

基于深度学习的步态识别 关键技术研究



答辩人：张 诚

导 师：马华东 教授

2016.3.10

提 纲

- 研究背景与意义
- 基于传统卷积神经网络的步态识别技术
- 步态识别中的距离度量学习模型
- 基于时空联合深度神经网络的步态识别技术
- 工作总结与展望

提 纲

□ 研究背景与意义

- 基于传统卷积神经网络的步态识别技术
- 步态识别中的距离度量学习模型
- 基于时空联合深度神经网络的步态识别技术
- 工作总结与展望

步态的定义



- 步态，即人行走的姿态。包括行走过程中上肢、下肢关节处呈现出有规律的运动趋势和变化。步态识别是一种生物识别技术，通过对步态的分析，可以得到身份和属性（性别、年龄、种族等）信息。
- **Gait analysis** is the systematic study of animal locomotion, more specifically the study of **human motion**, using the eye and the brain of observers, augmented by instrumentation for measuring **body movements**, **body mechanics**, and the **activity of the muscles**.



选题背景

- 现代智能监控/安防系统：“以人中心”
- 目标识别技术：外观/形态特征→融合生物特征
 - 对行人进行识别时，由于人体具有非刚性，外观在不同环境下变化剧烈，仅仅依靠外形特征难以识别，而诸如步态这种生物特征却十分有效。

不同场景监控视频样例



《碟中谍5》中的步态识别“黑科技”



步态特征的优势

□ 生物识别技术

■ 可远程获取

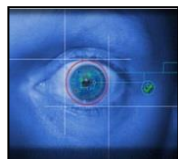
可以通过一定距离获取特定目标的步态信息，采用非接触性的方式进行隐蔽采集。（指纹、虹膜）

■ 鲁棒性

在低质监控画面下仍然有较好的识别效果。（人脸、声纹）

■ 安全性

难以模仿，在公共场合行人要刻意改变自己的步态，往往会变得更加可疑而引起注意。



虹膜识别



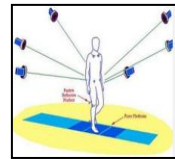
人脸识别



指纹识别



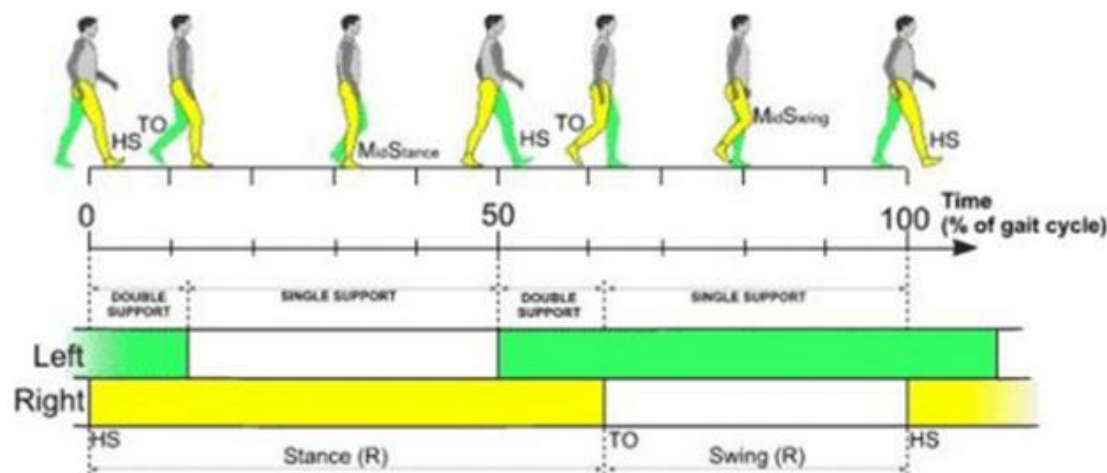
声纹识别



步态识别

步态模式的唯一性

□ 步态特征可识别性的生理学解释



人体关节点示例



步态是一种可识别的
周期性运动模式

[1] Murray M P, Drought A B, Kory R C. *Walking patterns of normal men*[J]. The Journal of Bone & Joint Surgery, 1964, 46(2): 335-360.

[2] Johansson G. Visual perception of biological motion and a model for its analysis[J]. Attention, Perception, & Psychophysics, 1973, 14(2): 201-211.

难点

- 不同行人之间的类间差异太小 (Intra-class)
 - 如何在步态识别场景下设计更加鲁棒的特征



不同行人侧面90° 视角特征轮廓图

- 相同行人不同场景下的类内差异太大 (Inter-class)

- 行走速度
- 观察视角
- 衣着
- 佩戴

相同行人不同穿戴示意图

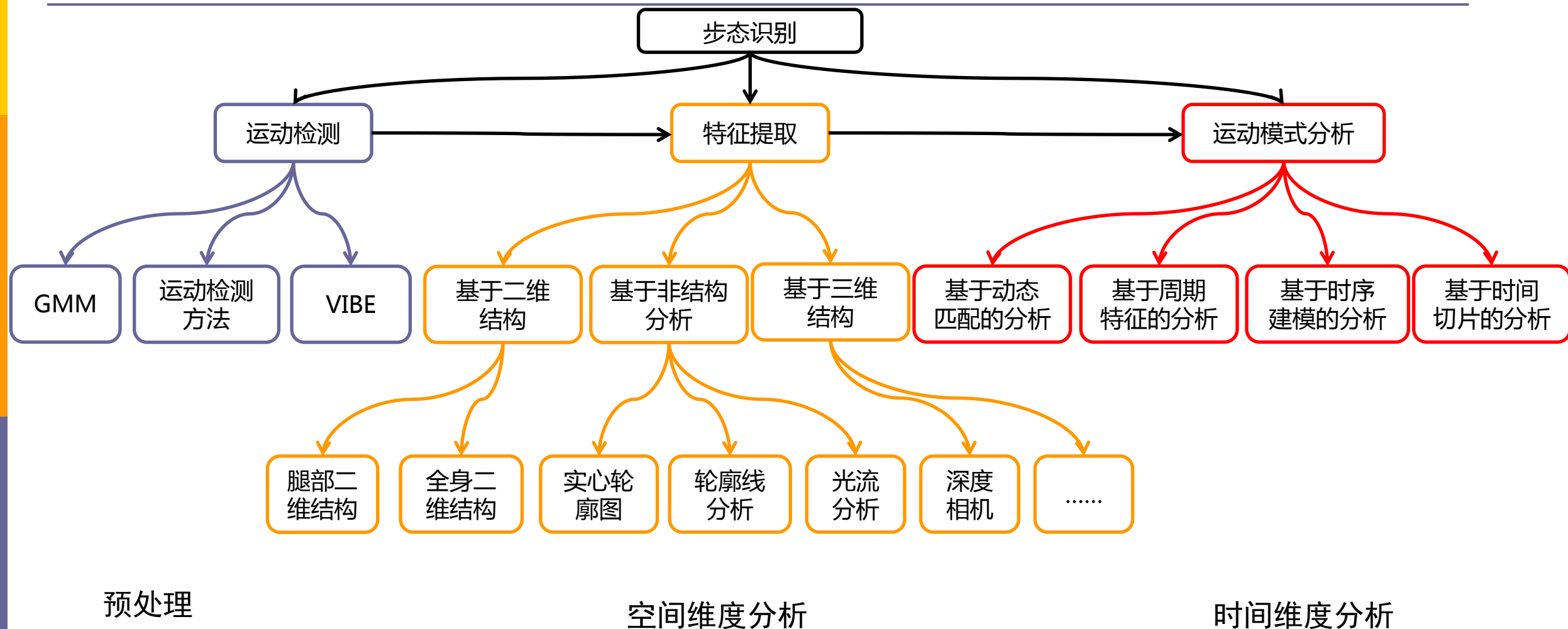


正常

大衣

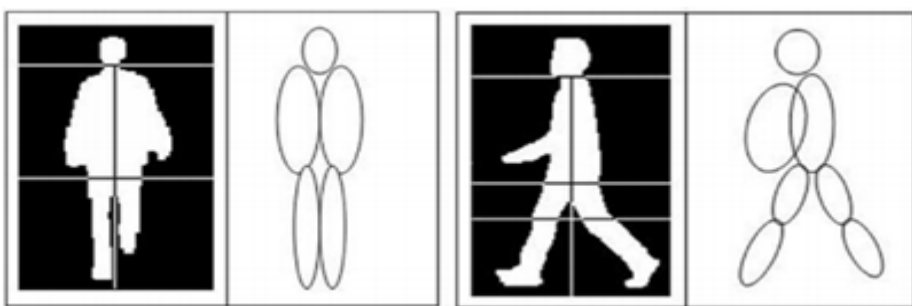
背包

研究现状



存在的问题

- 基于模型匹配的方法 (Model-based)
 - 从步态序列中抽取人体结构特征进行匹配
 - 对采集图像分辨率要求较高



椭圆拟合结果 [Huang et al]

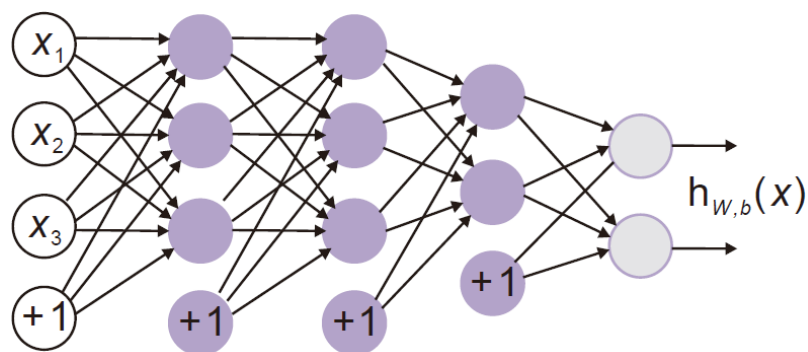


全身结构模型提取结果 [Wang et al]

