Supplementary Meterial of GB-CosFace

1 Details of Softmax-based Loss Derivation

For the objective design of face recognition, following the multi-classification pipeline, the direct idea is to constrain the target score larger than the maximum non-target score. For example, given a training sample and its label y, the base objective is as follows:

$$\mathcal{O}_{base} = ReLU(max(p_i) - p_y, 0) \tag{1}$$

Where p_y is the target score and p_i is the non-target score. ReLU is equivalent to the function $max(\cdot, 0)$, which is added to avoid overfitting.

For face recognition, as has been discussed in **Section 2** in the main text, the predicted score can be represented as the cosine of the face feature vector and the prototype, and usually, the margin is introduced for stricter constraints. Therefore, we can get the following training objective for face recognition:

$$\mathcal{O}_S = ReLU(max(cos\theta_i) - (cos(\theta_y + m_\theta) - m_p)) \tag{2}$$

However, generally, we use the softmax-based loss for training, which is the smooth form of $\mathcal{O}_S$ based on the following equations:

$$ReLU(x) = \lim_{s\to\infty} \frac{1}{s} log(1 + e^{sx}) \tag{3}$$

$$max(p_i) = \lim_{s\to\infty} \frac{1}{s} log \sum_{i=1}^{N} e^{sp_i} \tag{4}$$

The general softmax-based loss is as follows.

$$\mathcal{L}_S = -log \frac{e^{s(cos(\theta_y+m_\theta)-m_p)}}{e^{s(cos(\theta_y+m_\theta)-m_p)} + \sum_i e^{scos\theta_i}} \tag{5}$$

Substituting **Equ.** 3 and **Equ.** 4 into **Equ.** 2, we can reach the following equation.

$$\begin{aligned}
\mathcal{O}_S &= ReLU(max(cos\theta_i) - (cos(\theta_y + m_\theta) - m_p)) \\
&= \lim_{s\to+\infty} \frac{1}{s} log(1 + e^{(log \sum_{i=i,i\neq y}^{N} e^{scos\theta_i}) - s\cdot(cos(\theta_y+m_\theta)-m_p)}) \\
&= \lim_{s\to+\infty} \frac{1}{s} log(1 + \frac{\sum_{i=1,i\neq y}^{n} e^{scos\theta_i}}{e^{s\cdot(cos(\theta_y+m_\theta)-m_p}}) \\
&= \lim_{s\to+\infty} -\frac{1}{s} log \frac{e^{s(cos(\theta_y+m_\theta)-m_p)}}{e^{s(cos(\theta_y+m_\theta)-m_p)} + \sum_{i=1,i\neq y}^{n} e^{scos\theta_i}} \\
&= \lim_{s\to+\infty} \frac{1}{s} \mathcal{L}_S
\end{aligned} \tag{6}$$

In **Equ.** 6, if we take a fixed value of s rather than the limit of positive infinity, the form of the softmax-based loss can be reached.

2 Details of GB-CosFace Antetype Derivation

As has been discussed in **Section 3.1** in the main text, the original objective is as follows.

$$\begin{cases} \mathcal{O}_T = ReLU(p_v - (p_y - m)) \\ \mathcal{O}_N = ReLU(max(p_i) - (p_v - m)) \end{cases} \tag{7}$$

Similar to **Equ.** 6, we substitute **Equ.** 3 and **Equ.** 4 into **Equ.** 9 as follows.

$$\begin{aligned} \mathcal{O}_T &= ReLU(p_v - (p_y - m)) \\ &= \lim_{s \to +\infty} \frac{1}{s} log(1 + e^{s(p_v - (p_y - m))}) \\ &= \lim_{s \to +\infty} \frac{1}{s} log(1 + \frac{e^{sp_v}}{e^{s(p_y - m)}}) \\ &= \lim_{s \to +\infty} -\frac{1}{s} log(\frac{e^{s(p_y - m)}}{e^{s(p_y - m)} + e^{sp_v}}) \end{aligned} \tag{8}$$

$$\begin{aligned} \mathcal{O}_N &= ReLU(max(p_i) - (p_v - m)) \\ &= \lim_{s \to +\infty} \frac{1}{s} log(1 + e^{log \sum_i e^{sp_i} - s(p_v - m)}) \\ &= \lim_{s \to +\infty} \frac{1}{s} log(1 + \frac{\sum_i e^{sp_i}}{e^{s(p_y - m)}}) \\ &= \lim_{s \to +\infty} -\frac{1}{s} log(\frac{e^{s(p_y - m)}}{e^{s(p_y - m)} + \sum_i e^{sp_i}}) \end{aligned} \tag{9}$$

Taking parameter s as a fixed value rather than positive infinity, we can reach the following loss.

$$\begin{cases} \mathcal{L}_{T1} = -log\frac{e^{s(p_y - m)}}{e^{s(p_y - m)} + e^{sp_v}} \\ \mathcal{L}_{N1} = -log\frac{e^{s(p_v - m)}}{e^{s(p_v - m)} + \sum_i e^{sp_i}} \end{cases} \tag{10}$$

3 Proof of the Compatibility with CosFace

The final loss is as follows.

$$\begin{aligned} \mathcal{L}_{GB-CosFace} = &-\frac{1}{2} log\frac{e^{2s(p_y - m)}}{e^{2s(p_y - m)} + e^{2sp_v}} \\ &-\frac{1}{2} log\frac{e^{2s(p_v - m)}}{e^{2s(p_v - m)} + e^{2sp_n}} \end{aligned} \tag{11}$$

Where $p_n = \frac{1}{s} log \sum_i e^{sp_i}$. The parameter p_v is determined by the following equation.

$$p_v = \alpha p_{vg} + (1 - \alpha)\hat{p}_v \tag{12}$$

Where $\hat{p}_v = (p_y + p_n)/2$, and $p_{vg}$is the global boundary.

