Response to Reviewer 2 Comments

1. Summary

Thank you for your insightful analysis and questions. Below, we have answered each of your questions and provided additional explanations and experimental results to clarify any gaps. Please note that due to time constraints, we were unable to conduct all the additional experiments, but we prioritized those that would add the most to the quality of the paper.

2. Point-by-point response to Comments and Suggestions for Authors

Comments 1: How can the cosine similarity distance threshold for the pseudo labeling generation process be selected? Is that sensitive to the final performance?

Response 1: *We acknowledge and appreciate your point regarding the lack of explanation for the threshold values used in the pseudo-labeling process. In this manuscript, there are three primary thresholds used.*

The first is the Euclidean distance threshold, t_d, used in Algorithm 1, which employs the average performance value from the test to detect outliers among multiple pseudo landmark samples. This is discussed in the manuscript.

We realized that the other two thresholds lacked specific explanations, so we've written more about them in Section 3.1.3. These two thresholds are derived from the pre-training of the face recognition task on the VGGFace2 and MS1MV3 datasets, respectively. They can typically be measured using the LFW verification test, a common validation test in face recognition. Therefore, the thresholds, t_{MS} and, t_{VGG} were selected to maximize performance in the LFW verification test for each dataset.

There are generally two usable distance metrics: measuring either the Euclidean distance or the angular distance between feature vectors. The choice of distance metric depends on whether the Euclidean or angular distance was set as the loss function. Using a different distance metric than the one set during training can significantly reduce performance. Hence, we can only use cosine similarity in the pseudo-labeling framework, as we employed a large margin cosine loss function that utilizes angular distance during pre-training.

We could have used Euclidean distance, but the trend of recent face recognition tasks such as ArcFace and PartialFC mentioned in the manuscript shows that angular distance is better for feature clustering than Euclidean distance. We chose to use the angular distance metric in our experiments to keep the performance of all tasks as high as possible due to the multi-task learning process.

#page 9

We have described how to label pseudo-landmarks and pseudo-pose labels. Next, we describe a framework for distributing face IDs to integrate the dataset in the following algorithm 2. First, most FR datasets are composed of data on celebrities and other famous individuals. Consequently, some duplicate IDs exist between the FR datasets. To filter out duplicates between the VGGFace2 [28] and MS1MV3 [29] datasets, we calculated the similarity threshold values t_{MS} and t_{VGG} while pretraining

the face recognition task with each dataset. During the pretraining process, we used the LFW [65] verification test, commonly employed to evaluate the performance of face recognition tasks, to measure the optimal similarity threshold that maximizes verification performance. This threshold was then used as the similarity threshold for detecting duplicates in Algorithm 2. Both angular and Euclidean distances can be used to measure similarity; however, the same distance metric must be employed in the loss function used during pretraining. Since we used the large margin cosine loss [38] function in our proposed method, we employed angular distance for cosine similarity in detecting data duplicates. For multitask learning, we filtered out duplicate IDs between VGGFace2 [28] and MS1MV3 [29] and added new IDs to 300W-LP [22] to pseudo-labeled training dataset.

Comments 2: Pseudo-labeling is a very promising direction, and there are several works related to pseudo-labeling for image classification and segmentation problems. Please add the following reference in your introduction section, which will help readers understand the potential of this field.

[1] Lee, D.-H.Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In Proceedings of the Workshop on Challenges in Representation Learning, International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; Volume 3.

[2] Yu, X., Ouyang, B., Principe, J. C., Farrington, S., Reed, J., & Li, Y. (2019, October). Weakly supervised learning of point-level annotation for coral image segmentation. In *OCEANS 2019 MTS/IEEE SEATTLE* (pp. 1-7). IEEE.

[3] Caron, Mathilde, et al. "Deep clustering for unsupervised learning of visual features." *Proceedings of the European conference on computer vision (ECCV)*. 2018.

Response 2: *Thank you for your understanding of the potential of pseudo-labeling and for recommending related work. We've added each reference to this manuscript in the related work and introduction. You can find them below.*

.

#page 2

"However, these approaches assume label availability for multiple tasks within a dataset. Consequently, they inherently suffer from challenges as a fundamental solution to the lack of face IDs and overall volume in datasets. Alternative approaches exist, such as improving data quantity through research in pseudo-labeling methods [22–26]. Inspired by multitask learning methods such as all-in-one CNN [27] and SwinFace [21] as well as pseudo-labeling methods [22 ,23] such as those proposed by Pan et al., we hypothesized that using datasets such as those for face recognition tasks that have numerous images with a variety of face IDs would be feasible for a task with a scarce number of face IDs."