

# 基于四元数时空卷积神经网络的人体行为识别<sup>\*</sup>

孟 勃, 刘雪君, 王晓霖

(东北电力大学信息工程学院 吉林 132012)

**摘 要:**传统卷积神经网络(CNN)只适用于灰度图像或彩色图像分通道的特征提取,忽视了通道间的空间依赖性,破坏了真实环境的颜色特征,从而影响人体行为识别的准确率。为了解决上述问题,提出一种基于四元数时空卷积神经网络(QST-CNN)的人体行为识别方法。首先,采用码本算法预处理样本集所有图像,提取图像中人体运动的关键区域;然后将彩色图像的四元数矩阵形式作为网络的输入,并将 CNN 的空间卷积层扩展为四元数空间卷积层,将彩色图像的红、绿、蓝通道看作一个整体进行动作空间特征的提取,并在时间卷积层提取相邻帧的动态信息;最后,比较 QST-CNN、灰度单通道 CNN(Gray-CNN)和 RGB 3 通道 CNN(3Channel-CNN)3 种方法的识别率。实验结果表明,所提方法优于其他流行方法,在 Weizmann 和 UCF sports 数据集分别取得了 85.34% 和 80.2% 的识别率。

**关键词:** 四元数;卷积神经网络;码本;人体行为识别

**中图分类号:** TP391.4 TH164 **文献标识码:** A **国家标准学科分类代码:** 510.4050

## Human body action recognition based on quaternion spatial-temporal convolutional neural network

Meng Bo, Liu Xuejun, Wang Xiaolin

(School of Information Engineering, Northeast Electric Power University, Jilin 132012, China)

**Abstract:** Traditional CNN models are suitable for the feature extraction of gray image sequences or the color image separate channels, which ignores the interdependency among the channels and destroys the color features of real world objects, thereby affects the accuracy rate of human body action recognition. In order to solve this problem, a human body action recognition method is proposed based on quaternion spatial-temporal convolutional neural network (QST-CNN). Firstly, codebook algorithm is adopted to process all the images in the sample set and extract the key regions of human body motion in the images. Then, the quaternion matrix expression of the color images is taken as the input of the QST-CNN. The spatial convolutional layer of CNN is expended as a quaternion spatial convolutional layer. The values of the red, green, and blue channels of the color images are considered simultaneously as a whole in a spatial convolutional layer to conduct the extraction of the action spatial features, and avoid the loss of spatial relationships. The dynamical information of adjacent frames is extracted in a temporal convolutional layer. Finally, experiment was conducted, in which QST-CNN, gray single channel CNN (Gray-CNN) and RGB three channel CNN (3 Channel-CNN) were compared. The experiment result demonstrates that the QST-CNN boosts the performance of action recognition, the proposed method is superior to other popular methods and achieves the recognition rates of 85.34% and 80.2% in the Weizmann and UCF sports datasets, respectively.

**Keywords:** quaternion; convolutional neural network (CNN); codebook; human body action recognition

## 0 引 言

人体行为识别一直是计算机视觉领域内的研究热

点。在视频监控、人机交互和运动分析等诸多领域有着重要的应用价值<sup>[1-2]</sup>。人体行为识别是基于视频的复杂过程,它既包含单帧的空间信息又需考虑相邻帧之间的动态关系。因此,人体行为的特征提取和分类是一项具

收稿日期:2017-07 Received Date: 2017-07

<sup>\*</sup> 基金项目:国家自然科学基金青年项目(61602108)、吉林市科技局项目(20166016)资助

有挑战性的研究。

目前的人体行为特征提取方法主要分为两种:手工提取和自动学习。基于手工提取的人体行为识别方法包括特征提取,特征表示和动作分类。手工提取的动作特征有光流直方图(histograms of optical flows, HOF)<sup>[3]</sup>、方向梯度直方图(histograms of oriented gradients, HOG)<sup>[4]</sup>和时空兴趣点(space-time interest points, STIP)<sup>[5]</sup>等。虽然手工提取方法取得了很多成果,但特征提取过程的计算量大,对多样的动作视频集不具有泛化能力,导致其无法应用于复杂多变的真实环境。

