Unsupervised Person Re-identification by Soft Multilabel Learning

In proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition(CVPR)

Yunseon Byun
June 01, 2020

Contents

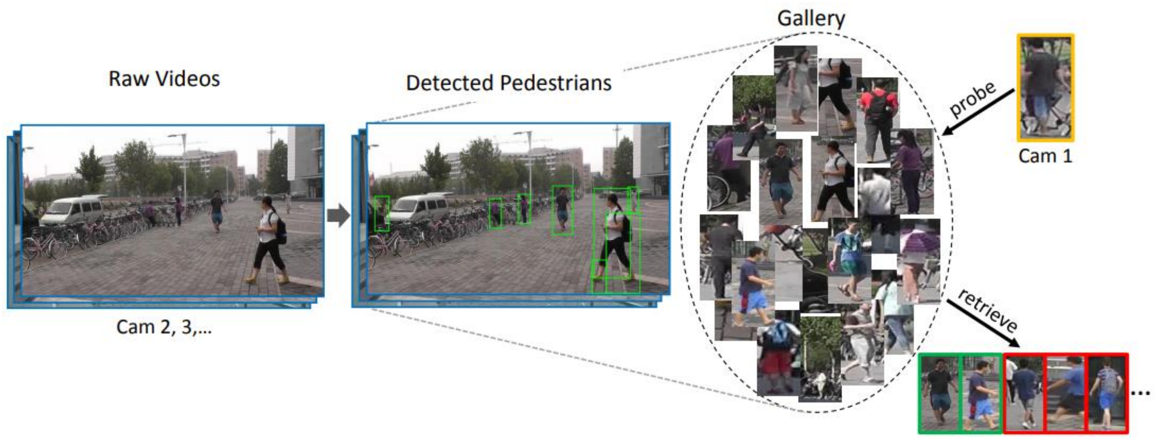
- 출처

Zheng, L., Yang, Y., & Hauptmann, A. G. (2016). Person re-identification: Past, present and future. *arXiv preprint arXiv:1610.02984*.

1. Introduction

What is person re-identification(Re-ID)?

Given an image/video of a person taken from one camera, re-identification is the process of identifying the person from images/videos taken from a different camera with non-overlapping fields of views. Re-identification is indispensable in establishing consistent labeling across multiple cameras or even within the same camera to re-establish disconnected or lost tracks.



An end-to-end person re-ID system

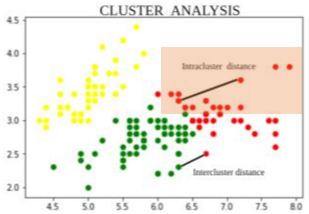
- ✓ Non-overlapped image로부터 Gallery 내의 일치하는 사람을 찾는 문제
- ✓ 몸 전체의 특징(의복, 체형, 보행 패턴 등)을 사용해 일치 여부 판정
- ✓ 동일한 카메라에서 찍은 동일한 사람의 다른 이미지, 다른 카메라에서 찍은 이미지에 대해서도 식별함

• 출처
http://rose1.ntu.edu.sg/PersonReId/,
https://medium.com/@niruhan/a-practical-guide-to-person-re-identification-using-alignedreid-7683222da644

