Graph-Based Self-Learning for Robust Person Re-identification (Supplementary Material)

Yuqiao Xian[1*], Jinrui Yang[2], Fufu Yu[2], Jun Zhang[2], and Xing Sun[2†]

[1] School of Computer Science and Engineering, Sun Yat-sen University, China
[2]Youtu Lab, Tencent
xianuyq3@mail2.sysu.edu.cn, {jinruiyang,fufuyu,bobbyjzhang,winfredsun}@tencent.com

Algorithm A1 Instance-Dependent Label Noise Generation

Input: Clean dataset $D = \{(x_i, y_i)\}_{i=1}^{n}$; Noise rate ρ_{noise}, network parameter $\mathbf{\Theta}$, training epoch T.
for $t = 1; t <= T; t++$ **do**
  Train network on clean dataset $D = \{(x_i, y_i)\}_{i=1}^{n}$ with cross-entropy loss and hard triplet loss.
end for
for $x_i \in D$ **do**
  Obtain identity predictions of x_i with the trained network: $\mathbf{p}_i = f(x_i; \mathbf{\Theta})$
  Record the secondary identity label of x_i by: $y_i' = \arg\max_{j \neq y_i} \mathbf{p}_i$
  Sample p from the uniform distribution $\mathcal{U}[0, 1]$.
  if $p \leq \rho_{noise}$ **then**
    Flip $\tilde{y}_i = y_i'$.
  else
    Keep $\tilde{y}_i = y_i$.
  end if
end for
Output: A corrupted dataset with IDN: $D = \{(x_i, \tilde{y}_i)\}_{i=1}^{n}$.

Algorithm A2 Tracklet-Level Label Noise Generation

Input: Clean dataset $D = \{(x_i, y_i)\}_{i=1}^{n}$; Noise rate ρ_{noise}, network parameter $\mathbf{\Theta}$, training epoch T.
for $t = 1; t <= T; t++$ **do**
  Train network on clean dataset $D = \{(x_i, y_i)\}_{i=1}^{n}$ with cross-entropy loss and hard triplet loss.
end for
for $x_i \in D$ **do**
  Obtain identity predictions of x_i with the trained network: $\mathbf{p}_i = f(x_i; \mathbf{\Theta})$
  Record the secondary identity label of x_i by: $y_i' = \arg\max_{j \neq y_i} \mathbf{p}_{ij}$
end for
for all tracklets $\mathbf{S} \in D$ **do**
  Obtain the tracklet-level secondary label by finding the most frequent secondary label of images in the tracklet: $y_{\mathbf{S}}' = MOST(\{y_i'\}_{x_i \in \mathbf{S}})$
  Sample p from the uniform distribution $\mathcal{U}[0, 1]$.
  for $x_i \in \mathbf{S}$ **do**
    if $p \leq \rho_{noise}$ **then**
      Flip $\tilde{y}_i = y_{\mathbf{S}}'$.
    else
      Keep $\tilde{y}_i = y_i$.
    end if
  end for
end for
Output: A corrupted dataset with TLN: $D = \{(x_i, \tilde{y}_i)\}_{i=1}^{n}$.

1. A. Generation of Label Noise.

In this section, we describe the generation of three types of label noise used in our experiments in detail.

Class-condintional noise (CCN): In our experiment of Sec X, we generate an uniformly distributed CCN on re-ID benchmarks. Given a noise ratio $\rho_{noise} \in (0, 1)$ and the number of identies C, the label of a sample x_i has a probability of $\frac{\rho_{noise}}{C-1}$ to be filped to a specific identity j ($j \neq y_i^*$, y_i^* is the ground thruth identity of x_i). This is similar to the symmetric noise [3] or random noise [4] in previous works. So we report the original results of PurifyNet [4] on random noise in Table 2.

Instance-dependent noise (IDN): There are different methods [5, 1, 2] to generate instance-dependent noise in existing research. Following [5], we firstly pretrain a model with a clean training set and then find the most similar identity in the training set based on the model prediction. Different from [2] which generates IDN based on the pixel-wise similarity of the images, our IDN generation algorithm uses a model to find semantically confusing identities, which better simulates the false human annotations. The process of IDN generation is shown **Algorithm A1**.

