

– Supplementary Material –

Unravelling the Effect of Image Distortions for Biased Prediction of Pre-trained Face Recognition Models

Puspita Majumdar
IIIT-Delhi

pushpitam@iiitd.ac.in

Surbhi Mittal
IIT Jodhpur

mittal.5@iitj.ac.in

Richa Singh
IIT Jodhpur

richa@iitj.ac.in

Mayank Vatsa
IIT Jodhpur

mvatsa@iitj.ac.in



Figure 1. Sample images from the MUCT dataset degraded with image distortions from (top to bottom): occlusion, Gaussian blur, brightness, Gaussian noise, salt and pepper noise, resolution. Left-most column shows the original image used for applying distortions.

1. Proposed Evaluation Framework

For the experiments, four pre-trained deep face recognition models: (i) LightCNN-29 [7] (ii) SENet50 [6] (iii) ResNet50 [5], and (iv) ArcFace [3] are used. The details of the pre-trained models is discussed in the following subsection.

1.1. Details of the Pre-trained Models

The four pre-trained models used for the experiments are trained on large-scale facial datasets. LightCNN-29 is pre-trained on the MS-Celeb-1M dataset [4]. The dataset contains approximately 10M images of 1M subjects. SENet50 and ResNet50 models are pre-trained (from scratch) on

the VGGFace2 dataset [2], which contains approximately 3.3M images corresponding to 9K subjects. ArcFace is pre-trained on the CASIA [8], VGGFace2 [2], MS1MV2 and DeepGlint-Face (including MS1M-DeepGlint and Asian-DeepGlint) [1] datasets. MS1MV2 is a semi-automatic refined version of the MS-Celeb-1M dataset. The CASIA dataset contains 0.5M images of 10K subjects. MS1M-DeepGlint and Asian-DeepGlint datasets contain 3.9M and 2.83M images corresponding to 87K and 94K subjects, respectively.

1.2. Details of Image Distortions

The details of the image distortions used in this research are discussed in the main paper. Figure 1 shows sample images of the MUCT dataset after applying image distortions of different intensities.

2. Analyzing the Effect of Distortions on Bias in Model Predictions

In this research, we evaluate the performance of pre-trained face recognition models in presence of real-world image distortions across different *gender* and *race* subgroups. Analysis is also performed across the intersectional subgroups of *gender* and *race*.

Role of facial regions in recognition across subgroups

Different facial regions are occluded to analyze the importance of these regions in recognition of a particular subgroup. The important results and analysis are already presented in the main paper. Here, we have shown the verification performance across the intersectional subgroups of *gender* and *race* corresponding to the MORPH dataset in Table 1.

Does model performance degrade equally across subgroups in presence of Gaussian blur?

Table 1. Verification accuracy and DoB (%) across the intersectional subgroups of *gender* and *race* under occlusion corresponding to the MORPH dataset.

		LCNN-29	SENet50	ResNet50	ArcFace
Eyes	{G1,R1}	97.67	74.20	62.30	95.83
	{G1,R2}	95.83	71.13	64.13	92.80
	{G2,R1}	90.37	40.60	25.43	85.53
	{G2,R2}	84.77	36.70	37.07	77.10
	DoB	5.82	19.74	19.07	8.35
Nose	{G1,R1}	96.17	73.60	78.40	95.17
	{G1,R2}	89.47	55.67	75.80	84.63
	{G2,R1}	81.13	62.23	48.93	78.43
	{G2,R2}	66.90	35.90	43.67	67.07
	DoB	12.61	15.81	17.94	11.75
Mouth	{G1,R1}	99.93	96.17	94.80	99.63
	{G1,R2}	99.97	97.00	97.53	99.53
	{G2,R1}	99.73	92.27	92.00	99.20
	{G2,R2}	99.50	90.13	87.73	98.73
	DoB	0.22	3.25	4.18	0.41
Forehead	{G1,R1}	99.93	97.00	97.93	99.93
	{G1,R2}	99.87	95.63	96.93	99.57
	{G2,R1}	99.77	85.83	85.43	99.43
	{G2,R2}	99.50	89.63	90.90	99.00
	DoB	0.19	5.22	5.81	0.38
Left Cheek	{G1,R1}	99.97	96.27	97.47	100.00
	{G1,R2}	100.00	98.40	98.40	99.97
	{G2,R1}	99.77	92.37	91.97	99.47
	{G2,R2}	99.90	94.80	94.77	99.50
	DoB	0.10	2.54	2.90	0.29
Right Cheek	{G1,R1}	99.97	96.70	95.60	99.97
	{G1,R2}	100.00	98.80	98.60	99.87
	{G2,R1}	99.80	91.37	91.60	99.47
	{G2,R2}	99.77	92.97	95.53	99.50
	DoB	0.12	3.40	2.87	0.25
Mask	{G1,R1}	94.53	71.90	55.07	89.93
	{G1,R2}	92.37	59.27	67.50	85.97
	{G2,R1}	85.30	61.37	40.57	80.97
	{G2,R2}	75.50	37.60	35.40	68.07
	DoB	8.58	14.39	14.53	9.51