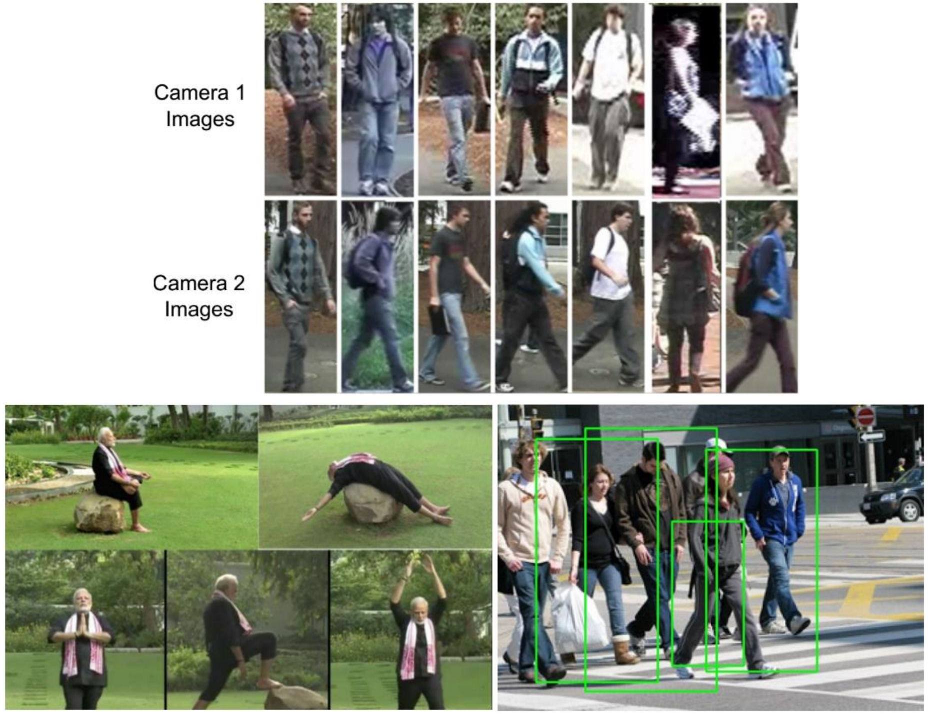
1. Introduction

Challenges

Intra-class variation control



a. Illumination changes : 일조도, 음영, 실내 및 실외 환경의 색상 변경, 다수의 camera view에서 동일한 물체의 색상 변경
b. Resolution : 저해상도 camera에서 사람을 구별하기 어려움
c. Occlusion : 혼잡한 화면에서 다른 사람이 개인의 일부 혹은 전체를 가릴 수 있으므로 전신 이미지를 얻기 어려움
d. Uniform clothing : 교복, 작업복 등의 복장을 착용한 사람들을 구분하기 어려움
e. Scalability : 다수의 camera에서 동일한 사람을 식별하기 어려움
f. Lack of annotated data : Labeled data 부족



Examples of person Re-ID dataset

• 출처 https://bismex.github.io/2019/07/04/MAR.html,

1. Introduction

Link to the multilabel reference learning(MAR) problem

Definition

✓ Unlabeled person에 대하여 auxiliary dataset(reference)의 labeled person과 비교함으로써 soft multilabel을 부여함

Related works

a. 시각적 특징의 유사도만을 사용
b. Auxiliary source domain으로부터 discriminative information을 뽑아 target domain으로 옮겨 적용하는 방식
→ Unlabeled source와 unlabeled target 사이의 domain shift 발생

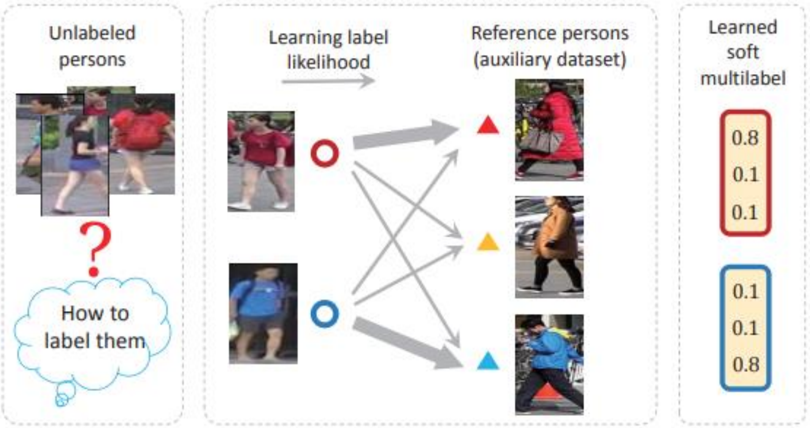


Illustration of soft multilabel concept

2. Methodology

Contributions

a. The unsupervised Re-ID problem by a novel soft multilabel reference learning method

