文章编号：2095-2163(2021)05-0053-07 中图分类号：TP391.41 文献标志码：A

基于形变卷积神经网络的行为识别

李君君，张彬彬，江朝晖
（合肥工业大学 计算机与信息学院，合肥 230601）

摘 要：人类行为识别作为视频分类中的重要问题，成为计算机视觉中的热门话题。由于卷积神经网络（CNN）的几何结构固定统一，这将会使得其几何变形建模受限，使得行为识别网络难以鲁棒性的识别行为类别。本文提出了一种融入可形变卷积的行为识别网络模型。首先，引入可形变卷积，构建了一种可协同学习空间外观和时间运动线索的模块，该模块分别学习视频数据 3 个正交视图特征进行融合；其次，在 ResNet 网络的基础上，用该模块将其网络中部分关键性卷积模块进行替换，产生一种新颖的改进版本的 3D-ResNet 网络，用于视频数据集的训练和测试；最后，在 UCF101 和 HMDB51 数据集训练和测试，得到识别精度优于现有的大多数先进方法。

关键词：行为识别；卷积神经网络；可形变卷积；ResNet

Action recognition based on deformation convolutional neural network

LI Junjun, ZHANG Binbin, JIANG Chaohui
（School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230601, China）

【Abstract】 As an important problem in video classification, Human action recognition is becoming a hot topic in computer vision. Due to the fixed geometric structure, convolutional neural network (CNN) is limited in its modeling of geometric deformation, which makes it difficult for the action recognition network to express its behavior robustly. In this paper, a action recognition network based on deformable convolution is proposed. Firstly, this paper introduces deformable convolution to construct a module that can learn spatial appearance and temporal motion cues cooperatively. The module learns the features of three orthogonal views of video data respectively and then fusion them. Secondly, on the basis of the ResNet network, some key convolution modules in the network are replaced with this module to produce a novel and improved 3D-ResNet network for the training and testing of video data - sets. We have trained and tested this method on UCF101 and HMDB51 datasets, and the results show that recognition accuracy of this method is more excellent than most of the existing advanced methods.

【Key words】 human action recognition; convolutional neural network; deformable convolutional; ResNet

0 引 言

行为识别技术是近年来计算机视觉研究领域被广泛关注的技术，受到国内外专家学者的广泛重视和深入研究，其相关技术在智慧监控、人机交互、视频序列理解、医疗卫生等领域发挥着越来越重要的作用。目的是通过研究人体在视频中的图像帧或图像序列的时空变化，利用计算机处理和分析视觉信息，自动识别出视频中的行为模式。由于人体行为类别多样，复杂多变的背景，视频视角的差异性等问题，网络模型难以鲁棒、准确对真实的视频行为动作进行辨别，因此行为识别亟待研究工作者深入地开展研究工作。

现有的深度学习模式对特征提取模型的训练多采用端到端的模式，使用训练好的神经网络模型参数去学习视频的显著特征，对行为进行分类识别。一些早前的相关研究工作主要专注于利用卷积神经网络（CNN）来学习视频帧连续序列中蕴含的行为的深度特征。主流的 CNN 网络模型包括双流结构的一系列的模型和 3DCNN 模型。然而，卷积神经网络通常有两个缺点：

（1）假设卷积计算的几何变换是固定的和已知的，一般是使用这些先验知识，来做数据的增强工作并且设计特性和算法，但是这种默认的规则，会导致算法不能对未知几何变换的新任务进行有效泛化，会导致任务建模的不正确或不恰当；

（2）相对更加复杂的变换来说，即使已经知道其固定的特征和算法，也难以用手工的方式进行设计[1]。

一般来说，对于卷积神经网络，卷积核具有固定几何会导致其对几何形变建模能力有限，标准卷积中的规则格点采样是网络难以适应几何变形和时间序列位移的根本原因。Dai 提出变形卷积，可以通过在传统卷积运算的基础上增加一个并行网络来预

作者简介：李君君（1995-），男，硕士研究生，主要研究方向：计算机视觉。

收稿日期：2021-01-29

测传统卷积采样点的偏移量,使每个采样点都有一定的偏移量,并学习自适应感受野,从而提高了对不同尺寸和形状物体的特征提取能力[1]。本文针对标准卷积建模能力有限的问题,在残差网络的基础上,提出了可变形卷积改进的残差网络,以提升网络识别的准确性。

