ARoFace: Alignment Robustness to Improve Low-Quality Face Recognition

Supplementary Material

West Virginia University

1 Defining Allowed Perturbation Set $\mathcal{S}$

Conventionally face detection and alignment are performed before training the FR module, ensuring the consistency of landmarks' positions across all training instances according to the alignment template. Therefore, the initial positions of landmarks (u_p, v_p) match that of the alignment template. Additionally, one can determine the final positions of landmarks after applying the invertible affine transformation T_θ, as described in Equation 4 of Manuscript. Consequently, the displacement vector's upper bound $\overline{\mathbf{f}}$ is calculable by considering the transformation components' upper bounds, namely the maximum permissible rotation, translation, and scaling:

$$\overline{\mathbf{f}}_p = (u_p^{'} - u_p, v_p^{'} - v_p); \quad (u', v') = T_{\overline{\theta}}^{-1}(P_u(j), P_v(i)), \tag{1}$$

where $\overline{\theta}$ denotes the upper bound for the transformation parameters, *I.e.,* maximum rotation, scaling and translation.

In the same way, the displacement vector is calculated during the training for defining $\mathcal{S}$:

$$\mathbf{f}_p^{\theta} = (u_p^{'} - u_p, v_p^{'} - v_p); \quad (u', v') = T_{\theta}^{-1}(P_u(j), P_v(i)). \tag{2}$$

Finally, $\overline{f}_p = ||\overline{\mathbf{f}}_p||_2$ and we define the allowed transformation using Equation 8 of Manuscript without requiring landmark estimation.

1.1 Experiment on Upper Bound of Individual Transformations

Table 1: Experiments on the upper bound of affine transformations. TinyFace performance is the Rank1 identification. Verification performance TAR@FAR=1e-4 is reported for IJB-B and IJB-C. Training data is CASIA-WebFace, backbone is ResNet-100 and ArcFace is the objective function.

	Scaling			Rotation			Translation		
	0.01	0.1	0.5	0.01	0.1	0.2	0.01	0.1	0.2
TinyFace	61.37	62.11	62.24	50.32	53.76	57.26	54.85	57.31	56.18
IJB-B	58.40	35.17	16.92	55.68	37.03	1.76	46.56	9.49	2.60
IJB-C	52.90	48.84	41.74	41.49	16.20	5.71	52.84	36.94	6.36

With the predefined face alignment template, the allowed perturbation set $\mathcal{S}$ is directly linked to $\overline{\theta}$, which defines the maximum allowed values for individual transformations. Here, in Table 1, we conduct experiments on the upper bounds of the components of transformation T_θ, *I.e.,* scaling, translation, and rotation. The results indicate consistent improvement in TinyFace performance

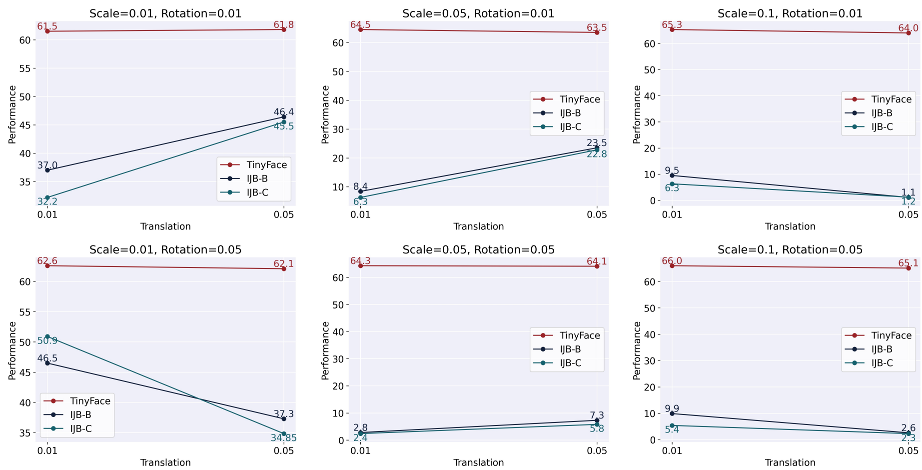

Fig. 1: Experiments on different perturbation budgets. Tinyface performance is the Rank1 identification. IJB-B and IJB-C performance is the TAR@FAR=$1e-5$.