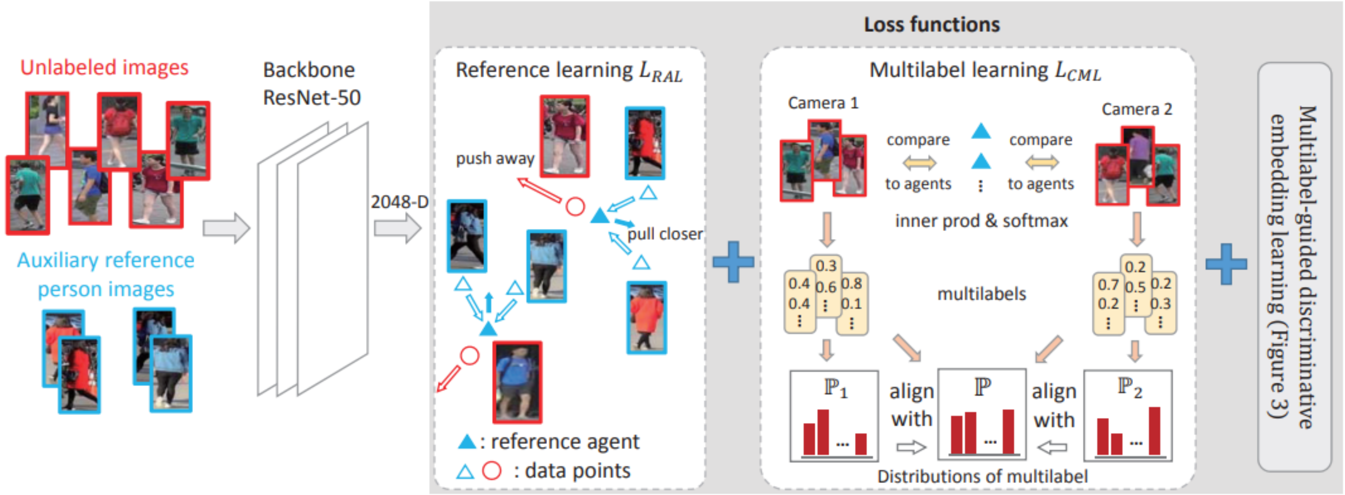
⇒ Unlabeled data에 대해서는 auxiliary source data와 reference data를 비교함으로써 potential label information을 뽑음

b. Deep soft multilabel reference learning (MAR)을 통해 하나의 모델로 3가지를 수행함

- *Soft multilabel-guided hard negative mining (MDL)*
  ⇒ Batch 마다 unlabeled data에 대한 positive, negative set을 확보한 후, positive set은 서로 가까워지고 negative set은 서로 멀어지도록 embedding feature를 학습
- *Cross-view consistent soft multilabel learning (CML)*
  ⇒ Camera view에 무관한 soft multilabel 학습을 위해 soft multilabel 분포를 일치시키는 방식
- *Reference agent learning (RAL)*
  ⇒ 효율적인 학습을 위해 joint feature embedding

2. Methodology

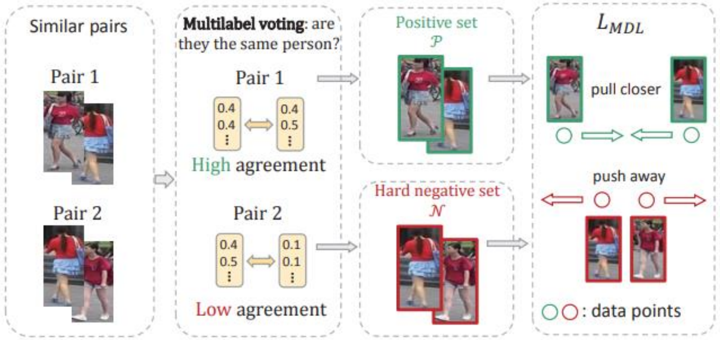
The overview of methodology : 3가지 loss term (MDL, CML, RAL) 정의



