

侦、社会安全、身份认证等领域逐渐发挥起重要作用^[1]。由此,步态识别在计算机视觉领域也掀起了一片研究热潮。

然而,步态识别仍然面临着很多挑战,拍摄视角、衣着、携带物、行走环境的变化等因素大大影响了步态识别的准确率。而在众多干扰因素中,最常见的就是拍摄视角的变化,因为在实际情境中人行走的方向和视频拍摄之间的角度是最难控制且变化最多的。解决跨视角步态识别问题,是本文研究的重点。

在深度学习领域,生成对抗网络^[2](Generative Adversarial Networks, GAN)自2014年提出以来就一直是学术界的研究热点,其原理为生成器和判别器的相互作用。在图像处理方面,对于修复受损图像、生成全新图像等应用场景,GAN的效果非常好。受此启发,本文通过对生成对抗网络改进,将任意视角或行走状态下的步态模板转换成特定视角且正常状态的步态模板,进而去除干扰,提取有效的步态特征,从而提高步态识别的准确率。

本文的结构安排如下,第1部分介绍了近年来国内外对于步态识别的研究成果,第2部分介绍所提出的基于生成对抗网络的跨视角步态特征提取方法,第3部分展示了所提出方法与其他相关方法的实验结果对比与分析,第4部分对全文进行了总结与展望。

1 国内外相关工作

近年来,随着深度学习的发展,越来越多的研究人员将深度学习用于步态识别技术中。如Shiraga等人将步态能量图作为模板,提出了基于卷积神经网络的步态识别模型GEINet^[3]。Wu等人在文献^[4]中提出的基于深度卷积神经网络的3种模型(deepCNNs)就取得了非常优异的效果,该模型输入为一对步态图像,经特征提取和计算,输出的是两个样本的相似度,文中还对不同情况下三种模型的性能进行的对比分析。

为了克服衣着和携带物对识别效果的影响,Liao等提出了一种从原始视频中提取人体姿态关键点的方法^[5],然后利用基于姿态的时空网络学习步态特征。Feng等提出了基于长短时记忆模块的人体姿态估计模型^[6],针对人体关节热图进行特征提取。

为了解决跨视角步态识别问题,一些研究人员试图利用3D重建的方法,以通过在所需视角下投影3D模型来生成任意的二维视图,但这种方法通常需要

严格的环境条件和巨大的计算代价。目前,解决跨视角问题最常见的方法是构建能够将步态特征从一个视图投影到另一个视图的转换模型。Makihara等^[7]提出了一种视图转换模型,主要是利用奇异值分解的方法计算每个步态能量图的投影矩阵并计算出视角特征值。受鲁棒的主成分分析在特征提取上的良好表现的启发,Zheng等^[8]提出了一种鲁棒的视图转换模型用于步态识别。Bashir等^[9]提出了一种基于高斯过程的框架来预估测试集中的步态视角的方法,然后运用典型相关分析得出不同视角的步态序列的相关性。此外,Yu等人在文献^[10]中提出了一种用自动编码器(SPAE)去除来自于携带物、衣着等影响因素,之后将不同视角的步态模板特征转换到统一侧面视角的模型。但训练自动编码器过程较为复杂,要按照任务拆分从下至上依次训练,逐层累加,各层累加完毕后要对整个网络重新调整。

受GAN在多领域图像翻译上的优秀效果的启发,本文提出了一种基于生成对抗网络跨视角的步态特征提取方法,主要思想是步态模板的多视角和多状态转换。生成对抗网络是一种深度学习模型,是近年来复杂分布上的无监督学习领域中最具前景的方法之一。例如,Choi等人在2017年提出的StarGAN^[11]模型在人脸表情、发色转换的效果非常惊人,堪称GAN在多领域图像翻译中的典型成功案例。GAN的实现方法是让生成器和判别器进行博弈,训练过程中通过相互竞争让这两个模型同时得到增强。Yu等在2017年提出的GaitGAN^[12]就是将GAN用于视角转换来进行步态特征提取,与传统GAN不同的是,GaitGAN使用了两个判别器来确保生成图片真实且保留有效的身份信息。但这种方法只是把任意视角的步态模板转换成了侧面视角的步态模板,并不能应对实际情况中复杂的多视角变换问题。此外,He等人在2019年提出的MGANs^[13]也是一种基于GAN的多任务视角转换模型,该文首次将流形学习的概念引入步态视图特征转换,同时还提出了一种新的步态模板周期能量图(PEI)。