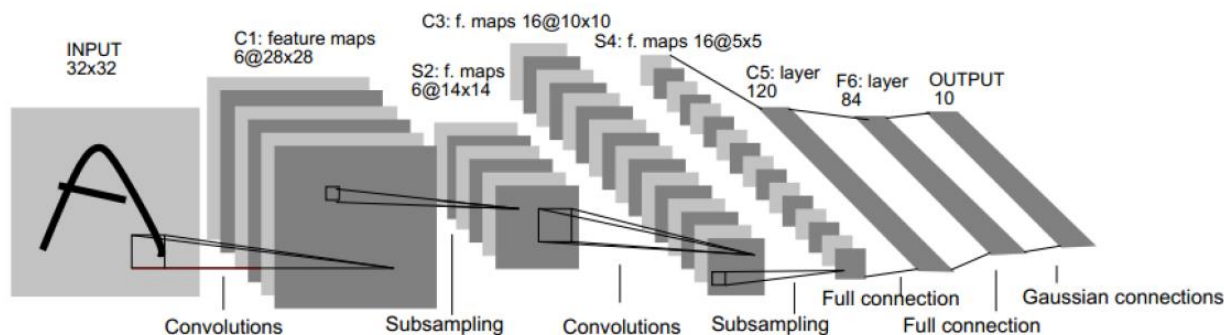
- 基于外观匹配的方法 (Appearance-based)
 - 适合于户外场景步态识别，对采集图像分辨率要求有所降低
 - 如何选择合适的特征是一大挑战 (SIFT, HOG, LBP等)
 - 人工特征很难打破特征表示的瓶颈

深度学习 (Deep Learning)

