中图法分类号：TP391　　**文献标识码**：A　　**文章编号**：1006-8961(2023)05-1487-12

论文引用格式：Ding Z Y，Shang Y F and Zhang C Y. 2023. Pedestrian attribute recognition method based on the progressive iterative optimization. Journal of Image and Graphics，28(05)：1487-1498(丁正彦，尚岩峰，张重阳. 2023. 渐进式迭代优化的行人属性识别. 中国图象图形学报，28(05)：1487-1498)[DOI：10.11834/jig.221064]

渐进式迭代优化的行人属性识别

丁正彦[1]，尚岩峰[1]，张重阳[2*]

1. 公安部第三研究所物联网技术研发中心，上海　201204；2. 上海交通大学电子信息与电气工程学院，上海　200240

摘　要：**目的**　现阶段行人属性识别任务存在的主要问题在于某些属性类别的样本分布严重不均衡，为了解决上述问题，提出了一种基于渐进式迭代优化的行人属性识别方法。**方法**　首先针对不均衡类别，采用马赛克自编码器进行数据增广，构建基于属性平衡化的数据生成模型(balanced attributes-data generation model，BA-DGM)，实现从通用大模型到专用小任务的迁移学习和知识增强；然后针对新生成的样本数据，采用判别模型进行一致性筛选，在与生成模型的相互对抗中实现启发式的注意力机制，从而构建基于特征注意力的数据判别模型(attention features-data discrimination model，AF-DDM)；最后通过数据生成与数据判别相互交替的循环迭代，实现行人属性识别模型和数据的渐进式优化，并针对均衡后的样本数据，采用知识蒸馏框架对不同轮次的判别模型进行融合，实现基于渐进式迭代的蒸馏融合模型(progressive iterations-distillation fusion model，PI-DFM)，进一步提高模型的泛化能力。**结果**　实验结果表明，所提出的渐进式优化方法在4个当前主流的评测数据集上均能有效提升模型准确率。在RAPv2(richly annotated pedestrian v2)数据集上，在模型复杂度不变的情况下，与已公开的最优行人属性识别模型相比，平均准确率(mean accuracy，mA)和平均$F1$分数分别提升了约5.0%和约1.7%；同时，经过多轮循环迭代后，原始数据中不均衡类别的个数减少为0，从而实现了数据集的渐进式优化。**结论**　本文提出的渐进式迭代优化策略与现有的改进方法之间具有良好的互补性，并有助于进一步提升模型的准确性指标。

关键词：行人属性识别；样本不均衡；渐进式迭代；马赛克自编码器；迁移学习；一致性筛选；知识蒸馏

Pedestrian attribute recognition method based on the progressive iterative optimization

Ding Zhengyan[1], Shang Yanfeng[1], Zhang Chongyang[2*]

1. Research Center on Internet of Things, the Third Research Institute of the Ministry of Public Security, Shanghai　201204, China;
2. School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai　200240, China

Abstract: **Objective** The pedestrian attribute recognition task is currently challenged for the sample distribution issue of some severe unbalanced attribute categories. To resolve the problems, we develop a method of progressive iteration optimization for pedestrian attribute recognition. **Method** First, data generation model based on masked autoencoder is used for data extension of the unbalanced categories distribution, and general large model-derived can be oriented to the small task. The balanced attributes-data generation model (BA-DGM) relevant masked autoencoder can be utilized to mask the original pedestrian images in terms of a random masking ratio and such newly generated images can be obtained for small-

收稿日期：2022-10-31；**修回日期**：2023-01-16；**预印本日期**：2023-01-23

*通信作者：张重阳　sunny_zhang@sjtu.edu.cn

基金项目：上海市青年科技英才扬帆计划资助项目(20YF1409300)

Supported by: Shanghai Sailing Program (20YF1409300)

amount categories. The potential information can be fully mined, such as the topological relationship of the visible area, and the latent features-derived pedestrian images can be more resilient. Furthermore, it demonstrates that the autoencoder model can effectively achieve the universal feature representation of the targeted pedestrian, including the consensus features like the relationship-interconnected between various key components of the pedestrian. Second, discrimination model is used for filtering-consistent for the newly generated sample data, and the heuristic attention mechanism is adopted and implemented to deal with generative adversarial networks (GANs). The newly attention features-data discrimination model (AF-DDM) can be utilized and the diversified sample can be achieved while the key features of the attributes are preserved, which can enhance the interpretability of the recognition model. At the same time, to learn effective features-related attributes, the filtered data is generated for training model. In the training process of the discrimination model, 50-layer residual network model is adopted as the backbone network to be trained on the original attribute recognition dataset, using a multi-label classification framework. And, in the reasoning process of the discrimination model, the whole attribute labels are divided into two categories: key attribute labels and other related attribute labels. For key attribute labels, to keep consistent with the original labels and preserve the relevant high confidence, the newly generated sample can be kept in consistency in terms of the predicted labels from discrimination model, but it cannot be vice versed. Finally, the pedestrian attribute recognition model and data-contextual can be optimized further based on the cyclical iteration of data generation and discrimination. To optimize generalization ability of the model, the knowledge distillation framework can be used to fuse the discrimination models of the balanced sample data as well. After multiple iterations, the progressive iterations-distillation fusion model (PI-DFM) based attribute discrimination models can be used as the teacher models and category balancing-afterward attribute recognition dataset is used as the training data. The above models are mutual-benefited in accordance with the datasets of different sample proportions. The network structure of the student model is consistent with the teacher model and the Kullback-Leibler (KL) divergence between the student output and the teacher output is calculated as the distillation loss function. In large-scale practical application scenarios, the sample proportion of test data and train data might be different. To improve the generalization ability of the model in an open uncertain scenario, teacher model can be trained by integrating different sample-proportion data in terms of the knowledge distillation framework. **Result** Experimental results are demonstrated that the proposed optimization method can effectively improve the accuracy of the model on the four popular evaluation datasets. The proposed metrics for attributes and samples are calculated, including 1) the mean accuracy of all attributes and 2) the $F1$ score of all samples, representing the harmonic average of the mean accuracy and the mean recall. For example, in the richly annotated pedestrian v2 (RAPv2) dataset, the mean accuracy is increased by about 5.0% and the average $F1$ score is increased by about 1.7% as well on the hypothesis of an unchanged model complexity. After several loops of cyclic iteration, the number of unbalanced categories in the original data is reduced to zero, and the optimization can be thus realized for the dataset. In the ablation studies, new samples are randomly generated for each positive sample image, and then the discrimination model is used to filter inconsistent samples. The probability of spatial distribution of the preserved details is analyzed experimentally in terms of the masked region analysis of the filtered samples. The heuristic attention mechanism is introduced and data discrimination model can retain the relevant features of the key attributes of the targeted pedestrian better, which demonstrates that the interpretability of the discrimination model can be further improved by deeply mining the distribution of related features for different attributes. **Conclusion** The progressive iterative optimization strategy proposed in this paper has good complementarity with the existing improvement methods, and is helpful to further improve the accuracy of the recognition model. To optimize the relationship modeling among multiple pedestrian attributes and improve the interpretability of the recognition model further, future research direction can be predicted and focused on universal feature representation-based masked autoencoder (MAE) model combined with such prior knowledge like human skeleton structure.

