

基于双重注意力和3DResNet-BiLSTM行为识别方法

闫雨寒 陈天 刘忠育 刘晓文

(中国矿业大学物联网(感知矿山)研究中心 江苏徐州 221008)

(中国矿业大学信息与控制工程学院 江苏徐州 221008)

摘要 针对利用视频数据进行行为识别容易受到多种动态杂乱背景信息的影响,存在时空特征提取较为困难的问题,提出一种基于双重注意力和3DResNet-BiLSTM混合模型的行为识别方法。利用通道加权融合的方式构建卷积注意力模块并嵌入3DResNet,用于提取原始视频数据中的时空特征,同时对重要特征进行加权和重定义,结合BiLSTM和时间注意力进一步深入提取时序特征,得到自适应特征实现行为识别。利用UCF101和HMDB51公开数据集进行实验,验证了该模型的有效性。

关键词 行为识别 卷积神经网络 长短期记忆网络 注意力机制

中图分类号 TP3 文献标志码 A DOI:10.3969/j.issn.1000-386x.2023.02.030

ACTION RECOGNITION METHOD BASED ON DOUBLE ATTENTION AND 3DRESNET-BILSTM

Yan Yuhan Chen Tian Liu Zhongyu Liu Xiaowen

(IOT Perception Mine Research Center, China University of Mining and Technology, Xuzhou 221008, Jiangsu, China)

(School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221008, Jiangsu, China)

Abstract When we use video data for behavior recognition, it is easy to be affected by a variety of dynamic and messy background information, and it is difficult to extract spatial-temporal features. Aiming at these problems, this paper proposes a behavior recognition method based on double attention and 3DResNet-BiLSTM hybrid model. It used the channel weighted fusion method to construct the convolutional attention module, and embedded 3DResNet to extract the spatial-temporal features in the original video data. And it weighted and redefined the important features. BiLSTM and time attention were combined to further extract the temporal features, so that adaptive features were obtained and behavior recognition was realized. The experiments were conducted on UCF101 and HMDB51 public data sets, which verified the effectiveness of the proposed model.

Keywords Action recognition CNN LSTM Attention mechanism

0 引言

行为识别在人机交互、目标检测、视频描述等多个领域具有巨大的应用潜力和价值,然而由于视频数据中存在运动风格复杂、背景杂乱、光照环境复杂等问题^[1-2],如何准确提取人员行为的时空特征并识别是目前极具挑战性的任务。

基于视频数据的行为识别需要考虑单帧图像的静

态特征以及连续多帧图像之间的动态特征,在深度学习取得突破性进展之前,手动提取特征获取人员行为的时空特征是行为识别的通常做法。受益于深度学习方法在特征提取方面的独特优势,目前利用卷积神经网络学习视频数据中的深度特征并实现行为识别的方法主要包括双流CNN、C3D、CNN-LSTM等。Han等^[3]在ImageNet上对深度残差网络进行预训练的基础上提出了一种深层双流卷积模型用于学习动作的复杂信息,在UCF101和KTH数据集上取得了良好效

果。Yang 等^[4]利用 3D 卷积能够有效从视频数据中提取时空特征的优势,提出一种用于行为识别的非对称 3D 卷积神经网络结构,并在预处理阶段融合了 RGB 图像和光流图像的有效信息。Ullah 等^[5]将具有人类显著活动特征的视频镜头用于 FlowNet2-CNN 网络来提取时间特征,最后通过多层 LSTM 学习时间光流特征实现行为识别。

