

# Hierarchical ICA Encoding Combined with Motion-Appearance Information for Anomaly Detection

Shilei Duan, Kewei Wu, Shen Tang, Huan Liang, Zhao Xie

School of Computer and Information, Hefei University of Technology, Hefei Anhui  
Email: shilei\_duan@yeah.net, wu\_kewei1984@163.com

Received: Apr. 2<sup>nd</sup>, 2017; accepted: Apr. 14<sup>th</sup>, 2017; published: Apr. 19<sup>th</sup>, 2017

---

## Abstract

Due to the shortcomings of the existing anomaly detection methods in terms of the representation of visual hierarchical perception, inspired by the regularities in perception encoding of bio-mimetic vision, a hierarchical ICA encoding approach combined with motion-appearance is presented for abnormal events detection. First of all, this method extends the Existing biological visual hierarchy coding framework with three-level layer-wise learning, and uses ICA statistical method to extract intra-layer visual perceptual coding patterns, and utilizes the HMAX mechanism to transmit the Hierarchical information. In addition, with the processing theories of double channels in the visual system, Each channel is proposed to separately complete three-layer coding pattern learning. Then, these double features are fused to represent the Anomaly patterns. Based on the joint representations, the one-class SVM model is used to predict the abnormal score for each input. The proposed method, whose properties of motion perceptual coding and the performance of anomaly detection, is evaluated on UCSD datasets, and the results demonstrate that the learned feature representations in this paper for anomalous patterns are superior to other traditional hand-crafted features and deep learning features.

## Keywords

Hierarchical ICA Encoding, Motion-Appearance Representations, One-Class SVM, Anomaly Detection

---

# 基于运动外观多通道层级ICA编码的异常检测

段士雷, 吴克伟, 唐 燊, 梁 欢, 谢 昭

合肥工业大学计算机与信息学院, 安徽 合肥  
Email: shilei\_duan@yeah.net, wu\_kewei1984@163.com

收稿日期: 2017年4月2日; 录用日期: 2017年4月14日; 发布日期: 2017年4月19日

## 摘要

针对现有异常表示方法对视觉感知层级关系描述能力的不足, 基于生物视觉感知编码特性启发, 本文提出一种基于运动外观多通道层级ICA编码模型, 实现复杂场景中的异常检测任务。首先, 对现有的生物视觉层级编码框架, 进行三级逐层学习拓展, 采用ICA统计方法提取层内视觉感知编码模式, 利用HMAX机制实现层级信息传递。其次, 借助视觉双通道处理机制, 各通道独立完成三层编码模式学习, 随后联合双通道特征构建异常模式表达, 最终, 利用单类支持向量机模型对正常和异常情况进行判定。在UCSD数据集上, 分别验证了本文方法的运动感知编码特性和异常检测的性能, 实验结果能够说明本文异常模式表达优于现有的手工设计特征, 以及深度学习特征。

## 关键词

层级独立成分分析编码, 运动外观联合表达, 单类支持向量机, 异常检测

Copyright © 2017 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

视频场景分析与理解研究已经吸引了来自计算机视觉领域众多研究者的关注, 其致力于研究新技术、新方法去更精确快速地分析、理解场景内容, 从而更有效地协助监控人员获取准确信息以及处理突发事件, 并最大限度地降低误报漏报, 起到监督管理的作用[1]。视频场景中的异常事件检测是其中一项重要的研究内容, 同时也是研究的热点和难点。

异常检测最经典的做法通常是基于训练数据学习描述正常运动模式的模型, 然后通过评估新视频中的运动模式偏离该模型的程度来判断其是否属于异常运动。如 Hu 等人[2]采用多目标追踪算法提取正常运动轨迹特征, 然后学习其统计分布, 充分考虑时空信息用于异常检测。Mehran 等人[3]首次采用表达行人交互力信息的社会力模型(Social Force Model, SFM)来检测视频中的异常行为。此外, 基于多尺度光流直方图的稀疏编码模型[4]也成功用于异常检测, 该模型采用稀疏重构代价(Sparse Reconstruction Cost, SRC)为判断准则。Li 等人[5]采用混合动态纹理模型(Mixture Dynamic Texture, MDT)对外观、运动以及空间尺度特征进行建模, 提出了时空异常的联合检测器。上述方法虽然能够实现异常检测, 但是其采用的是手工设计特征, 该类特征需要专业的先验知识, 而这在复杂的视频场景下难以实现, 也限制了检测性能的进一步提升[6]。

