

사람 탐색을 위한 Global Pooling 방법 분석

The Analysis of Global Pooling Strategies in Person Search

○Tang Qing¹, 최 제 환², 조 강 현^{3*}

¹⁾ 울산대학교 전기전자컴퓨터공학과 (TEL: 052-259-1664, E-mail: zucchini.tang@gmail.com)

²⁾ 울산대학교 전기전자컴퓨터공학과 (TEL: 052-259-1664, E-mail: jhchoi@islab.ulsan.ac.kr)

³⁾ 울산대학교 전기전자컴퓨터공학과 (TEL: 052-259-2208, E-mail: acejo@ulsan.ac.kr)

Abstract The specific person search is the foundation of a wide range of applications in intelligent security and surveillance systems. Person search aims at solving person detection and person re-identification (re-id) jointly. The common practice is to append the standard detection loss and re-id branches parallelly on Faster RCNN. For generating 2048-dim re-ID embedding features, global max pooling has been commonly adopted so far. In this paper, we analyze the performance of different pooling strategies, including global average pooling and global max pooling, for person search task on a widely used person search benchmark PRW. Then, a Global Pooling Attention (GPA) is proposed to further boost the performance of network by utilizing the self-attention information. The experimental results demonstrate the advantages of global max pooling in person search task and the effectiveness of the proposed GPA.

Keywords Person search, Pooling strategy, Person detection

1. 서론

지능형 보안 및 감시시스템은 최근 공공안전 수요 증가 및 공공장소에 설치된 카메라의 상호연결, 강도 높은 보안업무, 산업현장의 컴퓨터비전의 발전과 같은 이유로 주요 연구대상이 되었다. 지능형 보안 및 감시시스템의 기초가 되는 사람 탐색 (Person Search)은 특정 인물 탐색, 다수의 카메라를 사용한 여러 객체 추적[1] 및 인간 행동 분석[2]과 같은 실생활에서의 사람 재식별 (Re-Identification)에 대한 수요가 증가함에 따라 상당한 주목을 받고 있다.

사람 탐색은 사람 검출 (Person Detection)과 사람 재식별의 통합기술로 볼 수 있다. 일반적인 사람 탐색 시스템은 그림 1(a)에 묘사되어있다. 입력이미지에서 256개의 후보군이 선정되며 RoIAlign에 의해 7x7x2048의 RoI들로 정렬된다. 그 후, 일반적으로 [3, 4]와 같이 채널별로 global pooling이 적용되어 1x1x2048의 특징을 생성한다. 기존 방식 중 [4]은 average pooling과 함께 max pooling을 함께 사용하여 더 나은 결과를 보였다.

본 논문에서는 다른 global pooling 방법의 효율성을 분석한다. 실험결과를 통해 average pooling이 사람 탐색에 좋지 않은 결과를 도출한다는 것을 입증한다. 또한 max pooling을 사용한 attention 기법을 제안한다.

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

2. 본론

2.1 Baseline network

SeqNet (Sequential End-to-end Network)[4]의 기본구조를 본 논문의 토대로 적용했다. 기본구조는 그림 1(a)에 묘사되어있다. 입력이미지가 주어지면, RPN (Region Proposal Network)가 256개의 후보군을 선정하고 RoIAlign은 7x7x2048의 RoI로 정렬한다. global pooling은 RoI들을 2048차원의 특징으로 변형시키고 해당 특징들은 검출 손실값과 Re-ID를 계산할 때 사용된다. Re-ID 손실값은 NAE (Norm-Aware Embedding Method)[3]를 사용해서 계산된다.

2.2 GPA module

Attention map은 채널축을 따라 특징의 상호 채널 관계를 활용하여 생성된다. 효율적인 attention map 계산을 하기 위해서 그림 1(b)와 같이 hidden layer를 적용한 MLP (Multi-Layer Perceptron)가 사용된다. Attention map a 는 식 1과 같이 sigmoid 함수 $\sigma(\cdot)$ 와 함께 계산된다.

$$a = \sigma(MLP(f)) \quad (1)$$

그리고 attention map a 는 식 (2)와 같이 픽셀별 곱셈을 사용하여 feature embedding f 와 통합된다.