CNN 作为特征提取的有效深度模型,通常被应用于二维图像,然而 2DCNN 没有考虑多个连续帧之间的时间动态信息。3DCNN 通过在输入的另一位置进行三维卷积,能够有效捕获空间和时序特征,同时具有时间注意力模块的双向长短期记忆网络 (Bi-directional Long Short-Term Memory, BiLSTM) 能够充分获取上下文信息。在此基础上,本文提出一种基于双重注意力和 3DResNet-BiLSTM 的混合模型用于行为识别。首先将原始视频的连续帧作为 3DResNet 的输入,并利用卷积块注意力模块 (Convolutional Block Attention Module, CBAM) 聚焦空间和通道特征,抑制无关信息,为了更好地保留特征提取时的背景信息,提出一种加权池化融合系数对 CBAM 进行改进。随后利用嵌入时间注意力的 BiLSTM 进一步捕获时序特征实现行为识别。与基线模型相比,所提模型在 UCF101 和 HMDB51 数据集上的识别效果均具有较为明显的提升。

1 三维残差网络

2DCNN 通常在卷积层上进行二维卷积操作,从上一层的特征图中提取局部感受野中的特征,在应用加性偏置后通过非线性激活函数得到当前卷积层的输出特征图。在池化层中,通过下采样减少特征尺寸,增强不变性。CNN 模型通常以交替叠加多个卷积层和池化层的方式构建,利用反向传播算法实现权重参数更新。

由于 2DCNN 仅从空间维度计算特征,当对视频数据进行分析时,需要捕获多个连续视频帧中的动态时序信息,因此需要在卷积层进行 3D 卷积来计算空间和时间维度特征。3D 卷积通过一个三维卷积核对多个连续帧叠加构成的立方体进行卷积,将卷积层中的特征图连接到前一层中的多个相邻帧来获取时序信息^[6]。通常第 i 层网络中第 j 个特征图在 (x, y, z) 位置进行 3D 卷积的值可表示为:

$$v_{ij}^{xyz} = \tanh \left(b_{ij} + \sum_m \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} \sum_{r=0}^{R_i-1} \omega_{ijm}^{(pqr)} v_{(i-1)_m}^{(x+p)(y+q)(z+r)} \right) \quad (1)$$

式中: $\tanh(\cdot)$ 表示激活函数; b_{ij} 为加性偏置; $\omega_{ijm}^{(pqr)}$ 表示

前一层网络的第 m 个特征图上 (p, q, r) 位置的卷积值; P_i, Q_i, R_i 分别为卷积核的高度、宽度和深度。残差网络 (Residual Network, ResNet) 由多个残差块堆叠而成,采用快捷连接方式克服了深层网络的退化问题,在多种任务中具有出色表现,残差学习结构如图 1 所示, 3DResNet 由应用残差学习结构的 3D 卷积块构建得到。

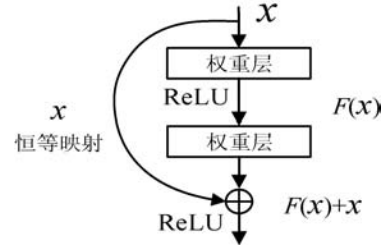


图 1 残差学习结构示意图

2 双向长短期记忆网络

递归神经网络 (Recurrent Neural Network, RNN) 通过将输入序列映射到隐藏层,再将隐藏层状态映射到输出,其计算过程可表示为:

$$h_t = \sigma(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \quad (2)$$

$$z_t = \sigma(W_{hz}h_t + b_z) \quad (3)$$

式中: $\sigma(\cdot)$ 表示激活函数; x_t 表示输入序列; $h_t \in \mathbf{R}^N$ 表示具有 N 个隐藏单元的隐藏层状态; z_t 表示 t 时刻的输出; W 和 b 分别表示权重矩阵和加性偏置。

RNN 能够将上下文信息融入到从输入到输出的映射中,但是在学习长期序列信息的过程中存在梯度消失或梯度爆炸的问题,导致其难以实现学习长期依赖。长短期记忆网络 (Long Short-Term Memory, LSTM) 是 RNN 的改进算法,通过输入门、遗忘门、输出门三个门控单元对信息流进行控制,有效克服了 RNN 存在的问题^[7-8]。BiLSTM 分别由前向 LSTM 和后向 LSTM 构成,能够有效获取上下文信息。LSTM 的单元结构如图 2 所示。

