

文章编号: 1671-251X(2023)03-0039-07

DOI: 10.13272/j.issn.1671-251x.18058

基于长短期存储的聚合增强型煤矸石视频识别模型

杨军

(国家能源集团宁夏煤业有限责任公司 枣泉煤矿, 宁夏 银川 750001)

摘要: 采用煤矸石图像识别技术进行煤矸石识别会错过一些关键目标的识别。视频目标识别模型比图像目标识别模型更贴近煤矸石识别分选场景需求, 对视频数据中的煤矸石特征可以进行更广泛、更有深度的提取。但目前煤矸石视频目标识别技术未考虑视频帧重复性、帧间相似性、关键帧偶然性对模型性能的影响。针对上述问题, 提出了一种基于长短期存储(LSS)的聚合增强型煤矸石视频识别模型。首先, 采用关键帧与非关键帧对海量信息进行初筛。对煤矸石视频帧序列进行多帧聚合, 通过时空关系网络(TRN)将关键帧与相邻帧特征信息相聚合, 建立长期视频帧和短期视频帧, 在不丢失关键特征信息的同时减少模型计算量。然后, 采用语义相似性权重、可学习权重和感兴趣区域(ROI)相似性权重融合的注意力机制, 对长期视频帧、短期视频帧与关键帧之间的特征进行权重再分配。最后, 设计用于存储增强的LSS模块, 对长期视频帧与短期视频帧进行有效特征存储, 并在关键帧识别时加以融合, 增强关键帧特征的表征能力, 以实现煤矸石识别。基于枣泉选煤厂自建煤矸石视频数据集对该模型进行实验验证, 结果表明: 相较于记忆增强全局-局部聚合(MEGA)网络、基于流引导的特征聚合视频目标检测(FGFA)、关系蒸馏网络(RDN)、视频识别的深度特征流(DFF)模型, 基于LSS的聚合增强型煤矸石视频识别模型的平均精度均值优于其他模型, 为77.12%; 模型视频目标运动速度与识别精度呈负相关, 基于LSS的聚合增强型煤矸石视频识别模型在慢速运动的目标检测上识别精度最高达83.82%。

关键词: 煤矸石视频识别; 视频帧多帧聚合; 关键帧; 非关键帧; 注意力机制; 时空关系网络; 长短期存储

中图分类号: TD67

文献标志码: A

Aggregation enhanced coal-gangue video recognition model based on long and short-term storage

YANG Jun

(Zaoquan Colliery, CHN Energy Ningxia Coal Industry Co., Ltd., Yinchuan 750001, China)

Abstract: Some key targets will be missed when using coal-gangue image recognition technology to recognize coal-gangue. Compared with the image target recognition model, the video target recognition model is closer to the requirements of the coal-gangue recognition and separation scene. The coal-gangue features in the video data can be extracted more widely and deeply. However, the influence of frame repetition, frame similarity and contingency of key frame on the model performance is not considered in the current coal-gangue video target recognition technology. In order to solve the above problems, this paper proposes an aggregation enhanced coal-gangue video recognition model based on long and short-term storage (LSS) model. Firstly, the key frames and non-key frames are used to screen the massive information. Multi-frame aggregation is carried out on the video frame sequence of the coal-gangue. The feature information of the key frame and the adjacent frame is aggregated through temporal relation networks (TRN), and a long-term video frame and a short-term video frame are established. The calculation amount of the model is reduced while the key feature information is not lost.

收稿日期: 2022-11-22; 修回日期: 2023-02-20; 责任编辑: 王晖, 郑海霞。

基金项目: 陕西省秦创原“科学家+工程师”队伍建设项目(2022KXJ-38)。

作者简介: 杨军(1982—), 男, 宁夏平罗人, 工程师, 主要从事煤矿智能化技术研究工作, E-mail: 273364857@qq.com。

引用格式: 杨军. 基于长短期存储的聚合增强型煤矸石视频识别模型[J]. 工矿自动化, 2023, 49(3): 39-44, 62.

YANG Jun. Aggregation enhanced coal-gangue video recognition model based on long and short-term storage [J]. Journal of Mine Automation, 2023, 49(3): 39-44, 62.