Key words: pedestrian attribute recognition; unbalanced sample; progressive iteration; masked autoencoder; transfer learning; consistency filtering; knowledge distillation

0　引　言

行人属性识别任务一直以来都是智能视频分析领域的研究热点。该任务在行人检测(罗艳 等，2022)的基础上，通过识别行人目标的视觉属性，包括性别、年龄、衣着以及携带物等语义内容，为目标结构化描述与快速检索提供支撑。

随着深度学习技术的不断发展，尤其是卷积神经网络模型在图像分类领域的广泛应用，研究人员提出了许多基于深度网络模型的行人属性识别优化方法。Li 等人(2015)引入多属性联合学习机制，通过单个骨干网络模型得到多属性共享的全局特征，从而有效挖掘属性之间的相互关系。Liu 等人(2017)引入视觉注意机制，采用多方向注意力模块对骨干网络模型的不同语义层提取注意力特征图，并借助多尺度融合模块实现特征优化。Tang 等人(2019)引入属性相关的弱监督定位机制，结合高层语义的指导信息，实现特定属性关注区域的多尺度自适应感知。

另一方面，研究人员通过采集不同场景下的行人目标数据，并标注相关属性信息，构建了多个面向行人属性识别任务的大规模公开数据集，例如 PA100K(pedestrian attribute 100K)(Liu 等，2017)、PETA(pedestrian attribute)(Deng 等，2014)、RAPv1(richly annotated pedestrian v1)(Li 等，2016)和 RAPv2(richly annotated pedestrian v2)(Li 等，2019)，涵盖各种全局目标属性和局部目标属性，其中全局属性主要包括性别、年龄等，局部属性主要包括头部属性、上身属性和下身属性等。

但是，现阶段行人属性识别任务存在的主要问题在于某些属性类别的样本分布严重不均衡，以 RAPv2 数据集为例，属性类别总数为 54 个，其中正样本所占比例低于 10% 的属性类别达到 34 个，占比超过 60%(34/54)，如图 1 所示(黄色标注)。

针对样本分布不均衡的问题，研究人员通常重点关注其中实例数较少的类别(记做少数类)，在模型训练过程中引入了多种优化策略。1)数据增广策略。通过对少数类进行数据混合(Zhang 等，2018)或数据噪声(Zhong 等，2020)等操作，随机生成新训练样本，从而增加相关类别的数量比例，该方法在目标实例分割等任务中也广泛使用(Ghiasi 等，2021)；

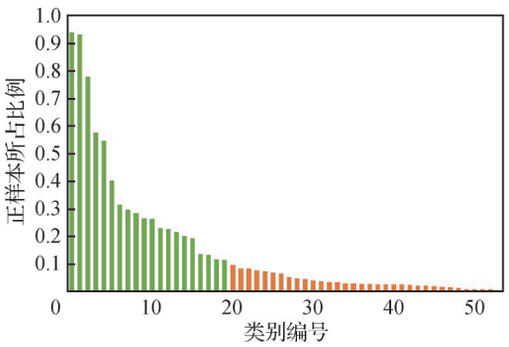


图 1　RAPv2 数据集的各个属性类别正样本比例统计情况

Fig. 1　Positive sample ratio for each attribute category of RAPv2 dataset

2)损失加权策略。在损失函数的设计过程中增加类别自适应的权重，从而选择性地提高少数类的样本重要程度(Jia 等，2020)；3)迁移学习策略，在样本均衡的相关数据集上进行模型预训练，并通过特征迁移实现少数类样本数据上的模型泛化(Liu 等，2019)。

虽然上述方法能够缓解样本分布不均衡问题，但是如何根据具体任务特点进行多种优化策略的互补融合仍然有待进一步研究。因此，本文针对行人多属性识别任务，提出一种基于渐进式迭代的优化方法，深入融合上述 3 种优化策略，具体改进如下：1)数据生成。针对行人目标中的少数类属性，采用马赛克自编码器(masked autoencoder，MAE)(He 等，2022)得到新生成的相关样本数据，通过提高少数类的正样本比例，构建基于属性平衡化的数据生成模型(balanced attributes-data generation model，BA-DGM)，同时实现从 MAE 通用大模型到专用小任务的迁移学习和知识增强。2)数据判别。针对新生成的行人目标数据，采用基于原有数据训练的属性判别模型进行数据筛选，根据属性预测的分布一致性剔除异常数据，通过启发式的注意力机制，构建基于特征注意力的数据判别模型(attention features-data discrimination model，AF-DDM)，其中判别模型在训练过程中需要根据样本比例自适应调整损失函数的权重，保证数据判别的有效性。3)渐进式迭代。采用数据生成与数据判别相互交替的循环迭代框架，在不增加数据标注代价的情况下，实现行人多属性识别数据的渐进式优化，同时针对均衡后的数据集，通过知识蒸馏机制融合多个不同轮次的数据判别模型，如图 2 所示，构建基于渐进式迭代的蒸馏融合模型(progressive iterations-distillation fusion model，PI-

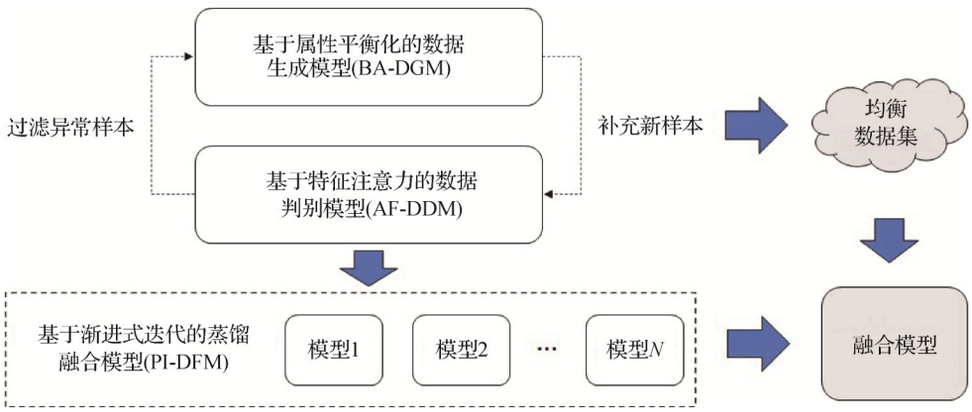


图2　基于渐进式迭代的行人属性识别优化框架

Fig. 2　Optimization framework based on progressive iteration for pedestrian attributes recognition

DFM),在不增加模型复杂度的情况下进一步提升行人属性识别精度。

在实际应用过程中,本文所采用的基于渐进式迭代的行人属性识别优化方法能够通过无监督学习机制,在海量无标注的行人属性识别数据集上进行自适应的数据均衡与模型融合,从而有效提升模型的泛化能力。