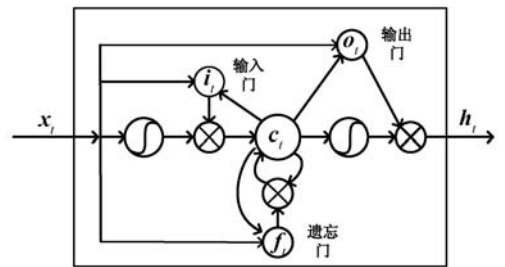


图 2 LSTM 单元结构

LSTM 单元状态更新过程如下。

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (4)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (5)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (6)$$

$$\mathbf{g}_t = \tanh(\mathbf{W}_{xc}\mathbf{x}_t + \mathbf{W}_{hc}\mathbf{h}_{t-1} + \mathbf{b}_c) \quad (7)$$

$$\mathbf{c}_t = \mathbf{c}_{t-1} \cdot \mathbf{f}_t + \mathbf{g}_t \cdot \mathbf{i}_t \quad (8)$$

$$\mathbf{h}_t = \mathbf{o}_t \cdot \tanh(\mathbf{c}_t) \quad (9)$$

式中: $\mathbf{i}_t, \mathbf{f}_t, \mathbf{o}_t, \mathbf{g}_t, \mathbf{c}_t, \mathbf{h}_t \in \mathbf{R}^N$ 分别表示输入门、遗忘门、输出门、输入调制门、记忆存储单元和隐藏单元。由于 $\mathbf{i}_t, \mathbf{f}_t, \mathbf{o}_t$ 均为 S 形的, 取值范围均在 $[0, 1]$ 之间, 因此 \mathbf{i}_t 和 \mathbf{o}_t 用于控制状态信息的输入和输出, \mathbf{f}_t 决定了上一单元中哪些状态信息会被遗忘。

3 注意力机制

3.1 卷积注意力模块

对于给定的输入三维特征图 $\mathbf{F} \in \mathbf{R}^{T \times H \times W \times C}$, H 和 W 分别表示特征图的长和宽, C 表示通道数量, T 表示时间尺度。依次沿通道和空间分别生成注意力特征权重 M_c 和 M_s , 然后对 t 时刻特征图 \mathbf{F}_t 进行重标定^[9], 其过程可表示为:

$$\mathbf{F}'_t = M_c(\mathbf{F}_t) \otimes \mathbf{F}_t \quad (10)$$

$$\mathbf{F}''_t = M_s(\mathbf{F}'_t) \otimes \mathbf{F}'_t \quad (11)$$

式中: \otimes 表示逐个元素相乘; \mathbf{F}''_t 为最终输出。对于每一个 3D 特征图重复进行该过程。CBAM 整体结构示意图如图 3 所示。

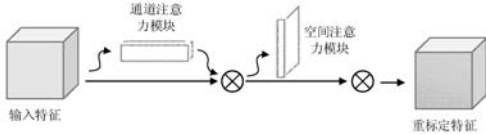


图3 卷积注意力模块示意图

通道注意力模块利用通道之间的相关性构建特征分布权重, 从而提高网络的特征表达能力。在计算通道注意力特征时, 首先对特征图进行挤压操作, 利用平均池化生成两种通道描述符用于汇聚空间特征。然后将描述符输入到单隐藏层的多层感知神经网络中, 生成通道特征权重, 最后将特征图与通道注意力权重 M_c 进行逐通道相乘得到通道注意力特征图。为了简化网络、减少参数计算量, 利用降维比 r 将隐藏层激活大小设置为 $\mathbf{R}^{1 \times 1 \times C/r}$ 。通道注意力的计算过程可表示为:

$$M_c(\mathbf{F}_t) = \sigma(\text{MLP}(\text{Avg}(\mathbf{F}_t))) = \sigma(\mathbf{W}_1(\mathbf{W}_0(\mathbf{F}_{t,\text{avg}}^c))) \quad (12)$$

