第 58 卷 第 20 期/2021 年 10 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

改进的编码-解码时序动作检测算法

王玥，苏寒松，刘高华*

天津大学电气自动化与信息工程学院，天津 300072

摘要 时序动作检测作为视频理解中的一项基本任务，被广泛应用于人机交互、视频监控、智能安防等领域。基于卷积神经网络，提出了一种改进的编码-解码时序动作检测算法。改进后的算法分两阶段进行：首先，替换特征提取网络，用残差结构网络提取视频帧的深度特征；之后，构建编码-解码时序卷积网络。采用联接的方式进行特征融合，改进上采样的形式，并运用新的激活函数 LReLU 进行训练，提高网络的检测精度。实验结果表明，所提算法在时序动作检测数据集 MERL Shopping 和 GTEA 上取得了优良的效果。

关键词 光计算；图像处理；动作检测；时序卷积神经网络；深度学习

中图分类号 TP391 **文献标志码** A **doi**：10.3788/LOP202158.2020001

Improved Encoder-Decoder Temporal Action Detection Algorithm

Wang Yue, Su Hansong, Liu Gaohua*

School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, *China*

Abstract Temporal action detection is a fundamental task in video understanding that is commonly used in the fields of human-computer interaction, video surveillance, intelligent security, and other fields. An improved encoder-decoder temporal action detection algorithm based on the convolutional neural network is proposed. The improved algorithm is applied in two stages: first, the feature extraction network is replaced and the residual structure network is used to extract the deep features of the video frame; and second, the encoder-decoder temporal convolutional network is constructed. The feature fusion is conducted via contact, and the method of upsampling is improved. To improve the detection accuracy of the network, the proposed algorithm employs the appropriate activation function LReLU for training. The experimental results show that the accuracy of the proposed algorithm on the temporal action detection datasets MERL Shopping and GTEA has improved.

Key words optics in computing; image processing; action detection; temporal convolutional neural network; deep learning

OCIS codes 200.4260; 100.4996; 100.3008; 150.1135

1 引 言

动作视频分析包括传统动作识别和时序动作检测。其中，传统动作识别从经裁剪的视频片段中识别动作类型；而时序动作检测从未裁剪的视频中对动作实例和其他信息进行分割，在传统动作识别的基础上，定位目标动作的开始帧和结束帧。在实际用于分析的数据中，长视频居多，并且包含多种动作实例和背景信息。因此，时序动作检测具有更高的应用价值，被广泛应用在人机交互、视频检索、视频监督和动作分析等领域[1]，具有重要意义。

时序动作检测主要分为两个阶段，确立动作阶段和分类阶段。其中，确立动作阶段从长视频中提取包含动作的片段，分类阶段识别动作的类型。现有的时序动作检测方法主要为多阶段检测，主要包括特征提取、动作定位和动作分类。早期的研究多

收稿日期：2020-09-24；**修回日期**：2020-11-05；**录用日期**：2020-12-08

通信作者：*suppig@126.com

基于时间或空间特征，使用滑动窗口进行检测。Gaidon 等[2]通过动作单元演示建模对动作序列进行配对，并应用于视频片段中，但该方法只适用于特定的动作，具有局限性。Singh 等[3]利用卷积神经网络提取每一帧的特征，将特征输入到长短记忆模型中，并应用非极大值抑制进行输出。基于穷举滑窗的方法计算效率低，并在一定程度上限制了时序动作的边界，模型不能精确定位动作的开始和结束时间。随着深度学习的发展，基于深度学习的时序动作检测方法展现出了优异的性能。Xiong 等[4]提出了一种递归神经网络，该网络利用视频帧图像的特征作为输入，以端到端的形式预测动作的起点和终点。但是该方法不支持与卷积神经网络进行联合训练，且耗时长、速度慢。Gao 等[5]提出了一种时间单元回归网络，在未剪辑的长视频中生成时间候选运动，并对它们进行分类。Xu 等[6]提出了 R-C3D 网络，网络的输入为视频帧图像，使用 C3D 网络提取特征图，将特征输入到动作时序提议网络，得到一系列粗略的动作序列；再经过精细分类网络调整，消除高度重叠的提议和评估得分低的提议，使得提议数量更少、质量更高，最终得到动作序列的开始和结束时间，及所包含的动作类别。总体而言，时序动作检测在分类阶段和传统动作识别类似，而动作识别的准确率已经达到了很高的水平。因此，目前时序动作检测的重点主要为定位动作的边界和提高检测的精度。

