基于深度学习和智能规划的行为识别

郑兴华[1],孙喜庆[1],吕嘉欣[1],鲜征征[2],李 磊[1]
(1. 中山大学数据科学与计算机学院,广东广州 510006;2. 广东金融学院互联网金融与信息工程学院,广东广州 510521)

摘 要: 现有行为识别方法在未能持续覆盖造成视频监控盲区所引起行为数据缺失的情况,难以有效实施特征分析、行为分类补全,无法准确识别出智能体完整的行为动作序列. 为此,本文提出一种基于深度学习和智能规划的行为识别方法. 首先,利用深度残差网络对图像进行分类训练,然后使用递归神经网络对图像特征进行提取深度信息以增强分类效果;其次,运用智能规划的 STRIPS(Stanford Research Institute Problem Solver)模型,将深度学习提取的图像特征命题信息转化为规划领域的模型描述文档,并使用前向状态空间搜索规划器推导出完整的行为动作序列. 在 HMDB51 等行为识别公共数据集中,本方法与生成式对抗网络、深度卷积逆向图网络、深度信念网络、支持向量机等同类先进方法相比展现出更好的性能.

关键词: 行为识别; 深度学习; 智能规划; 深度残差网络; 递归神经网络; STRIPS 规划模型; 前向状态空间搜索规划器

中图分类号: TP302 **文献标识码:** A **文章编号:** 0372-2112 (2019)08-1661-08

电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2019.08.008

Action Recognition Based on Deep Learning and Artificial Intelligence Planning

ZHENG Xing-hua[1], SUN Xi-qing[1], LU Jia-xin[1], XIAN Zheng-zheng[2], LI Lei[1]
(1. *School of Data and Computer Science, Sun Yat-sen University, Guangzhou, Guangdong* 510006, *China*;
2. *School of Internet Finance and Information Engineering, Guangdong University of Finance, Guangzhou, Guangdong* 510521, *China*)

Abstract: Currently, action recognition methods can hardly carry out feature analysis, behavior classification, and action completion, and are incapable of accurately identifying the complete behavioral action sequence of intelligent agent for the discontinuous and incomplete motion capture, behavioral data missing or even broken in the time dimension, which are resulted from sensor device not being continuous coverage caused by the monitoring blind area. In this regard, we put forward a method of action recognition based on deep learning and artificial intelligence planning. Firstly, a deep learning network is constructed, by which the image is classified and trained using DRN(Deep Residual Network). After that, the extraction depth information of image frame feature for recurrent neural network is trained to enhance the classification effect. Secondly, the STRIPS(Stanford Research Institute Problem Solver) planning model is used to extract the image feature of deep learning, transforming into the description document for domain model, which facilitates deriving the optimal planning solution by means of forward state-space search planner. In the experiment, we exhibit that our method outperforms baselines in the public datasets, e.g., DCIGN(Deep Convolutional Inverse Graphics Networks), GAN(Generative Adversarial Networks), DBN(Deep Belief Networks), and SVM(Support Vector Machine).

Key words: action recognition; deep learning; artificial intelligence planning; deep residual network; recurrent neural network; STRIPS planning model; forward state-space search planner

1 引言

行为识别(action recognition)是指从观察、捕捉到智能体的行为动作或者行为效果出发,准确完整识别出智能体的行为动作序列或其行为目标的过程. 当前,大多数的行为识别方法采取计算机视觉的技术对视

收稿日期:2018-05-18;修回日期:2019-01-02;责任编辑:蓝红杰
基金项目:广东省自然科学基金(No. 2017A030313391);广东省科技厅国际合作项目(No. 2017A050501042)

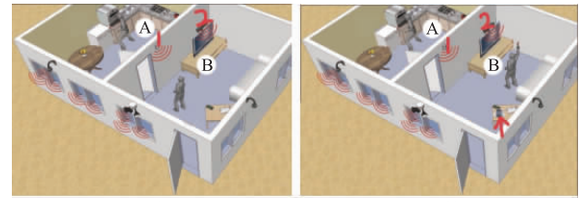
频、图像等多媒体数据进行分析处理,从中识别出智能体的行为动作,并按相应的环境信息对行为动作的目的、过程及效果进行语义描述[1,2]. 随着大数据的蓬勃发展,行为识别已成为机器学习、人机交互、模式识别及数据挖掘等领域的研究热点[3,4].

行为识别方法主要分为提取特征和动作识别及理解两个阶段. 现实中,其总会面对空间复杂性(智能体动作不连续)及时间差异性(时间维度上数据缺失)等问题. 如,面对传感器、摄像头等设备未能无缝或持续覆盖而造成的监控盲区,所引起的捕捉智能体动作不连续不完整、时间维度上行为数据缺失,从而无法准确识别出智能体完整的行为动作序列和目标过程;再如,卷积神经网络算法受自身结构及计算复杂度约束,并不适合处理行为识别的时间序列问题,而利用递归神经网络,则计算量不足较难实现大规模的运算,均会导致行为动作识别准确度降低,等等.

