文章编号:1671-251X(2020)07-0107-06 DOI:10.13272/j.issn.1671-251x.2019070074

煤矿探水卸杆动作识别研究

党伟超[1], 姚远[1], 白尚旺[1], 高改梅[1], 吴喆峰[2]

(1.太原科技大学 计算机科学与技术学院,山西 太原 030024;
2.精英数智科技股份有限公司,山西 太原 030006)

扫码移动阅读

摘要:针对煤矿井下探水作业监工人员通过观看视频来监控卸杆作业的方式存在效率低下且极易出错的问题,提出利用三维卷积神经网络(3DCNN)模型对探水作业中的卸杆动作进行识别。3DCNN 模型使用 3D 卷积层自动完成动作特征提取,通过 3D 池化层对运动特征进行降维,通过 Softmax 分类处理来识别卸杆动作,并使用批量归一化层提高模型的收敛速度和识别准确率。采用 3DCNN 模型对卸杆动作进行识别时,首先对数据集进行预处理,从每段视频中均匀抽取几帧图像作为某动作的代表,并降低分辨率;然后采用训练集对 3DCNN 模型进行训练,并保存训练好的权重文件;最后采用训练好的 3DCNN 模型对测试集进行测试,得出分类结果。实验结果表明,设置采样帧数为 10 帧、分辨率为 32×32、学习率为 0.000 1,3DCNN 模型对卸杆动作的识别准确率最高可达 98.86%。

关键词:煤矿防治水;煤矿探水;卸杆动作识别;三维卷积神经网络;3DCNN;批量归一化层
中图分类号:TD74 文献标志码:A

Research on unloading drill-rod action identification in coal mine water exploration

DANG Weichao[1], YAO Yuan[1], BAI Shangwang[1], GAO Gaimei[1], WU Zhefeng[2]

(1.School of Computer Science and Technology, Taiyuan University of Science and Technology, Taiyuan 030024, China; 2.Jingying Shuzhi Technology Co., Ltd., Taiyuan 030006, China)

Abstract: In view of low efficiency and error prone problems in the way that supervisors of underground water exploration operation realize monitoring of unloading drill-rod operation by watching video, 3D convolutional neural network (3DCNN) model is proposed to identify unloading drill-rod action in water exploration operation. In 3DCNN model, 3D convolution layer is used to automatically extract action features, 3D pooling layer is used to reduce dimension of motion features, softmax classification is used to identify unloading dirll-rod action, and batch normalization layer is used to improve convergence speed and identification accuracy of the model. When the 3DCNN model is used to identify unloading drill-rod action, firstly, the data set is preprocessed, and several frames of images are extracted from each video as representatives of an action, and the resolution is reduced; secondly, the training set is used to train the 3DCNN model, and the trained weight file is saved; finally, the trained 3DCNN model is used to test the test set, and the classification results are obtained. The experimental results show that when the number of sampling frames is 10, the resolution is 32×32, and the learning rate is 0.000 1, the highest recognition accuracy of the model can reach 98.86%.

Key words: coal mine water control; coal mine water exploration; unloading drill-rod action identification; 3D convolutional neural network; 3DCNN; batch normalization

---

收稿日期:2019-07-24;**修回日期**:2020-05-14;**责任编辑**:李明,郑海霞。
基金项目:山西省重点研发计划项目(201703D121042-1,201803D121048);山西省应用基础研究项目(201901D111266)。
作者简介:党伟超(1974—),男,山西运城人,副教授,博士,研究方向为智能计算、软件可靠性,E-mail:dangweichao@tyust.edu.cn。
引用格式:党伟超,姚远,白尚旺,等.煤矿探水卸杆动作识别研究[J].工矿自动化,2020,46(7):107-112.
DANG Weichao, YAO Yuan, BAI Shangwang, et al. Research on unloading drill-rod action identification in coal mine water exploration[J]. Industry and Mine Automation, 2020, 46(7): 107-112.

