

# 融合多特征与时间序列的人群行为识别模型<sup>①</sup>



吴敏忠, 王 雷, 盛 捷

(中国科学技术大学 信息科学与技术学院, 合肥 230031)

通信作者: 王 雷, E-mail: wangl@ustc.edu.cn

**摘 要:** 人群行为识别在公共安全等领域具有重要的应用价值. 现有研究分别考虑了人群情绪、人群类型、人群密度以及人群社会文化环境等因素对于人群行为的影响, 但少有综合考虑这些因素的模型, 导致模型性能受限. 本文综合考虑人群物理特征、社交特征、情绪人格特征和文化背景特征之间的相关性, 以及相结合之后对人群行为的影响, 提出一种融合多特征与时间序列的人群行为识别模型. 模型采用两个并行的网络层分别处理多特征相关性和时间序列依赖性对于人群行为的影响, 同时为提高模型可解释性, 网络层采用融合结构因果模型 (SCM) 与图神经网络 (GNN) 的因果图网络 (CGN). 通过在运动情感数据集 (MED) 上进行实验并与其他方法模型进行对比, 证明了本文方法能够成功识别人群行为, 并且优于目前最先进的方法.

**关键词:** 人群行为识别; 多特征融合; 图神经网络; 结构因果模型; 因果图网络; 时间序列

引用格式: 吴敏忠, 王雷, 盛捷. 融合多特征与时间序列的人群行为识别模型. 计算机系统应用, 2022, 31(11): 268-274. <http://www.c-s-a.org.cn/1003-3254/8788.html>

## Crowd Behavior Recognition Model Integrating Multi-feature and Time Series

WU Min-Zhong, WANG Lei, SHENG Jie

(School of Information Science and Technology, University of Science and Technology of China, Hefei 230031, China)

**Abstract:** Crowd behavior recognition has important application value in public safety and other fields. Existing studies have considered the influence of such factors on crowd behavior as crowd emotions, crowd types, crowd density, and social and cultural backgrounds of crowds separately, but few models comprehensively consider these factors, which limits model performance. This study comprehensively considers the correlation between the physical features, social features, emotional and personality features, and cultural background features of the crowd and the influence of the combination of these factors on crowd behavior. As a result, a crowd behavior recognition model that integrates multiple features and time series is proposed. The model uses two parallel network layers to deal with the influence of multi-feature correlation and time-series dependence on crowd behavior separately. Meanwhile, the network layer fuses the structural causal model (SCM) and the causal graph network (CGN) based on the graph neural network (GNN) to improve the interpretability of the model. The experiments on the motion and emotion dataset (MED) and the comparison with other state-of-the-art models demonstrate that the proposed method can successfully identify crowd behavior and outperform the state-of-the-art methods.

**Key words:** crowd behavior recognition; multi-feature fusion; graph neural network (GNN); structural causal model (SCM); causal graph network (CGN); time series

近些年来, 在人们聚集的地方, 影响公共安全的事件频繁发生, 对于人群行为分析和预测的需求也随之

上升, 因为如果能在危险事件发生之前识别出人群中异常的行为, 并提前采取防治措施, 就能防止可能的人

<sup>①</sup> 基金项目: 科技创新特区计划 (20-163-14-LZ-001-004-01)

