基于 DenseNet 和深度运动图的行为识别算法*

张　健,张永辉,何京璇
(海南大学,海南 海口 570228)

摘　要:结合深度信息以及 RGB 视频序列中丰富的纹理信息,提出了一种基于 DenseNet 和深度运动图像的人体行为识别算法。该算法基于 DenseNet 网络结构,首先获取彩色纹理信息和光流信息,然后从同步的深度视频序列获取深度信息,以增强特征互补性;再将空间流、时间流和深度流三种特征信息分别作为网络的输入;最后通过 LSTMs 进行特征融合和行为分类。实验结果表明,在公开的动作识别库 UTD-MHAD 数据集上,该算法识别准确率为 92.11%,与该领域中的同类算法相比表现优异。

关键词:行为识别;深度运动图像;DenseNet;光流
中图分类号:TP391.4　　**文献标识码**:A　　DOI: 10.19358/j.issn.2096-5133.2020.01.012
引用格式:张健,张永辉,何京璇. 基于 DenseNet 和深度运动图的行为识别算法[J]. 信息技术与网络安全,2020,39(1):63-69.

Action recognition algorithm based on DenseNet and depth motion map

Zhang Jian,Zhang Yonghui,He Jingxuan
(Hainan University,Haikou 570228,China)

Abstract:This paper proposes a human behavior recognition algorithm based on DenseNet and DMM,which integrates depth information and rich texture information in RGB video sequence. Based on the DenseNet network structure,the algorithm firstly obtains color texture information and optical flow information,and then obtains depth information from synchronous depth video sequence to enhance feature complementarity. Three kinds of characteristic information are used as the input of spatial flow network,temporal flow network and deep flow network. Then LSTMs is used for feature fusion and behavior classification. Experimental results show that the recognition rate of UTD-MHAD data set is 92.11%,which is an excellent performance compared with similar algorithms in this field.

Key words: action recognition;depth motion maps;DenseNet;optical flow

0　引言

近年来,有关人体行为识别的研究层出不穷,现如今已成为计算机视觉研究中日益关注的热点。其中,对视频中目标的行为识别一直以来都是一个非常活跃的研究领域[1]。虽然在对于静止图像识别的研究上取得了很大的成功,但是对视频类的行为识别如今仍是一个富有挑战性的课题。

在行为识别领域中,卷积神经网络得到了广泛的应用。早期的研究人员主要尝试融合光流与 RGB 视频帧来提高行为识别准确率。RGB 视频内的细节信息非常丰富,但缺乏深度信息,其识别准确率常常受光照变化、阴影、物体遮挡等因素的干扰。如文献[2]在 2014 年首次提出了创造性的双流网络,通过从 RGB 视频序列提取时空信息进行识别;文献[3]用基于长短期记忆的多尺度卷积神经网络来提取多层次表观特征,从而学习长周期的高层时空特征;文献[4]使用在 ImageNet 上进行预训练的 DenseNet 来搭建双流卷积神经网络,从中提取空间和时间特征,然后微调来进行单帧活动预测。

随着深度相机的出现,如 Kinect 设备,使深度信息更容易获取,它们不仅能够提供 RGB 视频序列,还可以提供深度视频序列。由于深度信息更容易获取,文献[5]将双流即空间流和时域流进行了训练,然后完成网络融合,提高了对识别的效

*基金项目:海南省自然科学基金资助项目(618MS027)

率。文献[6]将连续深度图的差异堆叠为深度运动图(DMM),再通过 HOG 将 DMM 中的相关特征提取出来。文献[7]利用 DMM 计算基于轮廓梯度方向直方图特征(CT-HOG)、局部二值模式特征(LBP)和边缘方向直方图特征(EOH),并决策级融合上述特征来避免光照变化和物体遮挡等不利环境因素的影响,取得了较好的识别效果。文献[8]提出了基于 CNN 和 LSTM 的模型 LRCN,该模型通过 CNN 提取空间信息,再通过一个 LSTM 网络提取时间信息,最后通过 Softmax 输出结果,LRCN 模型收敛慢,训练难度大。文献[9]提出了基于运动姿态描述子特征和最近算法的识别方法。文献[10]基于 RGB-D 的动作识别方法,将深度图序列表示为三对结构化动态图像(DIs),是一种简单有效的视频表示方法。这些文献都采用了深度视频序列包含的深度信息,虽然深度信息对光照不敏感,但是单一的深度信息缺少细节的描述。