1　基于属性平衡化的数据生成

针对行人属性识别任务中存在的样本不均衡问题,现有的数据增广策略主要通过数据混合和数据噪声等操作实现。如图3所示,以常用的随机擦除算法为例,随机掩盖部分图像块进行模拟(掩盖比例为0.3),可以发现随机擦除后的样本数据与真实场景数据存在明显差异,并且可能丢失某些局部属性相关的关键特征信息,导致训练稳定性下降。

本文提出引入马赛克自编码器进行图像复原,经过复原后的目标图像与真实场景的目标图像无明显差异(图3),可以作为少数类样本数据的补充,从而实现基于属性平衡化的数据生成。在具体实现过程中,采用一种改进的MAE算法,如图4所示。首先按照一定比例随机掩盖原始目标中的部分图像块,然后通过非对称的编码器与解码器结构进行复原操作,其中编码器模块采用基于Transformer结构的深度网络模型(Dosovitskiy等,2021)进行特征编码,而解码器模块则采用轻量级模型,MAE模型预先在海量无标注数据集上进行自监督训练,因此能够很好地实现行人目标的通用特征表示。另一方面,为了进一步优化模型对于行人目标属性特征的表示能力,本文采用大规模行人属性识别数据对MAE模型进行微调,并引入多个典型的行人属性类别标签作为监督信息,例如性别、年龄和衣着等。

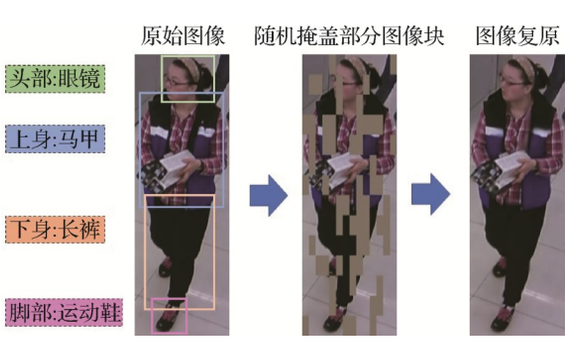


图3　随机掩盖部分图像块与复原结果对比

Fig. 3　Comparison between random masks and restoration results

1.1　BA-DGM模型训练

在生成模型的训练过程中,本文采用原始MAE模型作为预训练模型,通过增加行人属性识别相关的监督约束(图4),从而进一步强化模型对于属性相关的常识信息提取,训练阶段马赛克区域所占比例(masking ratio, MR)的随机取值范围记为[MR_{low}, MR_{high}],在保留关键特征信息的情况下增加生成数据的多样性。

1.2　BA-DGM模型推理

在生成模型的推理过程中,本文在[MR_{low}, MR_{high}]范围内随机选择MR值生成马赛克图像并进行复原操作,具体为

$$\boldsymbol{I}_{\text{gen}} = D_{\boldsymbol{\phi}}\left(E_{\boldsymbol{\theta}}\left(M \odot \boldsymbol{I}_{\text{ori}}\right)\right) \tag{1}$$

式中,E和D分别表示图像编码器和解码器,$\boldsymbol{\theta}$和$\boldsymbol{\phi}$分别为编码器和解码器的模型参数,$M\odot$为马赛克

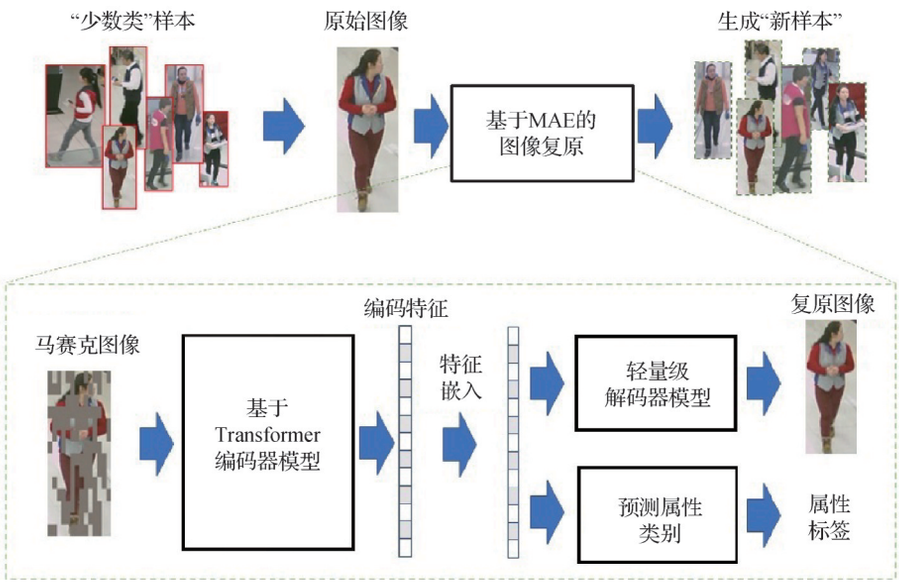


图4　针对少数类采用改进的MAE算法进行数据生成

Fig. 4　Data generation using an improved MAE algorithm for small-amount categories

操作，$\boldsymbol{I}_{\text{ori}}$和$\boldsymbol{I}_{\text{gen}}$分别为原始图像和生成图像。

1.3　少数类数据增广

本文采用上述数据生成策略，针对属性识别任务中存在的少数类数据，通过MAE模型生成新样本，即每个原始图像得到N幅生成图像，N的取值为

$$N = \max\{N_l\}, l \in [1, num]$$

$$N_l = \begin{cases} \lambda \times round\left(\exp\left(1 - p_l\right) - 1\right) & y_l = 1 \\ \lambda \times round\left(\exp\left(p_l\right) - 1\right) & y_l = 0 \end{cases} \tag{2}$$

式中，num为属性个数，N_l为第l个属性对应的生成图像个数，p_l为第l个属性的训练集正样本比例，y_l为原始图像的第l个属性标签，λ为数据增广系数，$round(\cdot)$为四舍五入操作。在行人多属性识别任务中，每个数据样本对应num个属性标签$(y_1, y_2, \cdots, y_{\text{num}})$。

1.4　从通用大模型迁移到专用小任务

在数据生成过程中，MR值的大小将对目标图像的复原结果产生影响。如图5所示，随着MR值的不断增加，所生成目标图像的细节特征逐渐减少。从对比结果可以看出，当MR大于0.9时，目标马赛克图像的可见部分仅为零散分布的少数局部区域，但是通过MAE模型仍然能够充分挖掘可见区域的位置关系等深层信息，从潜在特征中恢复出行人的基本样貌图像，同时证明了本文所采用的自编码器模型有效实现了行人目标的通用特征表示，其中包括各个关键部件之间的位置关系等常识特征。