收稿时间: 2022-03-02; 修改时间: 2022-04-02; 采用时间: 2022-04-11; csa 在线出版时间: 2022-07-07

群相关的灾难,保护人民的生命财产安全。

目前人群行为识别工作主要分为人群异常行为检测和人群行为分类两种,对于人群的异常行为检测,大多数研究工作都集中在人群特定场景的异常行为检测<sup>[1,2]</sup>,但是它们对于异常行为缺乏一个恰当的定义<sup>[3]</sup>,会导致很多正常行为被误检为异常行为。而人群行为分类通过考虑人群行为中几类特定行为的特征将他们区分开,对于跨场景人群行为研究具有较高的价值,因此可以将人群行为分类作为首选。另外,目前大多数人群行为识别工作都基于人群的单类型特征完成的,例如很多人群行为识别工作都是基于人群的物理运动特征来进行的,比如说梯度直方图(HOG)<sup>[4,5]</sup>、光流直方图(HOF)<sup>[6]</sup>和运动边界直方图(MBH)<sup>[4]</sup>等,但是这些特征都无法区分语义级别的差异<sup>[7]</sup>。随着深度学习的发展,人们通过深度学习网络提取出人群特征的高级表示来完成人群行为识别任务,例如Varghese等人通过神经网络提取出人群的情绪特征来识别人群行为<sup>[8]</sup>。但是综合来看上述这些研究方法都是考虑人群的单类型特征或者部分特征对于人群行为的影响进而完成人群行为识别任务,但是人群行为本质上是多种因素共同决定而产生的结果<sup>[9]</sup>,除了人群的物理运动特征和人群的情绪特征之外,人群行为还会受到人群的社交特征等因素的影响。因此,本文将综合考虑人群运动特征、社交特征、情绪人格特征以及文化特征,并根据这些特征之间的联系构建得到一个人群的多特征关系模型。另外,考虑到人群行为的时间序列依赖性,模型将采用两个并行的网络层来融合人群多特征相关性和人群行为时间序列依赖性对于人群行为的影响来完成人群行为识别任务。为了提高模型的可解释性,网络层结构采用的是一种融合GNN与SCM的因果图网络(CGN)模型,该网络融合了GNN的结构化模型的近似能力与SCM的因果机制表述能力。总结来说,主要贡献有:

(1) 综合考虑影响人群行为的多类特征,提出一种用于表述人群特征对于人群行为的影响以及人群特征之间相关性的人群多特征关系模型。

(2) 采用一种双重并行网络模型,融合人群行为识别中人群多特征对于人群行为的影响以及人群行为时间序列依赖性对于人群行为的影响。

(3) 模型中网络层采用因果图网络,该网络兼具图神经网络的推理能力以及结构因果模型的因果机制表述能力。

## 1 多特征关系模型

由于人群运动是人群行为最直观的展现,因此很多研究者的基于人群的物理运动特征来研究人群行为,Colque等人<sup>[10]</sup>基于人群的物理运动特征来检测人群中的异常行为。但是除此之外,还有很多其他类型的特征会对人群行为产生影响。情绪作为一种复杂的心理状态,是一种直接而强烈的行为动机<sup>[11]</sup>,可以影响人的行为和思想,Challenger等人<sup>[10]</sup>研究指出情绪在决定人群行为方面发挥重要作用,同时,人格特征能够反映出人群根据事件采取行动的行为模式。因此在人群行为识别工作上,人群情绪人格是一类不可忽略的特征。此外Forsyth<sup>[12]</sup>指出每个特定文化中的人群都会表现出独特的人际交往过程,文化差异会导致人群速度和密度等物理特征的变化<sup>[13]</sup>,同时,文化还影响着人群之间的交流方式,在人群行为的形成过程中发挥重要作用<sup>[14]</sup>。Hofstede提出描述人群文化特征的6维度文化模型<sup>[15]</sup>,并且研究中发现文化维度模型与人格特征具有显著联系<sup>[16]</sup>。最后,人群中个体与周围个体之间的社会关系会造成人群中个体之间的物理位置,距离以及速度上的差异,进而影响人群的行为,因此在人群行为识别问题上,考虑人群的社交特征是有益的。以上特征都会对人群行为产生影响并且各特征之间具有一定的相关性,因此,本文将人群的物理运动特征、社交特征、情绪人格特征以及背景文化结合,得到人群的多特征关系模型,如图1所示。下面将具体介绍上述4类特征。

### 1.1 物理特征

对于物理特征,本文主要考虑人群的运动熵、速度方差、运动混乱指数和人群密度。人群运动熵可以用来描述人群系统的物理状态,速度方差可以用来描述人群运动速度大小的一致性,人群运动混乱指数可以用来描述人群运动方向的一致性,人群密度可以用来描述人群的空间分布情况。为获取这些特征值,本文采用deep-sort算法<sup>[17]</sup>提取出视频画面中每个个体 $p_i$ 的各个时间的物理位置 $(x_i, y_i)$ ,并据此计算出其对应的运动速度 $v_i$ 与方向 $\alpha_i$ ,并计算人群平均速度 $\bar{v}$ ,进而计算得到人群的运动熵和速度方差。在人群运动混乱指数计算过程中,首先将空间划分为8个方向区域,每个方向区域夹角为 $45^\circ$ ,再根据人群中各个个体的运动方向 $\alpha_i$ 得到每个方向区域的人数 $h_i$  ( $i = 1, 2, \dots, 8$ ),从而获取人群运动混乱指数 $E$ 。人群密度可以通过采用维诺图来估计<sup>[18]</sup>。基于画面中各个个体的空间位置,画出维诺图并计算出每个多边形小区域的面积 $S_i$ ,并求取人群密度。