0 引言

据统计，2004—2013 年全国煤矿水害事故起数和死亡人数分别占事故总起数和总死亡人数的 3.1%和 8.1%[1]。对此，国家煤矿安全监察局在 2018 年发布的新版《煤矿防治水细则》中规定煤矿应采取“预测预报、有疑必探、先探后掘、先治后采”防治水原则。在探水作业中，煤矿普遍采用监工人员全程观看探水视频的方式来对卸杆作业情况进行监控，效率低下，且极易出错。针对该问题，本文提出利用深度学习技术中的三维卷积神经网络（3D Convolutional Neural Network，3DCNN）对探水作业中的卸杆动作进行识别，进而减少监工人员的工作量，同时达到减员增效、降低生产成本的目的。

传统动作识别方法一般分为动作特征表征和动作特征分类 2 个步骤[2]。动作特征表征方法包括 STIP[3]、SIFT[4-5]、HOG[6-7] 和改进稠密光流轨迹[8] 等。动作特征分类方法包括支持向量机[9] 和神经网络[10] 等。传统动作识别方法需要人工设计动作特征，随着深度学习技术的发展，出现了基于深度学习的动作识别方法。JI Shuiwang 等[11] 提出将包含时空信息的图片序列同时输入 3DCNN 进行动作识别。D. Jeff 等[12] 提出将包含动作信息的图片序列输入 CNN 网络进行特征学习，然后将学习结果输入到长短时记忆网络中实现动作识别。S. Karen 等[13] 提出了双流结构 CNN 网络，2 个 CNN 网络分别在时间与空间上提取特征，空间流通过帧图像进行识别，时间流则以光流密度为主，将 2 种特征融合后再进行分类处理。在上述 3 种基于深度学习的动作识别方法中，利用 3DCNN 模型来进行动作识别具有模型简单、训练计算量小的优点。

本文利用 3DCNN 模型来识别探水作业中的卸杆动作，并在 3DCNN 模型中使用批量归一化（Batch Normalization，BN）层，提高了识别准确率。

1 数据集准备

动作识别领域最常用的数据集是 UCF101。探水作业中的卸杆动作由一系列简单动作组合而成，属于组合动作。为保证每种类型动作分类准确、类间差别明显，需要对卸杆动作进行分解。将卸杆动作分解为 4 类，如图 1 所示。钻机拔杆：钻机将钻杆从煤岩中拔出。扳手拧杆：施工人员使用管钳将要卸下的钻杆拧松。拧杆：钻机或者施工人员（1 个或者 2 个）将钻杆完全拧下。转身放杆：施工人员双手或者单手将钻杆撤离钻机工作台并转身放到一起。

（a）钻机拔杆 （b）扳手拧杆

（c）拧杆 （d）转身放杆

图 1 卸杆动作分解

Fig. 1 Decomposition of unloading dirll-rods action

本文中视频数据来自 3 个煤矿的 11 个不同场景，筛选前视频总时长为 4.9 h。在制作数据集时兼顾不同的拍摄角度、拍摄距离、光照条件等，使模型具有较强的鲁棒性。在同一个场景中每类动作截取 10 段视频。经过筛选之后的数据集，每个类别的动作包括 110 段视频，共 440 段视频。得到的数据集中每类动作的视频数量与公共数据集 UCF101 基本一致，可用于验证模型的泛化能力。

2 3DCNN 模型

2.1 模型结构

构建的 3DCNN 模型主要由卷积层、池化层、BN 层等组成，模型结构见表 1，其中视频采样帧数为 10，卷积核尺寸为 3×3×3。

2.2 3D 卷积层

图像识别分类效果较好的 2DCNN 在结构上只考虑如何更好地收集单张图像的特征信息，并不能兼顾到各张图像间的动作变化关系。因此，JI Shuiwang 等首次提出使用可以同时学习空间及时间维度特征信息的 3DCNN 来进行动作识别[11]。3D 卷积公式为

$$v_{ij}^{xyz}=\tanh\left(b_{ij}\pm\sum_{r}\sum_{p=0}^{P_i-1}\sum_{q=0}^{Q_i-1}\sum_{o=0}^{O_i-1}\omega_{ijr}^{pqo}\vartheta_{(i-1)r}^{(x+p)(y+q)(z+o)}\right) \quad (1)$$

