DOI:10. 13382/j. jemi. B2508451

基于跨层融合语义增强特征的废钢图像分类方法

梁凯朔[1] 朱倍孝[2] 赵高鹏[1] 徐皓远[1]
(1. 南京理工大学自动化学院 南京 210094;2. 上海科技大学信息科学与技术学院 上海 201210)

摘 要:针对废钢堆叠严重的问题及对废钢精细化分类的需求,本文提出一种基于跨层融合语义增强特征的废钢图像细粒度分类方法。首先,采用运动检测实现从视频序列中检索出不包含抓斗等运动物体的废钢图像;其次,采用 Segment Anything(SAM)视觉大模型对不包含抓斗等运动物体的废钢图像进行语义分割,以分割出废钢图像中的废钢实例;最后,提出了一种基于跨层融合语义增强特征的废钢图像分类模型(efficientnetb5-cross layer fusion semantically enhanced feature, EfficientNetB5-CLFSEF),该模型采用 EfficientNetB5 模型的特征提取器,并且通过使用跨层融合特征语义增强特征模块(CLFSEF)实现废钢图像分类。CLFSEF 模块包括跨层特征融合(cross layer fusion, CLF)部分和语义增强特征(semantically enhanced feature, SEF)部分,CLF 通过融合来自特征提取器中不同层的特征,使模型在捕获深层语义信息同时,保留边界等低级语义信息;SEF 模块对融合特征按照各通道之间的语义相似性进行分组,并结合知识蒸馏技术和最大熵正则化技术提升模型对输入废钢图像中最具区分性部分的理解。本文在自制数据集上进行实验,实验结果表明,所提出的 EfficientNetB5-CFLSEF 模型能够对统废、剪料 1、剪料 2、炉料 1、炉料 2、钢板料和重废进行准确分类,该模型在测试集上的准确率为 90. 51%,优于相对比的分类模型。

关键词: 图像细粒度分类;跨层特征融合;语义增强特征;SAM;运动检测
中图分类号: TP399;TN911. 73;TP391 **文献标识码**: A **国家标准学科分类代码**: 510. 99

Classification method for scrap steel images based on cross-layer fusion semantic enhanced features

Liang Kaishuo[1] Zhu Beixiao[2] Zhao Gaopeng[1] Xu Haoyuan[1]
(1. School of Automation, Nanjing University of Science and Technology, Nanjing 210094, China;
2. School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China)

Abstract: In response to the severe stacking of scrap steel samples and the need for refined classification of scrap steel types, this paper proposes a scrap steel image classification method based on cross-layer fusion of semantic-enhanced features. The proposed method consists of several stages, aiming to optimize the accuracy and efficiency of scrap steel classification. The first stage is motion detection, which is used to extract scrap steel images without moving objects such as grapples from video sequences. This step ensures that the dataset excludes irrelevant objects, providing a more accurate foundation for subsequent analysis. Next, the state-of-the-art visual model "Segment Anything Model (SAM)" is applied to perform semantic segmentation on scrap steel images without moving objects such as grapples, to segment the instances in the scrap steel images. The core contribution of this paper lies in the design of a scrap steel image classification model, EfficientNetB5-CLFSEF, which can effectively handle the subtle differences between scrap steel categories and the significant morphological changes within each category. This model uses EfficientNetB5 as the feature extractor, as it is renowned for its efficiency and high performance in visual recognition tasks. Additionally, the model integrates a novel cross-layer fusion of semantic-enhanced features (CLFSEF) module, which is crucial for improving the classification accuracy of scrap steel images. The CLFSEF module consists of two key components: cross-layer feature fusion (CLF) and semantic-enhanced features (SEF). CLF fuses the features from different layers of the EfficientNetB5 feature extractor, enabling the model to capture deep semantic information and low-level details such as boundaries, which is crucial for distinguishing similar scrap steel categories. On the other hand, the SEF module groups the fused features based on semantic similarity between channels. This grouping process enables the model to focus on the most discriminative

收稿日期: 2025-06-10 Received Date: 2025-06-10

features in the image. Moreover, the SEF module also integrates knowledge distillation and maximum entropy regularization techniques to enhance the model's ability to recognize the most significant parts of the input scrap steel images. To validate the proposed method, experiments were conducted using a specially customized dataset for scrap steel classification. The benchmark EfficientNetB5 achieved an accuracy of 87.98% on the test set. After introducing the CLF module, the accuracy increased to 89.63%. Adding the SEF module resulted in an accuracy of 89.23%, and when the CLF and SEF modules are combined into the complete CLFSEF module, the accuracy increased to 90.51%. Compared to the benchmark classification model, these improvements increased by 1.65%, 1.25%, and 2.53% respectively. Moreover, the proposed model outperforms the comparison classification models.

Keywords: image fine-grained classification; cross-layer feature fusion; semantic enhancement features; SAM; motion detection

0 引 言

我国钢铁产业在全球居于领先地位,废钢作为重要的生产原料,其回收和再利用对钢铁工业至关重要[1-3]。然而,废钢来源复杂且质量参差不齐,不同钢铁企业的分类标准差异较大,导致废钢的分类回收面临巨大挑战。目前,钢铁企业主要依赖于人工经验进行废钢分类,但由于废钢形态复杂,人工分类效率低下且容易产生主观误判,从而造成经济损失[4-5]。