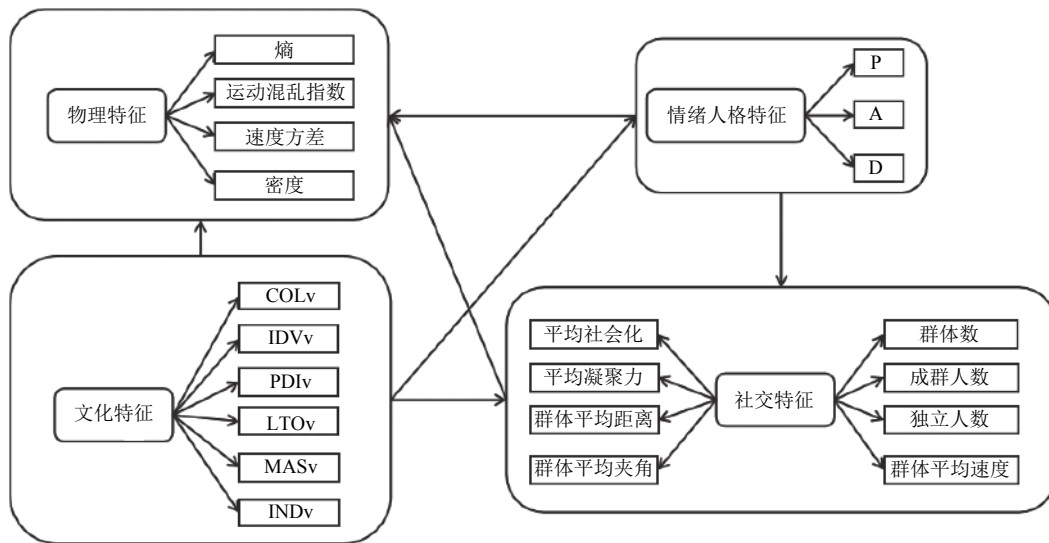


图1 多特征关系模型

### 1.2 社交特征

对于社会特征, 本文主要考虑人群中, 行人的集体化水平、孤立化水平以及社会化水平. 对于人群的社交特征, 无法避免的需要考虑人群中的小团体. 人群之间的小团体可以基于个体之间的距离, 速度以及方向来识别. Hall 提出近因学术语用来描述人群中个体交互时周围个人空间使用情况, 并且证明个体之间的社会距离与物理距离有关<sup>[19]</sup>, 他将人与人之间的人际距离划分为4个区域: 亲密距离、个人距离、社会距离和公共距离. Favaretto 基于 Hall 的理论, 提出判定人群中两个个体是否属于同一小团体的三要素<sup>[20]</sup>:  $d(\vec{x}_i, \vec{x}_j) \leq 1.2m$ ;  $\alpha(\alpha_i, \alpha_j) \leq 15^\circ$ ;  $s(s_i, s_j) \leq 0.05 \times \max\{s_i, s_j\}$ .

通过这3个要素判断两个个体是否属于同一小团体, 可以将人群划分成很多个二元组或者独立个体组, 再将具有交集的组合并就可以将群体划分成很多个小团体. 在获取到人群中的小团体数据之后, 进一步计算小团体数量 $n$ 、组成小团体的人数 $v$ 、未组成团体的人数 $\xi$ 、小团体之间的平均距离、平均速度、平均方向夹角、社会化程度以及群体凝聚力. 小团体之间的平均距离, 平均速度和平均方向夹角根据式(1)–式(3)计算:

$$\bar{d}_g = 2 \times \frac{\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} d(\vec{x}_i, \vec{x}_j)}{n(n-1)} \quad (1)$$

$$\bar{s}_g = 2 \times \frac{\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} s(s_i, s_j)}{n(n-1)} \quad (2)$$

$$\bar{\alpha}_g = 2 \times \frac{\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} o(\alpha_i, \alpha_j)}{n(n-1)} \quad (3)$$

下面主要讲述群体凝聚力和群体社会化程度.

