Accurately Classifying Out-Of-Distribution Data in Facial Recognition

Gianluca Barone [1], Aashirit Cunchala [2], Rudy Nunez [3]

[1] Rowan University, [2] University of Pittsburgh, [3] Emory University

Classification Problem & Fairness

It is an assumption in standard classification theory that the testing data and the training data come from the same distribution, however in many real world applications that is often not the case. When encountering out-of-distribution (OOD) data, a model can often struggle to accurately classify images. This causes issues in real world scenarios due to problems with fairness. A fair model would be able to classify different genders and races evenly. Often, this does not happen, as models contain bias which causes the majority group to overpower the minority group.

Outlier Exposure

Outlier exposure is a method to improve classifier accuracy by exposing it to a sample of OOD images. The goal of outlier exposure is to improve the model's ability to properly detect OOD images. To do this, it modifies the loss function to be:

$$\min_{\theta} \mathbb{E}_{P_{in}(\mathbf{x},y)}[L(f_\theta(\mathbf{x}),y)] + \lambda \mathbb{E}_{P_{out}(\mathbf{x}_{OE})}[L(f_\theta(\mathbf{x}^{OE}),U(y))], \qquad (1)$$

where θ are the hyperparameters that are being minimized, x is the image, y is the label, and $L(\cdot,\cdot)$ represents the cross entropy loss function.

- $\mathbb{E}_{P_{in}(\mathbf{x},y)}[L(f_\theta(\mathbf{x}),y)]$ represents the loss for in-distribution data.
- $\mathbb{E}_{P_{out}(\mathbf{x}^{OE})}[L(f_\theta(\mathbf{x}^{OE}),U(y))]$ represents the loss for the outlier exposure data. This represents the loss between a normal softmax distribution and $U(y)$, where $U(y)$ represents a uniform distribution

In our model, λ is a hyperparameter [1] in which we changed based on the differences between our image distributions. We found outlier exposure for various different OOD sets, and noticed that close image distributions reduce the effectiveness. To combat this, we reweighted the sample size with $\lambda(D)$ representing parameter as a function of the difference in distributions.

Datasets

UTKFace : Over 20,000 face images which are labeled by age, gender, and race.
Fairface : Over 100,000 face images with seven race labels which we collapsed into five labels for consistency
CIFAR-10 : 10 classes of aniamls and transportation vehicle
UTKFace Outlier : 20% furthest from the mean distribution of UTKFace chosen through Kullback-Leibler divergence
FairFace Outlier : 20% furthest from the mean distribution of FairFace chosen through Kullback-Leibler divergence