扫码移动阅读

Secondly, the feature weights among the long-term video frames, the short-term video frames and the keyframes are reallocated by using an attention mechanism that integrates semantic similarity weights, learnable weights and region of interest (ROI) similarity weights. Finally, the LSS module is designed to store the effective features of long-term video frames and short-term video frames. The module fuses them in the key frame recognition to enhance the characterization capability of the key frame features, so as to realize coal-gangue recognition. The model is tested based on the coal-gangue video data set in Zaoquan Coal Preparation Plant. The results show that in comparison with the memory enhanced global-local aggregation (MEGA) network, the flow-guided feature aggregation for video object detection (FGFA), the relation distillation networks (RDN) and deep feature flow for video recognition (DFF) model for video recognition, the mean average precision of the aggregation enhanced coal-gangue video recognition model based on LSS is 77.12 % and better than that of other models. The recognition precision of the modes is negatively correlated with the moving speed of the target in the video. The recognition precision of the model in this paper is 83.82% for the slow-moving target detection, and the performance is the best.

Key words: video recognition of coal-gangue; multi-frame aggregation of video frames; key frame; non-key frame; attention mechanism; temporal relation networks; long and short-term storage

0 引言

在煤炭开采过程中,不可避免地会混入大量热值低、含碳量低的矸石,影响煤炭质量。目前常用的煤矸石分选方法包括手动选矸、动筛跳汰分选、重介质分选和选择性破碎等。这些方法存在效率低、配置困难、工艺复杂等局限性^[1,2]。亟需一种智能程度高、识别性能好的煤矸石识别方法。

随着计算机视觉技术的发展,目标识别技术开始应用到煤矸石识别领域。基于计算机视觉的目标识别技术分为图像目标识别和视频目标识别。在煤矸石图像识别方面,文献[3]采用改进的卷积神经网络算法对煤和矸石图像进行数据扩充,并结合迁移学习方法对煤矸石图像进行处理,提升了模型识别精度;文献[4]基于改进的 tiny YOLO V3 算法对煤矸石进行识别,通过 SPPNet、SE 模块和扩张卷积的组合,提升了模型对煤矸石图像的识别精度。煤矸石图像识别技术对目标的识别精度较高,但通过采集图像进行煤矸石识别的方式会错过一些关键目标的识别,采集到的低质量图像也会影响系统整体的识别精度。视频目标识别模型比图像目标识别模型更贴近煤矸石识别分选场景需求,对视频数据中的煤矸石特征可以进行更广泛、更有深度的提取。在煤矸石视频目标识别方面,文献[5]提出基于光流的特征聚合方法,增强了当前帧特征表示能力,提高了模型泛化性,对煤矸石的识别精度达 73.4%;文献[6]提出一种基于视频处理的煤矸石识别方法,对视频中的原煤图像进行分类,在多工况识别实验中取得了较好的识别效果;文献[7]采用改进的高斯混合模

型实现煤矸石的识别,并使用粒子群算法对高斯混合模型参数进行优化与自整定,对矸石的检测准确率达 95.83%。上述研究通过巧妙的模型和实验设计获得了较好的研究结果,但未考虑视频帧重复性、帧间相似性、关键帧偶然性对模型性能的影响。

针对上述问题,本文提出了一种基于长短期存储(Long and Short-term Storage, LSS)的聚合增强型煤矸石视频识别模型。该模型设计 LSS 模块对长短期视频帧特征进行存储,以增强关键帧特征表征能力,结合多帧聚合、注意力机制减少冗余信息,增加重要特征权重,提升模型的识别性能。

1 关键帧选取

视频目标识别与图像目标识别相比,需要考虑的因素更多、更复杂^[8-10]。视频数据中大量的重复信息和丰富的时空信息为其识别带来了更大的挑战^[11]。因此,本文采用关键帧与非关键帧对海量信息进行初筛。

1.1 关键帧与非关键帧

将视频切片中每个视频帧按时间顺序叠加在一起,这时视频帧在图像坐标系中的特征矢量呈现出一个轨迹,与轨迹中特征值相对应的帧即为关键帧,其余为非关键帧^[12-13]。关键帧的选取和处理对于煤矸石视频识别有着重要作用。非关键帧通常为关键帧的检测提供辅助信息。煤矸石视频中的关键帧与非关键帧示例如图 1 所示。

