We appreciate the constructive comments and valuable points raised by the reviewers and the editor. We carefully read and discussed the opinions and doubts of each reviewer, and our responses are as follow.

The writing problems. The writing problems. Some expressions in this paper are not proper or brief enough. The writing needs to be more accurate, concise and academic. We will try our best to make the writing better.

The consistency between training and inference. Both outputs of training and inference are feature vectors of input images. The element weights are only used in the proposed element-weighted TriHard loss during training rather than components of the network. They make the loss function more accurate by focusing on different semantic parts among persons. It is an improvement on loss function without increasing the inference cost. Therefore, the consistency between training and inference is guaranteed.

Doubts on Table 1. Table 1 in our paper is the results of ablation study. These results fluctuate with or without some certain items of the proposed loss function to test the efficiency of each part. The analysis is detailed in chapter 5.1.

Results fluctuate with parameter t. Parameter t is a hyper-parameter of element-weighted TriHard loss. Like the threshold in triplet loss, such hyper-parameters influence the performance of networks dramatically which is obtained by experience and experiments. That makes the training complex. It is a point to replace t with a trainable parameter in our future work.

Doubts on mathematical symbols. We will add a parameter form into the next version of the paper.

The evidence of some claims. The claims in this paper without demonstration are proposed empirically. It will be more serious to cite previous work or attach experiment results to prove some of these claims. Such as the claim in L91-L93, we will attach an experiment result to it.

Without training dynamics. We are sorry to say that training dynamics figures are replaced by Figure 4, t-SNE visualization of feature distributions in our paper. Because of the limitation of pages and the effects they show, we made the decision.

Doubts on Figure 2. Figure 2 in our paper is plotted according to a classic visualization method CAM[1]. $D_{g1,g2}$ is not used.

The proposed method is simple. Our aim is to find a simple but serious method with good performance. Therefore, it is easy to follow and applied to other works.

To evaluate the method on larger datasets MSMT17. We replenish experiments on MSMT17[2] dataset, which contains more than 3 times identities than Market1501 and DukeMTMC-reID. The results are shown in Table 1.On AGW baseline, rank1 increases 3% and mAP 3.1% with $t = 0.3$. On BoT baseline, rank1 increases 4.7% and mAP 4.6% with $t = 0.3$. That proves the proposed loss improves the representational ability of the network to a great extent, which becomes more obvious on larger dataset.

Table 1: The performance of element-weighted TriHard loss on MSMT17

Method	MSMT17		Method	MSMT17	
	rank1	mAP		rank1	mAP
BoT	63.4%	45.1%	BoT+HNEWTH	**68.1%**	**49.7%**
AGW	68.2%	49.5%	AGW+HNEWTH	**71.2%**	**52.6%**

We demonstrate the importance of loss function in ReID networks. Only some simple improvements of TriHard loss will result in considerable enhancement of the performance. Without making any changes to the architecture of networks, it is easy to apply to other ReID or face recognition frameworks. The study of TriHard loss will also enlighten new researches on such commonly used loss functions in deep learning areas.

[1] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva and A. Torralba, "Learning Deep Features for Discriminative Localization," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) , 2016, pp. 2921-2929.

[2] L. Wei, S. Zhang, W. Gao and Q. Tian, "Person Transfer GAN to Bridge Domain Gap for Person Re-identification," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 79-88.