The overview of MAR model

a. 각 target unlabeled person image $f(x)$와 auxiliary reference persons a_i 비교를 통한 soft multilabel 학습
b. Similar pair의 positive / hard negative 여부 판정
c. Soft multilabel learning and reference learning

2. Methodology



1) Soft multilabel-guided hard negative mining

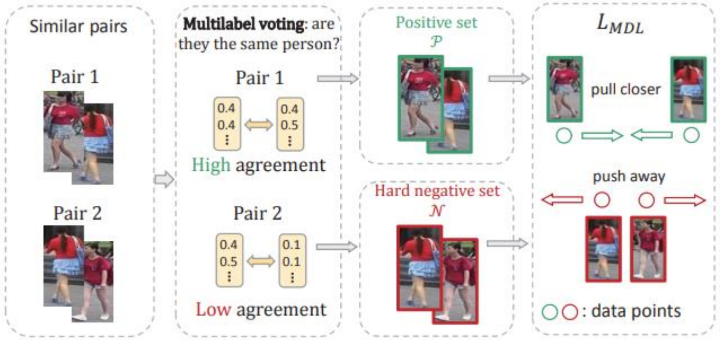
- Target 이미지를 표현하는 soft multilabel을 학습하고 이를 통해 discriminative feature 학습
- Unlabeled target domain에서 positive pair는 서로 가까워지도록, negative pair는 멀어지도록 embedding feature를 학습
  - ✓ Positive pair: similar feature & similar comparative characteristics (same soft multilabel)
  - ✓ Negative pair: similar feature & not similar comparative characteristics (different soft multilabel)

Inner product : $f(x_i)^T f(x_j)$

Soft multilabel function : $y^{(k)} = l(f(x), \{a_i\}_{i=1}^{N_p})^{(k)} = \dfrac{\exp(a_k^T f(x))}{\sum_i \exp(a_k^T f(x))}$

Soft multilabel agreement : $A(y_i, y_j) = \sum_k \min({y_i}^{(k)}, {y_j}^{(k)}) = 1 - \dfrac{\|y_i - y_j\|_1}{2}$

2. Methodology



1) Soft multilabel-guided hard negative mining

- Target 이미지를 표현하는 soft multilabel을 학습하고 이를 통해 discriminative feature 학습
- Unlabeled target domain에서 positive pair는 서로 가까워지도록, negative pair는 멀어지도록 embedding feature를 학습
  - ✓ Positive pair: similar feature & similar comparative characteristics (same soft multilabel)
  - ✓ Negative pair: similar feature & not similar comparative characteristics (different soft multilabel)

Assign (i, j) to the set :

$$P = \{(i,j) | f(x_i)^T f(x_j) \geq S, A(y_i, y_j) \geq T\}$$

$$N = \{(k,l) | f(x_k)^T f(x_l) \geq S, A(y_k, y_l) < T\}$$

The similarity threshold according to the feature similarity

- cosine similarity (inner product)

The threshold for the soft multilabel agreement

Loss for soft multilabel guided discriminative embedding learning :

$$\boldsymbol{L_{MDL}} = -log \frac{\bar{P}}{\bar{P} + \bar{N}}$$

$$\bar{P} = \frac{1}{|P|} \Sigma_P \exp(-\|f(z_i) - f(z_j)\|_2^2), \quad \bar{N} = \frac{1}{|N|} \Sigma_N \exp(-\|f(z_i) - f(z_j)\|_2^2)$$