The effect of Gaussian blur on the performance of models for *gender* subgroups corresponding to the MORPH dataset is discussed in the main paper. However, due to page limitation, we have shown the results for limited σ values. Here, in Table 2, we have reported the results for all σ values. Table 3 shows the verification performance corresponding to the MUCT dataset across *gender* subgroups. It is observed that the performance gap increases as we increase the intensity of blur. Therefore, the DoB increases from 0.00%, 0.05%, 0.02%, and 0.00% to 0.07%, 11.62%, 1.41%, and 5.49% corresponding to the LCNN-29, SENet50, ResNet50, and ArcFace, respectively at 0.01 False Accept Rate (FAR) with $\sigma = 4.0$. The verification accuracy across *race* subgroups of the MORPH dataset

Table 2. Verification accuracy and DoB (%) across different *gender* subgroups with varying intensities of Gaussian blur corresponding to the MORPH dataset. * represents a relatively high disparity in model performance across different subgroups on undistorted images.

σ		LCNN-29	SENet50	ResNet50	ArcFace
0.0	G1	100.00	97.90	98.27	99.90
	G2	99.83	91.97	93.30	99.67
	DoB	0.12	4.19*	3.51*	0.16
2.0	G1	99.83	91.23	94.30	99.70
	G2	99.40	79.40	89.80	98.07
	DoB	0.30	8.37	3.18	1.15
2.2	G1	99.80	87.77	92.87	99.57
	G2	98.77	76.00	86.33	97.70
	DoB	0.73	8.32	4.62	1.32
2.4	G1	99.73	84.53	91.47	99.30
	G2	98.43	73.60	84.33	96.93
	DoB	0.92	7.73	5.05	1.68
2.6	G1	99.60	81.73	88.47	98.90
	G2	98.00	69.50	82.07	96.07
	DoB	1.13	8.65	4.53	2.00
2.8	G1	99.33	78.37	86.77	98.37
	G2	97.77	64.70	77.37	94.83
	DoB	1.10	9.67	6.65	2.50
3.0	G1	98.83	74.00	82.83	97.07
	G2	96.40	61.70	73.40	92.67
	DoB	1.72	8.70	6.67	3.11
3.2	G1	98.00	65.40	78.37	95.47
	G2	93.77	52.43	63.67	88.17
	DoB	2.99	9.17	10.39	5.16
3.4	G1	96.87	60.83	74.10	93.70
	G2	91.37	49.53	56.83	85.80
	DoB	3.89	7.99	12.21	5.59
3.6	G1	93.47	56.73	63.83	90.53
	G2	87.63	46.90	46.60	78.17
	DoB	4.13	6.95	12.18	8.74
3.8	G1	89.90	53.67	57.57	86.4
	G2	79.87	39.60	40.73	71.30
	DoB	7.09	9.95	11.91	10.68
4.0	G1	84.70	50.57	52.27	81.07
	G2	72.27	35.57	33.77	63.53
	DoB	8.79	10.61	13.08	12.40

is reported in Table 4. It is observed that for most of the models, performance degradation is maximum for subgroup *R1*. Blur affects different model’s performance in a similar fashion across gender (or race) subgroups. However, the extent of bias introduced in model predictions is not consistent. The performance of the models across the intersectional subgroups of *gender* and *race* are reported in Table 5. A huge disparity in model performance across different subgroups is observed.