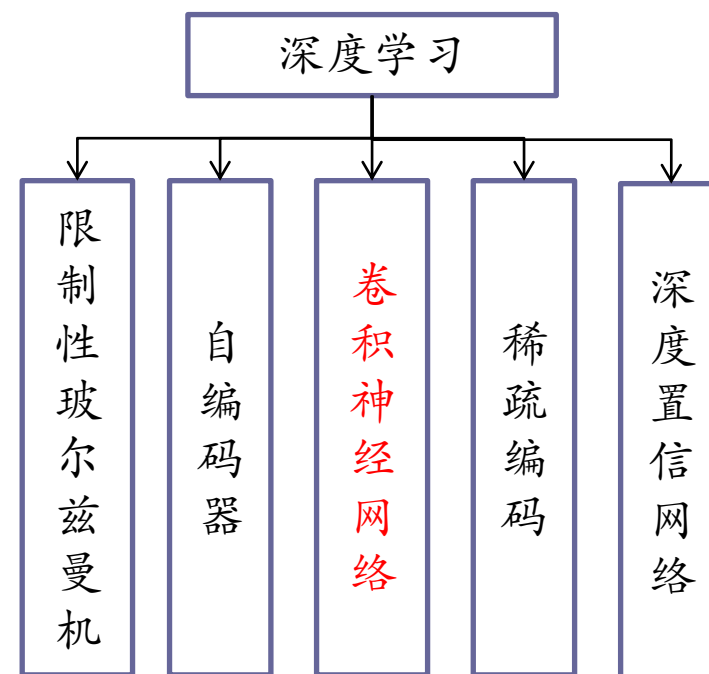
- 2012年以来，深度模型在语音、视觉和自然语言处理领域取得了巨大的成功



BP神经网络

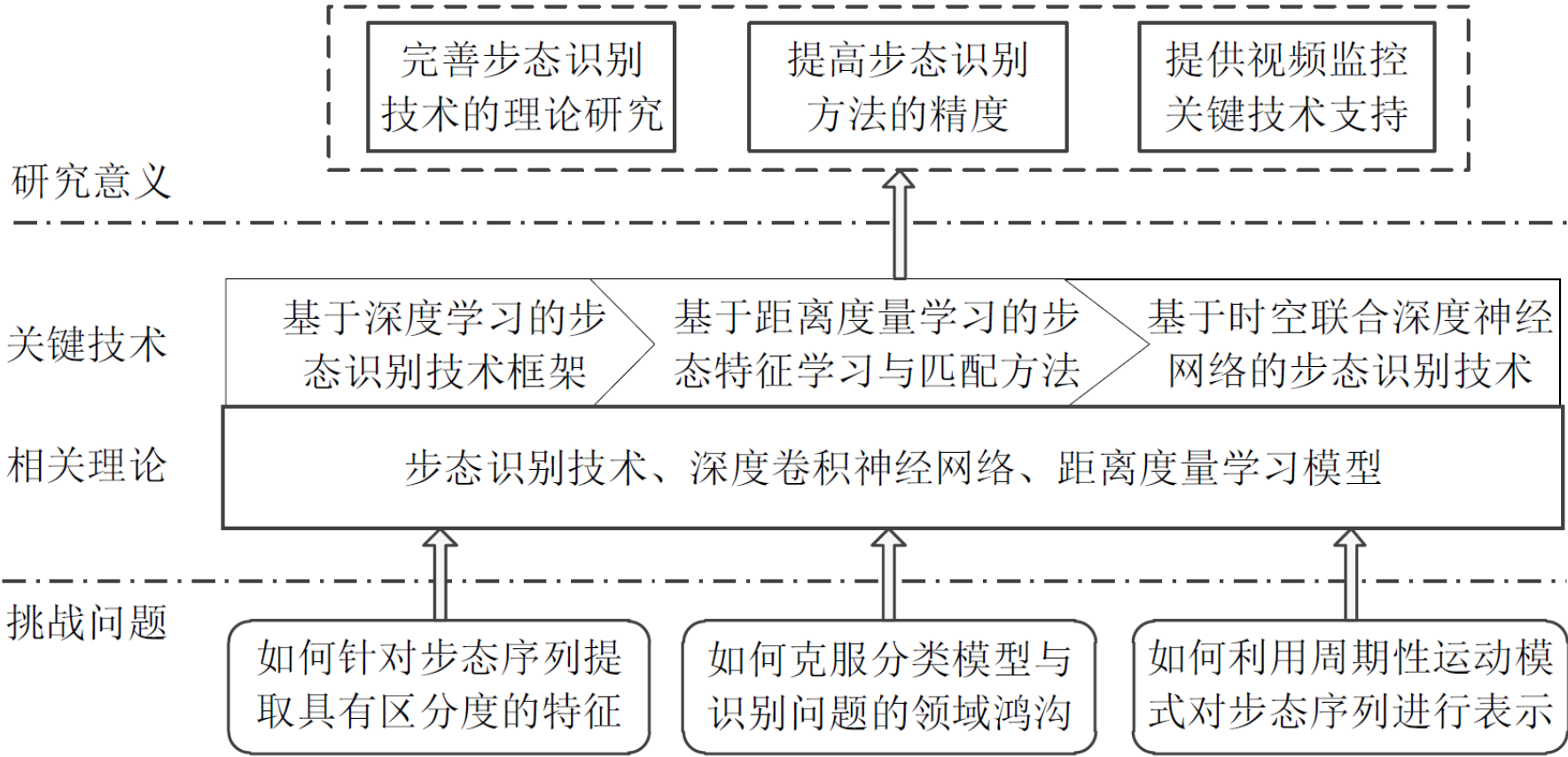


卷积神经网络LeNet

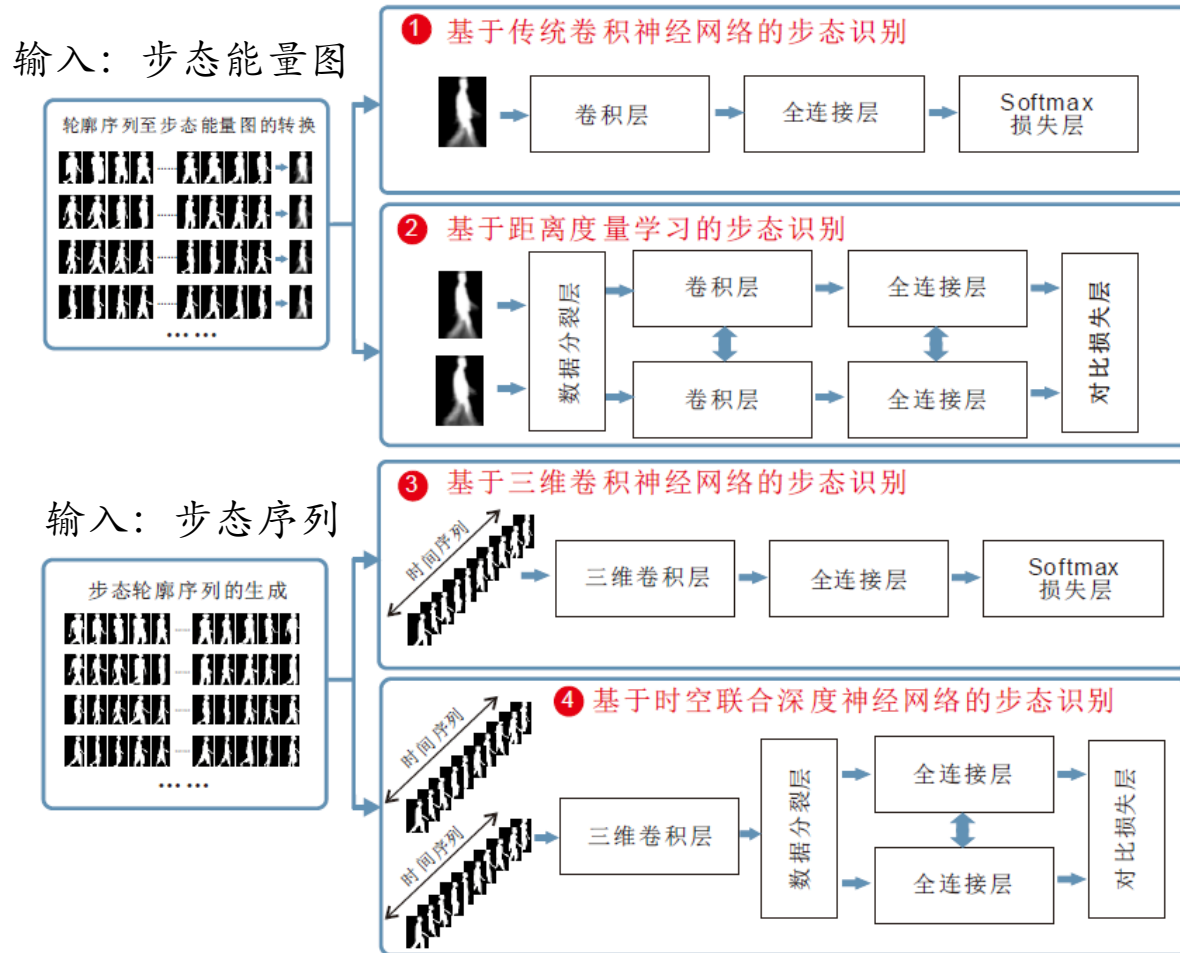


研究框架

基于深度学习的步态识别关键技术研究



技术路线图



领域鸿沟（分类 vs 问题）



Siamese Network 能够解决目标识别场景下的特征度量问题

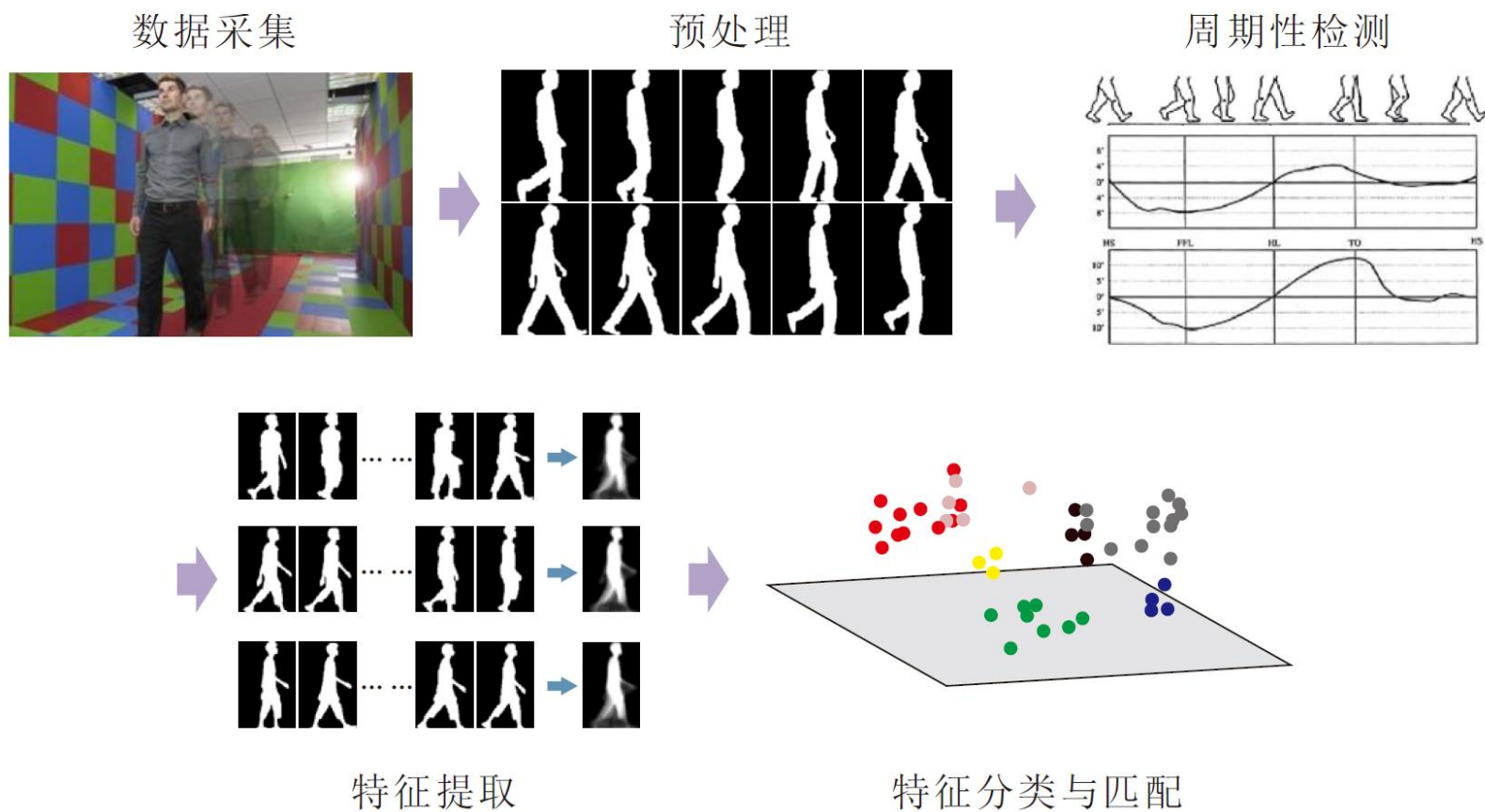


3D-Conv. Network 能够提取步态序列特征，充分利用步态的周期性特征辅助识别。

提 纲

- 研究背景与意义
- 基于传统卷积神经网络的步态识别技术
- 步态识别中的距离度量学习模型
- 基于时空联合深度神经网络的步态识别技术
- 工作总结与展望

步态识别的基本框架

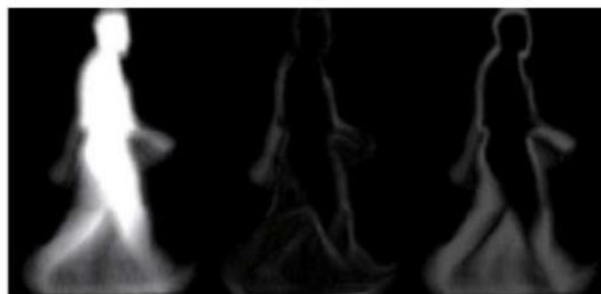


步态特征图的生成

- 目前广泛使用的步态特征包括以下几种：
 - GEI (Gait Energy Image)是最为有效的特征图



步态能量图



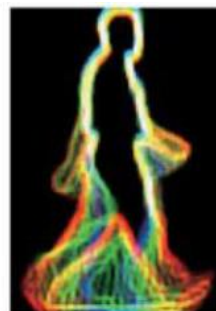
步态频繁域图



步态熵



步态掩模



步态时序图



步态流图

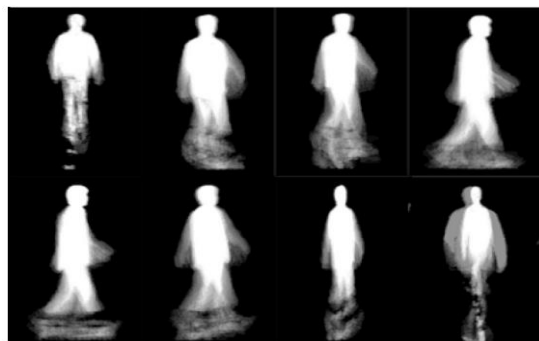
步态能量图

- 对一个周期内的步态序列进行加权平均
 - 计算简单
 - 最大化保留运动的时序信息
 - 对于单帧轮廓分割不佳带来的噪声具有很强的鲁棒性

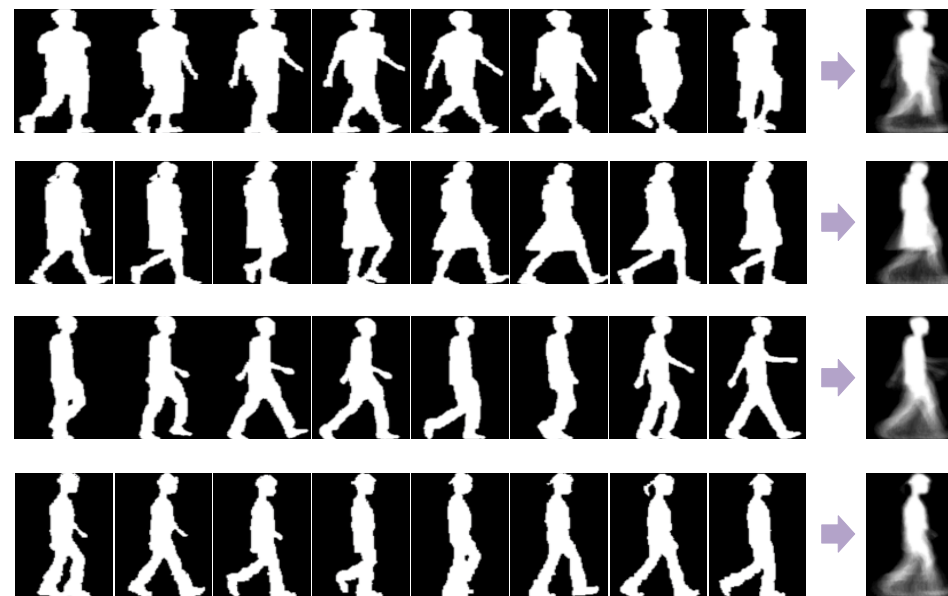
$$G(x, y) = \frac{1}{N} \sum_{t=1}^N I(x, y, t)$$