• 출처
https://bismex.github.io/2019/07/04/MAR.html,

2. Methodology

2) Cross-view consistent soft multilabel learning

- Camera view에 robust한 feature 학습을 위해 soft multilabel 분포를 일치시킴 (cross-view consistent)
- Soft multilabel이 camera view와 독립이기 위한 loss 정의
  - ✓ 'Soft multilabel의 분포는 camera view에 대해 독립적이다'라는 가정이 있음
  - ✓ Camera view에는 무관하고, target domain에서 사람의 외형 분포에만 dependent 해야 함
  - ✓ 분포 비교를 위해 KL-divergence, Wasserstein distance 등을 활용 가능함

Loss for cross-view consistent soft multilabel learning :

$$\boldsymbol{L_{CML}} = \sum_v d(P_v(y), p(y))^2$$

(Using 2-Wasserstein distance to simplify)

$$\boldsymbol{L_{CML}} = \sum_v \|\mu_v - \underbrace{\mu}_{\text{Mean vector}}\|_2^2 + \|\sigma_v - \underbrace{\sigma}_{\text{Standard deviation vector}}\|_2^2$$

2. Methodology

3) Reference agent learning

- Reference space와 target space에 대하여 동시에 discriminative한 feature를 학습하도록 함
- Agent learning(AL)과 reference agent joint embedding learning(RJ)을 고려한 loss 설정
  - ✓ Agent: reference data의 각각의 identity를 표현하는 feature

$$\boldsymbol{L_{RAL}} = L_{AL} + \beta L_{RJ}$$

Agent learning (AL)

: Reference space에 feature embedding 하였을 때 서로 다른 identity에 대해 discriminative한 학습을 진행함으로써
multilabel function $l(\quad)$에 validity 부여

- Soft label 값이 1이 되도록 학습
- Reference에서 같은 사람의 agreements 값은 1이 되고 다른 사람의 agreement 값은 0이 되도록 학습

$$L_{AL} = \sum_k -\log l(f(z_k), \{a_i\})^{(w_k)} = \sum_k -log \frac{\exp(a_{w_k}^T f(z_k))}{\sum_j \exp(a_j^T f(z_k))}$$

2. Methodology

3) Reference agent learning

- Reference space와 target space에 대하여 동시에 discriminative한 feature를 학습하도록 함
- Agent learning(AL)과 reference agent joint embedding learning(RJ)을 고려한 loss 설정
  - ✓ Agent: reference data의 각각의 identity를 표현하는 feature

Reference agent joint embedding learning (RJ)

: Unlabeled target data에 대해 discriminative한 feature를 학습하기 위함

$$L_{RJ} = \sum_i \sum_{j \in M_i} \sum_{k \in W_i} \left[m - \|a_i - f(x_j)\|_2^2\right]_+ + \|a_i - f(z_k)\|_2^2$$

- Reference agent마다 시각적으로 비슷한 unlabeled person을 찾은 후, discriminative하도록 학습
- 서로 비슷한 cross-domain pair가 멀어지도록 학습

- Reference agent와 해당 label에 대응되는 reference image를 가깝도록 학습

$$M_i = \{j \mid \|a_i - f(x_j)\|_2^2 < m\}\ ,\ W_i = \{k \mid w_k = i\}$$

2. Methodology

4) Model training and testing

- Compute the cosine feature similarity of each probe(query) – gallery pair
- Obtain the ranking list of the probe image against the gallery images
- Train the model by stochastic gradient descent (SGD)
- Minimize the below loss function