Does performance gap between subgroups increase for

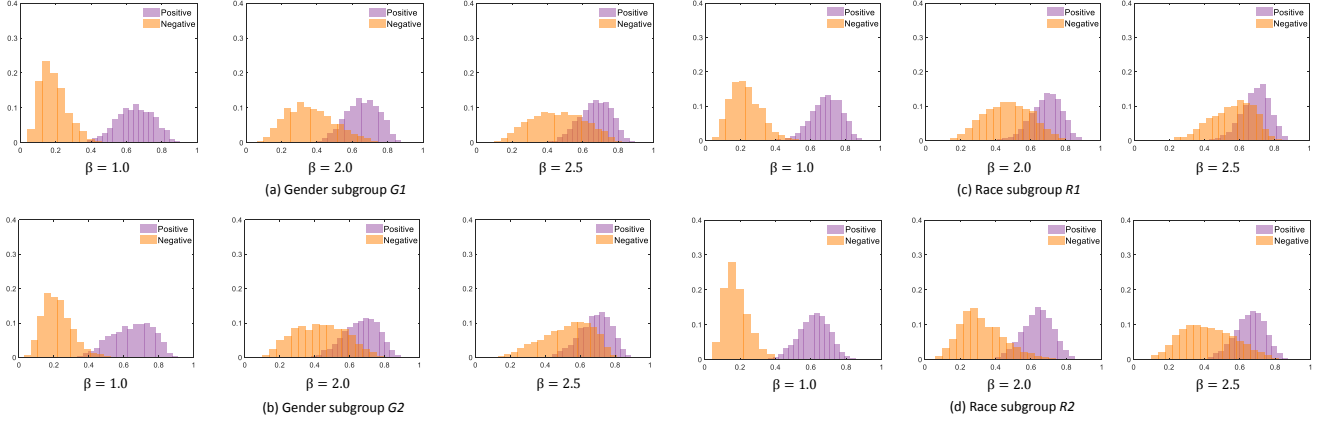


Figure 2. Score distribution of pre-trained SENet50 model across (a-b) *gender* and (c-d) *race* subgroups by varying the brightness factor β from 1.0 to 2.5 corresponding to the MORPH dataset.

brighter images?

As reported in the main paper, brightness affects the performance of subgroups *G2* and *R1* more severely than subgroups *G1* and *R2*, respectively. Figure 2 shows the score distribution corresponding to the SENet50 model across different subgroups of the MORPH dataset with varying intensities of brightness. It is observed that the overlap in distribution is more for subgroups *G2* and *R1* compared to subgroups *G1* and *R2*, respectively, when the same intensity of brightness is applied to the images. As a result, performance degradation is higher for these subgroups. Table 6 shows the verification performance across different *gender* subgroups corresponding to the MORPH dataset. The performance on the MUCT dataset is reported in Table 7. Model performance across the intersectional subgroups of *gender* and *race* with varying intensities of brightness is shown in Table 8. The results indicate that the model predictions become biased in presence of brighter images.

Do models perform differently across subgroups in presence of noise?

We have observed that Gaussian noise equally affects the performance of the models for different subgroups. The results across *gender* subgroups are reported in the main paper. Here, we have shown the results across *race* subgroups in Table 9. It is observed that the overall performance degrades, but the difference in performance across subgroups is not significant. Table 10 shows the performance across the intersectional subgroups. Salt and pepper noise affects the performance of most of the models for subgroups *G2* and *R1* more severely than subgroups *G1* and *R2*, respectively. The performance across *race* subgroups are reported in Table 11. The

DoB indicates that bias is introduced in model prediction with increasing intensities of salt and pepper noise. For example, the DoB increases from 0.25% to 14.80% corresponding to the ResNet50 with $p = 0.09$. The performance across the intersectional subgroups are reported in Table 12.

Does model performance differ across subgroups with varying image resolution?

As reported in the main paper, a significant degradation in performance is observed for low resolution images. Also, we have observed that bias is introduced in ResNet50 model predictions when the resolution of the images are reduced. Table 13 shows the performance across *race* subgroups corresponding to the MORPH dataset. The DoB increases from 0.47% to 12.59% corresponding to the ResNet50 model at 28×28 resolution. Table 14 shows the performance of the models across intersectional subgroups. The results of the MUCT dataset is summarized in Table 15. We have also observed that the ArcFace model does not perform well at low resolution and a sharp drop in accuracy is observed beyond 48×48 resolution.