基于以上问题，本文对编码-解码时序动作检测算法进行了改进。所提算法能够处理不同时长的视频，充分利用视频特征定位动作边界，防止信息遗漏，并且时序卷积结构占用内存小，处理速度快，在硬件需求和处理效率上具有一定的优势。首先对视频帧图像进行预处理，使用深度残差神经网络[7]提取丰富的深层特征，对视频进行更贴切的表达；之后将特征输入到编码-解码时序卷积网络中[8]，针对解码过程中的信息丢失问题，在解码模块中应用联接的特征融合方式，令解码层融合复用编码层信息，实现多尺度特征的结合，修正中间层特征[9]；加入改进后的插值上采样，用间隔插值代替单一重复的特征值，并使用新的 LReLU 激活函数[10]训练模型，进一步提升检测精度；最后经过全连接层，使用 One-Hot 编码输出每个视频帧的动作标签，实现动作类型的识别，并准确定位动作的边界。

2 编码-解码时序卷积网络

改进的编码-解码时序卷积网络由编码模块和解码模块构成，具有以下特点：卷积计算是分层进行的，即每一帧的特征同时被计算，而不是逐帧进行；卷积是跨时域进行的；可接收任意长度的输入序列，并将其映射为等长度的输出序列。

2.1 编码模块

在编码模块中，用来自同一视频的帧图像特征作为输入。令第 t 帧图片的特征 $X_t \in \mathbf{R}^{F_0}$，其中 F_0 是特征的维度，t 为帧数（$1 \leqslant t \leqslant T$），$T$ 为视频的时长。所提网络要求输入特征维度相同，对于时长不同的视频，对缺少的部分进行补零操作，以适应网络对输入的要求。

编码模块包括 4 层，每一层可定义为 $E^{(i)} \in \mathbf{R}^{F_i \times T_i}$，其中 F_i 是第 i 层的卷积核数量，T_i 是相应的时间步长。每层由卷积层、激活函数和跨时间的最大池化层组成。每层的卷积滤波器由权重 W 和偏置项 b 确定，由来自 $E^{(i-1)}$ 的信号计算 $E^{(i)}$ 的信号：

$$E^{(i)} = \text{maxpooling}\{f[(W * E^{(i-1)} + b)]\}, \quad (1)$$

式中：$*$ 是卷积符号。对于卷积操作，采用一维时序卷积。

2.2 解码模块

解码模块 $D^{(i)}$ 与编码模块类似，不同之处在于，使用上采样层代替最大池化层，并且顺序为上采样层、卷积层和激活函数。上采样层将输入数据中的每一个原始数值重复两次，即经过上采样层后，数值的维度扩大为原来的 2 倍。与编码模块相比，解码模块的索引顺序相反，因此编码模块第一层的滤波器个数与解码模块最后一层的滤波器个数相同。解码模块中的卷积层能够利用时序信息和中间层的激活值对动作作出预测，捕捉动作之间的成对转换，从而确定动作的边界，显著提高模型性能。

最后，将处理后的特征经过全连接层，得到网络预测的输出 $\boldsymbol{Y}_t = \text{softmax}[\boldsymbol{U}D_t^{(1)} + \boldsymbol{c}]$，其中 $\boldsymbol{U}$ 为权重矩阵，$\boldsymbol{c}$ 为常量矩阵。对于第 t 帧图像，向量 $\boldsymbol{Y}_t \in [0,1]^C$ 是经过矢量化后的 C 维向量，其中的一个索引为 1，其余索引为 0，对应一个动作类别，与神经网络训练出的 C 个概率值对应。概率值最大的输出即对应该动作的标签，并生成包括动作真实值和预测值的预测结果文件。