2 基于生成对抗网络的步态特征提取

2.1 算法整体框架

图1给出了整个算法的流程图。首先,所有测试集、训练集和验证集的步态视频序列都被统一处理成同一种步态模板。对于测试集中的一个 α 视角下的测

试样本 p^a , 验证集是含有 n 个样本的 b 视角下的某种行走状态的样本集合 $\{g_1^b, g_2^b, \dots, g_n^b\}$, 假定 p^a 一定与上述验证集中的某个样本具有相同身份, 我们的目标就是确定 p^a 的身份是 $\{1, 2, \dots, n\}$ 中的哪一个. 具体过程是, 测试样本的步态模板 p^a 首先经过一个编码器 E 编码得到隐特征 $z^p = E(p^a)$, 然后经过一个视角转换器 V 把 z^p 从视角 a 转换到视角 b , 得到视角转换后的特征隐表示 $z^s = V(z^p, a, b)$. 将 z^s 以及行走状态的标签联合输入到生成对抗网络中的生成器, 得到视角 b 下的正常行走状态的假特征模板的隐表示 \hat{z}^s . 在整个算法模型中, 对生成器和判别器的训练是尤为重要的, 通过对生成器与判别器进行迭代交叉训练, 使得所得到的 \hat{z}^s 在达到视图转换的同时最大化的保留身份信息. 对于验证集中的所有步态模板 g_i^b 同样经过 E 的编码得到隐特征 $z_i^g = E(g_i^b)$, 其中 $i \in \{1, 2, \dots, n\}$. 最终确定 p^a 的身份是视角 b 验证集中的 g_x^b , 利用最近邻分类器的计算公式如下:

$$x = \arg \min \|z^s - z_i^g\|_2, i \in \{1, 2, \dots, n\} \quad (1)$$

其中, $\|\cdot\|_2$ 代表 L2 范数.

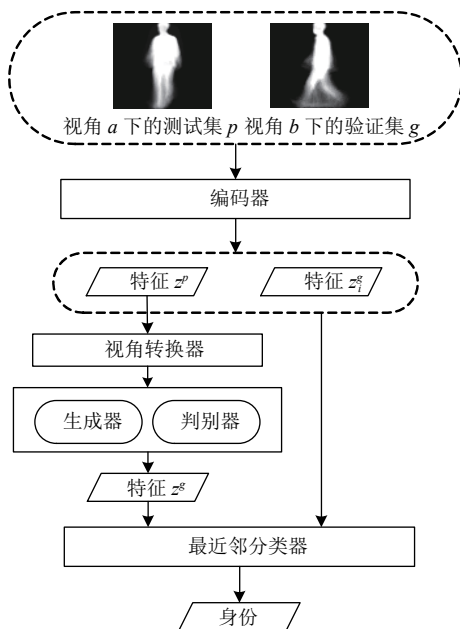


图1 算法流程图

2.2 步态能量图

本文提出的算法使用的步态模板为步态能量图^[14] (GEI). 在还没有运用深度学习的方法研究步态识别问题时, Han 等人在文献^[14]中提出的利用 GEI 和主成分

分析进行数据降维的方法 (GEI+PCA) 在正常行走状态下就取得较好的效果, 只是该方法较易受到衣着、携带物、视角变化等因素的影响. GEI 是步态识别技术中最常用的步态模板, 具有计算简便且能保留原有的步态序列的周期性的特点. 简单来说, 步态能量图就是一个完整周期内的所有步态剪影图的平均, 如图 2 所示.



图2 步态能量图的合成过程

GEI 中的像素值可以解释为一个步态周期内, 人体占据 GEI 中像素位置的概率. 由于 GEI 在步态识别中的成功应用, 本文也采用 GEI 作为输入的步态模板. 本文提取 GEI 的方法与 Zheng 等人在文献^[8]中提出的方法相同.

2.3 基于生成对抗网络的跨视角步态特征提取

本文提出的基于生成对抗网络的跨视角步态特征提取模型主要包含 4 个部分: (1) 基于卷积神经网络的编码器用于对 GEI 中的特征进行编码; (2) 视角转换器用来将一个视角的步态模板转换到另一个视角; (3) 生成器用于生成视角转换之后的假 GEI; (4) 判别器用于判断生成器生成的图像的真实性并判断其是否保留了原有输入的身份信息. 接下来对每部分进行具体介绍.