基于深度学习的图像处理技术在目标检测、语义分割、图像分类等领域表现出强大能力,并广泛应用于不同任务之中[6-8]。深度学习技术通过对图像数据的自动学习,可以有效解决传统图像处理方法在复杂场景下的局限性[9-10]。国内外已有少数研究学者应用深度学习技术对废钢进行分类与识别。文献[11]首先提出了一种基于改进 Faster-RCNN 的废钢检测评级方法。文献[12]针对厚度小于 3 mm、厚度 3~6 mm、厚度大于 6 mm 以及镀锌件等 7 种类别废钢的分类问题,基于深度学习构建了废钢分类评级模型 CCBFNet,该模型融合卷积注意力机制和双向金字塔网络进行特征提取和多尺度特征融合。文献[13]将跨阶段局部网络(CSP)结构用于废钢图像的特征提取,并且结合空间金字塔结构和挤压-激励注意力机制构建了 CSSNet 模型。文献[14]提出了一种使用深度学习技术进行废钢检测和分级的分层框架,该方法根据废钢的类型将废钢分为 0~5 级,并且提出了车厢注意模块、废钢检测模块和分级模块。文献[15]提出一种基于改进 MobileNet 的废钢识别方法,该方法针对钢筋、破料、压件和重废料 4 个类别的废钢进行分类,由于这 4 类废钢在视觉上差异较大,该方法难以对细粒度废钢实例进行准确分类。文献[16]提出基于 VA-Expo-WA 集成迁移学习的方法,该方法运用 VGG-16、ResNet-50、DenseNet-121 等深度卷积神经网络预训练模型对废钢分类数据集进行特征提取,并优化、训练多个模型的分类层得到迁移学习模型。该方法针对废槽钢、废钢筋和边角料这 3 种类别间差异较大的废钢进行分类。该模型在处理类别间差异较小的废钢时,容易受到类别间视觉特征重叠的影响导致分类性能下降。文献[17]针对铁路货车中废金属的分类问题,通过使用 InceptionNet、NAS 等预训练模型为铁路货车中废金属的自动化分类提供技术支撑。

弱监督图像细粒度分类方法仅使用图像类别标签。通过采用基于注意力机制的图像细粒度分类方法对废钢进行分类,易产生局部特征表示不足的问题,从而影响分类性能。因此,本文采用执行高阶特征交互以及设计特殊损失函数的方法对废钢进行分类,以有效捕捉废钢图像中最具区分性的部位。文献[18]首次提出双线性卷积神经网络,该网络使用两个特征提取器,并通过在输入图像的每个位置进行外积运算和池化来进行图像细粒度分类。基于此,文献[19]运用跨层双线性池化来捕获不同卷积层之间的部分关系,并通过层次双线性池化框架整合多个跨层双线性特征。文献[20]提出 Cross-X Learning 方法,该方法结合跨类别跨语义正则化和跨层正则化,有效利用图像之间和网络层之间的关系进行稳健地多尺度特征学习。文献[21]无需边界框或局部注释,驱动特征通道关注局部判别区域。文献[22]通过调节函数加权样本损失,避免模型过拟合困难样本,提高模型的泛化能力。

针对废钢实例堆叠严重及对废钢类别精细化分类的需求,本文提出一种基于跨层融合语义增强特征的废钢图像分类方法。首先,采用运动检测实现从视频序列中检索出不包含抓斗的废钢图像;其次,采用 SAM 分割模型对不包含抓斗的废钢图像进行语义分割,以分割出废钢图像中的废钢实例。本文结合深度卷积神经网络、跨层特征融合以及语义增强特征,提出了一种基于跨层融合语义增强特征的废钢图像分类模型(EfficientNetB5-Cross layer fusion semantically enhanced feature, EfficientNetB5-CLFSEF),并在自制数据集上进行了实验分析,实现对统废、剪料 1、剪料 2、炉料 1、炉料 2、钢板料和重废 7 类废钢的准确分类。

1 方法介绍

1.1 方法概述

本文提出一种基于跨层融合语义增强特征的废钢图

像分类方法,该方法的流程如图 1 所示。当废钢运输到卸载场地时,视觉传感器捕获卸载的过程。首先,对视觉传感器捕获的视频序列应用运动检测,以检索出不包含运动物体的废钢图像,从而避免抓斗对废钢分类精度的影响。其次,采用 SAM 模型对不包含抓斗的废钢图像中的实例进行分割,以解决废钢图像中实例堆叠严重导致的分类检测不准确问题。最后,将废钢实例送入到所提出的基于跨层融合语义增强特征的废钢图像分类模型 EfficientNetB5-CLFSEF 中,得到分类结果。

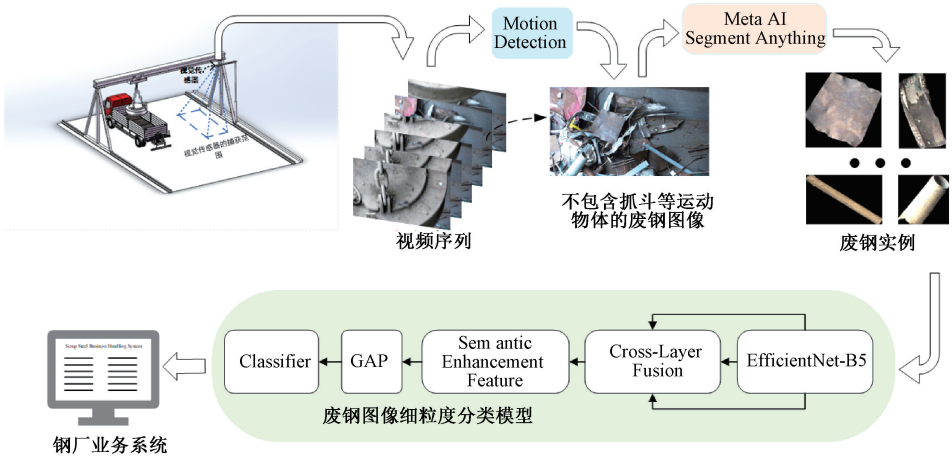


图 1　方法总体概述图

Fig. 1　Overall method overview diagram

1.2　运动检测

废钢运输车到达指定位置后进行卸载,位于卸载场地上方的视觉传感器捕获卸载过程中的视频图像。在卸载过程中,视觉传感器不仅会捕获到抓斗的运动过程,还可能捕捉到工人、鸟类等其他运动目标的运动。这些包含抓斗、工人以及其他干扰因素的废钢图像对后续的废钢分类带来干扰和影响。在两次抓取卸载之间存在几秒钟的时间间隔,因此,对此时间段内的视频序列通过运动检测的方法进行检索,提取一张不包含抓斗等运动目标的废钢图像,以提高废钢分类的准确率和可靠性。

首先,对视频序列中第 t 帧图像中所有位置的像素点经过灰度处理得到灰度图;其次,对灰度图中所有像素点采用基于高斯混合模型的背景减方法,以实现视频帧图像中前景与背景的分离。

