Generalization Study on Person Attribute Recognition

Jay Chaudhari[1], *Dinesh Nariani*[1], *Mehul S Raval*[1], *Paawan Sharma*[2]

[1]*Ahmedabad Univeristy*, [2]*PDEU, Gandhinagar*

Introduction

Person identification is a process for establishing the identity of a person based on unique attributes. A well trained network is necessary in developing end to end system using Natural Language Processing(NLP) and Computer Vision(CV) for person retrieval. The poster showcase the study on generalization ability of the network trained for Person Attribute Recognition(PAR).

Main Objectives

1. Understand the impact of difference in training and testing data on the network.
2. Understand challenge of CV in PAR.
3. Study generalization ability of network in PAR.

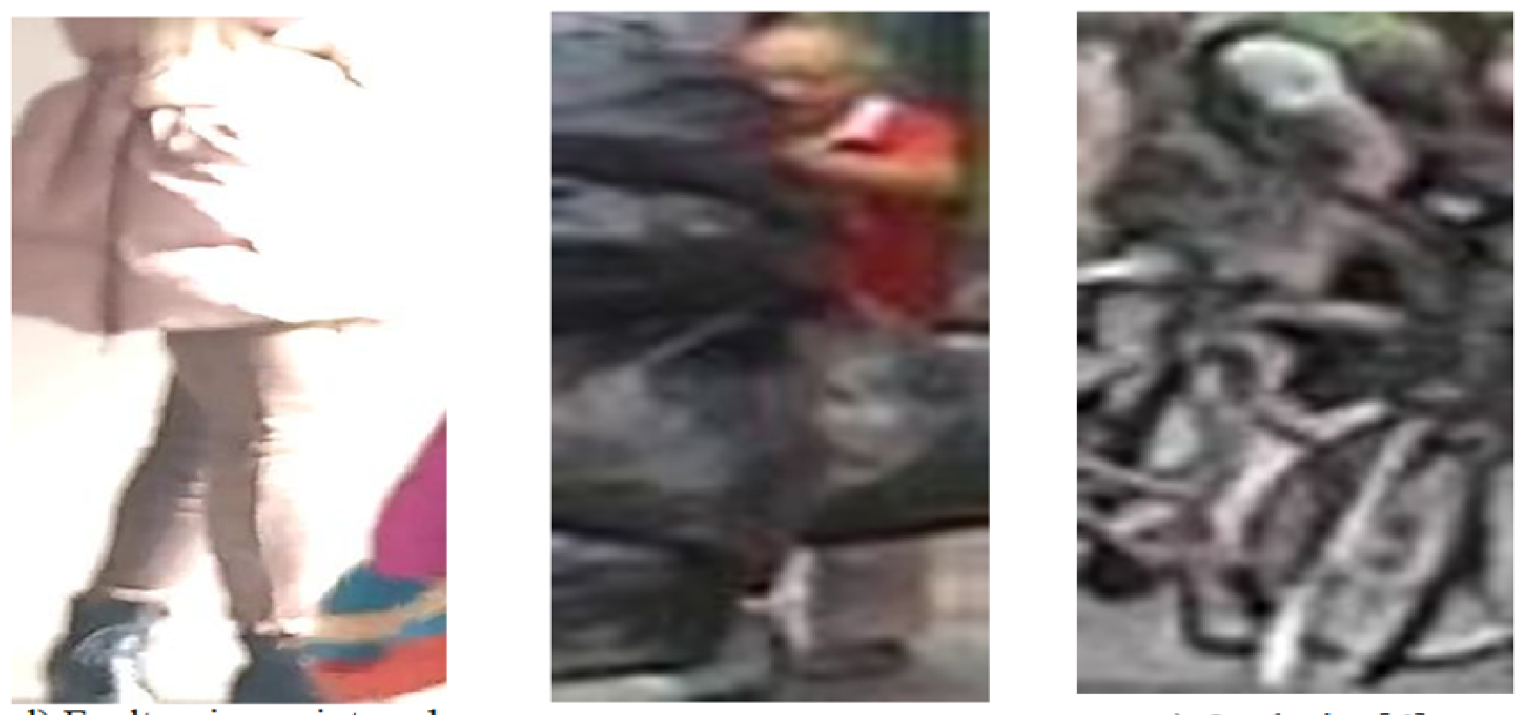
Challenges from CV Perspective

Three datasets have been used in this study are PA100K[1],PETA[3], MARKET1501[6].

a) Multiple person in frame with similar clothing [1]. Dimensions = 80 x 160

b) Distorted recording[1]. Dimensions = 69 x 225

c) Subject merging with background[3]. Dimensions = 92 x 226



d) Faulty viewpoint and excessive illumination[3]. Dimensions = 194 x 492

e) Partial view[6]. Dimensions = 64 x 128

c) Occlusion[6]. Dimensions: 64 x 128

Figure 1: Samples from the datasets depicting the various challenges.