近年来, 深度学习方法被成功应用于各项视觉任务, 证明了其强大的编码表达能力。深度学习本质思想也是以人类大脑对视觉信息的层次处理方式为基础, 构建多层次学习模型进行学习。如蔡瑞初等人[7]提出了一种基于多尺度时间递归神经网络的人群异常检测和定位方法。Xu 等人[6]则提出外观和运动深度网络(Appearance and Motion Deep Net, AMDN)学习运动、外观以及联合信息的特征表达用于异常检测。该方法采用堆栈去噪自动编码器网络进行特征学习, 并提出双融合策略, 分别是输入端的运动外观信息融合和输出端的异常得分融合, 其中, 前者融合是为了充分考虑外观和运动信息之间的互补性, 后者融合是为了得到运动区域的最终异常得分。但是, 在异常检测任务中, 这些深层框架仅仅被看作是输

入直接至输出的黑盒子模式的学习过程,即图1中红框部分,忽略了运动感知层次编码中潜在的编码特性,而且其层级较多,结构较为复杂,在学习过程中易产生过度拟合现象,从而导致结果不准确。

针对现有手工设计特征和端对端深度学习特征,在特征编码的层级关系描述能力的不足,即缺少对图1中红色框中的编码机制的深入研究,本文启发于仿生物视觉计算模型的(Hierarchical Max-pooling, HMAX)机制[8],对现有层级模型进行了扩展,采用具有稀疏特性的独立成分分析(Independent Component Analysis, ICA)统计方法,分析了运动感知编码中潜在的三层编码特性,并模仿人类视觉的双通道处理机制,采用独立的外观通道和运动通道分别学习高层结构特征,通过特征层融合来进行正常运动模式表示和异常检测。本文的异常检测框架如图1所示。与现有的异常检测方法相比,本文主要贡献如下:

1. 针对异常检测任务中的特征编码难题,对现有的生物视觉层级编码框架,进行三级逐层学习拓展,采用ICA统计方法提取层内视觉感知编码模式,利用HMAX机制实现层级信息传递,如图1中的层级ICA基元学习部分。

2. 借助视觉双通道处理机制,运动外观各通道独立完成三层编码模式学习,随后联合多通道信息构建异常模式表达,采用One-Class SVM对正常和异常情况进行判定,多通道特征层融合效果能够提升现有层级编码表达的准确性。

3. 在UCSD数据库上,与其他公认手工设计特征[3][4][5]和AMDN[6]深度学习特征进行了对比,实验表明本文方法的检测性能要优于对比方法,同时也直观解释了层级框架的三层运动感知编码特性。

## 2. 基于高层结构特征的异常检测

现有异常检测方法主要采用手工设计特征,如轨迹特征,光流直方图特征和3D时空梯度特征等等。这类特征不仅属于低层特征,也缺乏强大的编码表达能力,从而导致不佳的检测性能。AMDN方法首次用深层结构分别学习外观、运动及两者联合的编码特征用于异常检测。本文方法是生物视觉启发式的计算模型,采用层级学习结构,不同于传统手工设计特征提取,与AMDN方法在层级处理上也存在明显不同:1) 本文层级编码框架是模拟人类视觉系统的双通道处理机制,AMDN方法采用复杂的三通道学习机制;2) 本文层级框架包含三层学习结构,AMDN方法的深度更为复杂,层级与视觉通路的对应关系不明确;3) 运动外观通道融合策略不同:本文采用特征融合方式,AMDN方法采用得分融合方式;4) 本文能够有效解释层级结构的运动感知编码特性,AMDN方法只是当作输入至输出的黑盒子模式;5) 本文采用具备稀疏特性的ICA学习算法,而且在卷积之后增加了归一化和校正线性单元[9]等非线性操作以提高层与层之间的信息传输能力,AMDN方法采用传统稀疏编码学习算法。

### 2.1. 层级编码理论

人类视觉系统中存在结构和功能相对分离的两条通路[10]:一条是腹侧通路,主要感知目标形状和纹理等信息;另一条是背侧通路,主要感知目标位置和运动方向信息等。两条通路均呈现出明显的层次特性,层与层之间存在信息相互传递。图1中红框部分体现类似的双通道信息处理机制。两条通路拥有相似的层级学习过程[11],静态视觉信息具有层级编码过程[12],启发于此,本文的双通道机制采用相同的算法学习方式,如算法1所示。

