Fully Unsupervised Person Re-Identification via Centroids and Neighborhoods Joint Learning

1st Qing Tang
Graduate School of Electrical Engineering
University of Ulsan
Ulsan, Korea
zucchini.tang@hotmail.com

2nd Kang-Hyun Jo
Graduate School of Electrical Engineering
University of Ulsan
Ulsan, Korea
acejo@ulsan.ac.kr

2022 IEEE 31st International Symposium on Industrial Electronics (ISIE) | 978-1-6654-8240-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISIE51582.2022.9831570

***Abstract*—This paper considers that the challenge of unsupervised person re-identification (re-ID) is generating high-quality pseudo labels. Recent label prediction methods can be mainly divided into Clustering-based Label Prediction (C-LP) and Similarity Measurements-based Label Prediction (SM-LP) methods. The existing researches only focus on improving the accuracy of one of the label generation method. In this letter, we first point out three complementarities between C-LP and SM-LP, including (1) interval of the pseudo label prediction (2) feature learning directions, and (3) inliers and outliers processing. Based on these three complementarities, we proposed a Joint Label Prediction (Joint-LP) method that can give full play to complementary advantages of C-LP and SM-LP. Moreover, we discover that standard Binary Cross Entropy (BCE) loss forces the unsupervised model to overfit the noisy labels, thereby leading the model training to fail. Therefore, we further proposed a Rectified Binary Cross Entropy (ReBCE) loss that is robust to label noise. The experimental results confirm the effectiveness of the proposed Joint-LP and ReBCE loss on two mainstream person re-ID datasets, Market-1501 and DukeMTMC-reID.**

***Index Terms*—Person re-identification, fully unsupervised learning, pseudo label prediction**

I. Introduction

The person re-identification (re-ID) systems aim to retrieve images that contain the same person. The supervised methods [1]–[3] require substantial labeled training data for achieving satisfying performance. Therefore, people pay more attention on unsupervised person re-ID methods, which do not need any labeled data to train the re-ID network. To make unsupervised training possible, the unsupervised model needs to generate the pseudo labels. Unlike human-labeled annotation, such generated labels contain noisy labels that substantially hinder the model's capability, thus the performances of the unsupervised re-ID methods still fall behind the supervised re-ID methods.

Based on learning strategy, unsupervised person re-ID can be generally divided into Fully Unsupervised Learning (FUL) methods [4]–[14] and Unsupervised Domain Adaption (UDA) methods [15]–[19]. The UDA outperforms FUL because UDA used a labeled source dataset to learn to extract better feature representation or to generate better pseudo labels. However, UDA still requires labeled information. In this study, we focus on the FUL-based person re-ID system, which trains the model without any manually annotated labels.

Based on the pseudo label prediction methods, unsupervised person re-ID can be generally divided into Clustering-based Label Prediction (C-LP) [6], [8], [9], [16], [19] and Similarity Measurements-based Label Prediction (SM-LP) [4], [5], [10], [15], where the C-LP methods maintain state-of-the-art performance to date by introducing an additional unsupervised clustering algorithm.

The core idea of C-LP is performing a clustering algorithm on Convolutional Neural Network (CNN) features to generate pseudo labels for training. Fan et al. [19] can be seen as an original work studying C-LP methods. They proposed a Progressive Unsupervised Learning (PUL) method to iterate clustering and fine-tune CNN step by step until convergence. Because the clustering results may noisy, subsequent researches [6], [8], [9], [16] mainly focus on refining noisy labels. Ge et al. [16] first grouped the features into M_t classes by clustering algorithm k-means [20], then they trained the model using the hard and soft pseudo-classes jointly to mitigate the effects of noisy labels. Yang et al. [6] generated pseudo-classes by clustering algorithm DBSCAN [21], then it further proposed a Dynamic and Symmetric Cross-Entropy loss (DSCE) to deal with noisy samples. In this letter, DBSCAN [21] is adopted because of its strong robustness against noisy samples.