深度学习(deep learning)是一种基于对海量数据实施表征学习的机器学习方法[5],其通过重构含有多层隐藏层的机器学习模型以及学习大量的训练样本,获取更具价值的特征来进行精准分类或预测[6]. 自 Hinton 等人提出系统概念后[7],迅速成为行为识别领域前沿焦点[8,9]. 智能规划(artificial intelligence planning)指在实施某项行动或完成某件事情之前,对解决问题以及所选用的处理方式进行预判分析,并制订对应的步骤计划在行动前设计好操作步骤,是一种问题求解的科学方法,目的就是运用人工智能领域的理论、知识和技术,半自定或自动地生成一系列动作序列,以此实现期望或计划的目标[10,11]. 当前,使用 STRIPS(Stanford Research Institute Problem Solver)规划模型来解释动作模型间的逻辑关系[12],在以命题逻辑推导智能体动作过程方面极具优势[13]. 因此,本文选择深度学习的特征提取与智能规划的 STRIPS 模型[12,14]相结合的方法来解决行为识别问题.

2 相关工作

目前,研究行为识别问题较成熟的有三类方法[15]:概率推理方法[16]、模板匹配方法[17]、逻辑推理方法[18]. 但随着行为识别研究的持续深入,场景日益复杂多变,不断出现因传感器、摄像头未能无缝或持续覆盖而造成监控盲区的情况,上述方法就无法对待识别智能体活动做出准确完整的行为识别[1]. 以下场景为例,房 B 设两个摄像头,房 A 未设摄像头,在 $T1$ 和 $T2$ 时刻摄像捕捉到的图像为图 1(a)和(b)所示,即行为人移至电视机前,其右下角桌上增加了一杯咖啡,但上述方法无法识别出 $T1$ 至 $T2$ 间人的行为动作. 这里运用逻辑推理的知识,从 $T1$ 和 $T2$ 捕捉的状态变化推理出人行为动作是:从房 B 走到 A,从抽屉中取咖啡,把咖啡放进咖啡机中煮熟,再从房 A 走回 B,把咖啡放在桌上. 这种行为识别方法是可靠、有效的.



(a) (b)

图1 应用场景

在行为识别中常用的深度学习方法有四种[5]:一是有监督卷积神经网络(Convolution Neural Network, CNN):使用时空 3D 核对视频图像实施多次卷积、批量归一以及下采样等操作以获取全局特征表达[19,20];二是基于自编码的深度神经网络:采用“让输出等于输入”思想的无监督学习方法,当隐藏层多于输入时得到稀疏编码,当隐藏层小于输入时降维[2];三是基于受限玻尔兹曼机的深度置信网络:由隐藏的输出神经元和可见的输入神经元构成的概率生成模型[21,22]. 四是递归神经网络(Recurrent Neural Network,RNN):通过将状态在自身网络中循环传递的方式接受时间序列结构输入,从而解决随递归而引起的权重爆炸、消失以及无法捕捉长期时间关联等问题[23,24].

3 基于深度学习和智能规划的行为识别方法

本文提出一种基于深度学习和智能规划的行为识别方法(Action Recognition Based on Deep Learning and Artificial Intelligence Planning,ARDLAP),旨在利用深度残差网对图像进行分类训练,再使用递归神经网络对图像特征提取深度信息以增强分类效果,然后运用 STRIPS 模型,将深度学习提取的图像特征命题信息转化为规划领域的模型描述文档,最后通过前向状态空间搜索规划器推导出完整的行为动作序列.

3.1 算法框架

ARDLAP 算法框架分两部分:一是深度学习,目的是提取图像的特征强化分类;二是智能规划,目的是识别智能体的完整行为动作序列,具体如算法 1 所示:

算法 1 基于深度学习和智能规划的行为识别方法

输入:初始状态的图像和目标状态的图像
带有标签的训练集和测试集
对象类型文件、状态谓词文件
动作描述文件、状态命题文件
输出:从初始状态至目标状态间的行为动作序列
步骤:
1: 设置深度学习网络的参数(迭代次数、网络层数)

2:　利用深度残差网提取视频帧特征
3:　运用多层递归神经网络模型对特征向量强化分类
4:　　For each layer
5:　　设置各层节点数量,初始化权值和向量
6:　　在递归神经网络模型上重载训练数据
7:　　End for
8:　通过初始及目标状态的图像特征获得状态的命题信息
9:　使用 PDDL 语言建立各类 xml 文档
10:　生成领域描述文档以及问题描述文件
11:　通过前向状态空间搜索规划器推导最优规划解

3.2　深度学习的深网架构

深度学习分两阶段实施:一是使用卷积神经网络对视频图像进行特征提取;二是通过递归神经网络对提取的图像特征实施行为动作分类. 结构如图 2 所示:

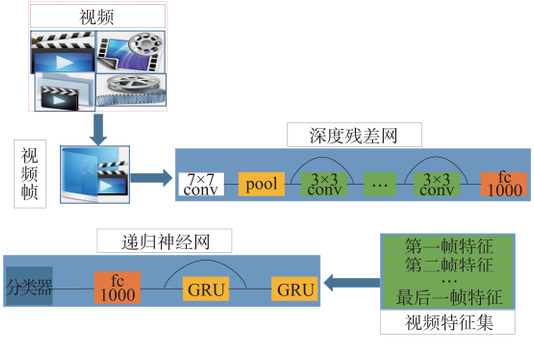


图2　深度学习的算法结构

3.2.1　使用卷积神经网络实施视频帧的特征提取

使用卷积神经网络的深度残差网[25](ResNet)进行深度学习,以此提取视频帧特征. 既考虑获取更好的拟合目标函数,又避免出现训练数据量不足情况,选取 ResNet50 作为图像特征提取的深度残差网络.