GEI的合成



不同视角GEI

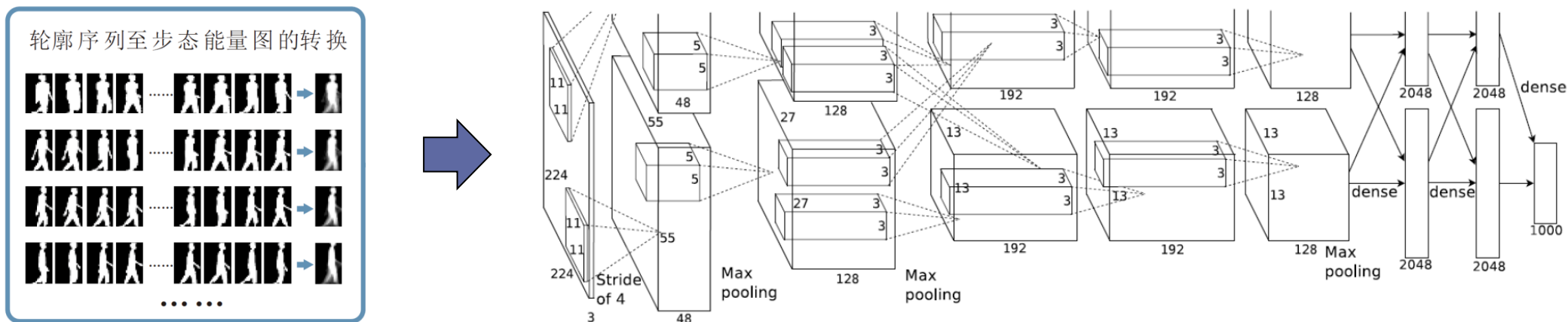


GEI的合成

基于传统CNN的步态特征提取

□ 实现方法：用分类模型来处理识别问题

- 网络配置：AlexNet
- 初始化模型：Caffe model
- 特征输出层类别：3835 subjects in OU-ISIR LP database
- 策略：参数微调，Fine-tuning
- 参数设置：固定卷积层参数，优调全连接层参数
- 特征选择：抽取CNN.FC1层参数，使用K近邻法进行行人匹配



基于传统CNN的步态特征提取

基于深度学习的层次化步态特征提取

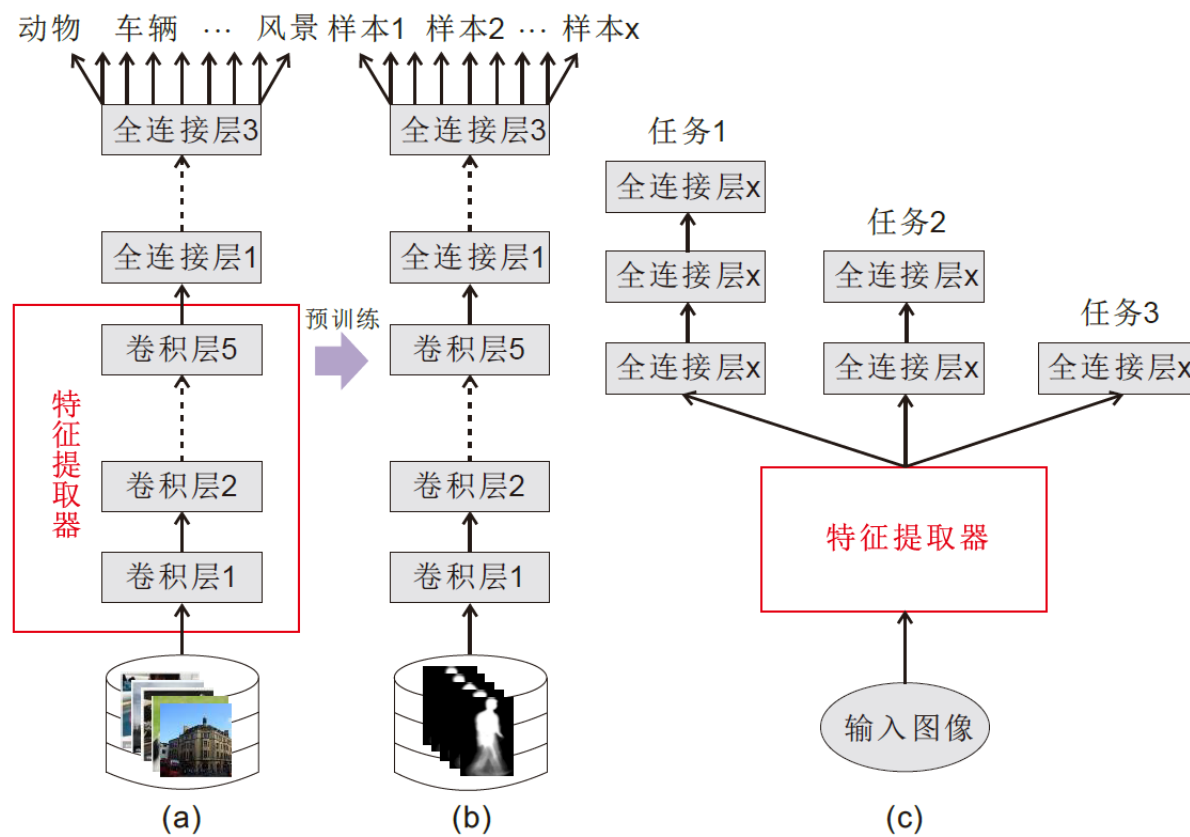


图 3-4 基于深度学习的模型训练与层次化特征提取框架。(a) 预训练的通用基础大类识别模型, (b) 经过参数微调后领域专家模型, (c) 特征提取模型。

两大优点

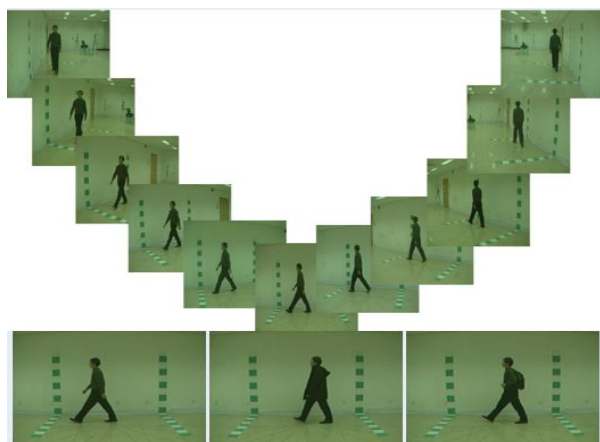
CNN能够从简单的抽象输入图像中自动提取出具有区分性的特征而**不需要了解领域知识**。

Fine-tuning能够**防止过拟合现象**以及随机参数训练可能导致的新模型不收敛问题, 同时**加快训练速度**。

数据库

□ CASIA-B

- 中科院自动化所于2005年1月采集
- 常用步态数据库
- 93名男性和31名女性
- 11个视角
- 3种行走条件（普通、大衣、背包）



□ OU-ISIR Large Population

- 日本大阪大学于2012年采集，持续更新中
- 世界上最大的步态数据库
- 使用范围广泛，最具挑战性的身份识别数据库
- 划分好的查询集和源数据集，适合算法对比

表 3-3 OU-ISIR LP 数据库样本数量统计

视角	55°	65°	75°	86°	All	总计
行人数量	3,706	3,770	3,751	3,249	3,141	3,835
男性/女性	1,977/1,729	2,007/1,763	1,995/1,756	1,688/1,561	1,644/1,497	2,032/1,803



[1] Wang L, Tan T, Hu W, et al. Automatic gait recognition based on statistical shape analysis[J]. IEEE T-IP, 2003, 12(9): 1120-1131.

[2] Iwama H, Okumura M, Makihara Y, et al. The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition[J]. IEEE T-IFS, 2012, 7(5): 1511-1521.

属性识别 vs 身份识别

	属性识别	身份识别
研究目标	粗分类，例如性别、年龄、种族、体型等	精确定位，中远距离、人数较少等场景中 对特定目标的查找和验证
问题模型方法	二分类或者类别较少	识别与验证问题
数据获取代价	数据广泛，统计建模	特定身份行人的多视角数据
解决方法的通用性	运动检测，特征特区，运动模式分析	
挑战	类内差异大、类间差异小	

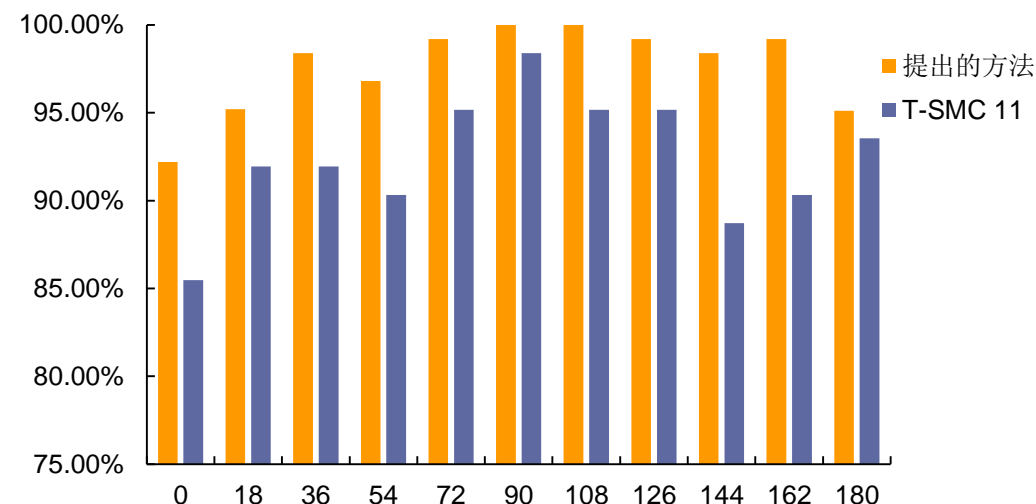
性别分类

- ❑ 数据库：CASIA-B
- ❑ 实验设置：随机选取了31名男性和所有的31名女性普通行走状态步态数据
- ❑ 评价指标：62段样本交叉验证的结果作为正确分类率
- ❑ 流程：前景背景分割 → 周期检测 → 步态能量图的生成 → CNN Fine-tune训练 → Softmax分类

