

基于多任务深度学习的铝材表面缺陷检测

沈晓海, 栗泽昊, 李敏, 徐晓龙, 张学武*

河海大学物联网工程学院, 江苏 常州 213022

摘要 针对工业铝材缺陷检测中由缺陷样本稀疏带来的训练过拟合、泛化性能差等问题, 提出一种基于多任务深度学习的铝材缺陷检测方法。先基于 Faster RCNN 设计一个包含铝材区域分割、缺陷多标签分类和缺陷目标检测的多任务深度网络模型; 再设计多任务损失层, 利用自适应权重对各项任务进行加权平衡, 解决了多项任务训练中的收敛不均衡问题。实验结果表明, 在有限的数据集支持下, 相较于单任务学习, 该方法能够在保持分割任务的均交并比 (MIoU) 指标最优的情况下, 分别提高多标签分类和缺陷目标检测的准确率, 解决了由铝材缺陷检测样本少引起的检测精度较低的问题。对于多任务应用场景, 该模型能够同时完成三个任务, 减少推断时间, 提高检测效率。

关键词 机器视觉; 工业检测; 缺陷检测; 多任务学习; 目标检测; 图像分类; 图像分割

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP57.101501

Aluminum Surface-Defect Detection Based on Multi-Task Deep Learning

Shen Xiaohai, Li Zehao, Li Min, Xu Xiaolong, Zhang Xuewu*

College of Internet of Things Engineering, Hohai University, Changzhou, Jiangsu 213022, China

Abstract In industrial aluminum defect detection, sparse defect samples always lead to the training overfit and poor generalization. This study describes a defect detection model based on multi-task deep learning. Based on Faster RCNN, a multi-task deep network model is designed, including the aluminum area segmentation, defect multi-label classification, and defect target detection. Then the multi-task loss layer is designed, and the weights are balanced by using adaptive weights to solve the problem of uneven convergence in multi-task training. Experiment results show that with the support of a limited dataset, the proposed method can improve the accuracy of multi-label classification and defect target detection while maintaining the optimal mean intersection over union (MIoU) index of the segmentation task, compared to single-task learning. The method solves the problem of low detection accuracy caused by fewer samples of aluminum defect detection. For multi-tasking application scenarios, the model can simultaneously complete three tasks, while reducing the inference time and improving the detection efficiency.

Key words machine vision; industrial inspection; defect detection; multi-task learning; object detection; image classification; image segmentation

OCIS codes 150.3040; 150.1835; 150.1135; 100.3008

1 引言

随着我国制造行业水平和人民消费水平的不断提升, 制造业对工业产品的质量有了更高的要求, 同时产品表面质量对产品的直接使用和再加工都有重要影响。目前, 铝型材行业已实现了生产环节的自动化, 但在质检环节仍依靠人工目视检测, 劳动强度大且效率低。因此, 在产品线的质量控制阶段引入机器

视觉和智能检测的方法可准确识别产品表面缺陷, 在工业流水线质检环节实现自动化具有重要意义^[1]。

缺陷检测任务通常有多种形式, 即分割、分类和检测。分割是对图像中的像素进行分类, 王森等^[2]融合了小波边缘检测和多尺度结构化森林, 提高了表面裂纹分割的精度; 张芳等^[3]基于语义分割网络 (U-Net)^[4]实现了对纳米颗粒的自动分割。分类是常见的缺陷检测任务, Zhang 等^[5]基于统计方法和光谱测

收稿日期: 2019-07-15; 修回日期: 2019-09-14; 录用日期: 2019-10-18

基金项目: 国家重点研发计划(2016YFC0401606, 2018YFC0407101)、国家自然科学基金(61671202, 61573128, 61701169)

* E-mail: lab_112@126.com

量提取了钢带缺陷的特征向量,并基于支持向量机搭建了在线分类系统;王理顺等^[6]基于深度卷积网络 GoogLeNet^[7]训练缺陷分类模型并设计了织物缺陷在线分类算法。检测是对图像中的缺陷进行定位和分类,通常难度较大,常海涛等^[8]基于目标检测框架 Faster RCNN^[9]实现了工业 CT (Computed Tomography) 图像的缺陷检测;Liong 等^[10]基于实例分割框架 Mask RCNN^[11]搭建了用于皮革的自动缺陷检测和分割系统。上述研究都是基于单一任务设计,实现了较高准确率的缺陷检测算法,但未考虑工业场景中部署深度学习时遇到的数据稀疏问题,样本数据少会引起检测模型训练过拟合,导致泛化性能差,最终影响缺陷检测的准确率。

近年来,多任务学习^[12-13]越来越受重视,可共享信息,提高泛化性能,广泛应用于深度学习各研究领域。Zhang 等^[14]基于循环卷积网络的文本分类任务中引入了多任务学习,设计了三种不同的信息共享机制,取得了较好的分类结果;Wan 等^[15]基于自动相关性来学习所有任务特征之间的相关性,以预测阿尔茨海默症神经影像测量数据的认知结果;Yuan 等^[16]提出了一种基于 Beta 过程的多任务稀疏模型,以学习动作识别字典。上述研究将多任务学习机制分别应用于文本分类、医学图像处理和动作识别等领域,取得了不错的成果,但这些方法关注的目标任务与工业图像缺陷检测不同,其算法流程无法直接应用于缺陷检测任务,对多任务学习机制在缺陷检测方向上的应用仍需探索。

