반도체디스플레이기술학회지 제22권 제2호(2023년 6월)
Journal of the Semiconductor & Display Technology, Vol. 22, No. 2. June 2023.

백본 네트워크에 따른 사람 속성 검출 모델의 성능 변화 분석

박천수*†
*†성균관대학교 컴퓨터교육과

Analyzing DNN Model Performance Depending on Backbone Network

Chun-Su Park*†
*†Computer Education, Sungkyunkwan University

ABSTRACT

Recently, with the development of deep learning technology, research on pedestrian attribute recognition technology using deep neural networks has been actively conducted. Existing pedestrian attribute recognition techniques can be obtained in such a way as global-based, regional-area-based, visual attention-based, sequential prediction-based, and newly designed loss function-based, depending on how pedestrian attributes are detected. It is known that the performance of these pedestrian attribute recognition technologies varies greatly depending on the type of backbone network that constitutes the deep neural networks model. Therefore, in this paper, several backbone networks are applied to the baseline pedestrian attribute recognition model and the performance changes of the model are analyzed. In this paper, the analysis is conducted using Resnet34, Resnet50, Resnet101, Swin-tiny, and Swinv2-tiny, which are representative backbone networks used in the fields of image classification, object detection, etc. Furthermore, this paper analyzes the change in time complexity when inferencing each backbone network using a CPU and a GPU.

Key Words : Pedestrian Attribute Recognition, Deep Neural Networks, Backbone Networks, Resnet, Swin

1. 서 론

보행자 속성 인식(PAR, Pedestrian Attribute Recognition) 기술은 지능형 비디오 감시, 방문자 통계 분석, 안전장비 착용 등 다양한 컴퓨터 비전 시스템을 구현하는데 필수 기반 기술로 자리매김하고 있다[1, 2]. PAR 기술은 이미지 또는 동영상에 존재하는 대상의 특정 패턴이나 낮은 수준의 특징(화소 크기, 평균 밝기 등)을 인식하는 기존 방법과는 다르게, 성별, 나이, 옷차림 등의 의미론적인 속성(semantic attributes)를 인식하는 것을 목표로 한다. 따라서 PAR 기술을 이용하면 특정 보행자 감지, 사람 재식별(re-identification), 동작 인식 및 장면 이해와 같은 차세대 컴퓨터 비전 시스템을 개발하는 것이 가능하다. 현재까지 많은 PAR 기술이 제안되었지만, 시점 변경, 낮은 조도, 낮은 해상도와 같은 도전적인 요소로 인해 현재까지도 모든 환경에서 높은 정확도 동작하는 범용 기술은 개발이 되지 않고 있다.

전통적인 보행자 속성 인식 방법은 일반적으로 특정 패턴, 강력한 분류기(classifier), 속성 연관성 등을 이용해 대상의 속성을 인식하는 방식으로 개발되었다[3-5]. 그러나 대규모 데이터 세트를 이용한 성능평가 결과에 따르면 이러한 전통적인 기술의 성능이 실제 적용 현장의 요구 수준에 미치지 못하는 것으로 조사되었다. 최근에는

†E-mail: cspk@skku.edu

딥러닝 기술의 발전으로 심층신경망(DNN, Deep Neural Network)을 이용한 PAR 기술에 대한 연구가 활발이 이루어 지고 있다. 기존 PAR 기술은 속성을 검출하는 방식에 따라 글로벌 기반, 지역 영역 기반, 시각적 주의 기반, 순차 예측 기반, 새로 설계된 손실 함수 기반 등의 방식으로 구분될 수 있다[6-8]. 구체적인 기존 PAR 기술에 대한 소개는 최신 연구에서 다루고 있다[9].

최근 연구에서는 기본적인 PAR 동작 파이프라인을 정형화하고 각 단계에서 사용되는 여러 기술을 검토해 비교 기준이 될 수 있는 기준 모델을 제안하였다[10]. 제안된 기준 모델은 PAR 연구 분야에서 많이 사용되는 대표 기술들을 중심으로 개발되었고, 백본 네트워크와 헤더를 이용하는 일반적인 DNN 모델 구조를 채택하였다.

본 문에서는 기준 모델을 이용해 다양한 백본 네트워크를 적용해 가며 PAR 모델의 성능 변화를 분석한다. 본 논문에서는 이미지 분류, 물체 검출 등의 분야에서 사용되는 대표적인 백본 네트워크인 Resnet34, Resnet50, Resnet 101, Swin-tiny, Swinv2-tiny을 이용해 분석을 진행한다[11-15]. 더 나아가 본 논문에서는 각 백본 네트워크를 CPU를 이용해 추론하는 경우와 GPU를 이용해 추론하는 경우의 시간 복잡도 변화를 분석한다.