The proposed GB-CosFace has the following property.

Property 3 For $\alpha = 0$, GB-CosFace with the scale parameter s and the margin parameter m is equivalent to CosFace with scale parameter s and the margin parameter $2m$.

Proof. To prove the compatibility with CosFace[1, 2], we calculate the gradient for the target score $\mathcal{G}^y_{CosFace}$ and the gradient for the non-target score $\mathcal{G}^i_{CosFace}$ under CosFace framework.

$$\mathcal{G}^y_{CosFace} = -\frac{s \cdot \sum_i e^{sp_i}}{e^{s(p_y - m)} + \sum_i e^{sp_i}} \tag{13}$$

$$\mathcal{G}^i_{CosFace} = \frac{s \cdot e^{sp_i}}{e^{s(p_y - m)} + \sum_i e^{sp_i}} \tag{14}$$

For GB-CosFace with $\alpha = 0$, $p_v = (p_y + p_n)/2$. The gradient for the target score $\mathcal{G}^y_{GB}$ is as follows. Note that p_v is a detached parameter which does not require gradients.

$$\begin{aligned}
\mathcal{G}^y_{GB} &= \frac{-s \cdot e^{2sp_v}}{e^{2s(p_y - m)} + e^{2sp_v}} \\
&= \frac{-s \cdot e^{2s((p_y + p_n)/2)}}{e^{2s(p_y - m)} + e^{2s((p_y + p_n)/2)}} \\
&= \frac{-s \cdot e^{s(p_y + p_n)}}{e^{2s(p_y - m)} + e^{s(p_y + p_n)}} \\
&= \frac{-s \cdot e^{sp_n}}{e^{s(p_y - 2m)} + e^{sp_n}} \\
&= -\frac{s \cdot \sum_i e^{sp_i}}{e^{s(p_y - 2m)} + \sum_i e^{sp_i}} = \mathcal{G}^y_{CosFace}
\end{aligned} \tag{15}$$

The gradient for the non-target score $\mathcal{G}^i_{GB}$ is as follows.

$$
\begin{aligned}
\mathcal{G}_{GB}^{i} &= \frac{\partial \mathcal{L}_{GB-CosFace}}{\partial p_n} \cdot \frac{\partial p_n}{\partial p_i} \\
&= \frac{s \cdot e^{2sp_n}}{e^{2s(p_v-m)} + e^{2sp_n}} \cdot \frac{e^{sp_i}}{\sum_i e^{sp_i}} \\
&= \frac{s \cdot e^{2sp_n}}{e^{s((p_y+p_n)-2m)} + e^{2sp_n}} \cdot \frac{e^{sp_i}}{\sum_i e^{sp_i}} \\
&= \frac{s \cdot e^{sp_n}}{e^{s(p_y-2m)} + e^{sp_n}} \cdot \frac{e^{sp_i}}{\sum_i e^{sp_i}} \\
&= \frac{s \cdot \sum_i e^{sp_i}}{e^{s(p_y-2m)} + e^{sp_n}} \cdot \frac{e^{sp_i}}{\sum_i e^{sp_i}} \\
&= \frac{s \cdot e^{sp_i}}{e^{s(p_y-2m)} + \sum_i e^{sp_i}} = \mathcal{G}_{CosFace}^{i}
\end{aligned}
\tag{16}
$$

4 Effect of the update rate γ

We expect to calculate $\hat{p}_v$ in **Equ.** 12 for each sample in the data set and get the mean value as the threshold p_v. The momentum update strategy is applied to update the mean of $\hat{p}_v$ for each batch.

$$p_{vg} = (1-\gamma)p_{vg} + \gamma p_{vb} \tag{17}$$

Where $\gamma \in [0,1]$ is the update rate, p_{vb} is the mean of p_v in a batch. According to our experience, as long as this γ is set in a small reasonable range, it will not have much impact on our experimental results. To evaluate the effectiveness of the update rate γ, we empirically set $s = 32$, $m = 0.16$, $\alpha = 0.15$ and focus on the setting of γ. As can be seen from the **Table** 1, when the γ value is set relatively small, around 0.01, the performance of the model can maintain relatively good results. Therefore, we empirically set γ to 0.01 is reasonable.

Table 1. The results of the proposed GB-CosFace under different settings of update rate γ.

Settings	IJB-C(TAR)
FAR=1e-4, R50, α=0.15, γ=0.001	96.25
FAR=1e-4, R50, α=0.15, γ=0.003	96.27
FAR=1e-4, R50, α=0.15, γ=0.005	96.19
FAR=1e-4, R50, α=0.15, γ=0.01	96.24
FAR=1e-4, R50, α=0.15, γ=0.05	96.33

References

1. Wang, F., Cheng, J., Liu, W., Liu, H.: Additive margin softmax for face verification. IEEE Signal Processing Letters (2018)

2. Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J., Li, Z., Liu, W.: Cosface: Large margin cosine loss for deep face recognition. In: Conference on Computer Vision and Pattern Recognition (CVPR). (2018)