近几年,基于深度学习的人体行为识别越来越引起关注。Ji SH. W. 等人<sup>[6]</sup>首次构建了7层结构的3D卷积神经网络(convolutional neural networks, CNN)人体行为识别模型,通过将2D卷积核扩展成3D卷积核获取动作的时间信息,但它依然依赖于梯度、光流等底层特征;Simonyan K 等人<sup>[7]</sup>提出双流CNN模型,分别提取原始图像中动作外观特征和光流图像中运动动态特征;Cheron G 等人<sup>[8]</sup>通过融合人体不同部位的运动特征构建了基于姿态的CNN模型,但是忽略了动作的动态信息。为了解决CNN网络容易陷入局部最优的问题,Ijina E. P. 等人<sup>[9]</sup>采用遗传算法初始化CNN网络参数,提高了UCF50动作数据集的识别率;Liu C 等人<sup>[10]</sup>结合CNN和条件随机场(conditional random fields, CRFs)构建了基于卷积神经随机场(convolutional neural random fields, CNRFs)的人体行为识别模型,其贡献在于实现了对无分割视频中行为的识别和预测,即不再局限于对只包含单个动作的视频进行行为识别,还能够对包含多种连续动作的视频进行行为识别和预测,推进了基于CNN的人体行为识别方法在现实环境中的应用。然而,上述方法均针对灰度图像序列,或在第1层卷积层采用3D卷积核将彩色图像的三通道信息直接转换成单通道的灰度图像,忽略了三通道间的空间关系和目标的颜色特征,减少了网络提取的动作特征,影响了行为识别的准确率。为了解决上述问题,本文提出一种基于四元数时空卷积神经网络(quaternion spatial-temporal convolutional neural network, QST-CNN)的人体行为识别方法。

## 1 算法的理论基础

### 1.1 四元数

1843年,Hamilton A. S. 等人<sup>[11]</sup>提出了四元数的概念,它是由1个实部和3个虚部组成的超复数,即:

$$q = r + ai + bj + ck \quad (1)$$

式中: $r, a, b$ 和 $c$ 是实数, $i, j, k$ 是虚数单元。它们遵循如下运算规则:

$$\begin{aligned} i^2 = j^2 = k^2 = -1, ij = -ji = k, jk = -kj = i, \\ ki = -ik = j \end{aligned} \quad (2)$$

由式(2)可知,四元数的乘法不满足交换律。当实部 $r = 0$ 时, $q$ 被称为纯四元数。四元数的共轭和模分别为:

$$q^* = r - ai - bj - ck \quad (3)$$

$$\|q\| = \sqrt{qq^*} = \sqrt{r^2 + a^2 + b^2 + c^2} \quad (4)$$

Sangwine S. J.<sup>[12]</sup>首次采用纯四元数矩阵表示彩色图像,它将图像的三通道看作是一个整体,而不是单独的组件。彩色图像 $Q$ 中 $(x, y)$ 位置的像素可以表示为:

$$Q(x, y) = Q_r(x, y)i + Q_g(x, y)j + Q_b(x, y)k \quad (5)$$

式中: $Q_r(x, y), Q_g(x, y), Q_b(x, y)$ 分别表示彩色图像的红绿蓝3个通道。另外,图像一个像素也可以表示为向量的形式,即:

$$Q(x, y) = (Q_r(x, y), Q_g(x, y), Q_b(x, y)) \quad (6)$$

因此,一张彩色图像可以表示为一个纯四元数矩阵。彩色图像的四元数表示已经成功应用于很多计算机视觉领域,如人体追踪<sup>[13]</sup>、人脸识别<sup>[14]</sup>,而对于人体行为识别领域,将彩色图像像素作为一个整体进行处理,保持三通道的相关性和整体性,对提高人体运动特征的提取性能具有重要的理论依据和现实意义。

### 1.2 码本算法

码本算法<sup>[15]</sup>是一种通过对视频背景建模检测运动目标的算法。一般分为两步:背景建模和检测。设 $X = \{x_1, x_2, \dots, x_n\}$ 为单个像素的训练序列,长度为 $n$ , $C = \{c_1, c_2, \dots, c_L\}$ 为该像素的一个码本,它包含 $L$ 个码字。每一个码字包含RGB向量的平均向量 $v_i = (\bar{R}_i, \bar{G}_i, \bar{B}_i)$ 和一个六元组 $aux_i = (\check{I}_i, \hat{I}_i, f_i, \lambda_i, p_i, q_i)$ , $\check{I}_i$ 和 $\hat{I}_i$ 分别代表码字建模期间亮度的最大值和最小值, $f_i$ 是该码字出现的频率, $p_i$ 和 $q_i$ 是首次和最后一次匹配的时间, $\lambda_i$ 是没有匹配的最大时间间隔。在进行前景目标检测时,首先经过 $T$ 帧的视频序列进行背景学习,得到背景模型的码本,然后根据图像像素与对应码本之中的码字进行匹配,如果像素值落在相应的码字内,则归类为背景,否则为前景,并对码本进行码字更新<sup>[16-17]</sup>。