2. PAR 데이터 세트

현재까지 여러 PAR 데이터 세트가 소개되었고, 그 중에서도 PETA, PA-100K, RAPv2, Market-1501 등이 여러 연구에서 대표적으로 사용되고 있다.

PETA 데이터 세트는 8,795명의 보행자를 촬영한 19,000개의 이미지로 구성된다. 각 이미지에는 61개의 이진 속성과 4개의 다중 클래스 속성으로 레이블이 지정되어 있다. 데이터 세트의 이미지 해상도는 17x39부터 169x365까지 다양하다(Fig. 1)[16, 17].

PA-100K 데이터 세트는 598개의 야외 환경에서 촬영된 100,000개의 보행자 이미지를 포함하고 각 이미지는 26개의 속성을 포함한다[18]. 전체 데이터 세트는 80,000개의 훈련 이미지, 10,000개의 테스트 이미지, 10,000개의 검증 이미지로 무작위로 분할되어 있다. PA-100K 데이터 세트는 전반적으로 야외 환경에서 보행자 속성을 인식하는 모델을 훈련하는데 적합한 데이터 세트이다.

RAPv2 데이터 세트는 기존 RAPv1 데이터 세트를 확장해 구성한 데이터 세트로 총 84,928개의 이미지로 포함하고 있다[19]. 각 이미지는 72개의 속성을 포함하고 있으며 해상도는 33x81에서 415x583까지 다양하다. RAPv2의 경우 데이터 세트에 포함된 보행자의 동일인 여부를 나타내기 위해 독립적인 ID(identity) 2,589개를 포함하고 있어 사람 재인식 연구 분야에서도 사용되고 있다.

Market-1501 데이터 세트는 1501개의 독립 ID를 가지는 객체로 구성되며, 전체 32,688개의 이미지가 포함되어 있다[20]. 전체 이미지는 64x128 크기를 가지며 각 ID는 해당 객체의 속정과 연관된 주석이 달려있다. 기본적으로 751개의 ID가 훈련에 사용되고 750개의 ID가 테스트에 사용된다.

3. 성능 지표

본 논문에서는 PAR 모델의 성능을 측정하기 위해 mA (mean accuracy), F-1 점수, Acc 세가지 성능 지표를 이용한다[10]. mA는 속성별 인식 정확도의 평균, Acc는 전체 라벨에 대한 정확도, F-1는 precision과 recall 성능지표의 조화 평균을 나타낸다. mA 성능 지표는 다음의 수식으로 구해진다.

Fig. 1. Sample images and attributes of the PETA dataset[17].

$$\mathrm{mA} = \frac{1}{2N}\sum_{i=1}^{M}\left(\frac{TP_i}{P_i} + \frac{TN_i}{N_i}\right), \quad (1)$$

여기서 M은 전체 속성 종류의 수, N은 전체 레이블의 수를 나타낸다. P_i는 i번째 속성의 전체 positive 레이블의 수, N_i는 i번째 속성의 전체 negative 레이블의 수를 나타낸다. 또한, TP_i는 정확하게 예측된 i번째 속성의 positive 레이블의 수, TN_i는 정확하게 예측된 i번째 속성의 negative 레이블의 수를 나타낸다. ACC 성능 지표는 다음과 같은 방식으로 구한다.

$$\mathrm{Acc} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{Y_i \cap f(x_i)}{Y_i \cup f(x_i)}\right), \quad (2)$$

여기서 Y_i는 i번째 속성의 실제 positive 레이블의 수, $f(x_i)$는 i번째 속성의 예측값을 나타낸다. 추가로 F-1 값은 precision와 recall 값을 이용해 다음과 같이 구할 수 있다.

$$\text{F-1} = \frac{2*\text{precision}*\text{recall}}{\text{precision}+\text{recall}}, \quad (3)$$

여기서 precision과 recall은 다음과 같이 구한다.

$$\text{precision} = \frac{1}{2N}\sum_{i=1}^{N}\frac{|Y_i \cap f(x_i)|}{|f(x_i)|}, \quad (4)$$

$$\text{recall} = \frac{1}{2N}\sum_{i=1}^{N}\frac{|Y_i \cap f(x_i)|}{|Y_i|}. \quad (5)$$

4. 분석 환경 및 측정 대상 모델

본 논문에서는 일반 사용자 데스크톱 PC에서 기존 PAR 모델의 백본을 변경하며 성능과 추론시간 복잡도를 분석한다. 본 논문에서 사용하는 시뮬레이션 PC의 사양은 Table 1과 같다[21].