1 本文方法

1.1 整体框架

在通常情况下,目前绝大多数双流网络不仅要融合光流信息,而且还需要融合彩色信息(RGB 视频帧)。彩色信息虽然容易获取并且具有丰富的细节信息,但是由于深度信息的缺乏,导致其准确率易受一些微小因素影响,如光照强度的变化、阴影、物体遮挡等。根据上述问题,为了更好地解决深度信息缺乏带来的问题,本文同时将光流信息、彩色信息和深度学习信息三者进行融合,这就为本文构造的框架提供了深度特征,从而增强了其特征互补性。因此,本文主要研究的是基于 DenseNet 的三种识别流网络,分别是空间流网络、时间流网络和深度流网络,三者之间相互独立;然后根据三种识别流网络对应输出的三种特征即空间特征、时间特征和深度特征,以此构造特征矩阵,再基于 LSTMs 系统进行特征融合操作,最终在 Softmax 层实现对融合之后的特征分类,其网络结构如图 1 所示。

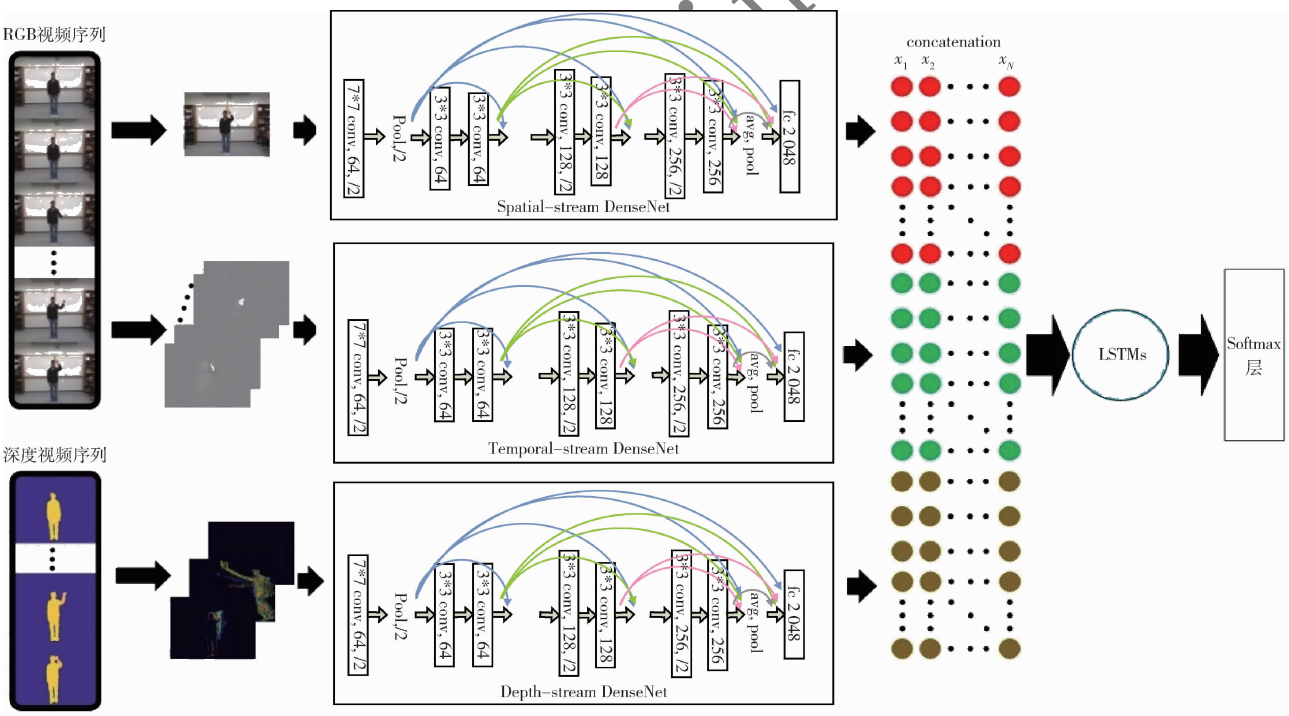


图 1 本文框架

1.2 DenseNet 网络结构

在深度学习网络中,梯度消失的问题会随着网络深度的加深愈加明显,现如今大量研究对此问题提出了解决方法,比如 ResNet[11]、Highway Networks、Stochastic depth[12]、FractalNets[13] 等,尽管这些算法的网络结构有所差别,但是核心都在于:创建层与层之间的全连接[12]。基于这个思路,DenseNet 网络结构可以直接将所有层进行连接,同时保证网络中每层间进行最大程度的信息传输。DenseNet 有效解决了梯度消失和梯度爆炸的问题,