1.2 关键帧选取框架

本文基于视频特征的得分矩阵设计了关键帧选取框架,如图 2 所示。将视频切片作为关键帧框架

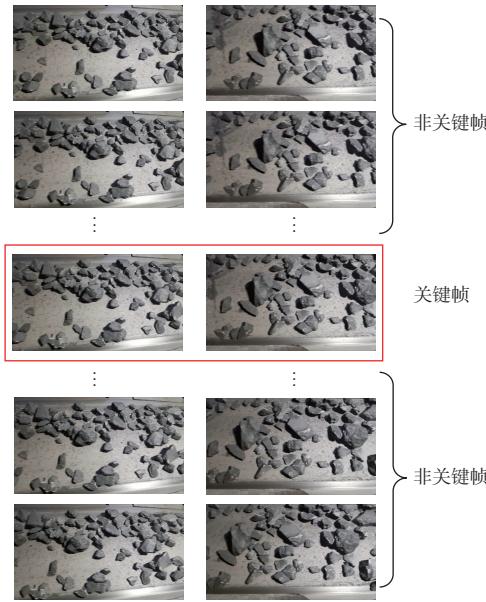


图 1 煤矸石视频中的关键帧与非关键帧

Fig. 1 Key frames and non-key frames in coal-gangue video

的输入,通过卷积层对视频切片进行深度特征提取,再将提取的特征值输入得分矩阵中,根据得分矩阵的输出结果来判断视频切片中的关键帧。

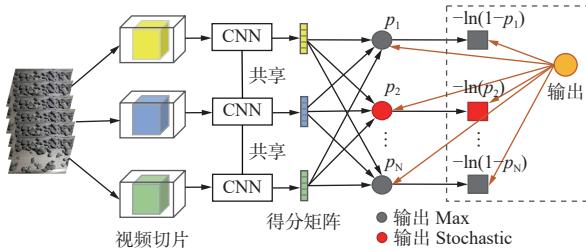


图 2 关键帧选取框架

Fig. 2 Key frame selection frame

不同视频切片之间共享深度特征,通过 Sigmoid 激活函数后得到一个得分矩阵:

$$\mathbf{S} = \{s_{k,n}\} \quad k = 1, 2, \dots, K; n = 1, 2, \dots, N \quad (1)$$

式中: $s_{k,n}$ 为第 n 个分类器在第 k 个视频切片处的响应; K 为视频切片总数; N 为分类器总数。

关键帧选取框架根据得分矩阵的响应进行更新,从得分矩阵的响应中得到含关键帧的概率:

$$p_n = \begin{cases} \text{Max}(s_{:,n}) & n \neq Y \\ \text{Stochastic}(s_{:,n}) & n = Y \end{cases} \quad (2)$$

式中: $s_{:,n}$ 为第 n 个分类器在所有视频切片处的响应; Y 为视频类别数; Stochastic 为随机选取函数。

2 模型设计

针对视频帧重复性、帧间相似性、关键帧偶然性等问题,本文提出的基于 LSS 的聚合增强型煤矸石视频识别模型如图 3 所示。该模型主要包括多帧聚合、注意力机制和 LSS 模块 3 个部分。首先,对煤矸

石视频帧序列进行多帧聚合,通过时空关系网络(Temporal Relation Networks, TRN)^[14]将关键帧与相邻帧特征信息相聚合,建立长期视频帧和短期视频帧;然后,采用语义相似性权重、可学习权重和感兴趣区域(Region of Interest, ROI)^[15]相似性权重融合的注意力机制对长期视频帧、短期视频帧与关键帧之间的特征进行权重再分配;最后,设计用于存储增强的 LSS 模块,对长期视频帧与短期视频帧进行有效特征存储,并在关键帧识别时加以融合,增强关键帧特征的表征能力^[16],以实现煤矸石的识别。

