Soft Pseudo 레이블의 유사도 기반 사람 탐색 연구

Person Search via similarity-based Soft Pseudo Labels

○최 제 환[1], Tang Qing[2], 조 강 현[3*]

[1] 울산대학교 전기전자컴퓨터공학과 (TEL: 052-259-1664, E-mail: jhchoi@islab.ulsan.ac.kr)
[2] 울산대학교 전기전자컴퓨터공학과 (TEL: 052-259-1664, E-mail: zucchini.tang@gmail.com)
[3] 울산대학교 전기전자컴퓨터공학과 (TEL: 052-259-2208, E-mail: acejo@ulsan.ac.kr)

Abstract This paper studies person search, which is an integrated technology of person detection and person re-identification. There are two different types of identities in the person search dataset, labeled identities, and unlabeled identities. Existing works only consider unlabeled identities as negatives and distinct identities for labeled identities by utilizing a Circular Queue (CQ), which ignores the relation among samples. In this paper, we proposed similarity-based soft pseudo labels by considering similarities between labeled and unlabeled identities. As a summary, the network is optimized under detection loss and two re-identification losses. (1) The OIM loss for labeled identities learning. (2) The Kullback–Leibler (KL) Divergence Loss using our proposed soft pseudo labels for unlabeled identities learning. The experimental results demonstrate the advantages of our proposed learning strategy.

Keywords Person search, Similarity computation, Unlabeled target learning

1. 서론

지능형 감시시스템은 최근 사회 치안 유지, 인공지능 CCTV의 발전과 상호작용, 컴퓨터비전 기술의 진보 등과 같은 이유로 최신 연구분야로 주목받고 있다. 사람 탐색은 사람 검출 (person detection)과 사람 재식별의 통합 기술로 볼 수 있다. 사람 검출이 우선적으로 진행된 후 재식별 기술을 통해 각 객체의 클래스를 결정한다. 학습을 할 때 검출된 사람은 각각 레이블 된 데이터와 레이블이 없는 데이터로 분류된다. OIM[1]에 따르면 레이블 된 데이터는 LUT(Lookup Table)을 사용하여 학습한다. 또한, 레이블이 없는 데이터를 개별 클래스로 지정하여 학습할 수 있는 CQ(Circular Queue)를 제안하였다. 해당 아이디어는 레이블 된 데이터와 레이블이 없는 데이터를 비교하여 레이블이 없는 데이터를 soft pseudo 기법으로 학습하기 위해 제안되었다.

2. 본론

2.1 Baseline network

본 논문에서 사용한 네트워크의 baseline은 SeqNet (Sequential End-to-end Network)[2]의 baseline을 적용했다. 이미지가 입력되면, RPN (Region Proposal Network)은 객체가 존재할 확률이 높은 후보를 선정하고 RoIAlign를 사용하여 7x7x2048 모양의 RoI로 정렬한다. Global pooling은 RoI를 한번 더 변형시키고 검출 손실값과 Re-ID를 계산할 때 사용된다. NAE (Norm-Aware Embedding Method)[3]를 사용해서 Re-ID 손실값이 계산된다.

※ 이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임.(No.2020R1A2C200897212)

2.2 Lookup table for labeled identities

학습데이터에 서로 다른 L명의 사람이 있다고 가정하면, 그것을 레이블된 인스턴스라고 하며 1부터 L까지의 class-id를 부여한다. 네트워크의 목표는 사람을 구별하는 것이며, 원초적으로 다른 사람간의 불일치성을 극대화하는 것이다. LUT는 모든 사람의 특징을 기억하기위해 사용된다. LUT는 $V \in R^{D \times L}$로 표시되며, D는 특징의 차원 개수이고 L은 클래스 개수이다. 역전파 알고리즘이 적용될 때, target class id가 t라면 LUT는 $v_t \leftarrow \gamma v_t + (1-\gamma)x, \gamma \in [0,1]$에 의해 t번째 열이 업데이트되며 v_t는 L2-norm에 의해 정규화된다.

2.3 Soft pseudo labels for unlabeled identities

One-hot label은 객체 간 유사성의 복잡함을 포착하기 어려우므로, 분류기 Cls(·)를 학습하기 위해 U와 V의 유사도를 soft label처럼 계산한다. 즉, 레이블이 없는 객체의 pseudo label은 레이블 된 객체와 레이블이 없는 객체의 유사도에 의해 생성된다. 유사도는 식 (1)과 같이 계산한다.

$$S = U^{\top} \times V \tag{1}$$

U는 미니배치의 레이블이 없는 특징이며, 크기는 $n_b \times D$(n_b는 미니배치의 레이블이 없는 특징의 개수)이다. 그러므로 S의 크기는 $n_b \times L$이고, 레이블이 없