深度残差网的参数设置和操作内容包括:(1)卷积层. 为有效提取图像特征,通过设置深度残差网的卷积层来实施卷积操作,并采取共享权值的方式来减少网络参数个数. (2)池化层. 为降低图像的维数,运用局部平移不变性原则,通过深度残差网的池化层进行降维. (3)激活函数. 为确保网络收敛更快的同时防止过度拟合,选取 ReLU 激活函数,使得深度残差网能有效处理非线性问题. (4)全连接层. 为将每个神经元全部有效地连接起来,通过设置深度残差网的全连接层来进行图像特征的融合. (5)批正则化. 为确保训练时深度残差网的参数持续改变,通过设置批正则化对数据进行平移和放缩操作,以解决泛化能力不足和梯度发散等问题.

具体步骤为:(1)将视频图像按规则分解成为视频帧的序列($P_1,\cdots,P_n$);(2)将标签按顺序标注在每段视频帧上;(3)将已注有标签的视频帧当作样本用于训练 ResNet50,并对其再次训练直至收敛;(4)将视频帧按顺序输入至重新加载训练完的 ResNet50 中;(5)将 ResNet50 全连接层所输出的向量($X_1,\cdots,X_n$)作为视频帧的特征向量,并按原视频帧的时间序列保存. 上述训练过程本质就是对视频帧实施特征分类的过程.

3.2.2　使用递归神经网络实施图像特征的强化分类

选取递归神经网(RNN)中的基于门控递归单元(GRU)实施行为动作的强化分类,以解决图像的时间序列相关问题,本质就是破解长期依赖的难题[26]. GRU 就是在长短期记忆网络[23]基础上,将输出信息与隐藏状态合在一起,收敛效果好且不影响输出结果.

递归神经网对图像特征实施强化分类的步骤为:(1)为发挥相邻视频帧图像的内容相似特性,在设置两个 GRU 后,再添加一个残差结构;(2)为有效提高深度学习中特征融合的成效,在添加的残差结构后,再并接一个全连接层;(3)为全面推广到多元分类的程度,在引入的全连接层后,再增加一个 Logistic 回归分类器,并选取反向传播算法实施训练;(4)为减少深度学习所产生的过度拟合问题,选取带偏好的剪枝方式来减少训练参数. 结构如图 3 所示:

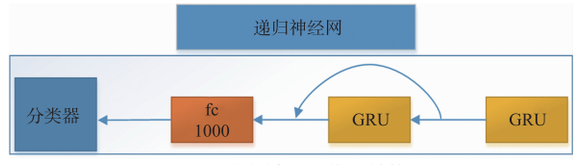


图3　递归神经网络的结构

3.3　智能规划的 PDDL 领域模型

深度学习获得图像特征分类后,智能规划分四个阶段实施:一是获取图像特征的命题信息;二是建立各类 xml 文件;三是生成领域模型描述文档;四是使用规划器推导行为序列.

3.3.1　获取图像特征的命题信息

使用分层表示的方法(即,以一个宏观的词组来描述同一类对象的行为),对数据集内的不同状态进行谓词识别. 如,选取 on、at 两个谓词表示多个智能体间的位置关系,故状态场景(图 4 所示)中的图像特征描述就转化为表 1 的命题信息来表示. 再将表中含 not 的命题去掉 not 后,相应改变该命题值域的真假值,并根据图像的次序至上而下列出命题值域所组成的二进制串“0001011101101010”. 故此,把命题信息与深度学习提取的标签形成一一对应关系,即训练网络的输出是二进制串长度个神经元,每个输出的神经元就代表一个命题值域,神经元的值就代表命题的真假.

图4 初始状态场景

表 1 图像的特征描述及相应的命题信息

图像特征描述	命题信息
(chair at room*A*)	0
(book on table*A* at room*A*)	0
(box ontable*A* at room*A*)	0
(poton table*A* at room*A*)	1
(curtain at room*A*)	0
(chair at room*B*)	1
(book on table*B* at room*B*)	1
(box on table*B* at room*B*)	1
(pot on table*B* at room*B*)	0
(curtain at room*B*)	1
(robot at room*A*)	1
(robot at room*B*)	0

3.3.2 各类 xml 文件的建立

STRIPS 规划模型使用 Planning Domain Definition Language(PDDL)语言[27]将图像特征的命题信息,分别写入 types. xml(对象分类)、actions. xml(动作描述)、predicates. xml(状态谓词)及 propositions. xml(状态命题)等文件中. 以下场景为例:有 *A*、*B* 两房,*A*、*B* 两桌子,1 个机器人. 机器人可在房 *A* 和 *B* 间移动,可用臂抓起或放下桌子. 在初始状态时机器人和桌子均在房 *A* 中,在目标状态时机器人将桌子移至房 *B* 中. 使用 PDDL 语言建立各类 xml 描述文件为:

(1)将对象分类写至 types. xml. 对象主要包含两类:对象类型(Houses、Tables、Robot)以及相应的对象名称(house*A* 和 house*B*,table*A* 和 table*B*,arms-left 和 arms-right). 故算法 2 描述如下:.

