

DOI: 10.11830/ISSN.1000-5013.201911047



长短时记忆网络的自由体操 视频自动描述方法

贺 凤^{1,2,3}, 张 洪 博^{1,2,3}, 杜 吉 祥^{1,2,3}, 汪 冠 鸿^{1,2,3}

- (1. 华侨大学 计算机科学与技术学院, 福建 厦门 361021;
2. 华侨大学 福建省大数据智能与安全重点实验室, 福建 厦门 361021;
3. 华侨大学 厦门市计算机视觉与模式识别重点实验室, 福建 厦门 361021)

摘要: 提出一种长短时记忆网络的自由体操视频自动描述方法. 在视频描述模型 S2VT 中, 通过长短时记忆网络学习单词序列和视频帧序列之间的映射关系. 引入注意力机制对 S2VT 模型进行改进, 增大含有翻转方向、旋转度数、身体姿态等关键帧的权重, 提高自由体操视频自动描述的准确性. 建立自由体操分解动作数据集, 在数据集 MSVD 及自建数据集上进行 3 种模型的对比实验, 并通过计划采样方法消除训练解码器与预测解码器之间的差异. 实验结果表明: 文中方法可提高自由体操视频自动描述的精度.

关键词: 长短时记忆网络; 注意力机制; 自由体操; 自动描述

中图分类号: TP 183 **文献标志码:** A **文章编号:** 1000-5013(2020)06-0808-08

Floor Exercise Video Automatic Description Method Using Long Short-Term Memory Network

HE Feng^{1,2,3}, ZHANG Hongbo^{1,2,3},
DU Jixiang^{1,2,3}, WANG Guanhong^{1,2,3}

- (1. College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China;
2. Fujian Key Laboratory of Big Data Intelligence and Security, Huaqiao University, Xiamen 361021, China;
3. Xiamen Key Laboratory of Computer Vision and Pattern Recognition, Huaqiao University, Xiamen 361021, China)

Abstract: An automatic description method of floor exercise video based on long short-term memory network was proposed. In the video description model S2VT, learning the mapping relationship between word sequence and video frame sequence through long and short-term memory network. The attention mechanism was introduced to improve the S2VT model, increase the weight of key frames including turning direction, rotation degree and body posture, and improve the accuracy of automatic description of floor exercise video. The data set of floor exercise decomposition was established, and three models were compared among the data set MSVD and self built data set. The difference between the training decoder and the prediction decoder was eliminated by the scheduled sampling method. The experimental results showed that the proposed method can improve the accuracy of automatic description of floor exercise video.

收稿日期: 2019-11-19
通信作者: 张洪博(1986-), 男, 副教授, 博士, 主要从事计算机视觉、模式识别和图像视频分析的研究. E-mail: hong-bobest@gmail.com.
基金项目: 国家自然科学基金资助项目(61871196, 61673186); 福建省自然科学基金资助项目(2017J01110, 2019J01082)

Keywords: long short-term memory network; attention mechanism; floor exercise; automatic description

在现代人日益注重健康、推崇体育运动的大背景下,体育视频的内容分析与识别成为研究热点.然而,现有的体育视频研究多集中于体育运动的语义分析及识别方面,在体育运动的分类细项研究中,除足球、篮球、高尔夫球、羽毛球等球类运动外,其他运动项目少有涉及^[1-8].体育视频自动描述是对体育视频中人体动作的自动描述,是计算机视觉研究中的难点和热点之一.通过计算机视觉和模式识别等技术手段对视频中的人体运动进行分析,可以对视频序列中存在的特定人体运动进行智能化表示和标记.自由体操是竞技体操的典型代表,其自由度最大,动作变化较多,需在规定时间内按照规则表演多个动作组合的成套动作^[9].自由体操视频的自动描述研究具有极强的现实意义.对非专业人士而言,自动描述不仅能够提升观赛感受,也能增进观众对自由体操运动的了解;对专业人士而言,可以通过自动描述进行动作数据分析,挖掘自由体操技术创新发展的规律性特征,从而辅助训练.

