

클러스터 일관성을 기반으로 한 비지도 도메인 적응 사람 재인식

오상엽, 조남익

서울대학교,

syup5@snu.ac.kr, nicho@snu.ac.kr

Unsupervised Domain Adaptive Re-identification based on Cluster Consistency

Oh, Sang-Yup Cho, Nam-Ik

Seoul National University

요 약

사람 재인식을 수행하기 위해서 많은 연구들이 진행되어 좋은 결과들을 보였다. 그러나 이 결과들은 라벨이 있는 도메인에서의 지도 학습으로 얻은 결과들이었다. 라벨이 없는 도메인에서의 사람 재인식의 성능은 아직 많이 부족한 상태이다. 사람 재인식을 수행하고자 하는 목표 도메인에 반해 주어진 소스 도메인에서는 라벨이 풍부하다. 지금까지의 논문에서는 소스 도메인에서의 사람 이미지를 목표 도메인의 이미지처럼 만들어서 소스 도메인에서 높은 성능을 보이는 사람 재인식기를 목표 도메인에서도 잘 동작하도록 학습하는 방법들이 주를 이루었다. 하지만 이 방법에서는 소스 도메인의 사람 이미지를 목표 도메인의 이미지와 비슷하게 만들기만하고 사람의 신원에 대한 일관성을 유지시키지는 못하였다. 본 논문에서는 비지도 도메인 적응 사람 재인식을 수행하기 위해 클러스터 일관성(cluster consistency)을 유지하는 기법을 제안한다. 제안한 방법은 사람의 신원에 대한 일관성을 유지시켜서 사람 재인식의 성능을 높인다

1. 서론

정보의 양이 많아지면서, 사람들은 정보들을 효과적으로 처리하기 위한 기술들이 필요해졌다. 어떤 정보를 사용하고자 할 때 필요한 정보를 찾아내기 위한 검색 기술은 필수적이다. 정보를 활용하는 기술도 중요하지만 그것보다도 정보를 검색하는 기술은 근본적으로 더 중요하다. 처음에는 글자를 기반으로 한 정보가 많았기 때문에 글자를 단서로 검색하는 기술들이 많이 개발되었다. 정보의 형태는 글자를 넘어서 더 다양한 형태로 발전하였다. 마찬가지로 이러한 정보를 활용하기 위해서는 정보 검색 기술의 발전이 필수 불가결하다.

사람 재인식(person re-identification)은 사람의 신원을 검색하기 위한 기술이다. 특정 사람의 영상(query image)이 주어졌을 때 다른 영상들 중에서 해당 영상의 사람과 동일한

신원을 가진 사람이 촬영된 영상을 찾아내는 것이 목표이다. 많은 CCTV 영상들을 데이터로 심층 신경망(Deep neural network)을 학습하면 사람 재인식에 좋은 성능을 보인다.

이렇게 좋은 성능을 보이는 신경망들은 보통 지도 학습(supervised learn)을 통해서 학습한다. 지도 학습이란 영상속 사람이 어떤 사람인지 신원 정보가 담겨있는 라벨을 이용하여 학습하는 학습방법을 의미한다. 지도 학습으로 신경망을 학습하는 데에는 두 가지 단점이 있다. 첫 번째로, 이러한 지도 학습 학습을 위해서는 라벨과 영상이 많이 필요하다. 필요한 라벨들을 취득하기 위해서는 많은 시간적, 경제적 비용이 발생한다. 두 번째, 이렇게 지도 학습으로 학습한 신경망도 학습한 영상들과 다른 환경에서 촬영한 CCTV 영상들로 사람을 검색하게 되면 제대로 된 성능을 발휘 하지 못한다.

이러한 한계점을 극복하여 라벨이 없는 도메인에서도 사람을

검색할 수 있는 심층 신경망을 학습시키는 방법이 비지도 도메인 적응(Unsupervised Domain Adaptive)학습방법이다. 이러한 방향으로 학습하면 많은 양의 이미지와 라벨이 없어도 사람의 신원을 검색할 수 있게 되고, 낯선 환경에서 촬영된 CCTV 영상에서도 사람을 검색할 수 있게 된다.