算法 2 STRIPS 规划问题的对象

```
1: (:objects houseA houseB
2: tableA tableB
3: arms-left arms-right )
```

(2)将状态谓词写至 predicate. xml. 如,at-robot(*x*)命题描述:若 robot 在房 *x* 中,则命题为真;at-table(*x z*)命题描述:若 *x* 是桌子,*y* 是房且 *x* 在 *y* 里,则命题为真;hold up(*x z*)命题描述:若 *x* 是臂抓,*z* 是桌且 *x* 举起 *z*,则命题为真. 故算法 3 ~ 5 描述如下:

算法 3 STRIPS 规划问题的谓词

```
1: (:predicates(at-robot ? -house)(at-table ? x-table ? z-house)
2: (free ? x-arms)(hold up ? x-arms ? z-table))
```

算法 4 STRIPS 规划问题的初始状态

```
1: (:init(free arms-left)(free arms-right)(at-robot house1)
2: (at-table table1 house1)(at-table table2 house1))
```

算法 5 STRIPS 规划问题的目标状态

```
1: (:goal(at-table table1 house2)(at- table table2 house2))
```

(3)将每个状态谓词的命题值域(二进制数)对应写至 predicates. xml 中每个状态谓词的命题属性(value),经重构后形成状态命题集合(propositions. xml).

(4)将动作描述写至 actions. xml. 如,移动操作指机器人在房 *x* 移至 *z*;举起操作指机器人 *z* 在房 *y* 中举起桌 *x*;放下操作指机器人 *z* 将桌 *x* 放在房 *y*. 定义一个动作需定义一个包含动作名称、前提条件及执行效果的三元组规划操作式,故算法 6 ~ 8 描述如下:

算法 6 STRIPS 规划问题的移动操作

```
1: (:action move to:parameters(? x-house ? z-house)
2: :precondition(at-robot ? x)
3: :effect(at-robot ? z)(not(at-robot ? x)))
```

算法 7 STRIPS 规划问题的举起操作

```
1: (:action hold up :parameters(? x- table ? y- house ? z-arms)
2: :precondition(at-table ? x ? y)(at-robot ? y)(free ? z)
3: :effect(hold up ? z ? x)(not(at-table ? x ? y)(not(free ? z))))
```

算法 8 STRIPS 规划问题的放下操作

```
1: (:action put down:parameters(? x-table ? y-house ? z-arms)
2: :precondition(and(hold up? z ? x)(at-robot ? y))
3: :effect(and(at-table? x ? y)(free ? z)(not(hold up? z ? x))))
```

3.3.3 领域模型描述文档的生成

生成领域模型文档分两部分完成:一是生成领域

描述文档(domain. pddl). 将上阶段建立的 types. xml、actions. xml、predicates. xml 等描述文件写入领域描述文档,分别表示为 D、T、A、P,则 $D = <T,A,P>$. 二是生成问题描述文档(problem. pddl). 将初始状态场景(图 4 所示)中所得命题信息逐位和状态命题集合(propositions. xml)的属性值域进行逻辑与操作,再逐位将属性值域是真的命题写至初始状态文档中,同理将目标状态场景(图 5 所示)对应属性值域是真的命题写至目标状态文档中,再分别将初始和目标状态文档对应读入问题描述文档的 Sinit 和 Sgoal 中.

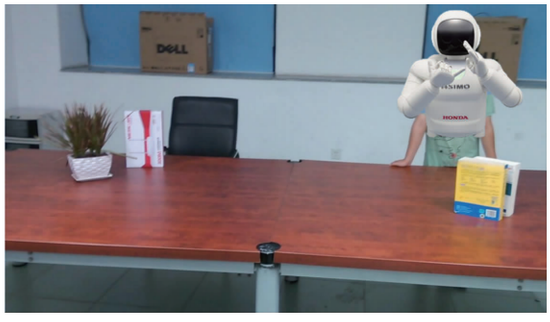


图5　目标状态场景

3.3.4　前向状态空间搜索算法的使用

将前阶段得到的领域描述文档和问题描述文档分别输入前向状态空间搜索规划器(Fast Forward State-Space Search Planner),以命题逻辑推理的方式推导运算[28,29]. 若输入的规划问题可解,那么规划器将输出最终的规划解,即一组从初始状态(Sinit)至目标状态(Sgoal)间智能体的行为动作序列,如表 2 所示:

表 2　规划器所输出的规划解

1	(robot_arm turn off curtain at room$A$)
2	(robot_arm hold up chair at room$A$)
3	(robot move room$A$ to room$B$)
4	(robot_arm put down chair at room$B$)
5	(robot_arm hold up box on table$B$ at room$B$)
6	(robot move room$B$ to room$A$)
7	(robot_arm put down box on table$A$ at room$A$)
8	(robot move room$A$ to room$B$)

4　实验设计与结果分析

实验在 Windows 7 下实施,8 核 3.6GHz 的 CPU,16GB 内存,GPU GTX1070,8GB 显存,实验语言 Python.