本文在 CoST 模块的基础上,构建了新颖的 DSTC(Deformable Spatio-Temporal Convolution)模块。该模块可在视频数据的 3 个正交视图执行 2D 可形变卷积,可分别学习空间外观和时间运动线索,增强了卷积核对感受野的适应能力,以适应不同特征图感受野的形状、大小等几何形变;在残差网络模型结构的基础上,提出了一种新网络模型,该模型融合了 DSTC 可变形卷积模块,并且能够将端对端训练的网络用于行为分类;在 UCF101 和 HMDB51 数据集上的实验结果,证明了本文方法在公开数据集测试中具有显著的行为识别性能,优于当前先进的方法,并且相对于 3D 卷积,大大减少了参数量,并提升了识别的精度。

1 相关工作

早期的行为识别工作主要使用一些传统方法,手工制作的行为表征被很好地用于视频行为识别。许多二维的图像特征描述符被推广到三维时间域,例如时空兴趣点(Space-Time Interest Points,STIP),SIFT-3D,时空 SIFT(Space-Time SIFT)和 3D 梯度直方图(3D Histogram of Gradient)。最成功的手工特征表征是稠密轨迹流(dense trajectories)和其改善版本,其通过光流引导的轨迹提取局部特征,是传统方法中最为鲁棒,效果最好的[2-3]。

在受到深度学习取得巨大成功的鼓舞下,特别是 CNN 模型在图像理解任务的成功,涌现了许多开发行为分类的深度学习方法的尝试。Karpathy 等人提出在每帧上独立应用 2D CNN 模型,并探讨了几种融合时序信息的策略,由于未考虑帧之间运动变化,性能不如基于手工特征的算法[4];Donahue 等人利用 LSTM,通过聚合 2D CNN 特征建模时序信息,高级别的 2D CNN 特征被用来学习时序关系[5]。现在通常利用两种方法来提升时序建模能力,第一个是基于 Simonyan 和 Zisserman 提出的双流体系结构,该体系包括一个空间 2D CNN 和时序 2D CNN,可分别建模帧的静止特征和帧间光流运动信息,并将其输出分类分数融合为最终预测,许多后续工作是对这个框架的拓展,探索了 2 个流特征的融合策略[6];另一个典型方法是基于 3D CNN 和其(2+1)D 变体,Tran 等人设计了一个 11 层 C3D 模型,以联合学习 Sports-1M 数据集上的时空特征,然而巨大的计算成本和 C3D 的密集参数使得深度模型难以训练[7]。Qiu 等人提出了伪 3D(P3D)模型,将 3×3×3 的 3D 卷积分解成 1×3×3 的 2D 卷积和 3×1×1 的 1D 卷积[8];Tran 等人在残差网络上分解 3D 卷积为(2+1)D 卷积,取得了优于 3DCNN 的识别效果[9];Carreira 等人提出膨胀三维卷积(Inflating 3D ConvNets,I3D),通过扩充预先训练的 C2D 模型的参数进行初始化[10]。

CNN 模型在行为任务领域已取得了许多优秀的成果,但大多将常规卷积作为先验知识,没有考虑到卷积计算的本质缺陷——卷积网络对视频行为目标的几何变化是未知的,这会导致模型和数据容量的低效利用。近期一些相关的工作想要通过变形建模来解决问题,Worrall 等人通过移位、旋转和反射等变形的设计,在网络中添加几何不变量[11];另一种思路是通过图像空间中的半参数化或完全自由形式采样来学习重新组合数据。Jaderberg 等人通过空间变换网络(Spatial Transformers Network,STN)学习二维仿射变换[12];Rocco 等人利用深度几何匹配器(Deep Geometric Matchers)学习薄板样条变换[13];Dai 利用可变形卷积学习自由形式的转换[1]。受到这些研究工作的引导和启发,在模型参数几乎不增加的前提下,本文的模型能够充分利用时空信息,有效地提取特征图中的重要特征。实验结果表明,该网络具有良好的识别精度。

2 模型框架

本文对 ResNet-50 网络进行了改进,网络框架如图 1 所示。ResNet 网络是在 VGG19 网络发展而来,在其网络基础上进行改进,添加了残差单元,3D-ResNet-50网络与本文的网络对应替换的模块示意如图 2 所示,将卷积模块和一致性模块中的 3×3×3 的卷积层替换成 DSTC 模块,形成了可变形卷积模块和可变形一致性模块。