3 改进的算法结构与原理

改进算法主要包括 3 个模块：特征提取网络、编

码模块和解码模块。改进后的编码-解码时序卷积网络[8]结构框架如图 1 所示。

1）在特征提取网络中，先用双线性插值法[11]对输入图像进行缩放，再用具有 50 层卷积结构的残差网络提取图片的深层特征。

2）编码模块主要由卷积层和池化层组成，编码模块从输入的帧序列图片中提取含有空间信息的深度特征。

3）解码模块主要由上采样层和卷积层组成，所提算法将单一重复的上采样改为插值运算，采用联接的方式融合编码模块的输出特征，最后通过全连接层对动作种类作出预测。

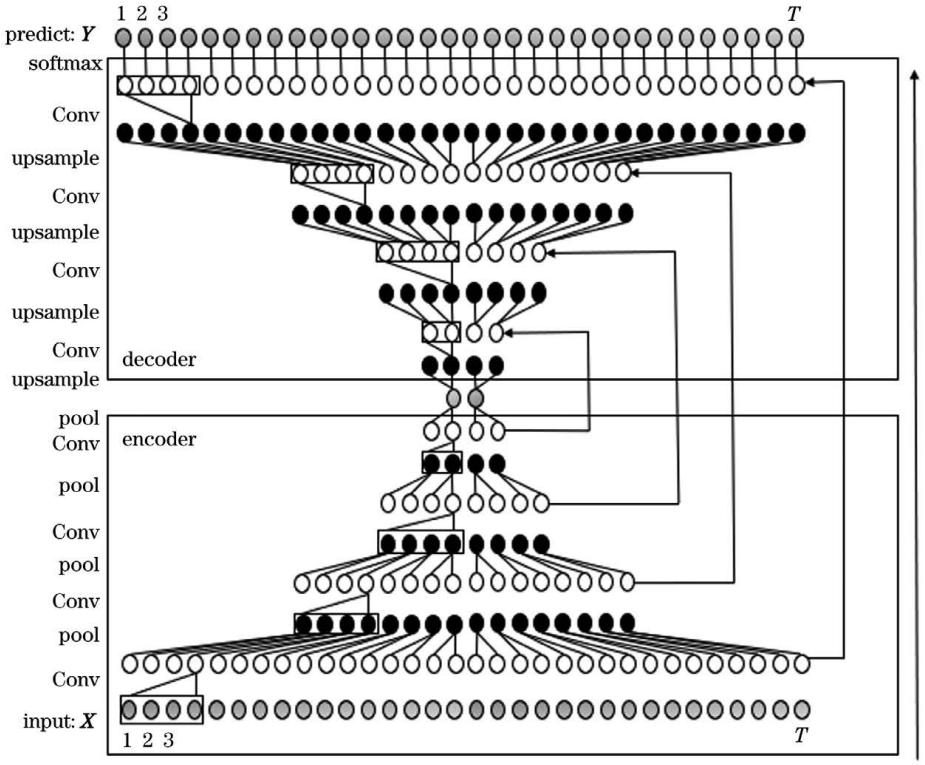

图 1　改进的编码-解码时序卷积网络结构

Fig. 1　Structure of the improved encoder-decoder temporal convolutional neural network

3.1　特征提取网络

近年来，用于提取特征的卷积神经网络发展迅速，从最初的 LeNet[12]、AlexNet[13]、Vgg-Net[14]等浅层网络不断向更深层次发展。然而，网络层数的增加并不意味着性能的持续提高，随着层数的增加，准确率反而会下降。并且，深层次网络在训练时往往会出现梯度消失或梯度爆炸的问题，由于信息丢失，网络的训练效果变差。针对以上现象，He 等[7]提出了“残差模块结构”，该结构使用跳跃连接直接将输入信息传到输出端，保持信息完整，避免信息丢失。这样一来，网络在学习过程中只需学习输入和输出的差异部分。残差模块的结构如图 2 所示。