4.1　实验数据建模

在 HMDB51 等数据集上选取了 128 组不同状态的视频图像(每段时长 1800s,40fps). 图像模拟智能体在房中移动、举或放物品等行为. 对于不同状态,任一当前状态可通过智能体实施系列行为来实现另一目标状态,满足了 STRIPS 规划模型的命题要求.

4.2　实验评价标准

为充分体现 ARDLAP 应对复杂场景的识别性能,本文选取经优化的杰卡德距离(Jaccard distance)作为衡量各行为识别方法间差异的评价标准,其长度为行为动作序列所含有的动作数量. 距离公式如下所示:

$$1 = \frac{L_{\text{totalActions}}}{L_{\text{actualActions}} + L_{\text{ardlapActions}} - L_{\text{totalActions}}} \tag{1}$$

在式(1)中,$L_{\text{actualActions}}$指实际观察所得的行为动作数量,$L_{\text{ardlapActions}}$指 ARDLAP 所识别出的行为动作数量,$L_{\text{totalActions}}$指 $L_{\text{ardlapActions}}$ 和 $L_{\text{actualActions}}$ 两个序列合计的行为动作数量,? 为行为动识别的准确度.

4.3　实验结果分析

在深度学习获取图像特征分类阶段,选取深度卷积逆向图网络(Deep Convolutional Inverse Graphics Networks,DIN)[30,31]、生成式对抗网络(Generative Adversarial Networks,GAN)[32]、深度信念网络(Deep Belief Networks,DBN)[33]、支持向量机(Support Vector Machine,SVM)[34]等 4 个主流的深度学习神经网络,并在此基础上形成基于深度卷积逆向图网络与智能规划的行为识别方法(ARDINAP)、基于生成式对抗网络与智能规划的行为识别方法(ARGANAP)、基于深信网络与智能规划行为识别方法(ARDBNAP)、基于支持向量机与智能规划的行为识别方法(ARSVMAP)用于对比实验. 在每组实验中,各使用 ARDLAP 与四类方法运行 40 次,以计算式(1)均值. 训练样本为 2000 张图像,微调迭代数量为 400 次,如图 6 所示:

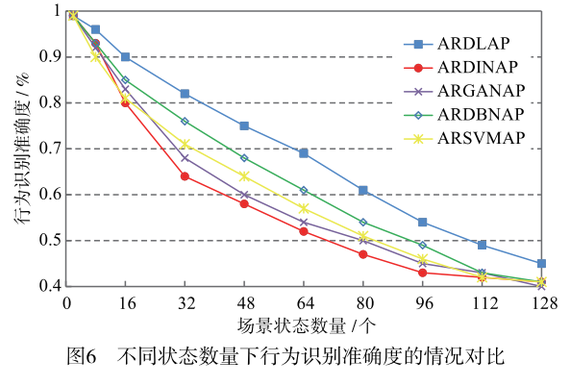


图6　不同状态数量下行为识别准确度的情况对比

从图 6 可知,当场景所含状态数量是 8 个时,ARDLAP 行为识别准确度超过 95%,明显比其余四类方法要高,且随着状态数量增加准确度也始终高于其余四类的行为识别准确度,从而验证了 ARDLAP 实效性. 故此,将进一步在场景状态数量不同情况下,分析准确度与图像数量、微调迭代数间的关系.

(1)行为识别准确度与样本数量的对比

在每组实验中,各使用 ARDLAP 与四类方法运行 40 次以计算(1)式平均值. 训练样本数量为 2000、1800、1600、1400、1200、1000、800 及 600,微调迭代次数为 400. 在场景所含状态数是 16、32、48、64、80 及 96 的情况下,样本数量与行为识别准确度间关系如图 7 ~ 12 所示:

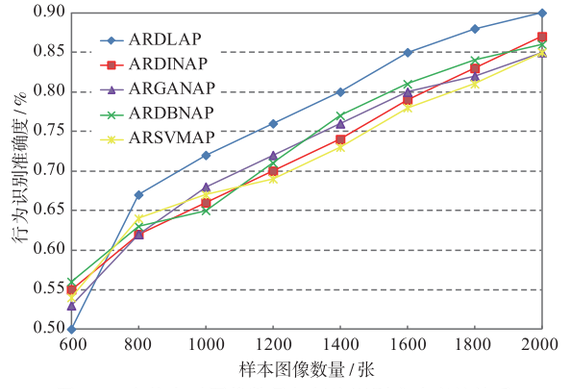


图7 16个状态时图像数量与行为识别准确度的关系

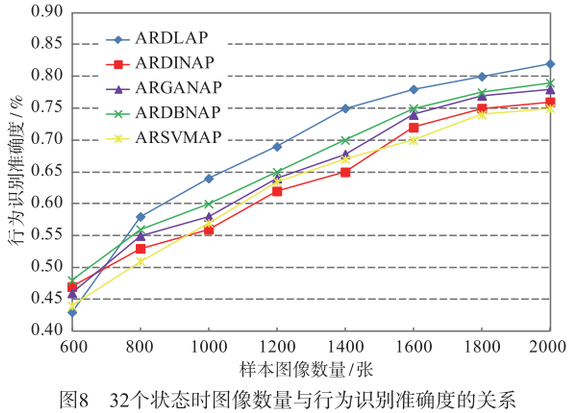


图8 32个状态时图像数量与行为识别准确度的关系

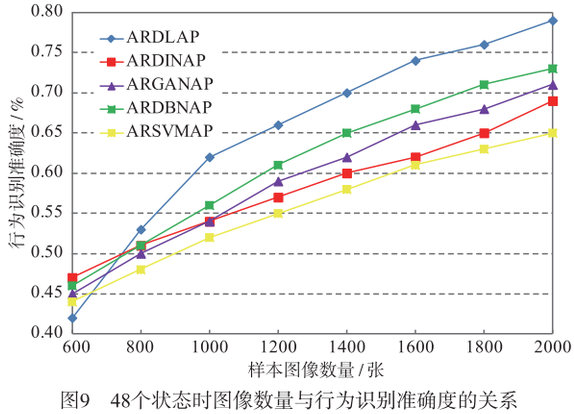


图9 48个状态时图像数量与行为识别准确度的关系

从图 7 ~ 12 可知,在不同场景状态个数下,随着样本所含图像数量的持续增加,ARDLAP 与其余方法的行为识别动作准确率均有所提升. 在样本数量达 800 张时,ARDLAP 行为识别准确率已超过其余方法,并随着

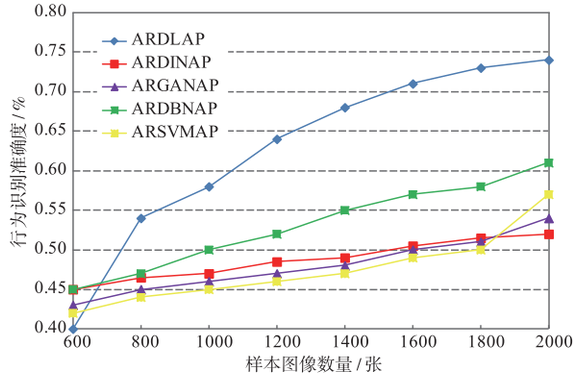


图10 64个状态时图像数量与行为识别准确度的关系

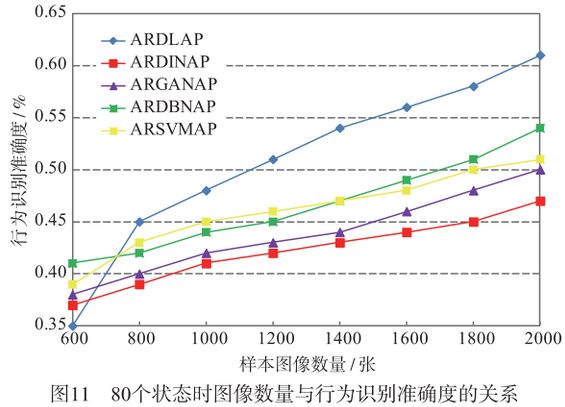


图11 80个状态时图像数量与行为识别准确度的关系

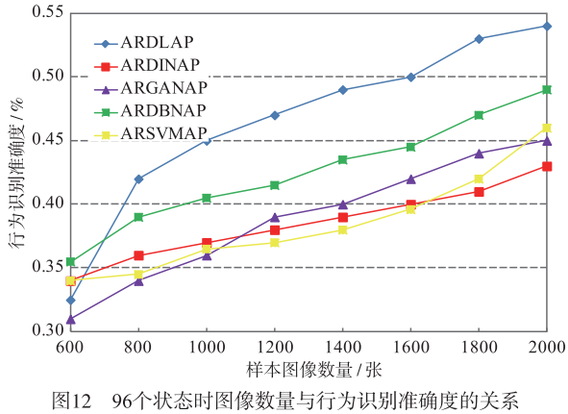


图12 96个状态时图像数量与行为识别准确度的关系

图像数量的增加,准确率差距亦越趋增大. 故此,ARDLAP在大数据、云计算背景下展现出更好性能.