如图 2(a)所示,在网络层之间加入短路机制,可以有效地解决深层网络的退化问题;将残差卷积模块和一致性模块(图 2(a)和图 2(c))中的 3D 卷积层替换成了可变形卷积模块(DSTC),进而构造可变形残差卷积模块和可变形一致性模块(图 2(b)和图 2(d)),从而构建了改进版本的 ResNet-50 网络。首先,将输入的视频帧堆叠的图像序列裁剪成固定

大小，通过三维卷积和最大池运算对数据进行初始化；其次，将初始化后的特征图像依次发送到 4 个大的卷积模块中（Layer1－Layer4），每个大模块依次由 3，4，6 和 3 个可变形残差模块组成，每个可变形残差模块中包括对特征图进行卷积运算，批量归一化和激活函数运算操作；最后，将特征图输入到分类层中依次执行 3D 平均池化、全连接层和 Softmax 操作得到行为的标签，得到的行为识别结果。

图 2 是 3D－ResNet－50 网络与本文的网络对应替换的模块示意图，将卷积模块和一致性模块中的 3×3×3 的卷积层替换成 DSTC 模块，形成了可变形卷积模块和可变形一致性模块。

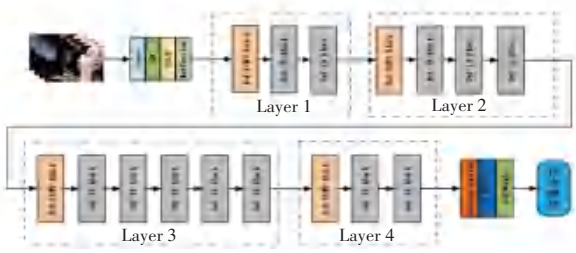
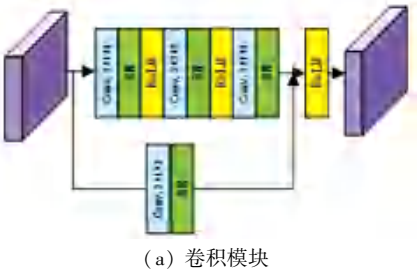


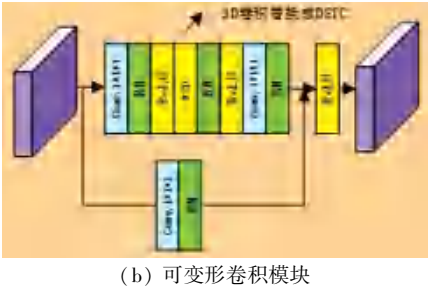
图 1　本文网络整体框架图

Fig. 1　The overall framework of the network in this paper



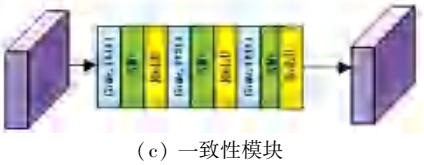
（a）卷积模块

（a）Convolution Block



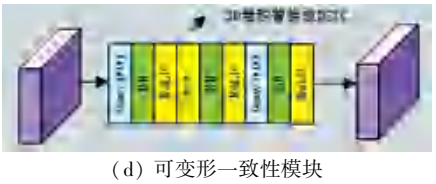
（b）可变形卷积模块

（b）Deformable Convolution Block



（c）一致性模块

（c）Identity Block



（d）可变形一致性模块

（d）Deformable Identity Block

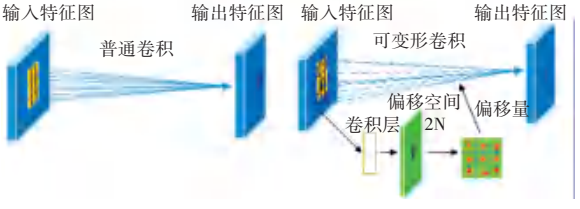
图 2　引入可变形卷积的模块

Fig. 2　Modules with deformable convolution

2.1　可形变卷积层

卷积网络对大尺寸多形变目标的建模存在固有的缺陷，因为卷积网络只对输入特征图的固定位置进行采样。例如，在同一层特征图中，所有特征点的感受野都是相同的，但不同的位置可能对应不同的尺度或变形对象，因此尺度或感受野大小的自适应学习是实现精确定位的必要条件。在模型中加入可变形卷积能够有效提升对目标形变的建模能力，使用一个平行卷积层学习 offset 偏移，在输入特征图上对应的任一卷积核的采样点位置上进行偏移，使得这些采样点更加集中在兴趣目标区域上，即增添一个偏移量在每个采样点对应位置，就可以打破常规卷积的规则网格的约束，在采样位置周边进行随意的采样。