Imbalance Attribute Distribution in Dataset.

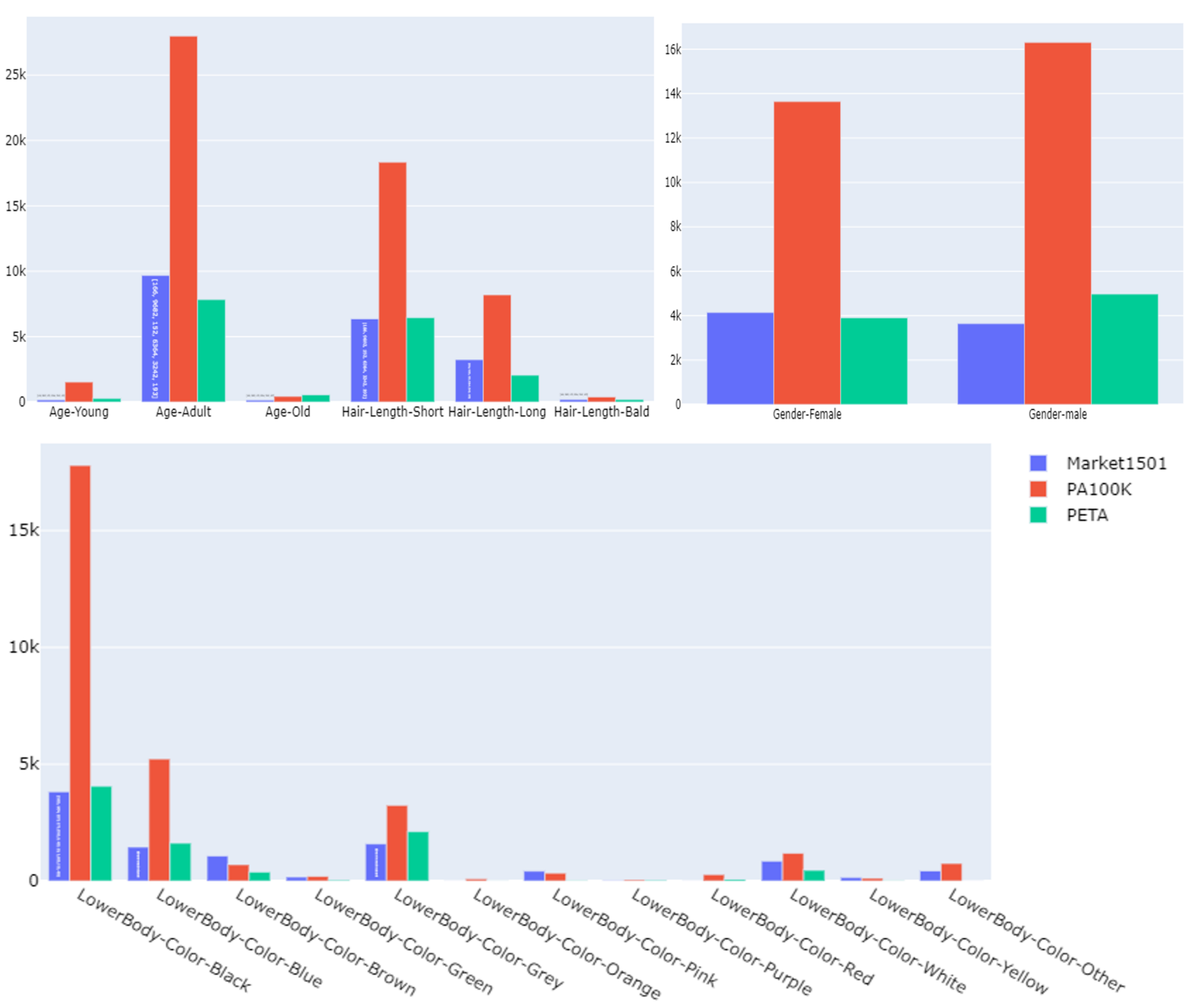
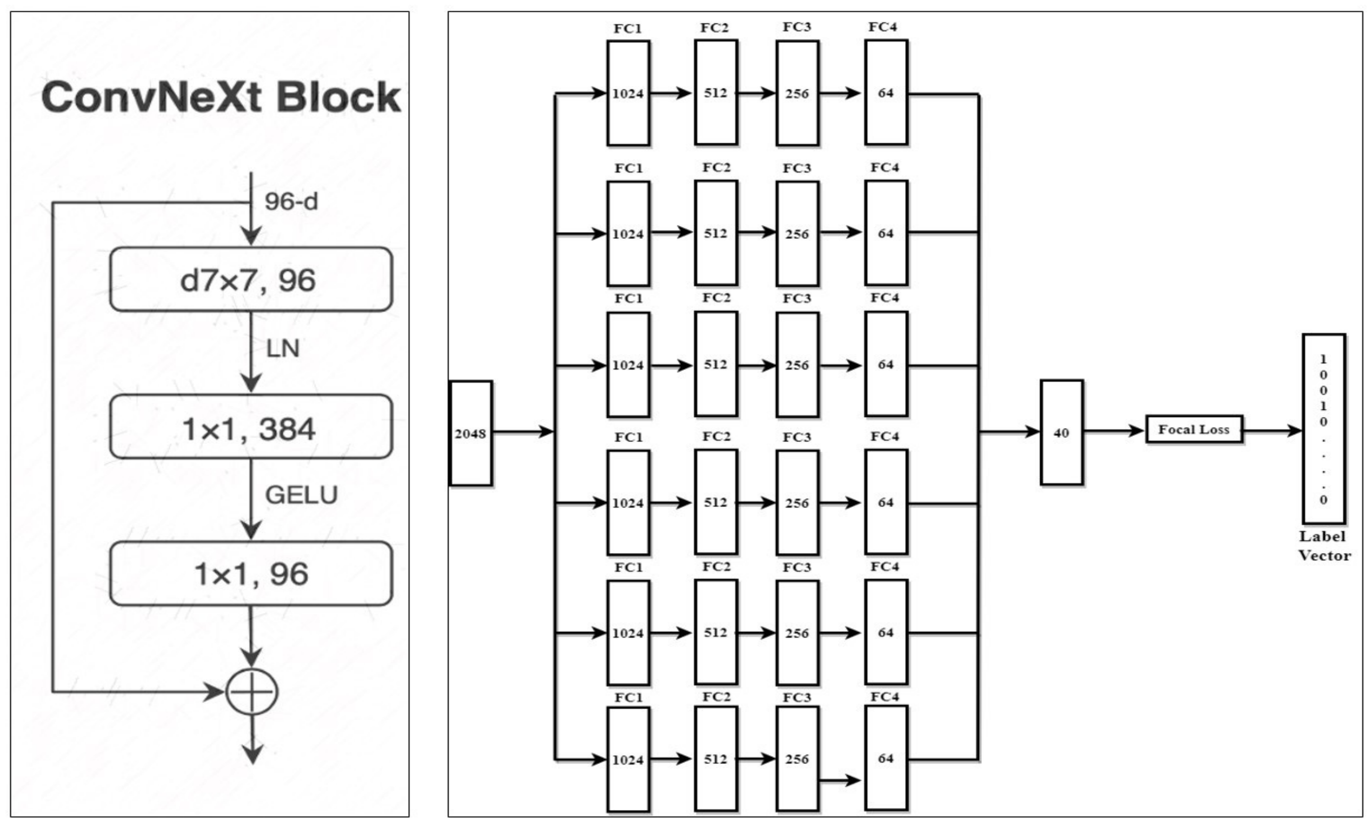