The loss objective of deep soft multilabel reference learning :

$$\boldsymbol{L_{MAR}} = L_{MDL} + \lambda_1 L_{CML} + \lambda_2 L_{RAL}$$

$$\boldsymbol{L_{MDL}} = -log \frac{\bar{P}}{\bar{P} + \bar{N}}$$

$$\boldsymbol{L_{CML}} = \sum_v d(P_v(y), p(y))^2$$

$$\boldsymbol{L_{RAL}} = L_{AL} + \beta L_{RJ}$$

3. Experiments

Dataset

a. Market-1501
   - 32,668 person images of 1,501 identities
   - 6 camera views

b. DukeMTMC-reID
   - 36,411 person images of 1,404 identities
   - 8 camera views

c. MSMT17 (Auxiliary)
   - 126,441 person images of 4,101 identities

Dataset examples

Parameter setting

- Batch size = 368
- λ_1 = 0.0002
- λ_2 = 50
- The mining ratio = 5%
- β = 0.2

3. Experiments

Results

Ablation study

Methods	Market-1501			
	rank-1	rank-5	rank-10	mAP
Pretrained (source only)	46.2	64.4	71.3	24.6
Baseline (feature-guided)	44.4	62.5	69.8	21.5
MAR w/o L_{CML}	60.0	75.9	81.9	34.6
MAR w/o L_{CML}&L_{RAL}	53.9	71.5	77.7	28.2
MAR w/o L_{RAL}	59.2	76.4	82.3	30.8
MAR	**67.7**	**81.9**	**87.3**	**40.0**

Methods	DukeMTMC-reID			
	rank-1	rank-5	rank-10	mAP
Pretrained (source only)	43.1	59.2	65.7	28.8
Baseline (feature-guided)	50.0	66.4	71.7	31.7
MAR w/o L_{CML}	63.2	77.2	82.5	44.9
MAR w/o L_{CML}&L_{RAL}	60.1	73.0	78.4	40.4
MAR w/o L_{RAL}	57.9	72.6	77.8	37.1
MAR	**67.1**	**79.8**	**84.2**	**48.0**

Comparison to the unsupervised results

For the Market-1501 dataset,

Methods	Reference	Market-1501		
		rank-1	rank-5	mAP
LOMO [20]	CVPR'15	27.2	41.6	8.0
BoW [59]	ICCV'15	35.8	52.4	14.8
DIC [16]	BMVC'15	50.2	68.8	22.7
ISR [21]	TPAMI'15	40.3	62.2	14.3
UDML [29]	CVPR'16	34.5	52.6	12.4
CAMEL [52]	ICCV'17	54.5	73.1	26.3
PUL [8]	ToMM'18	45.5	60.7	20.5
TJ-AIDL [48]	CVPR'18	58.2	74.8	26.5
PTGAN [50]	CVPR'18	38.6	57.3	15.7
SPGAN [7]	CVPR'18	51.5	70.1	27.1
HHL [62]	ECCV'18	**62.2**	**78.8**	31.4
DECAMEL [53]	TPAMI'19	60.2	76.0	**32.4**
MAR	This work	**67.7**	**81.9**	**40.0**

For the DukeMTMC-reID dataset,

Methods	Reference	DukeMTMC-reID		
		rank-1	rank-5	mAP
LOMO [20]	CVPR'15	12.3	21.3	4.8
BoW [59]	ICCV'15	17.1	28.8	8.3
UDML [29]	CVPR'16	18.5	31.4	7.3
CAMEL [52]	ICCV'17	40.3	57.6	19.8
PUL [8]	ToMM'18	30.0	43.4	16.4
TJ-AIDL [48]	CVPR'18	44.3	59.6	23.0
PTGAN [50]	CVPR'18	27.4	43.6	13.5
SPGAN [7]	CVPR'18	41.1	56.6	22.3
HHL [62]	ECCV'18	**46.9**	**61.0**	**27.2**
MAR	This work	**67.1**	**79.8**	**48.0**

4. Conclusion

Summary

- Unlabeled target Re-ID data에 대하여 potential label information을 찾기 위해 auxiliary source Re-ID data를 활용하였음
- Soft multilabel-guided hard negative mining, the cross-view consistent soft multilabel learning, the reference agent learning을 동시에 진행하는 MAR model 제안
- Domain adaption 분야의 문제 해결에 의의가 있음

Visual results of the soft multilabel-guided hard negative mining

Thank you for your listening ☺

Yunseon Byun
June 01, 2020