利用所提模型时，对于给定的视频帧序列，为满足特征提取网络对输入的要求，首先要对帧序列图像进行预处理，在保证图像内容与原图像一致的基础上，修改图像的尺寸。采用双线性插值的方法，把原图像的尺寸修改为 224×224。

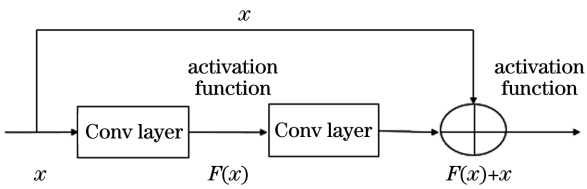

图 2　残差模块结构

Fig. 2　Structure of the residual module

特征提取网络采用 50 层的卷积神经网络，以残差模块的方式相连。为简化训练过程，充分提取图像特征，使用在 ImageNet 数据集上预训练好的模型。具体而言，该网络按照参数不同共分为 5 个模块，参数如表 1 所示。

3.2　激活函数

对神经网络进行特征提取的过程中，激活函数起着至关重要的作用。如果不使用激活函数，则整个网络的输入和输出完全是线性关系，即使增加网络的层数，网络也不具备很好的拟合能力。

表 1　特征提取网络的参数

Table 1　Parameters of feature extraction network

Block	Kernel size	Number of channels
Conv1	7×7	64
Conv2_x	$\begin{bmatrix}1\times1\\3\times3\\1\times1\end{bmatrix}\times3$	$\begin{bmatrix}64\\64\\256\end{bmatrix}\times3$
Conv3_x	$\begin{bmatrix}1\times1\\3\times3\\1\times1\end{bmatrix}\times4$	$\begin{bmatrix}128\\128\\512\end{bmatrix}\times4$
Conv4_x	$\begin{bmatrix}1\times1\\3\times3\\1\times1\end{bmatrix}\times6$	$\begin{bmatrix}256\\256\\1024\end{bmatrix}\times6$
Conv5_x	$\begin{bmatrix}1\times1\\3\times3\\1\times1\end{bmatrix}\times3$	$\begin{bmatrix}512\\512\\2048\end{bmatrix}\times3$

通常运用 ReLU 函数[15]来提高网络的非线性能力。ReLU 函数的正半轴输出值与输入值相同，负半轴输出值为 0。但 ReLU 作为激活函数也存在许多问题。因在反向传播的过程中需要进行求导操作，此时如果负半轴信号的梯度为 0，则神经元不会被更新，不再进行学习。为解决上述问题，所提网络引入 LReLU 函数，此函数在负半轴有一个较小但不为零的常数泄漏值，用该常数作为负半轴输入的导数值。引入该值后，导数总不为零，避免出现神经元学习停滞的情况，能够修正数据分布，提升网络的抗干扰能力。

经改进后，LReLU 函数的数学表达式为

$$y=\max(0,x)+a*\min(0,x),\tag{2}$$

式中：a 为泄漏值，设为 0.01。

3.3　特征融合

所提算法采用如图 3 所示的联接的特征融合方式。与将对应点的特征相加(add)不同，所提算法将编码模块激活函数输出的特征整合到解码模块，激活函数的输出特征之后，形成更厚的特征，再将特征向后传递。

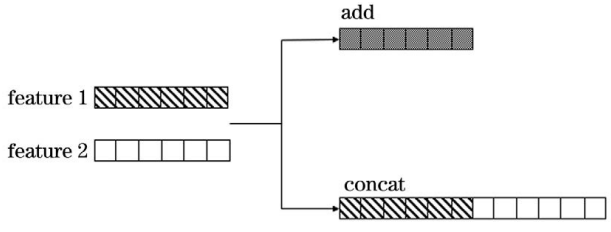

图 3　不同的特征融合方式

Fig. 3　Different feature fusion methods

在所提算法中，解码模块可以融合编码模块的输出，既有来自对应解码模块的同尺度特征，又有来自解码模块上采样输出的大尺度特征，实现了多尺度特征的融合。并且，由于编码模块和解码模块的结构对称，解码模块的特征可以得到补充。以这种对称的形式进行融合，能够避免随着卷积网络的深度增加，出现特征消失、特征不全面的情况，使网络能更加全面地学习图片特征。