Figure 2: Attribute wise distribution.

Methodology



a) ConvNeXt Architecture [4].

b) PAR Architecture [2].

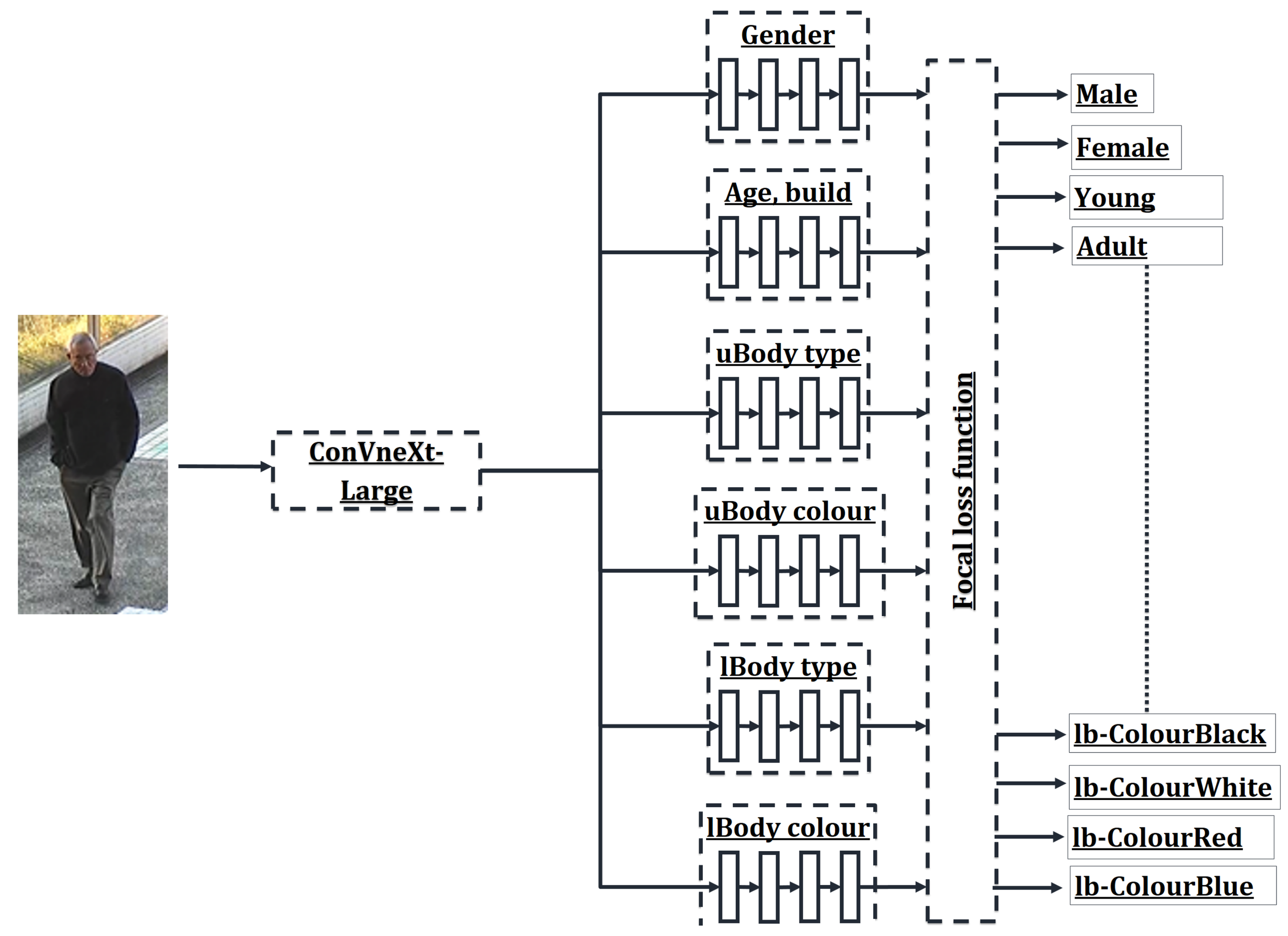


Figure 3: Network architecture for PAR model[2].

Results

(a) Results for average testing accuracy for all the attributes.

Dataset	Market1501	PA100K	PETA
Market1501	92.65	91.85	89.45
PA100K	93.45	94.64	92.5
PETA	87.83	89.52	92.03

(b) Results for average F1-Score of all the attributes.

Dataset	Market1501	PA100K	PETA
Market1501	79.42	82.8	67.74
PA100K	76.78	84.66	67.23
PETA	69.03	76.83	77.94