2.3.1 编码器

为了获得有用的特征进行步态识别, 本文提出的模型中引入了一个基于卷积神经网络的编码器, 具体结构如表 1 所示, 其主要由 4 个卷积层构成. 在特征提取的最初阶段, 输入是 a 视角下的步态能量图 p^a . 为了减少计算量, 提高训练速度, 图片尺寸缩放为 64×64 像素. 之后使用均值池化并经过全连接层得到步态特征隐表示 z^p . 非线性激活层采用 L-ReLU 函数.

表1 编码器的组成

卷积层	卷积核大小	过滤器数量	步长	是否归一化	激活函数
Conv1	7×7	32	2	是	L-ReLU
Conv2	5×5	64	2	是	L-ReLU
Conv3	5×5	128	2	是	L-ReLU
Conv4	3×3	256	2	是	L-ReLU

2.3.2 视角转换器

在跨视角的步态识别问题中, 如何在保留步态模

板的身份特征的情况下转换其视角非常关键. 根据流形学习的理论, 我们所观察到的数据实际上是由一个低维流形映射到高维空间的. 可以假设输入的步态模板位于一个低维的流形上, 沿着该流形移动的样本可以实现视角的转换^[13]. 从视角 a 到视角 b 的转换过程可以用如下公式表示:

$$z^b = z^a + \sum_{i=a}^{b-1} w_i \quad (2)$$

其中, z^a 和 z^b 是经过编码器编码后得到的步态特征隐表示, w_i 表示从视角 $i-1$ 到视角 i 的转换向量. 与传统的视角转换方法不同的是, 本文提出的模型是将视图的特征转换到一个隐空间. 具体的转换过程是通过没有偏置的全连接层来完成的, 从而减少重建视图所积累的错误. 全连接层的权重可以表示为 $W = [w_1, w_2, \dots, w_{n_b}]$, n_b 表示角度的标签, 每个角度对应一个权值. 因此, 编码视角 a 到 b 的过程用向量可以表示为 $e^{ab} = [e_1^{ab}, e_2^{ab}, \dots, e_{n_b}^{ab}]^T$, 其中 $e_i^{ab} \in \{0, 1\}$, 进而视图转换过程可以表示为如下公式:

$$z^b = z^a + W e^{ab} \quad (3)$$

2.3.3 生成器

视角转换器的输出就是生成器的输入, 主要由5个反卷积层构成, 具体参数如表2所示. 生成器的作用是生成与真实 GEI 难以区分的假图片, 但由于步态识别的结果受衣着、携带物等不同行走状态的影响, 所以本文方法中引入了行走状态的独热编码作为标签来控制输出的 GEI 的行走状态为正常状态, 从而消除穿大衣、背包等干扰因素. 定义生成器的输入是 $[z^b, c]$, 其中 z^b 是经过视角转换后的视角 b 下的步态特征隐表示, c 为需要生成的行走状态的独热编码, 通过生成器的反卷积过程得到重构后的步态能量图 y^b .

表2 生成器的组成

反卷积层	卷积核大小	过滤器数量	步长	是否归一化	激活函数
Deconv1	4×4	256	2	是	L-ReLU
Deconv2	3×3	128	2	是	L-ReLU
Deconv3	3×3	64	2	是	L-ReLU
Deconv4	3×3	32	2	是	L-ReLU
Deconv5	3×3	16	2	否	Tanh

2.3.4 判别器

判别器的主要任务有两个: (1) 判断生成图片的真实性; (2) 判别该生成图片的域是否为指定的角度域和

状态域. 与编码器类似, 判别器主要由4个卷积层构成, 具体参数如表3所示. 判别器的输入为生成器生成的假图片 y^b 和真实图 \hat{y}^b , 要注意这里的 \hat{y}^b 应是与生成的 y^b 所指定的角度域、状态域一致的真实 GEI. 通过对生成器和判别器的迭代训练, 可以确保生成器生成足够真实且保留原有身份信息的假 GEI.