加强了特征的传递，更有效地利用了特征，在一定程度上减少了参数量[13]。

图 2 是一个 5 层的 Dense Block 结构，每一层都接受前面所有的特征映射作为输入。在卷积神经网络中，网络层数就是连接的数量，若有 L 层就会有 L 个连接。但对于 DenseNet 网络，若有 L 层就会有 $L(L+1)/2$ 个连接。通俗讲，即每层的输入都是在上一层的全部层输出的基础上完成的。如图 2 所示，x_0 是输入，H_1 的输入是 x_0，H_2 的输入是 x_0 和 x_1（x_1 是 H_1 的输出）。

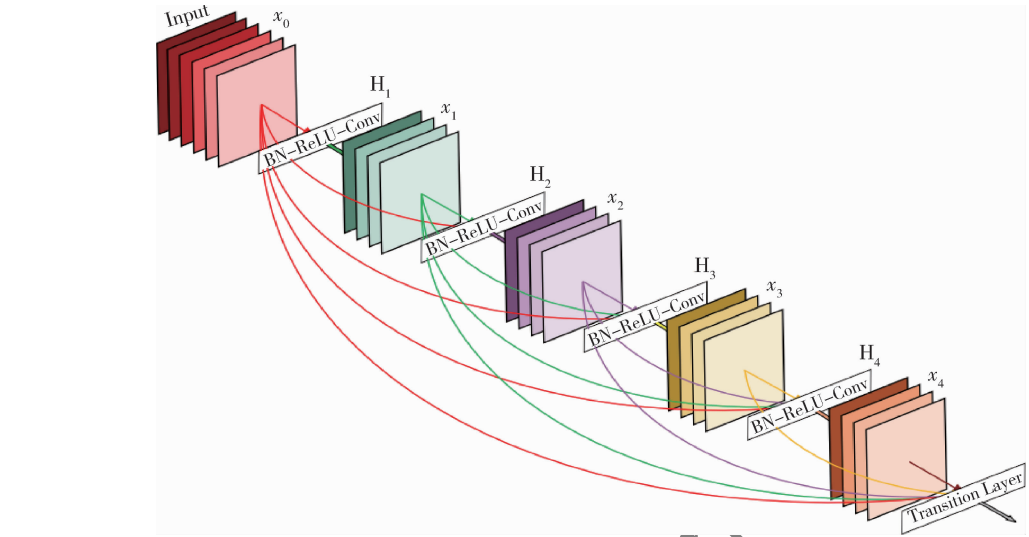


图 2　五层 Dense Block 结构图

DenseNet 具有网络更窄、参数更少的优点，其形成主要得益于这种 Dense Block 的设计。与其他宽度常处在成百上千以上的网络相比，Dense Block 中各卷积层输出的特征图的数量都非常小（小于 100）。这种连接方式的优点是有效地传递了特征和梯度，使网络训练变得更简单。网络层数加深往往导致梯度消失问题的出现，原因在于输入信息和梯度信息在很多卷积层之间传递，因此 DenseNet 网络的性能和效果更好。本文的网络结构利用 dense connection 将每一层直接与 input 及 loss 进行连接，有效解决网络加深和梯度加深带来的问题[14]。