普通卷积和可变形卷积的计算过程如图 3 所示。在普通卷积中，使用卷积核 w 对规则网格 $R(R=\{(-1,-1),(-1,0),\cdots,(0,1),(1,1)\})$ 中的采样点进行加权运算；在可变形卷积中，通过一个平行的卷积层，对输入特征图进行卷积，得到与输出特征图具有相同的分辨率的偏移量，输出通道数为 $3N$（N 为卷积核采样点个数），其中 $2N$ 为预测的 x，$y2$ 个维度上的偏移量；由于不同采样点对特征有不同的贡献，还要预测 N 个采样点的权重。到目前为止，已经有了输入特征图以及输入特征图上每个点对应的偏移量和权重，于是可以执可变形卷积运算。



（a）普通卷积示意图

（a）Schematic diagram of general convolution

（b）可形变卷积示意图

（b）Schematic diagram of deformable convolution

图 3　普通卷积和可变形卷积计算过程示意图

Fig. 3　Calculation process of general convolution and deformable convolution

在可变形卷积的操作中，延续了卷积运算的一般计算过程，只是在采样区域加入一个由能够通过

网络自动学习的参数 $\{\Delta p_n \mid n=1,\cdots,N\}$, $N=|R|$, 同时对每个采样点预测一个权重 Δm_n, 那么同样的位置 P_0 的值变为公式(1):

$$y(p_0)=\sum_{p_n\in R} w(p_n)\cdot x(p_0+p_n+\Delta p_n)\cdot \Delta m_n. \quad (1)$$

由于 Δp_n 通常是小数,因此需要通过双线性插值法计算 x 的值,公式(2)为:

$$x(p)=\sum_q G(q,p)\cdot x(q). \quad (2)$$

其中, p 代表位置,也就是公式中的 $p_0+p_n+\Delta p_n$, 列举了输入特征图 x 的空间位置,其中 $G(.,.)$ 表示双线性插值算法中的核函数,是二维的,可以被分为 2 个一维核,式(3):

$$G(q,p)=g(q_x,p_x)\cdot g(q_y,p_y). \quad (3)$$

因为绝大部分的参数都为 0,因此可以很快地计算出结果。

2.2 DSTC 模块

本文模块在 CoST 模块改进而来。$C3D_{3\times3\times3}$ 卷积操作利用 3×3 的三维卷积联合提取空间(沿 H 和 W) 和时间(沿 T) 特征。本文所提出的模块中,沿 $T\times H\times W$ 立体数据的 3 个视图 $H-W$、$T-H$ 和 $T-W$ 分别执行可变形 $2D_{3\times3}$ 卷积。值得注意的是,模块的三视图卷积计算的参数是共享的,这使得参数的数量与单视图二维卷积相同,这样可以大大降低参数的数量。随后,3 个生成的特征图依次加权求和,卷积计算的权值将在训练过程中以端到端的方式学习。

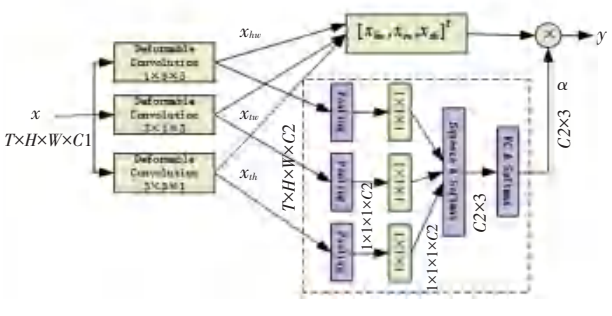


图 4 可变形时空卷积模块

Fig. 4 DeformableSpatial-Temporal Convolution Module

图 4 给出了 DSCT 模块的示意图,设 x 表示大小为 $T\times H\times W\times C1$ 的输入特征映射,其中 $C1$ 是输入通道的数目。来自不同视图的 3 组输出特征映射的计算方法是公式(4):

$$x_{hw}=x\otimes w_{1\times3\times3};\ x_{hw}=x\otimes w_{3\times1\times3};\ x_{hw}=x\otimes w_{3\times3\times1}. \quad (4)$$