式中: $\sigma(\cdot)$ 表示 S 型函数; $\mathbf{W}_0 \in \mathbf{R}^{C/r \times C}$ 和 $\mathbf{W}_1 \in \mathbf{R}^{C \times C/r}$ 分别为 MLP 网络的权重。其过程如图 4 所示。

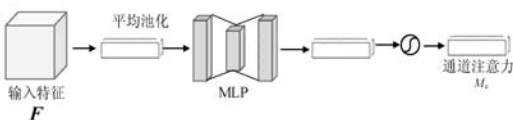


图4 通道注意力模块算法流程图示意图

由于 3D 卷积需要同时考虑连续帧之间的时空信息, 提取特征时的背景信息变得更为重要, 因此在挤压操作中通常利用平均池化保留细节信息。而最大池化能够保留纹理信息, 为了同时获得平均池化和最大池化的优势, 提出利用融合系数对挤压操作中的平均池化和最大池化进行加权融合。设最大池化融合系数为 λ , 则对通道注意力的最大池化操作应用融合系数得到:

$$M_c(\mathbf{F}_t) = \sigma(\text{MLP}(\text{Avg}(\mathbf{F}_t)) + \lambda \cdot \text{MLP}(\text{Max}(\mathbf{F}_t))) = \sigma(\mathbf{W}_1(\mathbf{W}_0(\mathbf{F}_{t,\text{avg}}^c)) + \lambda \cdot \mathbf{W}_1(\mathbf{W}_0(\mathbf{F}_{t,\text{max}}^c))) \quad (13)$$

空间注意力利用特征之间的空间关系生成空间注意力特征图, 用于聚焦特征图中的重要信息。首先沿通道维度应用平均池化进行挤压操作生成空间描述符, 通过卷积层后构建出空间注意力特征权重 M_s , 最后将特征图 \mathbf{F}_t 与空间特征权重 M_s 逐元素相乘得到空间注意力特征图。其计算过程可表示为:

$$M_s(\mathbf{F}_t) = \sigma(f^{7 \times 7}(\text{AvgPool}(\mathbf{F}_t))) = \sigma(f^{7 \times 7}(\mathbf{F}_{t,\text{avg}}^s)) \quad (14)$$

式中: $\sigma(\cdot)$ 表示 S 型函数; $f^{7 \times 7}(\cdot)$ 表示卷积核大小为 7×7 的卷积运算。

空间注意力模块算法流程如图 5 所示。

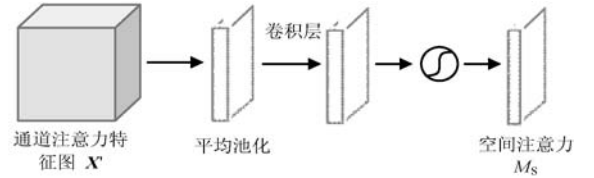


图5 空间注意力模块算法流程图示意图

3.2 时间注意力模型

软注意力机制通过模拟视觉注意力的分配过程, 最大化相关上下文编码信息, 减少无关信息的影响。对输入特征向量集合 \mathbf{X} 中的任一向量 \mathbf{x}_i , 时间注意力权重的计算过程如下:

$$\mathbf{a}_i = \frac{\exp(f(\mathbf{x}_i))}{\sum_{j=1}^r \exp(f(\mathbf{x}_j))} \quad (15)$$

式中: $f(\mathbf{x}) = \mathbf{W}^T \mathbf{x}$ 为评价函数, 用于反映特征的时序重要性, \mathbf{W} 为模型参数。注意力模块的输出为输入序列的加权和, 利用注意力权重融合特征向量得到具有时间关注度的输出特征 \mathbf{X}_a 。

$$\mathbf{X}_a = \sum_i \mathbf{a}_i \mathbf{x}_i \quad (16)$$

4 行为识别模型构建

本文提出一种基于卷积和时间双重注意力的

3DRAN-BiLSTM-Attention 模型。其中,三维残差注意力网络(3D Residual Attention Network,3DRAN)通过在 ResNet 网络中的第一个卷积层和最后一个卷积层后嵌入 CBAM 构建得到,残差注意力单元结构如图 6 所示。3DRAN 中卷积层卷积核大小均为 $3 \times 3 \times 3$,同时利用具有通道加权融合的 CBAM 聚焦空间和通道特征,提高网络的学习能力。将 16 个连续帧作为 3DRAN 的输入提取视频的时空特征,作为 BiLSTM 模型输入。