假设视频序列中第 t 帧和第 $t+1$ 帧图像经过背景分离后得到的前景掩码图像分别为 $\boldsymbol{F}_t$ 和 $\boldsymbol{F}_{t+1}$,将前景掩码图像 $\boldsymbol{F}_t$、$\boldsymbol{F}_{t+1}$ 作差并取绝对值,可得到前景掩码图像的差分图像 $\boldsymbol{D}_{t,t+1}$:

$$\boldsymbol{D}_{t,t+1} = |\boldsymbol{F}_t - \boldsymbol{F}_{t+1}| \tag{1}$$

式中:$\boldsymbol{D}_{t,t+1}$ 中每个像素位置 (x,y) 处的值表示前景掩码图像 $\boldsymbol{F}_t$、$\boldsymbol{F}_{t+1}$ 中对应位置 (x,y) 处的前景掩码差异。

对差分图像 $\boldsymbol{D}_{t,t+1}$ 求像素均值:

$$U_{t,t+1} = \frac{1}{N}\sum_{(x,y)} \boldsymbol{D}_{t,t+1}(x,y) \tag{2}$$

式中:N 表示像素总数;$U_{t,t+1}$ 表示像素均值。

当像素均值 $U_{t,t+1}$ 小于阈值 ζ 时,则认为第 $t+1$ 帧图像为不包含抓斗等运动物体的废钢图像。

1.3　基于 SAM 的废钢图像分割

经过运动检测所得的废钢图像中,各个废钢实例之间堆叠严重,且图像中可能包含多种类别废钢。因此需要对经过运动检测获取的不含运动物体的废钢图像进行语义分割。

SAM 语义分割模型[23]能够根据文本指令或图像内容实现对任意物体进行精准分割,本文采用 SAM 语义分割模型对经过运动检测得到的废钢图像中的实例进行分割。

尽管 SAM 模型提供了性能优越的预训练模型,但在废钢图像中,由于各个实例之间存在显著的堆叠,直接应用预训练模型难以获得理想的分割效果。本文对自有废钢图像数据进行标注并结合 SAM 模型的预训练权重,对其进行微调,以增强其对废钢图像中复杂实例关系的分割能力。

对于 SAM 模型的微调,本文采用 sam_vit_h. pth 预训练权重,epoch 为 10 轮,batch size 为 1,采用 AdamW 优化器,初始学习率为 0.000 1,其余训练配置选用默认训练配置。

微调前后的 SAM 模型对废钢图像的分割效果如图 2(a)~(b)所示。由对比分割效果图可知,微调后的 SAM 模型在背景区域没有进行错误分割,能够有效处理废钢实例之间的复杂关系。

(a) 微调前
(a) Before fine-tuning

(b) 微调后
(b) After fine-tuning

图2　微调前后的SAM模型分割效果图示例

Fig. 2　Examples of segmentation results after fine-tuning the SAM model

1.4　废钢图像分类模型 EfficientNetB5-CLFSEF

本文根据钢铁企业的需求,将废钢划分为统废、剪料1、剪料2、炉料1、炉料2、钢板料和重废7种类别。其中,剪料1和剪料2、炉料1和炉料2这两组废钢类别,不同类别的废钢实例形态上视觉差异微小,而同一类别内的废钢实例之间则有较大的视觉形态差异,基础分类网络难以对其进行准确分类。针对此问题,本文提出一种基于跨层融合语义增强特征的废钢图像分类模型,该模型的网络结构如图3所示,主要由特征提取网络(Backbone)、跨层融合语义增强特征模块(cross layer fusion semantically enhanced feature, CLFSEF)构成。

1)特征提取网络

本文鉴于EfficientNet[24]系列模型在模型性能与计算资源之间实现了最优平衡,选取EfficientNetB5作为废钢图像分类模型的特征提取网络。

EfficientNet系列模型主要由MBConv模块堆叠而

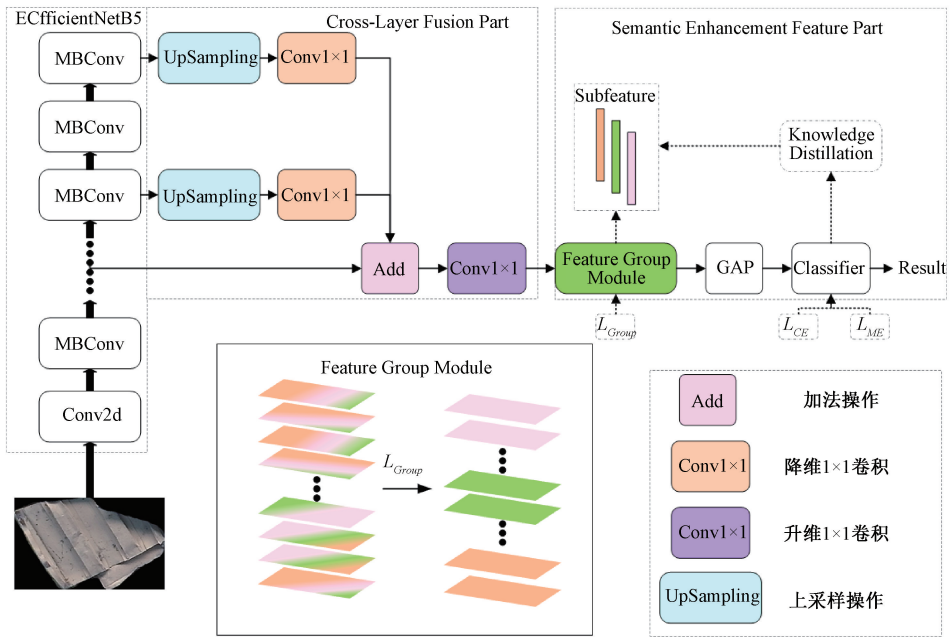


图3　EfficientNetB5-CLFSEF网络结构图

Fig. 3　Structure of efficientNetB5-CLFSEF

成,具备轻量化、表达能力强的特点。MBConv模块结合了深度可分离卷积与SE注意力机制,能够有效提取废钢图像中的关键特征。

2)CLFSEF模块及总体损失函数

文献[25]提出语义增强特征模块(Semantically Enhanced Feature,SEF),该模块通过对CNN最后一层卷积层的输出特征按照各个特征通道之间的语义相似性进行分组,以表示输入图像中的不同语义部分,并且通过知识蒸馏和最大熵技术增强模型对输入图像中不同语义部分的学习。然而,CNN深层的输出特征主要捕捉了输入图像的抽象语义信息,忽略了边界、纹理等低级语义信息。因此,针对这一局限,本文通过对CNN不同层的输出特征进行融合,结合低级与高级特征的优势。然后,根据特征通道之间的语义相似性对融合后的特征进行分组,以形成跨层融合语义增强特征。