90° 视角下性别分类准确率

方法	准确率
Li et al. [T-SMC 08]	93.28 %
Yu et al. [T-IP 09]	95.97 %
Hu et al. [ICPR 10]	96.77 %
Hu et al. [T-SMC 11]	98.38 %
提出的方法	100.00 %

不同视角下性别分类准确率



身份识别

- 数据库：OU-ISIR Large Population
- 实验设置：只使用Gallery库进行训练，其他测试条件同[1]
- 评价指标：Rank-1和Rank-5正确识别率
- 流程：前景背景分割 → 周期检测 → 步态能量图的生成 → CNN Fine-tune训练 → CNN特征提取 → K近邻查询

Rank-1正确率				
视角/数据	Gallery			
	55°	65°	75°	85°
Probe	55°	73.96 (84.7)	68.94	57.42
	65°	67.19	76.72 (86.63)	72.65
	75°	53.82	69.79	77.87 (86.91)
	85°	40.07	56.76	69.13

Rank-5正确率				
视角/数据	Gallery			
	55	65	75	85
Probe	55	86.64 (92.39)	85.35	79.09
	65	83.26	88.67 (92.84)	86.87
	75	75.47	85.55	89.39 (92.78)
	85	64.20	77.62	85.41

面临的问题

□ 训练数据量太小

- 深度学习需要依赖于庞大的训练数据。为了学习到丰富的特征，CNN需要每一类目中都包含大量的训练样本。与之矛盾的是，在步态识别领域的数据库中，行人的身份数据很多，但是同一样本的数据却很少。
- 步态数据库的数量远远小于ImageNet图像库，步态图像与ImageNet图像存在较大差异。

□ 领域鸿沟

- 分类模型和识别问题存在领域鸿沟。大部分的可以被用于预训练的深度模型都是来源于分类任务，而识别任务与分类任务的具有本质的区别。

提 纲

- 研究背景与意义
- 基于传统卷积神经网络的步态识别技术
- 步态识别中的距离度量学习模型
- 基于时空联合深度神经网络的步态识别技术
- 工作总结与展望

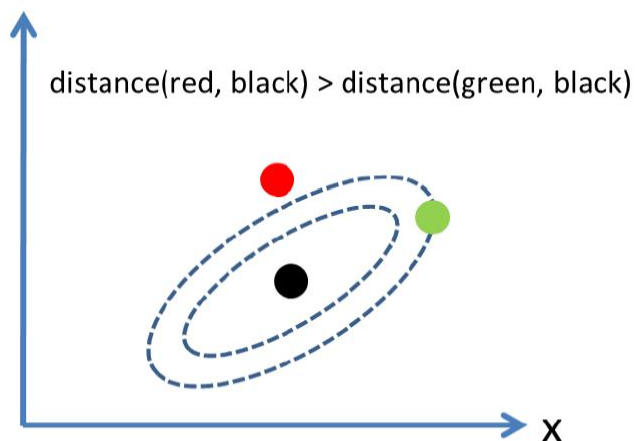
距离度量学习 (Distance Metric Learning)

□ 度量学习基本理论

- 从Euclidean距离到Mahalanobis距离
- 直接使用CNN存在领域鸿沟
- 将深度学习应用于小规模数据集

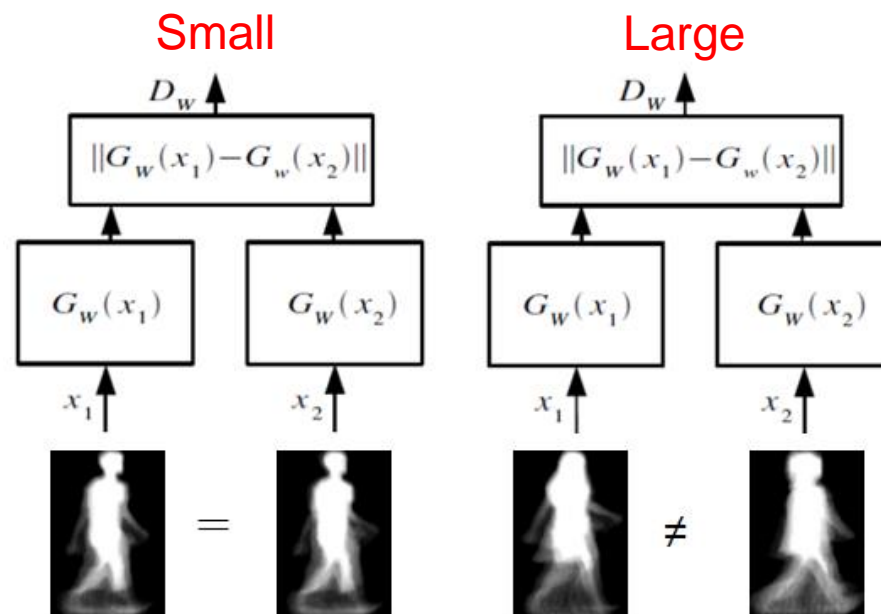
$$d(x, y) = d_A(x, y) = \|x - y\|_A = \sqrt{(x - y)^T A (x - y)}$$

$$\begin{aligned} \min_A \quad & \sum_{(x_i, x_j) \in \mathcal{S}} \|x_i - x_j\|_A^2 \\ \text{s.t.} \quad & \sum_{(x_i, x_j) \in \mathcal{D}} \|x_i - x_j\|_A \geq 1 \\ & A \succeq 0 \end{aligned}$$