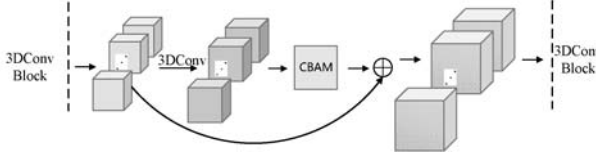


图 6 残差注意力模块结构

在解码部分,BiLSTM 分别利用前向和后向隐藏层节点获取全局上下文信息,得到输入特征序列 $H = \{h_1, h_2, \dots, h_{n-1}, h_n\}$ 。利用软注意力模块获取特征序列的权重分布,使关键特征信息具有更高的注意力权重,能够更好地聚焦重要时序特征,从而得到注意力加权特征序列 $A = \{a_1, a_2, \dots, a_{n-1}, a_n\}$ 。具体结构如图 7 所示。

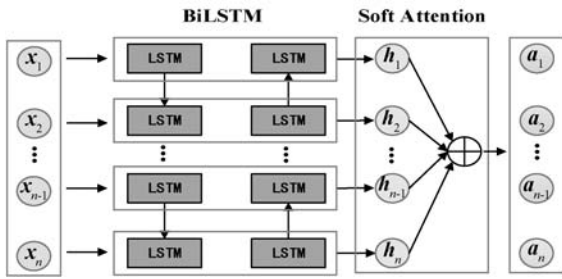


图 7 具有时间注意力的 BiLSTM 模型结构

最后,结合 3DRAN 和具有时间注意力的 BiLSTM 构建用于行为识别的混合模型,其整体结构如图 8 所示。

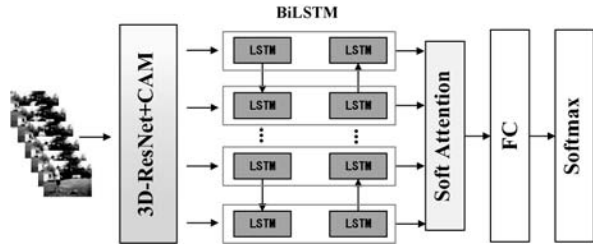


图 8 双重注意力和 3DResNet-BiLSTM 模型结构

5 实验

5.1 实验数据集

分别利用 UCF101 和 HMDB51 公共数据集进行实验。UCF101 从 YouTube 上收集得到的真实动作数据

集,共包含 101 个类别的 13 320 个短视频。HMDB51 中的视频来自电影片段和一些在线视频网站,共包含 55 个类别 6 849 个短视频。两个数据集的 70% 用于训练,30% 用于测试。

5.2 模型训练

本文实验运行环境为 Intel Core i7-8700K,主频为 3.70 GHz 的 6 核心 12 线程 CPU,16 GB 内存,GPU 为 GTX1080Ti。首先对训练数据集利用水平翻转、随机剪裁、亮度调整等方式进行扩充,并从视频中随机选取 16 个连续帧,如果视频长度不够则进行循环播放和选取,最后将所有视频帧的大小调整为 112×112 ,因此模型的输入尺寸为 $16 \times 3 \times 112 \times 112$,batch-size 大小为 256。为了减少模型过拟合,使用动量为 0.9 的随机梯度下降算法在 Kinetics 数据集上对 3DRAN 进行预训练,初始学习率设为 0.01,当验证损失达到饱和后将学习率除以 10。最后将预训练的 3DRAN 与具有时间注意力的 BiLSTM 结合构建得到本文的实验模型。

