李斌, 霍光磊, 黄承曦. 基于深度强化学习的人体动作分割方法 [J]. 智能计算机与应用,2026,16(5):49-53. DOI:10.20169/j.issn.2095-2163.24073101

基于深度强化学习的人体动作分割方法

李 斌[1], 霍光磊[2,3], 黄承曦[2,3]
(1 西安航天化学动力有限公司, 西安 710025;
2 福建(泉州)先进制造技术研究院, 福建 泉州 362200; 3 福建省智能运维机器人技术重点实验室, 福建 泉州 362200)

摘 要: 现有人体动作分割方法忽略了动作视频的固有时空上下文信息,导致在空间复杂性大、时间冗余性高的情况下动作分割精度不高。提出将动作分割建模为马尔可夫过程,并建立深度强化学习框架,捕捉动作视频的全局结构关联,并通过反馈机制迭代实现准确的动作分割。首先,构建一个时序图卷积网络来表示动作视频的时空相关性,并通过均匀采样进行动作分割过程的初始化;然后,通过自适应地学习奖励来调整初始策略,以实现模型的训练。该强化学习框架可提取动作视频的潜在深度特征,同时,为提高网络的效率,将每个动作视频的组成单位约束为图像序列。最后,通过融合特征选择策略和特征识别的损失函数进行联合学习。通过在 ActivityNetv1.3 和 THUMOS14 数据集上的实验验证了所提方法的可行性和优越性。

关键词: 动作分割; 强化学习; 深度学习; 图卷积神经网络
中图分类号: TP183 **文献标志码**: A **文章编号**: 2095-2163(2026)05-0049-05

Human motion segmentation based on deep reinforcement learning

LI Bin[1], HUO Guanglei[2,3], HUANG Chengxi[2,3]
(1 Xi′an Aerospace Chemical Propulsion Co., Ltd., Xi′an 710025, China;
2 Fujian (Quanzhou) Institute of Advanced Manufacturing Technology, Quanzhou 362200, Fujian, China;
3 Fujian Key Laboratory of Intelligent Operation and Maintenance Robot Technology, Quanzhou 362200, Fujian, China)

Abstract: Existing human action segmentation methods ignore the inherent spatiotemporal context information of action video, which leads to low precision of motion segmentation under the condition of large spatial complexity and high time redundancy. In this paper, action segmentation is modeled as a Markov process, and a deep reinforcement learning framework is established to capture the global structural relationship of action video, and the accurate action segmentation is achieved iteratively through a feedback mechanism. First, a Temporal Graph Convolutional Network (TGCN) is constructed to represent the spatiotemporal correlation of action videos, and the action segmentation process is initialized by uniform sampling. Then, the initial strategy is adjusted by adaptively learning the reward to achieve the training of the model. The proposed reinforcement learning framework can extract the potential depth features of action videos, and constrain the processing units of each action video to image sequences to improve the efficiency of the network. Finally, joint learning is carried out by fusing the feature selection strategy and the loss function of feature recognition. The feasibility and superiority of the method are verified on ActivityNetv1.3 and THUMOS14 datasets.

Key words: action segmentation; reinforcement learning; deep learning; graph convolutional neural network

0 引 言

基于视觉的动作分割是人体动作理解的关键环节,也是机器视觉领域的研究热点之一。动作分割的目的是从未修剪的视频中分割出动作片段并同时对每个片段进行分类,在机器人人机交互、安防、视频索引与监控等领域具有广泛的应用前景[1-4]。不同于动作识别任务,动作分割面对的是未经修剪的长视频,存在大量冗余与复杂信息,因此更具挑战性[5-6]。动作分割的核心问题是动作特征的时空表示,为从速度、长度、视角、背景及边界等存在极大差异的视频中提取有效的时空特征[7-8],国内外学者

基金项目: 福建省泉州市科技计划项目(2023C009R)。
作者简介: 李 斌(1976—),男,学士,高级工程师。主要研究方向:固体推进剂装药。
通信作者: 黄承曦(1987—),男,博士,高级工程师。主要研究方向:机器人感知技术。Email:chengxi_huang@163.com。
收稿日期: 2024-07-31

