



计算机应用
Journal of Computer Applications
ISSN 1001-9081, CN 51-1307/TP

《计算机应用》网络首发论文

题目：基于深度对比网络的印刷缺陷检测方法
作者：王佑芯，陈斌
收稿日期：2021-11-13
网络首发日期：2022-06-20
引用格式：王佑芯，陈斌. 基于深度对比网络的印刷缺陷检测方法[J/OL]. 计算机应用. <https://kns.cnki.net/kcms/detail/51.1307.TP.20220617.1629.004.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

基于深度对比网络的印刷缺陷检测方法

王佑芯^{1,2}, 陈斌^{2,3,4*}

(1.中国科学院 成都计算机应用研究所, 成都 610041;

2.中国科学院大学 计算机科学与技术学院, 北京 100049;

3.哈尔滨工业大学(深圳) 国际人工智能研究院, 广东 深圳 518055;

4.哈尔滨工业大学 重庆研究院, 重庆 401100)

(*通信作者电子邮箱 chenbin2020@hit.edu.cn)

摘要: 针对基于传统图像处理技术的印刷缺陷检测方法鲁棒性差与基于深度学习的目标检测方法不完全适用于印刷缺陷检测任务的问题, 将模板匹配方法中的对比思想与深度学习中的语义特征结合, 提出了用于印刷缺陷检测任务的深度对比网络 (CoNet)。首先, 提出了采用孪生结构的深度对比模块 (DCM), 通过在语义空间提取并融合检测图像与参考图像的特征图挖掘二者的语义关系。然后, 基于非对称的双通路特征金字塔结构, 提出了多尺度变化检测模块 (MsCDM), 用于定位并识别印刷缺陷。在公开的印刷电路板缺陷数据集上, CoNet 的平均精度 (mAP) 为 99.1%, 与模板匹配方法和 Faster R-CNN 相比分别提升了 9.8 个百分点和 1.5 个百分点; 与同样采用变化检测思路的两个基线模型最大分组金字塔池化 (MP-GPP) 和变化检测单次检测器 (CD-SSD) 相比, 分别提升了 0.5 个百分点和 0.8 个百分点。在更复杂的立金缺陷数据集上, 采用相同实验设定得出, CoNet 的 mAP 为 69.8%, 不仅比模板匹配方法和 Faster R-CNN 的 mAP 分别高了 12 个百分点和 5.3 个百分点, 而且相较 MP-GPP 和 CD-SSD 也分别提升 3.5 个百分点和 2.4 个百分点。此外, 当输入图像分辨率为 640×640 时, CoNet 的平均耗时为 35.7ms, 完全可以满足工业检测任务的实时性需求。

关键词: 印刷缺陷检测; 深度学习; 孪生卷积神经网络; 特征金字塔; 变化检测

中图分类号: TP391.4

文献标志码: A

Print defects detection method based on deep comparison network

WANG Youxin^{1,2}, CHEN Bin^{2,3,4*}

(1. Chengdu Institute of Computer Application, Chinese Academy of Sciences, Chengdu Sichuan 610041, China;

2. School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China;

3. International Institute for Artificial Intelligence, Harbin Institute of Technology (Shenzhen), Shenzhen Guangdong 518055, China;

4. Chongqing Research Institute, Harbin Institute of Technology, Chongqing 401100, China)

Abstract: In order to solve the problems of poor robustness of the print defects detection methods based on traditional image processing technology and the object detection methods based on deep learning are not completely suitable for the detection tasks of print defects, the comparison ideas in template matching method were combined with semantic features in deep learning, the Deep Comparison Network (CoNet) could be used for the detection tasks of print defects was proposed. Firstly, the Deep Comparison Module (DCM) based on twin structure was proposed to mine the semantic relationship between the detection image and the reference image through extracting and fusing the feature maps of them in the semantic space. Then, based on the feature pyramid structure composed of asymmetric double channels, the Multi-scale Change Detection Module (MsCDM) was proposed to locate and classify print defects. On the published dataset of print circuit board defects, the mean Average Precision (mAP) of CoNet is 99.1%, which is increased by 9.8 percentage points and 1.5 percentage points respectively compared with the template matching method and Faster Region-based Convolutional Neural Network (Faster R-CNN). Compared with the two baseline models Max-Pooling Group Pyramid Pooling (MP-GPP) and Change-Detection Single Shot Detector (CD-SSD), which also adopt the idea of change detection, the mAP of the proposed method is increased by 0.5 percentage points and 0.8 percentage points respectively. With the same experimental setting, the mAP of CoNet is 69.8% on a more complex dataset of Lijin defects, which is not only 12 percentage points and 5.3 percentage points higher than that of template matching method and Faster R-CNN, but also 3.5 percentage points and 2.4 percentage points higher

收稿日期: 2021-11-13; 修回日期: 2022-03-13; 录用日期: 2022-04-01。

作者简介: 王佑芯(1997—), 男, 江西宜春人, 硕士研究生, 主要研究方向: 工业缺陷检测、目标检测、视频多模态表征; 陈斌(1970—), 男, 四川广汉人, 研究员, 博士, CCF 会员, 主要研究方向: 工业检测、深度学习。

than that of MP-GPP and CD-SSD respectively. Besides, when the resolution of input image is 640×640 , the average time consumption of CoNet is 35.7ms, which can absolutely meet the real-time requirements of industrial detection tasks.