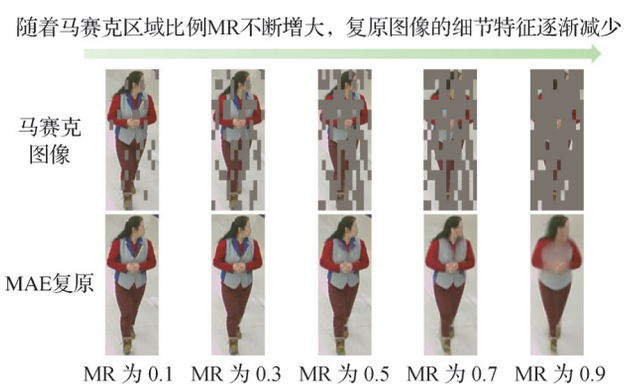


图5　不同MR值的数据生成结果对比

Fig. 5　Comparison of data generation results for different MR

通过数据生成，本文实现了从通用大模型（MAE自编码器）到专用小任务（行人属性识别）的迁移学习和知识增强，如图6所示，其中MAE模型采用自监督学习机制，从大规模无标注数据中自主获取目标的通用特征表示，包含行人目标姿态等先验知识，即数据背后的目标常识特征，有助于提升行人目标属性识别的准确性和泛化性。

2　基于特征注意力的数据判别

针对生成模型得到的新样本数据，由于存在某些细节特征丢失，可能导致相关属性类别发生变化。本文采用基于多标签分类框架的判别模型进行数据筛选，如图7所示，结合原始图像的属性标签信息，

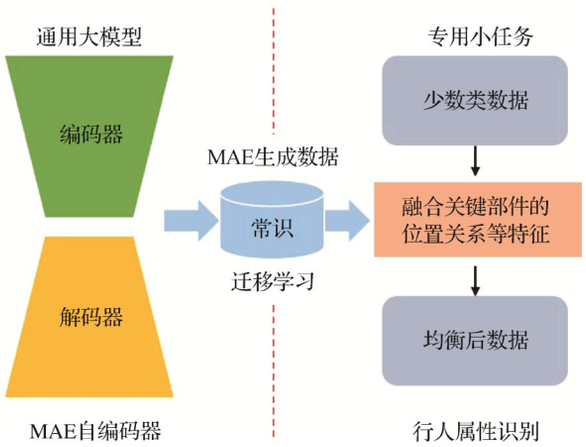


图6　从通用大模型到专用小任务的迁移学习

Fig. 6　Transfer learning from general big models to dedicated small tasks

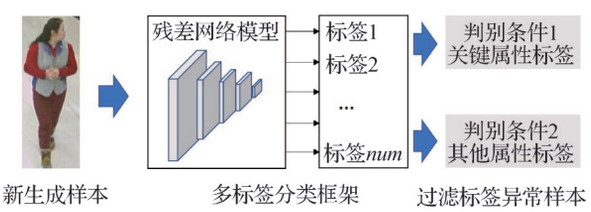


图7　针对新生成样本采用多标签分类框架进行数据判别

Fig. 7　Multi-label classification framework used for data discrimination of newly generated samples

根据预测分布的一致性，从中过滤标签异常的新生成样本。筛选后的数据能够保留与属性相关的关键特征，通过启发式的注意力机制，实现基于特征注意力的数据判别。

2.1　AF-DDM模型训练

在判别模型的训练过程中，本文基于多标签分类框架，采用50层的残差网络模型作为骨干网络，在原始的属性识别数据集上进行训练。在训练过程中采用类别自适应的权重（Jia等，2020），即根据样本比例自适应调整损失函数的权重，从而缓解原始数据不均衡问题对模型精度的影响。

2.2　AF-DDM模型推理

在判别模型的推理过程（图7）中，本文针对新生成样本进行数据筛选，首先将判别模型输出的多个属性标签分为关键属性标签和其他属性标签两类，分别采用判别条件1和判别条件2进行筛选，其中N_l取值参见式（2）。

判别条件1：对于关键属性标签（$N_l > 0$），若判别模型预测的标签与原始标签一致，且输出的相应置信度值高于阈值τ，则满足判别条件，否则不满足；

判别条件2：对于其他属性标签（$N_l = 0$），若判别模型输出的相应置信度值高于阈值τ，则满足判别条件，否则不满足。

根据上述判别条件对判别模型预测的所有属性标签进行逐一筛选，最后选择每个属性预测结果均满足一致性的新生成样本。

2.3　对抗学习与注意力机制

本文采用上述数据判别策略，针对新生成样本进行数据筛选，如图8所示，当新生成样本不满足一致性判别条件时，则丢弃该样本，并由生成模型随机得到新样本进行补充。通过生成模型与判别模型的相互对抗机制，使得新生成样本能够在保留属性关键特征的情况下实现样本多样化，同时筛选后的生成数据训练得到的模型将更加易于学习属性相关特征，从而引入启发式的模型注意力机制，进一步提升模型的可解释性与泛化能力。

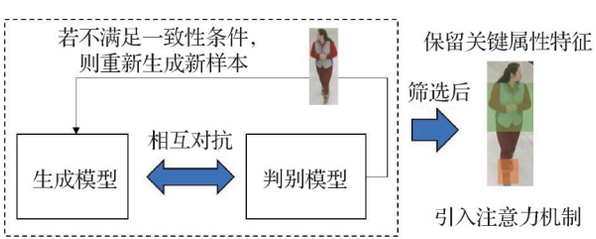


图8　通过判别模型引入对抗学习与注意力机制

Fig. 8　Introducing adversarial learning and attention mechanisms through discriminative model

3　基于渐进式迭代的蒸馏融合

为了更好地解决行人属性分类数据的样本不均衡问题，本文在数据生成与数据筛选相互结合的基础上提出渐进式迭代机制，通过两者相互交替的循环迭代，逐步优化得到类别均衡数据集，并通过知识蒸馏框架对不同轮次的判别模型进行有效融合，提升属性识别模型对多种应用场景的适应能力。

3.1　数据生成与数据判别的循环迭代

本文采用循环迭代算法，具体步骤如下：

输入：原始属性识别数据集$\boldsymbol{D}_{\text{ori}}$，迭代轮数$T$。

输出：类别均衡后的数据集$\boldsymbol{D}_{\text{new}}$，不同轮次训练得到的属性判别模型$\{\boldsymbol{M}_1, \cdots, \boldsymbol{M}_T\}$。

1）$\boldsymbol{D}_0 = \boldsymbol{D}_{\text{ori}}$；

2)for t = 1 to T do;

3) 通过生成模型扩充 $\boldsymbol{D}_{t-1}$ 的少数类，得到 $\boldsymbol{G}_t$。根据式(2)每个样本随机生成 N 个新样本；

4) 基于 $\boldsymbol{D}_{t-1}$ 训练得到属性判别模型 M_t；

5) 通过判别模型 M_t 筛选 $\boldsymbol{G}_t$，得到 $\boldsymbol{G}_t^*$。若新样本不满足判别条件1和2，则重新生成；