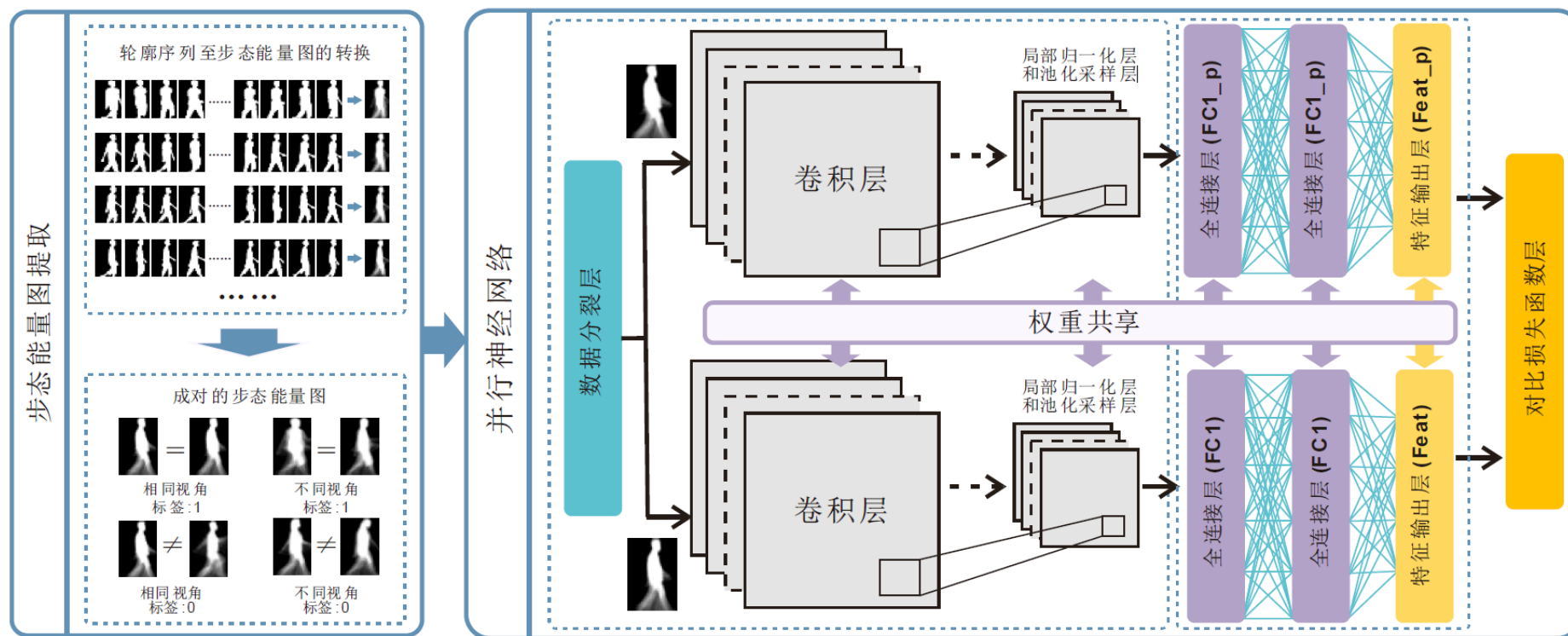
□ 度量学习在计算机视觉中的应用

- 人脸识别
- 手写字母识别
- 相似图像块匹配
- 图像检索中的排序

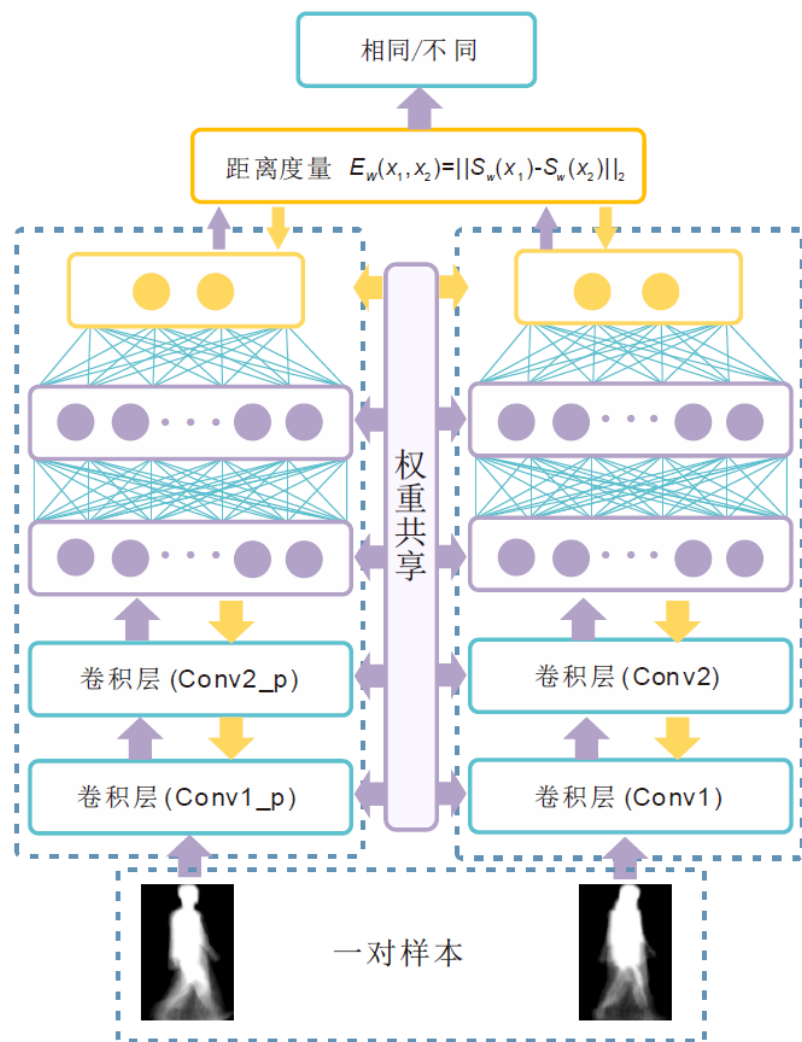


基于Siamese 网络的步态识别技术

- 传统的CNN将步态识别问题转化为行人分类问题，但是二者之间存在巨大的领域鸿沟
- 基于Siamese Neural Network的步态识别整体框架图



本文设计的并行深度神经网络



样本距离度量函数：

$$E_W(x_1, x_2) = \|S_W(x_1) - S_W(x_2)\|_2^2$$

对比损失函数：

$$\mathcal{L}(W) = \sum_{i=1}^P L(W, (y, x_1, x_2)^i)$$

Annotations for the loss function:

- 样本数量 (Sample quantity) points to P .
- 样本标签 (1/0) (Sample label (1/0)) points to y .
- 优化的参数集合 (Optimized parameter set) points to W .

$$L(W, (y, x_1, x_2)^i) = (1 - y) \cdot \max(m - E_W(x_1, x_2)^i, 0) + y \cdot E_W(x_1, x_2)^i$$

Margin

训练样本集的构建

□ 训练数据集是高度不平衡的

- 正负样本集须**保证平衡**，防止偏置（biased）。
- 负样本集须**保证多样性**，防止模型在某一类负样本例上发生过拟合。

□ 具体操作

- 从步态数据库中选择正负样本构建神经网络训练数据集，其中身份相同的两段步态序列为正样本对，同时设置标签为1；身份不同的两段步态序列为负样本对，同时设置标签为0。
- 从OU-ISIR LP数据库中的Gallery集合中抽取所有的20,000对正样本，并且使用了一个Sampler抽取了等量的负样本作为最终的训练集。

实验

- 数据库：OU-ISIR Large Population
- 评价指标：Rank-1和Rank-5正确识别率
- 对比方法：目前最优的步态识别方法，包括：GEI [TIFS12]、FDF[TIFS12]、HWLD [CVPR13]、VTM [TIP15]和RankSVM [ECCV12]
- 流程：前景背景分割 → 周期检测 → 步态能量图的生成 → CNN训练 → CNN特征提取 → K近邻查询

同一视角下步态识别

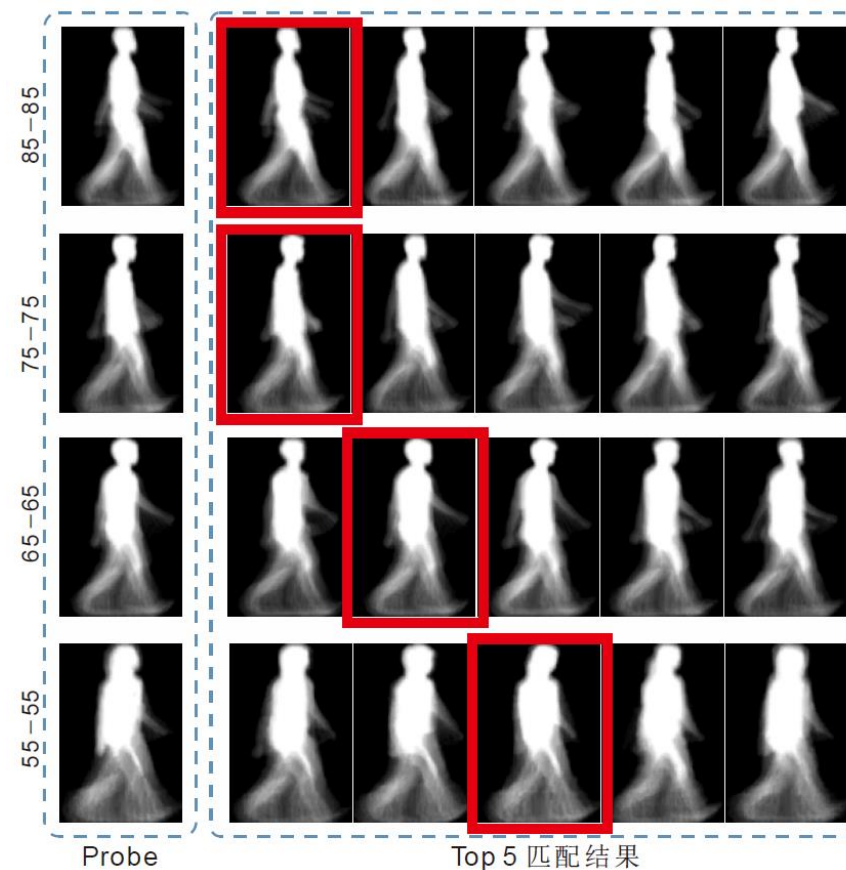
□ 同一视角下步态识别准确率

表 4-2 在 OU-ISIR LP 数据库上 Rank-1 识别准确率

	55°	65°	75°	85°	All
HWLD[98]	—	—	—	87.70	95.50
GEI[47]	84.70	86.63	86.91	85.72	94.24
FDF[47]	83.89	85.49	86.59	85.90	94.17
CNN.FC2	73.96	76.71	77.87	78.82	86.09
SiaNet.FC	90.12	91.14	91.18	90.43	96.02

表 4-3 在 OU-ISIR LP 数据库上 Rank-5 识别准确率

	55°	65°	75°	85°	All
HWLD[98]	—	—	—	94.70	98.50
GEI[47]	92.39	92.84	92.78	93.01	97.13
FDF[47]	91.53	92.81	92.88	92.83	97.10
CNN.FC2	86.64	88.67	89.39	90.09	93.56
SiaNet.FC	94.98	95.90	95.92	95.97	98.31



不同视角下步态识别

□ 交叉视角下步态识别Rank-1准确率

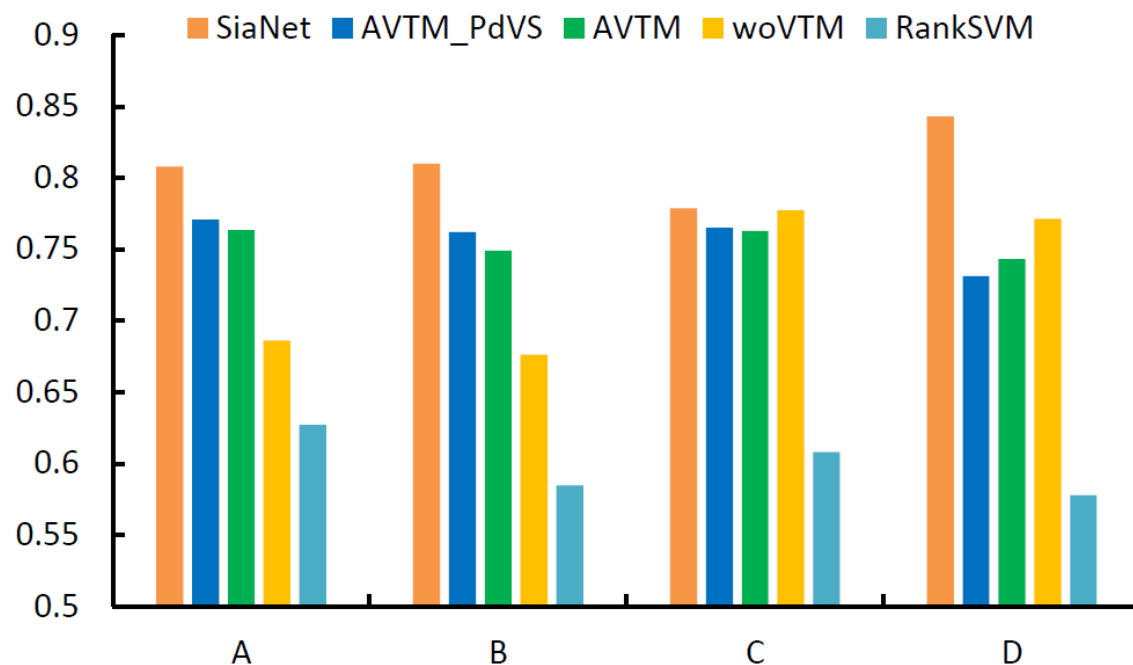
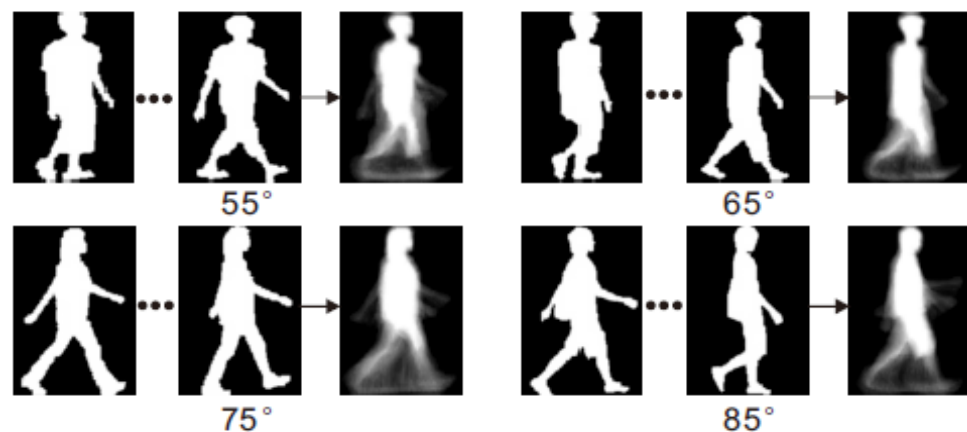