1.3　LSTMs 特征融合

作为一种时间循环类型的长短期记忆网络，LSTM（Long Short-Term Memory）擅长预测并处理一些时间序列中间隔和延迟相对较长的重要事件[15]。LSTM 包括三个输入信息，双输出信息，内部维护三个门状态，其运算如式（1）~式（5）所示，其中 f_t、i_t、o_t、c_t、h_t 分别为遗忘门、输入门、输出门、当前状态值、当前时刻输出值，W_f、W_i、W_o、W_c 分别表示遗忘门、输入门、输出门及当前状态的权重梯度，x_t 为当前输入值，b_f、b_i、b_o 表示遗忘门、输入门、输出门的偏置项，σ 表示 Sigmoid 函数。

$$f_t = \sigma_f(W_f \cdot [x_t, h_{t-1}, c_{t-1}] + b_f) \tag{1}$$

$$i_t = \sigma_i(W_i \cdot [x_t, h_{t-1}, c_{t-1}] + b_i) \tag{2}$$

$$o_t = \sigma_o(W_o \cdot [x_t, h_{t-1}, c_{t-1}] + b_o) \tag{3}$$

$$c_t = \sigma_c(W_c \cdot [x_t, h_{t-1}] + b_i) + f_t c_{t-1} \tag{4}$$

$$h_t = o_t \sigma_h c_t \tag{5}$$

遗忘门 f_t 决定从 c_{t-1} 中丢弃什么信息；输入门 i_t 控制让多少新的信息加入到当前状态 c_t 中；输入的三个信息分别为当前时刻的输入值 x_t、上一时刻的输出值 h_{t-1} 和上一时刻的单元状态 c_{t-1}；输出门 o_t 决定 c_{t-1} 多少信息保留到输出 h_t 中；LSTM 单元的输出为当前时刻状态 c_t 和当前时刻输出 h_t。

本文 LSTMs 网络层引入 LSTM 记忆单元对循环卷积网络进行拓展。鉴于 LSTMs 网络能很好地对长序列历史信息进行记忆和控制，本文将特征矩阵按时间维度分为多个时间片段，并将它们的输出按顺序输入 LSTMs 网络层；然后，通过 LSTMs 网络层对特征图在时间轴上的依赖关系进行建模，并同时实现特征融合，如图 3 所示。本文特征矩阵 $\boldsymbol{x} = \{x_1, x_2, \cdots, x_t, \cdots, x_N\} \in \mathbf{R}^{2\,048\times3\times N}$ 是由三个识别流网络输出的空间特征、时间特征和深度特征构造成的，其中 N 为一个视频中采样帧数。与自然图像不同的是，每个 x_t 中的元素之间几乎没有空间依赖关

系，但在不同的 x_t 之间具有时间相关性。

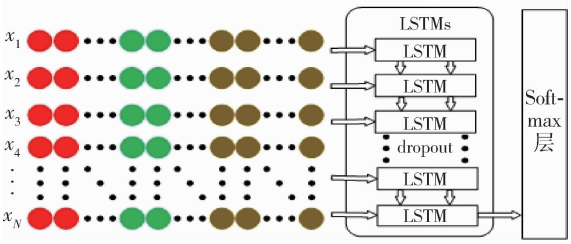


图 3　特征矩阵被分为多个时间片段作为 LSTMs 网络的输入

2　多流网络的数据预处理

2.1　空间流网络数据预处理

大量研究表明，为了获得较好的性能，可以采用 RGB 图像作为空间流的输入。本文识别模型在公开的数据集 UTD-MHAD[16] 上进行，利用空间流网络中的视频采样对静态 RGB 帧图像进行识别。本文在没有特殊指定的情况下，使用全采样的方式对 UTD-MHAD 数据集的 RGB 视频进行采样，原因是该数据集里的 RGB 视频时长均不超过 4 s，每个视频的帧数均不超过 70 帧。此外，为了去除无运动条件，本文对初始和结束帧进行了去除，即删除每个 RGB 视频序列的前 3 帧和后 3 帧，这是因为受试者在视频的开头和结尾处往往只有小幅度的肢体运动，大多数处于静止状态，这些初始帧和结束帧对视频序列的运动特性无影响。