开展了大量研究,然而,由于动作类间的相似性与类内的多样性,动作分割问题仍未得到有效解决。

动作分割技术可分为基于手工特征和基于深度学习两类方法。在基于手工特征的方法中,均匀采样和随机连续采样是在简单背景下广泛采用的有效方案。然而,在实际应用中面对动作未对齐的场景时,基于采样的方法准确性较低。因此,郑元杰等[9]提出将滑动窗口从空间域扩展到时间域进行序列选择,并通过结合描述符和分类器优化滑动窗口的设置。基于手工特征方法中,动作特征提取器通常采用经典的动作识别算法,例如,胡铭轩等[10]先提取人体动作的骨骼点特征,再通过高斯混合模型对动作行为进行建模,并结合动态时间规划算法对动作进行匹配和分割。鉴于深度神经网络具备优越的非线性表征能力,近期的动作分割方法大多基于深度学习技术,其中,卷积神经网络(CNN)和循环神经网络(RNN)以其卓越的空间和时间表征能力被广泛采用。由于动作是时空信号,吕淑平等[11]提出基于双流卷积神经网络对动作的光流和图像特征进行深度学习,以实现动作的时空表征和识别。毛琳等[12]通过对卷积神经网络进行时序扩展,构建多阶段边界参考网络实现动作分割。余金锁等[13]在卷积神经网络的基础上,结合双向长短期记忆(Long Short Term Memory, LSTM)网络对人体行为进行特征提取,并通过注意力机制实现动作的识别。尽管深度神经网络在动作识别领域取得了显著成果,但动作分割问题具有时空随机性,因此融合强化学习解决该问题具有更好的应用前景,例如,李学明等[14]利用分数阶网络对图像实例进行分割,并将分割结果输入到强化学习策略中进行奖励或惩罚,完成图像实例分割模型的构建。然而,现有基于深度强化学习的动作分割方法没有考虑动作间的相互关联,精度难以满足实际需求。

基于以上分析,本文重点关注动作片段间的时空结构关系,并将动作分割问题建模为强化学习过程,如图 1 所示。首先,将视频片段选择策略作为深度强化学习中需要优化的动作,定义奖励函数为各检测片段的分类结果,并通过均匀采样进行初始化。强化学习框架的观测信息由图卷积网络提取的时空特征提供,并通过使用卷积神经网络作为骨干网络来构建图卷积神经网络。图卷积神经网络的特征矩阵通过将每个片段定义为一个节点来构建,边表示片段间的连接,邻接矩阵则通过计算分割片段与真实标签之间的时序交集并结合欧几里得距离来构建。最后,通过全连接层后,使用 Softmax 和交叉熵的累积差异和分类损失来建立策略损失,实现网络参数的学习。本文主要贡献包括:(1)构建一种基于强化学习框架的动作分割框架;(2)提出一种表征动作片段间时空相互关系的图卷积神经网络;(3)在公共数据库上取得了良好的动作分割结果。

