영상 기반 사람 검색 주요 기술과 현황

□ 엄찬호 / 중앙대학교

요 약

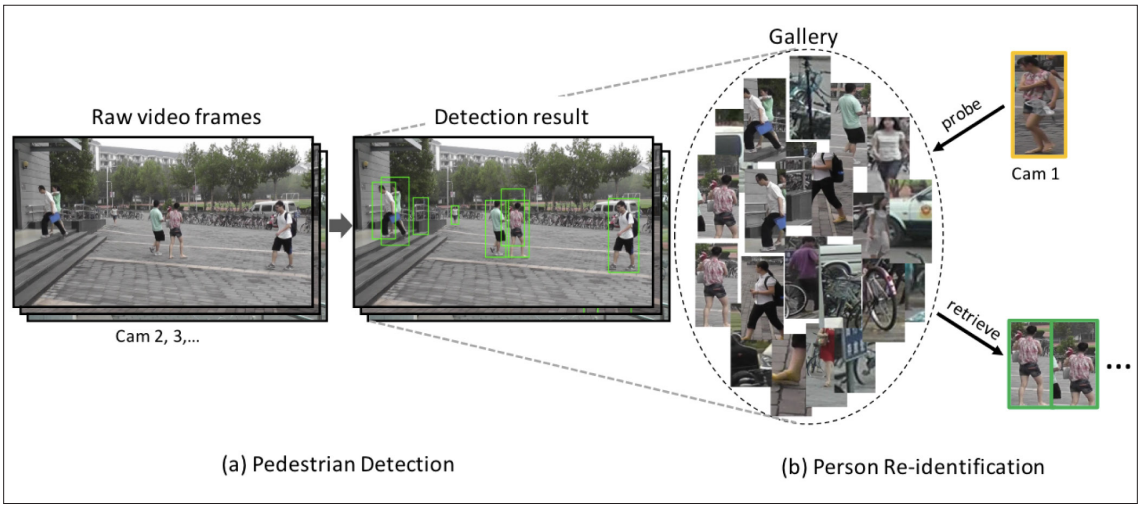
본 기고에서는 최근 컴퓨터 비전 분야에서 활발하게 연구되고 있는 여러 분야 중에 하나인 영상 기반 사람 검색의 주요 기술과 현재 동향을 분석하고자 한다. 다양한 컴퓨터 비전 기술들이 발전함에 따라 사람 검색 기술도 큰 발전을 이루었으며, 영상 기반 사람 검색 기술은 다양한 분야에서 중요한 역할을 하고 있다. 사람 재식별(re-identification)과 객체 탐지(object detection) 기술에 기반한 사람 검색은 보안, 감시, 소매 및 스마트 시티 등 여러 응용 분야에서 큰 주목을 받고 있다. 본 기고에서는 사람 검색 기술의 기술적 구성 요소와 최신 연구 방향에 대해 논의한다.

I. 서 론

영상 기반 사람 검색(person search)은 대규모 영상 데이터베이스에서 특정 인물을 찾아내는 기술로, 스마트 감시 시스템 및 방송 화면 기반 운동 경기 분석 등 여러 분야에서 활용되고 있다. 이러한 기술은 사람 재식별(re-identification)과 객체 탐지(detection) 기술에 기반을 두고 있다(〈그림 1〉). 본 기고에서는 영상 기반 사람 검색 기술의 주요 구성 요소와 최신 연구 동향을 소개하고자 한다.

사람 검색 기술은 최근 몇 년간 급격한 발전을 이루었다. 이러한 기술은 보안, 감시, 개인화된 서비스 제공 등의 다양한 응용 분야에서 중요한 역할을 한다. 예를 들어, 공항과 같은 보안이 중요한 장소에서는 실시간으로 의심스러운 행동을 감지할 수 있다. 또한, 소매점에서는 고객의 동선을 분석하여 맞춤형 서비스를 제공할 수 있다.

영상 기반 사람 검색 기술은 크게 두 가지 접근 방식으로 나눌 수 있다. 첫째, 전통적인 컴퓨터 비전 접근 방식으로서, 주로 특징자 추출과 매칭 기술을 사용한다.



