PEDESTRIAN ATTRIBUTE RECOGNITION BASED ON MTCNN WITH ONLINE BATCH WEIGHTED LOSS

Xingting He[1,2] Qiuyue Shi[1,2] Fei Su[1,2] Zhicheng Zhao[1,2] Bojin Zhuang[3]

[1]School of Information and Communication Engineering, [2]Beijing Key Laboratory of Network System and Network Culture, Beijing University of Posts and Telecommunications, Beijing, China,

[3]Ping An Technology (Shenzhen) Co., Ltd.

Introduction

Due to the large number and huge diversity of attributes, pedestrian attribute recognition in video surveillance scenarios is a challenging task in the field of computer vision. Different from most previous works which only focus on extremely imbalanced attribute distribution problem, a new grouping way of attributes based multi-task convolutional neural network (MTCNN) is put forward, which exploits the spatial correlations among attributes and guarantees some independence of each attribute as well. Meanwhile, we propose a novel online batch weighted loss to narrow the performance differences among attributes and boost the model to gain a higher average recognition accuracy. The whole network can be trained end to end, and experimental results on PETA and RAP datasets show that our method achieves significant performance, comparing with those state-of-the-art methods. ***Index Terms***----Pedestrian Attribute Recognition, Grouping, CNN, Multi-task Learning, Online Batch Weighted Loss

Main Contributions

- We develop a multi-task CNN for pedestrian attribute recognition, which involves our proposed new grouping way of all pedestrian attributes.
- We propose a novel online batch weighted loss, which enables attribute prediction networks converge faster and narrows the performance differences among attributes.
- Our model can be trained end-to-end, and we achieve superior performance on two benchmark datasets.

Network Architecture

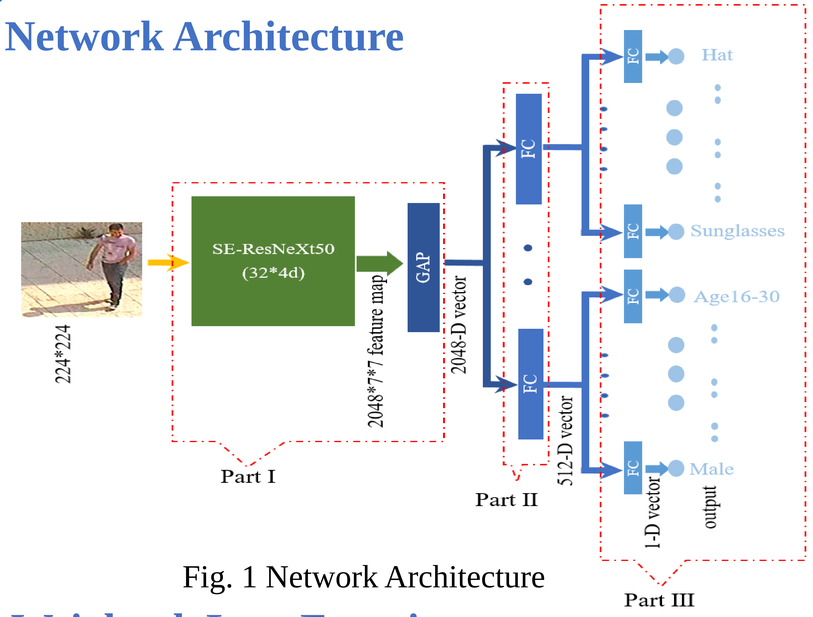

Fig. 1 Network Architecture

Weighted Loss Function

$$d(k) = \frac{1}{M}\sum_{i=1}^{M}|p_{ik} - y_{ik}| \quad (1)$$

$$f(k) = \begin{cases} 0, d(k) < th \\ d(k), d(k) > th \end{cases} \quad (2)$$

$$g(k) = \frac{e^{f(k)}}{\sum_{k=1}^{C} e^{f(k)}} \quad (3)$$

$$w(k) = e^{g(k)} \quad (4)$$

$$Loss_1 = \frac{1}{M}\sum_{i=1}^{M}\sum_{k=1}^{C} -[e^{1-\eta_k} y_{ik}\log p_{ik} + e^{\eta_k}(1-y_{ik})\log(1-p_{ik})] \quad (5)$$

$$Loss_2 = \frac{1}{M}\sum_{i=1}^{M}\sum_{k=1}^{C} -w(k)[y_{ik}\log(p_{ik}) + (1-y_{ik})\log(1-p_{ik})] \quad (6)$$

$$Loss_t = Loss_1 + \lambda Loss_2 \quad (7)$$

Experimental Results

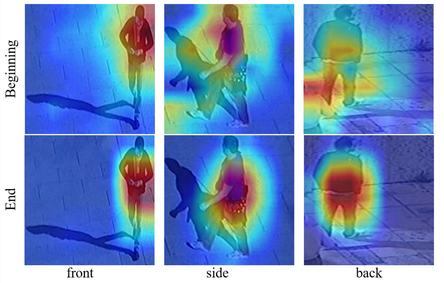

Fig. 2 Grad-CAM visualization results.

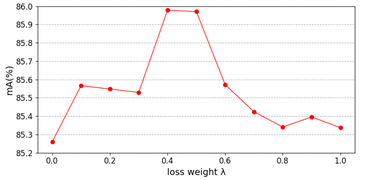

Fig.3 Evaluation on PETA validation set under different loss weight λ.

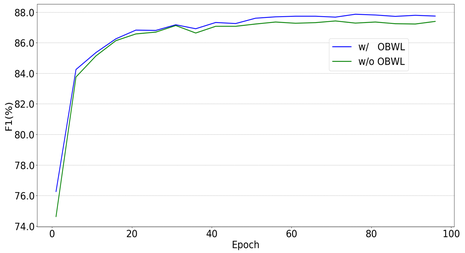

Fig.4: Evaluation on PETA validation set with and without OBWL.

Table 2: Ablation study to assess the impact of Grouping Scheme(GS), Online Batch Weighted Loss(OBWL) on PETA validation set

Methods		Metrics				
GS	OBWL	mA	Acc.	Pre.	Rec.	F1
	✓	85.39	80.68	88.16	87.27	87.71
✓		85.26	80.17	87.68	86.93	87.30
✓	✓	**85.98**	**80.90**	**88.21**	**87.46**	**87.83**

Table 1: Evaluation on PETA and RAP

Method	PETA					RAP				
	mA	Accuracy	Precision	Recall	F1	mA	Accuracy	Precision	Recall	F1
DeepMar [11]	82.89	75.07	83.68	83.14	83.41	73.79	62.02	74.92	76.21	75.56
VeSPA [12]	83.45	77.73	86.18	84.81	85.49	77.70	67.35	**79.51**	79.67	79.59
JRL [14]	85.67	-	86.03	85.34	85.42	77.81	-	78.11	78.98	78.58
VAA [13]	84.59	78.56	86.79	86.12	86.46	-	-	-	-	-
GRL [15]	**86.70**	-	84.34	**88.82**	86.51	81.20	-	77.70	80.90	79.29
Ours	85.73	**79.88**	**87.39**	86.79	**87.09**	**81.43**	**67.95**	78.46	**81.46**	**79.93**