## 2 人体运动的关键区域提取

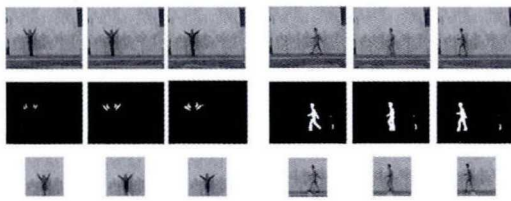
从视频图像序列中准确地检测并提取前景目标区域是人体行为识别的首要步骤,其精度会直接影响到后续行为分析与行为理解等处理步骤的性能<sup>[18-19]</sup>。本文采用码本算法进行背景建模的过程中,亮度标准定义为:

$$\begin{aligned} bn(I, <\hat{I}_i, \check{I}_i>) = \\ \begin{cases} \text{true}, & \alpha \hat{I}_i \leq \|x_i\| \leq \min\{\beta \hat{I}_i, \check{I}_i/\alpha\} \\ \text{false}, & \text{其他} \end{cases} \end{aligned} \quad (7)$$

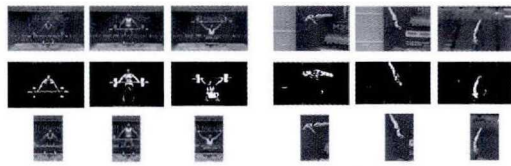
式中:  $\alpha(\alpha < 1)$  和  $\beta(\beta > 1)$  为亮度范围的阈值。另外, 颜色标准定义为:

$$cd(x_i, v_i) = \sqrt{(R^2 + G^2 + B^2) - \frac{(\bar{R}_i R + \bar{G}_i G + \bar{B}_i B)}{\bar{R}_i^2 + \bar{G}_i^2 + \bar{B}_i^2}} \quad (8)$$

当像素  $x_i$  满足  $bn(I, < \hat{I}_m, \hat{I}_m >) = \text{true}$  和  $cd(x_i, v_m) \leq \varepsilon$  这两个条件时, 则判定该像素属于背景, 否则, 属于前景。然后, 以前景像素的中心点画包围框, 提取人体运动的关键区域, 去除图像的冗余背景。因为不同数据集中的图像尺寸不同, 所设定的包围框大小也不同。如图 1 所示, 对于 Weizmann 和 UCF sports 数据集, 本文分别采用了  $90 \times 90$  和  $250 \times 400$  的包围框。图 1 所示为原图、码本检测结果图和人体运动关键区域图。



(a) Weizmann 数据集运动区域提取  
(a) Action regions extracted from Weizmann dataset



(b) UCF sports 数据集运动区域提取示例  
(b) Action regions extracted from UCF sports dataset

图 1 人体运动关键区域提取示例

Fig. 1 Examples of human body action key region extraction

### 3 基于 QST-CNN 的人体行为识别

#### 3.1 四元数空间卷积层

在空间卷积层, 将传统提取灰度图像特征的 2D 卷积

核扩展为纯四元数形式  $W = (W_r, W_g, W_b)$ 。假设一张彩色输入帧为  $Q = (Q_r, Q_g, Q_b)$ , 那 QST-CNN 第  $i$  层第  $j$  个特征图  $Z$  中  $(x, y)$  位置像素值的卷积运算为:

$$Z^{i,j}(x, y) = f(\sum_p \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \text{conv}(W_{i,j,p}^{n,m}, Q^{(i-1),p} \times (x+n, y+m)) + b^{i,j}) \quad (9)$$

$$\text{conv}(W, Q) = W \otimes Q + W \times Q \quad (10)$$

$$W \otimes Q = (W_r Q_r, W_g Q_g, W_b Q_b) \quad (11)$$

$$W \times Q = (W_g Q_b - W_b Q_g, W_b Q_r - W_r Q_b, W_r Q_g - W_g Q_r) \quad (12)$$

在式(9)中,  $f$  为 Sigmoid 激活函数,  $b^{i,j}$  是偏置,  $W_{i,j,p}^{n,m}$  为连接第  $i$  层第  $j$  个特征图和第  $(i-1)$  层第  $p$  个特征图的卷积核的  $(n, m)$  位置的权值向量,  $N$  和  $M$  为该卷积核的长和宽。式(10)中,  $\otimes$  操作是两个纯四元数向量元素对应相乘, 提取每个通道上的空间特征;  $\times$  操作是叉积操作, 提取不同颜色通道的空间关系。

#### 3.2 时间卷积层和下采样层

和姿态识别不同, 行为识别针对一个动态过程, 获取不同帧之间的运动信息是非常重要的。假设四元数空间卷积层的输出特征图序列为  $Z = \{Z_1, Z_2, \dots, Z_t\}$ , 那每个颜色通道的时间卷积结果为:

$$Z_{c,t}^i = f(\sum_{s=0}^S w_{c,s}^i z_{c,t-s}^{i-1} + b^i) \quad (13)$$

式中:  $f$  为 sigmoid 函数,  $b^i$  为第  $i$  层的偏置,  $c$  为指图像的第  $c$  个通道,  $s$  为时间卷积核的第  $s$  维, 共  $S$  维,  $w_{c,s}^i$  为第  $i$  层第  $c$  个通道第  $s$  时间维的权值; 假设输入序列的帧数为  $T$ , 那时间卷积层输出序列的帧数为  $T' = T - S + 1$ 。在下采样层, 本文采用平均值操作对每个通道进行采样。