비지도 도메인 적응 학습방법에서 도메인은 소스 도메인과 목표 도메인으로 나뉜다. 소스 도메인은 영상이 많고, 라벨도 많다. 그러나 목표 도메인에서는 라벨이 전혀 없고 영상만 주어진다. 비지도 도메인 적응 학습은 목표 도메인과 소스 도메인이 공유하고 있는 특징 공간을 찾아낼 수 있다는 이론을 기반으로 한다 [1]. 도메인과 관련이 없는(domain-invariant) 특징을 추출하기 위해 먼저 심층 신경망은 소스 도메인과 목표 도메인의 영상의 특징을 각각 추출한다. 그리고 특정 신경망을 추가하여 소스 도메인으로부터 온 특징과 목표 도메인으로부터 온 특징을 서로 구분할 수 없도록 학습한다. 이를 위해 본 논문에서는 적대적 학습(adversarial learning)기법을 활용한다. 특징 추출에는 백본(backbone)[2]을 사용한다. 적대적 학습[3]에서는 판별자(discriminator)와 백본을 각각 반대로 학습한다. 판별자는 소스로부터 온 특징과 목표로부터 온 특징을 구분하도록 학습하고, 백본은 소스로부터 온 특징과 목표로부터 온 특징을 서로 구분할 수 없도록 학습한다. 적대적으로 학습한 백본을 통해 소스 영상과 목표 영상은 도메인에 따른 차이가 없는 특징들이 추출된다. 이렇게 추출된 특징들을 신원 검색에 사용하면 소스 도메인에서 신원을 잘 검색하는 심층 신경망은 목표 도메인에서도 좋은 성능을 보일 것 이라는게 일반적인 접근법이다.

하지만 이 방법에는 한계가 있다. 소스 도메인의 특징을 목표 도메인의 특징과 비슷하게 만드는 과정에서 소스 도메인의 특징이 지니고 있는 신원에 대한 정보가 사라져서 성능이 저하된다는 것이다. 적대적 학습은 추출된 특징의 신원에 대한 정보 대신, 추출된 특징이 어느 도메인으로부터 왔는지에 대해서만 집중하여 학습하기 때문에 학습을 하는 과정에서 신원에 대한 정보가 사라지는 것은 당연하다.

본 논문에서는 클러스터 일관성(cluster consistency)을 유지하는 기법을 제시하여 신원 정보를 유지하도록 한다. 소스 영상을 추출한 특징을 밀도 기반으로 클러스터링을 하였을 때 그 결과가 라벨과 일치하지 않는다면 특징점들의 밀집 공간이 신원정보를 제대로 포함하고 있지 않는다는 것을 의미한다. 밀도 기반 클러스터링의 결과가 원래 라벨과 다른 데이터들을 이상점(outlier)으로 취급한다. 이 이상점들을 원래 라벨을 갖는 특징들이 분포하고 있는 중심점에 가깝도록 목적 함수(loss function)를 설계하고 학습하여 이상점들의 특징점이 원래 라벨과 같은 공간에 위치할 수 있도록 한다.

2. 학습 방법

2.1 신원 학습

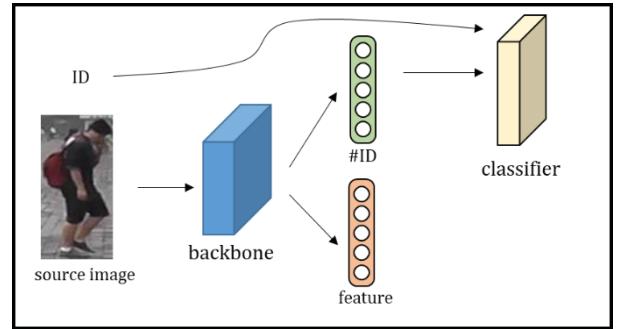


그림 1. 신원 학습 (Identification learning)

심층 신경망은 그림 1 에서 보이듯이 백본과 구별자(classifier)를 통해서 신원을 학습한다. 백본에서는 주어진 이미지의 특징들을 추출하고, 추출한 특징들을 완전연결 레이어(fully connected layer)에 통과시켜서 이미지의 신원(ID)를 얻는다. 이후 학습을 위한 목적함수는 다음과 같고,

$$L_{id} = E[-\log(p(y_s|x_s))] \quad (1)$$

여기서 x_s 는 소스 이미지를 의미하고 y_s 는 소스 도메인 이미지의 라벨을 의미한다. P 는 백본을 통해 얻은 신원 결과 확률값을 의미한다. 식 (1) 을 통해 구별자(classifier)는 소스 이미지의 라벨을 학습하여 주어진 이미지의 신원을 구별할 수 있게 된다.

2.2 적대 학습

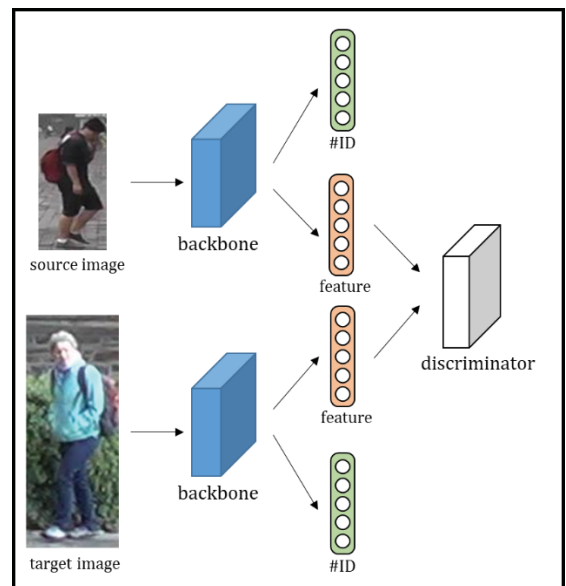


그림 2. 적대 학습 (adversarial learning)