with increased upper bounds across different components. Furthermore, scaling has a more significant impact on TinyFace performance, aligning with previous research on the effects of augmentation in FR performance [1,2,4]. Moreover, the results suggest that the performance on datasets containing both HQ and LQ instances is susceptible to the high values of transformation upper bounds, leading to a significant reduction in performance. We attribute this to the nature of the images in these datasets. Since both HQ and LQ samples are present in these benchmarks, excessive augmentation reduces the discriminative power of the FR model on HQ images, thereby diminishing performance in these evaluations.

2 Experiments on Perturbation Budgets

We conduct experiments to evaluate the impact of perturbation budgets α, corresponding to the components of spatial transformation, *I.e.,* scale, rotation, and translation, as shown in Figure 1. We use CASIA-WebFace as the training dataset, and ArcFace serves as the objective. The results demonstrate a positive correlation between the scaling budget and TinyFace performance; specifically, a larger scaling budget significantly enhances Rank1 TinyFace identification. This outcome is anticipated, as larger scaling perturbations produce smaller faces, thereby improving performance on LQ samples.

However, a consistent increase in the scaling budget leads to a significant decrease in verification performance on the IJB-B and IJB-C datasets. We expected these results since IJB-B and IJB-C consist of both HQ and LQ images and severe scaling results in reducing the discriminative power on the HQ images [1–3]. Aiming for a generalizable FR module, we have selected a scaling budget of 0.01. Furthermore, a rotation budget of 0.01 results in superior performance across IJB-B and IJB-C evaluation, indicating potential misalignment in HQ samples. Additionally, the translation budget of 0.01 results in balanced performance across TinyFace and other datasets. CNNs are somewhat translation-invariant, partly due to the nature of pooling operations. Thus, this minor perturbation aids in enhancing the model's robustness to misalignment. It is important to note that extreme translation might zero out essential parts of the face, rendering learning infeasible, unlike rotation, which does not remove any part of the image.

3 Experiments on PGD Steps k

In Table 2 we study the impact of k on the performance across TinyFace, IJB-B, and IJB-C. We experiment using $k \in \{1, 2, 3\}$. As Table 2 shows, the performance gain with the increase in the PGD steps is marginal. However, the computational overhead increases drastically with the increase in the number of PGD steps (detailed in Section 4.6 of Manuscript). Therefore, we employ $k = 1$ across our experiments. Also, the marginal reduction of the performance on IJB-B and IJB-C is in line with observations of [1–3] that these datasets contain both HQ and LQ instances and too much change in the input images, *I.e.,* augmentations, results in performance reduction in them.

Table 2: Experiments on the number of PGD steps.

k	TinyFace	IJB-B	IJB-C
1	65.26	58.47	52.88
2	65.28	58.54	52.68
3	65.35	58.50	52.48

References

1. Kim, M., Jain, A.K., Liu, X.: Adaface: Quality adaptive margin for face recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18750–18759 (2022)
2. Liu, F., Kim, M., Jain, A., Liu, X.: Controllable and guided face synthesis for unconstrained face recognition. In: Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XII. pp. 701–719. Springer (2022)
3. Saadabadi, M.S.E., Malakshan, S.R., Zafari, A., Mostofa, M., Nasrabadi, N.M.: A quality aware sample-to-sample comparison for face recognition. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 6129–6138 (2023)
4. Shi, Y., Jain, A.K.: Boosting unconstrained face recognition with auxiliary unlabeled data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2795–2804 (2021)