5.3 实验结果与分析

为了验证本文模型的有效性,分别设计了三组实验。首先第一组实验利用不同深度的嵌入卷积注意力的残差网络结构用于 3D 特征提取,对比不同深度模型的识别效果。由表 1 结果可以看出,随着残差注意力网络的加深,识别准确率呈上升趋势,表明深度网络能够提取到更多有效特征。

表 1 具有不同深度残差结构的模型识别准确率(%)

模型	UCF101	HMDB51
3DRAN18 + BiLSTM-Attention	88.7	61.3
3DRAN34 + Att-BiLSTM-Attention	89.5	62.2
3DRAN50 + Att-BiLSTM-Attention	91.3	63.3
3DRAN101 + Att-BiLSTM-Attention	92.7	64.8

卷积和时间注意力模块都是灵活的通用模块,其中卷积注意力模块能够利用通道和空间注意力模块的多种不同的组合顺序实现。第二个实验中对不同组合方式实现的注意力模型进行实验,并与无卷积和时间注意力的基础模型进行对比。为了节约计算资源和时间,均以 3DResNet18 为基础进行实验,实验结果如表 2 所示。结果表明,单通道注意力的效果要优于单个空间注意力模块,均好于基础模型。当同时使用通道和空间注意力时,通道注意力在前的顺序结构具有最好的识别效果。并且具有卷积注意力的模型在两个数据集上的识别效果与基础模型相比分别提升了 1.7 个百分点和 2.2 个百分点,说明卷积注意力模块能够有效抑

制背景运动。此外,仅使用时间注意力的模型的识别效果要略低于仅使用卷积注意力的模型,这是因为 Att-BiLSTM 要依赖于 3DResNet18 提取的特征。

表 2 不同注意力结构的模型识别准确率(%)

模型	UCF101	HMDB51
3DResNet18 + BiLSTM	85.6	57.9
3DResNet18-Channel-Spatia + BiLSTM	87.3	60.1
3DResNet18 + BiLSTM-Attention	86.7	59.2
3DResNet18-Channel + BiLSTM-Attention	87.6	60.7
3DResNet18-Spatial + BiLSTM-Attention	87.9	60.8
3DResNet18-Spatial-Channel + BiLSTM-Attention	88.3	61.1
3DResNet18-Channel-Spatial + BiLSTM-Attention	88.7	61.3

最后与其他流行的网络结构进行对比,由于模型训练和评估耗时较长,我们直接引用其他论文的实验结果。为了保证公平性,所有实验均采用 RGB 帧作为输入,结果如表 3 所示。可以看出,本文模型在 UCF101 和 HMDB51 数据集上的识别准确率分别达到 92.7% 和 64.8%,要明显优于标准 C3D、P3D、3DResNet101 等模型。并且具有加权融合的通道注意力模型较上述模型的识别准确率分别提升了 0.4 个百分点和 1.1 百分点,表明通过最大池化捕获特征纹理信息能够在一定程度上提升 3DResNet 的时空特征提取能力。此外,I3D 的性能较好于本文模型,因为 I3D 采用的 224×224 大小的视频帧作为输入,包含了更多的特征信息,同时也使得 I3D 参数量较大,计算复杂度较高。

表 3 不同模型的识别准确率对比(%)

模型	UCF101	HMDB51
C3D ^[10]	82.3	56.8
P3D ^[11]	88.6	—
3DResNet50 ^[12]	89.1	61.0
3DResNet101 ^[12]	88.9	61.7
I3D ^[13]	95.4	74.6
3DRAN101 + BiLSTM-Attention	92.7	64.8
3DRAN101 + BiLSTM-Attention ($\lambda = 0.5$)	93.1	65.9

6 结 语

本文提出一种基于双重注意力和 3DResNet-BiLSTM 的行为识别算法。3DResNet 借助残差结构不会显著增加计算成本,将卷积注意力模块嵌入 3DResNet 模型中,聚焦重要特征信息而抑制无关信息。并且采