There are two weaknesses in C-LP that are valuable to discuss but were ignored in the existing methods. (1) The intervals of the pseudo label prediction and model optimization are out of sync. More specifically, the model parameters are updated in every training iteration but label are predicted before every training epoch. This asynchronism hinders the model's performance because the model can not be updated based on sync updated labels. (2) The intra-class inliers can not perform intra-class differential learning because intra-class inliers share the same labels. As illustrated in Fig. 1(a), intra-class inliers are enforced class centroid-towards learning without considering neighborhood information. Moreover, how to deal with the un-clustered outliers is still an open question.

To tackle the weaknesses of C-LP, we propose a Joint Label Prediction (Joint-LP) to bound C-LP and SM-LP together to utilize the merits of SM-LP. Although SM-LP [4], [5], [15] achieve poorer performance than C-LP, SM-LP still enjoys three merits that are complementary to the C-LP. (1) The intervals of the label prediction and model optimization are

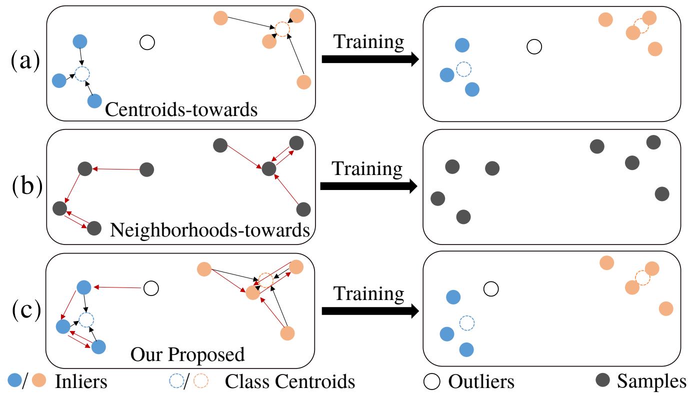

Fig. 1. The illustrations of (a) C-LP (class centroids-towards learning) (b) SM-LP (neighborhoods-towards learning) (b) proposed Joint-LP (both).

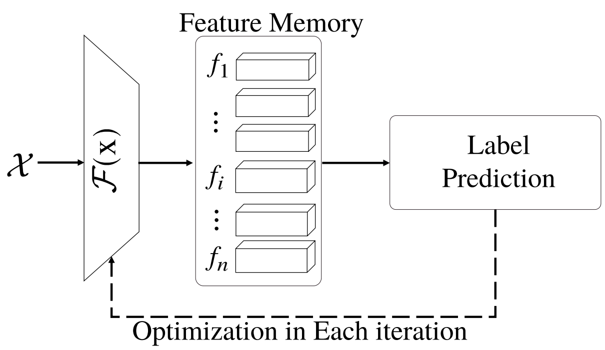

Fig. 2. General framework for FUL person re-ID methods.

synchronous. (2) the pseudo label of each sample is different to enforce samples learning towards their own nearest neighbor. (3) SM-LP assigns the label for every sample including the outliers. As illustrated in Fig. 1(b), every sample is enforced learning towards their own nearest neighbor.

Moreover, we discover that the traditional Binary Cross Entropy (BCE) loss achieved satisfying performance in supervised learning methods because of correct human-annotated labels, but BCE loss achieves poor performance in unsupervised learning methods because of extensive noisy pseudo labels. Therefore, to remedy this issue, we further propose a Rectified BCE (ReBCE) loss to make unsupervised training with BCE loss possible by alleviating model excessive attention on noise.

Our contributions are summarized as three-fold. (1) We propose a Joint-LP method to predict high-quality pseudo labels by utilizing complementarities between C-LP and SM-LP. (2) We propose a ReBCE loss to avoid the model pay more attention to noisy labels. (3) The proposed unsupervised person re-ID method achieves superior person Re-ID performance under the FUL setting on two large-scale datasets.

To the best of our knowledge, this letter is an original work studying and utilizing the complementarities between C-LP and SM-LP.

II. Proposed Method

A. Framework Overview

The general framework for FUL person re-ID is shown in Fig. 2. Given an unlabeled person image $x_{\{i|i=1,2,\ldots,n\}} \in \mathcal{X}$, d-dimensional feature f_i are extracted by backbone network $\mathcal{F}(\cdot)$ to form the feature memory $\mathcal{M}$, $f_{\{i|i=1,2,\ldots,n\}} \in \mathcal{M}$. i means the index of the image, which is fixed throughout training process, and n is the total numbers of images in $\mathcal{X}$. $\mathcal{M}$ store features for all images in $\mathcal{X}$, the size of $\mathcal{M}$ is $n \times d$.

