Table 1: The manually selected weighting coefficients λ from the experiments with the CelebA, UTKFace, and SIIM-ISIC melanoma classification data set.

(a) CelebA

#	Loss	λ	#	Loss	λ
2	L_{iou}	60.0	8	L_{iou}	75.0
3	$L_{eo}^{l_2}$	25.0	9	$L_{eo}^{l_2}$	25.0
4	L_{eo}^{mi}	25.0	10	L_{eo}^{mi}	50.0
5	$L_{dp}^{l_2}$	15.0	11	$L_{dp}^{l_2}$	15.0
6	L_{dp}^{mi}	30.0	12	L_{dp}^{mi}	75.0

(b) UTKFace

#	Loss	λ	#	Loss	λ
2	L_{iou}	55.0	8	L_{iou}	5.5
3	$L_{eo}^{l_2}$	6.0	9	$L_{eo}^{l_2}$	1.0
4	L_{eo}^{mi}	8.0	10	L_{eo}^{mi}	6.0
5	$L_{dp}^{l_2}$	1.5	11	$L_{dp}^{l_2}$	0.6
6	L_{dp}^{mi}	15.0	12	L_{dp}^{mi}	20

(c) SIIM-ISIC

#	Loss	λ
1	L_{ce}	-
2	L_{iou}	1500.0
3	$L_{eo}^{l_2}$	2.5
4	L_{eo}^{mi}	1.5
5	$L_{dp}^{l_2}$	5.0
6	L_{dp}^{mi}	1.5

A Experiments

Weighting Coefficients We select the weighting coefficients λ heuristically for the initial experiments with our fairness losses. We compare the final values of the cross-entropy loss L_{ce} and each fairness loss L_{fair} of the baseline model on the validation partition and select an initial value of λ, s.t. these losses are proportional. After a weighting coefficient is selected, we train a corresponding model and evaluate its classification performance and fairness quantitatively with existing performance measures (such as the accuracy and the AUC score) based on the validation partition. If none of the initial coefficients yield any improvements for the respective fair loss w.r.t. the baseline model, we select additional weighting coefficients from a similar range. This process is performed until the optimized fairness loss decreases by more than half from the baseline model. We report results for the weighting coefficients where the corresponding performance metric achieved is maximum, and the fairness loss decreased considerably. Table 1 shows the selected weighting coefficients for CelebA (1a), UTKFace (1b) and the SIIM-ISIC melanoma classification data set (1c).

Hyper-Parameters We train a baseline model for 100 epochs by minimizing the sole cross-entropy loss L_{ce} with the ADAM optimizer. The default optimization settings are used in all experiments (learning rate $\alpha = 0.001$, exponential decay factors $\beta_1 = 0.99, \beta_2 = 0.999$ and $\epsilon = 1 \times 10^{-8}$). The batch size is set to 256 samples to get enough samples for the fairness estimation. After that, we train by extending the cross-entropy loss with our fairness loss for 25 epochs.

SIIM-ISIC Melanoma Classification We performed multiple augmentation transformations on the training images for the SIIM-ISIC melanoma data set since the original amount of malignant samples in this data set is low. We applied random horizontal and vertical flips with probabilities $p = 0.5$, random rotations with uniform rotation angles between -180 to 180 degrees, and a random color jitter that changes the brightness, hue, and saturation by uniformly random sampled offsets 0.2, 0.2 and 0.3.