工业缺陷数据集标注成本高导致训练数据量较少,少量数据在训练单一缺陷检测任务时,检测模型复杂而样本稀少,容易发生过拟合,造成泛化性能低,因此在测试时检测准确率不高。针对这一问题,本文提出一个基于多任务深度学习的缺陷检测方法,构建多任务网络模型,引入铝材金属表面分割任务、缺陷图像多标签分类任务和铝材缺陷目标检测任务,共享多个任务间的特征提取网络参数,利用任务间的互相约束来降低过拟合风险,从而提高缺陷检测的精度和效率,最后使用铝材缺陷数据集进行实验验证。

2 基于多任务深度学习的铝材缺陷检测

2.1 研究的问题

研究的问题是利利用多任务深度学习机制检测铝材缺陷,其中包含铝材区域分割任务、缺陷图像分类

任务和铝材图像缺陷检测任务。

铝材的缺陷检测通常包括分类任务和检测任务。分类任务要解决图像包含哪些缺陷的问题,一幅图像可能包含多种缺陷,所以铝材缺陷图像分类属于多标签分类。检测任务要解决图像中缺陷的分类和定位问题,属于目标检测范畴。除此之外,由于光照不均和拍摄背景复杂,铝材缺陷图像在检测过程中易受背景干扰,因此增加铝材区域分割任务,提取的铝材区域可用于排除区域外的检测目标,提高准确率。

受数据集数目的限制,单一任务下的训练模型常由于过拟合而在测试集上无法达到较高的检测性能,而多任务学习是同时学习多个相关任务的过程,旨在利用任务间的相关性来提高泛化性能,共同学习这些监督任务,可在学习过程中共享特征信息,平衡各自任务的噪声差异,每项学习任务可对其他任务进行约束,从而降低了单任务过拟合的风险,最终提升全部或部分学习任务的性能^[17]。多任务学习机制如图 1 所示^[12]。

2.2 模型结构

为利用多任务学习机制来提升缺陷检测精度,基于 2.1 节三项任务,设计了一种用于铝材缺陷检测的多任务深度神经网络模型,如图 2 所示。

设计的神经网络模型基于二阶目标检测架构 Faster RCNN^[9]进行拓展。模型由四部分组成,一个用于共享提取特征的主干网络和分别用于金属区域分割、多标签分类和缺陷目标检测的三个单任务分支网络。本节主要介绍特征提取主干网络。

铝材图像缺陷的检测主要是纹理特征等低语义层次信息上的处理,包含的语义信息层次较低,所以不需要较深的网络进行高层次语义特征的提取,因此选用深度残差网络(ResNet50)^[18]作为检测网络的特征提取部分。ResNet50 共包含五个卷积阶段,即 conv0~conv4,每个卷积阶段的输出作为特征图,每个阶段的特征图经下采样缩小为输入图像尺寸的 1/2,以增大卷积核覆盖的特征图感受野,对图像进行更大范围的特征提取。考虑到铝材图像的纹理信息在特征图下采样过程中易丢失,实验采用特征金字塔网络(FPN)^[19]对特征提取网络进行改进,将最高特征层(conv4)的特征图进行 8 倍,4 倍,2 倍和 1 倍的上采样,并分别与 conv1~conv4 进行拼接融合,得到多个尺度的特征图,分别记为 fpn1, fpn2, fpn3 和 fpn4,每个尺度下的特征图既能保留卷积得到的底层特征,又能融合顶层特征,能够有效避免由下采样引起的信息丢失问题。