<그림 1> 사람 검색 기술 파이프라인 (Zheng et. al., 2017)

둘째, 최근의 기술을 활용한 방법으로서, 특히 트랜스포머 및 확산 모델을 통해 높은 성능을 보이고 있다. 이 기고에서는 영상 기반 사람 검색 정확도를 높이기 위해 시도되고 있는 다양한 접근 방식을 비교하고 분석하고자 한다.

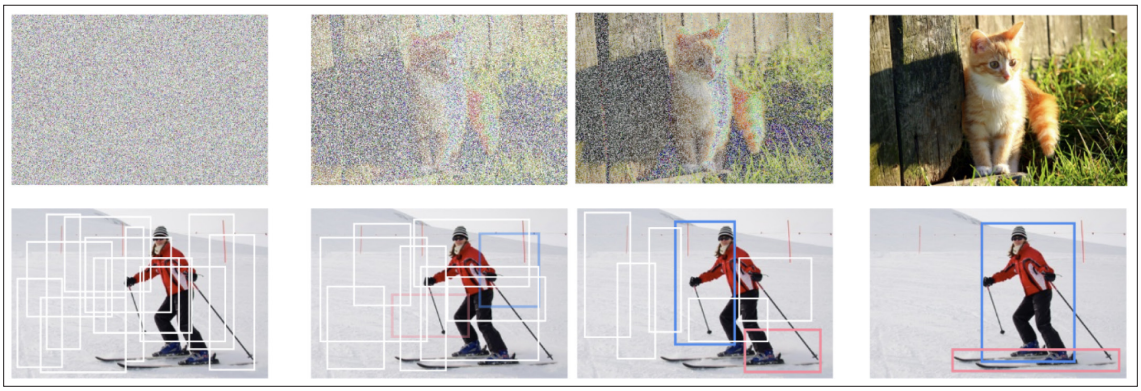
II. 기술적 구성 요소

1. 객체 탐지

객체 탐지(object detection) 기술은 이미지나 영상에서 객체의 위치를 탐지하고 분류하는 기술이다. 이 기술은 주로 사람 검출에 사용되며, 다양한 배경과 조건에서도 사람을 정확히 검출할 수 있다. Faster R-CNN(Ren et al., 2017)과 같은 방법론은 높은 정확도를 자랑하며, 다양한 응용 분야에서 활용되고 있다. 이 모델은 영역 제안(region proposal) 네트워크를 사용하여 객체의 후보 영역을 제안하고, 이를 기반으로 객체를 분류하고 위치를 결정한다. 객체 탐지 기술의 최근 연구는 실시간 성능 향상에 중점을 두고 있다. YOLO(Redmon et al., 2016)와 같은 모델은 단일 패스(single pass)로 객체를 검출하여 실시간 처리가 가능하다. 이러한 모델은 자율 주행, 실시간 감시 시스템 등 다양한 분야에서 큰 주목을 받고 있다. 최근에는 Diffusion 모델이 객체 탐지 분야에서 주목받고 있다. Diffusion 모델은 이미지 생성 및 변환 작업에서 사용되는 기법으로, 잡음이 추가된 이미지에서 점진적으로 잡음을 제거하며 원본 이미지를 복원하거나 새로운 이미지를 생성하는 모델이다. 최근 딥러닝 분야에서 뛰어난 성능을 보여 많은 주목을 받고 있는데, 이를 활용하여 객체 탐지를 시도한 방법이 DiffusionDet(Chen et. al., 2023)이다(〈그림 2〉). 기존 객체 탐지 방법들과 다르게 별도의 찾고자 하는 물체 개수를 설정하지 않아도 되는 장점이 있으면서도 기존 방법 대비 객체 검출에 강력한 성능을 보여주고 있다.