Using $\mathcal{M}$, the proposed label prediction method Joint-LP predicts the pseudo label for every image in $\mathcal{X}$. Finally, $\mathcal{F}(\cdot)$ is optimized progressively with the proposed ReBCE loss based on pseudo labels step by step. Consequently, the key to improving model performance is to generate high-quality pseudo labels which can represent the unlabeled data domain distribution.

B. Joint Label Prediction (Joint-LP)

To generate high-quality pseudo labels, we propose a Joint-LP in this letter. The structure of Joint-LP is shown in Fig. 3. The Joint-LP consists of three components: C-LP, the proposed Dimension Increment pseudo-class Encoding method (DIE), and SM-LP, which will be introduced one by one.

1) C-LP: The unsupervised clustering algorithm k-means [20] and DBSCAN [21] are widely used to generate pseudo labels in recent studies. Following the previous works [6], [8], [16], DBSCAN is used in here because it has stronger robustness against noisy samples. As illustrated in Fig. 3, given the feature memory $\mathcal{M}$, DBSCAN assigns pseudo-class $c_{\{i|i=1,2,\ldots,n\}}$ for every image $\in \mathcal{X}$ before every training epoch. DBSCAN assigns pseudo-class $c_i \geq 0$ to clustered inliers, and remains $c_i = -1$ to un-clustered outliers.

2) DIE Pseudo-class Encoding: The DBSCAN-based methods still face one challenge that the numbers of pseudo-classes keep changing during the whole training process. The centroid-based clustering algorithm K-means [20] generates certain cluster centroids, therefore a Fully Connected layer (FC-layer) can easily be adopted to output a probability vector for computing the cross-entropy classification loss or triplet

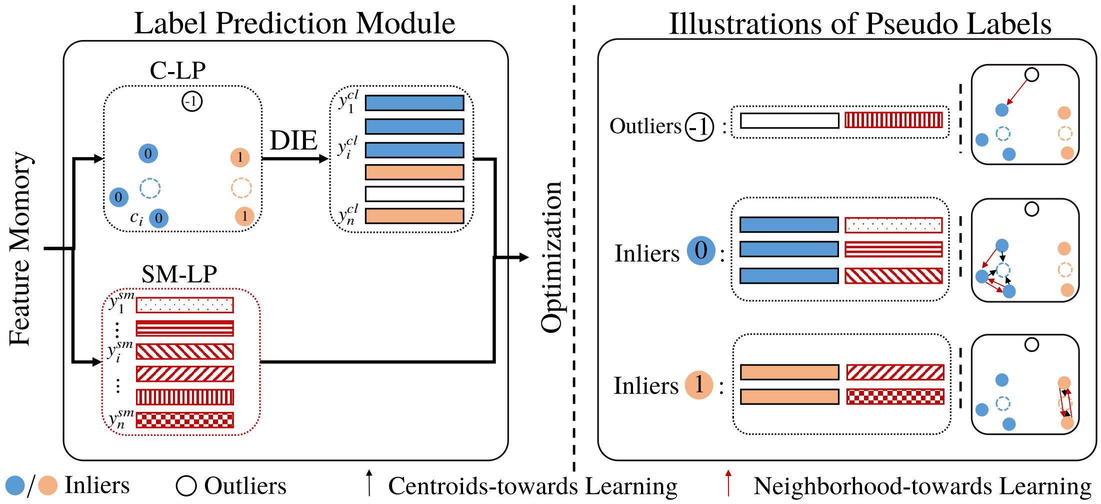

Fig. 3. The illustration of our proposed Jointly Label Prediction Module (Joint-LP).

loss [26] as [16], [19]. However, the number of pseudo-class predicted by DBSCAN [21] are constantly changing because DBSCAN only considers high-confident samples as clustered inliers. To address this issue, we proposed a Dimension Increment pseudo-class Encoding method (DIE) to equivalently encode 1-dimensional pseudo-class c_i to n-dimensional clustering-based pseudo label y_i^{cl}. Then, n independent binary classifiers can be adopted to compute loss functions easily.

