(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 113435319 B
(45)授权公告日 2022.05.10

(21)申请号 202110713283.4

(22)申请日 2021.06.25

(65)同一申请的已公布的文献号
申请公布号 CN 113435319 A

(43)申请公布日 2021.09.24

(73)专利权人 重庆邮电大学
地址 400065 重庆市南岸区南山街道崇文路2号

(72)发明人 杨春德 徐同耀 姜小明 吕明鸿 余毅 熊道文

(74)专利代理机构 重庆辉腾律师事务所 50215
专利代理师 王海军

(51)Int.Cl.
G06V 10/764(2022.01)
G06V 10/774(2022.01)
G06V 10/80(2022.01)
G06V 10/74(2022.01)
G06V 40/10(2022.01)
G06K 9/62(2022.01)

审查员 梁倩

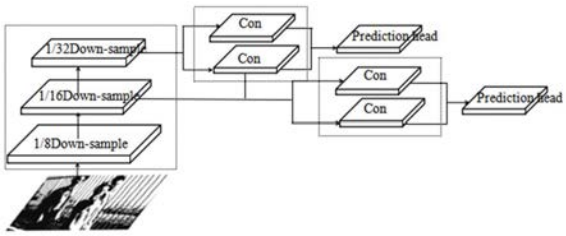
权利要求书3页 说明书8页 附图1页

(54)发明名称

一种联合多目标跟踪和行人角度识别的分类方法

(57)摘要

本发明属于多目标跟踪与行人角度识别领域,具体涉及一种联合多目标跟踪和行人角度识别的分类方法,该方法包括:将待检测的图像进行增强处理;将增强后的图像输入到训练好的分类模型中进行行人跟踪和角度的识别分类,根据分类结果对待检测图像进行标记;分类模型为改进的JDE多目标跟踪模型和行人角度识别模型;本发明通过特征共享的方式实现了多目标跟踪算法与行人角度识别的算法的结合,减小了模型参数数量,减小了计算量。本发明既能够对视频中出现的所有大目标以及中等目标进行跟踪,又能够对视频中出现的所有大目标以及中等目标的角度进行角度识别,同时该算法能够满足实时性要求。



CN 113435319 B

1.一种联合多目标跟踪和行人角度识别的分类方法,其特征在于,该方法包括:将待检测的图像进行增强处理;将增强后的图像输入到训练好的分类模型中进行行人跟踪和角度的识别分类,根据分类结果对待检测图像进行标记;分类模型为改进的JDE多目标跟踪模型和行人角度识别模型;所述改进的JDE多目标跟踪模型结构为将原始的JDE多目标跟踪模型的跟踪小目标分支结构删除,并增加两个预测大目标和中等目标行人角度识别信息分支;增加的预测大目标和中等目标行人角度识别信息分支分别采用了八个卷积层;

对分类模型进行训练的过程包括:

S1:获取原始数据集,根据原始数据集制作包含脸部角度信息与身体角度信息的多标签多分类数据集;

S2:对多标签多分类数据集中的数据进行增强处理,得到用于行人角度识别的训练数据集和测试数据集;

S3:根据用于行人角度识别的训练数据集获取跟踪数据集,得到用于多目标跟踪的训练数据集和测试数据集;

S4:将用于多目标跟踪的训练数据集输入到主干网络和跟踪分支中,得到行人的跟踪结果;

S5:根据跟踪结果计算与用于多目标跟踪的训练数据集中的真实标签之间的损失值,当损失值最小时,得到训练好的主干网络和跟踪分支;

S6:固定模型的主干网络和跟踪网络的参数;将多标签多分类训练数据集输入到分类分支中,得到分类结果,分类结果包括脸部角度、身体角度;

S7:根据分类结果计算与多标签多分类训练数据集中的真实标签之间的损失值,当损失值最小时,得到训练好的分类分支;

S8:将用于多目标跟踪的测试数据集中的数据输入到训练好的主干网络和跟踪分支中,得到跟踪功能的测试结果;

S9:将用于行人角度识别的测试数据集中的数据输入到训练好的分类分支中,得到分类功能的测试结果。

2.根据权利要求1所述的一种联合多目标跟踪和行人角度识别的分类方法,其特征在于,多标签多分类数据集包括采用DukeMTMC制作成含有975张图片的DukeFace数据集、采用数据集使用了Market-1501制作成含有5918张图片的FaceData1数据集、采用了Market-1501制作成含有3925张图片的FaceData2数据集、采用Mars制作成含有4439张图片的MarsFace数据集、采用MSMT17制作成含有5217张图片的Msmt17Face数据集以及PA-100K制作成含有3063张图片的Pa100kFace数据集。