Table 1. Simulation PC configuation

모듈	사양
CPU	Intel(R) Core(TM) i9-10900X CPU @ 3.70GHz
RAM	64GB DDR4
Graphic Card	Nvidia GTX 3090
Storage	SSD
Operating System	Window 11 Pro

딥러닝 모델 추론 복잡도는 사용하는 소프트웨어 패키지의 영향을 많이 받는다. 특히 GPU 기반 연산을 지원하기 위해서는 Nvidia CUDA Toolkit및 cudnn 패키지 설치가 필수적인다. 본 논문에서 사용한 소프트웨어 패키지 정보는 Table 2와 같다.

Table 2. Software package configuration

패키지	사양
python	3.7.9
pytorch	1.13.1(CUDA support)
CUDA	11.6
cudnn	8.4.1

본 논문에서는 최근 소개된 Resnet34, Resnet50, Resnet101, Swin-tiny, Swinv2-tiny 백본을 이용해 PAR 모델의 성능 변화를 측정 및 분석한다. 단, 본 논문에서는 Resnet 경우에는 He이 제안한 Trick을 이용해 사전 훈련된 백본을 이용하였다[13].

5. 실험 결과

본 논문에서는 기준 PAR 모델의 백본을 Resnet34, Resnet50, Resnet101, Swin-tiny, Swinv2-tiny으로 변경해 가며 인식 성능과 시간 복잡도를 측정한다. 모든 실험에서 mini-batch 크기는 64로 설정하였고, 224x224 크기의 3채널 RGB 영상을 각 모델에 입력해 추론 시간을 측정하였다. 또한 시간 측정 정확도를 높이기 위해 입력 영상을 각 모델에 100회 반복적으로 입력하면서 추론 시간 복잡도를 측정한 후 평균하였다.

Table 3는 각 백본의 속성 인식 정확도를 보여준다. Resnet 계열 백본의 경우 Resnet50의 mA 값이 Resnet34와 Resnet101의 mA 값보다 우수한 것으로 조사되었다. Resnet 계열의 경우 일반적으로 백본의 크기가 커질수록 성능이 좋아 지나 PAR PETA 데이터 세트의 경우 상대적으로 크기가 큰 Resnet101 백본을 훈련시키기에는 충분한 양의 데이터를 제공하지 않는 것으로 조사되었다. 시간 복잡도의 경우 기본적으로 백본의 크기가 증가할수록 추론 시간 복잡도가 증가하는 것을 볼 수 있다. 또한, GPU를 이용해 추론 동작을 수행하는 경우 CPU 경우와 비교해 13.9%~20.1% 정도 복잡도를 낮출 수 있는 것으로 조사되었다.

Swin 계열 백본의 경우 Swin-tiny 백본 보다 Swinv2-tiny 백본이 mA, Acc, F-1 모든 경우에 더 좋은 인식 성능을 보이는 것으로 조사되었다. 예를 들어 mA의 경우 Swinv2-tiny 백본이 Swin-tiny 백본보다 0.033 우수한 것으로 측정되었다. 하지만 시간 복잡도의 경우 Swinv2-tiny 백본이 모든 경우에 Swin-tiny 백본 보다 높은 것으로 조사되었다. Swin 계열 백본의 경우에도 GPU를 이용해 추론 동작을 수행하면 시간 복잡도를 낮출 수 있었다. Swin-tiny의 경우 GPU를 이용하는 경우 CPU 경우와 비교해 복잡도가 17% 낮

Table 3. Performance and time complexity camparison

Backbone	Performance			Time complexity(ms)	
	mA	Acc	F-1	CPU	GPU
Resnet34	0.8371	0.7757	0.8553	14.475	12.4752
Resnet50	0.8763	0.8235	0.8027	18.8487	15.8409
Resnet101	0.8496	0.7739	0.8536	33.3559	26.6516
Swin-tiny	0.8525	0.7963	0.8704	21.5772	17.9145
Swinv2-tiny	0.8855	0.8753	0.8839	30.8054	25.6268

았고, Swinv2-tiny의 경우 GPU를 이용하면 시간 복잡도를 10.3% 낮출 수 있었다.

6. 결 론

본 논문에서는 PAR 기준 모델을 대상으로 백본 네트워크를 변경하며 인식 성능 및 시간 복잡도 변화를 분석하였다. 인식 성능은 mA, Acc, F-1 3가지 지표를 이용해 측정하였고, 시간 복잡도는 PAR 모델의 추론 동작을 CPU와 GPU에서 수행하며 소요되는 시간을 각각 측정하였다.