The intuition of the proposed DIE is that, there may exist a number of inliers sharing the same pseudo-class, DIE directly sets these samples as mutual positive samples. For the inliers ($c_i \geq 0$), the c_i is encoded to y_i^{cl} using DIE as follows,

$$\text{Inliers:} \quad y_i^{cl}[j] = \begin{cases} 1 & c_j = c_i \\ -1 & c_j \neq c_i; \end{cases} \qquad i, j = 1, ..., n. \tag{1}$$

If a sample x_j has same pseudo-class with x_i ($c_j = c_i$), the x_j is a positive sample of x_i, therefore $y_i^{cl}[j]$ set to 1; Otherwise, $y_i^{cl}[j] = -1$. For the outliers ($c_i = -1$), DIE encodes c_i as follows,

$$\text{Outliers:} \quad y_i^{cl}[j] = \begin{cases} 1 & j = i \\ -1 & j \neq i; \end{cases} \qquad i, j = 1, ..., n. \tag{2}$$

where each outlier can be trained as an individual class. This operation is repeated until all samples in $\mathcal{X}$ are enumerated.

Finally, we obtain n numbers of n-dimensional clustering-based pseudo label y_i^{cl} as illustrated in Fig. 3. Our proposed DIE ensures equivalency between c_i and y_i^{cl}, and the value in y_i^{cl} points to the index of the samples that have the same pseudo-class with x_i in the meantime.

3) SM-LP: SM-LP methods [4], [5], [15] predicted positive labels by measuring the similarity among samples. Given the feature memory $\mathcal{M}$, the similarity of image x_i, notated as s_i, can be computed as:

$$s_i = \mathcal{M}[i] \times \mathcal{M}^\top \tag{3}$$

where s_i is an n-dimensional vector. $s_i[j]$ represents the similarity scores between x_i and the image $x_{\{j|j=1,...,n\}} \in \mathcal{X}$.

Existing positive sample selection strategies [4], [5], [15] selected positive samples for x_i based on its similarity s_i using some fixed rules. We use the latest and best positive sample selection methods MPLP [5]. The MPLP [5] used a pre-defined similarity threshold $t = 0.6$ and the cycle consistency to select positive neighbors for x_i. Finally, the similarity measurement-based pseudo label y_i^{sm} for x_i can be generated as:

$$y_i^{sm}[j] = \begin{cases} 1 & \text{if } x_j \text{ is a positive neighbor} \\ -1 & \text{Otherwise.} \end{cases} \tag{4}$$

where y_i^{sm} is an n-dimensional vector. SM-LP assigns distinct pseudo labels y_i^{sm} to every sample.

C. Rectified Binary Cross Entropy (ReBCE) Loss

To simplify the expression, we use asterisk symbol "*" to represent clustering-based information and similarity measurement-based information. For example, L^* can represent the loss of y^{cl} or y^{sm}, and y^* can represent y^{cl} or y^{sm}.

In supervised learning, the Binary Cross Entropy (BCE) loss with ground-truth labels has been well studied in previous researches [27], [28]. Inspired by [29], we discover that BCE loss poses a great challenge in unsupervised learning because of extensive noisy pseudo-labels. The BCE loss of classifying image x_i to its positive sample x_j can be computed as Eq.(5). The gradient of L_{bce}^* are represented as Eq.(6),

$$L_{bce}^* = -y_i^*[j] \times log(s_i[j]) \tag{5}$$

$$\frac{\partial L_{bce}^*}{\partial \theta} = -y_i^*[j] \times \frac{1}{s_i[j]} \times \partial_\theta s_i[j] \tag{6}$$

where θ means current network parameters. From Eq.(5) and Eq.(6), we can see a factor in BCE loss that samples with smaller similarity $s_i[j] \to 0$ are weighted more than higher similarity for gradient update. In supervised learning, this factor helps the model paying more attention to difficult samples. However, in unsupervised learning, small similarity samples may contain many false-positive noisy pseudo labels.

Therefore, using BCE loss might cause the model pay more attention to noises, thereby leading the model to fail.