6) 更新当前数据集：$\boldsymbol{D}_t = \boldsymbol{D}_{t-1} \bigcup \boldsymbol{G}_t^*$；

7)return $\boldsymbol{D}_{new} = \boldsymbol{D}_T$。

3.2　基于知识蒸馏框架实现模型融合

本文采用上述循环迭代策略，经过 T 轮迭代后，得到多个属性判别模型 $\{M_1, \cdots, M_T\}$，由于采用不同样本比例的数据集训练，模型之间具有较好的互补性。因此，本文实现了基于渐进式迭代的蒸馏融合模型 PI-DFM（progressive iterations-distillation fusion model），如图9所示，以不同迭代轮次的属性判别模型作为教师模型，并且采用类别均衡后的属性识别数据集 $\boldsymbol{D}_{new}$ 作为训练数据。

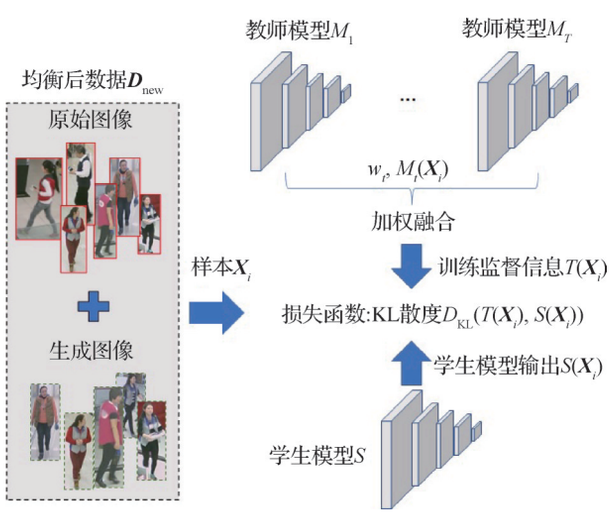


图9　基于渐进式迭代实现蒸馏融合模型

Fig. 9　Achieving distillation fusion model based on progressive iteration

对于每个训练样本 $\boldsymbol{X}_i$，首先通过对不同教师模型的输出结果 $M_t(\boldsymbol{X}_i)$ 进行加权融合，得到训练监督信息 $T(\boldsymbol{X}_i)$，具体为

$$T(\boldsymbol{X}_i) = \frac{\sum_{t=1}^{T}\left(w_t \times M_t(\boldsymbol{X}_i)\right)}{\sum_{t=1}^{T} w_t} \tag{3}$$

式中，w_t 为教师模型 M_t 所对应的权重，考虑到数据集的类别均衡性随着迭代轮次增加逐步优化，教师模型的准确率也同步实现渐进式提升，因此 w_t 的取值为

$$w_t = \exp(t - T),\ t \in [1, T] \tag{4}$$

然后，针对该样本预测得到学生模型的输出结果 $S(\boldsymbol{X}_i)$，其网络结构与教师模型一致，最后通过计算 $S(\boldsymbol{X}_i)$ 和 $T(\boldsymbol{X}_i)$ 之间的 KL(Kullback-Leibler)散度(Bagherinezhad 等，2018)作为蒸馏损失函数，具体为

$$Distill_Loss = D_{KL}\left(S(\boldsymbol{X}_i), T(\boldsymbol{X}_i)\right) \tag{5}$$

在大规模实际应用场景下，测试数据与已有训练数据的样本比例可能存在差异，本文通过知识蒸馏框架，融合不同样本比例数据训练得到的教师模型，能够有效提升模型在开放不确定场景的泛化能力。

4　实验结果分析

4.1　评测数据集及实验参数

实验在目前主流的4种行人属性识别评测数据集上进行。相关数据集的具体参数情况与常用评测方法(Jia等，2020)一致，如表1所示。

表1　行人属性识别评测数据集的参数情况

Table 1　Parameter configuration of benchmark datasets for pedestrian attributes recognition

数据集名称	属性/个	训练样本/幅	测试样本/幅
PA100K(Liu等，2017)	26	90 000	10 000
PETA(Deng等，2014)	35	11 400	7 600
RAPv1(Li等，2016)	51	33 268	8 317
RAPv2(Li等，2019)	54	67 943	16 985

实验中，采用装有4张 Nvidia Tesla A100(单张显卡的显存大小为40 GB)的GPU服务器，并基于Pytorch平台进行模型训练和推理。

1)生成模型。在模型训练阶段，参考MAE模型的训练超参数(He等，2022)，并增加属性预测机制，如图4所示，在固定编码器模型参数的情况下微调其他模块，进一步优化嵌入特征表示，同时马赛克区域所占比例MR的取值范围为[0.3, 0.7]；在模型推理阶段，数据增广系数 λ 的取值为2。

2)判别模型。在模型训练阶段，与常用评测方法(Jia 等，2020)的模型训练超参数保持一致，以

50层的残差网络模型作为骨干网络，输入图像尺度归一化为$H \times W = 256 \times 192$；在模型推理阶段，数据判别阈值$\tau$的取值为0.7。

3）渐进式迭代。迭代轮数T取值为3，通过生成模型与判别模型之间的T次循环迭代优化行人属性识别精度。在迭代过程中，判别模型的训练周期数根据生成数据的增加等比例减少。

在整个训练和推理过程中，本文没有对上述超参数进行调整，即针对不同数据集均采用一致的超参数。

4.2 测试结果对比

实验采用常用评测方法（Jia等，2020）使用的行人属性识别评测指标。针对不同属性，统计全部属性的平均准确率（mean accuracy，mA）；针对不同样本，统计全部样本的平均精度（precision，Prec）、平均召回率（Recall）和平均$F1$分数（准确率和召回率的调和平均，记做$F1$）。

在测试过程中，本文采用50层的残差网络模型作为基准模型（baseline），并提出基于渐进式迭代的优化方法，与多种目前主流的行人属性识别方法DeepMAR（deep multi attribute recognition）（Li等，2015）、VAC（visual attention consistency）（Guo等，2019）和RPAR（rethinking pedestrian attribute recognition）（Jia等，2020）分别在PA100K、PETA、RAPv1和RAPv2数据集上进行对比，对比算法均采用与行人属性识别任务常用评测方法（Jia等，2020）一致的参数配置，实验结果如表2—表5所示。可以看出，本文方法的mA和$F1$两个指标均有明显提升，表明本文方法能够有效优化行人属性识别结果。在不增加模型复杂度的情况下，mA和$F1$两项指标均优于目前最优的行人属性识别算法RPAR。

表2 不同算法在PA100K数据集上的识别结果

Table 2 Test results on PA100K dataset for different algorithms

/%

算法	骨干网络	mA	Prec	Recall	$F1$
DeepMAR	CaffeNet	72.70	82.24	80.42	81.32
RPAR	ResNet50	79.38	89.41	84.78	86.55
Baseline	ResNet50	79.00	89.17	84.71	86.88
本文	ResNet50	**81.61**	88.34	86.97	**87.65**