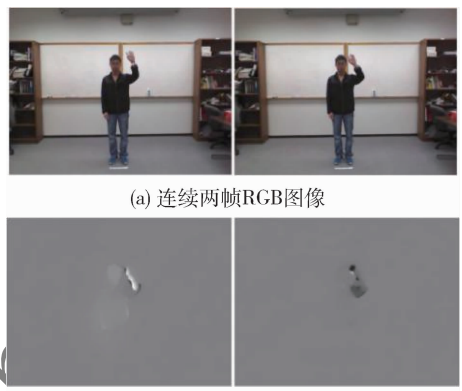
2.2　时间流网络数据预处理

在表达图像序列运动信息的各种方法中，光流较为简便且实用性高，因此在描述运动特征时，往往通过光流来实现，所以运动中的光流信息在行为识别方面起到重要作用。本文提供了不同的光流算法来为网络架构提供相关的光流信息。虽然大部分的研究主要使用的两种光流算法是 Brox[17] 与 TV-L1[18]，但是每种不同的光流算法之间仍存在一些差异。文献[4]从两种光流算法的差异出发进行了相关实验，比较了 Brox 和 TV-L1 的性能，实验结果表明 TV-L1 算法性能更优。基于此实验结论，本文在没有预先指明的条件下，使用 TV-L1 光流算法从 RGB 视频序列中提取光流信息。式(6)为此 TV-L1 光流算法的具体目标函数：

$$E_\theta = \int_\Omega \left\{ |\nabla U| + \frac{1}{2\theta}(U-V)^2 + \lambda |\rho(V)| \right\} \mathrm{d}X \tag{6}$$

其中，$\rho(V) = I_1(X+U_0) + (U-U_0)\nabla I_1 - I_0$，$I_0$ 与 I_1 表示连续两帧图像，$X = (x,y)$ 表示 I_0 上的像素点，V 是该算法引入的辅助变量，$U = (u,v)$ 表示二维光流向量，θ 表示一个很小的常量。

本文使用该方法计算水平和垂直两个方向的光流，并将 TV-L1 光流的水平和垂直分量进行了调整，其光流数值小于 0 值的都设置为 0，大于 255 的值都设置为 255。为了能够作为时域网络通道的输入，需要对其进行线性变换，最终将两个方向的光流保存为两张灰度图像，如图 4 所示。



(a) 连续两帧RGB图像

(b) 水平与垂直方向的光流图像

图 4　两帧连续的 RGB 图像对应的光流图像

为了有效提取视频的运动信息，本文采用 10 个连续帧的水平和垂直光流堆叠形成 20 个密集光流图像，并将其作为时间流网络的输入。本文采用预先在 ImageNet 上训练的 DenseNet 网络，并在 UTD-MHAD 数据集对其网络全局微调，但这些网络的输入是 RGB 图像，因此第一个卷积层的通道数为 3。但本文的时间流网络输入 20 个光流图像，与第一个卷积层的通道数不匹配，这里采用跨模态交叉预训练的方法，将第一个卷积层的 3 个通道的权值取平均，再将其复制 20 份作为时域网络第一个卷积层 20 个通道的权值，而时间流网络上的其他层的权值参数保持不变。

2.3　深度流网络数据预处理

深度图可以用来捕捉三维结构和形状信息。文献[6]提出动作的运动特征可以用投影到三个正交的笛卡尔平面上的深度帧表示。由于其计算较简单，因此本文采用了与文献[6]中相同的方法。如图 5 所示，本文方法是将三维深度图在三个正交的笛卡尔平面分别生成前视图 f、侧视图 s 和顶视图 t 的二维投影图，其投影图用 $\mathrm{map}_v(v \in \{f,s,t\})$ 表示。

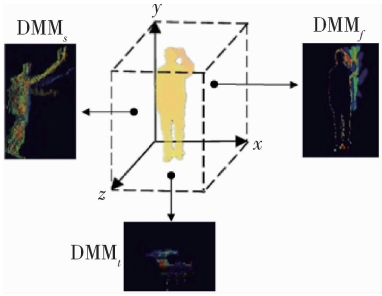


图 5　前视图 f、侧视图 s 和顶视图 t 的提取框架

该三个投影图中的像素值分别用 x、y、z 表示，其中 z 表示为深度坐标系中的深度值，则 (x,y,z) 表示为深度坐标系中的一个点。与文献[6]不同的是，对于每个投影映射，其计算的运动能量是两个连续映射之间的绝对差值，没有阈值。对于 N 帧深度视频序列，通过式(7)将整个深度视频序列的运动能量叠加得到深度运动图 DMM_v：

$$\mathrm{DMM}_v = \sum_{i=m}^{n} \left| \mathrm{map}_v^{i+1} - \mathrm{map}_v^{i} \right| \tag{7}$$