We hence propose a Rectified Binary Cross Entropy (ReBCE) loss to address the above issue. The ReBCE loss is formulated as,

$$L^*_{\text{Rebce}} = \begin{cases} -y^*[j] \times log(\alpha) & \text{if } \ s_j[j] < \alpha \\ -y^*[j] \times log(s_i[j]) & \text{if } \ s_j[j] \geq \alpha. \end{cases} \tag{7}$$

where $\alpha \in [0, 1]$ is a pre-defined rectified parameter to control the small similarity score amplify gradient excessively by rectifying very small $s_i[j]$ to α.

D. Overall Loss

The effectiveness of Memory-based Multi-label Classification Loss (MMCL) in unsupervised multi-label person re-ID task is demonstrated in previous research [5]. Therefore, the network is simultaneously optimized with respect to the MMCL L^*_{mmcl} and the proposed ReBCE loss L^*_{Rebce} to achieve optimal model performances. The overall loss L can be computed by combining Eq.7 and Eq.8 as follows,

$$L^*_{mmcl} = \|s_i[j] - y_i^{sm}[j]\|^2 \tag{8}$$

$$L = \frac{1}{\eta}(L^{cl}_{mmcl} + L^{cl}_{\text{Rebce}} + L^{sm}_{mmcl} + L^{sm}_{\text{Rebce}}) \tag{9}$$

where $\eta = 4$ is a normalized coefficient to normalize the scale of the overall loss.

III. Experimental Results

A. Experiment Setting

We perform experiments on the two person re-ID datasets, Market-1501 (Market) and DukeMTMC-reID (Duke). Market [22] has 32,668 person images of 1,501 identities in total. Duke [23], [24] has 36,411 person images of 1,404 identities in total. Two evaluation metrics are used to measure model performance. The first one is the Cumulative Matching Characteristic (CMC) curve which represents the probability of top-k ranked gallery samples containing the query identity. The CMCs (%) of Rank-k (R-k) are reported. Another evaluation metric is the Mean Average Precision (mAP) (%).

The experiments are performed using one NVIDIA 1080Ti GPU with 11 GB of memory. The ResNet-50 [25] are adopted as the backbone network, which is pre-trained on ImageNet. The setting of backbone network follows the same setting in [4], [5], [15]. The input images are resized to 256×128. The training batch size is 64. The total training epoch is 40. The initial learning rate is 0.03, and it is divided by 10 after 30 epochs. We set the rectified parameter $\alpha = 0.2$ in ReBCE loss to achieve the best performance.

B. Ablation Study

1) Importance of Outliers: As shown in Table I, directly discarding outliers from training data cannot achieve satisfying results on both datasets. There are two reasons. (1) discarding outliers leads to a poor initial model because there are many outliers during the whole training process, especially in early epochs. (2) discarding outliers inhibits the model learning

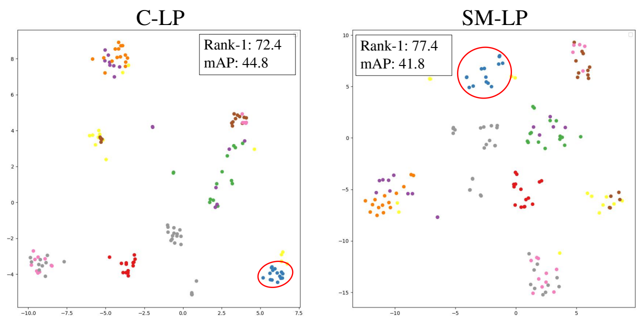

Fig. 4. The t-SNE [30] visualization on features representation of 10 identities. The different color points are donoted identities.

on difficult samples. Therefore, we train each outlier as an individual class, as mentioned in Eq. (2). The results verify the effectiveness of our proposed DIE, which treats each unclustered outlier as an individual class.

2) Effectiveness of Joint-LP: In order to verify the complementarities between C-LP and SM-LP, and the effectiveness of our proposed Joint-LP, we report comparison results of different label prediction methods in Table II and illustrate the t-SNE [30] visualization results in Fig. 4.