式中：v_{ij}^{xyz} 为第 i 层第 j 个特征映射经过卷积运算之后在视频序列中第 z 张图片位置 (x,y) 上的输出；b_{ij} 为第 i 层第 j 个特征映射的偏置；P_i，Q_i，O_i 分别为第 i 层图片的高度、宽度和所在视频序列中的索引；ω_{ijr}^{pqo} 为第 i 层第 j 个特征映射的第 r 张图片中 (p,q,o) 位置的权重；$\vartheta_{(i-1)r}^{(x+p)(y+q)(z+o)}$ 为第 i 层第 j 个特征映射的第 r 张图片中 (p,q,o) 位置的输入。

3D 卷积过程中，卷积核以固定步长对视频序列进行卷积操作。本文模型在第 1、2 次和第 3、4 次进行卷积操作时分别使用 32 个和 64 个不同的 3D 卷积核进行卷积操作，以获得视频中更多的特征值。

表 1　3DCNN 模型结构

Table 1　3DCNN model structure

层次	层(类型 & 核尺寸)	输出
第 1 层	输入层 InputLayer	(None,32,32,10,3)
第 2 层	3D 卷积层 Conv3d_1(3,3,3)	(None,32,32,10,32)
第 3 层	BN 层 BatchNormalization_1	(None,32,32,10,32)
第 4 层	3D 卷积层 Conv3d_2(3,3,3)	(None,32,32,10,32)
第 5 层	BN 层 BatchNormalization_2	(None,32,32,10,32)
第 6 层	3D 池化层 MaxPooling3d_1(3,3,3)	(None,11,11,4,32)
第 7 层	防过拟合层 Dropout_1(0.25)	(None,11,11,4,32)
第 8 层	3D 卷积层 Conv3d_3(3,3,3)	(None,11,11,4,64)
第 9 层	BN 层 BatchNormalization_3	(None,11,11,4,64)
第 10 层	3D 卷积层 Conv3d_4(3,3,3)	(None,11,11,4,64)
第 11 层	BN 层 BatchNormalization_4	(None,11,11,4,64)
第 12 层	3D 池化层 MaxPooling3d_2(3,3,3)	(None,4,4,2,64)
第 13 层	防过拟合层 Dropout_2(0.25)	(None,4,4,2,64)
第 14 层	一维化层 Flatten_1	(None,2048)
第 15 层	全连接层 Dense_1	(None,512)
第 16 层	BN 层 BatchNormalization_5	(None,512)
第 17 层	防过拟合层 Dropout_3(0.25)	(None,512)
第 18 层	全连接层 Dense_2	(None,4)

2.3　3D 池化层

池化层的作用是将特征向量聚合和降维,以达到降低网络复杂度和减少训练计算量的目的。虽然输出的特征映射数量与输入的特征映射数量一致,但每个特征映射中的特征维数成倍减少。对于处理多张图片数据来说,池化层同样需要扩展到三维。如果使用最大池化,那么在位置(m,n,l)的输出为

$$Y_{mnl}=\max_{0\leqslant\delta\leqslant S_1,0\leqslant\rho\leqslant S_2,0\leqslant\varpi\leqslant S_3}(x_{m\times s+\delta,n\times t+\rho,l\times w+\varpi}) \quad (2)$$

式中:δ,ρ,ϖ 分别为最大值运算中的迭代算子;S_1,S_2,S_3 分别为视频序列中图片的高、宽和所在视频序列的长度;$x_{m\times s+\delta,n\times t+\rho,l\times w+\varpi}$为池化区域内数值的集合;$s,t,w$ 分别为池化层的长、宽、高。

本文模型使用了 2 个最大池化层对特征映射进行降维处理。

2.4　BN 层

BN 层主要用于降低梯度对网络中参数值的依赖性,可使网络以较大的学习率进行训练,还可对每层的输出进行归一化[14],并减少 Dropout 层的使用。主要思路是对每个神经元引入 2 个可学习参数,其前向传导过程如图 2 所示。其中 d 和 $\hat{d}$ 分别为变换前后的输入数据;μ_B 为小批次(mini-batch)的均值;σ_B 为 mini-batch 的标准差;ε 为保证分母大于零引入的一个非常小的常数;γ,β 为需要学习的参数。