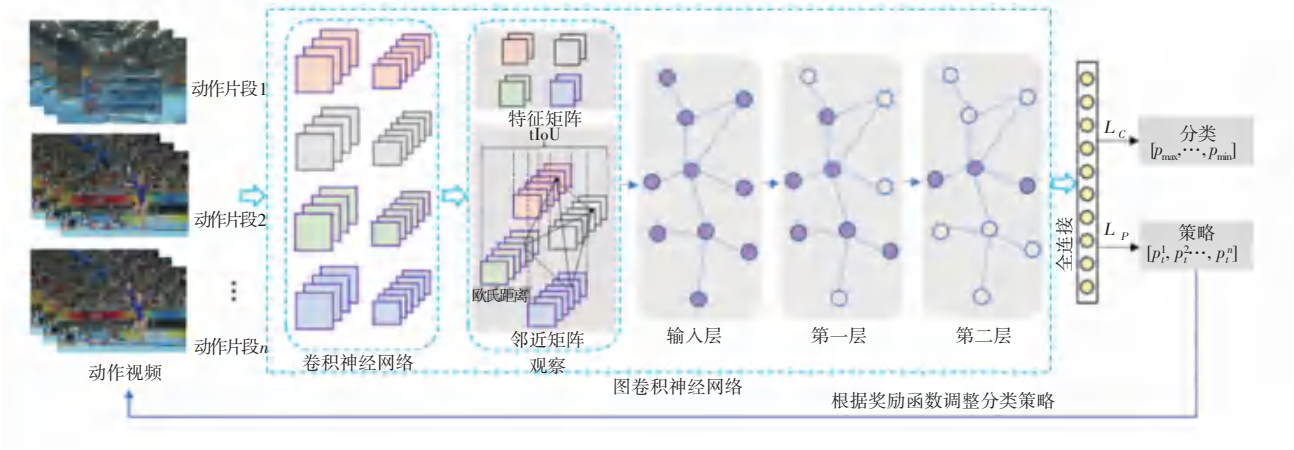


图 1 基于强化学习的动作分割框架

Fig. 1 Action segmentation framework based on reinforcement learning

1 动作分割模型观察

给定一个长度为 L、包含动作片段与间隙的未修剪视频 $V=\{I_t \in \mathbb{R}^{H\times W\times L}\}_{t=1}^{T}$,其中 I_t 是 t 时刻的帧,其图像高度和宽度分布为 H 和 W。首先,对于动作分割,通过均匀采样将动作片段初始化为 $V=(V_1,V_2,\cdots,V_T)$,其中 T 是由一系列动作片段或间隙的起始和结束时间确定的参数,选择方案可以详细描述为:

$$T=\{(t_1^s,t_1^e),(t_2^s,t_2^e),\cdots,(t_n^s,t_n^e)\} \quad (1)$$

其中，t_i^s 和 t_i^e 分别表示时间区间 $i \in [1,n]$ 的起始和结束帧的索引号。利用 ResNet-152 对 I_t 进行特征提取，可以获得每帧的深层特征，表示为 $P = \{p_i \mid p_i = (x_i,(t_i^s,t_i^e))\}_{i=1}^N$，其中 $x_i \in \Re^d$ 是特征映射。然后，为了描述动作帧的全局关系，采用图卷积神经网络 $G=(N,E)$ 构建 x_i 的拓扑图，学习用于分割选择策略和动作片段分类的图卷积神经网络特征 $\hat{y}_i$，将 P 定义为节点 N，而 E 被定义为 P 的连接，因此，定义强化学习的观察为：

$$F^{(l+1)} = \sigma(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\,\widetilde{D}^{-\frac{1}{2}}F^{(l)}W^{(l)}) \tag{2}$$

其中，σ 表示非线性激活函数；$F^{(l)}$ 表示第 l 层的卷积神经网络特征；$\widetilde{A} = A + I$，A 和 I 分别是邻接矩阵和对应的单位矩阵；$\widetilde{D}$ 是 A 的度矩阵；$W^{(l)}$ 是待学习的权重。根据图卷积神经网络的前向传播规则，对于每个 x_i，其映射函数为：

$$\{(\hat{t}_i^s,\hat{t}_i^e)\mid\hat{y}_i\}_{i=1}^N = \sigma(\mathrm{GCN}(\{x_i\}_{i=1}^N,G(P,E))) \tag{3}$$

其中，将每个特征映射作为图卷积神经网络的节点，对于任意两个节点特征 f_i 和 f_j，为了计算 A，在时间域内计算二者的欧几里得距离。此外，由于视频往往很长，导致计算量大，同时全局关系也未被充分捕捉，定义动作片段和间隙在一定程度上是连续的，将特征计算的对象从帧转换为序列，并将序列的长度定义为 5，以平衡精度与计算效率。同时，采用 tIoU 来描述节点之间的关系，通过融合 tIoU 和欧几里得距离，可以得到：

$$\xi_{ij}^D = (f_i)^{\mathrm{T}}\cdot f_j \times \frac{[\gamma + \mathrm{tIoU}(f_i,f_j)]}{[\chi + \|f_i - f_j\|_2]} \tag{4}$$