将CNN中的顶层特征、中间层特征和低层特征分别表示为 $\boldsymbol{F}_H \in \boldsymbol{R}^{C_H \times H_H \times W_H}$、$\boldsymbol{F}_M \in \boldsymbol{R}^{C_M \times H_M \times W_M}$ 和 $\boldsymbol{F}_L \in \boldsymbol{R}^{C_L \times H_L \times W_L}$。

首先,将顶层特征和中间层特征的特征尺寸上采样

到低层特征的特征尺寸,如式(3)、(4)所示:

$$\boldsymbol{F}'_H = \Gamma_{Upsampling}(\boldsymbol{F}_H) \in \boldsymbol{R}^{C_H \times H_L \times W_L} \tag{3}$$

$$\boldsymbol{F}'_M = \Gamma_{Upsampling}(\boldsymbol{F}_M) \in \boldsymbol{R}^{C_M \times H_L \times W_L} \tag{4}$$

式(3)、(4)中,$\Gamma_{Upsampling}$ 表示上采样运算。

其次,通过使用 1×1 卷积核将顶层特征和中间层特征的特征通道数降维到低层特征的通道数,如式(5)、(6)所示。

$$\boldsymbol{F}''_H = \Gamma_{conv1\times1,\downarrow}(\boldsymbol{F}'_H) \in \boldsymbol{R}^{C_L \times H_L \times W_L} \tag{5}$$

$$\boldsymbol{F}''_M = \Gamma_{conv1\times1,\downarrow}(\boldsymbol{F}'_M) \in \boldsymbol{R}^{C_L \times H_L \times W_L} \tag{6}$$

式(5)、(6)中,$\Gamma_{conv1\times1,\downarrow}$ 表示降维 1×1 卷积运算。

将顶层特征、中间层特征和低层特征按照在特征通道维度上进行相加并通过使用 1×1 卷积核将所得特征的特征通道数升维到顶层特征的特征通道数,以得到融合特征 $\boldsymbol{F}_{Fusion} \in \boldsymbol{R}^{C_H \times H_L \times W_L}$,如式(7)、(8)所示。

$$\boldsymbol{F}_{Add} = \boldsymbol{F}_L + \boldsymbol{F}''_M + \boldsymbol{F}''_H \in \boldsymbol{R}^{C_L \times H_L \times W_L} \tag{7}$$

$$\boldsymbol{F}_{Fusion} = \Gamma_{conv1\times1,\uparrow}(\boldsymbol{F}_{Add}) \in \boldsymbol{R}^{C_H \times H_L \times W_L} \tag{8}$$

式(8)中,$\Gamma_{conv1\times1,\uparrow}$ 表示 1×1 升维卷积运算。

对于融合特征 $\boldsymbol{F}_{Fusion}$,每个特征通道表示为 $\boldsymbol{F}^i_{Fusion} \in \boldsymbol{R}^{H_L \times W_L}$,其中 $i = [1,2,\cdots,C_H]$。将融合特征 $\boldsymbol{F}_{Fusion}$ 分成 G 组,并如式(9)所示,对分组后的融合特征 $\boldsymbol{F}'_{Fusion}$ 中各通道特 $\boldsymbol{F}^{i'}_{Fusion}$ 进行归一化:

$$\hat{\boldsymbol{F}}^{i'}_{Fusion} = \boldsymbol{F}^{i'}_{Fusion} / \| \boldsymbol{F}^{i'}_{Fusion} \|_2 \tag{9}$$

各通道特征之间的相似性可如式(10)确定,设 $\boldsymbol{D} \in \boldsymbol{R}^{G\times G}$ 为相关性矩阵,其中的元素如式(11)所示。

$$d_{i,j} = \hat{\boldsymbol{F}}^{i'^T}_{Fusion} \hat{\boldsymbol{F}}^{i'}_{Fusion} \tag{10}$$

$$D_{mn} = \frac{1}{C_m C_n} \sum_{i\in m, j\in n} d_{i,j} \tag{11}$$

式(11)中,D_{mn} 表示组 m 和组 n 中特征通道之间的平均相关性,C_m 和 C_n 分别表示组 m 和组 n 中的特征通道数,$m,n \in [1,\cdots,G]$。

因此,对融合特征 F_{Fusion} 的特征通道按照语义相似性进行分组可以由最大化各组间的相似性、最小化组内通道间的相似性确定,损失函数如式(12)所示。

$$L_{Group} = \frac{1}{2}(\| D \|_F^2 - 2 \| diag(D) \|_2^2) \tag{12}$$

通过对融合特征 F_{Fusion} 的特征通道按照语义相似性进行分组,可以使得不同组所代表的融合子特征专注于输入图像中不同语义信息的区域。然而,仅仅对融合特征的特征通道进行分组不能保证这些区域在分类任务中具有强区分性。因此,通过匹配融合子特征与融合特征的预测分布,增强融合子特征的区分性,从而提升模型在细粒度分类任务中的性能。具体而言,通过使用知识蒸馏技术来匹配融合子特征与融合特征的预测分布。

假设 $\boldsymbol{P}_\omega$ 为全局特征的预测分布,$\boldsymbol{P}_a$ 为子特征的预测分布,则知识蒸馏的损失函数如式(13)所示。

$$H(\boldsymbol{P}_\omega, \boldsymbol{P}_a) = -\sum \boldsymbol{P}_\omega \log \boldsymbol{P}_a \tag{13}$$

为了鼓励分类模型在做出决策前保持预测的不确定性,因此引入最大熵正则化损失,其计算公式如式(14)所示。

$$H(\boldsymbol{P}_\omega) = -\sum \boldsymbol{P}_\omega \log \boldsymbol{P}_\omega \tag{14}$$