3.4　改进的上采样

原算法中的上采样将原始特征点重复两次，扩大特征的维度，如图 4(a)所示。所提算法在此基础上提出一种改进的上采样形式，如图 4(b)所示。在改进后的上采样中，仍将特征数量扩大两倍，但特征的取值有所不同。其中一个特征值保留原始值，另一个特征值为两相邻原始值的均值，经过上采样后，原始值和均值交替出现。通过这种方式，上采样不再是对原始特征值的简单重复，既能恢复特征维度，也能保留更多的特征信息。实验结果表明，所提改进上采样在性能上优于原始单一值上采样，有助于提升检测精度。

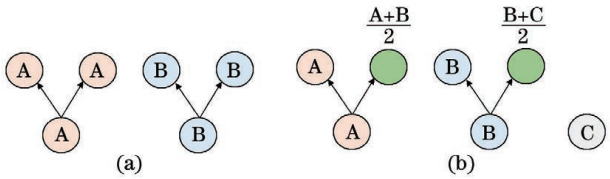

图 4　传统上采样和改进上采样的示意图。(a)传统上采样；(b)改进上采样

Fig. 4　Schematic diagram of traditional upsampling and improved upsampling. (a) Traditional upsampling; (b)improved upsampling

4　实　　验

4.1　训练数据库

在神经网络的训练过程中，数据库对训练结果起着至关重要的作用。对于时序动作检测，需要选择在同一视频中包含不同分动作的数据集，才能更好地完成这项任务。所提算法选择 MERL Shopping[16]和 Georgia Tech Egocentric Activities (GTEA)[17]数据集进行训练和测试。

MERL Shopping 数据集在 2016 年被推出，该数据集包括 106 个 2 分钟时长的视频，在实验室中模拟了商店购物的场景，均为俯视视角，由固定在上方的高清摄像机拍摄。视频中标记了 5 个不同种类动作的开始帧和结束帧，分别为伸手到货架、从货架

上收回手、手放在货架上、看商品、看货架。每个视频中包含多类动作，每种动作的持续时间从 0.5 s 到 1 min，能够使网络充分学习到不同长度、不同类别的动作。

GTEA 数据集包含 7 种类型的日常活动，每种活动由 4 名测试者进行，由安装在被拍摄对象的帽子上的高清相机拍摄，共有 28 个视频，其中包含 11 种动作，可用于训练动作分割和动作检测网络。

4.2　实验结果及评价指标

所提算法使用 Nvidia GTX 1080Ti 的 GPU，利用 Tensorflow[18] 深度学习框架，计算机操作系统为 Ubuntu 16.04 LTS，采用 Python 语言进行编程。

MERL Shopping 数据集和 GTEA 数据集的检测实例分别如图 5、6 所示，在一个视频中分别对 5 种和 7 种动作进行了标注，在图片中用不同的颜色区分。图片分为上下两个部分，上方为一个视频中各个动作在时间范围内的真实分布，下方为网络对测试数据做出的判断结果。准确率为能够被正确识别所包含动作类型的帧数占总帧数的比例。若图片上方的颜色与下方的颜色相同，即为能够被正确识别。可以看出：网络能够对视频中大部分动作的类别进行正确的识别，并对动作边界做出划分；对容易发生混淆的动作，如“伸手到货架”和“从货架上拿回手”的边界也能做出较好的判断。与原算法相比，所提算法在 MERL Shopping 数据集上的准确率提高了 2.3 个百分点，在 GTEA 数据集上的准确率提高了 4.2 个百分点。