其中，χ 和 γ 均为常数。因此，公式(2)可以化简为：

$$F^{(l+1)} = \sigma(W^{l+1}(F^{(l)})) \tag{5}$$

其中：

$$F^{(l)} = f_i + \sum_{j=1}^{N-1}\xi_{ij}T_{ij}(f_i) \tag{6}$$

其中，T_{ij} 表示将 f_j 映射到 f_i 的转换矩阵。

2　动作分割模型损失函数

为了训练构建的图卷积神经网络，本文在强化学习框架下设定总体目标：最大化动作分割选择的奖励，同时最小化分割动作的分类损失。通过从图卷积神经网络获取全连接层特征，对由 θ_π 和 θ_p 参数化的策略模块和分类模块进行训练，首先考虑预测标签 p_t 与真实标签 g_t 之间的差异，并将两个连续时刻之间的奖励表示为：

$$r_t = |p_t - g_t| - |p_{t-1} - g_{t-1}| \tag{7}$$

因此，从 0 到 t 的奖励函数为：

$$R_t = \sum_{i=0}^{T_{\mathrm{stop}}-t}\varepsilon^i r_{t+i} \tag{8}$$

其中，$\varepsilon \in (0,1]$ 表示尺度因子。给定视频的动作空间 U，深度强化学习动作策略的目标定义为：

$$J(\theta_\pi) = \sum_{\mu\in U}\sum_{i=1}^{t}\pi(\mu \mid s;\theta_\pi)R_t \tag{9}$$

利用梯度下降算法，可推导网络参数的梯度为：

$$\nabla_{\theta_\pi}J(\theta_\pi) = \sum_{\mu\in U}\sum_{i=1}^{t}\pi(\mu \mid s;\theta_\pi)\nabla_{\theta_\pi}\pi(t \mid s;\theta_\pi)R_t \tag{10}$$

然而，由于 U 的维度较高，$\nabla_{\theta_\pi}J(\theta_\pi)$ 无法直接求解，因此采用蒙特卡罗采样来估计带有 K 个切片的梯度：

$$\nabla_{\theta_\pi}J(\theta_\pi) \approx \frac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{t}\sum_{\mathrm{stop}_t=0}^{T}\nabla_{\theta_\pi}\log\pi(\mu_{t,k} \mid s_{t,k};\theta_\pi)R_t \tag{11}$$

因此，策略网络的损失函数为：

$$\mathcal{L}_J(\theta_\pi) = \frac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{t}\sum_{t=0}^{T_{\mathrm{stop}}}\log\pi(\mu_{t,k} \mid s_{t,k};\theta_\pi)R_t \tag{12}$$

为增强强化学习框架的探索能力，将 $\mathcal{L}_J(\theta_\pi)$ 与 $\mathcal{L}_E(\theta_\pi)$ 进行融合，定义总体损失为：

$$\mathcal{L}_p(\theta_\pi) = \mathcal{L}_J(\theta_\pi) + \lambda\mathcal{L}_E(\theta_\pi) \tag{13}$$

其中，λ 为平常超参数，而 $\mathcal{L}_E(\theta_\pi)$ 为探索损失，其完整形式为：

$$\mathcal{L}_E(\theta_\pi) = -\sum_{\mu\in U}\sum_{t=0}^{T_{\mathrm{stop}}}\pi(\mu_t \mid s_t;\theta_\pi)\log\pi(\mu_t \mid s_t;\theta_\pi) \tag{14}$$

此外，考虑分割分类损失 $\mathcal{L}_C(\theta_p)$，采用交叉熵来最小化预测标签 p 和真实标签之间的差异。最终，网络的损失函数为

$$\mathrm{Loss} = \mathcal{L}_C(\theta_p) + \delta\mathcal{L}_P(\theta_\pi) \tag{15}$$

其中，δ 为常数。

3　实验

3.1　数据库及配置

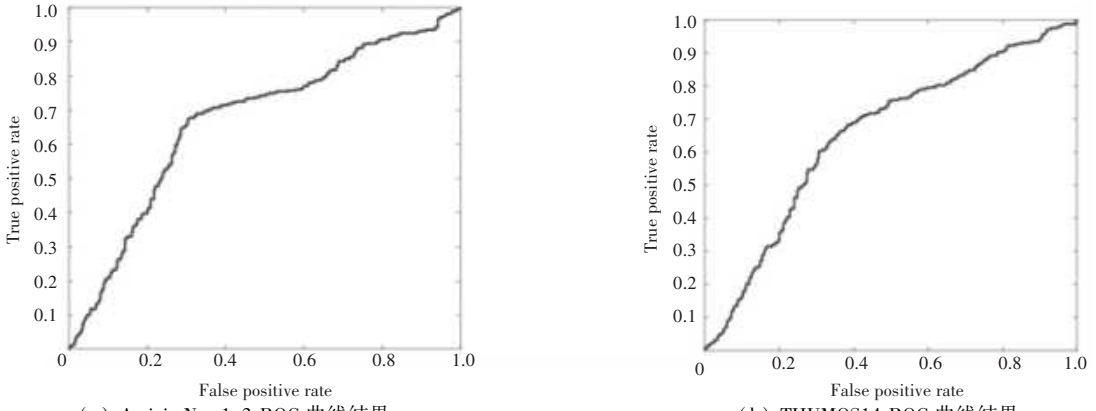
选取 ActivityNetv1.3[15] 和 THUMOS14[16] 作为基准来验证所提方法。ActivityNetv1.3 是一个包含 19 994 个视频的大型动作数据集，动作被分为 200 个类别，包括 849 h 的未分割视频，每个视频大约包