(2)行为识别准确度和微调迭代数量的对比

在每组实验中,各使用 ARDLAP 与四类方法运行 40 次以计算(1)式平均值. 训练样本数量为 2000,微调迭代次数是 400、350、300、250、200、150、100 及 50. 在场景所含状态数是 16、32、48、64 及 96 的情况下,微调迭代数量与行为识别准确度间关系图 13 所示:

从图 13 可知,ARDLAP 的行为识别准确度随微调

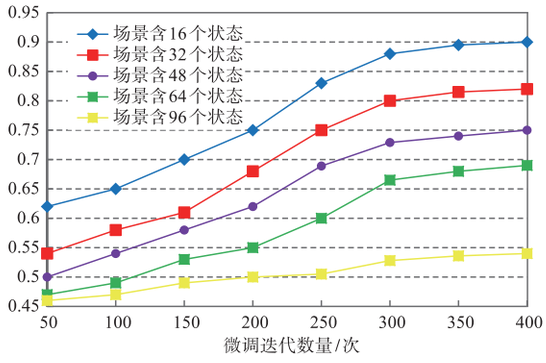


图13　微调迭代次数和行为识别准确度间关系图

次数增多而增大,在迭代次数达 300 次后,ARDLAP 的行为识别准确度曲线趋于平缓. 此时,由微调次数增多带来的时间代价已大于其所带来的准确度提高. 因此,选择合适的微调次数对整个算法性能至关重要.

5　总结

本文提出基于深度学习和智能规划的行为识别方法,首先利用深度残差网对图像进行分类训练,再使用递归神经网络深度提取图像特征信息以增强分类效果,然后运用 STRIPS 模型将深度学习提取的图像特征命题信息转化为规划领域的模型描述文档,最后通过规划器逻辑推导出完整的行为动作序列. 实验结果表明,本方法相比于同类先进方法展现出更好的性能. 下一步,对于行为未事先定义及场景过于开放等情况下,后续进入场景的智能体行为未能准确识别的问题,将研究如何消除假设条件,使方法具有更好的健壮性.

参考文献

[1] João Carreira, Andrew Zisserman. Quo Vadis. Action recognition? a new model and the kinetics dataset[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Washington, DC, USA: IEEE, 2017. 4724 – 4733.

[2] 张友梅,常发亮,刘洪彬. 基于 3D 人体骨架的动作识别[J]. 电子学报,2017,45(4):906 – 911.
ZHANG You-mei, CHAN Fa-liang, LIU Hong-bin. Action recognition based on 3D skeleton[J]. Acta Electronica Sinica, 2017, 45(4): 906 – 911. (in Chinese)

[3] Christoph Feichtenhofer, Axel Pinz, RichardWildes, Andrew Zisserman. What have we learned from deep representations for action recognition? [A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Washington, DC, USA: IEEE, 2018. 7844 – 7853.

[4] Du Tran, et al. A closer look at spatiotemporal convolutions for action recognition[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Washington, DC, USA: IEEE, 2018. 6450 – 6459.

[5] Chen B, Ting J A, De Freitas N. Deep learning of invariant spatio-temporal features from video[A]. Conference and Workshop on Neural Information Processing Systems[C]. Cambridge, Massachusetts, USA: MIT Press, 2010. 46 – 57.

[6] Jurgen Schmidhuber. Deep learning in neural networks: An overview[J]. Neural Networks, 2015, 61: 85 – 117.

[7] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 50: 313 – 325.

[8] Shaoqing Ren, Kaiming He, Ross B. Girshick, Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks[A]. Conference and Workshop on Neural Information Processing Systems[C]. Cambridge, Massachusetts, USA: MIT Press, 2015. 91 – 99.

[9] 李倩玉,蒋建国,齐美彬. 基于改进深层网络的人脸识别算法[J]. 电子学报,2017,45(3):619 – 625.
LI Qian-yu, JIANG Jian-guo, QI Mei-bin. Face recognition algorithm based on improved deep networks[J]. Acta Electronica Sinica, 2017, 45(3): 619 – 625. (in Chinese)

[10] 曾霖,卓汉逵,李磊. 基于智能规划的工作流任务识别算法[J]. 2018,46(4):871 – 877.
ZENG Lin, ZHUO Han-kui, LI Lei. Workflowtask recognition based on intelligent planning[J]. Acta Electronica Sinica, 2016, 44(8): 2025 – 2032. (in Chinese)

[11] 高洁,卓汉逵,刘亚松,李磊. 基于众包模式的开放式规划问题研究[J]. 电子学报,2016,44 (8):2025 – 2032.
GAO Jie, ZHUO Han-kui, LIU Ya-seng, LI Lei. Research on crowdsourced open planning[J]. Acta Electronica Sinica, 2016, 44 (8): 2025 – 2032. (in Chinese)

[12] Fikes, R. E., Nilsson, N. I. STRIPS: A new approach to the application of theorem proving to problem solving[J]. Artificial Intelligence, 1972, 2: 189 – 208.

[13] Hankz Hankui Zhuo, Qiang Yang, Subbarao Kambhampati. Action-model based multi-agent plan recognition[A]. Conference and Workshop on Neural Information Processing Systems[C]. Cambridge, Massachusetts, USA: MIT Press, 2012. 377 – 385.