References

- [1] DeepGlint. <http://trillionpairs.deepglint.com/overview>. 1
- [2] Qiong Cao, Li Shen, Weidi Xie, Omkar M Parkhi, and Andrew Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *IEEE FG*, pages 67–74, 2018. 1
- [3] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *IEEE CVPR*, pages 4690–4699, 2019. 1
- [4] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *ECCV*, pages 87–102. Springer, 2016. 1

Table 3. Verification accuracy and DoB (%) across different *gender* subgroups with varying intensities of Gaussian blur corresponding to the MUCT dataset.

σ		LCNN-29	SENet50	ResNet50	ArcFace
0.0	G1	100.00	100.00	100.00	100.00
	G2	100.00	99.93	99.97	100.00
	DoB	0.00	0.05	0.02	0.00
2.0	G1	100.00	99.87	99.73	100.00
	G2	99.97	99.40	99.03	99.93
	DoB	0.02	0.33	0.49	0.05
2.2	G1	99.97	99.87	99.60	100.00
	G2	99.93	99.03	98.73	99.90
	DoB	0.03	0.59	0.62	0.07
2.4	G1	99.93	99.70	99.33	99.97
	G2	99.93	98.60	98.40	99.80
	DoB	0.00	0.78	0.66	0.12
2.6	G1	99.93	99.50	98.77	99.87
	G2	99.80	98.03	97.63	99.5
	DoB	0.09	1.04	0.81	0.26
2.8	G1	99.77	99.10	98.13	99.63
	G2	99.77	97.50	96.77	99.33
	DoB	0.00	1.13	0.96	0.21
3.0	G1	99.63	99.1	97.57	99.33
	G2	99.63	96.83	95.3	99.07
	DoB	0.00	1.61	1.61	0.18
3.2	G1	99.30	97.73	95.17	98.13
	G2	99.07	94.43	92.90	97.73
	DoB	0.16	2.33	1.61	0.28
3.4	G1	98.40	95.87	92.00	96.20
	G2	98.13	90.70	90.07	97.07
	DoB	0.19	3.66	1.36	0.62
3.6	G1	95.97	92.27	85.83	89.37
	G2	96.20	84.40	84.97	94.37
	DoB	0.16	5.56	0.61	3.54
3.8	G1	93.83	88.80	82.97	84.20
	G2	93.90	73.83	79.27	91.73
	DoB	0.05	10.59	2.62	5.32
4.0	G1	89.70	82.43	75.43	79.50
	G2	89.60	66.00	73.43	87.27
	DoB	0.07	11.62	1.41	5.49

Table 4. Verification accuracy and DoB (%) across different *race* subgroups with varying intensities of Gaussian blur corresponding to the MORPH dataset.

σ		LCNN-29	SENet50	ResNet50	ArcFace
0.0	R1	100.00	98.53	98.80	99.97
	R2	100.00	99.27	99.17	99.93
	DoB	0.00	0.52	0.26	0.03
2.0	R1	100.00	90.93	94.40	99.67
	R2	99.90	95.67	96.03	99.30
	DoB	0.07	3.35	1.15	0.26
2.2	R1	99.90	88.07	92.77	99.47
	R2	99.83	93.30	95.00	98.70
	DoB	0.05	3.70	1.58	0.54
2.4	R1	99.90	84.27	90.23	99.27
	R2	99.80	91.70	93.23	98.67
	DoB	0.07	5.25	2.12	0.42
2.6	R1	99.87	77.90	86.97	98.70
	R2	99.73	89.17	89.63	97.83
	DoB	0.10	7.97	1.88	0.62
2.8	R1	99.63	72.27	80.13	97.97
	R2	99.60	84.53	87.50	96.73
	DoB	0.02	8.67	5.21	0.88
3.0	R1	99.50	70.33	74.17	97.30
	R2	99.13	80.07	85.63	95.63
	DoB	0.26	6.89	8.10	1.18
3.2	R1	98.43	60.87	61.60	95.30
	R2	97.93	71.67	79.63	93.20
	DoB	0.35	7.64	12.75	1.48
3.4	R1	97.10	56.10	52.30	93.00
	R2	96.73	65.93	76.53	91.10
	DoB	0.26	6.95	17.13	1.34
3.6	R1	93.47	47.50	40.37	86.30
	R2	94.10	52.23	71.40	85.73
	DoB	0.45	3.34	21.94	0.40
3.8	R1	89.23	42.20	31.20	81.90
	R2	91.47	40.80	63.83	82.07
	DoB	1.58	0.99	23.07	0.12
4.0	R1	81.70	36.90	26.13	76.87
	R2	88.40	34.97	52.97	77.53
	DoB	4.74	1.36	18.98	0.47