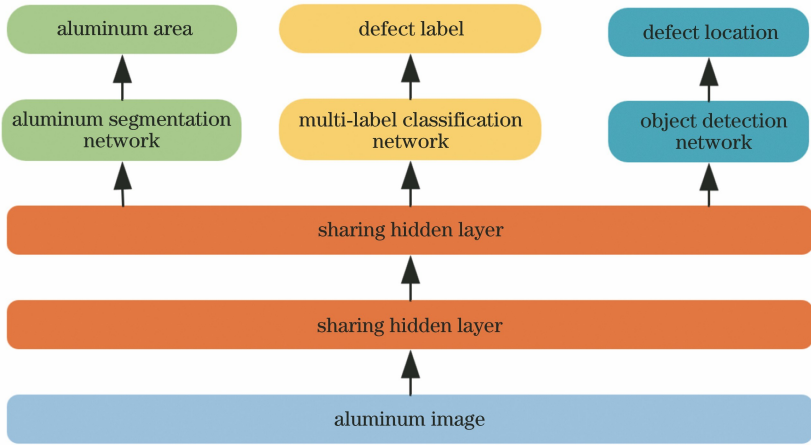


图1 多任务深度学习中的参数共享

Fig. 1 Parameters sharing in multi-task deep learning

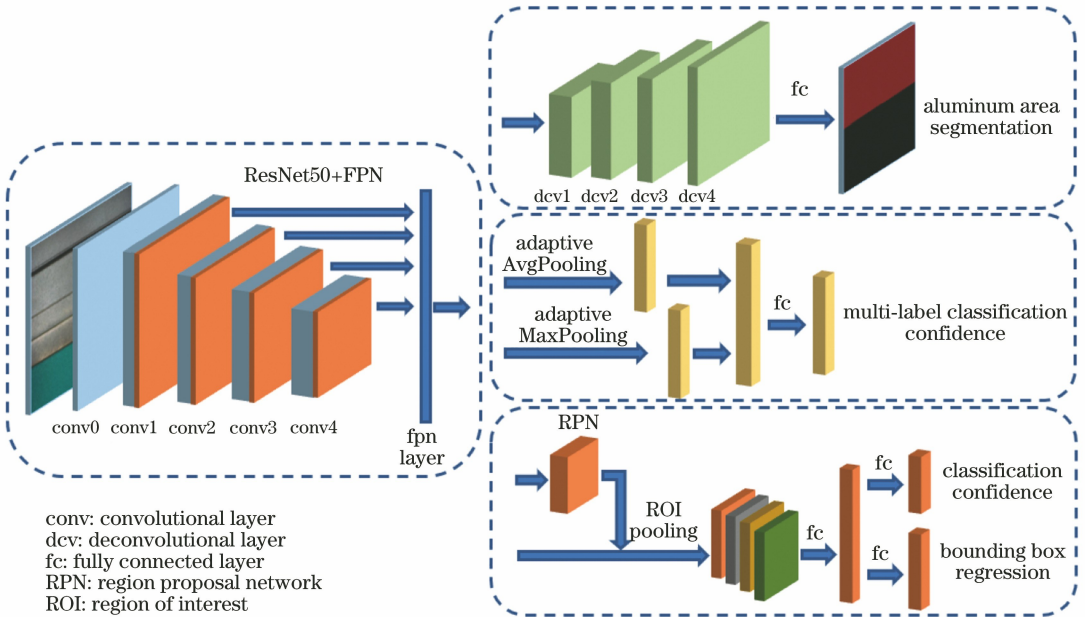


图2 铝材缺陷检测的多任务深度神经网络模型

Fig. 2 Multi-task deep neural network model of aluminum defect detection

特征提取主干网络由金属区域分割、多标签分类和缺陷目标检测三个单项任务网络所共享,该部分网络输入为铝材图像,输出为四个不同尺度的特征张量,即 $f_{pn1} \sim f_{pn4}$,多尺度的特征张量作为后续三项任务的输入特征。

2.3 金属区域分割网络

金属区域分割属于分割任务,金属区域分割任务的训练可增加特征提取网络对铝材前景和背景的区分能力,从而有助于提升特征提取网络在分类任务和检测任务上的性能。金属区域提取网络基于全卷积网络(FCN)^[20]架构设计,主要由输入、上采样和像素分类三部分构成。输入选择从特征提取网络

得到的特征图 f_{pn3} ,经连续多次上采样,每次上采样包含步长为 2 的反卷积、ReLU 激活函数和批归一化处理,得到尺寸与输入图像一致,通道数为 256 的多通道张量。再进行 1×1 卷积,将输出通道数设为 2 并采用 Sigmoid 函数进行激活,最后每个像素选取通道下标的最大值作为分类,得到一幅包含前景和背景的分割图像。

网络训练过程中将二分类交叉熵作为损失函数,公式为

$$L_{seg} = -\frac{1}{N} \sum_{i=0}^N y_i \exp(\hat{y}_i) + (1 - y_i) \exp(1 - \hat{y}_i), \quad (1)$$

式中： \hat{y}_i 和 y_i 分别为每个像素的预测分类和实际分类； N 为图像的像素数目。

2.4 多标签分类网络

多标签分类网络是指将主干网络提取的特征向量进行分类,判断图像中包含哪些种类的缺陷。该网络能够训练主干网络对铝材图像全局特征的感知,这些特征向量中既包含铝材前景和背景信息,又包含缺陷特征信息,从而有助于提升特征提取网络在分割任务和检测任务中的效果。采用多标签分类网络的输入选取特征提取网络得到的特征图 f_{pn4} ,分别经自适应最大池化层和自适应平均池化层得到两个一维张量。其中,自适应最大池化层用于捕获特征图中的显著性缺陷区域,自适应平均池化层用于捕获特征图中的大面积缺陷区域。再对两个张量进行拼接融合,经全连接层和 Sigmoid 激活函数得到每个类别的置信度,选取置信度大于 0.5 的类别作为分类结果。

多标签分类网络在训练过程中,计算每个样本在各个标签的损失,从而将多标签的损失计算转换成了每个标签二分类问题的损失计算,所以同样使用二分类交叉熵作为损失函数,公式为

$$L_{\text{label}} = -\frac{1}{M} \sum_{j=0}^M y_j \exp(\hat{y}_j) + (1 - y_j) \exp(1 - \hat{y}_j), \quad (2)$$

式中： \hat{y}_j 和 y_j 分别为每个标签的预测值与实际值； M 为数据集缺陷分类数。

2.5 缺陷目标检测网络

缺陷目标检测网络是指将主干网络提取的特征层进行目标检测,得到图像中缺陷的位置和分类。该网络能够训练主干网络对铝材图像局部特征的感知,从而在多标签分类任务训练过程中,提供包含局部缺陷特征的梯度信息,有助于提升分类效果。

缺陷目标检测网络基于 Faster RCNN 而设计,使用二阶(two-stage)检测框架对融合多尺度特征进行检测,采用特征提取网络得到的特征图 $f_{pn1} \sim f_{pn4}$ 作为输入特征,经区域推荐网络(RPN)计算预设锚框(anchor box)的前景置信度,通过非极大值抑制(NMS)算法筛选得到一定数目的锚框,将其对应的推荐区域特征图作为待检测对象。由于锚框的尺度存在差异,因此所有推荐区域特征图需经过感兴趣区域(ROI)池化层池化为统一维度的张量。对每个推荐区域特征图进行目标位置的回归预测和目标分类的概率预测,最终得到图像中每个缺陷的类