3.根据权利要求1所述的一种联合多目标跟踪和行人角度识别的分类方法,其特征在于,对模型的分类分支进行训练的过程包括:固定主干网络和两个跟踪分支的参数;将用于多目标跟踪的训练数据集输入到主干网络与跟踪分支中,经过卷积层运算,得到跟踪分支预测结果;从跟踪分支产生的预测结果中选取出置信度最大的位置;通过置信度最大的位置找到分类分支中对应的脸部角度与身体角度的预测结果;通过BCEWithLogitsLoss损失函数计算脸部角度与身体角度的真实标签与找到的模型对于脸部角度与身体角度的预测结果之间的损失值。

4.根据权利要求3所述的一种联合多目标跟踪和行人角度识别的分类方法,其特征在

于,对模型的跟踪分支进行训练过程中所采用的损失函数包括Cross Entropy Loss损失函数、Smooth-L1 Loss损失函数以及Cross Entropy Loss损失函数;采用Cross Entropy Loss损失函数计算目标类别的损失,目标类别包括行人与非行人;使用Smooth-L1 Loss损失函数计算边界框位置的回归损失;使用Cross Entropy Loss损失函数计算提取到的嵌入特征的损失,对所有的损失求和,得到跟踪分支损失函数。

5.根据权利要求1所述的一种联合多目标跟踪和行人角度识别的分类方法,其特征在于,对行人的跟踪和角度的识别分类的过程包括:

步骤1:将每帧图像输入到分类模型中,经过1/8、1/16和1/32的上采样后得到三张不同尺寸大小的特征图;

步骤2:将最小尺寸的特征图分别输入到预测大目标的分类分支和检测/嵌入特征提取分支中,得到大目标的预测信息;

步骤3:将最小尺寸大小的特征图与中等尺寸大小的特征图进行融合,并将融合后的特征图分别输入预测中等目标的分类分支与检测/嵌入特征提取分支,得到中等目标的预测信息;

步骤4:将得到的大目标的预测信息和中等目标的预测信息组合在一起,得到模型对当前帧图片的所有目标的最终预测信息;

步骤5:采用卡尔曼滤波对上一帧中的目标在当前帧中的最佳位置进行预测,得到上一帧中的目标在当前帧中的最佳位置;

步骤6:从模型的最终预测信息中提取出模型预测结果的当前帧中所有目标的嵌入特征,使用余弦距离计算轨迹的嵌入特征与所有目标的嵌入特征之间的嵌入特征相似度,并使用Jonker-Volgenant算法对嵌入特征相似度进行第一次匹配,得到部分已经匹配的目标和轨迹、未匹配的目标、未匹配的轨迹;

步骤7:从模型的最终预测信息中提取出模型预测的当前帧中所有目标的边界框位置,使用IOU距离计算轨迹在当前帧中的最佳位置与当前帧中所有目标的边界框位置之间的运动信息相似度,并使用Jonker-Volgenant算法对运动信息相似度进行第二次匹配,得到部分匹配的目标和轨迹、未匹配的目标、未匹配的轨迹;

步骤8:更新轨迹,并根据匹配到的目标和轨迹对当前帧中所有目标的跟踪ID、检测位置、脸部角度和身体角度进行标记。

6.根据权利要求5所述的一种联合多目标跟踪和行人角度识别的分类方法,其特征在于,每个网络分支得到的预测信息包含四个部分;四个部分包括分类信息、回归信息、特征嵌入信息以及分类结果;所述分类信息包括目标信息和非目标信息,即2A*W*H;所述回归信息为边界框位置,即4A*W*H;所述特征嵌入信息为重识别,即512*W*H;所述分类结果为行人属性识别,即6*W*H;其中,A为属性的数量,W、H分别为预测信息的宽度和高度。

7.根据权利要求5所述的一种联合多目标跟踪和行人角度识别的分类方法,其特征在于,采用卡尔曼滤波对上一帧中的目标在当前帧中的最佳位置进行预测的过程包括:首先根据轨迹的位置与速度得到轨迹在当前帧中的预测位置,然后从模型的预测信息中得到模型对当前帧中目标的观测位置,然后将从卡尔曼滤波得到的预测位置与从模型得到的观测位置进行加权平均,进而得到上一帧中的目标在当前帧中的最佳位置,其中所使用的权值为观测位置与最佳位置的均方误差。

8.根据权利要求5所述的一种联合多目标跟踪和行人角度识别的分类方法，其特征在于，计算嵌入特征相似度公式为：

$$dist = 1 - \frac{u \cdot v}{\|u\|_2 \|v\|_2}$$

其中u为所有轨迹的嵌入特征所组成的向量，v为所有目标的嵌入特征所组成的特征向量，$||||_2$为其参数的2范数；

运动信息相似度公式为：

$$IOU = \frac{S_A}{S_B}$$

$S_A = (\min\{x_{a2}, x_{b2}\} - \max\{x_{a1}, x_{b1}\}) \times (\min\{y_{a2}, y_{b2}\} - \max\{y_{a1}, y_{b1}\})$