用两种池化加权融合方式计算通道注意力能够在保留背景细节信息的同时获得纹理信息,有效提高了网络的时空特征学习能力。最后利用具有时间注意力的 BiLSTM 模型能够进一步学习时序信息从而实现行为识别。在 UCF101 和 HMDB51 数据集上进行的多组实验表明,在仅使用 RGB 帧作为输入的情况下,本文模型能够分别达到 93.1% 和 65.9% 的识别准确率。

参 考 文 献

- [1] 张小俊,李辰政,孙凌宇,等. 基于改进 3D 卷积神经网络的行为识别[J]. 计算机集成制造系统,2019,25(8):2000-2006.
- [2] 郑兴华,孙喜庆,吕嘉欣,等. 基于深度学习和智能规划的行为识别[J]. 电子学报,2019,47(8):1661-1668.
- [3] Han Y, Zhang P, Zhuo T, et al. Going deeper with two-stream ConvNets for action recognition in video surveillance[J]. Pattern Recognition Letters,2018,107:83-90.
- [4] Yang H, Yuan C, Li B, et al. Asymmetric 3D convolutional neural networks for action recognition[J]. Pattern Recognition, 2019,85:1-12.
- [5] Ullah A, Ahmad J, Muhammad K, et al. Action recognition in video sequences using deep Bi-Directional LSTM with CNN features[J]. IEEE Access,2018,6:1155-1166.
- [6] Ji S, Xu W, Yang M, et al. 3D convolutional neural networks for human action recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2013,35(1):221-231.
- [7] Donahue J, Hendricks L A, Guadarrama S, et al. Long-term recurrent convolutional networks for visual recognition and description[C]//Computer Vision and Pattern Recognition,2015:2625-2634.
- [8] Zhang L, Zhu G, Mei L, et al. Attention in convolutional LSTM for gesture recognition[C]//Neural Information Processing Systems,2018:1957-1966.
- [9] Woo S, Park J, Lee J, et al. CBAM: Convolutional block attention module[C]//European Conference on Computer Vision, 2018:3-19.
- [10] Tran D, Bourdev L, Fergus R, et al. Learning spatiotemporal features with 3D convolutional networks[EB]. arXiv:1412.0767, 2015.
- [11] Qiu Z, Yao T, Mei T, et al. Learning Spatio-Temporal representation with Pseudo-3D residual networks[C]//International Conference on Computer Vision,2017:5534-5542.
- [12] Hara K, Kataoka H, Satoh Y, et al. Can spatiotemporal 3D CNNs retrace the history of 2D CNNs and ImageNet[C]//Computer Vision and Pattern Recognition,2018:6546-6555.

字的可视化特征,以符合人类观察习惯的方式提取包含文字局部特征且按照固定顺序排列的可视化部件。实验表明,该算法准确性高、健壮性好,具有良好的尺度、平移和旋转不变性。并且,与 DCE 算法比较,VPE 算法得到的可视化部件较好地保留了东巴字的细节特征,从而为设计高效的东巴文字识别算法奠定基础。