$$f' = a \times f \quad (2)$$

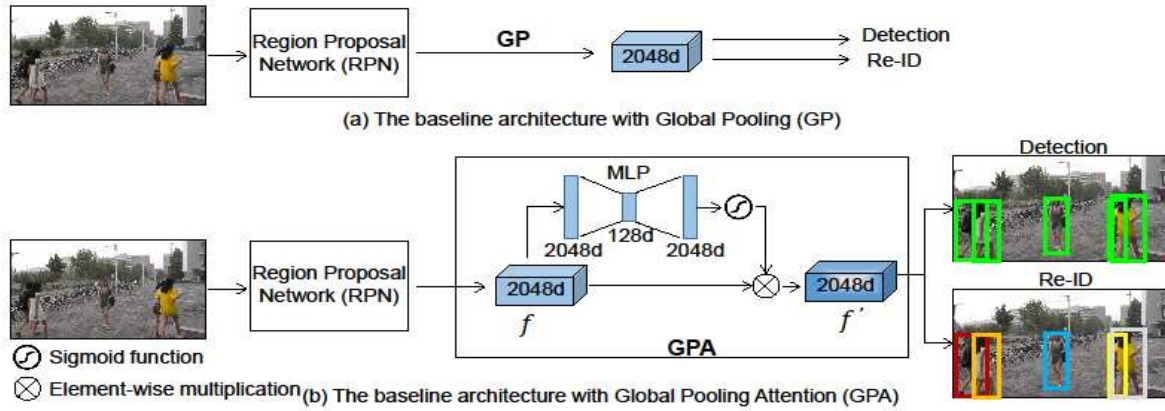


그림 1. Person search system의 네트워크 구조.

3. 실험

3.1 Dataset and Evaluation Metrics

PRW[6]은 총 11,816장의 거리 및 영화장면 이미지가 존재하며, 4,310명의 사람 BBox (Bounding Box)와 484개의 인식 레이블이 포함되어있다.

평가지표로는 mAP (Mean Average Precision)와 CMC (Cumulative Matching Characteristic) 커브가 사용되었으며 CMC의 가장 높은 값이 Top-1으로 기록된다.

3.2 Implementation Details

입력 이미지는 900 x 1500으로 조정되며, Batch size는 5로 설정한다. 네트워크는 학습률(Learning rate) 0.003인 SGD(Stochastic Gradient Descent)에 의해 학습되며, 7 epoch 시 0.0003로 감소한다. 실험은 총 10 epoch만큼 진행했으며, 32GB 메모리 NVIDIA Tesla V100으로 7시간동안 진행되었다.

3.3 Experimental Results

표 1. 다양한 pooling 방법 적용에 의한 결과

Global Pooling Strategy	Detection Acc. (%)		Re-ID Acc. (%)	
	Recall	AP	mAP	Top-1
Avg.	96.7	94.0	46.4	82.9
Avg. + Attention	96.5	93.7	46.7	83.6
Max	96.8	93.9	48.4	84.4
Max + Attention	97.0	94.1	48.9	85.2
Avg. + Max + Attention	97.0	94.2	48.1	84.0

본 실험은 기본 구조에 다른 global pooling 방법을 적용하여 진행했다. 결과는 표 1과 같으며, 먼저 global average pooling(Avg.)와 global max pooling(Max) 두 종류의 방법을 비교하였다. ‘Avg.’와 비교하였을 때, ‘Max’는 검출 정확도는 증가하지 않았지만 Re-ID 성능은 증가하였다. 특히 ‘Max’를 사용하면 mAP가 2.0% 증가했고, Top-1값도 1.5% 증가했다. ‘Max’의 Re-ID 성능증가는 사람 재식별 기법에서 독특한 특징이 더 중요하다는 것을 입증한다. 즉, average pooling 기능은

Re-ID 성능이 저하된다는 것을 확인하였다.

표 1에서는 global average pooling과 global max pooling이 attention module과 함께 사용된 것을 각각 ‘Avg. + Attention’, ‘Max + Attention’으로 표기하였다. 또한, 제안한 GPA가 일반적인 global pooling보다 효과적인 것을 볼 수 있다. ‘Avg. + Attention’와 ‘Max + Attention’은 일정하게 좋은 성능을 보였고, 특히 global max pooling에서 잘 나타난다.

4. 결론

본 논문은 end-to-end pooling 방법을 사용한 사람 탐지 모델연구를 진행했다. 먼저 서로 다른 global pooling 방법의 효율성을 비교하였고, attention module을 적용한 global pooling 기법을 제안했다. 실험결과는 상황에 맞는 pooling 기법을 적용하는 것의 중요성과 제안한 attention module의 효율성도 입증하였다.

참고문헌

- [1] C. Liang, Z. Zhang, Y. Lu, X. Zhou, B. Li, X. Ye, and J. Zou, “Rethinking the competition between detection and reid in multi-object tracking,” ArXiv, vol. abs/2010.12138, 2020.
- [2] L. Zheng, Y. Yang, and A. Hauptmann, “Person re-identification: Past, present and future,” ArXiv, vol. abs/1610.02984, 2016.
- [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] Z. Li and D. Miao, “Sequential end-to-end network for efficient person search,” in AACL, 2021.
- [5] Woo, S., Park, J., Lee, J., & Kweon, I. (2018). CBAM: Convolutional Block Attention Module. ECCV.
- [6] Zheng, Liang et al. “Person Re-identification in the Wild.” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017): 3346-3355.