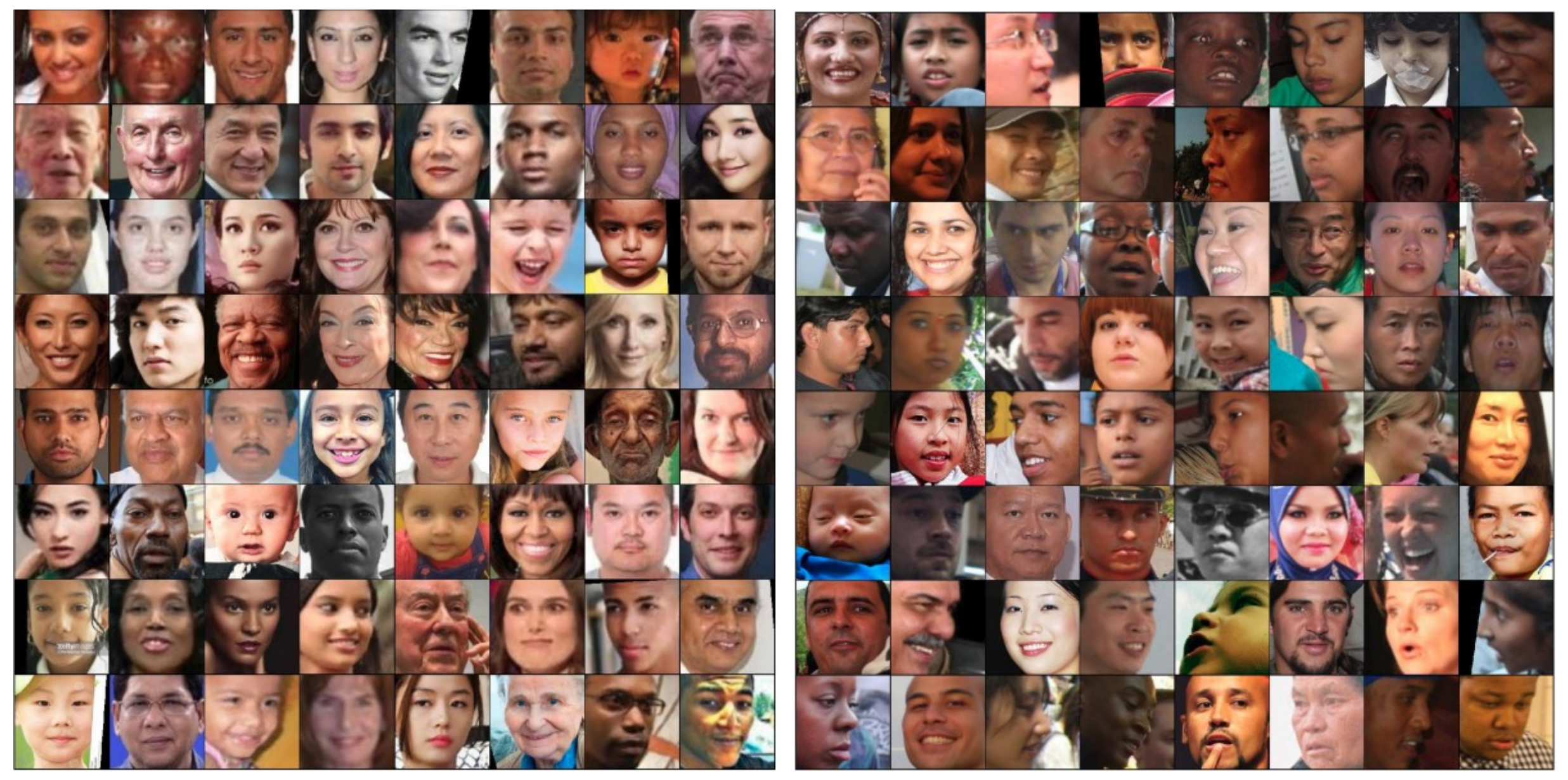

Fig. 1: Images From Datasets Left: UTKFace, Right: FairFace

Identifying and Quantifying Outlier Images and Datasets

We found outliers using the Kullback-Leibler (KL) divergence on the pixel values of images. We also rewrote λ as a function of difference in distributions These images were used in Outlier Exposure testing to see the effects of OOD data on training.
The KL distance is:

$$\mathbf{D} := \mathbf{KL}(\mathbf{P}||\mathbf{Q}) = \sum_{\mathbf{x}} \mathbf{P}(\mathbf{x}) \log\left(\frac{\mathbf{P}(\mathbf{x})}{\mathbf{Q}(\mathbf{x})}\right), \qquad (2)$$

and λ is:

$$\lambda(\mathbf{D}) = \tanh(\mathbf{D}), \qquad (3)$$

Then, λ was further modified to increase as the model got further in training so:

$$\lambda(\mathbf{D},\mathbf{i}) = \tanh(\mathbf{D})(\mathbf{1} - \cos(\pi\mathbf{i}/\mathbf{20})), \qquad (4)$$

where i represents the current epoch from 1 to 20.

Results

Outlier Sample	*Precision*	*Recall*	*Accuracy*	*F1*	*AUROC*
None	0.65	0.68	0.69	0.66	0.77
UTKFace Outliers	0.70	0.68	0.70	0.69	0.77
FairFace Outliers	**0.75**	**0.81**	**0.80**	**0.78**	**0.85**

Table 1: Results of increasing the samples for training by using the outlier sets as additional samples

Male Weight	*Female Weight*	*Precision*	*Recall*	*F1*	*AUROC*
1.0	1.0	0.57	**0.65**	0.61	**0.70**
1.0	1.5	**0.62**	0.64	**0.63**	**0.70**