2. 사람 재식별

사람 재식별(re-identification) 기술은 다양한 카메라 뷰(view)에서 촬영된 사람을 동일인으로 인식하는 기술이다. 이는 주로 감시 시스템에서 사용되며, 다양한 각도와 조명 조건에서도 동일 인물을 인식하는 데 중점을 둔

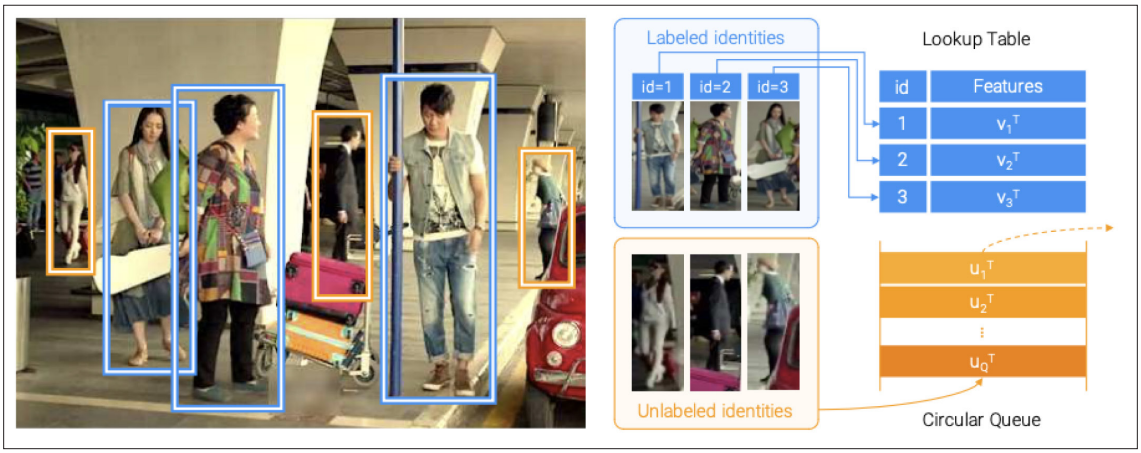


<그림 2> Diffusion 모델 기반 객체 탐지 (Chen et al., 2023)

다. 딥러닝 기반의 방법론을 사용하여 높은 성능을 보이고 있다. 사람 검색 시스템의 핵심은 효과적인 특징 추출에 있다. 즉, 서로 다른 조건에서 촬영된 동일 인물의 영상에서 공통적인 정보, 가령, 성별, 옷 색깔 및 특징 등의 정보를 추출하여 사람 표현자에 담는 것이 핵심이다. ResNet(Zheng et al., 2017)과 같은 신경망 구조는 강력한 특징 추출 능력을 제공하여 다양한 변형에도 견고한 성능을 보인다. 이러한 모델은 이미지의 저수준(low-level) 특징부터 고수준(high-level) 특징까지 계층적으로 학습하여, 다양한 조건에서도 높은 성능을 보인다. 특징 추출 과정에서 중요한 요소는 특징의 불변성(invariance)이다. 이는 다양한 각도, 조명, 배경 조건에서도 동일한 특징을 유지하는 능력을 의미한다. 최근의 연구에서는 불변성을 높이기 위해 여러 가지 기법을 도입하고 있다. 예를 들어, 다중 스케일 특징 추출(multi-scale feature extraction) 기법은 다양한 크기와 해상도의 특징을 동시에 학습하여, 더 견고한 성능을 제공한다.