别和位置。

在缺陷检测任务中,沿用 Faster RCNN 中的损失函数^[9],计算公式如下:

$$L_{\text{obj_cls}} = \frac{1}{N_{\text{obj}}} \sum_i B_{\text{CE}}(p_i, p_i^*), \quad (3)$$

$$L_{\text{obj_reg}} = \frac{1}{N_{\text{obj}}} \sum_i p_i^* s_{L_1}(t_i, t_i^*), \quad (4)$$

$$L_{\text{rpn_cls}} = \frac{1}{N_{\text{rpn}}} \sum_j B_{\text{CE}}(p_j, p_j^*), \quad (5)$$

$$L_{\text{rpn_reg}} = \frac{1}{N_{\text{rpn}}} \sum_i s_{L_1}(t_i, t_i^*), \quad (6)$$

$$L_{\text{obj}} = L_{\text{rpn_cls}} + L_{\text{rpn_reg}} + L_{\text{obj_cls}} + L_{\text{obj_reg}}, \quad (7)$$

式中： $L_{\text{rpn_cls}}$ 和 $L_{\text{rpn_reg}}$ 分别为区域推荐网络的前景分类损失和位置回归损失； $L_{\text{obj_cls}}$ 和 $L_{\text{obj_reg}}$ 分别为预测目标的分类损失和位置回归损失； N_{rpn} 为筛选前的锚框数目； p_j 和 p_j^* 分别为第 j 个锚框的预测前景置信度和实际前景置信度； N_{obj} 为筛选后的锚框数目； p_i 和 p_i^* 分别为第 i 个锚框的预测缺陷类别概率和实际缺陷类别概率； t_i 和 t_i^* 分别为第 i 个锚框预测的位置和实际位置； B_{CE} 为二分类交叉熵损失函数； s_{L_1} 为改进的 L_1 范数(Smooth L_1)。

2.6 自适应多任务损失层设计

多任务学习中需对各个任务的损失进行求和,再反向传播更新网络权重。在大多数多任务应用研究中^[21-23],多任务损失都是对各项任务进行简单的加权求和,各项任务的权重是均等的或是人为设定的。但实际训练发现,某些任务易占梯度下降算法的主导地位,无法充分优化其他任务,大大降低了多任务学习的优化效果。为此,Cipolla 等^[24]利用任务之间的同方差不确定性,引入任务噪声 σ ,将任务损失视为网络参数 W 对应的函数,并基于高斯过程对每项任务损失 $L(W)$ 进行正则化约束。单项任务的改进损失函数为

$$L_s = \frac{L(W)}{2\sigma^2} + \log_2 \sigma. \quad (8)$$

在设计的网络模型中,三项任务的损失分别为 L_{seg} 、 L_{label} 和 L_{obj} ,基于上述 Cipolla 等^[24]的研究,引入噪声 σ_1 、 σ_2 和 σ_3 ,按照(8)式计算总损失函数,公式为

$$L_{\text{total}} = \frac{L_{\text{seg}}}{2\sigma_1^2} + \frac{L_{\text{label}}}{2\sigma_2^2} + \frac{L_{\text{obj}}}{2\sigma_3^2} + \log_2 \sigma_1 + \log_2 \sigma_2 + \log_2 \sigma_3. \quad (9)$$

为进行权重 $1/2\sigma^2$ 和惩罚项 $\log_2 \sigma$ 的学习,设计了一个多任务损失层,将三项任务的输出作为损失

层的输入,将 $1/2\sigma^2$ 和 $\log_2\sigma$ 作为网络参数,多任务损失层如图 3 所示。

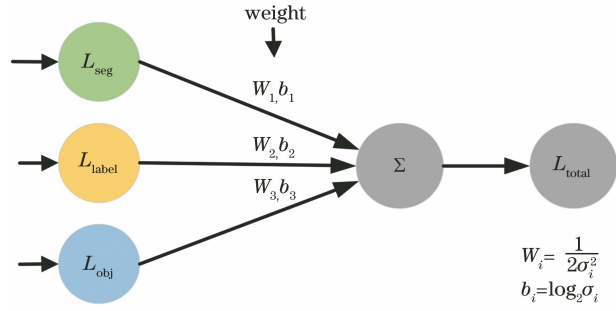


图 3 自适应多任务损失层

Fig. 3 Adaptive multi-task loss layer

3 实验与结果

3.1 软硬件环境

实验中,硬件配置:CPU 为 Intel Xeon E5-2680,主频为 3.3 GHz,内存为 128 GB,显卡为

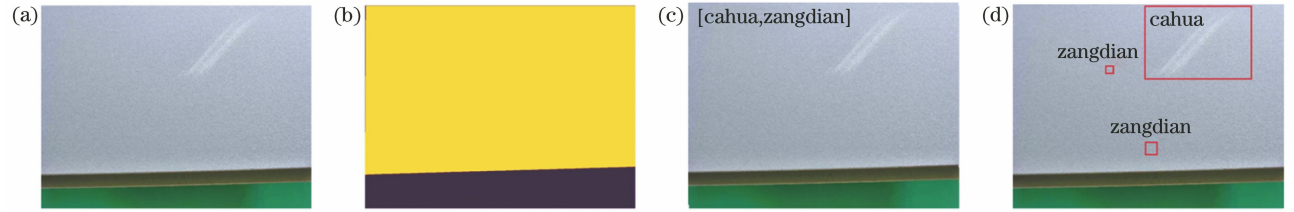


图 4 铝材缺陷图像和真值图。(a)原图;(b)分割标注;(c)多标签标注;(d)缺陷目标标注

Fig. 4 Aluminum defect image and ground truth. (a) Original image; (b) segmentation marked; (c) multi-label marked; (d) defect object marked