$S_B = (x_{a2} - x_{a1})(y_{a2} - y_{a1}) + (x_{b2} - x_{b1})(y_{b2} - y_{b1}) - S_B$；

其中，x_{a1}、y_{a1}分别为轨迹的左上顶点的横坐标、纵坐标，x_{a2}、y_{a2}分别为轨迹右下顶点的横坐标、纵坐标，x_{b1}、y_{b1}分别为模型预测的目标位置的左上顶点的横坐标、纵坐标，x_{b2}、y_{b2}分别为模型预测的目标位置的左上顶点的横坐标、纵坐标；min{}表示取其两个参数中较小的参数，max{}表示取其两个参数中较大的参数。

9.根据权利要求5所述的一种联合多目标跟踪和行人角度识别的分类方法，其特征在于，采用Jonker-Volgenant算法进行匹配的过程包括：在第一次匹配中，使用嵌入特征相似度组成的损失矩阵作为参数，输入到python中lap库的lapjv函数，得到匹配的轨迹与目标、未匹配的轨迹、未匹配的目标；第二次匹配中，使用运动信息相似度组成的损失矩阵作为参数，输入到python中lap库的lapjv函数，得到匹配的轨迹与目标、未匹配的轨迹、未匹配的目标。

一种联合多目标跟踪和行人角度识别的分类方法

技术领域

[0001] 本发明属于多目标跟踪与行人角度识别领域,具体涉及一种联合多目标跟踪和行人角度识别的分类方法。

背景技术

[0002] 多目标跟踪任务(Multi-object tracking)与行人属性识别任务(Pedestrian Attribute Recognition)是计算机视觉领域中常见的两个任务。多目标跟踪任务,其目的是输入一个视频,输出视频中所有目标的运动轨迹。行人属性识别任务,其目的是输入一张图片,输出图片中目标的多个属性。

[0003] 多目标跟踪任务与行人属性识别任务紧密相关,往往同时出现。在复杂的场景中,有时既需要对视频中出现的目标进行跟踪,又需要对视频中出现的每一个行人的属性进行识别。但是现实中这两个不同的计算机视觉任务往往被独立研究,很少有人考虑将它们结合在一起。

[0004] 独立执行多目标跟踪算法与行人属性识别算法时的过程为:使用一个多目标跟踪模型检测出视频中出现的所有目标的边界框以及每个目标的ID;使用一个行人属性识别模型对视频中出现的每一个目标的属性进行分类。

[0005] 这种分离式的做法依次使用了两个不同的模型。这种方式固然可以解决复杂场景下多目标跟踪与行人属性识别联合的问题,但是由于多目标跟踪模型和行人属性识别模型没有共享模型的特征提取部分,因此往往会出现计算量大、实时性差的问题。

发明内容

[0006] 为解决以上现有技术存在的问题,本发明提出了一种联合多目标跟踪和行人角度识别的分类方法,该方法包括:将待检测的图像进行增强处理;将增强后的图像输入到训练好的分类模型中进行行人跟踪和角度的识别分类,根据分类结果对待检测图像进行标记;分类模型为改进的JDE多目标跟踪模型和行人角度识别模型;

[0007] 对分类模型进行训练的过程包括:

[0008] S1:获取原始数据集,根据原始数据集制作包含脸部角度信息与身体角度信息的多标签多分类数据集;

[0009] S2:对多标签多分类数据集中的数据进行增强处理,得到用于行人角度识别的训练数据集和测试数据集;

[0010] S3:根据用于行人角度识别的训练数据集获取跟踪数据集,得到用于多目标跟踪的训练数据集和测试数据集;

[0011] S4:将用于多目标跟踪的训练数据集输入到主干网络和跟踪分支中,得到行人的跟踪结果;

[0012] S5:根据跟踪结果计算与用于多目标跟踪的训练数据集中的真实标签之间的损失值,当损失值最小时,得到训练好的主干网络和跟踪分支;

[0013] S6:固定模型的主干网络和跟踪网络的参数;将多标签多分类训练数据集输入到分类分支中,得到分类结果,分类结果包括脸部角度、身体角度;

[0014] S7:根据分类结果计算与多标签多分类训练数据集中的真实标签之间的损失值,当损失值最小时,得到训练好的分类分支;

[0015] S8:将用于多目标跟踪的测试数据集中的数据输入到训练好的主干网络和跟踪分支中,得到跟踪功能的测试结果;

[0016] S9:将用于行人角度识别的测试数据集中的数据输入到训练好的分类分支中,得到分类功能的测试结果。

[0017] 优选的,多标签多分类数据集包括采用DukeMTMC制作成含有975张图片的DukeFace数据集、采用数据集使用了Market-1501制作成含有5918张图片的FaceData1数据集、采用了Market-1501制作成含有3925张图片的FaceData2数据集、采用Mars制作成含有4439张图片的MarsFace数据集、采用MSMT17制作成含有5217张图片的Msmt17Face数据集以及PA-100K制作成含有3063张图片的Pa100kFace数据集。