由式(12)~(14)可知,EfficientNet-CLFSEF 模型的总体损失函数如式(15)所示。

$$L_{Total} = L_{CE} + \lambda H(\boldsymbol{P}_\omega) + \frac{\gamma}{G} H(\boldsymbol{P}_\omega, \boldsymbol{P}^g_a) + \mu L_{Group} \tag{15}$$

式中:L_{CE} 表示标准交叉熵损失函数;参数 λ、γ 和 μ 表示不同损失函数的权重系数。

2　实验结果及分析

2.1　数据集

本文原始图片数据由江阴西城钢铁通过海康 IDS-2DC7823IX-A/T3 型广角相机捕获废钢卸载过程的视频图像,通过运动检测从中提取不包含抓斗的废钢图像构成。以获取视频序列中的片段为例,抓斗抓取废钢的部分视频帧图像如图 4(a)~(f)所示,其中图 4(f)为视频片段经运动检测的结果图片,即不包含抓斗的废钢图像。

各类别的原始图片数据图例如图 5(a)~(g)所示。废钢按照厚度分为统废、剪料 1、剪料 2、炉料 1、炉料 2、钢板料和重废,各类别废钢的标签名称、划分标准及品种举例如表 1 所示。

对于原始图片数据,本文通过使用 SAM 模型对原始图片进行分割,并将所有类别的分割图片数据按照 8∶2 的比例划分成训练集、测试集,数据集统计如表 2 所示。

2.2　评价指标

EfficientNetB5-CLFSEF 模型的性能采用混淆矩阵和准确率(accuracy)来评估。

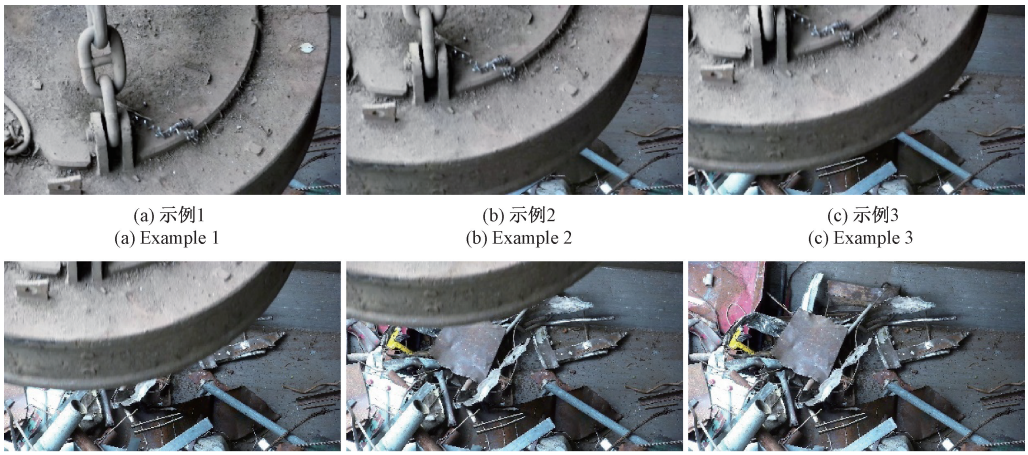
混淆矩阵主要由 TP、FN、FP、TN 组成,各参数的解释为:

TP:模型预测出为正类别且实际为正类别的目标个数;

FN:模型预测出为负类别且实际为正类别的目标个数;

FP:模型预测出为正类别且实际为负类别的目标个数;

TN:模型预测出为负类别且实际为负类被的目标个数。



(a) 示例1
(a) Example 1

(b) 示例2
(b) Example 2

(c) 示例3
(c) Example 3

(d) 示例4
(d) Example 4

(e) 示例5
(e) Example 5

(f) 示例6、运动检测结果图
(f) Example 6, Motion Detection Result

图 4　视频部分图像帧示例及运动检测结果

Fig. 4　Video frame image sequence example

(a) 统废
(a) Centralized scrap

(b) 剪料1
(b) Sheared scrap 1

(c) 剪料2
(c) Sheared scrap 2

(d) 炉料1
(d) Furnace burden 1

(e) 炉料2
(e) Furnace burden 2

(f) 钢板料
(f) Steel plate scrap

(g) 重废
(g) Heavy scrap

图 5　各类别废钢样例

Fig. 5　Scrap steel type legend

表 1　废钢类别划分标准

Table 1　Classification standard for scrap steel

标签名称	类别名称	厚度(d)划分标准	品种举例
type0	统废	1 mm≤d<1. 5 mm	车厢栏板、大栅管、工业铁皮等
type1	剪料 1	1. 5 mm≤d<3 mm	小方管、小圆管、摩托车车架等
type2	剪料 2	3 mm≤d<4 mm	各种工业废钢剪切加工料
type3	炉料 1	4 mm≤d<6 mm	角钢、槽钢、工字钢等
type4	炉料 2	6 mm≤d<8 mm	角钢、槽钢、工字钢等
type5	钢板料	8 mm≤d<10 mm	钢板裁切料、边角料等
type6	重废	d≥10 mm	大中型钢类、工业拆废类、马蹄铁等

表 2　数据集统计

Table 2　Dataset statistics

类别	统废	剪料 1	剪料 2	炉料 1	炉料 2	钢板料	重废
原始图片/张	67	46	52	68	52	58	79
分割图片/张	2 048	2 060	2 100	1 840	1 924	1 834	1 837

混淆矩阵如表 3 所示。

表 3　混淆矩阵

Table 3　Confusion matrix

Type	Positive	Negative
True	TP	FN
False	FP	TN

根据混淆矩阵，分类模型的准确率可由式(16)确定，其用来评估分类模型的全局准确程度。

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{16}$$

2.3　实验环境与设置

本文仿真实验在 Windows10 系统中，搭建 Pytorch-GPU 框架的废钢分割模型运行及废钢分类模型训练与测试的深度学习环境，使用 Nvidia RTX6000 24 G 显卡对模型进行加速，CPU 型号为 Intel(R) Xeon(R) Silver 4210@2.20 GHz。

对于 EfficientNetB5-CLFSEF 模型，本文在实验中将输入图片大小设置为 456×456，epoch 为 25 轮，batch size 为 8，采用 Adam 优化器，初始学习率为 0.000 1，并且当模型在验证集上的准确率连续 5 轮不上升时，将学习率缩小为原来的 10 倍，最小学习率为 0.000 000 1。