原始图像尺寸较大,受显存限制不能直接作为网络模型的输入。实验采用图像重叠分块的方式重新构建数据集:分块的尺寸设置为 $1024 \text{ pixel} \times 1024 \text{ pixel}$,横向和纵向的重叠率分别为 25.0% 和 12.5%,分割后得到约 18000 幅图像。实验中将数据集转换为 Pascal VOC 格式,对数据集进行随机划分,选取 4500 幅图像作为测试集,13500 幅图像作为训练集。此外,由于数据集类别不均衡,少数类图像的采样频率远小于多数类图像,影响小类别的检测精度,因此对训练集进行过采样处理,为少数类图像创建副本使其达到与多数类相同的样本数量。

3.3 评价指标

对于铝材区域分割任务,选用图像语义分割领域中最常用的评价指标,平均像素精度 (MPA, x_{MPA}) 和均交并比 (MIoU, x_{MIoU}),来衡量分割任务的精确度^[27]。MPA 和 MIoU 的计算公式分别为

$$x_{\text{MPA}} = \frac{1}{N_{\text{sample}}} \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij}}, \quad (10)$$

NVIDIA TITAN V,显存为 12 G;软件配置:系统为 Ubuntu 16.04,显卡驱动为 NVIDIA Driver 390.77 和 CUDA 9.0,深度学习训练框架为 PyTorch-1.0,数据标注工具为 LabelMe,图像处理工具为 OpenCV 3.4,编程语言为 Python 3.x,在 maskrcnn-benchmark^[25] 上进行网络模型的搭建和测试。

3.2 实验数据

数据集来自阿里云承办的 2018 年广东工业智造大数据创新大赛-铝型材表面瑕疵识别^[26],数据为广东省佛山市南海区某铝型材企业流水线上采集的铝型材图像。其中缺陷样本数量约为 3000 幅,图像尺寸为 $2560 \text{ pixel} \times 1920 \text{ pixel}$ 。数据集包含不导电、擦花、角位漏底、桔皮、漏底、喷流、漆泡、起坑、杂色和脏点共 10 类表面缺陷。对缺陷图像进行铝材区域标注、多标签分类标注和缺陷目标标注,得到三个用于单项任务的数据集。数据集中原始图像、铝材分割图像、多标签标注图像和缺陷目标标注图像,如图 4 所示。

$$x_{\text{MIoU}} = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}}, \quad (11)$$

式中: N_{sample} 为样本数目; k 为分类数目; p_{ij} 为本属于分类 i 但被预测为分类 j 的像素数目; p_{ji} 为本属于分类 j 但被预测为分类 i 的像素数目; p_{ii} 为预测正确的像素数目。

对于多标签分类任务,选用汉明损失 (h_{loss}) 衡量分类任务的精确度^[28],该指标衡量了预测标签与样本真实标签间的不一致程度。汉明损失的计算公式为

$$h_{\text{loss}} = \frac{1}{N_{\text{sample}} L} \sum_{i=1}^{N_{\text{sample}}} \sum_{j=1}^L X_{\text{or}}(Y_{ij}, P_{ij}), \quad (12)$$

式中: L 为标签数目; Y_{ij} 和 P_{ij} 分别为在第 i 个样本上第 j 类缺陷的实际置信度和预测置信度; X_{or} 为异或,即不同为 1,相同为 0。

对于缺陷检测任务,选用目标检测领域中常用的平均精度 (mAP, x_{mAP}) 作为检测任务的评价指标^[9],实验中利用 $x_{\text{MIoU}} > 0.5$ 的有效检测框计算精

度。mAP 的计算公式为

$$x_{\text{mAP}} = \frac{\sum_{q=1}^Q A_P(q)}{Q}, \quad (13)$$

式中: Q 为缺陷类别数目; $A_P(q)$ 为类别 q 的平均精度。

在上述指标中, MPA 值和 MIoU 值越大说明分割的效果越好, 汉明损失值越小说明多标签分类的准确率越高, mAP 值越大说明缺陷检测的准确率越高。

3.4 实验设置和训练过程

为了充分验证所提方法的准确性, 进行了两组: 1) 单任务与多任务对比; 2) 多任务网络与其他方法对比。实验 1) 依照三项任务的多任务组合情况, 分别设置三个单任务、三个双任务和一个三任务, 用于验证在提出的网络框架下多项任务共同训练后, 相比于单任务单独训练, 能否提高其中某项或多项任务的检测准确率。实验 2) 选取分割、分类和目标检测领域先进的深度学习方法进行, 将每项任务的训练结果与多任务学习训练后的结果进行对比, 用于验证所提方法相比于其他深度学习方法, 是否能

够提高铝材缺陷检测的准确率。

对于所有实验, 均在单块图形处理器上进行训练, 批量大小 (Batch Size) 设置为 2, 训练总迭代次数为 2×10^4 , 初始学习率设为 1.25×10^{-3} , 衰减因子 $\gamma = 0.9$, 并在迭代次数达到 5000、8000 和 12000 时, 对学习率进行更新。训练时优化方法采用随机梯度下降 (SGD), 动量因子 $\mu = 0.9$, 权重衰减系数 $\omega = 0.0001$ 。训练过程中, 采用 2.3~2.5 节中各项任务的训练损失进行收敛状态的跟踪, 训练完成时采用 3.3 节中各项任务的评价指标对训练模型进行性能评估。

对于多任务训练, 将任务权重作为网络参数进行学习, 对每项任务的损失进行加权求和作为总网络损失, 梯度反向传播时对权重进行更新, 训练完成后记录各项任务的权重。训练时为避免权重 $12/\sigma^2$ 的分母为零, 将 $\log 2\sigma$ 作为训练变量进行预测, 均初始化为 0。