注：加粗字体表示各列最优结果。

表3 不同算法在PETA数据集上的识别结果

Table 3 Test results on PETA dataset for different algorithms

/%

算法	骨干网络	mA	Prec	Recall	$F1$
DeepMAR	CaffeNet	82.89	83.68	83.14	83.41
VAC	ResNet50	83.63	87.63	85.45	86.23
RPAR	ResNet50	85.11	86.99	86.33	86.39
Baseline	ResNet50	83.10	90.72	82.40	86.36
本文	ResNet50	**85.27**	88.65	85.95	**87.28**

注：加粗字体表示各列最优结果。

表4 不同算法在RAPv1数据集上的识别结果

Table 4 Test results on RAPv1 dataset for different algorithms

/%

算法	骨干网络	mA	Prec	Recall	$F1$
DeepMAR	CaffeNet	73.79	74.92	76.21	75.56
RPAR	ResNet50	78.48	82.84	76.25	78.94
Baseline	ResNet50	77.26	83.08	71.78	77.02
本文	ResNet50	**80.00**	81.24	79.26	**80.24**

注：加粗字体表示各列最优结果。

表5 不同算法在RAPv2数据集上的识别结果

Table 5 Test results on RAPv2 dataset for different algorithms

/%

算法	骨干网络	mA	Prec	Recall	$F1$
VAC	ResNet50	76.74	80.42	78.78	79.24
RPAR	ResNet50	78.28	77.96	79.38	78.30
Baseline	ResNet50	76.83	81.50	77.98	79.70
本文	ResNet50	**83.28**	77.75	83.35	**80.45**

注：加粗字体表示各列最优结果。

实验进一步将本文算法与目前主流的引入额外模型复杂度的行人属性识别算法CAS-SAL-FR（cascaded split-and-aggregate learning with feature recombination）（Yang等，2021）和DBSAN（dual- branch self-attention network）（Liu等，2022）在PETA数据集上进行对比，并比较了模型复杂度情况，结果如表6所示。可以看出，本文所提出的算法在模型复杂度最小的情况下得到了较好模型准确率指标，其中$F1$指标在3种算法中结果最优。

表 6　3 种算法在 PETA 数据集上的识别结果

Table 6　Test results of three algorithms on PETA

算法	参数量/M	FLOPs/G	mA/%	*F*1/%
CAS-SAL-FR	35.20	5.58	86.40	87.18
DBSAN	36.63	5.03	**87.70**	86.77
本文	**23.56**	**4.08**	85.27	**87.28**

注：加粗字体表示各列最优结果。

本文提出的迭代优化方法是从属性类别均衡性的角度进行优化，与目前主流的行人属性识别方法具有较好的互补性。本文以 RAPv2 数据集为例，针对 34 个不均衡类别（正样本所占比例低于 10%）的属性识别结果进行统计，如图 10 所示，与基准模型结果相比，通过渐进式迭代优化后的属性识别模型能够有效缓解类别不均衡的问题，其中每个类别的平均准确率均有明显提升。

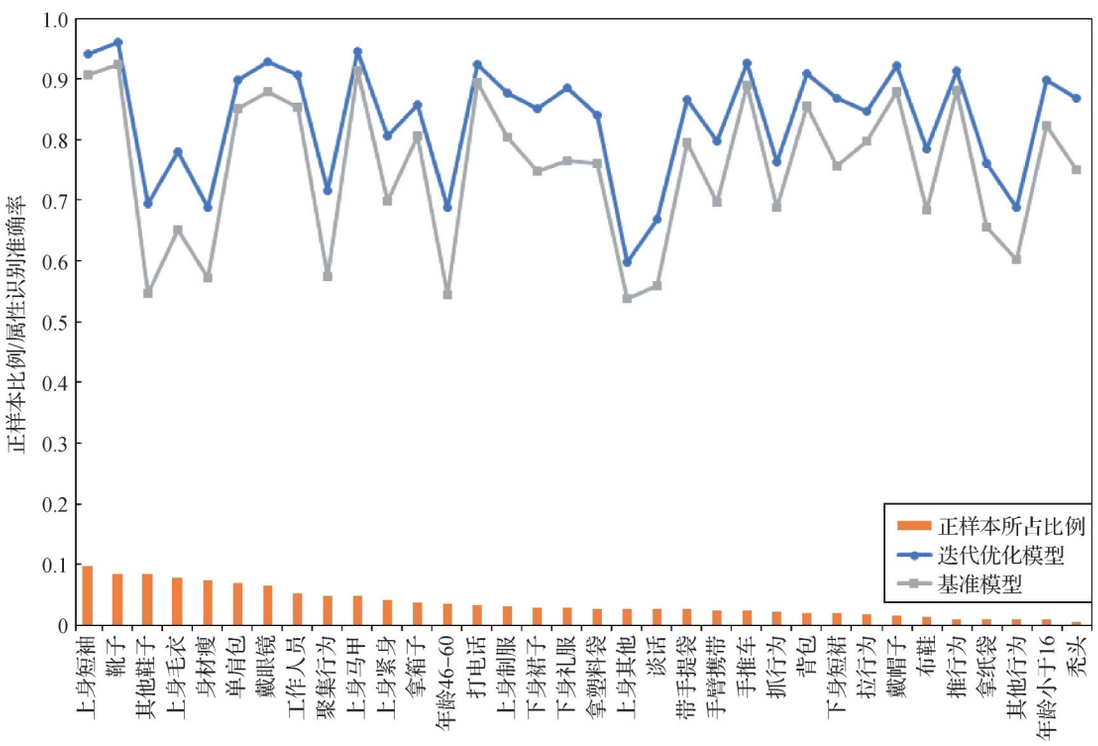


图 10　渐进式迭代优化模型与基准模型结果对比

Fig. 10　Results of progressive iterative optimization model and baseline model

4.3　消融实验

4.3.1　关键子模块选取

本文提出的渐进式迭代算法主要包括 BA-DGM 数据生成模块（记做模块 1）、AF-DDM 数据判别模块（记做模块 2）和 PI-DFM 蒸馏融合模块（记做模块 3）。为了进一步验证各子模块的有效性，本文分别采用第 1 轮数据生成后得到的数据（$\boldsymbol{D}_0 + \boldsymbol{G}_1$）和第 1 轮数据判别后得到的数据（$\boldsymbol{D}_0 + \boldsymbol{G}_1^*$）在 RAPv2 数据集上进行实验，并与经过 T 轮迭代和蒸馏融合后的最终模型对比，结果如表 7 所示，其中平均值表示 mA 和 $F1$ 指标的平均，反映了针对不同属性和针对不同样本评测指标的综合评价（参见 4.2 节）。实验结果表明，3 个主要模块对于属性识别的准确率均有提升作用。