表3 判别器的组成

卷积层	卷积核大小	过滤器数量	步长	是否归一化	激活函数
Conv1	7×7	32	2	是	L-ReLU
Conv2	5×5	64	2	是	L-ReLU
Conv3	5×5	128	2	是	L-ReLU
Conv4	3×3	256	2	是	L-ReLU

2.4 目标函数

2.4.1 重构损失

为了使生成图像尽可能多的保留原始身份信息, 最为重要的就是将真实图与生成图的逐像素损失最小化, 因此提出了重建损失, 公式如下:

$$\min_{E,V,G} L_{\text{rec}} = E \left\| G(V(E(p^a), a, b), c) - \hat{y}^b \right\|_1 \quad (4)$$

其中, $\|\cdot\|_1$ 代表 L1 范数, E 、 V 、 G 分别表示编码器, 视角转换器和生成器, p^a 是 a 视角下的原始输入 GEI, c 表示行走状态的独热编码, \hat{y}^b 是真实的 GEI.

2.4.2 对抗损失

对于生成器和判别器的训练, 我们采用的是常见的对抗损失, 公式如下:

$$\min_G \max_D L_{\text{adv}} = E \left[\log D(\hat{y}^b) \right] + E \left[\log(1 - D(G(V(E(p^a), a, b), c))) \right] \quad (5)$$

其中, E 、 V 、 G 、 D 分别代表编码器、视角转换器、生成器和判别器, p^a 是 a 视角下的原始输入 GEI, c 表示行走状态的独热编码, \hat{y}^b 是真实的 GEI.

2.4.3 目标损失函数

最终的关于 E 、 V 、 G 、 D 的最终目标损失函数定义如下:

$$\min_{E,V,G} \max_D L = L_{\text{rec}} + \lambda L_{\text{adv}} \quad (6)$$

其中, λ 是用来调节重建损失和对抗损失的所占比重的超参数, 在本文提出的算法训练过程中, $\lambda=0.0001$. 定义最终损失函数后, 使用反向传播算法对编码器、视角转换器、生成器和判别器进行训练.

3 实验与分析

3.1 数据集

3.1.1 CASIA-B 数据集

CASIA-B 步态数据集^[15]是中国科学院自动化研究所提供的开源跨视角数据集,其中包括 124 人(31 女, 93 男),拍摄视角有 11 个(从 0°到 180°,每 18 度为一个间隔),三种行走状态.每人有 10 个视频序列,6 个为正常行走状态(nm01-nm06),2 个为穿大衣行走状态(cl0, cl02),2 个为背包行走状态(bg01, bg02).

3.1.2 OUMVLP 数据集

OUMVLP 数据集^[16]是目前规模最大的多视角步态数据集,由日本大阪大学提出,其中包括 10 307 个人(5114 男和 5193 女),年龄范围从 2 到 87 岁,每人有 14 个视角(0°, 15°, ..., 90°, 180°, 195°, ..., 270°)和两个视频序列(#00 作为验证集,#01 作为测试集).

3.2 实验设计

对于 CASIA-B 数据集,实验中我们规定前 62 人(ID: 001-062)作为训练集,后 62 人(ID: 063-124)作为验证集,并将测试集分为正常行走(nm)、背包(bg)、穿大衣(cl)及 3 种状态混合(all)共 4 类,分别进行 4 组实验,具体见表 4.训练过程的迭代次数为 80 000 次,学习率为 1×10^{-5} .

表 4 CASIA-B 实验设计

数据集	ID	视频序列
训练集	001-062	nm01-nm06, bg01, bg02, cl01, cl02
验证集	062-124	nm01-nm04
测试集	063-124	nm: nm05-nm06
		bg: bg01, bg02
		cl: cl01, cl02
		all: nm05, nm06, bg01, bg02, cl01, cl02

对于 OUMVLP 数据集,为了验证模型的鲁棒性,便于与 CASIA-B 数据集做比较,实验中我们从所有人中随机抽取 62 人作为训练集,再从剩下的人中随机选择几组(62 人为一组)作为验证集(#01)和测试集(#00).

3.3 实验结果与分析

经过在 CASIA-B 数据集上的实验,3 种行走状态下与所有情况下的平均 rank-1 准确率如表 5 所示.

为了对实验结果更加直观的进行分析,我们选择了 3 种在非跨视角情况下取得较好识别效果的方法(GEI-PCA^[14], SPAE^[10], GaitGAN^[12]) 在同样数据集上的平均 rank-1 准确率进行了对比.图 3 为测试集与验证

集视角相同条件下识别准确率的柱状图,从图中可以看出在 nm 和 bg 状态下,本文提出的方法与其他 3 种方法没有较大优势,但在 cl 状态下,识别准确率有所提升,验证了本文方法对于衣着这个干扰因素具有更好的适应性.