3. 손실 함수

효과적인 손실 함수는 모델의 학습 성능에 직접적인 영향을 미친다. 손실 함수는 모델이 예측한 결과와 실제 값 간의 차이를 측정하여, 모델이 올바른 방향으로 학습할 수 있도록 한다. 사람 재식별 모델에서는 삼중항 손실(triplet loss)과 교차 엔트로피 손실(cross-entropy loss) 등이 일반적으로 사용된다. 삼중항 손실은 세 개의 샘플(양성, 음성, 앵커)을 사용하여, 양성 샘플과 앵커 샘플 간의 거리를 줄이고, 음성 샘플과 앵커 샘플 간의 거리를 늘리는 방식으로 학습한다. 이 방법은 특히 사람 재식별과 같은 응용 분야에서 유용하다. 교차 엔트로피 손실은 예측된 확률 분포와 실제 분포 간의 차이를 측정하여, 분류 문제에서 자주 사용된다. 사람 검색 모델의 경우 OIM (Online Instance Matching) 손실 함수가 사용되고 있다 (〈그림 3〉). 이 손실 함수는 사람 검색을 위한 효율적인 학습을 가능하게 하는 몇 가지 주요 특징을 가지고 있다. 첫째, OIM 손실 함수는 훈련 중에 인스턴스의 임베딩을 동적으로 업데이트한다. 이는 각 배치에서 새로 들어오는 샘플들의 임베딩 벡터를 사용하여 즉시 갱신하는 방식으로 이루어진다. 이렇게 하면 전체 데이터셋을 한 번에 처리하지 않아도 되기 때문에 메모리 효율성이 높아지고, 실시간 학습이 가능하다. 둘째, OIM 손실 함수는 두 가지 메모리 매핑을 사용한다. 첫 번째는 각 클래스(사람 ID)의 임베딩 벡터를 저장하는 ID 메모리이고, 두 번째는 최근에 본 인스턴스들의 임베딩 벡터를 저장하는 인스턴스 메모리이다. 이러한 메모리 매핑을 통해, OIM 손실 함수는 각 클래스의 대표 임베딩과 최근 인스턴스들의 임베딩 간의 거리를 줄이도록 학습한다. OIM 손실 함수는 결과적으로 비슷한 클래스 간의 분리도를 높이는 데 도움이 되어 다양한



<그림 3> OIM 손실 함수 도식화

각도, 조명 조건, 배경에서 촬영된 동일 인물을 정확하게 인식할 수 있는 모델을 학습하는 데 중요한 역할을 한다.

III. 최신 기술 동향

1. 주의 매커니즘

주의 메커니즘(attention mechanism)은 이미지 내 중요한 부분에 집중하여 검색 성능을 향상시킨다. 최근 연구에서는 주의(attention) 기법을 사용하여 사람 재식별 성능을 향상시키는 방법들이 많이 제안되었다. 대표적으로, ABD-Net(Chen et al., 2019b)이 있으며(〈그림 4〉), 이 모델은 글로벌 및 로컬 특징을 동시에 학습하여, 다양한 환경에서도 높은 인식률을 보인다. 주의 메커니즘은 특히 복잡한 장면에서 유용하며, 중요한 객체나 영역에 더 많은 가중치를 부여하여 모델이 더 정확한 예측을 할 수 있도록 한다(Chen et al., 2022). 주의 메커니즘의 또 다른 중요

<그림 4> 주의 기법 기반 사람 표현자 추출 (Chen et al., 2019b)

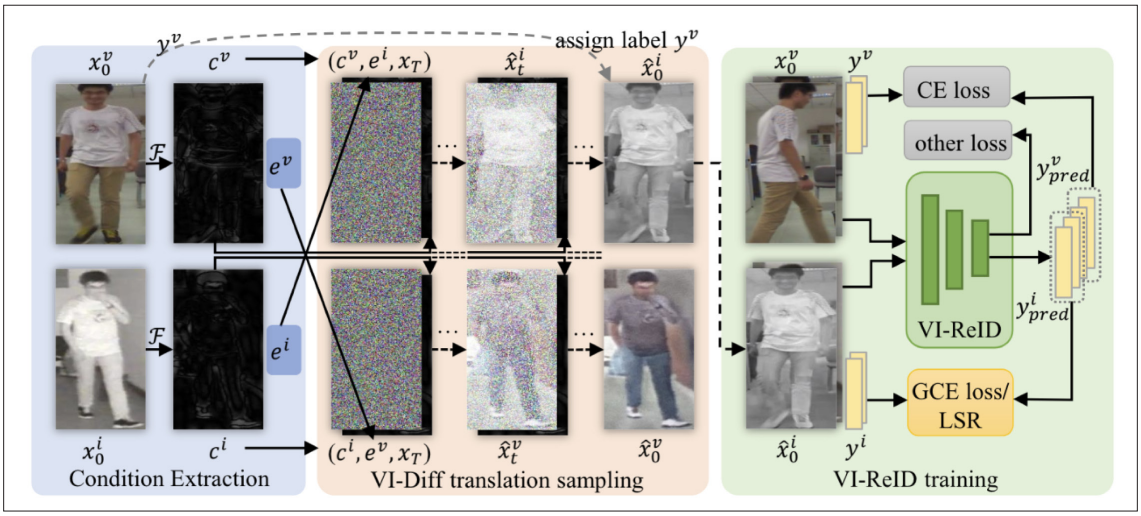
한 요소는 셀프 어텐션(self-attention) 기법이다. 이는 입력 이미지의 모든 부분 간의 상호작용을 고려하여, 더 풍부한 표현을 학습할 수 있도록 한다. 셀프 어텐션 기법은 특히 장거리 의존성(long-range dependency)을 학습하는 데 유용하다. 최근의 연구에서는 셀프 어텐션을 기반으로 한 트랜스포머(transformer) 모델이 주목받고 있다.