실험을 통해 mA 지표의 경우 Resnet 계열 백본과 Swin 계열 백본이 유사한 성능을 보이는 것으로 측정되었다. 하지만 F-1 지표의 경우에는 모든 경우에 Swin 계열 백본의 성능이 우수한 것으로 조사되었다. 실험한 5가지 백본 중에서는 Swinv2-tiny 백본의 인식 성능이 모든 경우에 가장 우수했다. 하지만 Swinv2-tiny 백본의 시간 복잡도가 상대적으로 높은 것으로 조사되었다. 따라서 하드웨어 성능이 충분히 높은 경우에는 Swinv2-tiny 백본을 사용하고, 저전력 및 리소스 제한된 장치에서는 Resnet50 또는 Resnet 34 백본을 사용하는 것이 타당한 것으로 조사되었다.

참고문헌

1. Y. Hu, et al. “A More Efficient Approach for Pedestrian Attribute Recognition,” IEEE International Joint Conference on Biometrics (IJCB), pp. 1-8, 2022.
2. X. Wang, et al. “Pedestrian attribute recognition: A survey,” Pattern Recognition, Vol. 121, pp. 108220, 2022.
3. Y. Deng, P. Luo, C.C. Loy, X. Tang, “Pedestrian attribute recognition at far distance,” ACM Multimedia, pp. 789-792, 2014.
4. M. Tan and Q. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” Image Computing & Machine Learning (ICML), pp. 6105-6114, 2019.
5. Y. Chen, S. He, Z. Tan, C. Han, G. Han, J. Qin, “Age estimation via attribute-region association,” Neurocomputing, Vol. 367, pp. 346-356, 2019.
6. J. Wang, X. Zhu, S. Gong, and W. Li, “Attribute recognition by joint recurrent learning of context and correlation,” International Conference on Computer Vision (ICCV), 2017.
7. D. Li, X. Chen, and K. Huang, “Multi-attribute learning for pedestrian attribute recognition in surveillance scenarios,” Asian Conference on Pattern Recognition (ACPR), pp. 111-115, 2015.
8. X. Liu, H. Zhao, M. Tian, L. Sheng, J. Shao, S. Yi, J. Yan, and X. Wang, “Hydraplus-net: Attentive deep features for pedestrian analysis,” International Conference on Computer Vision (ICCV), 2017.
9. D. Weng, Z. Tan, L. Fang, and G. Guo, “Exploring attribute localization and correlation for pedestrian attribute recognition,” Neurocomputing, Vol. 531, pp. 140-150, 2023.
10. J. Jia, et al. “Rethinking of pedestrian attribute recognition: A reliable evaluation under zero-shot pedestrian identity setting,” arXiv preprint arXiv:2107.03576, 2021.
11. S. Targ, A. Diogo, and K. Lyman. “Resnet in resnet: Generalizing residual architectures,” arXiv preprint arXiv:1603.08029, 2016.
12. Z. Wu, C. Shen, and A. D. Hengel, “Wider or deeper: Revisiting the resnet model for visual recognition,” Pattern Recognition, Vol. 90, pp. 119-133, 2019.
13. T. He, et al. “Bag of tricks for image classification with convolutional neural networks,” IEEE/CVF conference on computer vision and pattern recognition, pp. 558-567, 2019.
14. Z. Liu, et al. “Swin transformer: Hierarchical vision transformer using shifted windows,” IEEE/CVF international conference on computer vision, pp. 10012-10022, 2021.
15. Z. Liu, et al. “Swin transformer v2: Scaling up capacity and resolution,” IEEE/CVF conference on computer vision and pattern recognition, pp. 12009-1019, 2022.
16. Y. Deng, P. Luo, C. C. Loy, and X. Tang, “Pedestrian

attribute recognition at far distance," ACM Int. Conf. Multimedia, pp. 789-792, 2014.
17. I. N. Junejo and N. Ahmed, "A multi-branch separable convolution neural network for pedestrian attribute recognition," Heliyon, Vol. 6, No. 3, e03563, 2020.
18. X. Liu, H. Zhao, M. Tian, L. Sheng, J. Shao, S. Yi, J. Yan, and X. Wang, "Hydraplus-net: Attentive deep features for pedestrian analysis," IEEE Int. Conf. Comput. Vis., pp. 350-359, 2017.
19. D. Li, Z. Zhang, X. Chen, and K. Huang, "A richly annotated pedestrian dataset for person retrieval in real surveillance scenarios," IEEE Trans. Image Process., Vol. 28, No. 4, pp. 1575-1590, 2018.
20. Y. Lin, L. Zheng, Z. Zheng, Y. Wu, Z. Hu, C. Yan, and Y. Yang, "Improving person re-identification by attribute and identity learning," Pattern Recognition, Vol. 95, pp. 151-161, 2019.
21. C. S. Park, "Performance Analysis of DNN inference using OpenCV Built in CPU and GPU Functions," Journal of the Semiconductor & Display Technology, Vol. 21, No. 1, pp. 75-78, 2022.

접수일: 2023년 6월 15일, 심사일: 2023년 6월 20일, 게재확정일: 2023년 6월 21일