为了实现对自由体操动作的自动描述,可使用卷积神经网络对自由体操视频特征进行分析^[10-19];基于分析得出的特征,通过长短时记忆(LSTM)网络^[20-22]实现特征到自然语言的映射,自动生成自由体操分解动作的描述.在数据标注的基础上,自由体操视频的自动描述问题可看成是视频字幕的自动生成问题.视频字幕自动生成的任务是自动生成视频内容的自然语言描述.S2VT^[23]是第一个序列到序列的视频描述模型,可实现视频帧序列的输入及文本序列的输出,且输入的视频帧和输出的文本长度可变.然而,对自由体操视频而言,分解动作的关键在于翻转方向、旋转度数、身体姿态,故将包含这些关键动作的视频帧定义为关键帧,通过提取视频中判别力较高的关键帧,提高视频自动描述的精度.本文基于S2VT模型,提出一种长短时记忆网络的自由体操视频自动描述方法.

1 自由体操视频自动描述方法

自由体操视频自动描述的基本框架,如图 1 所示.

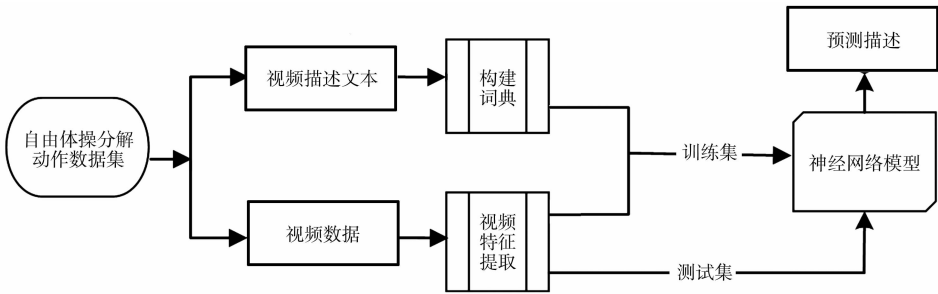


图 1 自由体操视频自动描述的基本框架

Fig. 1 Basic framework of automatic description of floor exercise video

1.1 特征提取

自由体操视频自动描述网络的输入数据包括视频序列和文本序列.通过卷积神经网络提取视频特征,并利用自然语言文本处理方式提取文本特征.

1.1.1 视频特征提取 卷积神经网络对几何变换、形变、光照具有一定程度的不变性,可用较小的计算代价扫描整幅图像^[24].因此,卷积神经网络被广泛地应用于图像特征提取.文中使用 VGG16^[25]卷积神经网络提取自由体操视频的特征.VGG16 网络结构图,如图 2 所示.图 2 中:ReLU 为线性整流函数;softmax 为激励函数.

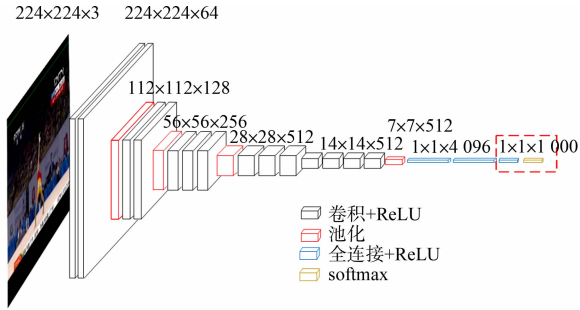


图 2 VGG16 网络结构图

Fig. 2 VGG16 network structure diagram

与 state-of-the-art 网络相比,VGG16 的错误率较低,拓展性较强,迁移到其他图片数据上的泛化性

非常好. VGG16 的结构简洁, 整个网络都使用相同大小的卷积核尺寸(3 px×3 px)和最大池化尺寸(2 px×2 px). VGG16 模型的输入为 227 px×227 px 固定尺寸的 RGB 图片, 输出为视频特征序列 $\mathbf{X}=(\mathbf{x}_1, \cdots, \mathbf{x}_n)$.

1.1.2 描述文本处理 通过 one-hot 向量编码的方式将自由体操视频的描述符转化为特征, 统计自由体操标注文本中的词语, 构建字典. 因为在自由体操分解动作描述中使用的字词种类数量不一, 所以在预处理时未对字词进行筛选.

one-hot 向量编码的输入是一句话, 输出是一个特征. one-hot 向量编码的计算方法: 先计算自由体操数据集中所有描述的单词总数 N ; 然后, 将每个单词表示为一个长 $1 \times N$ 的向量, 该向量中只有 0 和 1 两种取值, 且该向量中只有一个值为 1, 其所在位置即当单词在词表中的位置, 其余值为 0.