- [5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE CVPR*, pages 770–778, 2016. 1
- [6] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *IEEE CVPR*, pages 7132–7141, 2018. 1
- [7] Xiang Wu, Ran He, Zhenan Sun, and Tieniu Tan. A light cnn for deep face representation with noisy labels. *IEEE TIFS*, 13(11):2884–2896, 2018. 1
- [8] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li. Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*, 2014. 1

Table 5. Verification accuracy and DoB (%) across the intersectional subgroups of *gender* and *race* with varying intensities of Gaussian blur corresponding to the MORPH dataset.

σ		LCNN-29	SENet50	ResNet50	ArcFace	σ		LCNN-29	SENet50	ResNet50	ArcFace
0.0	{G1,R1}	100.00	97.53	98.20	100.00	3.0	{G1,R1}	99.23	70.60	77.83	97.00
	{G1,R2}	100.00	99.27	99.17	99.97		{G1,R2}	99.30	85.63	88.43	97.23
	{G2,R1}	99.97	96.40	95.53	99.77		{G2,R1}	97.30	57.23	53.30	93.60
	{G2,R2}	99.90	94.10	96.67	99.57		{G2,R2}	96.77	57.70	65.83	91.57
	DoB	0.05	2.17	1.61	0.20		DoB	1.31	13.41	15.17	2.75
2.0	{G1,R1}	99.97	89.67	93.97	99.73	3.2	{G1,R1}	98.43	61.23	71.53	94.57
	{G1,R2}	99.97	96.27	96.60	99.50		{G1,R2}	98.43	82.67	81.30	95.67
	{G2,R1}	99.50	84.30	77.60	99.07		{G2,R1}	94.30	48.23	44.43	89.83
	{G2,R2}	99.20	83.17	87.40	97.83		{G2,R2}	95.27	49.17	59.67	87.90
	DoB	0.38	5.99	8.46	0.85		DoB	2.14	16.03	15.89	3.72
2.2	{G1,R1}	99.83	86.80	92.13	99.33	3.4	{G1,R1}	96.93	55.50	64.20	93.60
	{G1,R2}	99.97	94.20	95.43	99.20		{G1,R2}	97.47	78.03	76.73	93.10
	{G2,R1}	99.40	80.60	73.87	98.73		{G2,R1}	90.93	44.10	36.97	87.37
	{G2,R2}	99.00	77.23	82.90	97.10		{G2,R2}	92.47	41.77	53.90	83.70
	DoB	0.44	7.47	9.72	1.03		DoB	3.24	16.58	16.82	4.76
2.4	{G1,R1}	99.67	84.73	90.47	98.90	3.6	{G1,R1}	93.97	44.83	54.87	90.70
	{G1,R2}	99.83	92.97	94.67	99.07		{G1,R2}	94.67	73.40	72.30	87.70
	{G2,R1}	99.30	76.93	74.03	98.23		{G2,R1}	85.83	38.17	26.37	78.67
	{G2,R2}	98.67	72.00	78.37	96.97		{G2,R2}	88.47	32.57	47.13	74.97
	DoB	0.52	9.18	9.77	0.95		DoB	4.29	18.14	19.04	7.41
2.6	{G1,R1}	99.60	80.50	86.27	98.17	3.8	{G1,R1}	90.00	37.87	47.43	85.60
	{G1,R2}	99.73	89.73	93.07	98.53		{G1,R2}	92.27	66.67	66.60	83.07
	{G2,R1}	98.83	69.83	64.70	97.03		{G2,R1}	78.87	34.40	18.90	71.20
	{G2,R2}	98.30	67.03	74.43	95.27		{G2,R2}	82.20	24.30	41.43	69.30
	DoB	0.67	10.41	12.58	1.47		DoB	6.34	18.18	19.65	8.23
2.8	{G1,R1}	99.30	75.40	82.03	97.53	4.0	{G1,R1}	85.30	36.57	41.43	78.60
	{G1,R2}	99.60	87.03	90.17	98.23		{G1,R2}	90.43	57.67	57.30	75.20
	{G2,R1}	98.47	63.67	56.03	96.27		{G2,R1}	69.23	29.90	15.27	64.37
	{G2,R2}	97.97	63.43	70.47	93.63		{G2,R2}	76.93	22.77	30.67	58.10
	DoB	0.75	11.25	14.83	2.03		DoB	9.33	15.06	17.71	9.50