表 5 3 种行走状态下的平均 rank-1 准确率(单位: %)

测试集角度(°)	nm	bg	cl	all
0	63.70	47.47	39.38	50.18
18	76.71	59.39	46.97	61.02
36	79.97	53.64	40.74	58.12
54	83.80	56.19	31.86	57.28
72	76.09	48.74	27.26	50.70
90	70.13	45.52	21.44	45.70
108	73.65	48.53	25.11	49.10
126	79.28	44.72	23.93	49.31
144	75.32	40.47	30.40	48.73
162	72.47	49.38	32.51	51.45
180	62.13	42.91	29.11	44.72
平均值	73.93	48.81	31.70	51.48

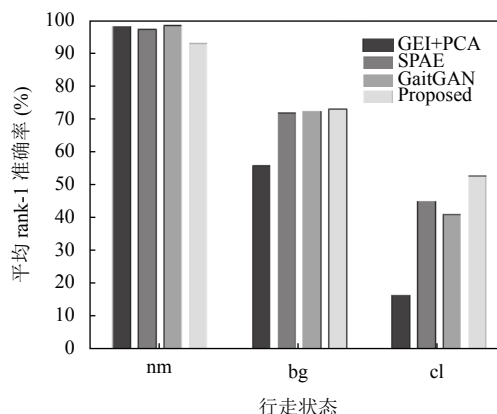


图 3 同视角下的平均 rank-1 准确率比较

为了更好验证本文方法是否在跨视角步态识别上更具优势,在排除相同视角的情况下,我们选取了 54°、90°和 126°三种典型视角与 deepCNNs^[4]、SPAEL^[10]、GaitGAN^[12]及 MGANs^[13]4 种方法进行了比较,这 4 种方法代表了近年来利用深度学习进行跨视角步态识别研究的最先进成果.具体结果如表 6 所示.

由于几年来在 OU-MVLP 数据集上的实验数据不多,所以下面只选择了 GEINet^[3]与本文提出的实验结果进行比较,如表 7 所示.

综合 2 个数据集上的实验结果可以得出结论,在测试集与验证集视角不一致的情况下,本文提出的方法的平均 rank-1 准确率具有大幅度的提升,因此,该方法能较好地解决跨视角步态识别问题.

表6 CASIA-B 排除相同视角下3种典型视角的识别准确率(单位: %)

方法	54°	90°	126°	平均
deepCNNs ^[4]	77.8	64.9	76.1	72.9
SPAE ^[10]	63.3	62.1	66.3	63.9
GaitGAN ^[12]	64.5	58.2	65.7	62.8
MGANS ^[13]	77.1	67.3	78.6	70.1
Proposed	82.2	67.9	77.0	78.0

表7 OU-MVLP 排除相同视角下的 rank-1 准确率(单位: %)

测试集角度(°)	GEINet ^[3]	Proposed
0	11.4	75.5
15	29.1	71.1
30	41.5	72.0
45	45.5	66.1
60	39.5	63.8
75	41.8	78.8
90	38.9	83.5
180	14.9	72.8
195	33.1	76.9
210	43.2	76.7
225	45.6	71.1
240	39.4	80.7
255	40.5	77.5
270	36.3	73.6
平均值	35.8	74.2

3.4 模型评估

结合 3.3 部分的实验结果, 不难发现, 在相同视角的步态识别问题上, 本文提出的模型虽然不是最优但并没有显著劣势, 且在穿着大衣的干扰因素下与其他方法相比具有更好的效果. 在跨视角步态识别问题上, 本文方法的平均 rank-1 准确率有了大幅度的提升. 这主要由于本文提出的模型在同视角识别的状态下, 与传统深度学习方法相比没有做出过多改进, 而在跨视角识别问题和携带物、衣着变化上则具有一定优势.

此外, 随着行走状态的改变, 从表 5 可以看出, 3 种情况对于步态识别准确率的影响程度是 $cl > bg > nm$, 这与以往研究结果一致. 这是由于步态识别最重要的部分在于提取腿部特征, 所以在穿着大衣的情况下能保留的腿部有效信息是最少的, 所以该情况下识别准确率会收到较大影响. 对于跨视角步态识别问题, 观察表 5 不难发现, 视角在 90°左右的识别效果较好, 而在 0°和 180°附近时识别效果较差. 这是由于在 90°状态下, 可以提取到的腿部变化的信息是最多的, 随着跨度增大, 提取的有效信息逐渐变少, 从而导致不同视角下的识别准确率差异较大.