层级编码特性中不同层次具有不同的感受野。感受野理论[13]是支持视觉信息分层串行处理的生物学基础理论,其主要是指视野某特定区域受到光线等刺激时,该区域可以引起相应神经元细胞的响应。研究表明,运动感知层级编码过程中,随着层级的不断提升,即算法1中参数 $l$ 的增加,感受野区域也在不断地增大,所提取的特征表达复杂度在不断增加。因此,根据感受野特征的不同,本文模型的编码特性可以分为以下三层:低层主要编码运动模式边缘和方向等信息,中层则主要编码运动模式的局部结构

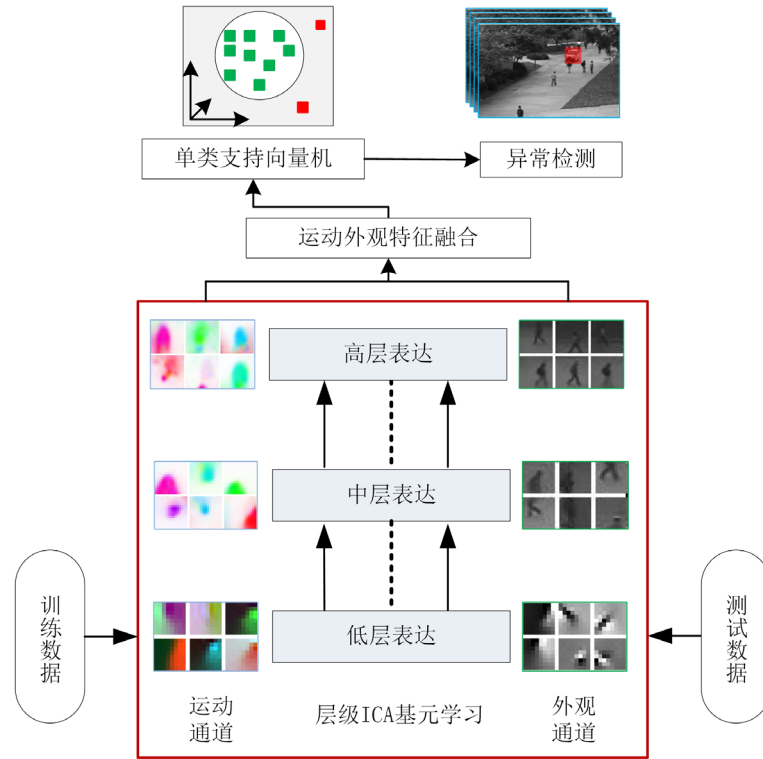


Figure 1. Framework of this paper for anomaly detection.  
图 1. 本文异常检测框架

**Algorithm 1.** Multi-channels hierarchical ICA encoding algorithm

**算法 1.** 多通道层级 ICA 编码算法

输入: 帧序列  $\{F_i^0\}_{i=1}^{N_f}$

输出: 高层特征集合  $P = \{p_i\}_{i=1}^{N_f}$

- 1: 对帧序列  $\{F_i^0\}_{i=1}^{N_f}$  提取光流特征  $\{Fm_i^0\}_{i=1}^{N_f}$
- 2: 运动通道层级 ICA 基元学习: For  $l=1: N_l$  ( $l$  表示是层级框架的第  $l$  层)
- 3: 对  $\{Fm_i^{l-1}\}_{i=1}^{N_f}$  随机采样, 构建训练集  $Xm_{l-1}$
- 4: 利用 ICA 进行层内基元模式学习  $\min \|Bm_l * \Psi(Xm_{l-1})\|, s.t. Bm_l Bm_l^T = I. \Rightarrow Bm_l$  表示基元模式
- 5: 利用基元模式实现层内编码  $Sm_l = \varphi(Bm_l * \Psi(Xm_{l-1}))$ , 其中  $\varphi(\cdot)$  表示归一化和校正线性单元等非线性操作, 增强层与层之间的信息传输能力
- 6: 层间信息传递,  $Fm_l^i = \max\text{-pooling}(Sm_l)$ , 通过局部邻域最大值池化操作增加局部平移不变特性
- 7: Endfor
- 8: 利用运动通道高层基元, 提取运动基元响应  $pm_i = \bigcup_{c=1}^{Nm_c} (dm_{(h,z)}^c)$ ,  $dm_{(h,z)}^c \in Fm_l^i$  表示坐标  $(h, z)$  处像素点,  $\bigcup_{\circ}$  表示像素点的层间级联串行操作,  $Nm_c$  表示  $Fm_l^i$  的通道数
- 9: 对运动区域  $\{F_i^{l-1}\}_{i=1}^{N_f}$  提取亮度特征  $\{Fa_i^{l-1}\}_{i=1}^{N_f}$ , 重复步骤 1-7, 学习外观通道的三层 ICA 基元  $Ba_i$
- 10: 利用外观基元  $Ba_i$  重复步骤 8, 提取外观基元响应  $pa_i$ , 两通道特征融合形成  $P = [\lambda pm_i, (1-\lambda) pa_i]$
- 11: 所有运动区域的运动外观多通道进行高层表达的矢量集合, 形成  $P = \{p_i\}_{i=1}^{N_f}$