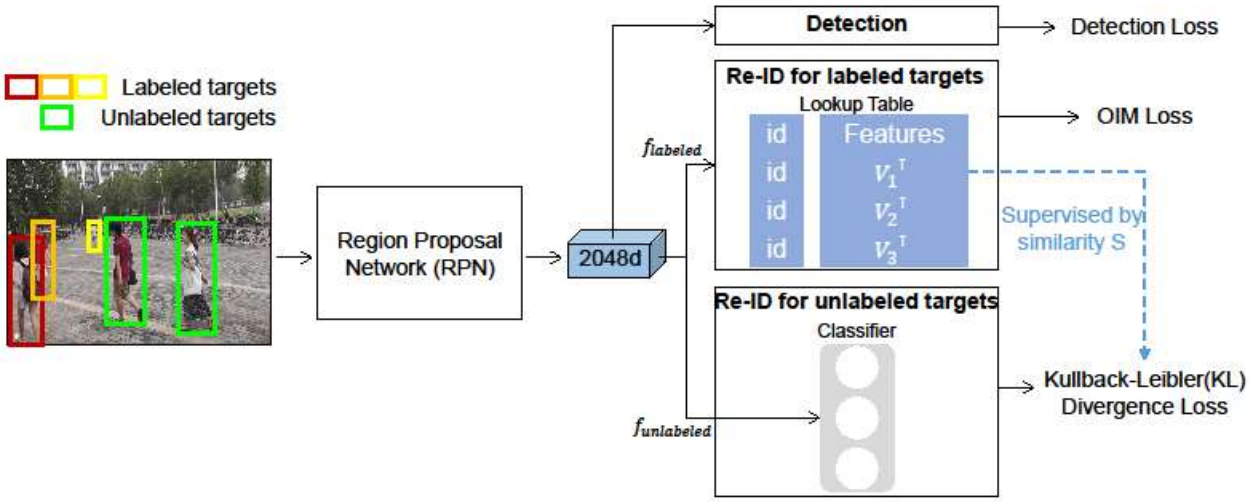

그림 1. 제안 네트워크 구조

는 단일객체의 크기는 $1\times L$이다. 파라미터수와 분류기 Cls(·)는 식 (2)와 같이 KL (Kullback-Leibler) 기울기 손실 알고리즘을 사용하여 최적화된다. Cls(·)의 크기는 LUT의 크기는 L로 동일하며, 2048d와 Cls(·) 사이에는 FC(Fully Connected) layer로 연결된다.

$$L_{Cls} = D_{KL}(softmax(Cls(f_{unlabeled})/\tau, softmax(S/\tau))) \quad (2)$$

L_{Cls}는 레이블이 없는 특징 손실값이고, $f_{unlabeled}$는 레이블이 없는 특징이다. τ는 higher temperature으로 확률적 분배에 따른 hyper parameter이다.

3. 실험

3.1 Dataset and Evaluation Metrics

PRW[4]는 Market-1501 데이터를 레이블링한 데이터셋으로 총 11,816장의 이미지와 4,310명의 Bounding Box, 484개의 레이블이 포함되어있다.

평가지표는 mAP (Mean Average Precision)와 CMC (Cumulative Matching Characteristic) 커브를 사용했다. CMC커브의 가장 높은 값을 Top-1으로 표기했다.

3.2 Implementation Setup

실험은 총 10 epoch을 진행했다. 네트워크의 학습률 (Learning rate)는 초기 0.003으로 설정 후 7 epoch 시작과 함께 0.0003이 되도록 설정했다. Batch size는 5로 설정하였으며. 입력이미지는 900x1500 크기로 조정하였다.

3.3 Experimental Results

표 1. τ에 따른 mAP와 Top-1 비교

τ	Re-ID Acc. (%)	
	mAP	Top-1
0.03	42.57	80.21
0.05	46.49	83.81
0.1	48.07	83.91
0.3	**49.49**	**84.25**
0.5	48.43	84.1

표 2. CQ 기법과 제안방법의 정확도 비교

Method	Re-ID Acc. (%)	
	mAP	Top-1
CQ[3]	46.3	83.37
Proposed	**49.49**	**84.25**

실험결과는 표 1과 표 2에서 확인할 수 있듯이, CQ기법으로 손실값을 계산하는 것보다 레이블의 유사도 기반 soft pseudo 기법이 더 높은 정확도를 도출했다. 특히, mAP에서 확연한 정확도의 증가를 나타냈다. 그리고 τ를 0.3로 설정했다. 그 결과, 가장 좋은 결과를 보였다.

4. 결론

본 논문은 지능 감시시스템의 기반이 되는 사람 탐색 기법에 대한 연구를 진행하였으며, 특히 CQ 기법 대신 soft pseudo 레이블의 유사도 계산법을 사용한 방법을 적용하였다. 그 결과, 기존 방법보다 3.19% 높은 정확도를 보였으며, 0.88% 높은 Top-1값을 도출했다.

참고문헌

[1] T. Xiao, S. Li, B. Wang, L. Lin, and X. Wang, “Joint detection and identification feature learning for person search,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3376 - 3385, 2017.

[2] Z. Li and D. Miao, “Sequential end-to-end network for efficient person search,” in AAAI, 2021

[3] D. Chen, S. Zhang, J. Yang, and B. Schiele, “Norm-aware embedding for efficient person search,” 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12 612 - 12 621, 2020.

[4] Woo, S., Park, J., Lee, J., & Kweon, I. (2018). CBAM: Convolutional Block Attention Module. ECCV.