其中 i 代表帧索引，map_v^i 为投影视图 v 下的第 i 帧投影图，$m \in \{1,\cdots,N\}$ 和 $n \in \{1,\cdots,N\}$ 表示起始帧和结束帧索引。将提取的每个 DMM 中非零区域作为前景。DMM 以一种可区分的方式从三个投影视图有效地捕捉运动的特征。这就是这里使用 DMM 作为动作识别的特征描述符的原因。由于不同动作视频序列的 DMM_v 可能会有不同的尺寸，为了减小类内变异性，例如由于受试者身高的不同，在同一投影视图下，使用双三次插值将所有 DMM_v 调整为固定的尺寸。值得注意的是，由于这里并没有像文献[6]中那样计算 DMM 的 HOG 特征，从而降低了特征提取过程的计算复杂度。每个 DMM 的固定尺寸被简单地设置为所有尺寸平均值的一半，对训练特征集和测试特征集采用 PCA 方法进行降维。利用训练特征集计算 PCA 变换矩阵，并应用于测试特征集，该降维步骤有助于提高分类的计算效率。在本文的实验中，保留了最大的 85% 的特征值。

3　实验结果与分析

3.1　实验数据集

本文使用 UTD-MHAD 数据集[16]进行实验，该数据集包含 RGB 视频序列、深度信息、骨架和惯性数据，包含了 27 种不同动作。每个动作分别由 8 个人执行 4 次，移除 3 个已损坏的序列后，总共有 861 个行为序列。

3.2　实验参数设置

本文实验在基于 PyTorch1.1 深度学习框架的 Linux 系统下进行，采用预训练的网络模型对 UTD-MHAD 数据集进行行为识别，同时为了利于网络的训练应用了较小的学习速率。空间流网络的初始学习速率设置为 0.000 5；时间流网络由于其输入数据为光流图像，与 RGB 图像存在一定差异，设置相对较大的初始学习速率将有利于网络的快速收敛，实验中设置为 0.01；深度流网络与时间流网络初始学习速率一样设置为 0.01。优化时学习速率采用自适应方法，即学习速率根据学习结果自动更新。基于对内存容量、使用率以及收敛速度等方面的考虑，将空间、时间、深度网络的 Batch-size 分别设置为 25、32 和 32。此外，为了使网络收敛速度更快，实验中将动量设置为 0.9。

在对三个通道网络进行训练时，需要选定合适的优化目标函数，在这里均选择使用交叉熵损失函数，使用随机梯度算法进行优化。在本实验中，每一轮训练都要进行 500 次迭代，每次迭代中随机选取 32 个视频作为训练样本，并采用前面所描述的算法进行数据增强，之后将其裁剪成尺寸大小为 224×224 的网络输入，再对其进行归一化操作。本实验的特征矩阵由三个识别流构建而成，以时间维度为划分依据，将其分为 32 个时间片段，同时使用了 32 个 LSTM 单元，将每一个时间片段上特征图的值作为 LSTM 单元的输入，进行 32 次迭代更新；LSTMs 网络的 dropout 概率值为 0.5。完成最后一次迭代后，把 LSTM 单元的输出进行分类处理，这里通过送入 Softmax 层进行分类来实现，为了有效地检验学习模型的性能，在每次轮回的训练后对测试集进行测试，测试时遵循 THUMOS13 挑战机制[19]。

3.3　UTD-MHAD 数据集的行为识别结果分析

本文采用基于 DenseNet 的三通道网络模型对沿着视线方向的动作进行识别，并融合光流、RGB 和深度运动图等三种特征信息，在颇具挑战性的数据集 UTD-MHAD 上进行试验，其准确率为 92.11%。由表 1 可知，通过融合时空深三通道的特征信息能取得更好的识别效果，它相比现有的 STSDD 识别方法，能将准确率提升近 1 个百分点，与文献中的其他 6 种方法相比较，准确率提高了 1.96% ~ 13.01%。此外，从该数据库上的混淆矩阵中(图 6)可以看出，

表 1　本文方法与现有方法的比较

方法	准确率/%
DMM-CRC[16]	79.1
PSEUDO-COLOR MHI[20]	84.3
GF + LF[21]	84.4
DMM-CTHOG-LBP-EOH[7]	88.40
DEPTH-INERTIAL SENSING[22]	86.3
S^2DDT[23]	89.04
STSDD[10]	91.16
本文方法	92.11