信息, 而高层则编码运动模式的全局结构信息, 更趋于类别信息。

为了在实际数据集上充分验证上述编码特性, 本文将采用 ICA 学习算法进行运动模式的学习, 算法

1 描述了模型的学习过程。其中,  $Bm_l$  表示第  $l$  层的神经元卷积核。 $\Psi(\cdot)$  表示白化操作, 此外, 受上述启发可知,  $N_l = 3$ , 表示低中高三层编码。该方法与传统稀疏编码方法相比, 具有更高效的计算速度, 也具备很好的稀疏特性, 被广泛应用于生物视觉模拟。层级框架通过逐层非线性变换学习, 提取运动模式的高层全局特征。

## 2.2 独立成分分析

独立成分分析(ICA)方法是利用统计独立作为目标来分离独立成分的信号分解技术, 有利于小目标以及小类别信息的保留, 其目的是把观测到的混合信号分解为相互之间统计独立的成分, 然后利用该独立成分信息对其他信号进行线性组合表达, 即 ICA 通过在特征空间上寻找能使数据互相独立的空间, 将  $n$  维随机信号分解成一组统计独立的随机变量的线性组合。此外, 相互独立的各个分量之间需要满足正交的约束限制[14]。假定一组观测信号  $X = \{x_1, \dots, x_m\}$  是源信号  $S = \{s_1, \dots, s_n\}$  的观测值, 假设第  $i$  个观测信号是由  $n$  个独立分量  $S$  线性混合而成, 公式表达如下:

$$x_i = a_{i1}s_1 + a_{i2}s_2 + \dots + a_{in}s_n, i = 1, \dots, nm \quad (1)$$

上式可以用矢量表达为  $X = AS$ , 其中  $A = [a_1, \dots, a_n]$  表示混合矩阵,  $a_i$  表示混合矩阵的基向量。而 ICA 方法就是仅通过数据观测信号估计出独立源信号或混合矩阵数据。

稀疏编码模型学习到的基矩阵属于超完备基, 维数较大, 基与基之间也存在相关性。而 ICA 算法学习到的基不仅线性无关, 还会保持相互正交的特性。ICA 算法的目标函数为  $J(W) = \|Wx\|_1$ , 其加入正交约束后的优化策略如下:

$$\text{minimize } \|Wx\|_1 \quad \text{s.t. } WW^T = I \quad (2)$$

此外, 在迭代优化过程中, 如果每次用梯度下降法更新权值  $w$  后, 都需要对其进行正交约束, 更新方式如下面公式所示:

$$W \leftarrow (WW^T)^{-\frac{1}{2}} W \quad (3)$$

其中权值  $w$  矩阵中基的维数低于输入数据的维数, 并且在采用 ICA 算法进行学习时, 要对输入数据进行白化预处理操作, 目的是为了去除数据维度之间的相关性, 保证特征间的协方差矩阵为单位矩阵。

## 2.3. 异常判断

本文把异常检测问题看成是基于数据块的二分类问题, 为了降低计算代价及提高计算效率, 异常检测过程中只考虑包含有运动区域的数据块, 忽略背景信息块, 即给定含有运动区域的测试块数据, 采用训练学习所构建的模型来预测该数据块包含的运动模式是属于正常模式还是异常模式。综上, 本文模拟人类视觉系统中的双通道机制, 采用外观和运动两个独立编码框架, 分别学习运动模式的高层表达。为了充分考虑视频场景中具有互补性的外观和运动两类信息, 需要对运动和外观特征进行融合操作共同表达运动模式。本文采用线性加权的融合方式进行整合, 即按照公式  $P = [\lambda pm_i, (1-\lambda) pa_i]_{i=1}^{N_f}$  进行融合操作获得最终的特征矢量, 描述运动模式信息, 框架如图 1 所示。基于该高层融合特征, 采用单类支持向量机[15]对正常运动模式进行学习建模, 评估其分布, 建模优化策略如下:

$$\begin{aligned} \min_{w, \rho} \quad & \frac{1}{2} \|w\|^2 + \frac{1}{uN} \sum_{i=1}^N \xi_i - \rho \\ \text{s.t.} \quad & w^T \Phi(p_i) \geq \rho - \xi_i; \xi_i \geq 0 \end{aligned} \quad (4)$$