其中, $\otimes$ 表示三维卷积, w 是 3 个视图之间共享的大小为 3×3 的卷积滤波器。为了将 w 应用于图像帧的不同视图,在不同的维度上插入一个尺寸为 1 的附加维度,由此产生的 w 的变体,即 $w_{1\times3\times3}$、$w_{3\times1\times3}$ 和 $w_{3\times3\times1}$ 分别学习 $H-W$、$T-W$ 和 $T-H$ 视图的特征,然后对 3 组特征映射加权求和,式(5):

$$y=[\alpha_{hw},\alpha_{tw},\alpha_{th}]\begin{bmatrix}x_{hw}\\ x_{tw}\\ x_{th}\end{bmatrix}. \quad (5)$$

其中, $\alpha=[\alpha_{hw},\alpha_{tw},\alpha_{th}]$ 的维度为 $C2\times3$; $C2$ 为输出通道数;3 表示 3 个视图。为了避免来自多个视图的响应的大小爆发式增长, α 沿每行用 *Softmax* 函数归一化, α 的系数由网络乘以 α 的特征图来学习得到,这种设计是受近来机器翻译的注意力机制的启发。在这种情况下,每个样本的系数取决于样本本身,可以用公式(6)表达:

$$[\alpha_{hw},\alpha_{tw},\alpha_{th}]=f([x_{hw},x_{tw},x_{th}]). \quad (6)$$

虚线内的计算块表示方程中的函数 f。对于每个视图,首先使用全局最大池化层将尺度为 $T\times H\times W\times C2$ 的特征映射沿着 T, H, W 3 个维度减少到 $1\times1\times1\times C2$; 然后,在池化特征上应用 $1\times1\times1$ 卷积,其权重也由所有 3 个视图共享,这种卷积将维数 $C2$ 的特征仍然映射回 $C2$, 可以捕获不同信道之间的上下文信息;这 3 组特征被连接并输入到一个全连接(FC)层中。相对于 1×1×1 卷积,这个全连接(FC)层被应用于 C2×3 矩阵的每一行,它捕捉不同视图之间的上下文信息; 最后,通过 *Softmax* 函数对输出进行归一化,得到 α, 将归一化后的参数 α 与 $[x_{hw},x_{tw},x_{th}]$ 相乘得到输出特征值。

2.3 损失函数

本文用标准交叉熵损失函数来评价网络性能。对于不同网络分支来说,损失函数如式(7):

$$L=-\sum_{i=1}^{n}(y_i\log P_i). \quad (7)$$

其中, y_i 是动作所属类的真实标签; p 是训练后的模型进行预测属于不同的类别分数; i 表示不同行为类别的种类数;计算得到的总损失 L 通过反向传播算法将所有网络参数不断优化。为了验证光流对行为识别准确率的影响,本文构建了双流的结构进行试验,将损失函数 L_o 表示光流的损失函数, L_R 表示 RGB 帧的损失函数,模型的损失函数表示为式(8):

$$TotalLoss=L_o+L_R. \quad (8)$$

3 实 验

3.1 数据集

UCF101 是于 YouTube 收集而来,包含大量真

实动作视频，用于动作识别任务，共 101 个动作类别。视频的分辨率是 320×240，数据集分成 101 个行为类别，这些动作类别又被分成 25 个小组，每个小组又分别包含 4-7 个动作视频，一共包含13 320 个视频，占用存储空间约为 6.5G。101 个动作类型由五类组成：人人间的互动、人物间的互动、人体做出的行为动作、乐器演奏表演和体育竞技运动。UCF101 的宗旨是对一些实际行动类别进行学习，并探索未知的行为类型，鼓励推进行动识别工作的研究与发展，对研究视频行为分类工作意义重大。

HMDB-51 数据集共含 51 个人类行为动作，任一类别包含 101 个视频段，共计6 766个拍摄视频，对于每一个视频，有平均 3 s 左右持续时长。每一个剪辑进行了多轮的手动注释，剪辑多来自于电影中，小部分来自于一些公开数据集。广泛的面部动作如微笑，咀嚼等；常规的身体动作，如散步，摆手；人物交互的动作，如打球，梳头发，拔剑以及人类间的交互动作，如接吻，拥抱等。