[14] Hankz Hankui Zhuo, Tuan Nguyen and Subbarao Kambhampati. Refiningincomplete planning domain models through plan traces[A]. International Joint Conferences on Artificial Intelligence[C]. San Jose, California, USA: Margan Kaufmann, 2013. 2451 – 2457.

[15] Thomas B. Moeslund, et al. A survey of advances in vision-based human motion capture and analysis[J]. Computer Vision and Image Understanding, 2006, 104(2): 90 – 126.

[16] E. Charniak and R. P. Goldman. A Bayesian model of plan recognition[J]. Artificial Intelligence, 1993, 64: 53 – 79.

[17] Tao Gu, et al. SICAR: An emerging patterns based ap-

proach to sequential, interleaved and concurrent activity r. ecognition[A]. International Conference on Pervasive Computing and Communications[C]. Washington, DC, USA: IEEE Computer Society, 2009. 142 – 151.
[18] Kautz H, Allen J F. Generalized plan recognition[A]. AAAI Conference on Artificial Intelligence[C]. Palo Alto, California, USA: AAAI, 1986. 167 – 178.
[19] Ji S, Xu W, Yang M, et al. 3D convolutional neural networks for human action recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 221 – 231.
[20] Simonyan K, Zisserman A. Two-stream convolutional networks for action recognition in videos[A]. Conference and Workshop on Neural Information Processing System[C]. Cambridge, Massachusetts, USA: MIT Press, 2014. 568 – 576.
[21] Bengio Y, Delaalleau O. On the expressive power of deep architectures[A]. International Conference on Algorithmic Learning Theory[C]. Berlin, GER: Springer, 2011. 18 – 36.
[22] Taylor G W, Hinton G E. Factored conditional restricted Boltzmann machines for modeling motion style[A]. International Conference on Machine Learning[C]. San Jose, California, USA: ACM, 2009. 1025 – 1032.
[23] Shuai Zheng, et al. Conditional random fields as recurrent neural networks[A]. IEEE International Conference on Computer Vision[C]. Washington, DC, USA: IEEE, 2015. 1529 – 1537.
[24] Jamil Ahmad, Khan Muhammad, Muhammad Sajjad, Sung Wook Baik. Action recognition in video sequences using deep bi-directional LSTM with CNN features[J]. IEEE Access, 2018, 6(1): 1155 – 1166.
[25] A. Krizhevsky, I. Sutskever, G. Hinton. ImageNet classification with deep convolutional neural networks[A]. Conference and Workshop on Neural Information Processing Systems[C]. Cambridge, Massachusetts, USA: MIT Press, 2012. 1273 – 1291.
[26] LeCun Y, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural Computation, 1989, 1(4): 541 – 551.
[27] Ghallab M, Aeronautiques C, Isi C K, et al. PDDL-The Planning Domain Definition Language[M]. Boston: Auerbach Publications, 1998.
[28] Henry A. Kautz, Bart Selman. Pushingthe envelope: planning, propositional logic and stochastic search[A]. AAAI Conference on Artificial Intelligence[C]. Palo Alto, California, USA: AAAI, 1996. 1194 – 1201.
[29] Javier Segovia Aguas, Sergio Jiménez Celorrio, Anders Jonsson. Hierarchical finite state controllers for generalized planning[A]. International Joint Conference on Artificial Intelligence[C]. San Jose, California, USA: Margan Kaufmann, 2016. 3235 – 3241.
[30] Chao Dong, Chen Change Loy, Kaiming He, Xiaoou Tang. Imagesuper-resolution using deep convolutional networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(2): 295 – 307.
[31] 王泽宇, 吴艳霞, 张国印, 布树辉. 基于空间结构化推理深度融合网络的 RGB-D 场景解析[J]. 电子学报, 2018, 46(5): 1253 – 1258.
WANG Ze-yu, WU Yan-xia, ZHANG Guo-yin, BU Shu-hui. RGB-D scene parsing based on spatial structured inference deep fusion networks[J]. Acta Electronica Sinica, 2018, 46(5): 1253 – 1258. (in Chinese)
[32] Xinhua Liu, Yao Zou, Chengjuan Xie, Xiaolin Ma. Bidirectional face aging synthesis based on improved deep convolutional generative adversarial networks[J]. Information, 2019, 10 (2): 69 – 84.
[33] Joe Yue-Hei Ng, et al. Beyond short snippets: Deep networks for video classification[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Washington, DC, USA: IEEE, 2015. 4694 – 4702.
[34] P. Shih, C. Liu. Face detection using discriminating feature analysis and support vector machine[J]. Pattern Recognition, 2006, 39 (11): 260 – 276.

作者简介

郑兴华 男, 1983 年生于广东广州. 中山大学数据科学与计算机学院博士. 主要研究方向为智能规划、数据挖掘、模式识别、神经网络等.
E-mail: zhengxh5@mail3.sysu.edu.cn

鲜征征(通信作者) 女, 1977 年出生于四川阆中. 中山大学数据科学与计算机学院博士, 现为广东金融学院讲师, CCF 会员. 主要研究方向为数据挖掘、隐私保护、模式识别等.
E-mail: xianzhengzheng@126.com