1.2 自由体操自动描述系统

通过长短时记忆网络实现视频特征到文本特征的学习. 循环神经网络(RNN)在反向传播过程中易产生梯度消失现象, 导致网络参数难以持续优化^[26], 而 LSTM 网络作为一种特殊的循环神经网络, 可以有效地解决该问题.

LSTM 在每个序列索引位置 t 时刻除了向前传播与 RNN 相同的隐藏状态 \mathbf{h}_t 外, 还多了另一个隐藏状态, 即细胞状态 \mathbf{C}_t . 此外, LSTM 网络通过增加门控单元控制即时信息对历史信息的影响程度, 使神经网络模型能够较长时间地保存并传递信息.

因此, LSTM 单元可以看作一个尝试将信息存储较久的记忆单元, 记忆单元被遗忘门(forget gate)、输入门(input gate)和输出门(output gate)保护, 并以此为基础, 实现有效信息的更新和利用.

LSTM 单元的更新方式为

$$\mathbf{f}_t = \sigma(\mathbf{W}_{f_1} \mathbf{h}_{t-1} + \mathbf{W}_{f_2} \mathbf{X}_t + \mathbf{b}_f), \tag{1}$$

$$\mathbf{i}_t = \sigma(\mathbf{W}_{i_1} \mathbf{h}_{t-1} + \mathbf{W}_{i_2} \mathbf{X}_t + \mathbf{b}_i), \tag{2}$$

$$\mathbf{a}_t = \tanh(\mathbf{W}_{a_1} \mathbf{h}_{t-1} + \mathbf{W}_{a_2} \mathbf{X}_t + \mathbf{b}_a), \tag{3}$$

$$\mathbf{C}_t = \mathbf{f}_t \odot \mathbf{C}_{t-1} + \mathbf{i}_t \odot \mathbf{a}_t, \tag{4}$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_{o_1} \mathbf{h}_{t-1} + \mathbf{W}_{o_2} \mathbf{X}_t + \mathbf{b}_o), \tag{5}$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh \mathbf{C}_t. \tag{6}$$

式(1)~(6)中: $\mathbf{f}_t, \mathbf{i}_t, \mathbf{a}_t, \mathbf{C}_t, \mathbf{o}_t, \mathbf{h}_t$ 分别为 t 时刻各状态的输出值; σ 为 Sigmoid 激活函数; $\mathbf{W}_{f_1}, \mathbf{W}_{i_1}, \mathbf{W}_{a_1}, \mathbf{W}_{o_1}, \mathbf{W}_{f_2}, \mathbf{W}_{i_2}, \mathbf{W}_{a_2}, \mathbf{W}_{o_2}$ 均为权重向量; $\mathbf{h}_{t-1}, \mathbf{C}_{t-1}$ 分别为 $t-1$ 时刻的隐藏状态和细胞状态; \mathbf{X}_t 为 t 时刻输入的视频序列; $\mathbf{b}_f, \mathbf{b}_i, \mathbf{b}_a, \mathbf{b}_o$ 均为偏移量; \tanh 为双曲正切激活函数; \odot 为向量元素的 Hadamard 积.

长短时记忆网络结构, 如图 3 所示. 图 3 中: $\mathbf{h}_N, \mathbf{h}_{t+1}, \mathbf{h}_n$ 分别为不同时刻的隐藏状态; $\mathbf{X}_N, \mathbf{X}_{t-1}, \mathbf{X}_t, \mathbf{X}_{t+1}, \mathbf{X}_n$ 分别为不同时刻的视频输入序列.

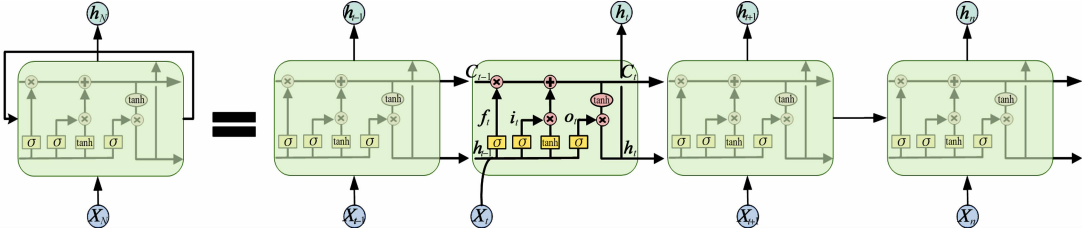


图 3 长短时记忆网络结构

Fig. 3 Long short-term memory network structure

将固定维度的自由体操分解动作特征向量 $\mathbf{X}=(\mathbf{x}_1, \cdots, \mathbf{x}_n)$ 编码成特征序列, 由此得到对应隐藏层的输出 $\mathbf{H}=(\mathbf{h}_1, \cdots, \mathbf{h}_n)$. 已知 LSTM 网络的输出取决于前一个输入序列, 将特征向量按顺序输入 LSTM 网络, 可得到一个序列向量的编码映射输出.