From the comparison between C-LP and SM-LP, three observations are obtained. 1) In table II, C-LP achieves better performance in mAP on two datasets. 2) SM-LP achieves better performance in Rank-k accuracy on two datasets. 2) In Fig. 4, C-LP generates closer and more compacter intra-class features than SM-LP. The reasons are that C-LP enforces centroid learning by assigning the same pseudo labels to the samples in the same cluster, therefore C-LP obtains higher clustering accuracy (in mAP) than SM-LP. Conversely, SM-LP enforces neighborhood learning by mining reliable positive samples around the sample, therefore SM-LP achieves higher ranking accuracy (in R-k) than C-LP. These results demonstrate that C-LP and SM-LP lead model to learn in different directions, and thus they can be complementary to each other in R-k accuracy and in mAP to achieve better performance. It is an important discovery for the current and future object re-ID research.

Based on the above discovery, we propose the Joint-LP in this letter. The proposed Joint-LP achieves the best performance by enforcing centroid-towards and neighborhood-towards learning collaboratively. It is also interesting to observe that, with the help of SM-LP, the upper bounds of mAP are also increased from 44.7% to 51.3% on Market-1501, and from 40.6% to 42.8% on DukeMTMC-reID. Same improvements are also observed that, with the help of C-LP, the upper bounds of ranking accuracy 77.4% and 63.6% are also increased on two datasets, respectively. The improvements further demonstrate the complementarity between C-LP and SM-LP, and the proposed Joint-LP can overcome their demerits and utilize their merits at the same time.

TABLE I. Ablation study on outliers. "✗": Training without outliers. "✓": Training each outlier as an individual class as Eq(2).

Outliers	Methods	Market-1501				DukeMTMC-reID			
		R-1	R-5	R-10	mAP	R-1	R-5	R-10	mAP
✗	C-LP	63.9	79.7	84.8	35.4	57.1	69.8	73.5	31.9
	Joint-LP (ours)	75.7	87.9	90.9	48.1	64.5	75.5	79.6	40.3
✓	C-LP	72.4	86.0	89.9	44.7	63.2	75.0	79.0	40.6
	Joint-LP (ours)	**78.4**	**88.7**	**91.7**	**51.3**	**66.2**	**77.6**	**81.1**	**42.8**

TABLE II. Comparison with different label prediction methods

Methods	Market-1501				DukeMTMC-reID			
	R-1	R-5	R-10	mAP	R-1	R-5	R-10	mAP
C-LP	72.4	86.0	89.9	44.7	63.2	75.0	79.0	40.6
SM-LP	77.4	87.3	90.3	41.8	63.6	75.0	79.4	39.0
Joint-LP (ours)	**78.4**	**88.7**	**91.7**	**51.3**	**66.2**	**77.6**	**81.1**	**42.8**

TABLE III. Comparison with different loss function

Methods (Loss function)	Market-1501				DukeMTMC-reID			
	R-1	R-5	R-10	mAP	R-1	R-5	R-10	mAP
MMCL	78.4	88.7	91.7	51.3	66.2	77.6	81.1	42.8
BCE	1.0	3.9	6.0	0.5	0.0	0.3	1.1	0.1
BCE + MMCL	0.9	2.7	4.7	0.3	0.6	1.6	2.4	0.3
ReBCE (ours)	78.8	89.5	92.8	50.6	66.1	77.8	81.5	42.3
ReBCE + MMCL (ours)	**80.3**	**90.8**	**93.2**	**55.1**	**67.8**	**78.4**	**81.6**	**44.0**

TABLE IV. Performance comparison with other FUL person re-ID methods on Market-1501 and DukeMTMC-ReID. "*": The MetaCam algorithm in DSCE [6] is not considered in this table, because MetaCam requires camera IDs. This paper performs comparison experiments in unknown camera IDs environment.