[0018] 优选的,改进的JDE多目标跟踪模型结构为将原始的JDE多目标跟踪模型的跟踪小目标分支结构删除,并增加两个预测大目标和中等目标行人角度识别信息分支;增加的预测大目标和中等目标行人角度识别信息分支分别采用了八个卷积层。

[0019] 优选的,对模型的分类分支进行训练的过程包括:固定主干网络和两个跟踪分支的参数;将用于多目标跟踪的训练数据集输入到主干网络与跟踪分支中,经过卷积层运算,得到跟踪分支预测结果;从跟踪分支产生的预测结果中选取出置信度最大的位置;通过置信度最大的位置找到分类分支中对应的脸部角度与身体角度的预测结果;通过BCEWithLogitsLoss损失函数计算脸部角度与身体角度的真实标签与找到的模型对于脸部角度与身体角度的预测结果之间的损失值。

[0020] 进一步的,对跟踪分支进行训练过程中所采用的损失函数包括Cross Entropy Loss损失函数、Smooth-L1 Loss损失函数以及Cross Entropy Loss损失函数;采用Cross Entropy Loss损失函数计算目标类别的损失,目标类别包括行人与非行人;使用Smooth-L1 Loss损失函数计算边界框位置的回归损失;使用Cross Entropy Loss损失函数计算提取到的嵌入特征的损失,对所有的损失求和,得到跟踪分支损失函数。

[0021] 优选的,对行人的跟踪和角度的识别分类的过程包括:

[0022] 步骤1:将每帧图像输入到分类模型中,经过1/8、1/16和1/32的上采样后得到三张不同尺寸大小的特征图;

[0023] 步骤2:将最小尺寸的特征图分别输入到预测大目标的分类分支和检测/嵌入特征提取分支中,得到大目标的预测信息;

[0024] 步骤3:将最小尺寸大小的特征图与中等尺寸大小的特征图进行融合,并将融合后的特征图分别输入预测中等目标的分类分支与检测/嵌入特征提取分支,得到中等目标的预测信息;

[0025] 步骤4:将得到的大目标的预测信息和中等目标的预测信息组合在一起,得到模型对当前帧图片的所有目标的最终预测信息;

[0026] 步骤5:采用卡尔曼滤波对上一帧中的目标在当前帧中的最佳位置进行预测,得到上一帧中的目标在当前帧中的最佳位置;

[0027] 步骤6:从模型的最终预测信息中提取出模型预测结果的当前帧中所有目标的嵌入特征,使用余弦距离计算轨迹的嵌入特征与所有目标的嵌入特征之间的嵌入特征相似度,并使用Jonker-Volgenant算法对嵌入特征相似度进行第一次匹配,得到部分已经匹配的目标和轨迹、未匹配的目标、未匹配的轨迹;

[0028] 步骤7:从模型的最终预测信息中提取出模型预测的当前帧中所有目标的边界框位置,使用IOU距离计算轨迹在当前帧中的最佳位置与当前帧中所有目标的边界框位置之间的运动信息相似度,并使用Jonker-Volgenant算法对运动信息相似度进行第二次匹配,得到部分匹配的目标和轨迹、未匹配的目标、未匹配的轨迹;

[0029] 步骤8:更新轨迹,并根据匹配到的目标和轨迹对当前帧中所有目标的跟踪ID、检测位置、脸部角度和身体角度进行标记。

[0030] 进一步的,每个网络分支得到的预测信息包含四个部分;四个部分包括分类信息、回归信息、特征嵌入信息以及分类结果;所述分类信息包括目标信息和非目标信息,即2A*W*H;所述回归信息为边界框位置,即4A*W*H;所述特征嵌入信息为重识别,即512*W*H;所述分类结果为行人属性识别,即6*W*H;其中,A为属性的数量,取值为6,W、H分别为预测信息的宽度和高度,大目标的预测信息的W、H分别为34、17,中等目标的预测信息的W、H分别为68、34。

[0031] 进一步的,采用卡尔曼滤波对上一帧中的目标在当前帧中的最佳位置进行预测的过程包括:首先根据轨迹的位置与速度得到轨迹在当前帧中的预测位置,然后从模型的预测信息中得到模型对当前帧中目标的观测位置,然后将从卡尔曼滤波得到的预测位置与从模型得到的观测位置进行加权平均,进而得到上一帧中的目标在当前帧中的最佳位置,其中所使用的权值为观测位置与最佳位置的均方误差。