其中  $w$  表示权重矢量,  $\rho$  表示偏差,  $u \in (0,1]$  为平衡控制参数,  $\Phi(\cdot)$  表示特征映射函数。然后对测试样例判断其是否符合正常运动模式的分布并采用已训练好的权重和偏差计算相应异常得分  $As(p_i) = \rho - w^T \Phi(p_i)$ 。此外, 根据接收器操作特性曲线计算取得最大曲线围成的面积时的阈值作为判断的最终阈值。当异常得分  $As(p_i)$  大于阈值时, 表示该样例是异常运动模式, 反之, 其属于正常模式。

### 3. 实验结果及分析

本文验证实验是基于 UCSD 数据集[16], 该数据集包含有两部分: Ped1 和 Ped2, 分别描述两个不同的场景。两个子集中异常运动模式包括骑自行车的人、踩滑板的人、汽车等等, 而正常运动模式都是人行道上正在行走的行人。为了提高计算效率, 本文保持分辨率长宽比不变, 统一把视频分辨率缩小为  $210 \times 140$  像素。三层卷积核个数设置为 100、100 和 210, 卷积核尺寸分别为  $8 \times 8$ ,  $4 \times 4$  及  $3 \times 3$ 。此外, 本文所采用的学习框架是采用无监督方式进行学习。测试过程中以帧层和像素层的接受器操作特性曲线 (Receiver Operating Characteristic, ROC) 以及等错误率 (Equal Error Rate, EER) 作为最终的评价标准来验证本文方法的检测性能。其中, EER 表示在漏检率等于假阳率情况下的检测错误率。该评价标准沿用了 Mahadevan 等人[15]提出的评估方式, 即当某个像素被认为是异常的时候就判定该帧是异常帧; 而如果引起异常的对象有 40% 的像素被检测到, 就认为异常定位准确。

为了更有效评估方法的检测性能, 本文采用以下四种对比方法: 社会力模型[3], 混合动态纹理模型[4], 稀疏重构代价模型[5]及深度学习模型 AMDN[6]。该四类方法按其所采用特征可以分为两大类: 1) 非深度学习方法, 采用不同的传统手工设计特征来检测异常。其中, 社会力模型是受物理学启发的用于异常检测的经典模型, 主要通过交互力和环境因素来描述场景动态; 混合动态纹理模型充分考虑场景外观和运动信息, 基于局部特征表达来检测异常; 稀疏重构代价模型是基于光流特征的稀疏表达思想, 采用重构误差来评估运动的异常程度。2) 深度模型。AMDN 方法是第一个用于异常检测的深度模型, 采用自动去噪编码器层级结构。这四种方法代表不同的特征编码方法, 也能够全面地反映出本文框架高效性。

**实验一: 验证层级编码框架的运动感知编码特性。**图 2 表示层级学习框架中三层编码特征的可视化。针对图 2, 可以得出以下结论: 1) 低层主要编码运动边缘和方向等信息; 2) 中层则主要表征运动模式的局部结构信息, 具有一定的类别信息, 如头部运动、脚部运动等; 3) 高层更倾向于表达不同的运动模式类别信息, 包含运动模式的全局结构信息。此外, 深度学习的本质思想也是模拟人类视觉系统的视觉信息层次处理机制, 从海量的数据中进行学习提取有用的特征。因此, 无论其他检测方法使用了几层的处理机制, 本文实验揭示的三层编码规律都应该能够去解释其编码特性。