当最后一帧图片的特征向量输入后, LSTM 网络的输出即为帧序列编码. 解码阶段的 LSTM 送入起始符, 促使其将收到的隐藏状态解码成单词序列, 输出的单词序列 $\mathbf{Y}=(\mathbf{y}_1, \cdots, \mathbf{y}_m)$, 概率为 $p(\mathbf{y}_1, \cdots, \mathbf{y}_m | \mathbf{x}_1, \cdots, \mathbf{x}_n)$, 即

$$p(y_1, \cdots, y_m \mid x_1, \cdots, x_n) = \prod_{t=1}^m p(y_t \mid h_{n+t-1}, y_{n-1}). \tag{7}$$

解码阶段进行训练时,在已知帧序列的隐藏状态及之前输出单词的条件下,求预测句子的对数似然性. 训练目标是使待估参数值 θ 达到最大值 θ^* , 即

$$\theta^* = \arg \max_{\theta} \sum_{t=1}^m \log p(y_t \mid h_{n+t-1}, y_{n-1}, \theta). \tag{8}$$

采用随机梯度下降算法对整个训练数据集进行优化,使 LSTM 网络学习更合适的隐藏状态. 第二层 LSTM 网络的输出 z 在 Y 中寻找最大可能性的目标单词 y , 其概率 $p(y \mid z_t)$ 为

$$p(y \mid z_t) = \frac{\exp(W_y z_t)}{\sum_{y' \in Y} \exp(W_{y'} z_t)}. \tag{9}$$

式(9)中: z_t 为 t 时刻的输出; $W_y, W_{y'}$ 均为权重向量.

视频描述的编解码结构图,如图 4 所示.

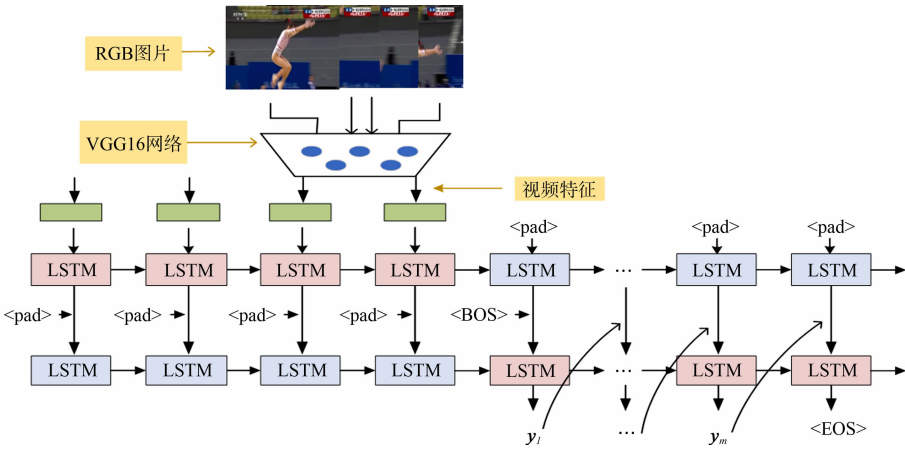


图 4 视频描述的编解码结构图

Fig. 4 Encoding and decoding structure diagram of video description

1.3 注意力机制

人脑对不同部分的注意力是不同的. 注意力机制是人类视觉特有的大脑信号处理机制. 人类视觉通过快速扫描全局图像,获得需要重点关注的目标区域,即注意力焦点,并对这一区域投入更多的注意力资源,以获取更多需要关注目标的细节信息,抑制其他无用信息.