表 7　3 个改进模块在 RAPv2 数据集上的识别结果

Table 7　Test results of three improvements on RAPv2

/%

算法	数据	mA	*F*1	平均值
基准模型	$\boldsymbol{D}_0$	76.83	79.70	78.27
增加模块 1	$\boldsymbol{D}_0 + \boldsymbol{G}_1$	80.03	79.53	79.78
增加模块 1+2	$\boldsymbol{D}_0 + \boldsymbol{G}_1^*$	80.11	79.85	79.98
增加模块 1+2+3	$\boldsymbol{D}_T$	**83.28**	**80.45**	**81.87**

注：加粗字体表示各列最优结果。

4.3.2　数据判别阈值 τ 的参数分析

数据判别阈值的选择需要考虑多个属性的平均准确率和召回率情况，实验分别针对 4 个训练集（具体情况参见表 1）进行验证，从中随机抽取 1/5 的样

本数据作为验证集，由其余4/5的样本数据训练得到判别模型，并统计在不同阈值下的属性判别结果平均准确率，结果如图11所示。实验结果表明，当判别阈值大于0.7时，属性判别的平均准确率较高（94%～98%），继续增加阈值后准确率的提升较小（小于1%）；另一方面，增加判别阈值将导致满足过滤条件的属性样本减少，以RAPv2数据集为例，当判别阈值由0.7调整为0.8时，样本的平均检出率由77.9%下降至49.8%。因此，根据上述参数分析结果，本文采用判别阈值$\tau = 0.7$，可以同时满足准确率和召回率要求。

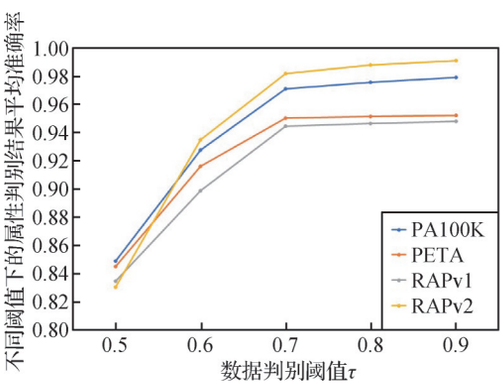


图11　不同阈值下的属性判别结果平均准确率

Fig. 11　Mean accuracy of attribute discrimination results under different thresholds

4.3.3　MR取值范围的参数分析

MR取值范围的选择需要考虑不同马赛克区域比例下的新样本生成情况，本文以RAPv2数据集为例，首先分别对比了不同MR取值下的新生成样本通过数据判别的比例，如图12所示，当MR大于0.7时，由于丢失了大量细节特征，导致属性识别结果下降，样本通过数据判别的比例低于5%（图中红色标注），因此MR取值范围的上限MR_{high}可以确定为0.7。

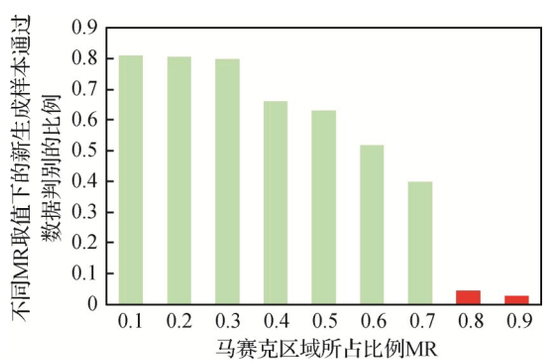


图12　不同MR取值下的新生成样本通过数据判别的比例

Fig. 12　Passing ratios of data discrimination for newly generated samples under different MR values

进一步地，实验分别针对不同MR_{low}取值下训练得到的判别模型结果（mA）进行对比分析，结果如表8所示，经过第1轮迭代，当$MR_{low} < 0.3$时，由于新样本的多样性下降导致模型结果变差，而当$MR_{low} > 0.3$时，由于新样本的低质量比例增加导致模型结果变差，因此MR取值范围的下限MR_{low}可以确定为0.3。

表8　不同MR_{low}取值下的模型结果mA对比

Table 8　mA comparison of model results under different MR_{low} values

MR_{low}	mA/%
0.1	79.50
0.2	79.73
0.3	**80.11**
0.4	80.04
0.5	79.53
0.6	78.51

注：加粗字体表示各列最优结果。

4.3.4　数据增广系数λ的参数分析

数据增广系数的选择需要考虑生成数据集的规模大小和样本均衡性，如表9所示，本文以RAPv2数据集为例，在不同的λ值下，分别针对第1次迭代后和第T次迭代后的模型结果（mA）进行对比。实验结果表明，当$\lambda < 2$时，由于生成数据集的样本均衡性变差导致模型精度下降，而当$\lambda > 2$时，由于生成数据集的样本规模变大影响模型训练效率，导致经过T轮渐进式迭代优化后的最终模型精度下降。因此，数据增广系数λ的取值可以确定为2。

表9　不同λ值下的模型结果mA对比

Table 9　mA comparison of model results under different λ values

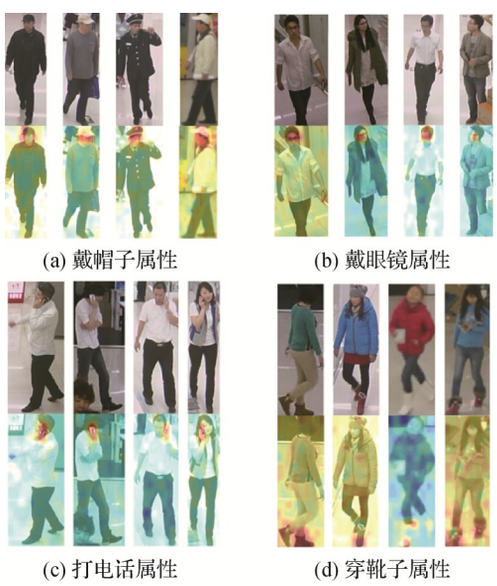
/%

λ值	第1次迭代后	第T次迭代后
1	78.99	80.57
2	80.11	**83.28**
3	80.44	81.92
4	**80.51**	81.59

注：加粗字体表示各列最优结果。

4.4　分析与讨论

本文对数据判别的筛选情况进行可视化验证，如图 13 所示，以戴帽子、戴眼镜、打电话和穿靴子 4 个典型属性为例，首先针对每个正样本图像随机生成 N 个（N 取 100）新样本，然后使用判别模型进行筛选，最后对通过筛选的新样本的马赛克区域进行实验分析，统计其中保留的细节特征的空间概率分布。通过实验结果可以发现，数据判别能够较好地保留行人目标关键属性的相关特征，从而引入启发式的注意力机制；另外，通过深入挖掘不同属性的相关特征分布情况，能够进一步提升判别模型的可解释性。



(a) 戴帽子属性　(b) 戴眼镜属性

(c) 打电话属性　(d) 穿靴子属性

图 13　4 种典型属性的数据判别筛选情况

Fig. 13　Filtering results of data discrimination for four typical attributes((a)wearing hat attribute; (b)wearing glasses attribute; (c)phoning attribute; (d)wearing boots attribute)