3.2　实验细节和评价标准

在实验准备工作中，用 FFmpeg 工具将视频数据按照设定的帧率分割成视频帧，并记录每个视频的视频帧数量。为了合理地挑选训练样本，采纳了均匀采样的提取方式，设定时间位置，在其周边选取视频中的视频帧。为了满足 16 帧的需求，有时候需要对视频进行多次循环采样。在设定时间位置连续地取若干个视频帧以构成三维(H，W，T)视频信息。紧接着对视频帧进行时空裁剪操作，选取空间位置依照的规则是选取视频提取帧的中心或四个边角点位置之一。输入的原始视频帧尺寸为224×224，网络将其裁剪成 112×112 的大小，训练一次取 16 帧，由于训练数据是 RGB 图像，取信道数为 3。

实验中采用交叉熵损失函数，参数的微调工作将通过反向传播算法来开展，将权重衰减参数和动量参数分别设置为 0.9 和 0.001。训练网络起初，将学习率 lr 设定为 0.2，当验证损失趋于饱和后，将学习率减少到其十分之一大小；在网络进入微调的阶段时，学习率 lr 参数改变为 0.01，权重衰减参数改变为 $1e^{-6}$。本文在深度学习框架 PyTorch 上进行实验设计，实验工作站配置为 i7 6800k 酷睿 6 核、2 块 NIVDIA GTX1080Ti 8GB 显卡、64G 内存，256G 固态硬盘。

Top-N 准确率被采用来评价行为识别的性能。评判依据是：在测试视频数据的前 N 大分类概率中，判断正确的分类是否被包括其中，如果是，则认定为识别成功。

3.3　实验结果分析

本文提出的可变形卷积模块(DSTC)对行为识别性能产生的影响，见表 1，可明显看出，在引入可变形卷积模块(DSTC)后，本文所提出的改进的网络模型所取得的效果显著，能够有效地运用于行为分类任务。

表 1　UCF101 数据集上可变形卷积模块对实验性能的影响

Tab. 1　The impact of experiment result by deformable convolution factor on UCF101 dataset

method	acc (top-1)
ResNet-50	89.3
ResNet-50+DSTC	89.9

在 UCF101 和 HMDB51 数据集上，分别观察与对比模型的识别效果，将本文方法和当前一些优秀方法进行比较，见表 2 和表 3。

表 2　UCF101 数据集上使用不同网络模型的识别性能

Tab. 2　Recognition performance of different networks on UCF101 dataset

method	dimension	acc (top-1)
C3D	3D	82.3
P3D	3D	88.6
TDD	2D	90.3
TSN-RGB	2D	85.7
C3D	3D	82.3
P3D	3D	88.6
Two-stream I3D	3D	98.0
T3D	3D	90.3
3D-ResNet-50	3D	89.3
Ours-RGB	2D+3D	89.9
Ours-RGB+optical flow	2D+3D	90.3

表 3　HMDB51 数据集上使用不同网络模型的识别性能

Tab. 3　Recognition performance of different networks on HMDB51 dataset

method	dimension	acc (top-1)
TDD	2D	63.2
TSN	2D	68.5
Two-stream I3D	3D	80.7
T3D	3D	59.2
3D-ResNet-50	3D	61.0
Ours-RGB	2D+3D	61.8
Ours-RGB+optical flow	2D+3D	62.5

表 2 和表 3 表明，相比于一些现有的效果良好的方法，本文提出方法最终得到了相对更高的识别正确率。实验证明，通过对网络设置并行支路来处理光流信息，可加强网络的识别性能，进一步证明本文的方法有深远的研究价值。

在 UCF-101 数据集上，DSTC 方法训练和验证过程中交叉熵损失函数的缓慢变化，如图 5 所示。

随着训练和验证过程的进行,交叉熵损失值逐渐减小,DSTC 模型的识别效果逐渐变好。

为了能够更直观地观察本文方法的细节效果,从 UCF101 和 HMDB51 数据集中选取了 6 个差异比较显著的行为类别进行可视化研究,展示了 DSTC 方法在不同类别上的注意力热图,颜色越深代表该区域的特征显著性越强,模型对其关注度更高。从图中可以发现,我们的方法能够更好的动态适应特征的形变,更加有效地关注视频中更重要的特征区域,能够捕获到有效的时空信息进行学习,以提升行为识别的准确率, 如图 6 所示。