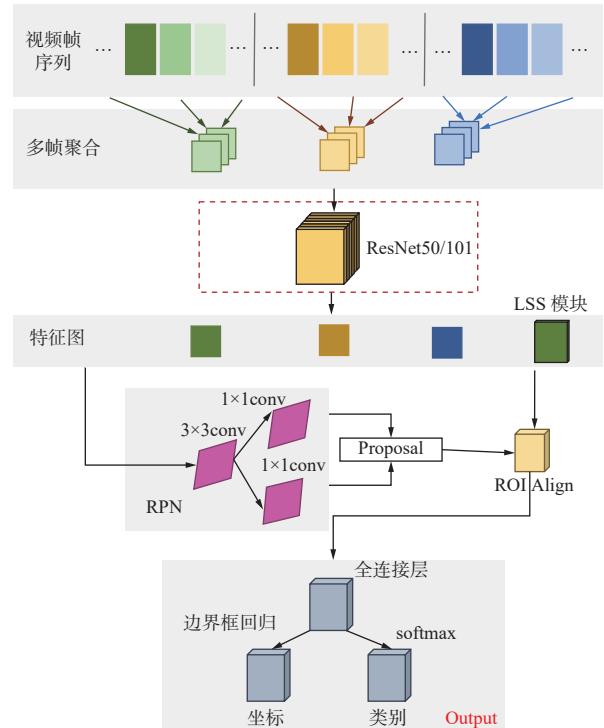


图 3 基于 LSS 的聚合增强型煤矸石视频识别模型

Fig. 3 Aggregation enhanced coal-gangue video recognition model based on LSS

2.1 多帧聚合

多帧聚合操作可以密集计算每帧特征,同时将视频帧特征进行不同尺度的加权聚合。煤矸石视频帧具有小目标多、帧间相似度高、画面冗余度较高等特性,逐帧计算耗费算力且效率不高。因此,采用多帧聚合对煤矸石视频帧进行初步处理,以减少煤矸石视频帧中小目标图像信息的丢失。该方法通过调整参数将多个视频帧的重要特征进行聚合,在不丢失视频帧特征信息的同时,减轻模型的计算负担和处理工作量。

TRN 方法简单有效,可以在获得良好特征融合效果的同时减少计算量,平衡模型的识别精度与速度,因此采用 TRN 进行多尺度视频帧聚合,融合过程如图 4 所示。首先,TRN 每次处理 15 帧视频帧作

为输入, 经过卷积之后得到 15 张特征图; 其次, 选取前 3 帧、中间 5 帧和后 7 帧这 3 个尺度的视频帧分

别进行特征融合; 最后, 将上述 3 个尺度的特征融合结果相加得到最终特征。

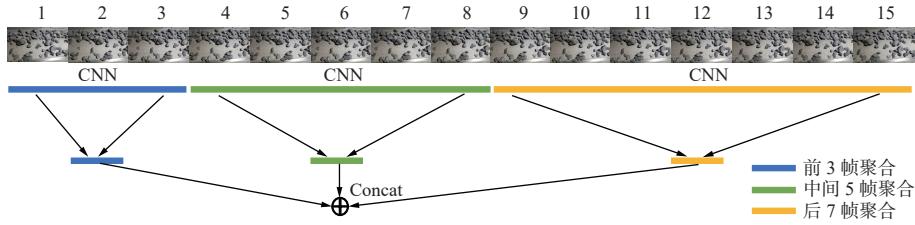


图 4 TRN 特征融合过程

Fig. 4 The TRN feature fusion process

2.2 注意力机制

基于 LSS 的聚合增强型煤矸石视频识别模型隶属于两阶段识别模型, 先通过区域提议网络(Region Proposal Network, RPN)^[17] 提取输入煤矸石视频帧序列的特征, 并生成候选框, 再进行识别目标的分类与回归。在相同环境、相同拍摄设备、固定拍摄角度及相同传送带的外部因素下, 所采集视频呈现出视频帧相似性较高、有效视频帧特征较少的特性。因此, 采用注意力机制将长期视频帧特征、短期视频帧特征与关键帧特征按照语义相似度权重、可学习权重和 ROI^[18] 相似度权重进行重新融合, 降低图像中相似度较高区域的语义相似性权重, 增加有效特征区域的权重, 以提高融合后特征图的有效性。

注意力机制计算原理如图 5 所示。首先, 采用多头注意力机制对长期视频帧特征、短期视频帧特征和关键帧特征同时进行注意力池化。然后, 将这 3 组注意力池化的输出拼接在一起, 通过其他可学习的线性投影进行变换, 以获得最终的融合特征。最后, 将这些视频帧特征进行存储, 在后续关键帧处理时加以融合, 可有效增强关键帧表征能力。

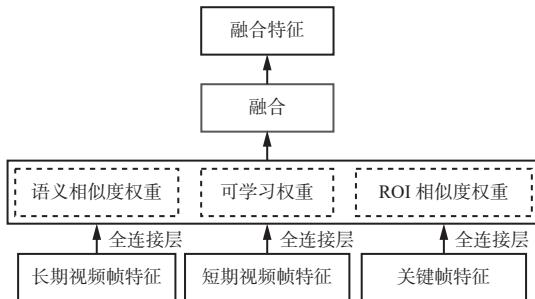


图 5 注意力机制计算原理

Fig. 5 The tational principle of attention mechanism

注意力机制通过参数调节长短期视频帧对于关键帧的权重, 形成一个权重渐小的循环机制。

$$h_i = f(W_i^{(q)} q, W_i^{(e)} e, W_i^{(v)} v) \in \mathbf{R}^{P_v} \quad (3)$$