本文算法在许多动作下都取得了不错的识别效果，但“右手绘制 x”、“右手绘制圆圈(顺时针)”和“右手绘制圆圈(逆时针)”等动作识别率相对偏低，其原因是这些动作具有较高的相似度，从而导致近0.2的误识别率。

4　结束语

本文提出一种基于光流和深度运动图的人体行为识别算法，用于三维动作识别。该算法不仅保留了时空信息，还增加了深度信息，并利用 LSTMs 网络的长期记忆的特点对本文特征矩阵进行融合，以进行最终动作识别。本文算法已经在具有挑战性的 UTD-MHAD 数据集上进行了评估，并获得了较好的识别效果，且算法对相似动作也具有一定的分辨能力。但本文算法在识别速度与实时性方面存在不足，因此如何把相关的光流算法集成到本文网络框架上是笔者未来的研究方向。

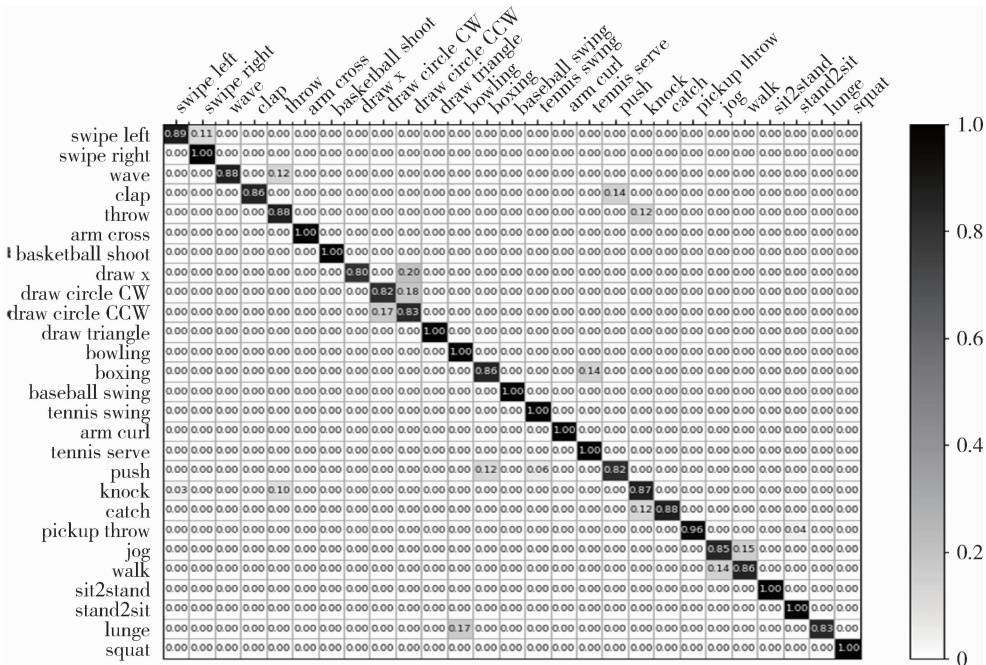


图 6　混淆矩阵

参考文献

[1] LUO H L, WANG C J, LU F. Survey of video behavior recognition[J]. Journal on Communications, 2018, 39(6): 169-180.

[2] SIMONYAN K, ZIZZERMAN A. Two-stream convolutional networks for action recognition in videos[C]. Advances in Neural Information Processing Systems. Montreal, QC: Neural Information Processing Systems Foundation, 2014: 568-576.

[3] Luo Yangxing. Research on multi-modal dynamic gesture recognition algorithm based on vision[D]. Guangzhou: South China University of Technology, 2018.

[4] MA C Y, CHEN M H, KIRA Z, et al. TS-LSTM and temporal inception: exploiting spatiotemporal dynamics for activity recognition[J]. Signal Processing-Image Communication, 2019, 71(2): 76-87.

[5] Xue Luqiang. Research on human behavior recognition based on dual-stream fusion convolutional neural network[D]. Hefei: Anhui University, 2018.

[6] Li Zhifei, Zheng Zhonglong, Lin Feilong, et al. Action recognition from depth sequence using depth motion maps-based local ternary patterns and CNN[J]. Multimedia Tools and Applications, 2019, 78(14): 19587-19601.