对自由体操分解动作自动描述具有决定性作用的是关键帧,故关键帧的权重应该更大. 在网络结构中引入注意力机制^[27],可允许解码器对自由体操视频的每个时间特征向量进行加权. 采用时间特征向量的动态加权和公式为

$$\varphi_t(X) = \sum_{i=1}^n \alpha(t)_i x_i. \tag{10}$$

式(10)中: $\sum_{i=1}^n \alpha(t)_i = 1, \alpha(t)_i$ 为 t 时刻的隐藏层输出与整个视频表示向量的匹配得分占总体得分的比值,其计算公式为

$$\alpha(t)_i = \frac{\exp(\text{score}(x_i, h_i))}{\sum_{j=1}^n \exp(\text{score}(x_i, h_j))}. \tag{11}$$

式(11)中: $\text{score}(x_i, h_j)$ 为第 i 个隐藏层的输出 h_j 在视频特征向量 x_i 中所占的分值; $\text{score}(x_i, h_i)$ 为第 i 个隐藏层的输出 h_i 在视频特征向量 x_i 中所占的分值,分值越大,说明这个时刻的输入在该视频中的注意力越大.

$\text{score}(x_i, h_i)$ 的计算公式为

$$\text{score}(x_i, h_i) = w^T \tanh(Wx_i + Uh_i + b). \tag{12}$$

式(12)中: w, W, U 均为权重向量; b 为偏移量.

P 在训练的过程中是变化的. 一开始训练不充分, P 可以小一些, 尽量使用真实标注数据作为下一个预测的输入; 随着训练的进行, P 逐渐增大, 可以多采用自身的输出作为输入; 随着 P 越来越大, 训练解码器最终趋于预测解码器. 通过计划采样方法, 可缩小训练解码器和预测解码器之间的差异.

2.3 损失函数

使用 TensorFlow 开源软件库中的 TensorBoard 可视化工具, 读取 TensorFlow 训练后保存的事件文件, 可展示各个参数的变化. S2VT 模型(原始模型)和文中模型(经文中方法改进的模型)训练完成后, 通过损失函数了解 2 种模型的损失值随迭代次数的变化情况, 结果如图 6 所示. 图 6 中: d 为迭代次数; L 为模型训练损失值. 由图 6 可知: 随着迭代次数的增加, S2VT 模型与文中模型的函数损失值均逐渐下降, 最终趋于稳定收敛状态; 相较于 S2VT 模型, 文中模型因复杂度增加, 起始的损失值较大, 但其收敛速度有所提高. 因此, 文中模型具有可训练性及有效性.

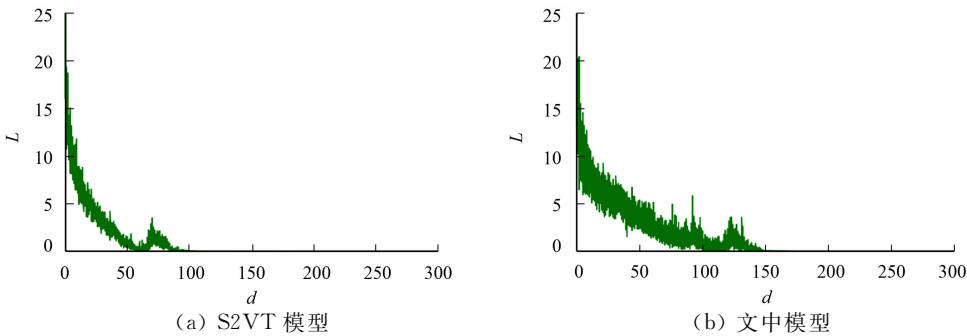


图 6 损失值随迭代次数的变化情况

Fig. 6 Change of loss value with number of iterations

2.4 评价指标及性能对比

自由体操视频自动描述的结果是自由体操分解动作的描述语句, 是一种自然语言, 因此, 评价自由体操视频自动描述结果可参考自然语言领域中机器翻译质量的评价指标, 即双语评估候补(BLEU). BLEU^[28]是目前最接近人类评分的评价指标, BLEU 重点关注相同文本下, 机器译文和参考译文的相似度, 其值越大, 则翻译质量越好.

表 2 不同模型的 BLEU 对比

Tab. 2 BLEU comparison of different models

数据集	BLEU		
	S2VT 模型	文中模型	计划采样模型
MSVD	17.2	17.9	18.8
OURS(1)	8.7	9.3	10.4
OURS(2)	10.9	11.2	12.3

在数据集 MSVD^[29] 和自建数据集的两个语料库 OURS(1), OURS(2) 中, 对 S2VT 模型、文中模型和计划采样模型(引入计划采样方法的文中模型)的性能进行对比. OURS(1) 采用最直接的自然语言; OURS(2) 则根据专业术语对描述语句进行调整. 不同模型的 BLEU 对比, 如表 2 所示. 由表 2 可知: 相较于 S2VT 模型, 文中模型和计划采样模型的性能更优.