对于相对比的分类模型，图片大小采用默认训练配置，其余同 EfficientNetB5-CLFSEF 模型相同。

2.4　消融实验结果与分析

本文通过消融实验来验证所提出的 EfficientNetB5-CLFSEF 模型中各部分的有效性，各组实验在测试集上的混淆矩阵如图 6 所示。

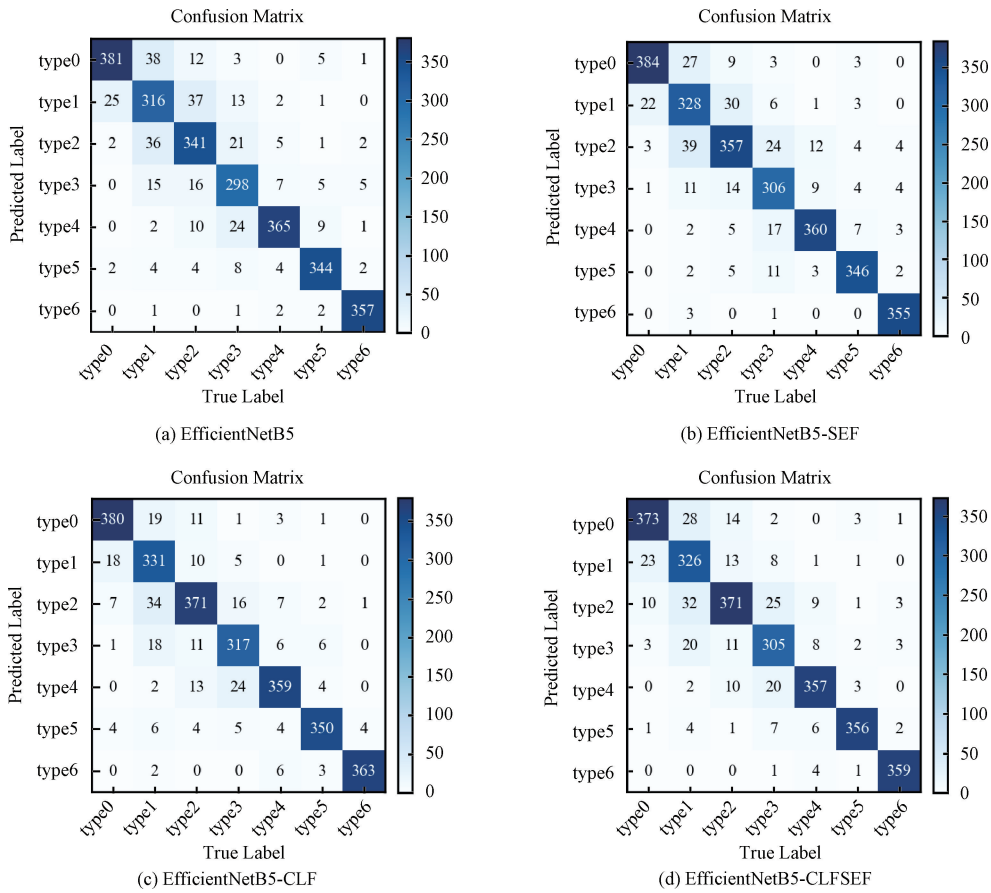


图 6　消融实验中各模型的混淆矩阵

Fig. 6　Confusion matrices of each model in the ablation experiment

由混淆矩阵计算出各组实验的准确率，其中，EfficientB5-SEF 表示仅在 EfficientNetB5 中添加 SEF 模块；EfficientNetB5-CLF 表示仅在 EfficientNetB5 模型中田间 CLF 模块。消融实验结果如表 4 所示。

消融实验的混淆矩阵以及准确率结果表明，本文所提出的 EfficientNetB5-CLFSEF 模型对各类别废钢，尤其

表 4　消融实验结果表

Table 4　Results of the ablation experiment

模型	准确率/%	模型大小/MB
EfficientNetB5	87.98	118.0
EfficientNetB5-SEF	89.23(+1.25)	130.0
EfficientNetB5-CLF	89.63(+1.65)	121.0
EfficientNetB5-CLFSEF(**ours**)	90.51(+**2.53**)	130.3

是剪料1、剪料2、炉料1和炉料2四类细粒度类别废钢的分类展现了显著优势。具体而言,相较于基础分类模型EfficientNetB5,该模型的分类准确率由87.98%提升至90.51%,提高了2.53%。这一提升主要得益于CLFSEF模块的有效性。其中,跨层特征融合CLF部分通过融合不同层的特征,使模型能够在捕获深层语义信息的同时,保留废钢实例边界等低级语义信息;语义增强特征SEF部分通过对融合特征按照通道间语义相似性进行分组,并结合知识蒸馏技术和最大熵正则化技术,进一步增强了模型对废钢图像中最具区分性部分的理解。因此,EfficientNetB5-CLFSEF模型能够准确地对统废、剪料1、剪料2、炉料1、炉料2、钢板料和重废7类废钢进行准确分类,有效解决因废钢类别实例间视觉差异微小且各类内形态差异较大的分类问题。同时,引入的CLFSEF模块并未显著增加模型的参数量。

2.5　对比实验结果与分析

为进一步证明本文所提出的废钢图像分类模型EfficientNetB5-CLFSEF对于本文所涉及废钢类别的分类性能,选取EfficientNetV2[26]模型、Vision Transformer(ViT)[27]模型和Swin Vision Transformer(Swin-ViT)[28]模型进行对比实验。各组实验的混淆矩阵如图7(a)~(d)所示。