式中: h_i 为第 i 个注意力头; f 为注意力池化函数; $W_i^{(q)}$, $W_i^{(e)}$, $W_i^{(v)}$ 分别为查询 q 、键 e 、值 v 对应的权重参

数; P_v 为值 v 的维度。

2.3 LSS 模块

在模型预测过程中, 关键帧对预测结果的影响较大, 但关键帧图像具有偶发性, 可能会遗漏掉重要特征从而影响识别效果^[19]。因此, 本文设计 LSS 模块将非关键帧中长期视频帧、短期视频帧特征进行存储, 在预测时增强关键帧特征, 提升模型整体的识别精度。

通过 LSS 模块将煤矸石视频光线变化帧、遮挡前后帧、模糊前后帧的图像特征进行存储, 在关键帧预测中与其图像特征相融合。LSS 模块在设计上受长短期记忆(Long Short-Term Memory, LSTM)算法^[20] 门控思想的启发, 在特征数据输入、输出部分采用门控处理, LSS 模块更注重数据的存储。

通过固定时长采样选取长短期视频帧, 经过多此实验验证可知, 当长期视频帧选取关键帧之前第 18~20 帧图像, 短期视频帧选取关键帧之前 3~5 帧图像时模型性能最优。随着关键帧的遍历, 长短期视频帧也相应发生改变。

LSS 模块包括继承、存储、输出 3 个部分, 如图 6 所示。继承主要是将上一时刻的输出数据进行承接, 作为当前时刻的输入数据来源; 存储是对输入的数据进行有选择的处理, 摒弃无效数据; 输出是将保存数据与上一时刻的输入数据进行整合处理, 处理结果作为当前时刻输出数据。

1) 继承。是每个 LSS 单元的输入, 将上一时刻 $t-1$ 的输出状态继承下来, 作为当前时刻 t 的输入。

$$g_t = \sigma(\omega_g [C_{t-1}, u_{t-1}, x_t] + b_g) \quad (4)$$

式中: g_t 为继承的输出信息; σ 为 Sigmoid 函数; ω_g 为当前输入的权重; C_{t-1} 为上一时刻的继承信息; u_{t-1} 为上一时刻的输出信息; x_t 为当前时刻输入的新信息; b_g 为当前时刻输入的偏置。

2) 存储。先采用 \tanh 函数对 C_{t-1} 与 x_t 中的有效信息 \tilde{C}_t 进行提取, 再通过 Sigmoid 函数控制存储信息 I_t 。

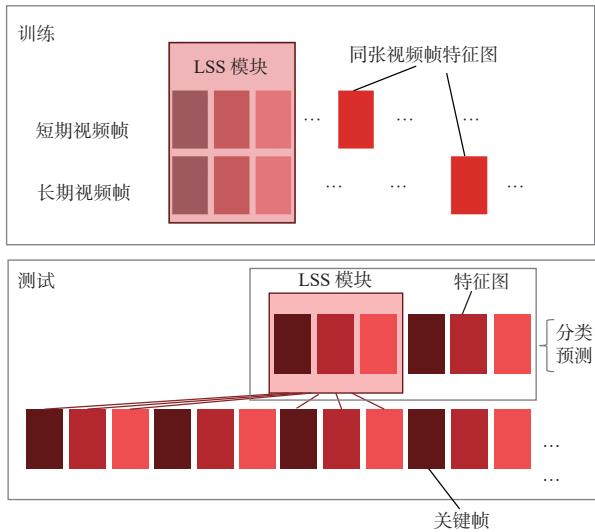


图 6 LSS 模块设计原理

Fig. 6 The LSS module design principle

$$\tilde{C}_t = \tanh(\omega_c [u_{t-1}, x_t] + b_c) \quad (5)$$

$$I_t = \sigma(\omega_l [u_{t-1}, x_t] + b_l) \quad (6)$$

式中: ω_c 为有效信息中输入的权重; b_c 为有效信息中输入的偏置; ω_l 为存储信息中输入的权重; b_l 为存储信息中输入的偏置。

3) 输出。输出层需将当前时刻存储信息 C_t 与上一时刻输出信息 u_{t-1} 进行整合。

$$r_t = \sigma(\omega_r [C_{t-1}, u_{t-1}, x_r] + b_r) \quad (7)$$

$$C_t = g_t * C_{t-1} + i_t * \tilde{C}_t \quad (8)$$

$$u_t = r_t * \tanh C_t \quad (9)$$

式中: r_t 为当前时刻的整合信息; ω_r 为整合信息中输入的权重; b_r 为整合信息中输入的偏置; u_t 为当前时刻的输出信息; C_t 为当前时刻的继承信息。