[0032] 进一步的,计算嵌入特征相似度的过程包括:使用余弦距离计算轨迹的嵌入特征与所有目标的嵌入特征之间的嵌入特征相似度,使用到的余弦距离公式具体如下所示:

[0033] $$dist = 1 - \frac{u \cdot v}{\|u\|_2 \|v\|_2}$$

[0034] 其中u为所有轨迹的嵌入特征所组成的向量,v为所有目标的嵌入特征所组成的特征向量,$||\ ||_2$为其参数的2范数。

[0035] 进一步的,计算运动信息相似度的过程包括:计算轨迹在当前帧中的最佳位置与当前帧中所有目标的观测位置之间的运动信息相似度,使用到的IOU距离公式具体如下所示:

[0036] $$IOU = \frac{S_A}{S_B}$$

[0037] $S_A = (\min\{x_{a2}, x_{b2}\} - \max\{x_{a1}, x_{b1}\}) \times (\min\{y_{a2}, y_{b2}\} - \max\{y_{a1}, y_{b1}\})$;

[0038] $S_B = (x_{a2} - x_{a1})(y_{a2} - y_{a1}) + (x_{b2} - x_{b1})(y_{b2} - y_{b1}) - S_B$;

[0039] 其中,x_{a1}、y_{a1}分别为轨迹的左上顶点的横坐标、纵坐标,x_{a2}、y_{a2}分别为轨迹右下顶点的横坐标、纵坐标,x_{b1}、y_{b1}分别为模型预测的目标位置的左上顶点的横坐标、纵坐标,x_{b2}、y_{b2}分别为模型预测的目标位置的左上顶点的横坐标、纵坐标;min{}表示取其两个参数中较小的参数,max{ }表示取其两个参数中较大的参数。

[0040] 进一步的,采用Jonker-Volgenant算法进行匹配的过程包括:在第一次匹配中,使

用嵌入特征相似度组成的损失矩阵作为参数，输入到python中lap库的lapjv函数，得到匹配的轨迹与目标、未匹配的轨迹、未匹配的目标；第二次匹配中，使用运动信息相似度组成的损失矩阵作为参数，输入到python中lap库的lapjv函数，得到匹配的轨迹与目标、未匹配的轨迹、未匹配的目标。

[0041] 本发明通过特征共享的方式实现了多目标跟踪算法与行人角度识别的算法的结合，减小了模型参数数量，减小了计算量。本发明既能够对视频中出现的所有大目标以及中等目标进行跟踪，又能够对视频中出现的所有大目标以及中等目标的角度进行角度识别，同时该算法能够满足实时性要求。

附图说明

[0042] 图1为本发明的网络整体结构图；

[0043] 图2为本发明的网络头部结构图。

具体实施方式

[0044] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0045] 一种联合多目标跟踪和行人角度识别的分类方法，该方法包括：将待检测的图像进行增强处理；将增强后的图像输入到训练好的分类模型中进行行人跟踪和角度的识别分类，根据分类结果对待检测图像进行标记；分类模型为改进的JDE多目标跟踪模型和行人角度识别模型。

[0046] 对分类模型进行训练的过程包括：

[0047] S1：获取原始数据集，根据原始数据集制作包含脸部角度信息与身体角度信息的多标签多分类数据集；

[0048] S2：对多标签多分类数据集中的数据进行增强处理，得到用于行人角度识别的训练数据集和测试数据集；

[0049] S3：根据用于行人角度识别的训练数据集获取跟踪数据集，得到用于多目标跟踪的训练数据集和测试数据集；

[0050] S4：将用于多目标跟踪的训练数据集输入到主干网络和跟踪分支中，得到行人的跟踪结果，跟踪结果包括目标ID、边界框坐标、目标类别(行人、非行人)；

[0051] S5：固定模型的主干网络和跟踪网络的参数；

[0052] S6：将多标签多分类训练数据集输入到分类分支中，得到分类结果；

[0053] S7：计算模型的损失函数，当损失函数最小时，得到训练好的模型；

[0054] S8：将测试集中的数据输入到训练好的模型中，得到测试结果。

[0055] 多标签多分类数据集包括采用DukeMTMC制作成含有975张图片的DukeFace数据集、采用数据集使用了Market-1501制作成含有5918张图片的FaceData1数据集、采用了Market-1501制作成含有3925张图片的FaceData2数据集、采用Mars制作成含有4439张图片的MarsFace数据集、采用MSMT17制作成含有5217张图片的Msmt17Face数据集以及PA-100K

制作成含有3063张图片的Pa100kFace数据集。

[0056] 本发明在加载数据时对数据进行了增强处理，以提高模型的泛化能力，使用到的数据增强方式包括：模糊图像、裁剪、随机应用仿射变换(平移，缩放，旋转)、网格失真、弹性变换、随机色调变化、随机饱和度变化、随机亮度变化、随机对比度变化、重新排列输入RGB图像的通道、随机擦处、光学畸变、运动模糊、中心模糊、高斯模糊、增加高斯噪声。