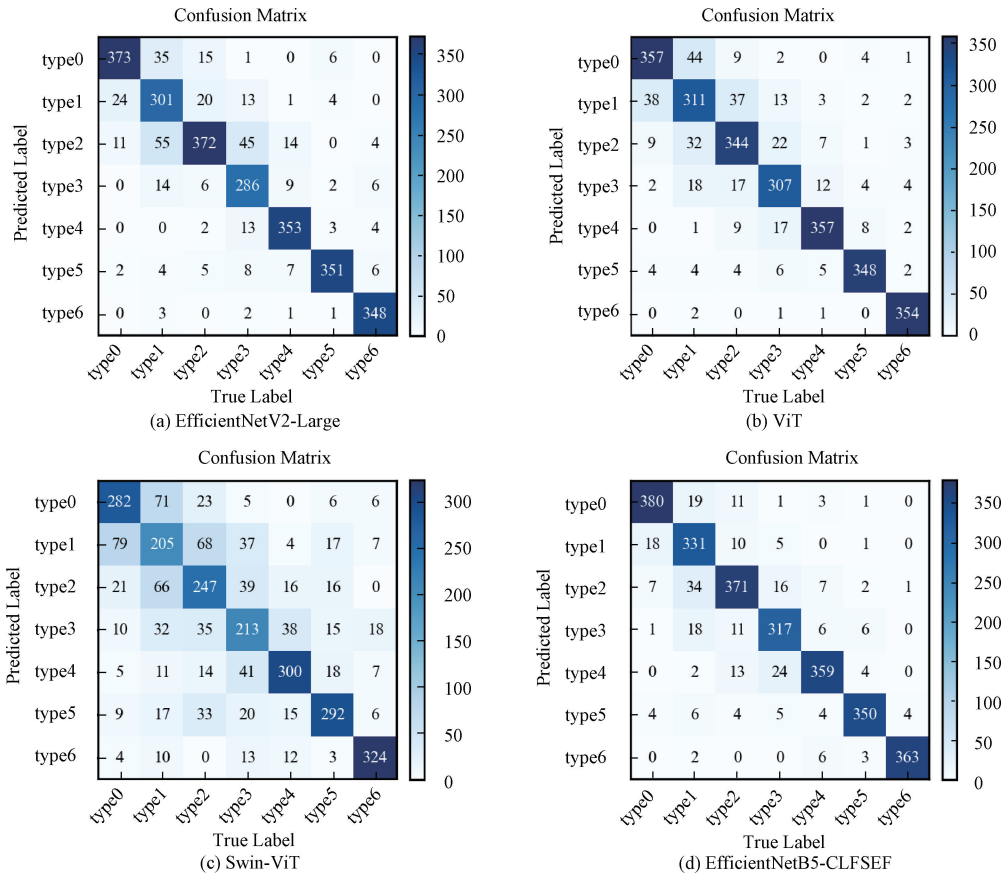


图7　对比实验中各模型的混淆矩阵

Fig. 7　Confusion matrices of each model in the comparative experiment

根据图7所示的各模型的混淆矩阵可知,本文所提出的EfficientNetB5-CLFSEF模型在统废、剪料1、炉料1、炉料2、统废这5类废钢中表现出最优的分类效果,而在钢板料类别中则展示了次优的分类性能。尽管EfficientNetB5-CLFSEF模型在所有废钢类别中并非始终取得最优分类效果,但其在钢板料类别和剪料2类别中分别呈现出接近最优的分类性能,且两者的偏差都位于可接受范围之内。同时,本文所提出的模型在剪料1、剪料2、炉料1和炉料2这4类细粒度废钢类别的综合分类效果优于相对比的分类模型。

由图7所示的混淆矩阵得到的各模型在测试集上的分类准确率如表5所示,其中标注加粗表示最优指标。

表5所示的对比实验结果表明,EfficientNetB5-CLFSEF模型展现出最优的分类准确率,优于相对比的方法,并且模型在保持较高分类准确率同时,避免过度增大模型的复杂度。

表 5　对比试验结果表

Table 5　Results of the comparative experiment

模型	准确率/%	模型大小/MB
EfficientNetV2-Large	87.33	455.0
ViT	87.11	327.0
Swin-ViT	68.24	331.0
EfficientNetB5-CLFSEF	**90.51**	**130.3**

3　结　论

针对废钢实例堆叠严重及对废钢类别精细化分类的需求,本文提出了一种基于跨层融合语义增强特征的废钢图像分类方法。首先,采用运动检测实现从视频序列中检索出不包含抓斗等运动物体的废钢图像;其次,采用 SAM 视觉大模型对不含抓斗等运动物体的废钢图像进行语义分割,以分割出废钢图像中的废钢示例;最后,本文提出了一种基于跨层融合语义增强特征的废钢分类模型,通过结合跨层融合与语义特征分组形成跨层融合语义增强特征,实现了对统废、剪料 1、剪料 2、炉料 1、炉料 2、钢板料和重废 7 个废钢类别的分类,进而满足对废钢类别精细化分类的需求。在自制数据集上的实验结果表明,所提出的废钢图像分类模型的准确率达到 90.51%,优于相对比的方法。

参考文献

[1] 闫甜甜. 我国废钢市场发展现状与趋势分析[J]. 冶金信息导刊, 2024, 61(6):5-7, 56.
YAN T T. Analysis of current situation and trend of scrap market development in China [J]. Metallurgical Information Review, 2024, 61(6):5-7, 56.

[2] 王国栋,张龙强,付静,等. “双碳”背景下我国废钢资源高质循环利用战略研究[J]. 中国工程科学, 2024, 26(3):63-73.
WANG G D, ZHANG L Q, FU J, et al. Research on high quality recycling strategy of scrap steel resources in China under the background of “double carbon” [J]. Strategic Study of CAE, 2024, 26(3):63-73.

[3] 张琦,田硕硕,李星宇,等. 碳中和目标下中国废钢资源量预测及高质利用策略[J]. 钢铁, 2024, 59(9):205-214.
ZHANG Q, TIAN SH SH, LI X Y, et al. Prediction and high quality utilization strategy of scrap steel resources in China under carbon neutral target[J]. Iron and Steel, 2024, 59(9):205-214.

[4] 江润芹. 关于废钢回收加工企业生存现状的思考[J]. 资源再生, 2025(2):24-27.
JIANG R Q. Thinking about the existing situation of scrap recycling and processing enterprises[J]. Resource Recycling, 2025(2):24-27.

[5] 赵忠阳. 论钢铁企业的废钢管理策略[J]. 冶金管理, 2019(9):144, 151.
ZHAO ZH Y. Management strategy of scrap steel in iron and steel enterprises[J]. China Steel Focus, 2019(9):144, 151.