#### 3.3 网络的整体结构

本文设计了一个多层的 QST-CNN, 由四元数时空卷积层、时间卷积层、下采样层和全连接层组成, 最后采用 Softmax 对动作进行分类。由于不同数据集的图像大小、视频长度不同, QST-CNN 的输入图像序列的尺寸也存在差异, 因此对于不同数据集, 网络的参数是不同的。以 Weizmann 数据集为例, 构建的 QST-CNN 的结构如图 2 所示。

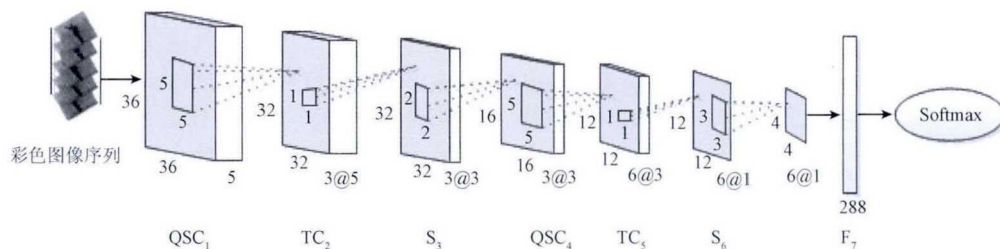


图 2 QST-CNN 的网络结构

Fig. 2 Network architecture of QST-CNN

- 1)采用最近邻插值法将图像缩放至 $36\times 36$ 像素,输入大小为 $36\times 36\times 5$ 的彩色图像序列,其中,5为图像序列的帧数;
- 2)在第1层四元数空间卷积层 $QSC_1$ 中,卷积核尺寸为 $5\times 5\times 3$ ,其中 $5\times 5$ 为图像分辨率,3为彩色图像的通道数,用3个卷积核产生3个 $32\times 32\times 5$ 的特征图序列;
- 3)在第2层时间卷积层 $TC_2$ 中,将第1层特征图中的每个通道用 $1\times 1\times 3$ 卷积核对进行时间卷积操作,其中,3为时间维度,得到3个 $32\times 32\times 3$ 的特征图序列;
- 4)在第3层下采样层 $S_3$ 中,采用 $2\times 2$ 窗口对第2层特征图中的每个通道进行下采样操作,得到3个 $16\times 16\times 3$ 的特征图序列;
- 5)在第4层四元数空间卷积层 $QSC_4$ 中,卷积核尺寸为 $5\times 5\times 3$ ,用6个卷积核对第3层的特征图进行四元数卷积操作,产生6个 $12\times 12\times 3$ 的特征图序列;
- 6)在第5层时间卷积层 $TC_5$ 中,将第4层特征图中的每个通道用 $1\times 1\times 3$ 卷积核对进行时间卷积操作,产生6个 $12\times 12\times 1$ 的特征图序列;
- 7)在第6层下采样层 $S_6$ 中,采用 $3\times 3$ 窗口对第5层特征图中的每个通道进行下采样操作,得到6个 $4\times 4\times 1$ 的特征图序列,即96个纯四元数向量;
- 8)在第7层全连接层 $F_7$ 中,将第6层每一个特征图都拉伸成一个特征向量,得到288维度的特征向量;
- (9)采用Softmax对特征向量进行分类,得到行为标签。

4 实验结果及分析

为了验证本文方法的有效性和准确性,本文基于Deep learning toolbox 框架,采用 MATLAB 2015a 进行程序编译,选择在 Weizmann<sup>[20]</sup>和 UFC sports<sup>[21]</sup>数据集上比较 QST-CNN 和灰度单通道 CNN (gray single channel CNN, Gray-CNN)、三通道 CNN(3 Channel-CNN)以及其他流行方法的识别率。如图 2 所示,Gray-CNN 是针对灰度图像序列的 5 层网络,由卷积层、下采样层和全连接层构成,3 Channel-CNN 的输入是彩色图像序列,将红绿蓝分通道处理,最后采用平均融合方法得到动作的类别。

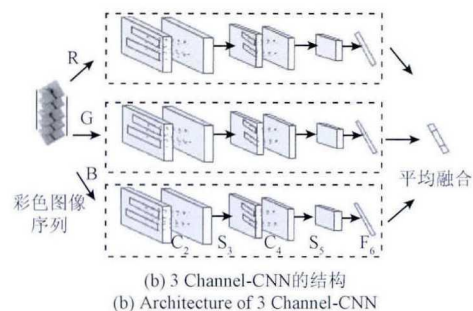
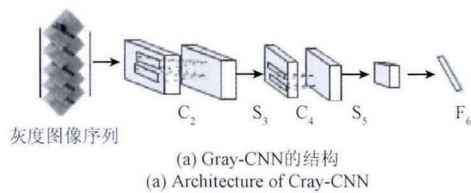