[0057] 本发明的多标签多分类数据集包括采用DukeMTMC制作成含有975张图片的DukeFace数据集、采用数据集使用了Market-1501制作成含有5918张图片的FaceData1数据集、采用了Market-1501制作成含有3925张图片的FaceData2数据集、采用Mars制作成含有4439张图片的MarsFace数据集、采用MSMT17制作成含有5217张图片的Msmt17Face数据集以及PA-100K制作成含有3063张图片的Pa100kFace数据集。

[0058] 本发明改进了JDE(Jointly learns the Detector and Embedding model)多目标跟踪模型，具体包括：减去了跟踪小目标的分支，增加了两个预测大目标以及中等目标行人角度识别信息的分支，其中，增加的两个预测大目标以及中等目标行人角度信息的分支分别使用了八个卷积层。

[0059] 如图1所示，修改后的网络包括两个预测大目标信息与中等目标信息的分支，每个分支又包括预测行人角度识别信息以及检测信息与嵌入特征信息的两个子分支。

[0060] 如图2所示，网络头部结构输出了模型最终的预测信息，包括分类信息，检测信息，嵌入特征，角度识别信息。

[0061] 本发明的训练模型的主干网络和跟踪分支部分，具体内容如下：

[0062] 本发明使用跟踪数据集训练模型的主干网络以及两个跟踪分支，对模型的踪分支进行训练过程中所采用的损失函数包括Cross Entropy Loss损失函数、Smooth-L1 Loss损失函数以及Cross Entropy Loss损失函数；采用Cross Entropy Loss损失函数计算目标类别的损失，目标类别包括行人与非行人；使用Smooth-L1 Loss损失函数计算边界框位置的回归损失；使用Cross Entropy Loss损失函数计算提取到的嵌入特征的损失，将所有的损失相加，得到跟踪分支损失函数。

[0063] 计算损失的具体过程包括：

[0064] 将跟踪数据集中的图片输入模型，得到模型对于目标类别、边界框位置、嵌入特征的预测结果。

[0065] 步骤1：使用Cross Entropy Loss计算目标类别的损失，公式具体如下所示：

[0066] $$L=-\frac{1}{n}\left[y\cdot log(p)+(1-y)\cdot log(1-p)\right]$$

[0067] 其中，p为模型判断图像中目标属于行人的概率，n为分类总数，取值为2，y为数据集中行人的标签，行人标签为1，非行人标签为0。

[0068] 步骤2：使用Smooth-L1 Loss计算边界框位置的回归损失，公式具体如下所示：

[0069] $$L=\frac{1}{n}\sum_{i=1}^{n}\begin{cases}0.5(x_i-y_i)^2, & |x_i-y_i|<1\\ |x_i-y_i|-0.5, & |x_i-y_i|\geq 1\end{cases}$$

[0070] 其中，n为左上角横坐标、纵坐标与右下角横坐标、纵坐标的总数量，取值为4，x_i表示跟踪数据集中行人位置中的第i个标签值，y_i表示模型预测的行人位置中的第i个标签

值。

[0071] 步骤3:使用在Cross Entropy Loss计算提取到的嵌入特征的损失,公式如步骤1所示。其中,在本步骤中,p为模型判断图像中目标属于第几个行人的概率,y为数据集中行人的ID标签,n为行人总数,取值为512。

[0072] 步骤4:将步骤1、步骤2、步骤3得到的损失函数累加起来,作为最终的损失值。

[0073] 本发明训练模型的分类分支部分,具体内容如下:

[0074] 由于行人属性识别数据集没有位置信息,因此本发明采用跟踪分支预测的位置信息,用来指导分类分支的训练。即训练过程包括:

[0075] 步骤1:本发明固定主干网络与两个跟踪分支的参数。

[0076] 步骤2:本发明从跟踪分支产生的预测结果中选取出置信度最大的位置。

[0077] 步骤3:本发明通过置信度最大的位置找到分类分支中对应的分类结果。

[0078] 步骤4:本发明通过BCEWithLogitsLoss损失函数计算脸部角度与身体角度的真实标签与找到的模型对于脸部角度与身体角度的预测结果之间的损失值,其中BCEWithLogitsLoss的具体公式如下所示:

[0079] $L=\{l_1,l_2,...,l_N\}^T$

[0080] 其中,N为属性个数6;l_n公式如下所示:

[0081] $l_n=-w_n[y_n \cdot \log\sigma(x_n)+(1-y_n) \cdot \log(1-\sigma(x_n))]$

[0082] 其中,w为权重,y为标签值,x为模型预测值,n为第几个属性,α()为Sigmod函数。

[0083] 在对主干网络、跟踪分支、分类分支进行训练过程中由于跟踪信息与分类信息耦合在一起不利于扩展到其他任务,且分类数据集远远多于跟踪数据集,可以更加充分的利用目前已有的分类数据集;因此,本发明采用了独立的跟踪数据集与分类数据集。