Table 6. Verification accuracy and DoB (%) across different *gender* subgroups with varying intensities of brightness corresponding to the MORPH dataset. * represents a relatively high disparity in model performance across different subgroups on undistorted images.

β		LCNN-29	SENet50	ResNet50	ArcFace
1.0	G1	100.00	97.90	98.27	99.90
	G2	99.83	91.97	93.30	99.67
	DoB	0.12	4.19*	3.51*	0.16
1.5	G1	99.00	79.97	55.33	94.53
	G2	97.17	56.40	24.17	83.27
	DoB	1.29	16.67	22.03	7.96
2.0	G1	88.13	39.50	0.83	60.03
	G2	79.97	25.00	1.10	31.03
	DoB	5.77	10.25	0.19	20.51
2.5	G1	64.57	16.77	0.30	30.67
	G2	46.40	18.03	1.43	12.07
	DoB	12.85	0.89	0.80	13.15

Table 7. Verification accuracy and DoB (%) across different *gender* subgroups with varying intensities of brightness corresponding to the MUCT dataset.

β		LCNN-29	SENet50	ResNet50	ArcFace
1.0	G1	100.00	100.00	100.00	100.00
	G2	100.00	99.93	99.97	100.00
	DoB	0.00	0.05	0.02	0.00
1.5	G1	100.00	99.47	98.70	100.00
	G2	100.00	99.60	97.50	100.00
	DoB	0.00	0.09	0.85	0.00
2.0	G1	99.97	91.93	44.50	99.77
	G2	99.93	89.83	57.97	99.83
	DoB	0.03	1.48	9.52	0.04
2.5	G1	97.10	76.33	0.97	90.67
	G2	99.07	64.83	2.87	90.23
	DoB	1.39	8.13	1.34	0.31

Table 8. Verification accuracy and DoB (%) across the intersectional subgroups of *gender* and *race* with varying intensities of brightness corresponding to the MORPH dataset.

β		LCNN-29	SENet50	ResNet50	ArcFace
1.0	{G1,R1}	100.00	97.53	98.20	100.00
	{G1,R2}	100.00	99.27	99.17	99.97
	{G2,R1}	99.97	96.40	95.53	99.77
	{G2,R2}	99.90	94.10	96.67	99.57
	DoB	0.05	2.17	1.61	0.20
1.5	{G1,R1}	99.37	78.33	45.17	95.17
	{G1,R2}	99.97	94.13	88.57	99.30
	{G2,R1}	96.77	53.97	13.63	78.97
	{G2,R2}	98.90	76.57	58.80	97.20
	DoB	1.39	16.53	31.10	9.28
2.0	{G1,R1}	85.20	27.67	1.53	54.70
	{G1,R2}	98.63	54.67	25.47	87.13
	{G2,R1}	73.60	15.57	1.27	30.77
	{G2,R2}	87.37	12.97	1.97	68.90
	DoB	10.26	19.07	11.94	23.78
2.5	{G1,R1}	50.17	12.40	0.93	23.43
	{G1,R2}	91.43	19.70	1.47	59.17
	{G2,R1}	31.10	8.10	1.13	12.97
	{G2,R2}	53.23	5.73	1.23	33.50
	DoB	25.27	6.13	0.22	19.80

Table 10. Verification accuracy and DoB (%) across the intersectional subgroups of *gender* and *race* subgroups with varying intensities of Gaussian noise corresponding to the MORPH dataset.