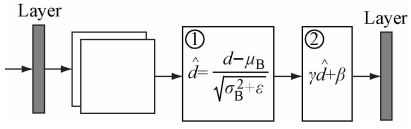

图 2　BN 层前向传导过程

Fig. 2　Forward conduction process of batch normalization layer

BN 层实现方法是在每层输入时,插入归一化层,将输入数据归一化为均值是 0、方差是 1 的标准正态分布。图 2 中阶段①实现输入数据的归一化,由于层与层之间进行归一化处理后会改变数据分布,破坏学习特征,所以,利用阶段②实现变换重构,以恢复该层学习特征。

训练阶段用每个 mini-batch 中的方差和均值来代替整个训练集的方差和均值,作为对整体的估计。测试阶段,均值和方差不是针对某个批次(batch),而是针对整个数据集。因此,在训练中除了正常的前向传播和反向求导外,还需记录每个 batch 的均值和方差,以便训练完成后按式(3)与式(4)计算整体的均值和方差。

$$E[d]=E_B[\mu_B] \quad (3)$$

$$\mathrm{Var}[d]=\frac{f}{f-1}E_B[\sigma_B^2] \quad (4)$$

式中:$E[d]$为训练集的均值;$E_B[\mu_B]$为所有 mini-batch 的均值;$\mathrm{Var}[d]$为训练集的方差;f 为 mini-batch 的数量;$E_B[\sigma_B^2]$为 mini-batch 标准差平方的均值。

2.5 分类层

基于 3DCNN 的动作识别方法在 UCF101 数据集上进行分类，种类多达 101 种。煤矿探水作业动作识别视为 4 分类问题，将 3DCNN 模型全连接层节点减少为 4 个，同时使用 Softmax 函数作为分类层的激活函数[15]，确定探水作业卸杆动作最终的识别结果。

2.6 模型训练和识别

首先，对数据集进行预处理，从每段视频中均匀抽取几帧图像作为此动作的代表，并将这些图片进行一定程度的缩小，处理为相同尺寸，以降低其分辨率。然后，对视频进行训练，并保存训练好的权重文件。最后，将权重值导入 3DCNN 模型对测试集进行测试，得出视频的分类结果和此次训练的准确率。3DCNN 模型训练和识别过程如图 3 所示。

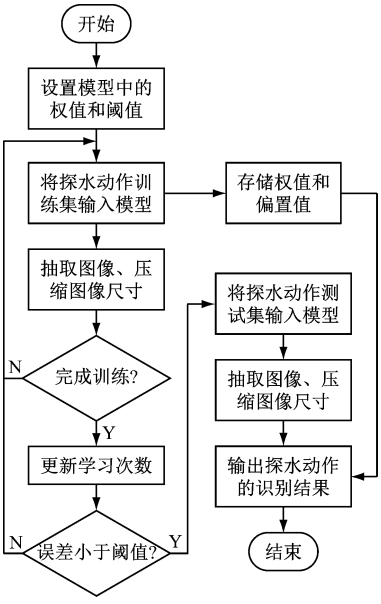

图 3 3DCNN 模型训练和识别过程

Fig. 3 3DCNN model training and recognition process

3 实验结果及分析

实验中采用了基于 Python 的深度学习框架 Keras(以 tensorflow 为后端)构建 3DCNN 模型，同时使用 GPU(型号为 NVIDIA GeForce RTX 2080Ti)对模型进行加速训练。将数据集按照每个动作类型 4∶1的比例组成训练集和测试集，其中训练集为 352 个，测试集为 88 个。

在实验过程中，采用 Xavier 方法对模型中权重和偏置参数进行初始化。假设 3DCNN 中某层的输入维度为 a，输出维度为 c，则该层的参数将均匀分布在 $\left[-\sqrt{\frac{6}{a+c}},\sqrt{\frac{6}{a+c}}\right]$ 之间。该方法可以使每层输出的方差几乎相等，从而使模型收敛速度更快。

3.1 不同采样帧数的准确率分析

在 4 类动作中，因作业人员工作习惯、钻机型号等不同，视频长度区别较大，采样帧数越多，提供给模型的信息越详细，但也增加了模型的复杂度和训练时长。为此，从每个视频中分别抽取 5，10，15，20，25，30 帧图像，再利用 OpenCV 将每帧图像尺寸转换为 32×32。在进行模型训练时，设置 batch-size 为 20，模型迭代训练 500 次，结果见表 2。