2. 데이터 증강

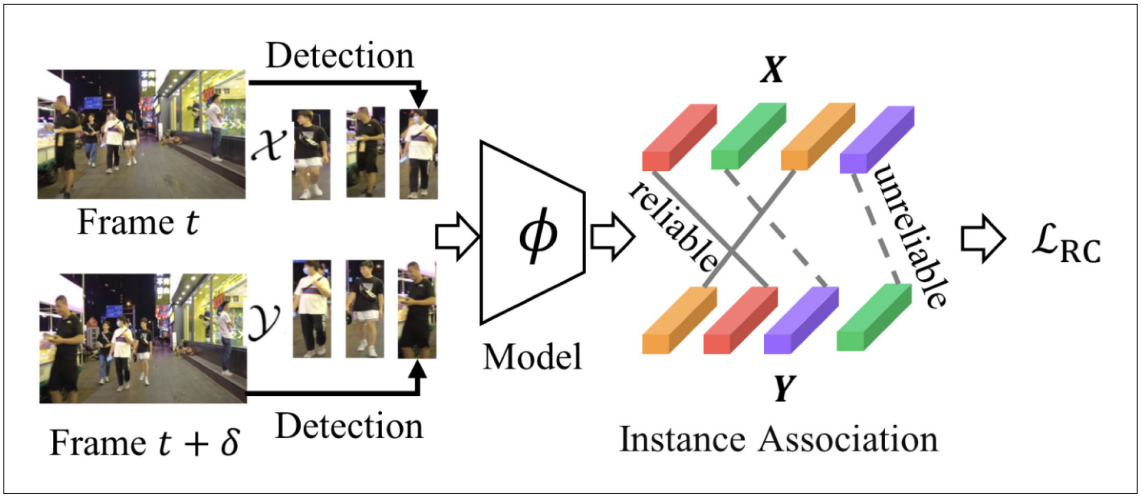
사람 재식별 기술의 또 다른 중요한 요소는 데이터 증강(data augmentation)이다. 데이터 증강(data augmentation) 기법은 학습 데이터를 다양화하여 모델의 일반화 성능을 향상시킨다. 예를 들어, 무작위 지우기(random erasing) 기법은 훈련 데이터의 일부를 임의로 삭제하여 모델이 더 견고하게 학습되도록 한다(Zhong et al., 2020). 이 기법은 특히 과적합(overfitting)을 방지하는 데 효과적이다. 데이터 증강의 또 다른 중요한 기법은 스타일 전이(style transfer)이다. 이는 한 이미지의 스타일을 다른 이미지에 적용하여, 다양한 환경 조건에서의 학습 데이터를 생성하는 방법이다. 예를 들어, 날씨 조건, 시간대, 조명 변화 등을 모방하여 다양한 학습 데이터를 생성할 수 있다. 또한, 더 나아가 학습 데이터에 없던 새로운 사람의 이미지들을 생성하고 이를 추가적인 학습 데이터로 이용하는 방식들 또한 많이 제안되었다. 대표적으로 Vi-Diff(Huang et. al., 2023)라는 모델이 있으며, RGB-IR 멀티 모달리티 환경에서 특정 모달리티를 이용해 다른 모달리티에서 촬영됐을 이미지를 Diffusion 모델을 통해 생성하여 학습에 이용하는 방식이다(〈그림 5〉).