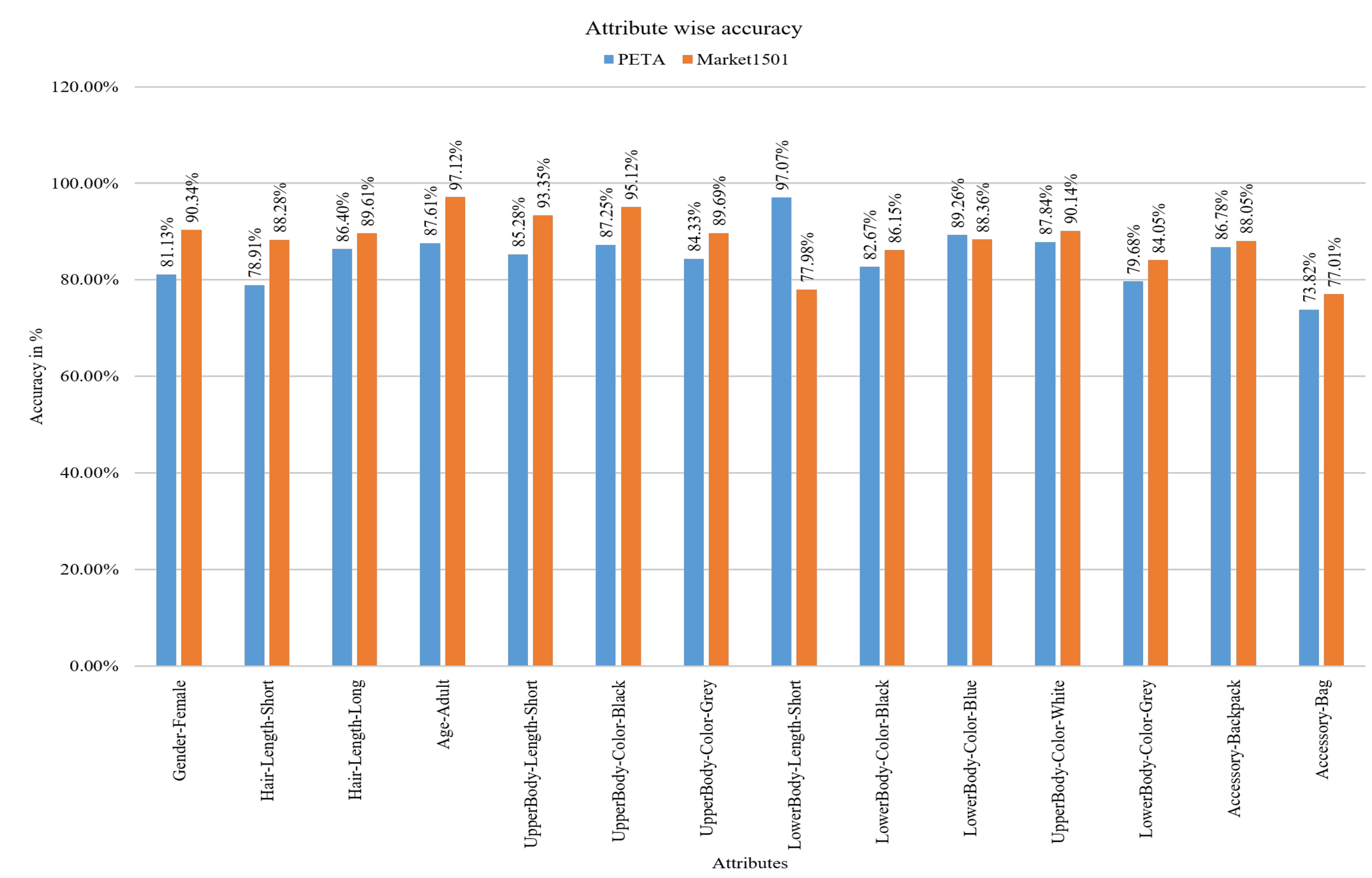


Figure 4: Attribute wise accuracy for network trained on PA100K and tested over PETA and Market1501.

Conclusion

- The imbalanced dataset can lead to biased training of the network, which in-turn favor the majority samples of the data.
- The use of focal loss allows the model to handle the data imbalance in the attribute distribution.
- With PA100K having more diverse attributes compared to PETA and Market1501, the model generalizes well.
- The results as shown in Table 1a and Table 1b show that model trained on PA100K has a best F1-score of **82.80** followed by Market1501 and PETA.

Acknowledgement

This research was supported by Gujarat Council on Science and Technology(GUJCOST)[Reference No: GUJCOST/STI/2021-22/3858]. The authors also acknowledge the help from Dr. Hiren Galiyawala and Mr. Harsh Tripathi for providing feedback and support in developing this project.

References

[1] Yubin Deng, Ping Luo, Chen Change Loy, and Xiaoou Tang. Pedestrian attribute recognition at far distance. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 789–792, 2014.

[2] Hiren Galiyawala, Mehul S Raval, and Meet Patel. Person retrieval in surveillance videos using attribute recognition. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–13, 2022.

[3] Xihui Liu, Haiyu Zhao, Maoqing Tian, Lu Sheng, Jing Shao, Junjie Yan, and Xiaogang Wang. Hydraplus-net: Attentive deep features for pedestrian analysis. In *Proceedings of the IEEE international conference on computer vision*, pages 1–9, 2017.

[4] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[5] Andreas Specker, Mickael Cormier, and Jürgen Beyerer. Upar: Unified pedestrian attribute recognition and person retrieval. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2023.

[6] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE international conference on computer vision*, pages 1116–1124, 2015.