图 4-4 不同视角下步态识别准确率。其中 A~D 组实验分别代表 $(65^\circ, 75^\circ)$ 、 $(75^\circ, 65^\circ)$ 、 $(75^\circ, 85^\circ)$ 和 $(85^\circ, 75^\circ)$ 实验设置

讨论



从视频帧到特征图的转换
存在信息损失



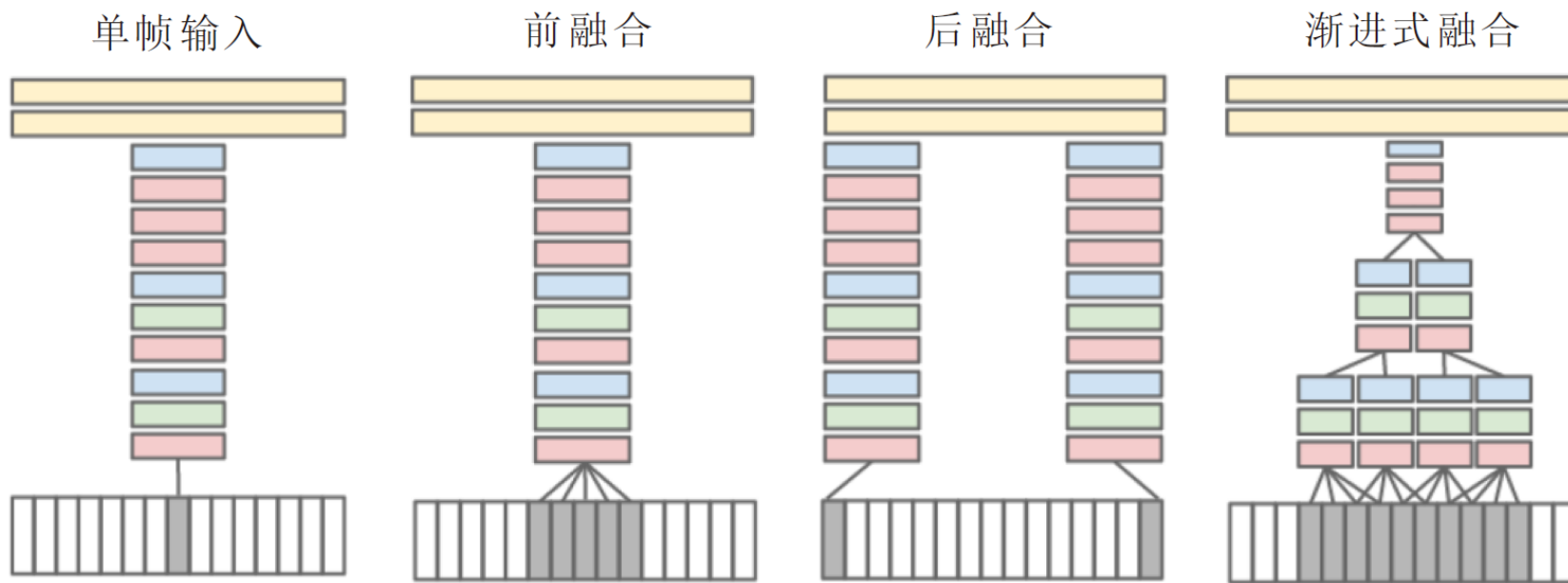
步态是一种周期性运动模式
具有空间和时间两个维度的特征

提 纲

- 研究背景与意义
- 基于传统卷积神经网络的步态识别技术
- 步态识别中的距离度量学习模型
- 基于时空联合深度神经网络的步态识别技术
- 工作总结与展望

基于CNN的视频处理方法

- CNN对视频单帧进行处理的不同方式。



三维卷积神经网络

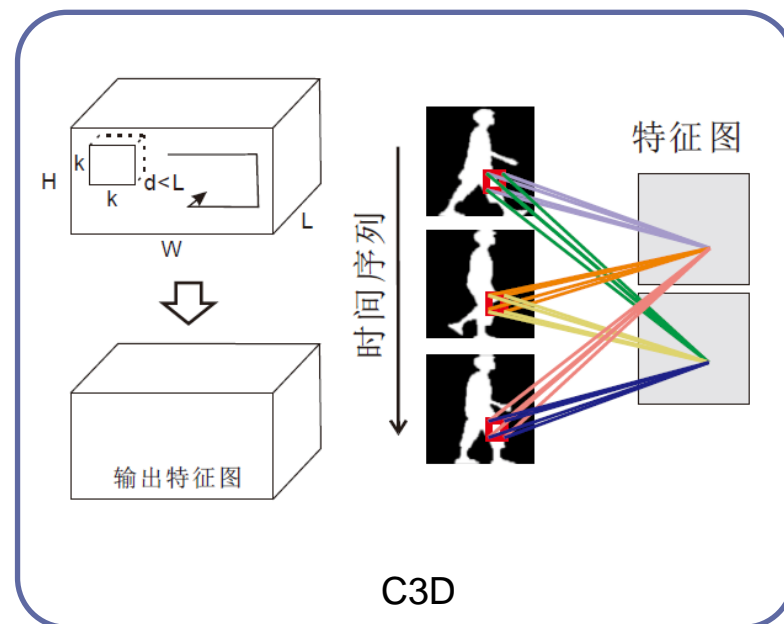
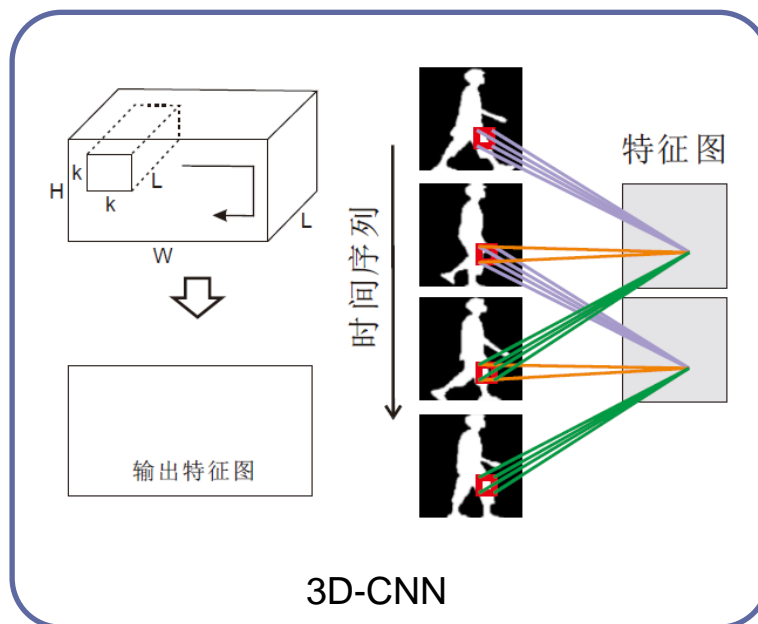
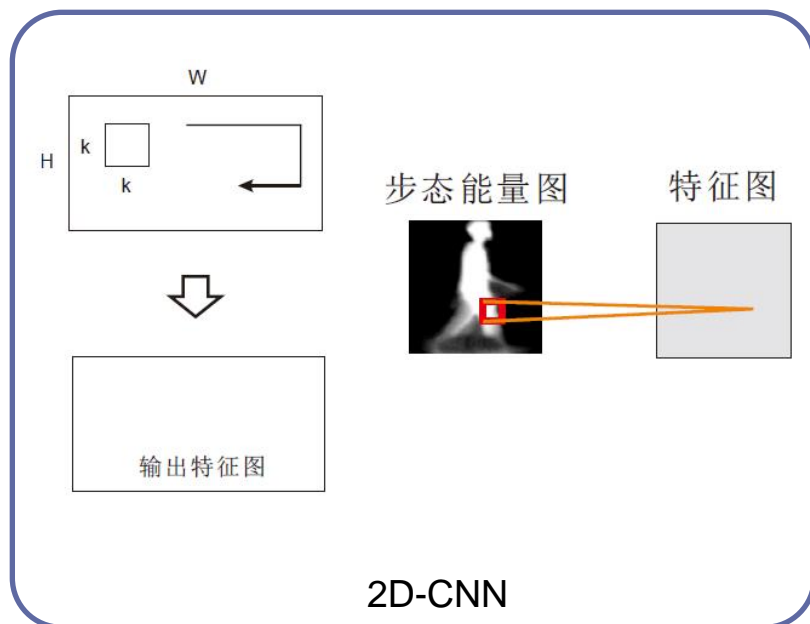
- C3D能够利用步态的周期性模式，直接基于步态序列提取更为丰富的特征。