3.5 实验结果和分析

实验 1) 中各实验训练完成后, 在测试集上进行测试, 实验结果如表 1 所示, 其中 \checkmark 表示该项任务学习时使用自适应权重。

表 1 单任务训练与多任务训练在测试集上的表现

Table 1 Performance of single task training and multi-task training on test sets

Task	Task weight			MIoU(MPA) / %	$h_{\text{loss}} / \%$	mAP / %
	Seg.	Clas.	Obj.			
Segmentation	1	0	0	98.59(99.57)	—	—
Classification	0	1	0	—	5.24	—
Object detection	0	0	1	—	—	71.46
Seg. + Clas.	\checkmark	\checkmark		98.43(99.41)	4.13	—
Seg. + Obj.	\checkmark		\checkmark	98.82(99.73)	—	74.72
Clas. + Obj.		\checkmark	\checkmark	—	1.01	75.97
Seg. + Clas. + Obj.	\checkmark	\checkmark	\checkmark	98.57(99.48)	0.96	75.65

从表 1 可以看到, 相比于单任务训练结果, 每项任务在参与多任务学习后均获得了不同程度的精度提升, 其中多标签分类任务和缺陷检测任务的准确度提升较大。铝材区域分割任务的衡量指标 MIoU 和 MPA 在单任务学习和多任务学习中变化不大, 对此解释如下: 1) 铝材区域分割任务相对简单, 均已达到 99% 的准确率, 属于可接受范围; 2) 分割任务精度与 100% 的误差在 0.5% 左右, 这可能来自数据集的标注误差。采用相同的网络架构和训练参数, 多标签分类任务和缺陷检测任务在参与多任务训练后, 性能指标均相较其在单任务训练后的指标高 2~5 个百分点, 原因在于多任务训练中, 多个任务间共同使用一个特征提取网络, 增加了对底层特

征提取的约束, 能够在有限样本数目的限制下避免学习与目标域无关的特征, 大大降低了过拟合风险, 从而增加了这些任务的检测性能。这些任务中, 铝材区域分割任务提供了铝材前景的约束, 多标签分类任务提供了图像级别的特征提取信息, 同时缺陷目标检测任务提供了图像局部细节的有效信息, 三者共享信息从而提高了检测精度。

除此之外, 实验中发现各项任务选取不同层次的特征图对多任务训练结果有影响。举例来说, 在铝材分割和多标签分类的双任务中, 分割网络输入特征图选取 f_{p1} 相比选取 f_{p4} , 训练的多标签分类模型效果更优。这是由于对于较为简单的分割任务, 浅层特征图能够满足分割网络的需求, 而分类任

务则需要高层特征图,如果将高层特征图同样用于分割任务,则会引起两者训练特征提取选择的冲突,导致分类任务性能降低。这一现象说明多任务学习在共享网络参数相互约束减小过拟合的同时,也会出现相互制约导致其他任务欠拟合的风险,因此在设计多任务网络模型时,仍需要考虑两个问题:1)网络容量是否足够各项任务同时进行特征提取;2)各项任务需选择哪些特征层进行共享。

选取一组测试集图像的输出结果进行分析,如图5所示。从图5可以看到,提出的方法可以非常

好地分割出铝型材区域;对于缺陷目标检测任务,模型可以很好地检测出大型缺陷,如擦花和不导电,也能够检测出大部分微小瑕疵,如脏点,但仍存在部分不足,有部分脏点未被检出,有部分正常区域被预测为脏点,导致准确率下降;对于缺陷多标签检测任务,模型均能够准确预测。由此可见,在参与所提的多任务模型训练后,多任务的准确率有所提高,对于目标检测任务,尽管提升了准确率,但未能解决缺陷检测上的一些难点,如小瑕疵的准确检出,仍需在目标检测网络上针对这类难点作出改进。

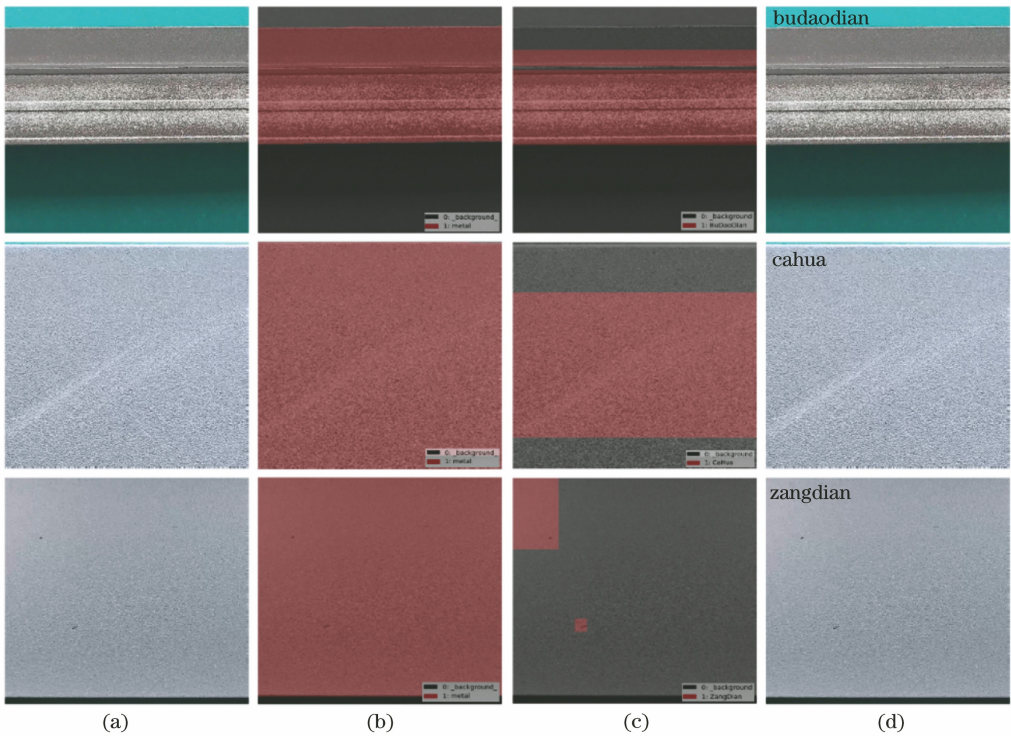


图5 测试集上的模型输出结果。(a)原图;(b)分割;(c)缺陷检测;(d)缺陷多标签检测

Fig. 5 Results of model output on test sets. (a) Original images; (b) segmentation; (c) defect detection; (d) defect multi-label detection