3. 비지도 및 자가 지도 학습

자가 학습(Self-supervised Learning) 기술은 라벨이 없는 데이터로부터 유용한 표현을 학습하는 방법으로, 사람 검색 분야에서도 그 잠재력이 크다(〈그림 6〉). 기존의 자가 학습 기법들은 주로 이미지나 비디오 데이터를 활용하여 특징을 추출하고, 이를 기반으로 사람을 식별하는 데 초점을 맞추고 있다. 예를 들어, BERT와 같은 모델은 자연어 처리 분야에서 높은 성능을 보여주었으며, 이러한 접근법이 컴퓨터 비전 분야에서도 적용되고 있다. 라벨이 없는 데이터에서 유용한 특징을 학습하는 비지도 학습(unsupervised learning) 기법도 활발히 연구되고 있다. 이는 대규모의 라벨링되지 않은 데이터로부터 의미 있는



<그림 5> Diffusion 모델을 활용한 데이터 증강 (Huang et. al., 2023)



<그림 6> 자가 학습을 활용한 기법 예시 (Dou et. al., 2023)

정보를 추출하는 데 유용하다. 비지도 학습 기법은 특히 데이터 라벨링 비용을 절감하고, 다양한 환경에서의 일반화 성능을 향상시킬 수 있다. 비지도 학습의 대표적인 방법은 클러스터링(clustering)이다. 이는 유사한 특징을 가진 데이터 포인트를 그룹화하여, 의미 있는 패턴을 발견하는 방법이다. 최근의 연구에서는 딥 클러스터링(deep clustering) 기법이 주목받고 있으며, 이는 딥러닝 모델을 사용하여 더 높은 수준의 특징을 학습할 수 있다.

4. 사람 외형 특징

속성 인식(attribute recognition)은 사람의 옷차림, 소지품 등의 속성을 인식하여 추가적인 정보를 제공하는 기술이다. 이는 사람 재식별의 보조 역할을 하며, 다양한 속성을 학습하여 재식별 성능을 높일 수 있다. 속성 인식은 특히 복잡한 환경에서 사람을 더 정확하게 인식하는 데 유용하다. 예를 들어, 촬영된 사람의 속성을 인식하고 이를 활용하여 재식별 성능을 향상시켰다. 속성 인식 기술은 또한 다중 태스크 학습(multi-task learning) 방식으로 발전하고 있다. 이는 한 번의 학습 과정에서 여러 가지 속성을 동시에 학습하여, 더 효율적이고 정확한 인식을 가능하게 한다. 최근의 연구에서는 딥러닝 기반의 속성 인식 모델이 높은 성능을 보이고 있다.

IV. 결 론

본 기고에서는 영상 기반 사람 검색 기술의 주요 구성 요소와 최신 연구 동향을 소개하였다. 이 분야는 빠르게 발전하고 있으며, 향후 더 많은 연구가 필요할 것이다. 특히, 딥러닝 기반의 방법론은 높은 성능을 제공하며, 다양한 응용 분야에서 큰 잠재력을 가지고 있다. 그러나 여전히 해결해야 할 문제들이 많이 남아 있으며, 향후 연구를 통해 이러한 문제들을 해결해 나갈 필요가 있다.