图 3 Gray-CNN 和 3 Channel-CNN 的结构  
Fig.3 Architectures of Gray-CNN and 3 Channel-CNN

4.1 Weizmann 数据集

Weizmann 数据集包含了 10 种动作,分别为弯腰、跳跃击掌、跳跃、双腿跑跳、跑步、侧向移动、单腿跑跳、行走、单手挥动和双手挥动,每种动作由 9 个人执行。本文采用了交叉验证的方法,随机选择 5 个人的动作数据作为训练集,剩余的动作数据为测试集。

网络的训练采用 BP 算法,网络的学习率、批量大小和迭代次数均为经验值。其中,学习率的取值范围为(0~1),一般设置为 0.1,姜枫等人<sup>[22]</sup>已经验证了此参数值的适用性。根据 Weizmann 数据集的样本数,选择实验的批量样本数为 50。为了设定合适的迭代次数,比较了 QST-CNN 分别迭代 1 次、5 次和 10 次时的识别准确率和网络训练时间,如表 1 所示。通过实验结果可以看出,迭代 5 次和 10 次的准确率相差不大,但迭代 10 次的训练时间比迭代 5 次的多了接近 2 倍,因此将迭代次数设置为 5。

表 1 不同迭代次数的实验结果

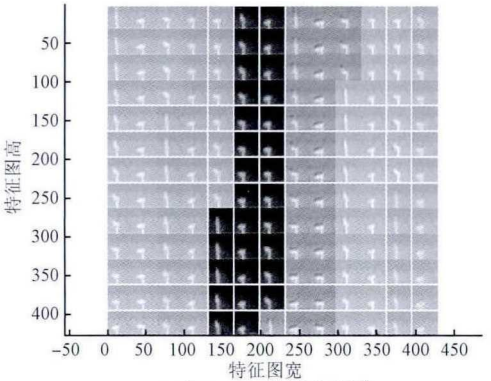
Table 1 Experiment result for different number of iterations

迭代次数	准确率/%	训练时间/min
1	69.70	2.37
5	85.34	9.85
10	85.67	25.82

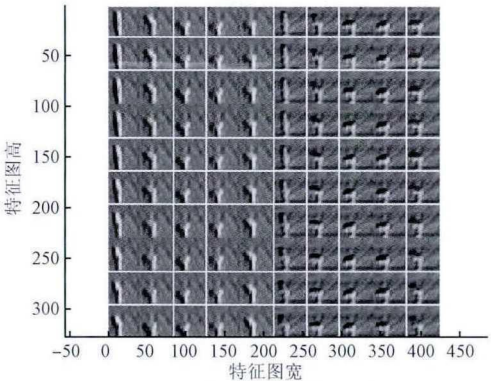
图 4 所示为 Gray-CNN 和 QST-CNN 第 1 层卷积层输出特征图的可视化结果。为了克服主观评价的不足,本文采用平均梯度、边缘强度和信息熵对图像特征提取的效果进行评价。平均梯度反映特征图的细节对比度,边缘强度反映人体轮廓特征的强度,信息熵用于评价图像的颜色特征。这里,平均梯度、边缘强度和信息熵均为图 4 中所有特征图的平均值。如表 2 所示,采用 QST-CNN 提取的特征图的平均梯度、边缘强度和信



于 Gray-CNN,说明 QST-CNN 对彩色图像的特征提取有明显的优势。表 3 对比了 Weizmann 数据集上 3 种方法的平均识别准确率和标准差。这里,准确率均是 5 次实验结果的平均值。Gray-CNN 的识别准确率为 78.00%,3Channel-CNN 的识别准确率为 66.67%,而本文提出的 QST-CNN 获得了 85.34% 的准确率,说明采用四元数进行彩色图像的特征提取能够有效提高人体行为识别的准确率。另外,通过对比准确率的标准差可知,QST-CNN 的标准差最小,说明该方法的稳定性最高。



(a) Visualization of example feature diagram for Gray-CNN



(b) Visualization of example feature diagram for QST-CNN

图 4 Gray-CNN 和 QST-CNN 第 1 层卷积层输出特征图可视化

Fig.4 Visualization of example feature diagram for the first convolutional layer outputs of Gray-CNN and QST-CNN

表 2 特征图客观评价数据			
Table 2 Objective evaluation data of feature diagram			
方法	平均梯度	边缘强度	信息熵
Gray-CNN	3.54	35.08	5.51
QST-CNN	4.37	35.92	6.47