Keywords: print defects detection; deep learning; Siamese convolutional neural network; feature pyramid; change detection

0 引言

现代印刷技术发展迅猛、应用广泛,从日用品外包装生产到钞票印刷,各种产品的生产过程都会应用到印刷技术。但因为设备参数、印刷材料、电气特性等方面的因素,可能会出现一些存在印刷缺陷的产品,从而导致印刷品视觉效果不一致,对产品质量产生负面影响。因此,印刷品质量检测已经成为印刷品生产过程中的一个重要环节。

作为印刷品质检环节的一项重要技术,印刷缺陷检测同样是工业检测领域的一个经典问题。目前,基于机器视觉的工业检测技术^[1]已取得了较好的发展,并且已经应用于金属表面划痕检测、印刷电路板缺陷检测等问题中。相较于人工抽样检测,基于视觉的自动化检测技术在保证产品全覆盖的同时,还具有速度快、精度高、人力成本低等优点,因此具有较高的研究与应用价值。

然而,即使在人工智能领域迅速发展的今天,现有的印刷缺陷检测技术依旧没有达到完全令人满意的水平。在现有技术中,依赖传统图像处理技术的视觉检测系统^[2-4]对成像时的光照条件、机械误差等影响因子极度敏感、鲁棒性较差,并且需要专业人员不断根据真实环境对系统参数进行调整,往往难以达到预期的检测效果。近几年来,得益于卷积神经网络(Convolutional Neural Network, CNN)^[5]强大的表征能力,基于深度学习^[6]的通用目标检测方法^[7-9]在诸多工业缺陷检测问题中都取得了较好的结果,但对于具有“内容相关性”的印刷缺陷检测问题,这些方法依旧面临着无法解决的语义矛盾。“内容相关性”引起的语义矛盾可以理解为,当缺陷与印刷内容的视觉特征相同时,YOLOv3 (You Only Look Once v3)^[7]等通用目标检测方法无法在只有检测图像信息的条件下完成检测。例如,模型无法将印刷内容缺失处的背景判别为缺陷,因为背景本身并非缺陷。图1展示了两种具有“内容相关性”的印刷缺陷。



(a) “BANK” 出现偏移



(b) 红框中漏印“A”

图1 具有“内容相关性”的缺陷样例

Fig. 1 Defect samples with "content correlation".

针对上述问题,本文将传统模板匹配技术中的对比思想与深度学习中的语义特征结合,提出了一种基于深度对比神经网络的印刷缺陷检测模型,最后在两个不同数据集上对该方法进行了全面评估。具体来说,本文的主要工作如下:

1) 首先,本文基于孪生卷积神经网络设计了一个端到端的印刷缺陷检测模型——深度对比网络(Deep Comparison Network, CoNet)。该模型在语义空间对比检测图像与参考图像,不仅解决了传统方法鲁棒性较差的问题,而且避免了基于深度学习的通用目标检测方法在印刷缺陷检测问题中面临的语义矛盾。

2) 然后,本文提出了一种非对称的双通路特征金字塔结构,并将其用于CoNet的多尺度变化检测模块(Multi-scale Change Detection Module, MsCDM)。这种网络结构可以在尽可能少的增加计算量的条件下,为检测模块引入更多的有效信息,从而提升检测性能。在公开的印刷电路板缺陷数据集(Print Circuit Board Defect Dataset, DeepPCB)^[10]和本文收集的立金缺陷数据集上的实验结果表明,相较于目前性能最优的两种印刷缺陷检测方法,CoNet的检测精度更高,并且检测速度也可以满足工业检测任务的实时性要求。

1 相关工作

1.1 传统工业检测算法

作为机器视觉领域的一个重要分支,基于传统图像处理技术的工业检测系统已取得了较好的发展,并且已成功应用于大量的工业场景当中,如印刷品^[11-14]、电路板^[15]、纺织和纹理^[16-17]检测等。

在印刷缺陷检测问题上,Shankar等^[12]利用动态滤波器检测图像边缘,然后通过阈值化与变换算法进一步分析,实现了可用于卷筒胶印设备的实时检测系统;Sun等^[13]通过增量主成分分析算法建模不同模式的正样本,并利用主成分为不同的测试样本重建模板,进而将测试样本与重建模板的差异视作印刷缺陷;Wang等^[14]针对传统差值方法存在的伪残差问题,提出了灰度差值和梯度差值相结合的方法,由于有效消除了伪残差,该方法的检出率明显高于传统差值方法。此外,针对印刷电路板缺陷检测问题,Malge等^[15]基于形态学算法提出了图像分割与局部像素对比相结合的缺陷检测方法;而Tsai等^[16]则利用傅里叶变换在频域对图像进行分析,通过一维霍夫变换检测并过滤高频分量后,再反变换得到只有缺陷区域被清晰保留的重构图像。