此外，本文以 RAPv2 数据集为例，验证 T 轮迭代后不均衡类别（正样本所占比例低于 10%）的个数变化情况，如表 10 所示，当 T 的取值为 3 时，经过 T 轮迭代，不均衡的类别数由 34 减少到 0，最终实现了数据集的渐进式优化。

表 10　经过 T 轮迭代后的不均衡类别的变化情况

Table 10　Variation of uneven categories after the T epochs

迭代轮次	不均衡类别个数	不均衡类别比例/%
迭代前	34	63
第 1 轮	27	50
第 2 轮	7	13
第 3 轮	**0**	**0**

注：加粗字体表示最优结果。

5　结　论

围绕行人属性识别任务中存在的样本不均衡问题，本文提出了一种基于渐进式迭代的模型和数据优化方法，通过数据生成与数据判别的相互对抗，引入启发式注意力机制，并构建了从 MAE 通用大模型到专用小任务的迁移框架，深入挖掘人体目标的常识特征，有效提升了模型的泛化能力。通过与目前主流的行人属性识别方法进行实验对比，结果表明本文提出的数据均衡化和渐进式迭代方法与现有的改进方法之间具有良好的互补性，并有助于进一步提升模型的准确性指标。

但是，本文方法存在以下不足之处，需要进一步深入研究。1）在实际场景下行人目标存在相互遮挡等问题，影响数据生成和模型迭代效果；2）行人目标的不同属性之间存在相关性，影响行人多属性判别模型的识别效果；3）行人动作姿态的样本多样性对于某些特定属性（如打电话等动作）的识别准确率影响较大，限制了模型精度的进一步提高。

下一步工作将基于 MAE 模型的通用特征表示，融合目标质量评价和人体骨架结构等先验知识，优化行人多属性之间的关系建模，进一步提升模型的可解释性。

参考文献（References）

Bagherinezhad H, Horton M, Rastegari M and Farhadi A. 2018. Label refinery: improving ImageNet classification through label progression [EB/OL]. [2022-09-24]. https://arxiv.org/pdf/1805.02641.pdf

Deng Y B, Luo P, Loy C C and Tang X O. 2014. Pedestrian attribute recognition at far distance//Proceedings of the 22nd ACM International Conference on Multimedia. Orlando, USA: ACM: 789-792 [DOI: 10.1145/2647868.2654966]

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J and Houlsby N. 2021. An image is worth 16 × 16 words: transformers for image recognition at scale [EB/OL]. [2022-09-24]. https://arxiv.org/pdf/2010.11929.pdf

Ghiasi G, Cui Y, Srinivas A, Qian R, Lin T Y, Cubuk E D, Le Q V and Zoph B. 2021. Simple copy-paste is a strong data augmentation method for instance segmentation//Proceedings of 2021 IEEE/CVF

Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 2917-2927 [DOI: 10.1109/CVPR46437.2021.00294]

Guo H, Zheng K, Fan X C, Yu H K and Wang S. 2019. Visual attention consistency under image transforms for multi-label image classification//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 729-739 [DOI: 10.1109/CVPR.2019.00082]

He K M, Chen X L, Xie S N, Li Y H, Dollár P and Girshick R. 2022. Masked autoencoders are scalable vision learners//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 15979-15988 [DOI: 10.1109/CVPR52688.2022.01553]

Jia J, Huang H J, Yang W J, Chen X T and Huang K Q. 2020. Rethinking of pedestrian attribute recognition: realistic datasets with efficient method [EB/OL]. [2022-09-22]. https://arxiv.org/pdf/2005.11909.pdf

Li D W, Chen X T and Huang K Q. 2015. Multi-attribute learning for pedestrian attribute recognition in surveillance scenarios//Proceedings of the 3rd IAPR Asian Conference on Pattern Recognition. Kuala Lumpur, Malaysia: IEEE: 111-115 [DOI: 10.1109/ACPR.2015.7486476]

Li D W, Zhang Z, Chen X T and Huang K Q. 2019. A richly annotated pedestrian dataset for person retrieval in real surveillance scenarios. IEEE Transactions on Image Processing, 28 (4): 1575-1590 [DOI: 10.1109/TIP.2018.2878349]

Li D W, Zhang Z, Chen X T, Ling H B and Huang K Q. 2016. A richly annotated dataset for pedestrian attribute recognition [EB/OL]. [2022-09-20]. https://arxiv.org/pdf/1603.07054.pdf

Liu X H, Zhao H Y, Tian M Q, Sheng L, Shao J, Yi S, Yan J J and Wang X G. 2017. HydraPlus-net: attentive deep features for pedestrian analysis//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 350-359 [DOI: 10.1109/ICCV.2017.46]

Liu Z W, Miao Z Q, Zhan X H, Wang J Y, Gong B Q and Yu S X. 2019. Large-scale long-tailed recognition in an open world//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 2532-2541 [DOI: 10.1109/CVPR.2019.00264]

Liu Z Y, Zhang Z, Li D, Zhang P and Shan C F. 2022. Dual-branch self-attention network for pedestrian attribute recognition. Pattern Recognition Letters, 163: 112-120 [DOI: 10.1016/j.patrec.2022.10.003]

Luo Y, Zhang C Y, Tian Y H, Guo J and Sun J. 2022. An overview of deep learning based pedestrian detection algorithms. Journal of Image and Graphics, 27(7): 2094-2111 (罗艳, 张重阳, 田永鸿, 郭捷, 孙军. 2022. 深度学习行人检测方法综述. 中国图象图形学报, 27(7): 2094-2111) [DOI: 10.11834/jig.200831]

Tang C F, Sheng L, Zhang Z X and Hu X L. 2019. Improving pedestrian attribute recognition with weakly-supervised multi-scale attribute-specific localization//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 4996-5005 [DOI: 10.1109/ICCV.2019.00510]

Yang Y, Tan Z C, Tiwari P, Pandey H M, Wan J, Lei Z, Guo G D and Li S Z. 2021. Cascaded split-and-aggregate learning with feature recombination for pedestrian attribute recognition. International Journal of Computer Vision, 129(10): 2731-2744 [DOI: 10.1007/s11263-021-01499-z]

Zhang H Y, Cisse M, Dauphin Y N and Lopez-Paz D. 2018. mixup: beyond empirical risk minimization [EB/OL]. [2022-09-22]. https://arxiv.org/pdf/1710.09412.pdf

Zhong Z, Zheng L, Kang G L, Li S Z and Yang Y. 2020. Random erasing data augmentation. Proceedings of 2020 AAAI Conference on Artificial Intelligence, 34(7): 13001-13008 [DOI: 10.1609/aaai.v34i07.7000]

作者简介

丁正彦,男,助理研究员,主要研究方向为智能视频监控。
E-mail: dzy_wlw@163.com
张重阳,通信作者,男,研究员,主要研究方向为图像目标检测与识别、视频内容分析与描述。
E-mail: sunny_zhang@sjtu.edu.cn
尚岩峰,男,副研究员,主要研究方向为物联网技术。
E-mail: aysyf@126.com