3 实验结果及分析

3.1 实验环境及数据选择

在显存为 32 GiB 的 NVIDIA GeForce RTX 2080 Ti 硬件环境中进行实验设计、实验验证及数据分析。

实验用煤研石自建视频数据集(VID 数据集和 DET 数据集)来自国家能源集团宁夏煤业有限责任公司枣泉选煤厂, 该数据集来源于真实煤研石分选场景, 涵盖了不同大小、形态的煤研石视频切片。该数据集共 3 028 个视频切片, 包含视频帧 90 840 张, 其中 80% 作为训练集和验证集, 20% 用作测试集。

根据加载训练产生的预测模型进行视频数据的检测。根据目标运动速度的不同将其分为快、中、慢 3 个等级, 分别计算检测精度。

3.2 结果分析

采用平均精度均值 (mean Average Precision,

mAP) 和损失值 2 个指标来评估基于 LSS 的聚合增强型煤研石视频识别模型的性能。

记忆增强全局-局部聚合 (Memory Enhanced Global-local Aggregation, MEGA) 网络模型参考了全局语义特征和局部位置特征来实现关键帧目标识别, 以获得更加全面的特征信息。基于流引导的特征聚合视频目标检测 (Flow-Guided Feature Aggregation for video object detection, FGFA) 模型主要根据流导向, 将光流导向的特征用在光流聚合道附近帧的特征图上。关系蒸馏网络 (Relation Distillation Networks, RDN)^[21] 模型通过聚合目标特征, 传播目标关系以增强目标特征。视频识别的深度特征流 (Deep Feature Flow for video recognition, DFF)^[22] 模型主要采用光流法, 通过流场将目标深度特征图传播到其他帧, 显著提高了识别速度。在视频目标识别中 MEGA, FGFA, RDN, DFF 模型的识别性能相对显著。

为了验证本文模型的优越性, 将本文模型与 MEGA, FGFA, RDN, DFF 模型在煤研石视频数据集上进行对比, 结果见表 1, 其中 “101” 表示骨干网络为 ResNet-101 模型, “50” 表示骨干网络为 ResNet-50 模型。可看出本文模型的 mAP 优于其他模型, 为 77.12%。本文模型在慢速运动的目标检测上识别精度最好, 达 83.82%, 这是因为门控存储增强模块提高了慢速运动目标的识别精度。各模型快速运动目标识别精度低于中速运动目标识别精度、中速运动目标识别精度低于慢速运动目标识别精度, 说明模型视频目标运动速度与识别精度呈负相关。

表 1 本文模型与 MEGA, FGFA, RDN, DFF 模型 mAP 对比

Table 1 The mAP comparison of the proposed model and MEGA, FGFA, RDN, DFF models

模型	识别精度			mAP
	快速运动目标	中速运动目标	慢速运动目标	
本文模型	55.12	76.02	83.82	77.12
MEGA-101	55.63	76.24	82.39	76.65
MEGA-50	49.53	70.58	79.83	72.63
RDN-101	51.65	71.95	82.10	74.68
RDN-50	45.27	67.46	80.22	70.40
FGFA-101	43.97	69.86	81.07	71.91
FGFA-50	40.75	66.57	78.90	68.68
DFF-101	37.47	66.87	79.32	68.42
DFF-50	35.65	62.19	74.14	63.50

模型训练时设置训练迭代次数为 120 000, 批尺寸为 64, 长期视频为 18~20 帧, 短期视频为 3~5 帧。本文选取训练中前 80 000 次迭代数据绘制损

失函数曲线,如图 7 所示。可看出在迭代过程中,损失值在训练开始后明显下降,并在 30 000 次迭代后趋于平稳且最终收敛于 0.23,说明了本文模型具有良好的鲁棒性和拟合性。