Table 2: Results of changing the weights used for the loss function

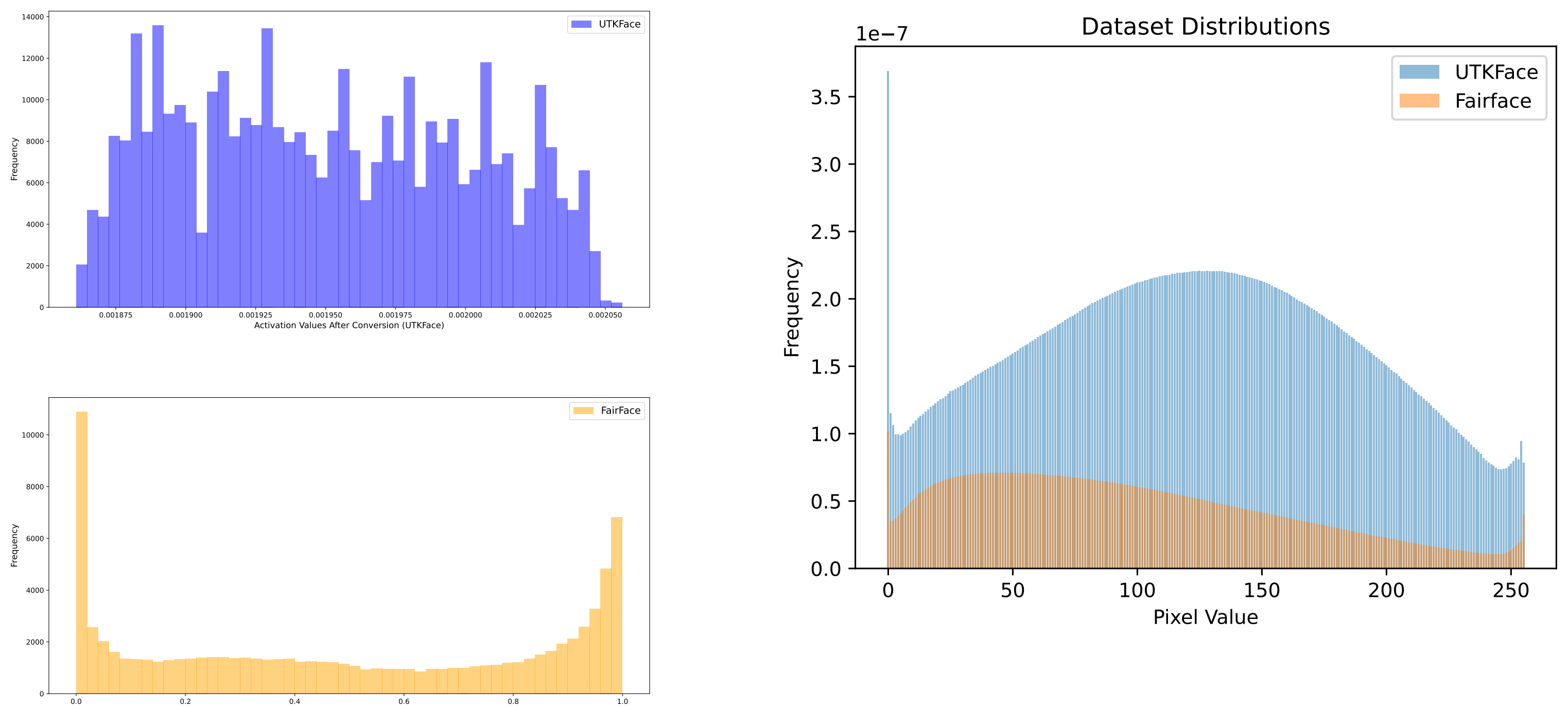

Fig. 2: Distribution of pixel frequency in UTKFace (blue) and FairFace (orange)

The pixel mean is close to 120 for UTKFace and close to 70 for FairFace implying that FairFace is on average darker images but with a KL divergence of 0.088 when measured by pixels the two datasets are rather similar.
Activation features are the output maps of intermediate layers of a CNN and help record different features in the images such as textures and color gradients. For FairFace our range is much larger and is bimodal, while UTKFace has a more narrow range. Therefore, it is harder for the model to train on UTKFace and test on FairFace compared to the other way around.

Results

Our method of testing involved training on UTKFace and testing on FairFace. For all trials λ was determined by the KL distance between UTKFace and FairFace, **except for** the trial without outlier exposure.

Outlier Group	*Precision*	*Recall*	*Accuracy*	*F1*	*AUROC*
None	**0.65**	0.68	0.69	**0.66**	**0.77**
UTKFace	0.62	**0.72**	**0.71**	**0.66**	0.76
FairFace	0.53	0.52	0.55	0.52	0.60
UTKFace Outliers	0.61	0.71	0.70	**0.66**	0.75
FairFace Outliers	0.58	0.66	0.66	0.62	0.70
CIFAR-10	0.58	0.71	0.69	0.64	0.75

The table above summarizes our results from using outlier exposure when classifying based on gender. The choice for the outlier group is incredibly significant, and causes accuracy and other metrics to vary widely.

Conclusion

- We implemented outlier exposure on a facial recognition machine learning network and found limited results for datasets that have similar distributions
- We made the parameter in our loss function trainable with respect to image distributions

Future work:

- Use feature norms of the images and other outputs of the CNN to sort the images and improve classification
- Adapt the activation features code into the outlier exposure code to see if it is more accurate than using the KL divergence based on pixels
- Alter λ by including KL divergence dependent on the distribution between the training and outlier set

Acknowledgements

This work is supported in part by the US NSF award DMS-2051019. We would like to thank our mentor, Dr. Nicole Yang, along with all of the mentors at the Emory University REU site.

References

[1] Dan Hendrycks, Mantas Mazeika, and Thomas Dietterich. *Deep Anomaly Detection with Outlier Exposure.* 2019. arXiv: 1812.04606 [cs.LG].

[2] Kimmo Karkkainen and Jungseock Joo. "FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age for Bias Measurement and Mitigation". In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision.* 2021, pp. 1548–1558.

[3] Alex Krizhevsky, Geoffrey Hinton, et al. "Learning multiple layers of features from tiny images". In: (2009).