含 1.4 个动作片段。该数据集被划分为 10 000 个视频的训练集、6 000 个视频的测试集和 500 个视频的验证集。按照默认设置选择视频来训练和测试所提网络，并使用均值平均精度（mean Average Precision，mAP）和受试者工作特征（Receiver Operating Characteristic，ROC）曲线来对不同方法进行比较。THUMOS14 由 413 个动作视频组成，共分为 20 个类别。该数据集包含 213 个视频组成的测试集和 200 个视频组成的验证集。

为获取用于图卷积神经网络输入的卷积神经网络特征，选择 ResNet-152 作为骨干网络。在 Kinetics 数据集上对 ResNet-152 进行预训练，并使用 Adam 优化器进行网络训练。学习率设置为 0.000 1，批量大小设置为 16。将动作片段数量 N 初始化为 10，将图卷积神经网络的隐含层数量设置为 1 024。因子常数 γ、χ 和 δ 分别设置为 0.9、1 和 1。对于不同长度的测试视频，随机采样 5 帧序列来构建剪辑，并确保剪辑对齐良好，手动删除长于 N 的剪辑，并使用重复采样来补偿短剪辑。

3.2 实验结果及分析

图 2 中给出了所提方法在 ActivityNetv1.3 和 THUMOS14 上利用分割的动作片段进行分类的 ROC 曲线，其 mAP 分别为 85.96% 和 81.28%。为对比所提方法与其他方法的性能，首先考虑改变骨干网络，采用不同的架构，并尝试不同的预训练数据集，其结果见表 1。由表 1 可知，使用 ResNet-152 的结果优于使用 BN-Inception，在相同预训练数据集下，ResNet-152 的特征提取能力更强。同时，基于 Kinetics 数据集预训练的结果远优于同架构在 ImageNet 上的预训练结果，表明基于相似任务的数据集的迁移学习对模型性能提升更为重要。



（a）ActivityNetv1.3 ROC 曲线结果　　（b）THUMOS14 ROC 曲线结果

图 2 ROC 曲线结果

Fig. 2 ROC curve results

表 1 不同框架及预训练策略的动作分割结果对比（mAP）

Table 1 Comparison of action segmentation results between different frameworks and pre-training strategies (mAP)

框架	预训练数据集	ActivityNetv1.3	THUMOS14
BN-Inception	ImageNet	81.99	76.73
ResNet-152	ImageNet	83.54	78.36
BN-Inception	Kinetics	85.01	80.88
ResNet-152	Kinetics	85.96	81.28

不同算法的比较见表 2，由表 2 可知 C3D[17] 的结果最差（在 ActivityNetv1.3 上为 67.68%，在 THUMOS14 上为 63.92%），表明基于三维卷积的迁移学习不足。通过采用在 Kinetics 上预训练的架构，R-C3D[18] 获得了更有前景的结果（mAP 分别为 78.86% 和 74.29%）。基于二维卷积，TSN[19] 的表现不及 RRA（分别为 76.56%、83.42% 和 70.34% vs 80.46%）。

表 2 不同方法在数据库上的动作分类结果（mAP）

Table 2 Action classification results of different methods on the database (mAP)

方法	骨干网络	ActivityNetv1.3	THUMOS14
C3D[17]		67.68	63.92
TSN[18]	BN-Inception	76.56	70.34
R-C3D[19]	ResNet-152	78.86	74.29
RRA[20]	ResNet-152	83.42	80.46
本文方法	ResNet-152	85.96	81.28