σ		LCNN-29	SENet50	ResNet50	ArcFace
0	{G1,R1}	100.00	97.53	98.20	100.00
	{G1,R2}	100.00	99.27	99.17	99.97
	{G2,R1}	99.97	96.40	95.53	99.77
	{G2,R2}	99.90	94.10	96.67	99.57
	DoB	0.05	2.17	1.61	0.20
20	{G1,R1}	100.00	96.73	95.87	100.00
	{G1,R2}	99.93	98.37	97.77	99.80
	{G2,R1}	99.77	93.5	87.77	99.47
	{G2,R2}	99.77	90.90	91.00	99.07
	DoB	0.12	3.33	4.56	0.41
30	{G1,R1}	99.97	95.10	93.67	99.60
	{G1,R2}	99.80	96.63	93.23	99.13
	{G2,R1}	99.70	88.27	72.97	98.23
	{G2,R2}	98.93	88.33	82.03	97.87
	DoB	0.46	4.41	9.93	0.80
40	{G1,R1}	99.93	90.63	90.10	98.57
	{G1,R2}	99.83	87.10	79.37	93.27
	{G2,R1}	99.30	83.43	57.77	92.80
	{G2,R2}	98.40	82.27	65.50	92.30
	DoB	0.70	3.79	14.39	2.92

Table 9. Verification accuracy and DoB (%) across different *race* subgroups with varying intensities of Gaussian noise corresponding to the MORPH dataset.

σ		LCNN-29	SENet50	ResNet50	ArcFace
0	R1	100.00	98.53	98.80	99.96
	R2	100.00	99.26	99.16	99.93
	DoB	0.00	0.52	0.26	0.03
20	R1	100.00	96.73	95.70	99.90
	R2	100.00	97.96	97.96	99.80
	DoB	0.00	0.87	1.60	0.07
30	R1	100.00	95.26	89.96	99.30
	R2	99.99	95.70	93.06	98.86
	DoB	0.01	0.31	2.19	0.31
40	R1	99.76	89.70	74.76	97.43
	R2	99.56	88.53	70.10	94.66
	DoB	0.14	0.83	3.30	1.96

Table 11. Verification accuracy and DoB (%) across different *race* subgroups with varying intensities of salt and pepper noise corresponding to the MORPH dataset.

p		LCNN-29	SENet50	ResNet50	ArcFace
0.00	R1	100.00	98.53	98.80	99.96
	R2	100.00	99.26	99.16	99.93
	DoB	0.00	0.52	0.26	0.03
0.03	R1	99.96	93.56	90.93	55.36
	R2	99.83	96.36	96.50	71.20
	DoB	0.09	1.98	3.94	11.20
0.06	R1	99.53	81.26	66.83	7.30
	R2	99.76	88.86	83.23	22.66
	DoB	0.16	5.37	11.60	10.86
0.09	R1	98.60	69.76	34.60	2.26
	R2	99.10	75.86	55.53	3.76
	DoB	0.35	4.31	14.80	1.06

Table 12. Verification accuracy and DoB (%) across the intersectional subgroups of *gender* and *race* with varying intensities of salt and pepper noise corresponding to the MORPH dataset.

p		LCNN-29	SENet50	ResNet50	ArcFace
0.00	{G1,R1}	100.00	97.53	98.20	100.00
	{G1,R2}	100.00	99.27	99.17	99.97
	{G2,R1}	99.97	96.40	95.53	99.77
	{G2,R2}	99.90	94.10	96.67	99.57
	DoB	0.05	2.17	1.61	0.20
0.03	{G1,R1}	99.87	93.53	93.97	63.70
	{G1,R2}	99.93	96.27	96.13	69.53
	{G2,R1}	99.67	85.90	76.17	48.57
	{G2,R2}	99.10	87.17	87.37	71.87
	DoB	0.38	4.98	8.97	10.48
0.06	{G1,R1}	99.70	83.70	79.80	10.03
	{G1,R2}	99.73	89.13	83.67	25.97
	{G2,R1}	97.97	68.13	41.03	7.23
	{G2,R2}	98.33	70.70	59.33	17.37
	DoB	0.92	10.12	19.75	8.39
0.09	{G1,R1}	98.57	70.23	54.77	2.33
	{G1,R2}	99.37	75.03	51.33	6.87
	{G2,R1}	94.70	56.50	24.23	3.70
	{G2,R2}	95.60	53.17	33.40	4.43
	DoB	2.26	10.55	14.55	1.90

Table 13. Verification accuracy and DoB (%) across different *race* subgroups with varying resolution corresponding to the MORPH dataset.

