degradation on underrepresented races - Discuss context, methods, results, and social impact on the final report - Plot graphs, identify correlations - Construct hypothesis for results based on model traits - Finish first draft of the paper
W15 (11/30~12/6)	Finalize results for review; Peer review other teams	Shou-Jen	- Review other teams’ papers - Provide feedback on reviewed papers
W16 (12/7~12/12)	Final Report & Presentation	Justin	- Finalize edits and content on the report - Include same information on presentation with more engaging delivery