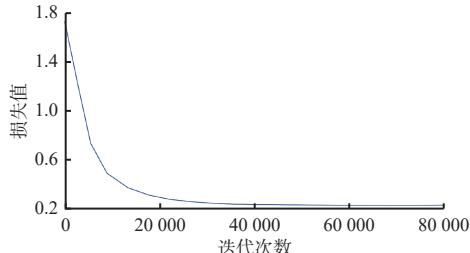


图 7 损失函数曲线

Fig. 7 Loss function curve

4 结论

1) 针对视频帧冗余度高的问题,采用煤矸石视频前后多帧聚合的处理思路,将多个视频帧按照权重进行聚合,加快模型整体的识别速度。采用注意力机制将图像中相似度较高区域的语义相似性权重降低,将有效特征区域的权重增加,以提高融合后特征图的有效性。将非关键帧中长期视频帧、短期视频帧特征进行存储,进而在关键帧预测时加以增强其特征,提升模型整体的识别精度。

2) 相较于 MEGA, FGFA, RDN, DFF 模型,基于 LSS 的聚合增强型煤矸石视频识别模型的 mAP 优于其他模型,为 77.12%;所有模型视频目标运动速度与识别精度呈负相关,且基于 LSS 的聚合增强型煤矸石视频识别模型在慢速运动的目标检测上识别精度最高,达 83.82%。

参考文献(References):

- [1] SHARMA V, GUPTA M, KUMAR A, et al. Video processing using deep learning techniques: a systematic literature review[J]. *IEEE Access*, 2021, 9: 139489-139507.
- [2] AICH A, ZHENG M, KAEANAM S, et al. Spatio-temporal representation factorization for video-based person re-identification[C]. International Conference on Computer Vision, Montreal, 2021: 152-162.
- [3] 孙立新. 基于卷积神经网络的煤矸石识别方法研究[D]. 邯郸:河北工程大学, 2020.
SUN Lixin. Research on coal gangue recognition method based on convolutional neural network[D]. Handan: Hebei University of Engineering, 2020.
- [4] PAN Hongguang, SHI Yuhong, LEI Xinyu, et al. Fast identification model for coal and gangue based on the improved tiny YOLO V3[J]. *Journal of Real-Time Image Processing*, 2022, 19(3): 687-701.
- [5] ZHU Xizhou, WANG Yujie, DAI Jifeng, et al. Flow-guided feature aggregation for video object detection[C]. IEEE International Conference on Computer Vision, Venice, 2017, 408-417.
- [6] 张勇. 基于视频处理的煤矸石识别研究[D]. 徐州:中国矿业大学, 2018.
ZHANG Yong. Research on gangue identification based on video processing[D]. Xuzhou: China University of Mining and Technology, 2018.
- [7] 程健, 王东伟, 杨凌凯, 等. 一种改进的高斯混合模型煤矸石视频检测方法[J]. 中南大学学报(自然科学版), 2018, 49(1): 118-123.
CHENG Jian, WANG Dongwei, YANG Lingkai, et al. An improved Gaussian mixture model for coal gangue video detection[J]. *Journal of Central South University (Science and Technology)*, 2018, 49(1): 118-123.
- [8] LEI Xinyu, PAN Hongguang, HUANG Xiangdong. A dilated CNN model for image classification[J]. *IEEE Access*, 2019, 7: 124087-124095.
- [9] PAN Hongguang, WEN Fan, HUANG Xiangdong, et al. The enhanced deep plug-and-play super-resolution algorithm with residual channel attention networks[J]. *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, 2021, 41(2): 4069-4078.
- [10] ZHU Xizhou, DAI Jifeng, YUAN Lu, et al. Towards high performance video object detection[C]. The IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018: 7210-7218.
- [11] WANG Shiyao, ZHOU Yucong, YAN Junjie, et al. Fully motion-aware network for video object detection[C]. Proceedings of the European Conference on Computer Vision, Munich, 2018: 542-557.
- [12] WU Haiping, CHEN Yuntao, WANG Naiyan, et al. Sequence level semantics aggregation for video object detection[C]. The IEEE/CVF International Conference on Computer Vision, Seoul, 2019: 9217-9225.
- [13] FEICHTENHOFER C, PINZ A, ZISSERMAN A. Detect to track and track to detect[C]. The IEEE International Conference on Computer Vision, Venice, 2017: 3038-3046.
- [14] ZHOU Bolei, ANDONIAN A, TORRALBA A. Temporal relational reasoning in videos[C]. European Conference on Computer Vision, Munich, 2018: 803-818.
- [15] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017: 936-944.