Label prediction methods	Method	Reference	Market				Duke			
			R-1	R-5	R-10	mAP	R-1	R-5	R-10	mAP
C-LP	BUC [13]	AAAI19	66.2	79.6	84.5	38.3	47.4	62.6	68.4	27.5
	DBC [14]	BMVC19	69.2	83.0	87.8	41.3	51.5	64.6	70.1	30.3
	DSCE* [6]	CVPR21	74.9	-	-	53.9	62.8	-	-	43.4
SM-LP	SSL [4]	CVPR20	71.7	83.8	87.4	37.8	52.5	63.5	68.9	28.6
	MLCReID [5]	CVPR20	80.3	89 4	92.3	45.5	65.2	75.9	80.0	40.2
	NNCT [10]	ICIP21	**82.0**	90.0	92.9	48.4	64.8	75.7	79.2	40.7
Combined	**Joint-LP** + **ReBCE** (ours)		80.3	**90.8**	**93.2**	**55.1**	**67.8**	**78.4**	**81.6**	**44.0**

3) Effectiveness of ReBCE Loss: We bring out that the traditional BCE loss cannot be directly adopted in the unsupervised multi-label classification task because of extensive noisy pseudo-labels. To demonstrate the above conjecture, we report the experimental results of different loss functions in Table III. Table III shows that using BCE loss (w/ or w/o MMCL) leads the experiments to fail on two datasets. The main reason is that BCE forces the model to pay more attention to noisy labels which leads to serious overfitting on noisy labels. Table III shows that the proposed ReBCE can solve this problem well, and ReBCE with or without MMCL achieves satisfying performance. These results demonstrate the effectiveness of the proposed ReBCE in the unsupervised multi-label classification task.

C. Comparision with Other FUL Methods

As shown in Table IV, we compare our method with other FUL person re-ID methods, including three C-LP based methods and three SM-LP based methods.

Compared to the SM-LP based method MLCReID [5], our method significantly outperforms it in mAP by 9.6% on Market-1501 and by 3.8% on DukeMTMC-reID because of the adding centroids-towards learning. The best SM-LP based method NNCT [10] achieved the best R-1 accuracy 82.0% in Market-1501. NNCT cannot achieve satisfying performance in mAP because it lacks the clustering information to enforce centroids-towards learning.

Compared to the best C-LP based method DSCE [6], our method significantly outperforms DSCE in R-1 accuracy by 5.4% on Market-1501 and by 5.0% ion DukeMTMC-reID because our method measures similarities among samples. It is noteworthy that, these specific and consistent improvements again demonstrate the importance of combining C-LP and SM-LP for the current and future object re-ID research. Finally, our method achieves the best performance with Rank-1 $= 80.3\%$, mAP $= 55.1\%$ on Market-1501, and Rank-1 $= 67.8\%$, mAP $= 44.0\%$ on DukeMTMC-reID.

IV. Conclusion

In this letter, we have presented a superior fully unsupervised person re-ID method. To the best of our knowledge, this letter is an original work that (1) investigates the relation and difference between different label prediction methods, C-LP and SM-LP, (2) demonstrates the failure reason of BCE loss in unsupervised learning is because BCE loss leads the noisy labels are weighted more for gradient update. Finally, comparisons with recent FUL methods demonstrate the superiority of our method.

V. Acknowledgment

This results was supported by “Regional Innovation Strategy (RIS)” through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (MOE) (2021RIS-003).

References

[1] K. Zhou, Y. Yang, A. Cavallaro, and T. Xiang. Omniscale feature learning for person re-identification. In 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pages 3701-3711, 2019.

[2] Q. Zhou, B. Zhong, X. Lan, G. Sun, Y. Zhang, B. Zhang, and R. Ji. Fine-grained spatial alignment model for person re-identification with focal triplet loss. IEEE Transactions on Image Processing, 29:7578-7589, 2020.

[3] H. Huang, W. Yang, J. Lin, G. Huang, J. Xu, G. Wang, X. Chen, and K. Huang. Improve person re-identification with part awareness learning. IEEE Transactions on Image Processing, 29:7468-7481, 2020.

[4] Y. Wu C. Yan Y. Lin, L. Xie and Q. Tian. Unsupervised person re-identification via softened similarity learning. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), page 3390-3399,2017.

[5] D. Wang and S. Zhang, “Unsupervised Person Re-Identification via Multi-Label Classification,” in *IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 10978-10987.

[6] Yang, Fengxiang and Zhong, Zhun and Luo, Zhiming and Cai, Yuanzheng and Li, Shaozi and Nicu, Sebe. Joint Noise-Tolerant Learning and Meta Camera Shift Adaptation for Unsupervised Person Re-Identification. In CVPR, 2021.