Tracklet-level noise (TLN). Similar to the generation of IDN, to find the most similar identity of each tracklet, we firstly pretrain a model with a clean training set. Second,

*This work was done when Yuqiao was an intern at Tencent Youtu Lab.
†Corresponding author.

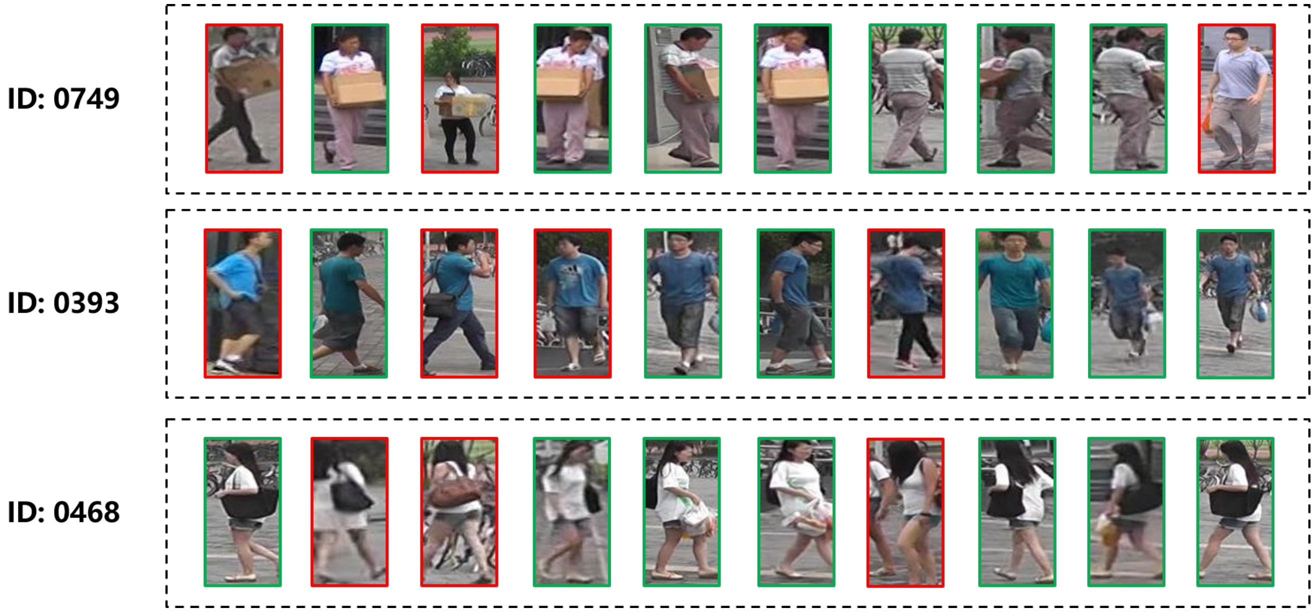

Figure B1. Examples of generated IDN label noise. Each row contains images with the same identity in the corrupted dataset with IDN. Images with red boxes are noisy samples that have different ground truth identity with the samples with green boxes (clean samples). Only a part of the clean samples are shown due to space limitation. We can observe that the noisy samples have many similar features with the clean samples, indicating that our IDN generation can produce confusing label noise using a pretrained model.