#### 1.2.1 群体凝聚力

凝聚力 $C_g$ 是由于群体内部成员之间的联系而产生的, 反映的是一个群体内部人员希望留在团队的程度, 可通过团体的稳定性来量化. 群体凝聚力 $C_g$ 可通过式(4)计算:

$$C_g = \frac{1}{n_g} \sum_{i=1}^n CO_i + CV_i + CP_i \quad (4)$$

$$CO_i = \frac{w - |\bar{\alpha}_g - \alpha_i|}{w} \quad (5)$$

$$CV_i = \frac{\gamma - |\bar{s}_g - v_i|}{\gamma} \quad (6)$$

$$CP_i = \log n_g \quad (7)$$

其中,  $n_g$ 表示小团体中成员数量,  $CO_i$ 是与运动方向有关的凝聚力,  $CV_i$ 是与速度有关的凝聚力,  $CP_i$ 是与群体规模有关的凝聚力.

#### 1.2.2 群体社会化程度

群体社会化程度反映的是一个群体中个体之间的相似性, 社会化程度越高的群体, 个体之间的行为模式相似程度也会越高. 可以通过式(8)计算得到:

$$\Psi = \frac{1}{N_a^2} \sum_{i=1}^{N_a} \sum_{j=1}^{N_a} e^{(-\beta w(i, j)^2)} \quad (8)$$

其中,  $N_a$ 是行人数量,  $\beta = 0.3$ ,  $\omega(i, j)$ 是相似度衰减函数,  $\omega(i, j) = s(s_i, s_j) + \alpha(\alpha_i, \alpha_j)$ .

### 1.3 情绪人格特征

情绪是人群行为的一种直接且强烈的动机, 对于人群行为模式具有十分重要的影响. Ortony 等人<sup>[21]</sup>提出情绪的 OCC 模型, 将情绪分为 22 种不同类型, 从动作、事件和对象 3 个概念出发成功定义了情绪, 但是, 由于人类行为的不可预测性和不确定性, 通过情绪特征反映出人群行为, 需要进一步的研究. Mebrabian<sup>[22]</sup>提出一种心理状态模型: 愉悦-唤醒-支配 (PAD) 模型, 该模型通过愉悦、唤醒和支配 3 个维度来表征情绪, 这 3 个维度的值对于管理不同行为的情绪倾向上具有重要作用, 因此若是将情绪表征成 PAD 空间的值, 将使得人群行为识别任务更加简单. 实验使用的数据集 MED 对于视频中每帧画面都有相应的人群情绪标注, 标注将人群情绪分为生气、悲伤、兴奋、害怕、高兴以及中立 6 类, 因此本文直接使用数据集自带的标注的人群情绪数据, 并利用 OCC 模型与 PAD 模型之间的映射关系<sup>[23]</sup>, 将人群情绪映射到 PAD 空间.

另外由于人格特征也会对人群的行为模式产生影响, 因此可以通过利用人格特征进一步细化前面得到的 PAD 值来将人格特征融入到 PAD 值. 具体的, 本文考虑刻画人格特征的 OCEAN 五因素模型<sup>[24]</sup>, 该模型从 5 个维度描述人格特征: 开放性、责任性、外向型、亲和性以及神经质. 利用 OCC 模型与 OCEAN 模型之间的映射关系<sup>[23]</sup>, 可以计算出 5 个维度的值  $F = [F_o, F_C, F_E, F_A, F_N]$ , 之后基于 OCEAN 模型与 PAD 的映射关系将其映射到 PAD 空间得  $[P_F, A_F, D_F]$ , 最后即可计算得到最终所需的 PAD 值:

$$P_{\text{final}} = P + P_F \quad (9)$$

$$A_{\text{final}} = A + A_F \quad (10)$$

$$D_{\text{final}} = D + D_F \quad (11)$$

### 1.4 背景文化特征

本文采用 Hofstede 文化维度模型<sup>[15]</sup>来描述人群的背景文化特征, 该模型通过权力距离指数 (PDI)、个人主义与集体主义 (IDV)、不确定性规避指数 (UAI)、男性化社会与女性化社会 (MAS)、长期导向与短期导向 (LTO) 和放纵与克制 (IND) 6 个维度来描述人群文化特征. 本文考虑其中除 UAI 之外的 5 个维度的指标特征. PDI 维度指标表示的是社会地位的差异对于人