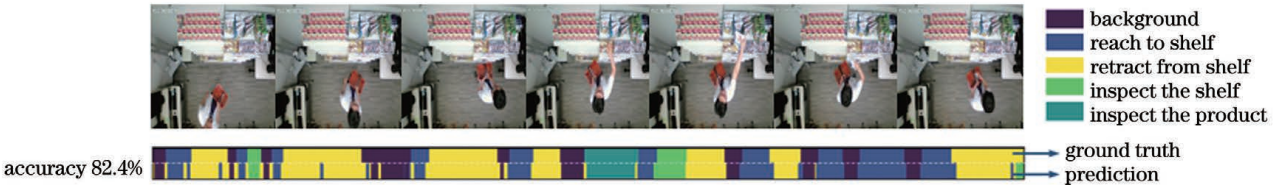

图 5　MERL Shopping 数据集检测实例

Fig. 5　Detection example of MERL Shopping dataset

图 6　GTEA 数据集检测实例

Fig. 6　Detection example of GTEA dataset

表 2 是在 MERL Shopping 数据集中，网络对各个分动作进行正确识别的比例。可以看出：对于“手放在货架上”“看商品”“看货架”等较为明确的动作，网络进行识别的准确率最高；对于两个容易混淆的动作，如“伸手到货架”和“从货架上收回手”，网络对动作的边界判断可能会出现差异，但也达到了较高的准确率。

表 2　各个动作的识别准确率

Table 2　Recognition accuracy rate of each action

Dataset	Action	Accuracy /%
MERL Shopping	Reach to shelf	77.8
	Retract from shelf	79.3
	Hand in shelf	81.6
	Inspect the product	80.4
	Inspect the shelf	81.2

表 3 对比了在交并比(IoU)为 0.5 时，不同模块对算法精确率(mAP)的影响，分别在 MERL Shopping 和 GTEA 数据集上进行了训练和测试。采用两种不同的特征 VggNet16 和 ResNet50 提取网络，并对比分析了它们在已有的编码-解码时序卷积网络和经过改进的网络上的实验结果。实验结果表明，新的特征提取网络 ResNet50 和改进后的时序卷积网络结合后，测试的效果最好。在 IoU 为 0.5 时，在 MERL Shopping 数据集上的最高精度 mAP 为 29.3%，在 GTEA 数据集上的最高精度为 mAP 为 30.2%。

F1 score 是精确率和召回率的调和平均数，是二分类问题中的常用指标。针对时序动作中的多分

表 3 不同模块对算法的影响

Table 3 Effectiveness of various module on the algorithm

Dataset	VggNet16	ResNet50	ED-TCN	Improved ED-TCN	mAP /%
MERL Shopping	√		√		24.3
MERL Shopping		√	√		25.6
MERL Shopping		√		√	29.3
GTEA	√		√		25.8
GTEA		√	√		27.2
GTEA		√		√	30.2

类问题，Lea 等[8]基于 F1 score，提出了适用于多分类任务的分段 F1 分数(Seg-F1@k)，$k=10\times P_{\mathrm{IoU}}$。使用相同的特征提取网络，对改进后的时序卷积神经网络和原有时序卷积神经网络在不同数据集上的 Seg-F1 数值进行对比，结果如表 4 所示。可以看出，改进后的时序卷积神经网络的 Seg-F1 有了不同程度的提高，能够对动作进行更精确的检测与识别。

表 4 不同算法在不同数据集上的 Seg-F1

Table 4 Seg-F1 of different algorithms on different datasets

Dataset	ED-TCN	Improved ED-TCN	Seg-F1@10	Seg-F1@25	Seg-F1@50
MERL Shopping	√		86.7	85.1	72.9
MERL Shopping		√	89.2	87.4	74.8
GTEA	√		72.2	69.3	56.0
GTEA		√	76.8	71.9	58.5

在 MERL Shopping 数据集上，对比改进后的算法和其他算法的检测效果，如表 5 所示，指标包括准确率(accuracy)、mAP 和 Seg-F1。所提改进算法在 3 个检测指标上的表现相较于现有算法均有提高，但相较于 MSN Det 算法在检测精度上还存在一定差距。这是因为 MSN Det 算法使用稀疏动作帧进行预测，这种特性有利于获得较高的检测精度，但也存在过度细分的缺陷；同时，由于破坏了动作的完整性，整体准确率较低，不利于界定动作边界。所提算法能在保证准确率的基础上提升精度，性能更加优越。

表 5 不同算法在 MERL Shopping 数据集上的结果对比

Table 5 Comparison of results of different algorithms on the MERL Shopping dataset