实验2)将所提方法与目前先进的深度网络进行对比测试,对比方法分别为,铝材区域提取任务选用FCN系列的分支FCN-8s^[20]和U-Net^[4]作为参照;多标签分类任务选用深度残差网络(ResNet)系列的ResNet50^[18]和ResNet101^[18]作为主干网络,并加以全连接层进行分类;铝材缺陷目标检测任务则选用Faster RCNN^[9]和密集检测网络(RetinaNet)^[29]作为参照。各实验在训练集上训练完成后,在测试集上进行测试,实验结果如表2所示。

从表2可以看到,所提模型在铝材缺陷数据集上的工作效果优于其他深度学习方法。其中,金属区域分割网络与效果最优的FCN-8s方法取得了相似结果;多标签分类网络相较于仅采用ResNet50

表2 多种深度学习方法在测试集上的表现

Table 2 Performance of several deep learning methods on test sets

Method	MIoU (MPA) /%	h_{loss} /%	mAP /%
FCN-8s	98.75(99.49)	—	—
U-Net	97.24(98.80)	—	—
ResNet50	—	6.41	—
ResNet101	—	4.67	—
Faster RCNN	—	—	69.78
RetinaNet	—	—	72.13
Proposed	98.57(99.48)	0.96	75.65

和ResNet101的方式, h_{loss} 值减小了5.45个百分点和3.71个百分点,分类效果得到显著提升;缺陷目标检测网络相较于原始的Faster RCNN和

RetinaNet, mAP 值提高了 5.87 个百分点和 3.52 个百分点,可以看出检测效果有了较大改善。

多任务模型中,由于多个任务共享了特征提取的隐藏层,即多项任务只需进行一次图像特征提取,大大地节省时间开销,提高效率。为进行验证,对实验 1) 的各组合实验在测试集上统计推断时间。在每个模型下,对测试集进行推断,统计测试集所有图像的完成时间,除以测试集图像数目即为单张图像的推断时间,进行多轮测试计算平均值,得到该模型下的平均推断时间。对于多任务下的推断时间,对各项任务时间求和作为参照,结果如表 3 所示。

表 3 单任务模型与多任务模型的推断时间统计

Table 3 Inference time statics on single task model and multi-task model

Task	Average inference time per image /ms	Sum of single task inference time /ms
Segmentation	14.960	—
Classification	13.791	—
Object detection	57.347	—
Seg. + Clas.	15.437	28.751
Seg. + Obj.	68.212	72.307
Clas. + Obj.	66.025	71.138
Seg. + Clas. + Obj.	69.435	86.098

从表 3 可以看到,多任务模型的推断时间均小于各项任务推断时间之和,说明在多任务的推断场景中,多任务模型可提高检测效率。考虑到实验所用的高性能图形处理器成本和近 70 ms 的单张图像推断时间,该方法无法直接应用于实际工业检测场景,若实际应用需同时使用模型压缩技术对训练的模型进行压缩以减少参数量和硬件成本,并结合在线推断技术搭建实时的缺陷检测系统以提升检测效率。但在单任务的应用场景中,可选择合适的辅助任务,搭建多任务模型进行训练优化,再仅对目标任务进行推断,能够在不影响推断速度的基础上提高检测精度。

4 结 论

提出一种基于多任务深度学习的铝材缺陷检测框架,将铝材区域分割任务、缺陷多标签分类任务和缺陷目标检测任务引入检测框架中,并在训练时利用自适应权重对各项任务损失进行加权计算,利用多任务学习共享任务间信息的优势对各项任务进行优化。实验结果表明,相比于单任务学习,所提方法参与多任务训练后,在保持分割任务 MIoU 指标最

优的情况下,多标签分类的准确率和缺陷检测的准确率均有较大提高。所提方法可应用于其他工业产品缺陷检测领域,针对单任务应用场景,可利用多任务学习选取最合适的辅助学习任务,以较少的标注成本来达到较优的模型精度。目前,对于各项任务共享特征层的选取,仍需通过实验验证,这是所提方法存在的缺点。今后的工作将考虑如何将特征共享的选择作为优化问题,引入检测框架中。