综上所述, 本文提出的模型能较好的解决跨视角步态识别问题, 且在行走状态变化情况下具有一定的鲁棒性, 实验结果也验证了该方法是有效且可行的.

4 结论与展望

本文提出了一种基于生成对抗网络的步态特征提取方法, 能有效缓解拍摄视角、衣着和携带物变化等因素对步态识别结果的影响, 在 CASIA-B 和 OUMVLP 数据集上的实验结果验证了本文提出的方法可以对步态模板进行任意角度的转化并提取出包含有效身份信息的步态特征. 同时, 与现有的其他方法的比较也表明本文的方法是有效且可行的.

然而, 目前所做的工作还不够, 实验所用的数据集的规模还比较小, 识别准确率也还不够高. 今后, 在训练阶段要使用更大规模的数据集, 使用网络结构更复杂且更具鲁棒性的模型. 相信未来生成对抗网络的发展会更好, 使得包括步态识别在内的其他问题取得新的研究突破.

参考文献

- 何逸炜, 张军平. 步态识别的深度学习: 综述. 模式识别与人工智能, 2018, 31(5): 442-452.
- Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada. 2014. 2672-2680.
- Shiraga K, Makihara Y, Muramatsu D, et al. GEINet: View-invariant gait recognition using a convolutional neural network. Proceedings of 2016 International Conference on Biometrics. Halmstad, Sweden. 2016. 1-8.
- Wu ZF, Huang YZ, Wang L, et al. A comprehensive study on cross-view gait based human identification with deep CNNs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(2): 209-226. [doi: 10.1109/TPAMI.2016.2545669]
- Liao RJ, Cao CS, Garcia EB, et al. Pose-based temporal-spatial network (PTSN) for gait recognition with carrying and clothing variations. Proceedings of the 12th Chinese Conference on Biometric Recognition. Shenzhen, China. 2017. 474-483.
- Feng Y, Li YC, Luo JB. Learning effective gait features using LSTM. Proceedings of the 2016 23rd International Conference on Pattern Recognition. Cancun, Mexico. 2016. 325-330.

- 7 Makihara Y, Sagawa R, Mukaigawa Y, *et al.* Gait recognition using a view transformation model in the frequency domain. Proceedings of the 9th European Conference on Computer Vision. Graz, Austria. 2006. 151–163.
- 8 Zheng S, Zhang JG, Huang KQ, *et al.* Robust view transformation model for gait recognition. Proceedings of the 2011 18th IEEE International Conference on Image Processing. Brussels, Belgium. 2011. 2073–2076.
- 9 Bashir K, Xiang T, Gong SG. Cross-view gait recognition using correlation strength. Proceedings of 2010 British Machine Vision Conference. London, UK. 2010. 1–11.
- 10 Yu SQ, Wang Q, Shen LL, *et al.* View invariant gait recognition using only one uniform model. Proceedings of the 2016 23rd International Conference on Pattern Recognition. Cancun, Mexico. 2016. 889–894.
- 11 Choi Y, Choi M, Kim M, *et al.* StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 8789–8797.
- 12 Yu SQ, Chen HF, Reyes EBG, *et al.* GaitGAN: Invariant gait feature extraction using generative adversarial networks. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu, HI, USA. 2017. 532–539.
- 13 He YW, Zhang JP, Shan HM, *et al.* Multi-task GANs for view-specific feature learning in gait recognition. IEEE Transactions on Information Forensics and Security, 2019, 14(1): 102–113. [doi: [10.1109/TIFS.2018.2844819](https://doi.org/10.1109/TIFS.2018.2844819)]
- 14 Han J, Bhanu B. Individual recognition using gait energy image. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(2): 316–322. [doi: [10.1109/TPAMI.2006.38](https://doi.org/10.1109/TPAMI.2006.38)]
- 15 Yu SQ, Tan DL, Tan TN. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. Proceedings of the 18th International Conference on Pattern Recognition. Hong Kong, China. 2006. 441–444.
- 16 Takemura N, Makihara Y, Muramatsu D, *et al.* Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. IPSJ Transactions on Computer Vision and Applications, 2018, 10(1): 4. [doi: [10.1186/s41074-018-0039-6](https://doi.org/10.1186/s41074-018-0039-6)]