(下转第 62 页)

- hierarchical adaptive k-means clustering[J]. *Journal of Intelligent Information Systems*, 2015, 44(1): 133-158.
- [10] PALMER C R, FALOUTSOS C. Density biased sampling: an improved method for data mining and clustering[J]. *ACM SIGMOD Record*, 2000, 29(2) : 82-92.
- [11] HUANG Jianbin, SUN Heli, KANG Jianmei, et al. ESC: an efficient synchronization-based clustering algorithm[J]. *Knowledge-Based Systems*, 2013, 40: 111-122.
- [12] MINAEI-BIDGOLI B, PARVIN H, ALINEJAD-ROKNY H, et al. Effects of resampling method and adaptation on clustering ensemble efficacy[J]. *Artificial Intelligence Review*, 2014, 41(1): 27-48.
- [13] AGGARWAL A, DESHPANDE A, KANNAN R. Adaptive sampling for K-means clustering[C]. 12th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems/13th International Workshop on Randomization and Computation, Berkeley, 2009: 15-28.
- [14] KUMAR K M, REDDY A R M. An efficient K-means clustering filtering algorithm using density based initial cluster centers[J]. *Information Sciences*, 2017, 418/419: 286-301.
- [15] SÁEZ J A, KRAWCZYK B, WOŹNIAK M. Analyzing the oversampling of different classes and types of examples in multi-class imbalanced datasets[J]. *Pattern Recognition*, 2016, 57: 164-178.
- [16] 胡欢. 多维数据上近似聚集和最近邻查询的高效算法[D]. 哈尔滨: 哈尔滨工业大学, 2021.
- [17] 李建忠. 面向社交网络的科技领域事件检测系统的研究与实现[D]. 西安: 西安电子科技大学, 2019.
- [18] 周萌. 基于多粒度级联森林哈希学习的图像检索[D]. 重庆: 重庆邮电大学, 2019.
- [19] ZHOU Meng. Multi-grained cascade forest based hashing for image retrieval[D]. Chongqing: Chongqing University of Posts and Telecommunications, 2019.
- [20] NELSON K P, THISTLETON W J. Comments on "Generalized Box-Müller method for generating q-Gaussian random deviates"[J]. *IEEE Transactions on Information Theory*, 2021, 67(10): 6785-6789.
- [21] 谷晓忱, 张民选. 一种基于FPGA的高斯随机数生成器的设计与实现[J]. *计算机学报*, 2011, 34(1): 165-173.
- [22] GU Xiaochen, ZHANG Minxuan. Design and implementation of a FPGA based Gaussian random number generator[J]. *Chinese Journal of Computers*, 2011, 34(1): 165-173.
- [23] ARNAIZ-GONZÁLEZ Á, DÍEZ-PASTOR J-F, RODRÍGUEZ J J, et al. Instance selection of linear complexity for big data[J]. *Knowledge-Based Systems*, 2016, 107: 83-95.
- [24] 谷晓忱. 基于FPGA的高斯随机数生成器的设计与实现[D]. 重庆: 重庆邮电大学, 2019.
- [25] ARNAIZ-GONZÁLEZ Á, DÍEZ-PASTOR J-F, RODRÍGUEZ J J, et al. Instance selection of linear complexity for big data[J]. *Knowledge-Based Systems*, 2016, 107: 83-95.
- [26] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [27] DENG Jiajun, PAN Yingwei, YAO Ting, et al. Relation distillation networks for video object detection[C]. The IEEE/CVF International Conference on Computer Vision, Seoul, 2019: 7023-7032.
- [28] ZHU Xizhou, XIONG Yuwen, DAI Jifeng, et al. Deep feature flow for video recognition[C]. The IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017: 2349-2358.
- (上接第 44 页)
- [16] HE Kaiming, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[C]. International Conference on Computer Vision, Venice, 2017: 2961-2969.
- [17] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [18] CHEN Yihong, CAO Yue, HU Han, et al. Memory enhanced global-local aggregation for video object detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, 2020: 10337-10346.
- [19] GIRSHICK R. Fast R-CNN[C]. IEEE International Conference on Computer Vision, Santiago, 2015, 1440-1448.
- [20] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [21] DENG Jiajun, PAN Yingwei, YAO Ting, et al. Relation distillation networks for video object detection[C]. The IEEE/CVF International Conference on Computer Vision, Seoul, 2019: 7023-7032.
- [22] ZHU Xizhou, XIONG Yuwen, DAI Jifeng, et al. Deep feature flow for video recognition[C]. The IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017: 2349-2358.