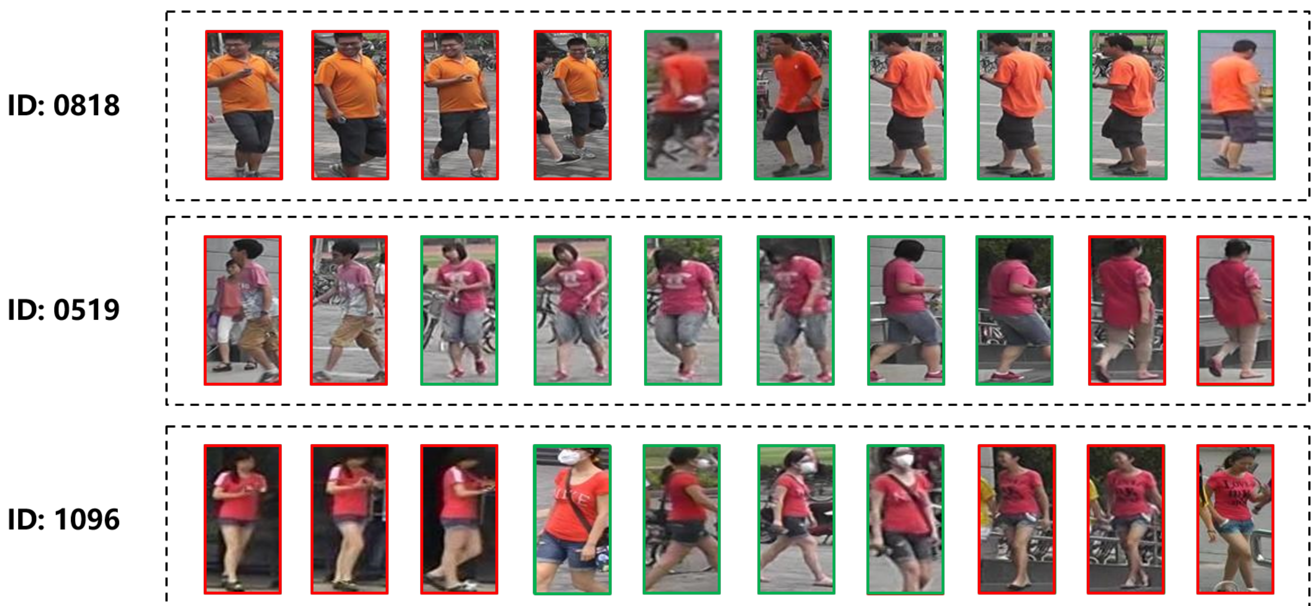

Figure B2. Examples of generated TLN label noise. Similar to IDN, the noise samples of an identity may come from multiple other identities. We can see that pedestrians in the noisy samples have highly similar clothing to the person in the clean samples. Different from samples of IDN, noise samples of a person are multiple images from one or more tracklets.

we record the secondary identity label of each image based on the model class prediction. Then, we determine the secondary label for each tracklet by finding the most frequent secondary label of all images in that tracklet. Finally, we sample a value from the uniform distribution from 0 to 1 and flip all the image-level labels in the tracklet to the same secondary labels if the value is less than the noise ratio. The detailed process of TLN generation is shown in **Algorithm A2**.

2. B. Visualization of Noisy Samples

To evaluate whether our IDN and TLN generation algorithms can simulate realistic label noise produced by false human annotations, we visualize some examples of the simulated dataset (Market-1501) with IDN in Fig. B1 and and TLN in Fig. B2. We can see that the noisy samples generated by our algorithm are highly similar to the clean samples. For example, we can see two different men with a white t-shirt in Fig B1, ID 0749, moving a box. They look very similar and we can only distinguish them by the color of their pants.

References

[1] Pengfei Chen, Junjie Ye, Guangyong Chen, Jingwei Zhao, and Pheng-Ann Heng. Beyond class-conditional assumption: A primary attempt to combat instance-dependent label noise. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 11442–11450, 2021.

[2] Hao Cheng, Zhaowei Zhu, Xingyu Li, Yifei Gong, Xing Sun, and Yang Liu. Learning with instance-dependent label noise: A sample sieve approach. *arXiv preprint arXiv:2010.02347*, 2020.

[3] Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor Tsang, and Masashi Sugiyama. Co-teaching: Robust training of deep neural networks with extremely noisy labels. *Advances in neural information processing systems*, 31, 2018.

[4] Mang Ye and Pong C Yuen. Purifynet: A robust person re-identification model with noisy labels. *IEEE Transactions on Information Forensics and Security*, 15:2655–2666, 2020.

[5] Tianyuan Yu, Da Li, Yongxin Yang, Timothy M Hospedales, and Tao Xiang. Robust person re-identification by modelling feature uncertainty. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 552–561, 2019.