参 考 文 献

- [1] Tang B, Kong J Y, Wu S Q. Review of surface defect detection based on machine vision[J]. Journal of Image and Graphics, 2017, 22(12): 1640-1663.
汤勃, 孔建益, 伍世虔. 机器视觉表面缺陷检测综述[J]. 中国图象图形学报, 2017, 22(12): 1640-1663.
- [2] Wang S, Wu X, Zhang Y H, et al. Surface crack segmentation based on multi-scale wavelet transform and structured forest[J]. Acta Optica Sinica, 2018, 38(8): 0815024.
王森, 伍星, 张印辉, 等. 基于多尺度小波变换和结构化森林的表面裂纹分割[J]. 光学学报, 2018, 38(8): 0815024.
- [3] Zhang F, Wu Y, Xiao Z T, et al. Nanoparticle segmentation based on U-Net convolutional neural network [J]. Laser & Optoelectronics Progress, 2019, 56(6): 061005.
张芳, 吴玥, 肖志涛, 等. 基于 U-Net 卷积神经网络的纳米颗粒分割[J]. 激光与光电子学进展, 2019, 56(6): 061005.
- [4] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[M] // Navab N, Hornegger J, Wells W, et al. Medical image computing and computer-assisted intervention. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.
- [5] Zhang X W, Ding Y Q, Yan-Yun L, et al. A vision inspection system for the surface defects of strongly reflected metal based on multi-class SVM[J]. Expert Systems with Applications, 2011, 38(5): 5930-5939.
- [6] Wang L S, Zhong Y, Li Z D, et al. On-line fabric defect recognition algorithm based on deep learning [J]. Journal of Computer Applications, 2019, 39(7): 2125-2128.
王理顺, 钟勇, 李振东, 等. 基于深度学习的织物缺陷在线检测算法[J]. 计算机应用, 2019, 39(7): 2125-2128.

- [7] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston. New York: IEEE, 2015: 1552-3970.
- [8] Chang H T, Gou J N, Li X M. Application of Faster R-CNN in image defect detection of industrial CT[J]. Journal of Image and Graphics, 2018, 23(7): 1061-1071.
常海涛, 苟军年, 李晓梅. Faster R-CNN 在工业 CT 图像缺陷检测中的应用[J]. 中国图象图形学报, 2018, 23(7): 1061-1071.
- [9] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C] // Advances in neural information processing systems, December 7-12, 2015, Montreal, Quebec, Canada. New York: Curran Associates., 2015: 91-99.
- [10] Liong S T, Gan Y, Huang Y C, et al. Automatic defect segmentation on leather with deep learning [EB/OL]. (2019-03-28) [2019-10-18]. <https://arxiv.org/abs/1903.12139>.
- [11] He K, Gkioxari G, Dollár P, et al. Mask R-CNN [C] // Proceedings of the IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 2961-2969.
- [12] Ruder S. An overview of multi-task learning in deep neural networks [EB/OL]. (2017-01-15) [2019-10-18]. <https://arxiv.org/abs/1706.05098>.
- [13] Zhang Y, Yang Q. A survey on multi-task learning [EB/OL]. (2018-07-27) [2019-10-17]. <https://arxiv.xilesou.top/abs/1707.08114>.
- [14] Zhang H L, Xiao L Q, Wang Y K, et al. A generalized recurrent neural architecture for text classification with multi-task learning [C] // Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, August 19-26, 2017, Melbourne, Australia. California: IJC AI, 2017: 3385-3391.
- [15] Wan J, Zhang Z L, Yan J W, et al. Sparse Bayesian multi-task learning for predicting cognitive outcomes from neuroimaging measures in Alzheimer's disease [C] // 2012 IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2012, Providence, RI. New York: IEEE, 2012: 940-947.
- [16] Yuan C F, Hu W M, Tian G D, et al. Multi-task sparse learning with beta process prior for action recognition[C] // 2013 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2013, Portland, OR, USA. New York: IEEE, 2013: 423-429.
- [17] Zhang Y, Yang Q. A survey on multi-task learning [EB/OL]. (2018-07-27) [2019-10-18]. <https://arxiv.org/abs/1707.08114>.
- [18] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.
- [19] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 2117-2125.
- [20] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 3431-3440.
- [21] Gibert X, Patel V M, Chellappa R. Deep multitask learning for railway track inspection [J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(1): 153-164.
- [22] Yu C M, Li H N, An L. Analysis of text emotion cause based on multi-task deep learning [J]. Journal of Guangxi Normal University (Natural Science Edition), 2019, 37(1): 50-61.
余传明, 李浩男, 安璐. 基于多任务深度学习的文本情感原因分析[J]. 广西师范大学学报(自然科学版), 2019, 37(1): 50-61.
- [23] Yang H X, Chen F, Gan W F. Deep face recognition algorithm based on multitask learning [J]. Laser & Optoelectronics Progress, 2019, 56(18): 181005.
杨恢先, 陈凡, 甘伟发. 基于多任务学习的深层人脸识别算法[J]. 激光与光电子学进展, 2019, 56(18): 181005.
- [24] Cipolla R, Gal Y, Kendall A. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 7482-7491.
- [25] GitHub, Inc. Faster R-CNN and Mask R-CNN in PyTorch 1.0 [EB/OL]. (2018) [2019-10-18]. <https://github.com/facebookresearch/maskrcnn-benchmark>.

- [26] 2018 Guangdong industrial intelligence big data innovation competition- competition and data[EB/OL]. (2018-11-16) [2019-10-18]. <https://tianchi.aliyun.com/competition/entrance/231682/information>.
2018 广东工业智造大数据创新大赛—赛题与数据[EB/OL]. (2018-11-16) [2019-10-18]. <https://tianchi.aliyun.com/competition/entrance/231682/information>.
- [27] Garcia-Garcia A, Orts-Escolano S, Oprea S, et al. A review on deep learning techniques applied to semantic segmentation [EB/OL]. (2017-04-22) [2019-10-18]. <https://arxiv.org/abs/1704.06857>.
- [28] Wu X Z, Zhou Z H. A unified view of multi-label performance measures[C] // Proceedings of the 34th International Conference on Machine Learning, August 6-11, 2017, Sydney, NSW, Australia. Massachusetts: MIT Press, 2017: 3780-3788.
- [29] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice. New York: IEEE, 2017: 2980-2988.