[0084] 为了提高模型的泛化能力,本发明使用了标签平滑技术,具体公式如下所示:

[0085] $$\hat{y}=\begin{cases}1-\alpha, i=target\\ \alpha/K, i\neq target\end{cases}$$

[0086] 其中,K为属性个数,取值为6;α为超参数,取值为0.1。

[0087] 本发明对行人的跟踪与角度的识别的具体过程包括:

[0088] 步骤1:将每帧图像输入到分类模型中,经过1/8、1/16和1/32的上采样后得到三张不同尺寸大小的特征图;

[0089] 步骤2:将最小尺寸的特征图分别输入到预测大目标的分类分支和检测/嵌入特征提取分支中,得到大目标的预测信息;

[0090] 步骤3:将最小尺寸大小的特征图与中等尺寸大小的特征图进行融合,并将融合后的特征图分别输入预测中等目标的分类分支与检测/嵌入特征提取分支,得到中等目标的预测信息;

[0091] 步骤4:将得到的大目标的预测信息和中等目标的预测信息组合在一起,得到模型对当前帧图片的所有目标的最终预测信息;

[0092] 步骤5:采用卡尔曼滤波进行预测,得到上一帧中的目标在当前帧中的最佳位置;

[0093] 步骤6:从模型的最终预测信息中得到当前帧中所有目标的嵌入特征,并使用余弦距离计算轨迹的嵌入特征与所有目标的嵌入特征之间的嵌入特征相似度,并使用Jonker-Volgenant算法进行第一次匹配,得到部分已经匹配的目标和轨迹、未匹配的目标、未匹配

的轨迹；

[0094] 步骤7：从模型的最终预测信息中直接取出当前帧中所有目标的边界框位置，并使用IOU距离计算轨迹在当前帧中的最佳位置与当前帧中所有目标的边界框位置之间的运动信息相似度，并使用Jonker-Volgenant算法进行第二次匹配，得到部分匹配的目标和轨迹、未匹配的目标、未匹配的轨迹；

[0095] 步骤8：更新轨迹，并根据匹配到的目标和轨迹对当前帧中所有目标的跟踪ID、检测位置、脸部角度和身体角度进行标记。

[0096] 进一步的，每个网络分支得到的预测信息包含四个部分；即1、分类信息(目标、非目标)：2A*W*H；2、回归信息(边界框位置)：4A*W*H；3、特征嵌入信息(重识别)：512*W*H；4、分类信息(行人属性识别)：6*W*H；其中A为属性的数量，取值为6，W、H分别为预测信息的宽度和高度，大目标的预测信息的W、H分别为34、17，中等目标的预测信息的W、H分别为68、34。

[0097] 进一步的，采用卡尔曼滤波对目标的位置进行预测的过程包括：采用卡尔曼滤波对上一帧中的目标在当前帧中的最佳位置进行预测的过程包括：首先根据轨迹的位置与速度得到轨迹在当前帧中的预测位置，从模型的预测信息中得到模型对当前帧中目标的观测位置，然后将从卡尔曼滤波得到的预测位置与从模型得到的观测位置加权平均，进而得到上一帧中的目标在当前帧中的最佳位置，其中所使用的权值为观测位置与最佳位置的均方误差。

[0098] 进一步的，计算嵌入特征相似度的过程包括：将轨迹的嵌入特征与所有目标的嵌入特征输入余弦距离公式，得到轨迹的嵌入特征与所有目标的嵌入特征之间的嵌入特征相似度，使用到的余弦距离公式具体如下所示：

[0099] $$dist = 1 - \frac{u \cdot v}{\|u\|_2 \|v\|_2}$$

[0100] 其中u为所有轨迹的嵌入特征所组成的向量，v为所有目标的嵌入特征所组成的特征向量，$||||_2$为其参数的2范数。

[0101] 进一步的，计算运动信息相似度的过程包括：计算轨迹在当前帧中的最佳位置与当前帧中所有目标的观测位置之间的运动信息相似度，使用到的IOU距离公式具体如下所示：

[0102] $$IOU = \frac{S_A}{S_B}$$

[0103] $S_A = (\min\{x_{a2}, x_{b2}\} - \max\{x_{a1}, x_{b1}\}) \times (\min\{y_{a2}, y_{b2}\} - \max\{y_{a1}, y_{b1}\})$；

[0104] $S_B = (x_{a2} - x_{a1})(y_{a2} - y_{a1}) + (x_{b2} - x_{b1})(y_{b2} - y_{b1}) - S_B$；