[7] Q. Tang, G. Cao and K. -H. Jo, ”Fully Unsupervised Person Re-Identification via Multiple Pseudo Labels Joint Training,” in IEEE Access, vol. 9, pp. 165120-165131, 2021.

[8] Yixiao Ge, Feng Zhu, Dapeng Chen, Rui Zhao, and Hongsheng Li. Self-paced contrastive learning with hybrid memory for domain adaptive object re-id. In NeurIPS, 2020.

[9] Shiyu Xuan and Shiliang Zhang. Intra-Inter Camera Similarity for Unsupervised Person Re-Identification. In CVPR, 2021.

[10] Q. Tang and K. -H. Jo, ”Unsupervised Person Re-Identification Via Nearest Neighbor Collaborative Training Strategy,” 2021 IEEE International Conference on Image Processing (ICIP), 2021, pp. 1139-1143.

[11] H.-X. Yu, A. Wu, and W.-S. Zheng, “Cross-view asymmetric metric learning for unsupervised person re-identification,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 994-1002.

[12] H.-X. Yu, A. Wu, and W.-S. Zheng, “Unsupervised person reidentification by deep asymmetric metric embedding,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 956-973, Apr. 2020.

[13] Y. Lin, X. Dong, L. Zheng, Y. Yan, and Y. Yang, “A Bottom-up Clustering Approach to Unsupervised Person Re-Identification,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, pp. 8738-8745.

[14] G. Ding, S. Khan, Q. Yin, and Z. Tang. “Dispersion based clustering for unsupervised person re-identification,” in *BMVC*, 2019.

[15] Z. Zhong, L. Zheng, Z. Luo, S. Li, and Y. Yang. “Invariance matters: Exemplar memory for domain adaptive person re-identification,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 598-607.

[16] Yixiao Ge, Dapeng Chen, and Hongsheng Li. Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification. In ICLR, 2020.

[17] H. Yu, W. Zheng, A. Wu, X. Guo, S. Gong, and J. Lai. Unsupervised person re-identification by soft multilabel learning. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 2143–2152, 2019.

[18] Zhun Zhong, Liang Zheng, Shaozi Li, and Yi Yang. Generalizing a person retrieval model hetero- and homogeneously. In Proceedings of the European Conference on Computer Vision (ECCV), September 2018.

[19] Unsupervised person re-identification: Clustering and finetuning. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 14(4):1-18, 2018.

[20] Pankaj K. Agarwal and Nabil H. Mustafa. k-means projective clustering. In PODS ’04: Proceedings of the twenty-third ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, pages 155-165, New York, NY, USA, 2004. ACM Press.

[21] Martin Ester, Hans-Peter Kriegel, Jorg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In KDD, pages 226-231, 1996.

[22] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In Proc. ICCV, 2015.

[23] Z. Zheng and L. Zheng and Y. Yang, “Unlabeled Samples Generated by GAN Improve the Person Re-identification Baseline in vitro,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 3774-3782.

[24] E. Ristani, F. Solera, R. Zou, R. Cucchiara and C. Tomasi, “Performance Measures and a Data Set for Multi-Target, Multi-Camera Tracking,” in *European Conference on Computer Vision workshop on Benchmarking Multi-Target Tracking*, 2016.

[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770-778, 2016.

[26] F. Schroff, D. Kalenichenko, and J. Philbin, “Facenet: A unified embedding for face recognition and clustering,” in CVPR, 2015.

[27] Min-Ling Zhang and Zhi-Hua Zhou. A review on multi-label learning algorithms. IEEE transactions on knowledge and data engineering, 26(8):1819-1837, 2013.

[28] Thibaut Durand, Nazanin Mehrasa, and Greg Mori. Learning a deep convnet for multi-label classification with partial labels. In CVPR, 2019.

[29] Feng, Lei, Senlin Shu, Zhuoyi Lin, F. Lv, L. Li and Bo An. “Can Cross Entropy Loss Be Robust to Label Noise?” IJCAI (2020).

[30] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. Journal of machine learning research, 9(Nov):2579-2605, 2008.