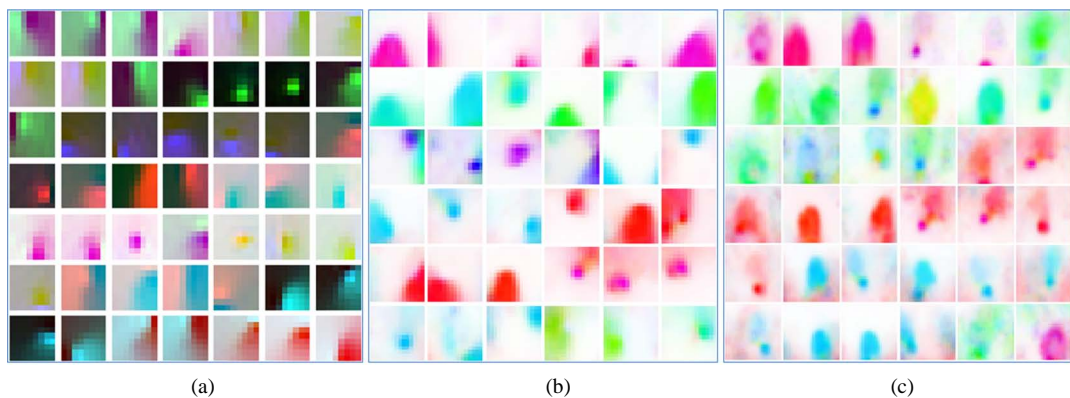
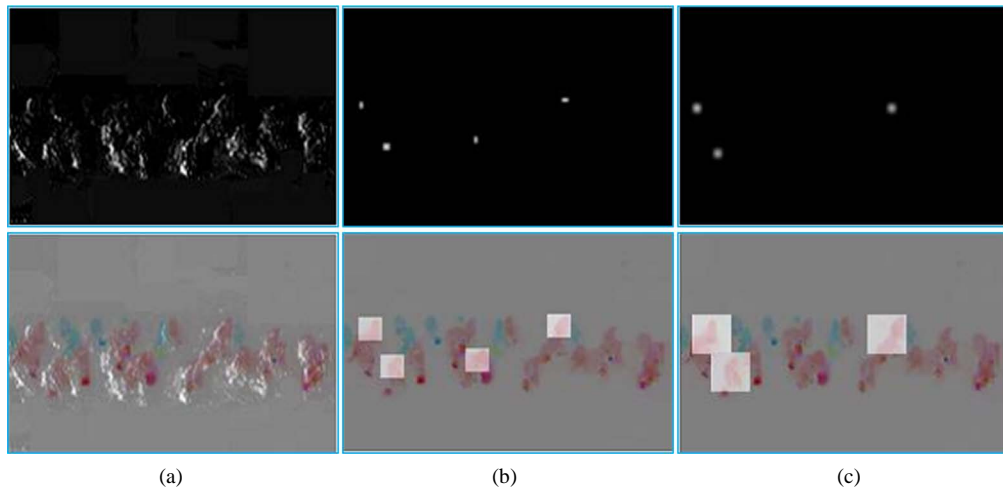


Figure 2. Visualization of feature representations for three-level neurons, (a) Low-level; (b) Mid-level; (c) High-level  
图 2. 三层神经元的特征表达可视化, (a) 低层; (b) 中层; (c) 高层

图 3 表示某卷积核特征响应图及其在原始数据上的感受野区域。针对图 3，可以看到：1) 低层编码的边缘等信息与运动模式类别无关，具有共性，上述三层编码规律再次得到了充分验证；2) 中层局部结构信息的特征表达适用于分析某些拥挤场景下局部运动模式；而高层编码特征，可以有效用于运动模式检测和行为识别等视觉任务。基于此，本文将高层全局结构特征来建模场景中的运动模式用于异常检测。

**实验二：评估三层学习框架在异常检测任务中的检测性能。**图 4 和表 1 给出了不同方法在 UCSD 数据集上的异常检测和异常区域定位的定量结果。由于现有大多数方法都没有给出 Ped2 子数据集上的异常区域定位结果，因此，图 4 中也没有给出异常区域定位曲线。针对定量结果，可以得到：1) 异常帧检测层面：本文方法在 Ped1 子集上取得了与最好方法相当的性能，但明显优于其他对比方法；在 Ped2 子集上本文方法的性能是最好的，相比于其他方法，有了较大的性能提升。2) 异常区域定位层面：本文方法在 EER 和 AUC 两个标准下都获得最好的性能。此外在 UCSD 数据集上的一些异常检测实例如图 5 所示。