表 2 不同采样帧数的准确率

Table 2 Accuracy of different sample frames

采样帧数	准确率/%	参数量/10^5	训练时长/s
5	94.32	7.26	206
10	98.86	12.5	305
15	96.59	12.5	411
20	98.86	17.7	511
25	97.73	17.7	616
30	98.86	23.0	707

从表 2 可看出，采样帧数为 5 时准确率最低，随着采样帧数的增加，准确率会有一定程度的提高，而模型参数量和训练时长都大幅增加。对于该模型来说，当采样帧数为 10 时，既能保证准确率，又能兼顾较少的模型参数量和训练时长。

3.2 不同采样图像分辨率的准确率分析

在确定采样帧数的情况下，每帧图像的分辨率越高，提供给模型的信息越丰富，但也会增加运算的开销和模型复杂度。为此，从每段视频中抽取10 帧图像，利用 OpenCV 将这些图像尺寸处理为 16×16，32×32，64×64，128×128，并分别进行训练。设置 batch-size 为 20，模型迭代训练 500 次，结果见表 3。

表 3 不同分辨率的准确率

Table 3 Accuracy of different resolutions

分辨率	准确率/%	参数量/10^5	训练时长/s
16×16	89.77	4.64	167
32×32	98.86	12.5	305
64×64	97.73	44.0	989
128×128	98.86	149.47	3 679

从表 3 可看出，分辨率为 16×16 时，准确率最低；随着分辨率的提高，准确率也会提高，但也会增加模型参数量和训练时长。综合考虑准确率、模型参数量和训练时长，将图像的分辨率定为 32×32。

3.3 不同学习率的准确率分析

为对比学习率对准确率和收敛速度的影响，从每个视频中均匀抽取 10 帧图像，并将这些图像尺寸处理为 32×32。在训练时，设置 batch-size 为 20，

学习率为 0.000 1,0.000 5,0.001,模型迭代训练 500 次,结果如图 4 所示。

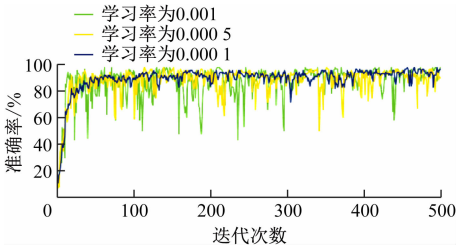

图 4　不同学习率的准确率

Fig. 4　Accuracy of different learning rates

从图 4 可看出,学习率对迭代次数的影响可以忽略不计;学习率为 0.000 1 时,模型准确率更为稳定,最高准确率为 98.86%。

3.4　加入 BN 层的准确率分析

为了验证加入 BN 层的效果,从每个视频中抽取 10 帧图像,将图像尺寸处理为 32×32。在训练时,设置 batch-size 为 20,学习率为 0.000 1,模型迭代训练 500 次,结果如图 5 所示。

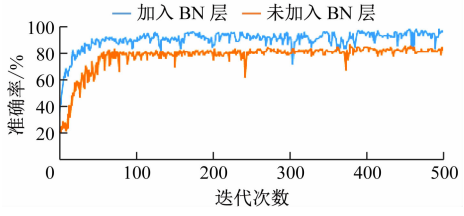

图 5　加入和未加入 BN 层的准确率

Fig. 5　Accuracy of whether to add BN layer

从图 5 可看出,在网络中加入 BN 层,模型在迭代 50 次之后,准确率基本稳定,可以认为模型已经达到最佳的拟合效果,准确率最高可达98.86%;未加入 BN 层时,准确率最高为 86.36%。

4　结语

提出一种基于 3DCNN 的煤矿探水作业卸杆动作识别方法。3DCNN 模型使用 3D 卷积层自动完成动作特征的提取,通过 3D 池化层对动作特征进行降维,通过 Softmax 分类处理实现煤矿探水作业中典型动作的识别,并使用 BN 层提高了模型收敛速度和识别准确率。实验对比了不同采样帧数、分辨率、学习率及是否加入 BN 层时模型的准确率。结果表明,设置采样帧数为 10 帧、分辨率为 32×32、学习率为0.000 1 时,模型在加入 BN 层之后,3DCNN 模型对卸杆动作的识别准确率最高可达 98.86%。

参考文献(References):

[1] 孙继平. 煤矿事故分析与煤矿大数据和物联网[J]. 工矿自动化,2015,41(3):1-5.
SUN Jiping. Accident analysis and big data and Internet of things in coal mine[J]. Industry and Mine Automation,2015,41(3):1-5.