群行为的影响, 人群中个体相互靠近的群体社会地位观念比较淡泊, 而群体中个体比较分散的群体社会地位观念比较强烈. 因此可以将 PDI 指标取值为群体中各个小团体的平均距离. IDV 维度指标是与群体中人际空间与个人速度相关的指标, 可以通过人群中组成小团体和未组成小团体的人数百分比来衡量. LTO 维度反映的是群体的长期目标导向程度, MAS 维度指标反映的是群体的合作偏好, IND 维度指标可以通过群体的平均速度和群体集体主义程度来衡量, 它们的计算方法如下:

$$LTO_v = \begin{cases} O_v, & \text{if } O_v \geq 50 \\ 100 - O_v, & \text{otherwise} \end{cases} \quad (12)$$

$$MAS_v = \chi_1 GC_v + (1 - \chi_1) LTO_v \quad (13)$$

$$IND_v = \rho_1 S_v + (1 - \rho_1) COL_v \quad (14)$$

其中,  $\chi_1 = \rho_1 = 0.5$ 均为经验参数,  $O_v$ 是群体中所有小团体运动夹角的平均值与最大夹角的百分比,  $GC_v$ 为群体凝聚力与经验最大值的比值,  $S_v$ 是群体平均速度与经验最大速度的百分比.

## 2 融合多特征与时间序列的人群行为识别模型

本文提出一种融合多特征与时间序列的人群行为识别模型. 模型的整体框架如图 2 所示. 模型首先通过特征提取部分提取出第 1 节介绍的 4 类人群特征, 然后通过两个并行的因果图网络层融合多特征相关性以及时间序列依赖性对于人群行为的影响向量, 之后将这两个向量与多特征向量拼接耦合, 接着用 GRU 来捕捉时间序列中的序列模式, 最后采用一个全连接的神经网络作为识别网络来对人群行为进行识别分类. 下面主要详细介绍模型中两个并行的因果图网络层.

### 2.1 因果图网络

因果关系一直以来都是人类认知的核心, 在人群行为识别任务中, 如果能发掘特征与人群行为之间的因果关系, 将对人群行为识别任务具有巨大的影响. 根据 Pearl 因果层次理论, 因果推理分为 3 个层次: 关联、干预和反事实推理, 传统的神经网络依靠数据之间的关联性完成任务, 只能覆盖第一层. 负责数据生成的结构因果模型 (SCM) 作为因果推理的核心模型, 其引入了因果层次理论的 3 个层次, 能够反映出图形节点之间的因果关系. 因此本文将图神经网络与结构因果模型结合, 得到一类网络, 称之为因果图网络 (CGN). CGN 按照如下方法构造.

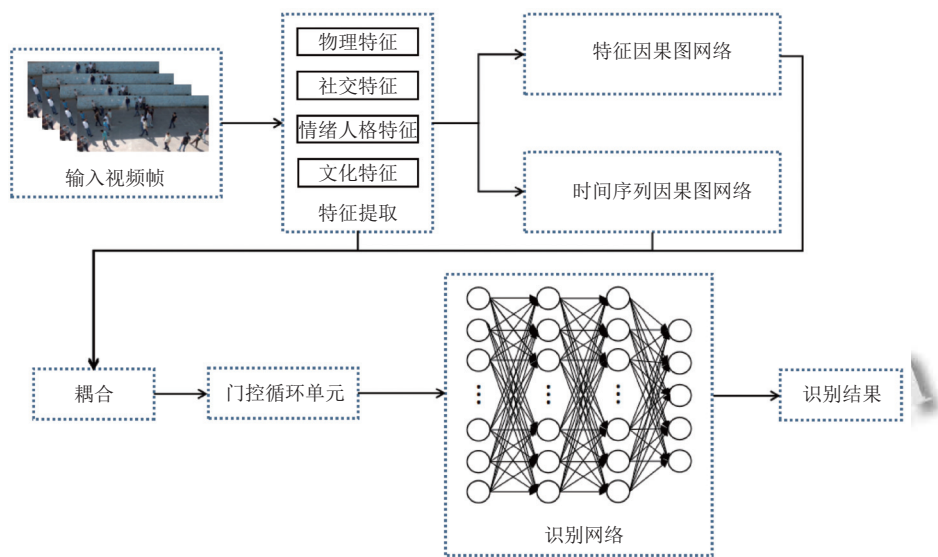


图2 本文模型框架

对于任意一个 SCM:  $\varsigma = (S, P(U))$ , 其结构方程为  $f_i(pa(i), U_i) = \sum_{j \in pa(i) \cup U_i} f_{ij}(V_j)$ , 其中  $f_i \in S$ , 那么令:

$$F = V \cup U = F' \tag{15}$$

$$\varphi(i, pa(i)) = U_i + \sum \psi(i, j) \tag{16}$$

$$\psi(i, j) = f_{ij} \tag{17}$$

最后, 令:

$$h_i = \varphi(d_i, \oplus_{j \in N_i^G} \psi(d_i, d_j)) = U_i + \sum_{j \in pa(i)} f_{ij}(V_j) = f_i \tag{18}$$

通过这样, 得到了一种特殊的 GNN, 即 CGN, 该网络兼具 GNN 的图计算能力和 SCM 的因果表示能力.