由于篇幅所限, 仅给出一个实例作为参考. 自建数据集 OURS 上, 自由体操视频自动描述实例, 如图 7 所示. 图 7 中: 正确描述为 forward stretched twist three forward stretched twist one; S2VT 模型为 twohalf forward onehalf forward; 文中模型为 onehalf forward one stretched forward; 计划采样模型为 two stretched <PAD> stretched <PAD>. 由图 7 可知: 文中模型在翻转方向方面的测试结果与正确描述相近, 如“forward”; 计划采样模型在身体姿势方面的测试结果与正确描述相近, 如“stretched”.

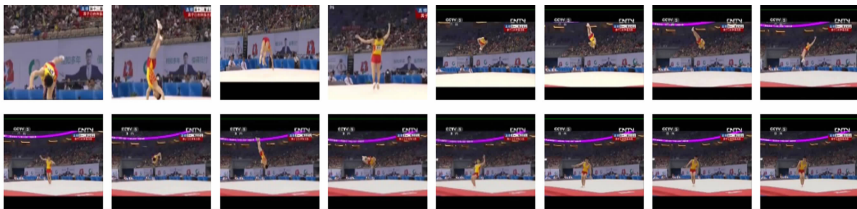


图 7 自由体操视频自动描述实例

Fig. 7 Example of floor exercise video description

3 结束语

提出一种长短时记忆网络的自由体操视频自动描述方法,但该研究仍需进一步优化.在数据层面,应进一步丰富自由体操自动描述数据集的建设,均匀不同分解动作的数据量及描述量.在视频特征的提取层面,可以使用目前较为流行的三维卷积操作对自由体操视频进行时域和空域的特征提取^[30].此外,可以借助知识迁移更有效地识别、定位体育视频中运动员的动作,并将自由体操视频自动描述方法运用于其他体育项目中,扩大研究成果的使用范围.

参考文献:

- [1] 张世杰. 基于深度学习的体育视频关键姿态检测[D]. 北京:北京工业大学,2017.
- [2] 茅洁,谷倩. 深度学习优化蚁群算法的羽毛球项目技战术决策研究[J]. 运动,2016(18):5-6. DOI:10.3969/j.issn.1674-151x.2016.18.003.
- [3] 王皓蜀. 基于 Mean Shift 算法的网球运动视频目标跟踪研究[J]. 现代电子技术,2017,40(13):73-76. DOI:10.16652/j.issn.1004-373x.2017.13.0190.
- [4] 杨彬,王同喜. 基于 DSP-FPGA 的嵌入式篮球运动视频目标跟踪算法实现[J]. 湘潭大学学报(自然科学版),2018,40(6):104-108.
- [5] 付裕. 神经网络在足球比赛中的胜负预测[J]. 科技风,2018(23):215. DOI:10.19392/j.cnki.1671-7341.2018.23214.
- [6] 王英英. 基于体育训练的运动视频分析系统设计与开发[J]. 自动化技术与应用,2019(8):35.
- [7] 马月洁,冯爽,王永滨. 基于深度学习的足球球员跟踪算法研究[J]. 中国传媒大学学报(自然科学版),2018,25(3):60-64. DOI:10.3969/j.issn.1673-1328.2017.17.046.
- [8] 杨斌. 足球比赛视频中的目标检测与跟踪算法研究[J]. 计算机测量与控制,2017,25(9):266-268. DOI:10.16526/j.cnki.11-4762/tp.2017.09.068.
- [9] 朱施成,李旭. 2017—2020 女子自由体操评分规则变化趋势及其影响因素探析[J]. 湖北师范大学学报(哲学社会科学版),2018,38(1):75-80. DOI:10.3969/j.issn.2096-3130.2018.01.017.
- [10] KOJIMA A, TAMURA T, FUKUNAGA K. Natural language description of human activities from video images based on concept hierarchy of actions[J]. International Journal of Computer Vision,2002,50(2):171-184. DOI:10.1023/A:1020346032608.
- [11] GUADARRAMA S, KRISHNAMOORTHY N, MALKARNENKAR G, *et al.* Youtube2text: Recognizing and describing arbitrary activities using semantic hierarchies and zero-shot recognition[C]//Proceedings of the IEEE International Conference on Computer Vision. Sydney:IEEE Press,2013:2712-2719. DOI:10.1109/ICCV.2013.337.
- [12] ROHRBACH M, QIU Wei, TITOV I, *et al.* Translating video content to natural language descriptions[C]//Proceedings of the IEEE International Conference on Computer Vision. Sydney:IEEE Press,2013:433-440. DOI:10.1109/ICCV.2013.61.
- [13] XU Ran, XIONG Caiming, CHEN Wei, *et al.* Jointly modeling deep video and compositional text to bridge vision and language in a unified framework[C]//Twenty-Ninth AAAI Conference on Artificial Intelligence. Austin:AAAI Publications,2015:2346-2352.
- [14] PAN Yingwei, YAO Ting, LI Houqiang, *et al.* Video captioning with transferred semantic attributes[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu:IEEE Press,2017:6504-6512. DOI:10.1109/CVPR.2017.111.
- [15] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]//Proceedings of the 25th International Conference on Neural Information Processing Systems. New York:Curran Associates,2012:1097-1105. DOI:10.1145/3065386.
- [16] RUSSAKOVSKY O, JIA Deng, SU Hao, *et al.* Imagenet large scale visual recognition challenge[J]. International Journal of Computer Vision,2015,115(3):211-252. DOI:10.1007/s11263-015-0816-y.
- [17] VENUGOPALAN S, XU Huijuan, DONAHUE J, *et al.* Translating videos to natural language using deep recurrent neural networks[EB/OL]. [2019-10-15]. <https://arxiv.org/abs/1412.4729>.
- [18] SHETTY R, LAAKSONEN J. Frame-and segment-level features and candidate pool evaluation for video caption generation[C]//Proceedings of the 24th ACM International Conference on Multimedia. New York:[s. n.],2016:

- 1073-1076. DOI:10.1145/2964284.2984062.
- [19] JIN Qin, CHEN Jia, CHEN Shizhe, *et al.* Describing videos using multi-modal fusion[C]//Proceedings of the 24th ACM International Conference on Multimedia. New York: [s. n.], 2016: 1087-1091. DOI:10.1145/2964284.2984065.
- [20] SUNDERMEYER M, NEY H, SCHLÜTER R. From feedforward to recurrent LSTM neural networks for language modeling[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2015, 23(3): 517-529. DOI:10.1109/TASLP.2015.2400218.
- [21] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780. DOI:10.1162/neco.1997.9.8.1735.
- [22] GERS F A, SCHMIDHUBER J, CUMMINS F J N C. Learning to forget: Continual prediction with LSTM[J]. 1999, 12(10): 2451-2471. DOI:10.1162/089976600300015015.
- [23] VENUGOPALAN S, ROHRBACH M, DONAHUE J, *et al.* Sequence to sequence-video to text[C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE Press, 2015: 4534-4542. DOI:10.1109/ICCV.2015.515.
- [24] SHARIF R A, AZIZPOUR H, SULLIVAN J, *et al.* CNN features off-the-shelf: An astounding baseline for recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Columbus: IEEE Press, 2014: 806-813. DOI:10.1109/CVPRW.2014.131.
- [25] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. [2019-10-15]. <https://arxiv.org/abs/1409.1556>.
- [26] ZHAO Haitao, SUN Shaoyuan, JIN Bo. Sequential fault diagnosis based on lstm neural network[J]. IEEE Access, 2018(6): 12929-12939. DOI:10.1109/ACCESS.2018.2794765.
- [27] YAO Li, TORABI A, CHO K, *et al.* Describing videos by exploiting temporal structure[C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE Press, 2015: 4507-4515. DOI:10.1109/ICCV.2015.512.
- [28] PAPINENI K, ROUKOS S, WARD T, *et al.* BLEU: A method for automatic evaluation of machine translation [C]//Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Stroudsburg: [s. n.], 2002: 311-318.
- [29] CHEN D L, DOLAN W B. Collecting highly parallel data for paraphrase evaluation[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: [s. n.], 2011: 190-200.
- [30] 张颖, 袁和金. 基于 3D 卷积神经网络的人体行为识别方法[J]. 软件导刊, 2017, 16(11): 9-11. DOI:10.11907/rjdk.172515.

(责任编辑: 钱筠 英文审校: 吴逢铁)