[6] 李季桐,刘杰,杨娜,等. 基于层次化多尺度特征融合的金属缺陷分类模型[J]. 仪器仪表学报, 2025, 46(3):206-218.
LI J T, LIU J, YANG N, et al. Enhanced hierarchical multi-scale feature fusion model for metal defect classification [J]. Chinese Journal of Scientific Instrument, 2025, 46(4):206-218.

[7] 李强,马超,黄民. 基于注意力的多尺度残差卷积网络轴承故障诊断[J]. 电子测量技术, 2025, 48(9):19-26.
LI Q, MA CH, HUANG M. Attention-based multi-scale residual convolutional network for bearing fault diagnosis[J]. Electronic Measurement Technology, 2025, 48(9):19-26.

[8] 王天洋,刘路,王太勇,等. 基于改进 YOLOv8s 的轻量级 PCB 缺陷检测算法[J]. 电子测量与仪器学报, 2025,39(3):44-52.
WANG T Y, LIU L, WANG T Y, et al. Lightweight PCB defect detection algorithm based on improved YOLOv8s[J]. Journal of Electronic Measurement and Instrumentation,2025,39(3):44-52.

[9] 孙志军, 薛磊, 许阳明,等. 深度学习研究综述[J]. 计算机应用研究, 2012, 29(8):2806-2810.
SUN ZH J, XUE L, XU Y M, et al. Overview of deep learning research [J]. Application Research of Computers, 2012, 29(8):2806-2810.

[10] 刘建伟, 刘媛, 罗雄麟. 深度学习研究进展[J]. 计算机应用研究, 2014, 31(7):1921-1930, 1942.
LIU J, LIU Y, LUO X L. Research progress of deep learning[J]. Application Research of Computers, 2014, 31(7):1921-1930, 1942.

[11] QIN Y, CHEN W, ZHANG P, et al. Research on scrap steel evaluation technology based on faster-RCNN[C]. ICMLCA 2021; 2nd International Conference on Machine Learning and Computer Application. VDE, 2021:1-4.

[12] 肖鹏程,徐文广,常金宝,等. 基于深度学习的废钢分类评级方法研究[J]. 工程科学与技术, 2023, 55(2):184-193.
XIAO P CH, XU W G, CHANG J B, et al, Research on classification and rating method of scrap steel based on deep learning [J]. Advanced Engineering Sciences,

2023, 55(2):184-193.
[13] 肖鹏程,徐文广,张妍,等. 基于SE注意力机制的废钢分类评级方法[J]. 工程科学学报, 2023, 45(8):1342-1352.
XIAO P CH, XU W G, ZHANG Y, et al, Classification method of scrap steel based on SE attention mechanism[J]. Chinese Journal of Engineering, 2023, 45(8):1342-1352.
[14] TU Q, LI D, XIE Q, et al. Automated scrap steel grading via a hierarchical learning-based framework[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71:1-13.
[15] 官世杰,席晨馨. 基于改进MobileNet的废钢识别方法研究[J]. 湖北大学学报(自然科学版), 2024, 46(5):621-628.
GUAN SH J, XI CH X. Research on scrap identification method based on improved MobileNet[J]. Journal of Hubei University (Natural Science), 2024, 46(5):621-628.
[16] 周婧,秦伦明. 基于VA-Expo-WA集成迁移学习的废钢分类[J]. 信息技术, 2022(5):18-24.
ZHOU J, QIN L M. Scrap classification based on VA-Expo-WA integrated transfer learning[J]. Information Technology, 2022(5):18-24.
[17] SMIRNOV N V, RYBIN E I. Machine learning methods for solving scrap metal classification task[C]. 2020 International Russian Automation Conference (RusAutoCon). IEEE, 2020:1020-1024.
[18] LIN T Y, ROYCHOWDHURY A, MAJI S. Bilinear CNN models for fine-grained visual recognition[C]. Proceedings of the IEEE International Conference on Computer Vision. 2015:1449-1457.
[19] YU CH J, ZHAO X Y, ZHENG Q, et al. Hierarchical bilinear pooling for fine-grained visual recognition[C]. Proceedings of the European Conference on Computer Vision (ECCV). 2018:574-589.
[20] LUO W, YANG X, MO X, et al. Cross-x learning for fine-grained visual categorization[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019:8242-8251.
[21] CHANG D, DING Y, XIE J, et al. The devil is in the channels: Mutual-channel loss for fine-grained image classification[J]. IEEE Transactions on Image Processing, 2020, 29:4683-4695.
[22] LIANG Y, ZHU L, WANG X, et al. Penalizing the hard example but not too much: A strong baseline for fine-grained visual classification[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 35(5):7048-7059.
[23] KIRILLOV A, MINTUN E, RAVI N, et al. Segment anything[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023:4015-4026.
[24] TAN M, LE Q. Efficientnet:Rethinking model scaling for convolutional neural networks[C]. International Conference on Machine Learning. PMLR, 2019:6105-6114.
[25] LUO W, ZHANG H, LI J, et al. Learning semantically enhanced feature for fine-grained image classification[J]. IEEE Signal Processing Letters, 2020, 27:1545-1549.
[26] TAN M, LE Q. Efficientnetv2:Smaller models and faster training[C]. International Conference on Machine Learning. PMLR, 2021:10096-10106.
[27] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words:Transformers for image recognition at scale[J]. ArXiv preprint arXiv:2010.11929, 2020.
[28] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021:10012-10022.

作者简介

梁凯朔,2023年于南京工程学院获得学士学位,现为南京理工大学硕士研究生,主要研究方向为视频图像处理、目标检测与深度学习。
E-mail:123110223263@njust.edu.cn

Liang Kaishuo received his B. Sc. degree from Nanjing Institute of Technology in 2023. He is now a M. Sc. candidate at Nanjing University of Science and Technology. His main research interest includes video image processing, object detection and deep learning.

赵高鹏(通信作者),2010年于南京理工大学获得博士学位,现为南京理工大学副研究员,主要研究方向为航天器制导、导航与控制、自主无人系统以及计算机视觉应用。
E-mail:zhaogaopeng@njust.edu.cn

Zhao Gaopeng (Corresponding author) received his Ph. D. degree from Nanjing University of Science and Technology in 2010. He is now an associate professor at Nanjing University of Science and Technology. His main research interests include spacecraft guidance navigation and control, autonomous unmanned system, computer vision applications.