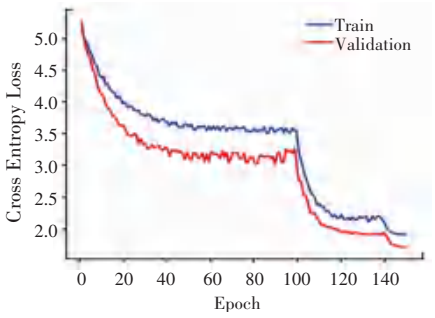


图 5 训练及验证过程损失函数变化

Fig. 5 The loss function of Training and Validation process

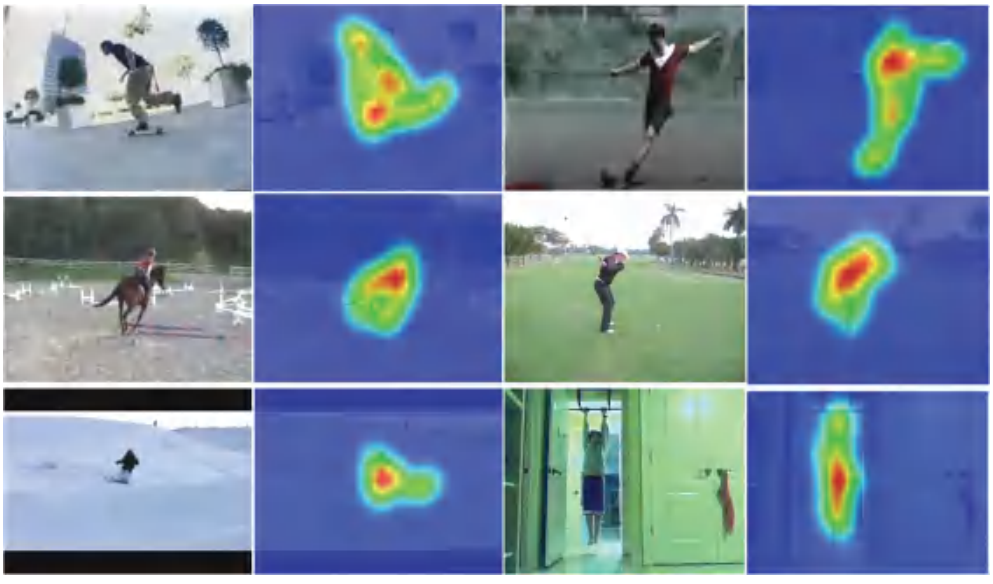


图 6 几种类别的注意力热图可视化

Fig. 6 Visualization of heat maps of attention for several categories

4 结束语

本文提出一种基于可变形卷积的改进型 3D-ResNet 网络,用于视频中的行为识别,通过引入形变卷积,构建了一个可自适应地协同学习视频三维信息的模块,将该模块替换 3D-ResNet 网络中部分卷积模块,提高行为识别效率。同时,融合了光流信息进行实验,证明了光流信息的引入可进一步提升模型的准确率,说明方法仍具有深远的研究价值。实验结果表明,与现有的一些效果显著的方法相比较而言,本文方法能拥有更准确的识别性能。

参考文献

[1] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks[C]//Proceedings of the IEEE international conference on computer vision. 2017: 764-773.

[2] WANG H, KLASER A, SCHMID C, et al. Action recognition by dense trajectories[C]// Computer Vision and Pattern Recognition (CVPR), 2011: 3169-3176.

[3] WANG H, SCHMID C. Action Recognition with Improved Trajectories[C]//IEEE International Conference on Computer Vision. 2013:3551-3558.

[4] Andrej Karpathy, George Toderici, Sanketh Shetty, et al. Large-scale video classification with convolutional neural networks [C]// In CVPR, 2014: 1725-1732.

[5] Jeffrey Donahue, Lisa Anne Hendricks, Sergio Guadarrama, et al. Long-term recurrent convolutional networks for visual recognition and description[C]// CVPR, 2015: 2625-2634.

[6] Karen Simonyan, Andrew Zisserman. Two-stream convolutional networks for action recognition in videos[C]//In NIPS, 2014: 568-576.

[7] TRAN D, BOURDEV L D, FERGUS R, et al. Learning spatiotemporal features with 3d convolutional networks [C]//International Conference on Computer Vision, 2015:4489-4497.

[8] qiu z, yao t, mei t. Learning spatio-temporal representation with pseudo-3d residual networks [C]//proceedings of the IEEE International Conference on Computer Vision. 2017: 5533-5541.

(下转第 64 页)