二维卷积特征图的计算：

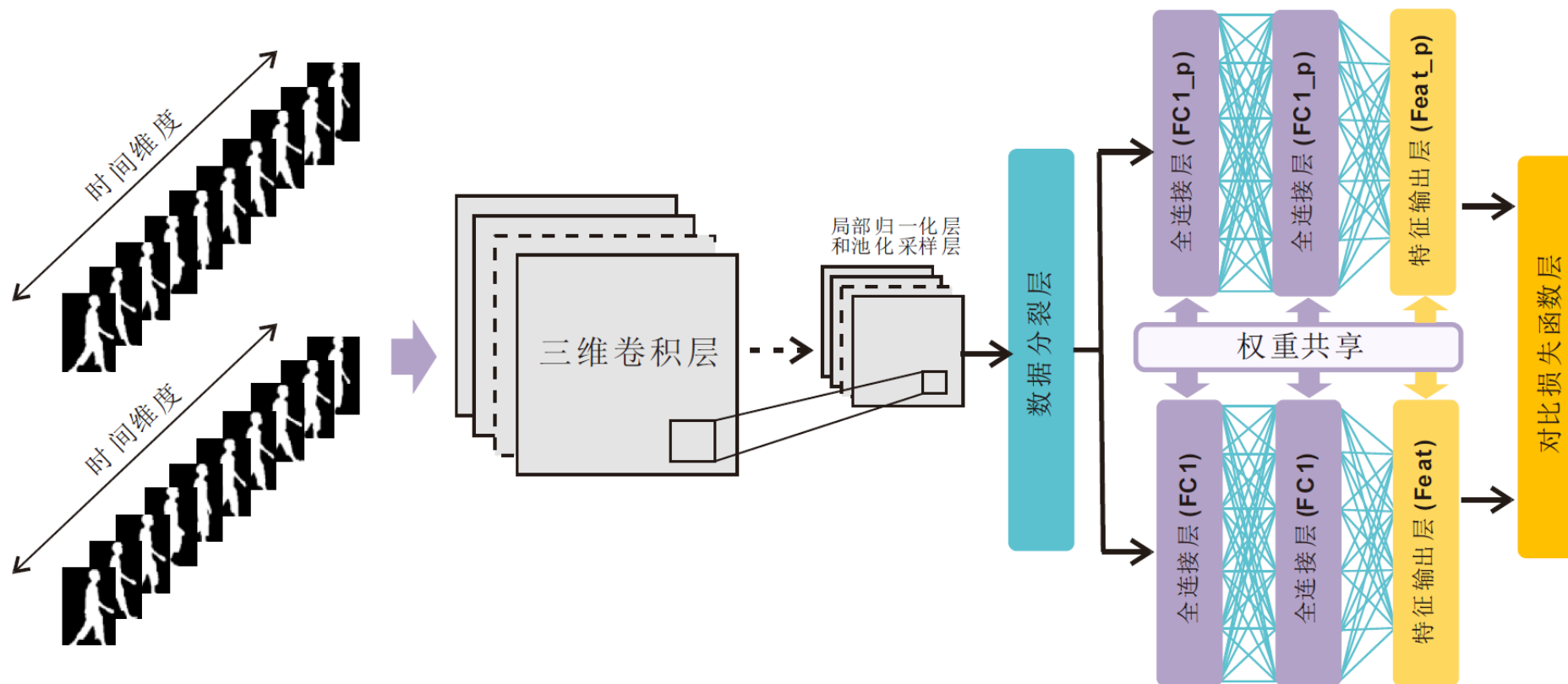
$$v(x, y) = f(b + \sum_{w=0}^{W-1} \sum_{h=0}^{H-1} C \cdot v(x+w, y+h))$$

三维卷积特征图的计算：

$$v(x, y, z) = f(b + \sum_{d=0}^D \sum_{w=0}^{W-1} \sum_{h=0}^{H-1} C \cdot v(x+w, y+h, z+d))$$



基于3D-Siamese结构的步态识别技术



实验

- 数据库：OU-ISIR Large Population
- 评价指标：Rank-1和Rank-5正确识别率
- 对比方法：目前最优的步态识别方法，包括：GEI [TIFS12]、FDF[TIFS12]和HWLD [CVPR13]
- 流程：前景背景分割 → 周期检测 → 3D-Siamese训练 → CNN特征提取 → K近邻查询

实验结果

在OU-ISIR LP上的身份识别准确率

表 5-2 在 OU-ISIR LP 数据库上 Rank-1 识别准确率

	55	65	75	85	All
HWLD[98]	—	—	—	87.70	95.50
GEI[47]	84.70	86.63	86.91	85.72	94.24
FDF[47]	83.89	85.49	86.59		
CNN.FC2	73.96	76.71	77.81		
SiaNet.FC[97]	90.12	91.14	91.18		
3D-SiaNet	68.23	70.53	92.05		

时空联合深度神经网络能够
进一步提高侧面视角的
步态识别准确率

表 5-3 在 OU-ISIR LP 数据库上 Rank-5 识别准确率

	55	65	75	85	All
HWLD[98]	—	—	—	94.70	98.50
GEI[47]	92.39	92.84	92.78	93.01	97.13
FDF[47]	91.53	92.81	92.88	92.83	97.10
CNN.FC2	86.64	88.67	89.39	90.09	93.56
SiaNet.FC[97]	94.98	95.90	95.92	95.97	98.31
3D-SiaNet	72.01	73.80	96.32	95.36	96.18



提 纲

- 研究背景与意义
- 基于传统卷积神经网络的步态识别技术
- 步态识别中的距离度量学习模型
- 基于时空联合深度神经网络的步态识别技术
- 工作总结与展望

工作总结与展望

□ 贡献

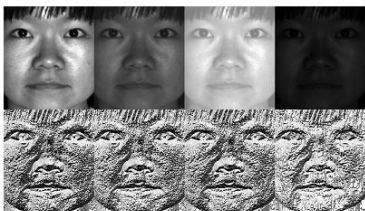
- 设计了一个基于深度学习的步态识别技术框架。
- 提出了一种基于Siamese神经网络的步态识别方法。将步态能量图作为卷积神经网络的输入对网络进行调优，解决了步态数据库样本量小以及分类和识别任务的领域鸿沟。
- 提出了一种基于时空联合深度神经网络的步态识别方法。从连续的周期性步态序列中捕捉空间维度和时间维度的信息，进一步提高了步态识别的准确率和实用性
- 在目前最大的步态识别数据库上表现出比目前最优方法更高的识别准确率，为步态识别走向应用提供了坚实的技术支持。

□ 未来工作

- 大规模真实场景步态数据库
- 端到端的步态识别系统
- 半周期或不完全周期步态识别
- 多模态生物识别

人脸识别

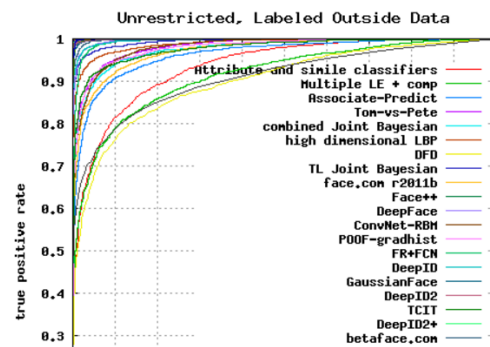
人工特征



Labeled Faces in the Wild



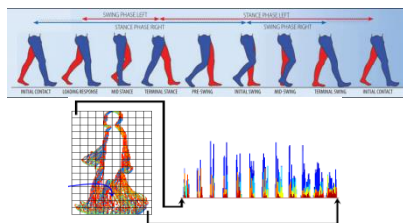
深度学习



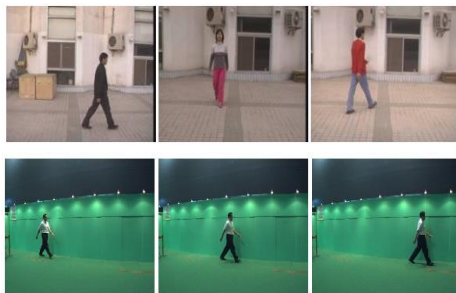
刷脸时代



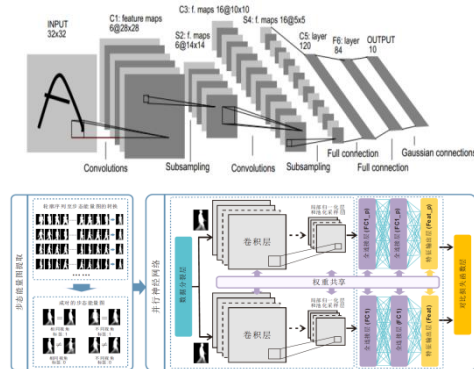
步态识别



从生理学研究 到视觉研究



从几人到几千人的步态数据库



深度学习



“刷退时代”

“High'st Queen of state, Great Juno comes; I know her by her *gait*”
—— *The Tempest* [Act 4 Scene 1], Shakespeare

谢 谢
请各位老师批评指正