表 3 Weizmann 数据集不同方法的准确率  
Table 3 Recognition accuracy for the Weizmann dataset with different methods

方法	准确率/%	标准差
Gray-CNN	78.00	2.54
3Channel-CNN	66.67	3.87
QST-CNN	85.34	1.63

4.2 UCF sports 数据集

UCF sports 数据集包含了通过 BBC 和 ESPN 广播电视频道收集的 10 种动作,分别为跳水、打高尔夫、踢球、举重、骑马、跑步、滑板运动、鞍马、单双杠和行走。和 Weizmann 数据集中的动作相比,UCF sports 数据集中的动作更加复杂,识别更加困难。对于 UCF sports 数据集,本文构建的 QST-CNN 的具体参数为通过最近邻插值法将 400×250 的原图像缩放至 80×50,输入 80×50×7 的彩色图像序列,3 个四元数空间卷积层的卷积核尺寸分别为 7×7×3、5×5×3、5×5×3,数量分别为 3、6、12,时间卷积层的时间维度为 3,两个下采样窗口为 2×2 和 3×3,最后在全连接层获得 504 维的特征向量,并通过 Softmax 分类器进行分类。本实验选择数据集的一半数据训练网络,剩余一半数据测试网络。

图 5 所示为本文方法在 UCF sports 数据集上进行动作识别的混淆矩阵,可以看出,跳水、滑板、鞍马和行走四种动作的识别率超过了 90%,而混淆情况主要发生在打高尔夫和行走、踢球和鞍马等动作之间,其识别率最低的动作是骑马(65%),所有动作的识别率是 80.2%。从表 4 中可以看出,本文提出的方法的识别性能高于文献[21,23]提出的方法和另外两种网络模型,而且其稳定性比 Gray-CNN 和 3 Channel-CNN 模型强。为了比较 Gray-CNN、3Channel-CNN 和 QST-CNN 3 种方法的分类性能,输入易混淆的行走、打高尔夫、踢球 3 种动作的数据训练和测试网络。如图 6 所示,随机可视化每种动作的 20 组视频片段(一个片段包含 7 帧)的输出情况,前 20 组为行走样本的输出,21~40 组是打高尔夫样本的输出,最后是踢球样本的输出。通过实验结果可以看出,图 6(c)中正确动作的曲线峰值更接近于 1,而且 30~40 的混淆情况也明显减弱,说明 QST-CNN 能够提高正确动作的输出概率,有效识别数据集中的人体行为。

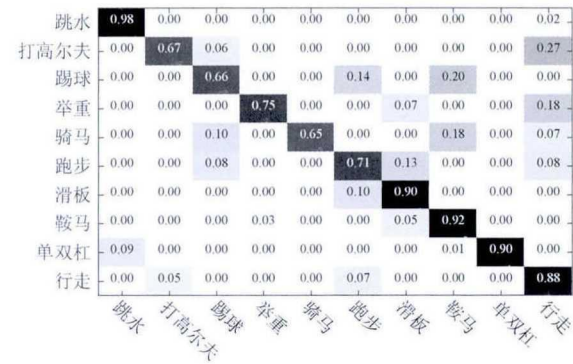
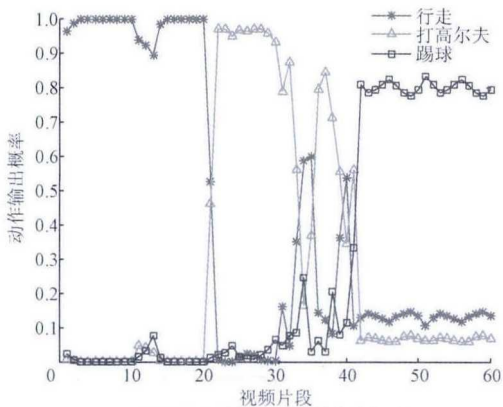


图 5 UCF sports 数据集上本文方法的混淆矩阵  
Fig. 5 Confusion matrix for the UCF sports dataset with the proposed method

表 4 UCF sports 数据集不同方法的准确率  
Table 4 Recognition accuracy for the UCF sports dataset with different methods

方法	准确率/%	标准差
文献[21]	69.2	无
文献[23]	79.2	无
Gray-CNN	65.4	2.59
3Channel-CNN	59.8	2.65
QST-CNN	80.2	1.44