### 2.2 特征因果图网络

该层网络主要是为了捕捉多个特征之间的相关性以及它们对于人群行为的影响, 因此这里将多特征时间序列看做一个完整的图, 图中的每个节点的表示一个特征, 节点的特征向量为对应特征的时间序列向量, 每条边反映的是特征之间的相关性. 具体来说, 图中每个节点  $i \in \{1, 2, \dots, k\}$  的特征向量  $x_i$  是对应特征的时间序列, 即  $x_i = \{x_{i,t} | t \in [0, n]\}$ , 其中  $k$  为特征的数量,  $n$  为选择的时间窗口的长度. 如图 3 所示为一个特征因果图网络的结构示例, 图中  $F_i, i = 1, 2, \dots, 5$  即表示节点,  $\alpha_{i5}$  表示节点  $F_i$  对于节点  $F_5$  的因果影响因子,  $h_5$  表示节点  $F_5$  的输出.

### 2.3 时间序列因果图网络

该层网络主要是为了捕捉特征在时间序列上的演

变对于人群行为的影响, 因此该层网络将每个时间戳的特征向量作为一个图节点, 图中的边反映的是时间序列的相关性. 具体的对于图中的每个节点  $t \in \{1, 2, \dots, n\}$ , 其特征向量的表示为  $c_t = \{c_{t,i} | i \in [0, k]\}$ . 如图 4 所示为一个时间序列因果图网络的结构示例.

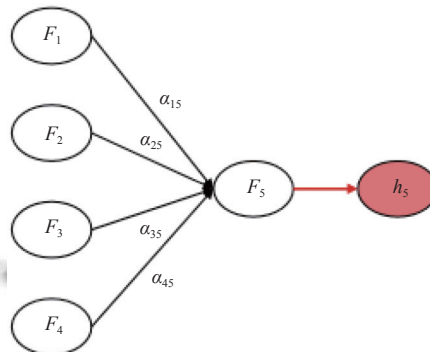


图3 特征因果图网络结构示例

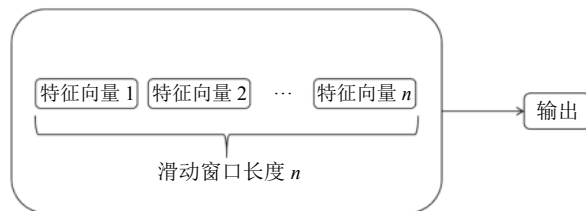


图4 时间序列因果图网络结构示例

## 3 实验

### 3.1 数据集与实验设置

实验采用 motion emotion dataset (MED) 数据集,

该数据集包含了 31 个视频, 共有大概 44 000 个视频剪辑, 视频中人群行为明显且对于每帧画面都有相应的人群行为标注. 正如该数据集中对于人群行为的划分, 本文将人群行为分为惊慌失措、打架、拥挤、异常物体和正常行为这 5 类.

在实验过程中, 首先对 MED 中的视频片段进行预处理筛选, 将不含有人群的片段删除, 预处理之后, 获取到有效的视频片段共 39 166 个. 在所有的训练过程中, 将所有有效数据按照 8:1:1 的比例分割成训练数据集, 验证数据集以及测试数据集, 另外, 实验中的时间窗口均设置为 64, 初始学习率设置为 0.000 6.

### 3.2 实验结果与分析

实验对比模型包括基于模糊识别的认知深度预测模型 (FCM-CDPM)<sup>[8]</sup> 和基于 3DCNN 和 SVM 的人群行为识别模型 (3DCNN-SVM)<sup>[25]</sup>. 对于模型的性能, 可以从模型收敛性, 识别准确率和 Kappa 系数来衡量. 此外, 实验中还对模型中两个并行的因果图网络层对于模型性能的影响进行了研究.