[2] 徐光祐,曹媛媛. 动作识别与行为理解综述[J]. 中国图象图形学报,2009,14(2):189-195.
XU Guangyou, CAO Yuanyuan. Action recognition and activity understanding: a review[J]. Journal of Image and Graphics,2009,14(2):189-195.

[3] ZHU Yu,CHEN Wenbin,GUO Guodong. Evaluating spatiotemporal interest point features for depth-based action recognition[J]. Image and Vision Computing,2014,32(8):453-464.

[4] 杨世沛,陈杰,周莉,等. 一种基于 SIFT 的图像特征匹配方法[J]. 电子测量技术,2014,37(6):50-53.
YANG Shipei, CHEN Jie, ZHOU Li, et al. Image feature matching method based on SIFT [J]. Electronic Measurement Technology, 2014, 37(6): 50-53.

[5] 宋健明,张桦,高赞,等. 基于多时空特征的人体动作识别算法[J]. 光电子激光,2014,25(10):2009-2017.
SONG Jianming, ZHANG Hua, GAO Zan, et al. Human action recognition based on multi-spatio-temporal features[J]. Journal of Optoelectronics · Laser,2014,25(10):2009-2017.

[6] ZHU Yu, CHEN Wenbin, GUO Guodong. Fusing multiple features for depth-based action recognition[J]. ACM Transactions on Intelligent Systems and Technology,2015,6(2):1-20.

[7] 卜庆志,裘君,胡超. 基于 HOG 特征提取与 SVM 驾驶员注意力分散行为检测方法研究[J]. 集成技术,2019,8(4):69-75.
BU Qingzhi,QIU Jun,HU Chao. Research on driver's distracted behavior detection method based on histogram of oriented gradient feature extraction and support vector machine[J]. Journal of Integration Technology,2019,8(4):69-75.

[8] 孙玉玲. 基于改进稠密轨迹的人体行为识别方法研究[D]. 天津:天津工业大学,2017.
SUN Yuling. Research on human behavior recognition method based on improved dense trajectory[D]. Tianjin:Tianjin Polytechinc University,2017.

[9] 沈舒,吴聪,李勃,等. 基于优化 SVM 的高速公路交通事件检测[J]. 电子测量技术,2012,35(5):40-44.
SHEN Shu, WU Cong, LI Bo, et al. Freeway traffic incident detection based on an optimized SVM model [J]. Electronic Measurement Technology, 2012, 35(5):40-44.

[10] 罗志增,熊静,刘志宏. 一种基于 WPT 和 LVQ 神经网络的手部动作识别方法[J]. 模式识别与人工智能,2010,23(5):695-700.
LUO Zhizeng, XIONG Jing, LIU Zhihong. Pattern

recognition of hand motions based on WPT and LVQ[J]. Pattern Recognition and Artificial Intelligence, 2010,23(5):695-700.

[11] JI Shuiwang, XU Wei, YANG Ming, et al. 3D convolutional neural networks for human action recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 495-502.

[12] JEFFREY D, LISA A H, MARCUS R, et al. Long-term recurrent convolutional networks for visual recognition and description[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(4):677-691.

[13] KAREN S, ANDREW Z. Two-stream convolutional networks for action recognition in videos [C]// Advances in Neural Information Processing Systems, 2014:568-576.

[14] SERGEY I, CHRISTIAN S. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]//Proceedings of the 32nd International Conference on Machine Learning, 2015: 448-456.

[15] 王俊,郑彤,雷鹏,等. 基于卷积神经网络的手势动作雷达识别方法[J]. 北京航空航天大学学报, 2018, 44(6):1117-1123.
WANG Jun, ZHENG Tong, LEI Peng, et al. Hand gesture recognition method by radar based on convolutional neural network[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(6):1117-1123.