(c) QST-CNN3种动作的输出概率  
(c) Output probabilities of three actions for QST-CNN

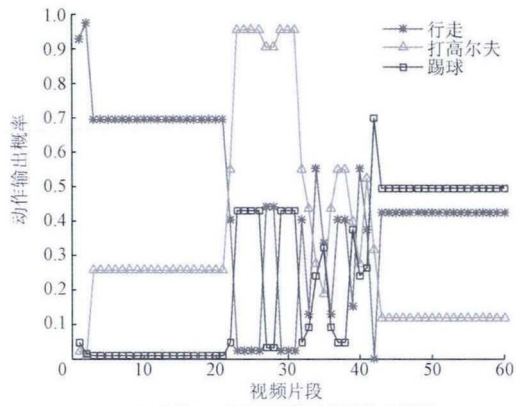
图 6 三种网络的动作输出概率  
Fig. 6 Output probabilities of three actions for the three networks

5 结 论

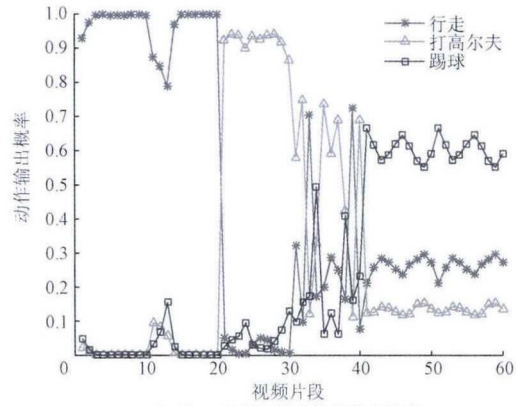
基于传统 CNN 的人体行为识别方法忽略了彩色图像三通道之间的空间关系,丢失彩色特征。本文构建了基于 QST-CNN 的人体行为识别方法,并对 Weizmann 和 UCF sports 数据集中的行为进行分类识别。实验结果证明,本文方法优于其他流行方法,在 Weizmann 和 UCF sports 数据集上的识别准确率分别为 85.34% 和 80.2%。由于本文方法增加了 CNN 的网络参数,使网络的训练时间变长,所以下一步的研究主要集中在如何在保证准确率的同时,加快网络的训练速度。

参考文献

[ 1 ] 王磊,江伟建,孙朋,等. 改进 D-S 证据理论在变电站人体跌倒检测的应用[J]. 电子测量与仪器学报, 2017,31(7):1090-1098.  
WANG L, JIANG W J, SUN P, et al. Application of improved D-S evidence theory in human fall detection of transformer substation [ J ]. Journal of Electronic Measurement and Instrumentation, 2017, 31 ( 7 ): 1090-1098.  
[ 2 ] 王恬,李庆武,刘艳,等. 利用姿势估计实现人体异常行为识别[J]. 仪器仪表学报, 2016, 37(10): 2366-2372.  
WANG T, LI Q W, LIU Y, et al. Abnormal human body behavior recognition using pose estimation[J]. Chinese Journal of Scientific Instrument, 2016, 37 ( 10 ): 2366-2372.



(a) 3 Channel-CNN3种动作的输出概率  
(a) Output probabilities of three actions for 3 Channel-CNN