#### 3.2.1 与其他方法的对比

实验结果如表 1 所示. 从结果可以看到, 相对于其它方法, 本文模型在人群行为识别准确率上有很明显的提升. 但是在实验过程中, 各种类的人群行为数据集数量存在的明显的差异, 因此对于数据数量较多的行为类型, 该种类行为识别准确率会对模型整体识别准确率起到一个主导作用. 另外从各模型识别结果的 Kappa 系数可以看出本文模型在人群行为识别工作上的一致性性能相对其他方法也有较明显的提升. 图 5 给出了 3 个模型在训练过程中损失函数的收敛曲线, 从损失值的降低以及曲线的收敛速度可以看出本文模型相对于其他两种模型的优越性.

表 1 本文模型与其它模型的准确率与 Kappa 系数

方法	准确率 (%)	Kappa系数
FCM-CDPM <sup>[8]</sup>	93.8	0.8801
3DCNN-SVM <sup>[25]</sup>	90.9	0.8446
本文模型去除时间序列CGN	97.9	0.9582
本文模型去除特征CGN	98.4	0.9697
本文模型	99.1	0.9835

#### 3.2.2 双重并行因果图网络的影响

本节实验中, 通过分别去除特征因果图网络和时间序列因果图网络来检验这两个因果图网络对于人群行为识别的作用, 实验结果如表 1 所示, 可以看到, 去

除特征因果图网络层对于模型的整体性能影响比较大, 准确率下降了 1.2%, Kappa 系数也降低了 0.025 3. 去除时间序列图网络层也会降低模型的识别能力, 准确率下降了 0.7%, Kappa 系数降低了 0.013 8. 这说明这两个并行的因果图网络层对于模型识别能力都具有促进作用, 起到一个相辅相成的效果.

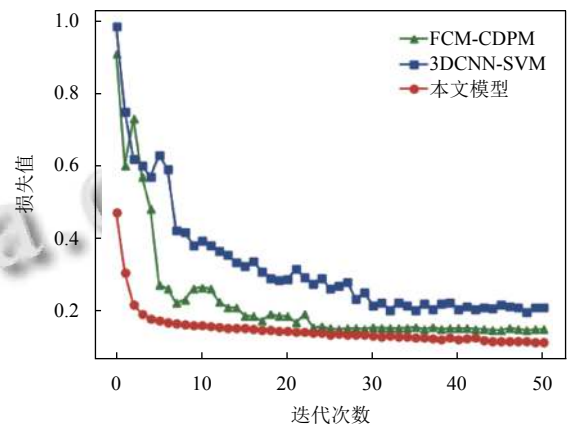


图 5 模型损失值收敛曲线

## 4 总结与展望

本文提出了一种融合多特征与时间序列的人群行为识别模型, 模型综合考虑人群的物理运动特征, 社交特征, 情绪人格特征以及背景文化特征对于人群行为的影响以及特征之间的相关性, 得到一个人群特征关系模型, 并在模型中采用两个并行的因果图网络层, 融合多特征相关性与时序依赖性对于人群行为的影响来识别人群行为. 在 MED 数据集进行实验, 并与其他文献中提到的优秀的人群行为识别模型框架对比, 发现本文模型在各方面表现都优于其他模型.

尽管本文模型在实验中有良好的表现, 但是还需要进一步的改进. 在本文的研究中, 考虑的是群体的整体行为, 不可否认的是群体中会有不同于群体的个体存在, 后续研究需要综合考虑群体行为与个体行为, 综合考虑两者来识别群体更加准确的发展趋势.

### 参考文献

- 1 Wu S, Yang H, Zheng SB, *et al.* Crowd behavior analysis via curl and divergence of motion trajectories. *International Journal of Computer Vision*, 2017, 123(3): 499–519. [doi: 10.1007/s11263-017-1005-y]
- 2 Thyagarajmurthy A, Ninad MG, Rakesh BG, *et al.* Anomaly