参 考 文 献

- [1] 和力民. 试论东巴文化的传承[J]. 云南社会科学,2004(1):83-87.
- [2] 和金光. 纳西族东巴文化研究发展趋势[J]. 云南民族大学学报(哲学社会科学版),2007,24(1):81-84.
- [3] 戈阿干. 东巴文化揽胜[J]. 民族艺术研究,1999(2):71-80.
- [4] 方国瑜,和志武. 纳西象形文字谱[M]. 昆明:云南人民出版社,2005.
- [5] 王敬婷. 东巴文象形字研究[D]. 上海:华东师范大学,2017.
- [6] Yang Y, Kang H. Study on the automatic classification algorithms of Dongba hieroglyphic[C]//International Conference on Intelligent and Interactive Systems and Applications, 2018:328-333.
- [7] Li X, Guo H, Suo G J, et al. The design and realization of NAXI pictograph character recognition preprocessing system [C]//International Workshop on Computer Science for Environmental Engineering and EcoInformatics,2011:54-59.
- [8] Yang Y, Kang H. A novel algorithm of contour tracking and partition for Dongba hieroglyph[C]//Chinese Conference on Image and Graphics Technologies,2018:157-167.
- [9] 杨玉婷,康厚良. 东巴象形文字特征曲线提取算法研究[J]. 图学学报,2019,40(3):591-599.
- [10] 杨玉婷,康厚良,廖国富. 东巴象形文字特征曲线简化算法研究[J]. 图学学报,2019,40(4):697-703.
- [11] 杨萌,徐小力,吴国新,等. 东巴象形文字识别方法[J]. 北京信息科技大学学报(自然科学版),2014,29(3):72-76.
- [12] 王海燕,王红军,徐小力. 基于支持向量机的纳西东巴象形文字符识别[J]. 云南大学学报(自然科学版),2016,38(5):730-736.
- [13] Da M J, Zhao J Y, Suo G J, et al. Online handwritten Naxi pictograph digits recognition system using coarse grid[C]//International Workshop on Computer Science for Environmental Engineering and EcoInformatics,2011:390-396.
- [14] 杨玉婷,康厚良. 基于网格分解的东巴象形文字分类算法研究[J]. 软件导刊,2019,18(9):196-198.
- [15] Homan D D, Richards W A. Parts of recognition[J]. Cognition,1984,18(1/3):65-95.
- [16] Siddiqi K, Tresness K J, Kimia B B. Parts of visual form: Psychophysical aspects[J]. Perception,1996,25(4):399-424.
- [17] Latecki L J, Lakamper R. Convexity rule for shape decomposition based on discrete contour evolution[J]. Computer Vision and Image Understanding,1999,73(3):441-454.
- [18] Latecki L J, Lakaemper R, Wolter D. Optimal partial shape similarity[J]. Image and Vision Computing,2005,23(2):227-236.
- [19] 周瑜,刘俊涛,白翔. 形状匹配方法研究与展望[J]. 自动化学报,2012,38(6):889-910.
- [20] 杨玉婷,康厚良. 基于 CDPM 的东巴象形文字预处理算法[J]. 计算机与数字工程,2019,47(2):417-422.
- [21] 薛理,杨树文,王中辉,等. 基于八区域的简单多边形顶点凸凹性识别算法[J]. 计算机应用与软件,2018,35(1):330-333.
- [22] Yang Y, Kang H. A local feature curve extraction algorithm for Dongba hieroglyphs[C]//8th International Conference on Computer Engineering and Networks, Advances in Intelligent Systems and Computing,2019:409-416.
- [23] Latecki L J, Lakamper R. Shape similarity measure based on correspondence of visual parts[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2000,10(22):1-6.

(上接第196页)

- [13] Carreira J, Zisserman A. Quo vadis, action recognition? A new model and the kinetics dataset[C]//Computer Vision and Pattern Recognition,2017:4724-4733.
- [14] Feichtenhofer C, Pinz A, Zisserman A. Convolutional two-stream network fusion for video action recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2016.
- [15] Du W, Wang Y, Qiao Y, et al. RPN: An End-to-End recurrent pose-attention network for action recognition in videos [C]//International Conference on Computer Vision,2017:3745-3754.
- [16] Hara K, Kataoka H, Satoh Y, et al. Learning Spatio-Temporal features with 3D residual networks for action recognition [C]//International Conference on Computer Vision, 2017:3154-3160.
- [17] Du W, Wang Y, Qiao Y, et al. Recurrent Spatial-Temporal attention network for action recognition in videos[J]. IEEE Transactions on Image Processing,2018,27(3):1347-1360.
- [18] Ouyang X, Xu S, Zhang C, et al. A 3D-CNN and LSTM based multi-task learning architecture for action recognition [J]. IEEE Access,2019,7:40757-40770.
- [19] Ning J, Zhao H, Lan L, et al. A Computer-Aided detection system for the detection of lung nodules based on 3D-ResNet [J]. Applied Sciences,2019,9(24):5544.