(b) Gray-CNN 3种动作的输出概率  
(b) Output probabilities of three actions for Gray-CNN

- [ 3 ] CHAUDHRY R, RAVICHANDRAN A, HAGER G, et al. Histograms of oriented optical flow and Binet-Cauchy kernels on nonlinear dynamical systems for the recognition of human actions [ C ]. IEEE Conference on Computer Vision & Pattern Recognition, 2009, 94 ( 10 ): 1932-1939.
- [ 4 ] YANG X, ZHANG C, TIAN Y L. Recognizing actions using depth motion maps-based histograms of oriented gradients [ C ]. ACM International Conference on Multimedia, 2012:1057-1060.
- [ 5 ] PENG X J, QIAO Y, PENG Q. Motion boundary based sampling and 3D co-occurrence descriptors for action recognition [ J ]. Image & Vision Computing, 2014, 32(9):616-628.
- [ 6 ] JI SH W, YANG M, YU K. 3D convolutional neural networks for human action recognition [ J ]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2013, 35(1):221-31.
- [ 7 ] SIMONYAN K, ZISSERMAN A. Two-stream convolutional networks for action recognition in videos[J]. Advances in Neural Information Processing Systems, 2014, 1(4):568-576.
- [ 8 ] CHERON G, LAPTEV I, SCHMID C. P-CNN: pose-based CNN features for action recognition [ C ]. IEEE International Conference on Computer Vision, 2015: 3218-3226.
- [ 9 ] IJINA E P, CHALAVADI K M. Human action recognition using genetic algorithms and convolutional neural networks [ J ]. Pattern Recognition, 2016, 59(11):199-212.
- [ 10 ] LIU C, LIU J, HE Z, et al. Convolutional neural random fields for action recognition [ J ]. Pattern Recognition, 2016, 59(1):213-224.
- [ 11 ] HARDY A S, CHRISTIE A S. Elements of quaternions [ J ]. Science, 1969, 2(75):564.
- [ 12 ] SANGWINE S J. Fourier transforms of color images using quaternion or hypercomplex, numbers [ J ]. Electronics Letters, 1996, 32(21):1979-1980.
- [ 13 ] 殷梦娇,俞亚萍,张之江. 基于四元数模型的密集人群视频特征提取[J]. 电子测量技术, 2016, 39(7):72-75.  
YIN M J, YU Y P, ZHANG ZH J. Feature extraction of dense crowd video based on quaternion mode [ J ]. Electronic Measurement Technology, 2016, 39 ( 7 ): 72-75.
- [ 14 ] WANG J W, LE N T, LEE J S, et al. Color face image enhancement using adaptive singular value decomposition in fourier domain for face recognition [ J ]. Pattern Recognition, 2016, 57(C):31-49.
- [ 15 ] KIM K, CHALIDABHONGSE T H, HARWOOD D, et al. Real-time foreground-background segmentation using codebook model[J]. Real-Time Imaging, 2005, 11(3): 172-185.
- [ 16 ] 杨燕妮,吴向前,刘鹏. 基于帧间差分与码本模型的运动车辆检测算法[J]. 新疆大学学报:自然科学版, 2016, 33(2):203-208.  
YANG Y N, WU X Q, LIU P. Moving vehicle detection algorithm based on frame difference and codebook model [ J ]. Journal of Xinjiang University: Natural Science Edition, 2016, 33(2):203-208.
- [ 17 ] 罗子安,毛征,孟灿,等. 基于一致性特征点匹配的目标跟踪算法实现 [ J ]. 国外电子测量技术, 2016, 35(9):45-49.  
LUO Z AN, MAO ZH, MENG C, et al. Based on consensus feature point matching of object tracking algorithm[J]. Foreign Electronic Measurement Technology, 2016, 35(9):45-49.
- [ 18 ] 李响,谭南林,王天雷,等. 复杂场景下基于局部运动补偿的目标检测[J]. 仪器仪表学报, 2014, 35(7): 1555-1563.  
LI X, TAN N L, WANG T L, et al. Object detection based on local motion compensation in complex scenes[J]. Chinese Journal of Scientific Instrument, 2014, 35(7):1555-1563.
- [ 19 ] ZHANG W, LIANG D, TIAN Z, et al. Optical generation, detection and non-destructive testing applications of terahertz waves [ J ]. Instrumentation, 2016, 3(1):1-20.
- [ 20 ] GORELICK L, BLANK M, SHECHTMAN E, et al. Actions as space-time shapes [ J ]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2005, 29(12): 2247.
- [ 21 ] RODRIGUEZ M D, AHMED J, SHAH M. Action MACH a spatio-temporal maximum average correlation height filter for action recognition [ C ]. IEEE Conference on Computer Vision & Pattern Recognition, 2008:1-8.
- [ 22 ] 姜枫,张丽红. 基于随机 Dropout 卷积神经网络的人体行为识别方法研究 [ J ]. 测试技术学报, 2016, 30(1):17-22.

JANG F, ZHANG L H. Research of human action recognition method based on random Dropout convolutional neural network [J]. Journal of Test and Measurement Technology, 2016,30(1):17-22.

- [23] YEFFET L, WOLF L. Local trinary patterns for human action recognition[C]. IEEE International Conference on Computer Vision, 2009,30(2):492-497.

#### 作者简介



孟勃,2003 年于长春大学获得学士学位,2008 年于中国科学院长春光学精密机械与物理研究所获得博士学位,现为东北电力大学副教授,主要研究方向为计算机视觉、目标识别与追踪、行为识别、机器人视觉。  
E-mail:mengbo\_nannan@163.com

**Meng Bo** received her B. Sc. degree from Changchun University in 2003, received her Ph. D. degree from Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences in 2008. Now she is an associate professor in Northeast Electric Power University. Her main research interest includes computer vision, object detection and tracking,

action recognition and robot vision.



刘雪君(通讯作者),2015 年于滨州学院获得学士学位,现为东北电力大学硕士研究生,主要研究方向为深度学习、行为识别。  
E-mail:liuxuejun\_0828@163.com

**Liu Xuejun** (Corresponding author) received her B. Sc. degree from Binzhou University in 2015. Now she is a M. Sc. candidate in Northeast Electric Power University. Her main research interest includes deep learning and action recognition.



王晓霖,2016 年于河北金融学院获得学士学位,现为东北电力大学硕士研究生,主要研究方向为图像处理与模式识别、行为识别。  
E-mail:wangxiaolin\_0508@126.com

**Wang Xiaolin** received his B. Sc. degree from Hebei Finance University in 2016. Now he is a M. Sc. candidate in Northeast Electric Power University. His main research interest includes image processing and pattern recognition, and action recognition.