적대 학습은 그림 2 에서처럼 판별자가 필요하다. 백본을

통해 추출된 소스 특징과 목표 특징을 각각 판별자에 넣고 소스에서 왔는지, 목표로부터 왔는지를 판별한다. 백본은 이 판별자가 제대로 특징을 판별할 수 없도록 반대로 학습한다. 식(2)는 이를 나타내며, 이 백본에서 특징 추출에 해당하는 f 는 식을 최소화 하도록 학습하고 판별자 D 는 이 식을 최대화 시키도록 학습한다.

$$\min_f \max_D E[\log(D(f(x_t))) + \log(1 - D(f(x_s)))] \quad (2)$$

2.3 클러스터 일관성

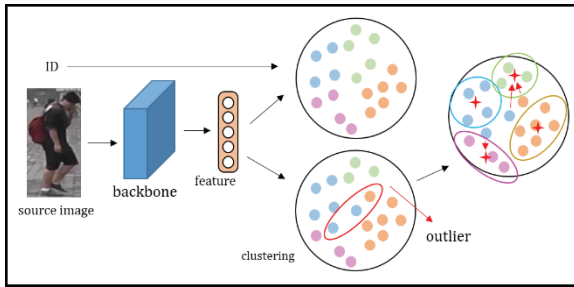


그림 3. 클러스터 일관성 (cluster consistency)

그림 3 에서 위의 원 속의 점들은 신원 라벨(ID) 에 의해 클러스터 된 특징들을 의미하고 아래의 원 속의 점들은 밀도 기반으로 클러스터링한 특징들을 의미한다.

그림 3 의 빨간 원처럼 일치하지 않는 특징점을 이상점(outlier)으로 설정한다. 그림 3 의 오른쪽 원에서는 클러스터의 중심점(prototype)을 빨간색 별표로 나타낸다. 클러스터링을 통해 찾은 이상점과 클러스터의 중심점을 다음과 같은 L1 으로 계산하여 최소화하도록 백본을 학습한다

$$\min_f E_{x \sim \text{outlier}(x)} |f(x) - \text{prototype}(x)|_1 \quad (3)$$

3. 결과

표 1. 실험에 사용한 데이터

Datasets	Training		Testing	
	#IDs	#Images	#IDs	#Images
Market	751	12,936	750	19,732
DukeMTMC	702	16,522	702	19,889

실험에 사용된 데이터셋은 Market-1501[4] 과 DukeMTMC-reID[5]이다. 각각의 데이터셋은 표 1 과 같이 구성되어 있다. 실험은 Market-1501 데이터를 소스 도메인으로 하고

DukeMTMC-reID 데이터를 목표 도메인으로 한 실험과 그 반대의 실험을 진행하여 각각 결과를 얻었다.

표 2. 실험 결과

Method	Duke → Market				Market → Duke			
	rank1	rank5	rank10	mAP	rank1	rank5	rank10	mAP
A	43.3	61.5	68.2	19.4	25.7	45.2	54.5	12.7
B	50.8	68.2	75.1	25.0	35.1	51.8	57.1	20.6
C	54.2	71.9	77.9	26.6	52.8	70.6	77.6	25.5

표 3. 실험 환경 설정

Method	Setting
A	소스 도메인에서 신원만 학습
B	소스 도메인에서 신원만 학습 + 적대 학습
C	소스 도메인에서 신원만 학습 + 적대 학습 + 클러스터 일관성 유지

실험은 세 가지 상황에서 진행되어 각각 결과를 비교한다. 첫 번째 A 는 소스 도메인에서 신원 학습만 진행하고 목표 도메인에서 바로 사람 재인식을 수행한 결과이다. B 는 클러스터 일관성을 유지시키지 않고 신원 학습과 적대 학습으로만 학습한 결과이다. 마지막으로 C 는 신원 학습과 적대 학습을 병행하며 본 논문에선 제안하는 클러스터 일관성을 유지시켜서 얻은 결과이다.

결과를 통해 클러스터 일관성을 유지시켰을 때 실험 결과가 제일 좋다는 것을 확인할 수 있다. 또한 Market1501 → Duke 실험 결과의 변화가 더 크다는 것도 확인할 수 있다.

4. 감사의 글

이 논문은 2020 년도 BK21 FOUR 정보기술 미래인재 교육연구단에 의하여 지원되었음.

5. 참고문헌

- [1] Song, Liangchen, et al. "Unsupervised domain adaptive re-identification: Theory and practice." Pattern Recognition 102 (2020): 107173.
- [2] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- [3] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- [4] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and

Q. Tian. Scalable Person Re-identification: A Benchmark. 2015 IEEE International Conference on Computer Vision (ICCV), pages 1116–1124, Dec 2015.

- [5] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo 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