Supplementary Information: A Differentiable Distance Approximation for Fairer Image Classification

Nicholas Rosa[1 [0000−0002−6079−8291]], Tom Drummond[1,2 [0000−0001−8204−5904]], and Mehrtash Harandi[1 [0000−0002−6937−6300]]

[1] Monash University, Australia
[2] The University of Melbourne, Australia

1 Training Procedure

All models were trained with the optimization hyper-parameters described in Table 1. As a default we used a batchsize of 512, a learning rate of 0.0001 and a cosine learning rate schedule [2]. However, in some cases we found extra performance could be found by varying these hyper-parameters. The exact values that were used for each experiment is shown in Table 2. Model weights were initialised using models from [4], which are pretrained upon the ImageNet dataset [1].

Table 1. Optimization hyper-parameters used during training.

Parameter	Value
Optimizer	AdamW [3]
Weight Decay	0.02
Epochs	50

Data augmentation was also used to assist in regularization of the models. The data augmentation scheme, shown in Table 3, was used for all methods and tasks. At inference time the images were resized, and then centre cropped.

References

1. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
2. Loshchilov, I., Hutter, F.: SGDR: Stochastic Gradient Descent with Warm Restarts. In: 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings. OpenReview.net (2017), https://openreview.net/forum?id=Skq89Scxx
3. Loshchilov, I., Hutter, F.: Decoupled Weight Decay Regularization. In: International Conference on Learning Representations (2019), https://openreview.net/forum?id=Bkg6RiCqY7

Table 2. Learning rate and batch size used during training.

Dataset	Task	Method	Batchsize	Learning Rate	Schedule	Image Size
UTKFace	Age	Naive	512	0.0001	Cosine	176
UTKFace	Age	Naive Balanced	512	0.0001	Cosine	176
UTKFace	Age	MFD	128	0.001	None	176
UTKFace	Age	AD	128	0.0001	Cosine	176
UTKFace	Age	BASE	512	0.0001	Cosine	176
UTKFace	Gender	Naive	512	0.0001	Cosine	176
UTKFace	Gender	Naive Balanced	512	0.0001	Cosine	176
UTKFace	Gender	MFD	512	0.0001	Cosine	176
UTKFace	Gender	AD	128	0.0001	Cosine	176
UTKFace	Gender	BASE	512	0.0001	Cosine	176
CelebA	Gender	Naive	512	0.0001	Cosine	128
CelebA	Gender	Naive Balanced	512	0.0001	Cosine	128
CelebA	Gender	BASE	512	0.0001	Cosine	128
Fairface	Age	Naive	512	0.0001	Cosine	176
Fairface	Age	Naive Balanced	512	0.0001	Cosine	176
Fairface	Age	MFD	512	0.0001	Cosine	176
Fairface	Age	AD	128	0.0001	Cosine	176
Fairface	Age	BASE	512	0.0001	Cosine	176
Fairface	Gender	Naive	512	0.0001	Cosine	176
Fairface	Gender	Naive Balanced	512	0.0001	Cosine	176
Fairface	Gender	MFD	512	0.0001	Cosine	176
Fairface	Gender	AD	128	0.0001	Cosine	176
Fairface	Gender	BASE	512	0.0001	Cosine	176
FairfaceS	Gender	Naive	512	0.0001	Cosine	176
FairfaceS	Gender	BASE	512	0.0001	Cosine	176

Table 3. Data augmentation transformations used during training.

Transformation
Horizontal Flip ($p = 0.5$)
Random Resized Crop (Scale $[0.08, 1]$, Ratio $[0.75, 1.33]$)
Color Jitter (factor 0.4, $p = 0.8$)

4. Wightman, R., Touvron, H., Jegou, H.: ResNet strikes back: An improved training procedure in timm. In: NeurIPS 2021 Workshop on ImageNet: Past, Present, and Future (2021), https://openreview.net/forum?id=NG6MJnVl6M5