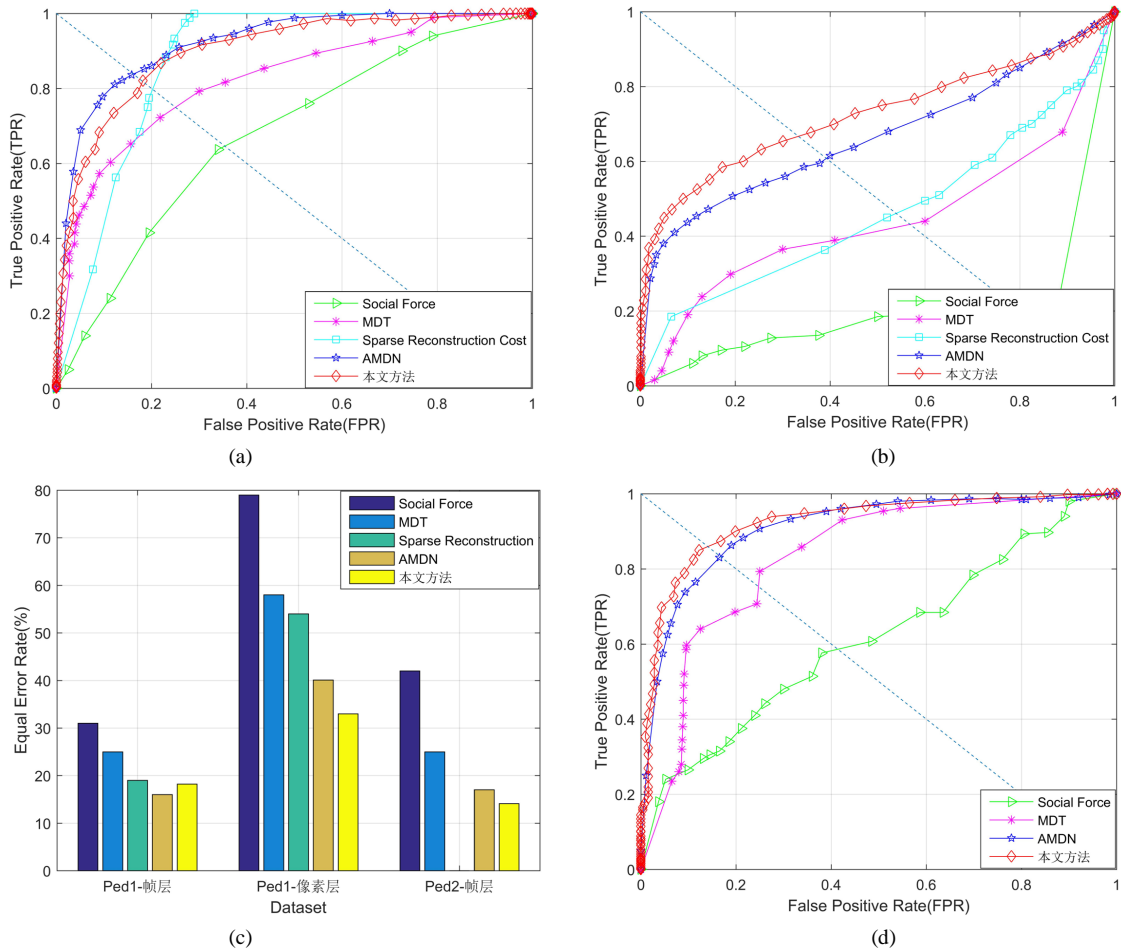
针对上述获得的优异定量结果进行分析：1) 相比于采用手工设计特征的传统方法，本文方法是采用学习机制提取高层全局特征，不需要专业的先验知识，并且特征具有更强的表达能力，因此本文方法的异常帧检测和异常区域定位性能都优于该类方法。2) 相比于基于深度学习的检测方法，AMDN 是第一个用于异常检测的采用深度框架的方法，从结果不难发现，其取得了优异的检测性能。但是该方法采用的去噪自动编码器网络结构较为复杂，层级较多，忽略了运动感知层级编码特性。而本文方法采用结构较为简单的三层学习框架，模拟生物视觉的双通道机制，充分考虑运动感知层级编码规律。尽管 AMDN 方法在 Ped1 子集上的异常帧检测取得最好的性能，但是本文方法同样也取得了相当的性能，而且异常区域定位和 Ped2 子集上异常帧检测性能上都有一定提升，综合来看，本文方法的异常检测优势更明显。



**Figure 3.** Responsive feature maps and associated receptive field region  
**图 3.** 响应特征图及其对应的感受野区域

**Table 1.** Quantitative comparison for anomaly detection on UCSD dataset  
**表 1.** UCSD 数据集上的异常检测定量评比结果

检测方法	Ped1 帧层	Ped1 像素层	Ped2 帧层
	AUC (%)	AUC (%)	AUC (%)
Social Force [3]	67.5	19.7	55.6
MDT [4]	81.8	44.1	82.9
Sparse Reconstruction [5]	86.0	46.1	—
AMDN [6]	<b>92.1</b>	67.2	90.8
本文方法	89.8	<b>72.5</b>	<b>92.3</b>



**Figure 4.** Quantitative curves for anomaly detection on UCSD dataset. (a) ROC curves of frame-level for detection results on Ped1; (b) ROC curves of pixel-level for detection results on Ped1; (c) EER results for UCSD dataset; (d) ROC curves of frame-level for detection results on Ped2