- detection in surveillance video using pose estimation. In: Sridhar V, Padma MC, Rao KAR, eds. *Emerging Research in Electronics, Computer Science and Technology*. Singapore: Springer, 2019. 753–766.
- 3 Chu WQ, Xue HY, Yao CW, *et al.* Sparse coding guided spatiotemporal feature learning for abnormal event detection in large videos. *IEEE Transactions on Multimedia*, 2019, 21(1): 246–255. [doi: [10.1109/TMM.2018.2846411](https://doi.org/10.1109/TMM.2018.2846411)]
- 4 Rabiee H, Haddadnia J, Mousavi H, *et al.* Novel dataset for fine-grained abnormal behavior understanding in crowd. *Proceedings of the 2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. Colorado Springs: IEEE, 2016. 95–101.
- 5 Hassner T, Itcher Y, Kliper-Gross O. Violent flows: Real-time detection of violent crowd behavior. *Proceedings of 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. Providence: IEEE, 2012. 1–6.
- 6 Su H, Yang H, Zheng SB, *et al.* The large-scale crowd behavior perception based on spatio-temporal viscous fluid field. *IEEE Transactions on Information Forensics and Security*, 2013, 8(10): 1575–1589. [doi: [10.1109/TIFS.2013.2277773](https://doi.org/10.1109/TIFS.2013.2277773)]
- 7 Thampi SM, Varghese EB. IoT-based smart surveillance: Role of sensor data analytics and mobile crowd sensing in crowd behavior analysis. In: Al-Sakib KP, ed. *Crowd-Assisted Networking and Computing*. Boca Raton: CRC Press, 2018. 45–76.
- 8 Varghese E, Thampi SM, Berretti S. A psychologically inspired fuzzy cognitive deep learning framework to predict crowd behavior. *IEEE Transactions on Affective Computing*, 2020. [doi: [10.1109/TAFFC.2020.2987021](https://doi.org/10.1109/TAFFC.2020.2987021)]
- 9 Zou HS, Su H, Song SH, *et al.* Understanding human behaviors in crowds by imitating the decision-making process. *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*. New Orleans: AAAI, 2018. 7648–7656.
- 10 Challenger W, Clegg WC, Robinson AM. Understanding crowd behaviours: Guidance and lessons identified. *Technical Report*, Leeds: University of Leeds. 2009. 11–13.
- 11 Baumeister RF, Vohs KD, Nathan DeWall C, *et al.* How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation. *Personality and Social Psychology Review*, 2007, 11(2): 167–203. [doi: [10.1177/108868307301033](https://doi.org/10.1177/108868307301033)]
- 12 Forsyth DR. *Group Dynamics*. 7th ed., Belmont: Cengage Learning, 2018.
- 13 Chattaraj U, Seyfried A, Chakroborty P. Comparison of pedestrian fundamental diagram across cultures. *Advances in Complex Systems*, 2009, 12(3): 393–405. [doi: [10.1142/S0219525909002209](https://doi.org/10.1142/S0219525909002209)]
- 14 Nazir A, Enz S, Lim MY, *et al.* Culture-personality based affective model. *AI & Society*, 2009, 24(3): 281–293.
- 15 Hofstede G. *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations*. 2nd ed., Thousand Oaks: Sage Publications, 2001.
- 16 Hofstede G. Dimensionalizing cultures: The Hofstede model in context. *Online Readings in Psychology and Culture*, 2011, 2(1): 1–26. [doi: [10.9707/2307-0919.1014](https://doi.org/10.9707/2307-0919.1014)]
- 17 Wojke N, Bewley A, Paulus D. Simple online and realtime tracking with a deep association metric. *Proceedings of 2017 IEEE International Conference on Image Processing (ICIP)*. Beijing: IEEE, 2017. 3645–3649.
- 18 Liu XH, Ma L, Guo JW, *et al.* Parallel computation of 3D clipped Voronoi diagrams. *IEEE Transactions on Visualization and Computer Graphics*, 2022, 28(2): 1363–1372. [doi: [10.1109/TVCG.2020.3012288](https://doi.org/10.1109/TVCG.2020.3012288)]
- 19 Hall ET. *The Hidden Dimension*. New York: Doubleday, 1966.
- 20 Favaretto RM, Musse SR, Costa AB. Emotion, personality, and cultural aspects in crowds. In: Favaretto RM, Musse SR, Costa AB, eds. *Emotion, Personality and Cultural Aspects in Crowds*. Cham: Springer, 2019. 23–33.
- 21 Ortony A, Clore GL, Collins A. *The Cognitive Structure of Emotions*. Cambridge: Cambridge University Press, 1990.
- 22 Mehrabian A. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 1996, 14(4): 261–292. [doi: [10.1007/BF02686918](https://doi.org/10.1007/BF02686918)]
- 23 Gebhard P. ALMA: A layered model of affect. *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multiagent Systems*. The Netherlands: ACM, 2005. 29–36.
- 24 Wiggins JS. *The Five-factor Model of Personality: Theoretical Perspectives*. New York: The Guilford Press, 1996.
- 25 Varghese EB, Thampi SM. A deep learning approach to predict crowd behavior based on emotion. *Proceedings of the 1st International Conference on Smart Multimedia*. Toulon: Springer, 2018. 296–307.

(校对责编: 孙君艳)