[0105] 其中，x_{a1}、y_{a1}分别为轨迹的左上顶点的横坐标、纵坐标，x_{a2}、y_{a2}分别为轨迹右下顶点的横坐标、纵坐标，x_{b1}、y_{b1}分别为模型预测的目标位置的左上顶点的横坐标、纵坐标，x_{b2}、y_{b2}分别为模型预测的目标位置的左上顶点的横坐标、纵坐标；min{}表示取其两个参数中较小的参数，max{}表示取其两个参数中较大的参数。

[0106] 进一步的，采用Jonker-Volgenant算法进行匹配的过程包括：

[0107] 在第一次匹配中，使用嵌入特征相似度组成的损失矩阵作为参数，输入到python中lap库的lapjv函数，得到匹配的轨迹与目标、未匹配的轨迹、未匹配的目标；第二次匹配中，使用运动信息相似度组成的损失矩阵作为参数，输入到python中lap库的lapjv函数，得

到匹配的轨迹与目标、未匹配的轨迹、未匹配的目标。

[0108] 本发明生成含有脸部角度与身体角度的多标签多分类数据集的详细信息包括:本发明根据已有的行人重识别数据集以及行人属性识别数据集制作了含有23537张图片的多标签多分类数据集,其中每张图片含有两个标签(Face角度、Body角度),每个标签有三个分类(FaceFront、FaceSide、FaceBack,BodyFront、BodySide、BodyBack)。

[0109] 本发明使用的行人重识别数据集有Market-1501、Mars、MSMT17,使用的行人属性识别的数据集有DukeMTMC、PA-100K。

[0110] 以上所举实施例,对本发明的目的、技术方案和优点进行了进一步的详细说明,所应理解的是,以上所举实施例仅为本发明的优选实施方式而已,并不用以限制本发明,凡在本发明的精神和原则之内对本发明所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

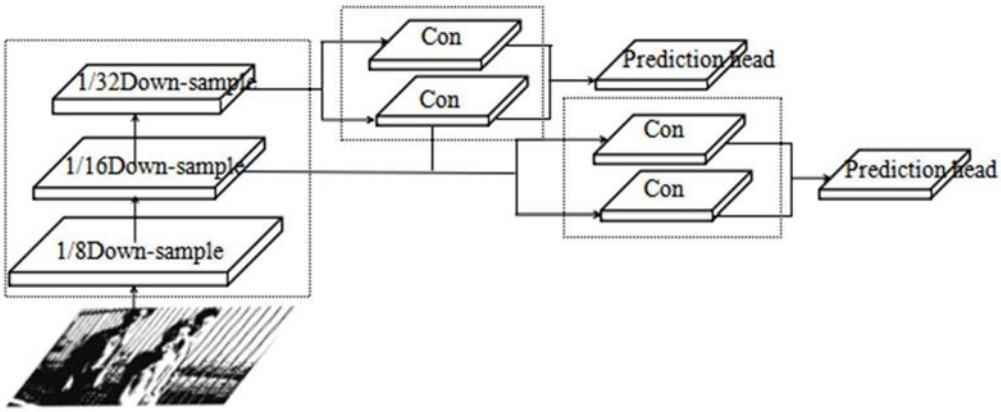

1/32Down-sample
1/16Down-sample
1/8Down-sample
Con
Con
Prediction head
Con
Con
Prediction head

图1

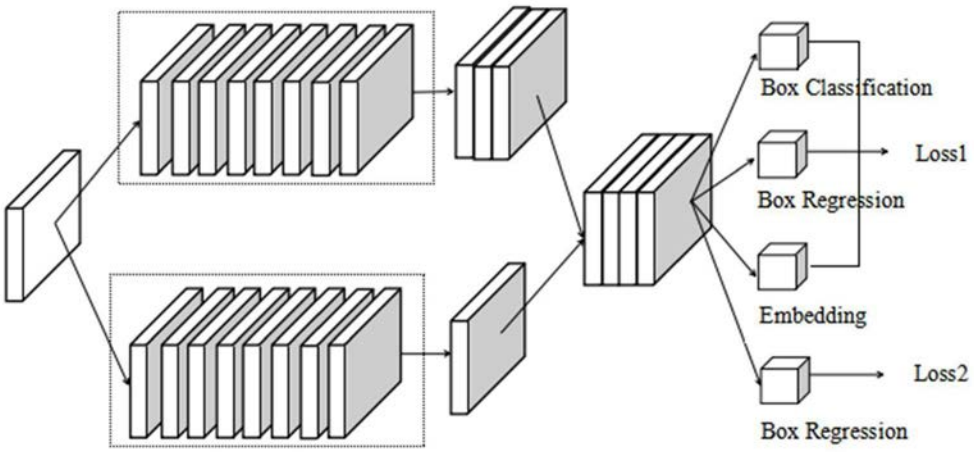

Box Classification
Loss1
Box Regression
Embedding
Loss2
Box Regression

图2