**图 4.** UCSD 数据集上异常检测结果定量曲线。(a) Ped1 帧层检测结果 ROC 曲线; (b) Ped1 像素层检测结果的 ROC 曲线; (c) UCSD 数据集上 EER 结果; (d) Ped2 帧层检测结果 ROC 曲线



**Figure 5.** Examples for anomaly detection on UCSD dataset

**图 5.** UCSD 数据集上异常检测实例



## 4. 结束语

现有异常检测方法在至关重要的特征编码层面存在不足, 忽略了运动感知编码特性。针对该问题, 本文模拟生物视觉系统的信息处理机制, 通过采用具备稀疏特性的 ICA 算法构建 HMAX 计算模型, 研究其运动感知层级编码特性, 揭示了低中高三层编码规律, 然后, 采用高层运动全局结构特征来对视频场景中的正常运动模式进行建模, 用于异常运动模式发现。UCSD 数据集实验已经充分证明了本文方法的有效性, 本文的异常检测方法获得的性能也优于其他传统方法和 AMDN 方法。

## 致 谢

感谢国家自然科学基金(No. 61273237; No. 61503111; No. 61501467)的支持。

## 参考文献 (References)

- [1] 余昊, 孙铨锋, 蒋兴浩. 基于光流块统计特征的视频异常行为检测算法[J]. 上海交通大学学报, 2015, 49(8): 1199-1204.
- [2] Hu, W., Xiao, X., Fu, Z., *et al.* (2006) A System for Learning Statistical Motion Patterns. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **28**, 1450-1464. <https://doi.org/10.1109/TPAMI.2006.176>
- [3] Mehran, R., Oyama, A. and Shah, M. (2009) Abnormal Crowd Behavior Detection Using Social Force Model. *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, 20-25 June 2009, 935-942.
- [4] Cong, Y., Yuan, J. and Liu, J. (2013) Abnormal Event Detection in Crowded Scenes Using Sparse Representation. *Pattern Recognition*, **46**, 1851-1864.
- [5] Li, W., Mahadevan, V. and Vasconcelos, N. (2014) Anomaly Detection and Localization in Crowded Scenes. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **36**, 18-32. <https://doi.org/10.1109/TPAMI.2013.111>
- [6] Xu, D., Yan, Y., Ricci, E., *et al.* (2017) Detecting Anomalous Events in Videos by Learning Deep Representations of Appearance and Motion. *Computer Vision & Image Understanding*, **156**, 117-127.
- [7] 蔡瑞初, 谢伟浩, 郝志峰, 等. 基于多尺度时间递归神经网络的人群异常检测[J]. 软件学报, 2015, 26(11): 2884-2896.
- [8] Riesenhuber, M. and Poggio, T. (1999) Hierarchical Models of Object Recognition in Cortex. *Nature Neuroscience*, **2**, 1019. <https://doi.org/10.1038/14819>
- [9] Nair, V. and Hinton, G.E. (2010) Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair. *International Conference on Machine Learning*, Haifa, 21-24 June 2010, 807-814.
- [10] Ungerleider, M.L.G. (1982) Two Cortical Visual Systems. *Analysis of Visual Behavior*, **35**, 549-586.
- [11] Giese, M.A. and Poggio, T. (2003) Neural Mechanisms for the Recognition of Biological Movements. *Nature Review Neuroscience*, **4**, 179. <https://doi.org/10.1038/nrn1057>
- [12] Hu, X., Zhang, J., Li, J., *et al.* (2014) Sparsity-Regularized HMAX for Visual Recognition. *PLoS ONE*, **9**, e81813. <https://doi.org/10.1371/journal.pone.0081813>
- [13] Hubel, D.H. and Wiesel, T.N. (1968) Receptive Fields and Functional Architecture of the Monkey Striate Cortex. *Journal of Physiology*, **195**, 215-243. <https://doi.org/10.1113/jphysiol.1968.sp008455>
- [14] 冯燕, 何明一, 宋江红, 等. 基于独立成分分析的高光谱图像数据降维及压缩[J]. 电子与信息学报, 2007, 29(12): 2871-2875.
- [15] Schölkopf, B., Platt, J.C., Shawetaylor, J., *et al.* (2001) Estimating the Support of a High-Dimensional Distribution. *Neural Computation*, **13**, 1443. <https://doi.org/10.1162/089976601750264965>
- [16] Mahadevan, V., Li, W., Bhalodia, V., *et al.* (2010) Anomaly Detection in Crowded Scenes. *Computer Vision and Pattern Recognition*, San Francisco, 13-18 June 2010, 1975-1981.