4 结束语

面向机器视觉中的动作分割任务，本文提出了一种基于图卷积神经网络的深度强化学习框架。首先，构建了一个图卷积网络以表示时空特征作为强

化学习的观测信息，用于捕捉动作片段之间的全局内在关联。将分割及分类结果作为奖励对动作分割策略进行定义和反馈，将分割视频与真实视频的时序交集和欧几里得距离的融合作为损失函数。最终，通过策略网络与分类网络的联合学习完成模型训练。在 ActivityNetv1.3 和 THUMOS14 数据集上，所提方法的 mAP 分别达到 85.96%和 81.28%，实验结果与对比分析验证了方法的有效性。后续研究将进一步探索所提方法在更多动作分割及识别数据集上的有效性，并验证模型的泛化能力。

参考文献

[1] PARK D, KIM H, KEMP C C, Multimodal anomaly detection for assistive Robots[J]. Autonomous Robots, 2019, 43: 611-629.

[2] 李国朋, 王连清, 韩鹏, 等. 多图融合约束半非负矩阵分解的动作分割方法[J]. 智能系统学报, 2023, 18(6): 1223-1232.

[3] BAI Yue, WANG Lichen, LIU Yunyu, et al. Human motion segmentation via velocity-sensitive dual-side auto-encoder[J]. IEEE Transactions on Image Processing, 2023, 32(2): 524-536.

[4] 赵琴, 赵团结, 郑新桥, 等. 基于运动历史图像的异常行为识别算法研究[J]. 智能计算机与应用, 2023, 13(10): 56-59.

[5] TAO Z, HUA Z F, CHEN G, et al. Consistency and diversity induced human motion segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(1): 197-210.

[6] 闫航, 陈刚, 崔莉亚, 等. 基于单目视觉的在线人体康复动作识别[J]. 计算机应用与软件, 2021, 38(2): 171-178.

[7] LI S, FARHA Y A, LIU Y, et al. MS-TCN++: Multi-stage temporal convolutional network for action segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(6):6647-6658.

[8] 陈雪健, 秦水介, 白忠臣, 等. 深度学习行为识别的挖掘机生产效率监控系统[J]. 智能计算机与应用, 2023, 13(4): 111-116.

[9] 郑元杰, 黄俊, 陈州全. 基于时空注意的毫米波雷达人体活动识别网络[J]. 计算机应用研究, 2023, 40(8): 2549-2554.

[10] 胡铭轩, 乔钧, 张执南. 连续康复训练动作分割与评估[J]. 上海交通大学学报, 2023, 57(5): 533-544.

[11] 吕淑平, 黄毅, 王莹莹. 基于双流卷积神经网络的人体动作识别研究[J]. 实验技术与管理, 2021, 38(8): 144-148.

[12] 毛琳, 曹哲, 杨大伟, 等. 多阶段边界参考网络的动作分割[J]. 光学精密工程, 2022,30(3): 340.

[13] 余金锁, 卢先领. 基于分割注意力的特征融合 CNN-Bi-LSTM 人体行为识别算法[J]. 电子测量与仪器学报, 2022, 36(2): 89-95.

[14] 李学明, 吴国豪, 周尚波, 等. 基于分数阶网络和强化学习的图像实例分割模型[J]. 计算机应用, 2022, 42(2): 10.

[15] HEILBRON F, ESCORCIA V. ActivityNet: A large-scale video benchmark for human activity understanding[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 961-970.

[16] JIANG Y, LIU J. Thumos challenge: Action recognition with a large number of classes[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2014: 472-479.

[17] TRAN D, BOURDEV L. Learning spatiotemporal features with 3D Convolutional Networks [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 4489-4497.

[18] ZHU C, TAN X. Fine-grained video categorization with redundancy reduction attention[C]// Proceedings of European Conference on Computer Vision. Cham: Springer, 2018: 139-155.

[19] WANG L, XIONG Y. Temporal segment networks: Towards good practices for deep action recognition[C]//Proceedings of European Conference on Computer Vision. Cham: Springer, 2016: 20-36.

[20] KIM J, CHI S. Action recognition of earthmoving excavators based on sequential pattern analysis of visual features and operation cycles[J]. Automation in Construction, 2019, 104: 255-264.