		LCNN-29	SENet50	ResNet50	ArcFace
96×96	R1	100.00	98.10	98.40	100.00
	R2	100.00	99.10	99.07	99.87
	DoB	0.00	0.71	0.47	0.09
64×64	R1	100.00	97.63	97.87	99.77
	R2	100.00	99.07	98.87	99.90
	DoB	0.00	1.02	0.71	0.09
48×48	R1	100.00	95.63	95.97	98.73
	R2	99.93	97.37	96.73	99.30
	DoB	0.05	1.23	0.54	0.40
32×32	R1	99.83	76.50	76.10	66.60
	R2	99.73	82.53	85.73	73.43
	DoB	0.07	4.26	6.81	4.83
28×28	R1	98.70	51.47	43.07	11.00
	R2	99.03	50.03	60.87	12.60
	DoB	0.23	1.02	12.59	1.13

Table 14. Verification accuracy and DoB (%) across the intersectional subgroups of *gender* and *race* with varying resolution corresponding to the MORPH dataset.

		LCNN-29	SENet50	ResNet50	ArcFace
96×96	{G1,R1}	100.00	97.10	98.23	100.00
	{G1,R2}	100.00	99.27	99.13	99.93
	{G2,R1}	99.97	95.63	94.53	99.53
	{G2,R2}	99.87	95.13	96.33	99.40
	DoB	0.06	1.86	2.05	0.29
64×64	{G1,R1}	100.00	97.17	97.20	99.90
	{G1,R2}	99.97	99.00	98.73	99.90
	{G2,R1}	99.97	94.20	90.40	99.33
	{G2,R2}	99.83	93.97	93.97	99.23
	DoB	0.08	2.43	3.69	0.36
48×48	{G1,R1}	99.90	94.30	95.90	99.23
	{G1,R2}	100.00	97.80	97.70	99.33
	{G2,R1}	99.87	91.70	83.63	97.13
	{G2,R2}	99.53	88.80	90.80	98.03
	DoB	0.20	3.83	6.30	1.05
32×32	{G1,R1}	99.67	82.70	83.37	68.40
	{G1,R2}	99.90	90.67	89.27	74.60
	{G2,R1}	99.30	71.00	52.50	58.63
	{G2,R2}	98.80	59.40	64.03	65.90
	DoB	0.48	13.67	17.04	6.61
28×28	{G1,R1}	99.00	56.67	57.47	9.80
	{G1,R2}	98.67	70.80	73.37	11.93
	{G2,R1}	95.40	38.07	26.17	9.77
	{G2,R2}	95.67	26.30	34.80	10.67
	DoB	1.91	19.7	21.48	1.01

Table 15. Verification accuracy and DoB (%) across different *gender* subgroups with varying resolution corresponding to the MUCT dataset.

		LCNN-29	SENet50	ResNet50	ArcFace
96×96	G1	100.00	100.00	100.00	100.00
	G2	99.57	99.96	99.96	100.00
	DoB	0.30	0.03	0.03	0.00
64×64	G1	100.00	100.00	100.00	100.00
	G2	100.00	99.93	99.96	100.00
	DoB	0.00	0.05	0.03	0.00
48×48	G1	100.00	99.93	99.93	99.96
	G2	99.96	99.66	99.73	99.93
	DoB	0.03	0.19	0.14	0.02
32×32	G1	98.68	98.33	98.30	90.93
	G2	99.86	97.16	94.66	88.56
	DoB	0.83	0.83	2.57	1.68
28×28	G1	99.86	94.16	88.10	13.86
	G2	99.26	84.63	80.50	25.60
	DoB	0.42	6.74	5.37	8.30