참 고 문 헌

[1] Bi, X., & Wang, H. (2024). Appearance-pose joint coordinates information collaboration model for clothes-changing person re-identification. Expert Syst. with Appl., 241, 122473.
[2] Cao, Y.-T., Wang, J., & Tao, D. (2020). Symbiotic adversarial learning for attribute-based person search. In Proc. Eur. Conf. Comput. Vis. (pp. 230-247).
[3] Chen, D., Li, H., Liu, X., Shen, Y., Shao, J., Yuan, Z., & Wang, X. (2018). Improving deep visual representation for person re-identification by global and local image-language association. In Proc. Eur. Conf. Comput. Vis. (pp. 54-70).
[4] Chen, G., Lin, C., Ren, L., Lu, J., & Zhou, J. (2019a). Self-critical attention learning for person re-identification. In Proc. IEEE Int. Conf. Comput. Vis. (pp. 9637-9646).
[5] Chen, T., Ding, S., Xie, J., Yuan, Y., Chen, W., Yang, Y., Ren, Z., & Wang, Z. (2019b). ABD-net: Attentive but diverse person re-identification. In Proc. IEEE Int. Conf. Comput. Vis. (pp. 8351-8361).
[6] Chen, X., Fu, C., Zhao, Y., Zheng, F., Song, J., Ji, R., & Yang, Y. (2020). Salience-guided cascaded suppression network for person re-identification. In Proc. IEEE Conf. Comput. Vis. Pattern Recog. (pp. 3300-3310).
[7] Chen, Y., Wang, H., Sun, X., Fan, B., Tang, C., & Zeng, H. (2022). Deep attention aware feature learning for person re-identification. Pattern Recognit., 126, 108567.
[8] Deng, Y., Luo, P., Loy, C. C., & Tang, X. (2014). Pedestrian attribute recognition at far distance. In Proc. ACM Int. Conf. on Multimedia (pp. 350-359).
[9] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.
[10] Loshchilov, I., & Hutter, F. (2016). Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983.
[11] Luo, H., Gu, Y., Liao, X., Lai, S., & Jiang, W. (2019). Bag of tricks and a strong baseline for deep person re-identification. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshop.
[12] Maaten, L. v. d., & Hinton, G. (2008). Visualizing data using t-sne. J. Mach. Learn. Res., 9(Nov), 2579-2605.
[13] Nguyen, B. X., Nguyen, B. D., Do, T., Tjiputra, E., Tran, Q. D., & Nguyen, A. (2021). Graph-based person signature for person re-identifications. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (pp. 3492-3501).
[14] Ni, X., Fang, L., & Huttunen, H. (2021). Adaptive l2 regularization in person re-identification. In Int. Conf. Pattern Recognit. (pp. 9601-9607).
[15] Quispe, R., & Pedrini, H. (2021). Top-db-net: Top dropblock for activation enhancement in person re-identification. In Int. Conf. Pattern Recognit. (pp. 2980-2987).
[16] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. (2021). Learning transferable visual models from natural language supervision. In Proc. Int. Conf. Mach. Learn (pp. 8748-8763).
[17] Ren, M., He, L., Liao, X., Liu, W., Wang, Y., & Tan, T. (2021). Learning instance-level spatial-temporal patterns for person re-identification. In Proc. IEEE Int. Conf. Comput. Vis. (pp. 14930-14939).
[18] Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell., 39(6), 1137-1149.
[19] Zheng, L., Zhang, H., Sun, S., Chandraker, M., Yang, Y., & Tian, Q. (2017). Person re-identification in the wild. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (pp. 1367-1376).
[20] Chen, S., Sun, P., Song, Y., & Luo, P. (2023). Diffusiondet: Diffusion model for object detection. In Proc. IEEE Int. Conf. Comput. Vis. (pp. 19830-19843).
[21] Xiao, T., Li, S., Wang, B., Lin, L., & Wang, X. (2017). Joint detection and identification feature learning for person search. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (pp. 3415-3424).
[22] Huang, H., Huang, Y., & Wang, L. (2023). Vi-diff: Unpaired visible-infrared translation diffusion model for single modality labeled visible-infrared person re-identification. arXiv preprint arXiv:2310.04122.
[23] Dou, Z., Wang, Z., Li, Y., & Wang, S. (2023). Identity-seeking self-supervised representation learning for generalizable person re-identification. In Proc. IEEE Int. Conf. Comput. Vis. (pp. 15847-15858).

저 자 소 개

엄 찬 호

- 2012년 ~ 2017년 : 연세대학교 전기전자공학과 학사
- 2017년 ~ 2023년 : 연세대학교 전기전자공학과 박사
- 2023년 : 삼성전자종합기술원 (SAIT) 책임연구원
- 2023년 ~ 현재 : 중앙대학교 첨단영상대학원 조교수
- 주관심분야 : 컴퓨터 비전, 인공지능, 영상 처리 등