[7] KHAIRE P, KUMAR P, IMRAN J. Combining CNN streams of RGB-D and skeletal data for human activity recognition[J]. Pattern Recognition Letters, 2018, 115:

107-116.

[8] DONAHUE J, HENDRICKS L A, ROHRBACH M, et al. Long-term recurrent convolutional networks for visual recognition and description [C]. Proceedings in IEEE Comput Soc Conf Comput Vision Pattern Recognit. Boston, MA: IEEE Computer Society, 2015: 2625-2634.

[9] ALELUYA E R M, VICENTE C T. Faceture ID: face and hand gesture multi-factor authentication using deep learning [J]. Procedia Computer Science, 2018, 135: 147-154.

[10] Hou Yonghong, Li Zhaoyang, Wang Pichao, et al. Skeleton optical spectra-based action recognition using convolutional neural networks [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 28 (3): 807-811.

[11] Wu Zifeng, Shen Chunhua, HENGEL A V D. Wider or deeper: revisiting the ResNet model for visual recognition [J]. Pattern Recognition, 2019, 40 (10): 2395-2401.

[12] Ren Meng. An overview of image vision recognition algorithms [J]. Intelligent Computers and Applications, 2019, 9 (3): 294.

[13] Feng Xinjie, Yao Hongxun, Zhang Shengping. An efficient way to refine DenseNet [J]. Signal, Image and Video Processing, 2019, 13 (5): 959-965.

[14] BRAHIMI S, AOUN N B, BENOIT A, et al. Semantic segmentation using reinforced fully convolutional densenet with multiscale kernel [J]. Multimedia Tools and Applications, 2019, 78 (15): 22077-22098.

[15] LIU T L, QIAO Q W, WAN J W, et. al. Human action recognition via spatio-temporal dual network flow and visual attention fusion [J]. Electronic and Informatics, 2018, 40 (10): 2395-2401.

[16] CHEN C, JAFARI R, KEHTARNAVAZ N. UTD-MHAD: a multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor [C]. Proceedings of the 2015 IEEE International Conference on Image Processing. Quebec City, QC: IEEE Computer Society, 2015: 168-172.

[17] BROX T, BRUHN A, PAPENBERG N, et al. High accuracy optical flow estimation based on a theory for warping [C]. Lecture Notes in Computer Science. Prague, Czech republic: Springer Verlag, 2004: 25-36.

[18] Huang Youwen, Wan Chaolun, Feng Heng. Multi- feature fusion human behavior recognition algorithm based on convolutional neural network and long short term memory neural network [J]. Laser & Optoelectronics Progress, 2019, 56 (7): 243-249.

[19] JIANG Y G, LIU J, ZAMIR A, et al. Competition track evaluation setup, the first international works- hop on action recognition with a large number of classes [EB/OL]. (2018-05-20) [2019-10-01]. https://www. crcv. ucf. edu/ICCV13-Action-Workshop/.

[20] Chun Qiuping, Zhang Erhu. Human action recognition based on improved motion history image and deep convolutional neural networks [C]. Proceedings of the 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics. Shanghai, China: Institute of Electrical and Electronics Engineers Inc., 2018: 1-5.

[21] ESCOBEDO E, CAMARA G. A new approach for dynamic gesture recognition using skeleton trajectory representation and histograms of cumulative magnitudes [C]. Proceedings of the 29th SIBGRAPI Conference on Graphics, Patterns and Images, 2017: 209-216.

[22] DAWAR N, KEHTARNAVAZ N. Real-time continuous detection and recognition of subject specific smart TV gestures via fusion of depth and inertial sensing [J]. IEEE Access, 2017, 6 (12): 7019-7028.

[23] Wang Pichao, Wang Shuang, Gao Zhimin, et al. Structured images for RGB- D action recognition [C]. Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops. Venice, Italy: Institute of Electrical and Electronics Engineers Inc., 2018: 1005-1014.

(收稿日期: 2019-10-18)

作者简介:

张健(1974 -),男,博士研究生,讲师,主要研究方向:视频识别、嵌入式智能系统与工程。

张永辉(1974 -),通信作者,男,博士生导师,教授,主要研究方向:嵌入式智能系统与工程等。E-mail: yhzhang@hainanu. edu. cn。

何京璇(1997 -),女,本科,主要研究方向:机器学习。

版权声明