Algorithm	Accuracy /%	mAP /%	Seg-F1@10	Seg-F1@25	Seg-F1@50
MSN Det	64.6	29.5	46.4	42.6	25.6
MSN Seg	76.3	24.2	80.0	78.3	65.4
Dilated TCN	76.4	26.3	79.9	78.0	67.5
ED-TCN	79.0	25.5	86.7	85.1	72.9
Improved ED-TCN	82.4	29.3	89.2	87.4	74.8

5 结 论

针对未分割长视频中的时序动作检测任务，基于时序卷积网络，提出了一种改进的编码-解码时序动作检测算法。采用残差神经网络提取帧图像的特征；同时构建编码-解码时序卷积网络，运用联接的方式融合特征，将单一上采样改为插值上采样；并选用新的激活函数训练网络，增强模型的抗干扰性。经实验验证，所提算法在时序动作检测数据库中取得了良好的效果。

参 考 文 献

[1] Wu Y C, Yin J Q, Wang L, et al. Temporal action detection based on action temporal semantic continuity[J]. IEEE Access, 2018, 6: 31677-31684.

[2] Gaidon A, Harchaoui Z, Schmid C. Actom sequence models for efficient action detection[C]//2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2011, Colorado Springs, CO, USA. New York: IEEE Press, 2011: 3201-3208.

[3] Singh B, Marks T K, Jones M, et al. A multi-stream bi-directional recurrent neural network for fine-grained action detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1961-1970.

[4] Xiong Y J, Zhao Y, Wang L M, et al. A pursuit of temporal accuracy in general activity detection[EB/OL]. (2017-03-08)[2020-09-23]. https://arxiv.org/abs/1703.02716.

[5] Gao J Y, Yang Z H, Sun C, et al. TURN TAP: temporal unit regression network for temporal action proposals[C]//2017 IEEE International Conference on Computer Vision (ICCV), October22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 3648-3656.

[6] Xu H J, Das A, Saenko K. R-C3D: region convolutional 3D network for temporal activity detection[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 5794-5803.

[7] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.

[8] Lea C, Flynn M D, Vidal R, et al. Temporal convolutional networks for action segmentation and detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1003-1012.

[9] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W M, et al. Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.

[10] Xu B, Wang N Y, Chen T Q, et al. Empirical evaluation of rectified activations in convolutional network[EB/OL]. (2015-05-05)[2020-09-23]. https://arxiv.org/abs/1505.00853.

[11] Wang J J, Jian M W, Liu X Y, et al. Video saliency detection based on 3D full ConvLSTM neural network[J]. Computer Science, 2020, 47(8): 195-201.
王教金, 蹇木伟, 刘翔宇, 等. 基于 3D 全时序卷积神经网络的视频显著性检测[J]. 计算机科学, 2020, 47(8): 195-201.

[12] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[13] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[14] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04)[2020-09-23]. https://arxiv.org/abs/1409.1556.

[15] Wang M, Su H S, Liu G H, et al. Classroom face detection algorithm based on convolutional neural network[J]. Laser & Optoelectronics Progress, 2019, 56(21): 211501.
王萌, 苏寒松, 刘高华, 等. 基于卷积神经网络的教室人脸检测算法[J]. 激光与光电子学进展, 2019, 56(21): 211501.

[16] Singh B, Marks T K, Jones M, et al. A multi-stream bi-directional recurrent neural network for fine-grained action detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1961-1970.

[17] Fathi A, Ren X F, Rehg J M. Learning to recognize objects in egocentric activities[C]//2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2011, Colorado Springs, CO, USA. New York: IEEE Press, 2011: 3281-3288.

[18] Liu F, Liu P Y, Li B, et al. Deep learning model design of video target tracking based on TensorFlow platform[J]. Laser & Optoelectronics Progress, 2017, 54(9): 091501.
刘帆, 刘鹏远, 李兵, 等. TensorFlow 平台下的视频目标跟踪深度学习模型设计[J]. 激光与光电子学进展, 2017, 54(9): 091501.