整体上看,基于传统图像处理技术的方法通常在像素空间或低维特征空间进行检测,例如直接计算检测图像与参考图像之间的差值^[12],或者在检测时增加梯度^[14]、频率^[16]等信息。可以发现,这些方法无法建模图像的语义信息,而图像

的灰度、梯度等信息容易受成像条件等因素影响,因此鲁棒性十分不足。

1.2 基于深度学习的目标检测方法

近年来,基于深度学习的计算机视觉算法发展迅猛,尤其在目标检测问题上取得了惊人的进展。

首先, Ren 等^[8]提出的二阶段模型 Faster R-CNN (Faster Region-based Convolutional Neural Network) 以 CNN 提取的语义特征图为基础,先利用两个 1×1 卷积层生成粗粒度的候选框,再通过感兴趣区域池化 (Region of Interest Pooling, RoIPooling) 提取候选框内特征,最后将所得特征用于预测目标的类别与边界框。与 Faster R-CNN 不同, Redmon 等^[7]提出的 YOLOv3 直接以骨架网络提取的多尺度特征图为输入,利用三个相互独立的 1×1 卷积层分别预测三种尺度目标的置信度、类别与边界框,得益于不用预先提取候选框,该方法不仅能保证检测精度,还可以满足实时检测的需求。

基于经典 Faster R-CNN, He 等^[9]提出的 Mask R-CNN (Mask Region-based Convolutional Neural) 将 Faster R-CNN 的骨架网络改进为 Lin 等^[18]提出的特征金字塔网络 (Feature Pyramid Network, FPN),并增加了一个全卷积掩码预测头,用于分割检测结果中的目标与背景,提升检测性能的同时还较好地解决了实例分割问题。马佳良等^[19]基于其提出的有效交并比 (Effective Intersection over Union, EIoU) 和特征重分配模块 (Feature Reassignment Module, FRM),设计了改进的目标检测框架。而 Yang 等^[20]则利用关键点检测的思路,设计了一个不依赖锚框的目标检测模型 RepPoints (Point set Representation),该模型为特征图的每个位置预测一组点集,并利用该点集输出检测框,是一种与上述基于锚框的方法完全不同的范式。此外, Sparse R-CNN (Sparse Region-Based Convolutional Neural Network)^[21]等候选框稀疏的目标检测方法也逐渐被关注,它们不用像大多数方法一样设置稠密分布的候选框,因此更符合人类思维。

上述方法在公开目标检测数据集上已经取得十分优秀的效果,并在许多工业检测问题中得以应用,例如 Liong 等^[22]将 Mask R-CNN 用于皮革缺陷检测问题; He 等^[23]将改进后的 Faster R-CNN 用于钢表面缺陷检测问题;冯涛等^[24]提出的角度加权交并比与 Mask R-CNN 结合后,较好地解决了染色体实例分割问题。

之所以基于深度学习的方法能全面超越基于特征描述子和形态学算法等技术的目标检测方法,主要是因为 CNN 具有更强大的表征能力。通过卷积层、池化层与激活层等线性与非线性变换运算的堆叠, CNN 将图像从像素空间映射到语义空间,然后利用反向传播算法^[5]优化模型参数,使模型可以在语义空间自适应地表征各种复杂的模式。

但在印刷缺陷检测问题上,上述以单张图像为输入的方法因为缺少参考图像的信息,所以无法完全解决具有“内容

相关性”的缺陷,例如印刷内容的偏移、缺失,以及与印刷内容视觉特征相同的缺陷。与通用目标检测方法相比,本文提出的 CoNet 通过孪生的网络结构让模型得到正样本(即参考图像)的信息,然后通过语义空间比较检测图像与参考图像完成缺陷检测,从而避免了“内容相关性”缺陷带来的问题。因此, CoNet 相较 Faster R-CNN 等通用目标检测方法更适合解决依赖参考图像的检测问题。

1.3 变化检测与印刷缺陷检测

与目标检测问题不同,变化检测的目标是对图像之间的差异进行判别与定位,需要以两张或多张图像为输入。一直以来,变化检测都是遥感图像分析领域的一个重要研究问题,并且也有许多优秀的研究成果。

Zhao 等^[25]首次将深度学习用于变化检测问题,其通过比例对数算子计算输入图像之间的差异图,然后利用深度学习完成差异图的后继判别。相反的, Zhan 等^[26]将孪生卷积神经网络用于图像对特征图提取阶段,然后以欧氏距离作为相似性度量计算特征图之间的距离,最后通过阈值化和 K 近邻得到检测结果。与前两种方法不同, Daudt 等^[27]基于 UNet^[28]提出了三种变化检测模型,这三种模型分别采用原图拼接、特征图拼接与特征图差值三种信息融合的方式,并且都是端到端的全卷积神经网络。而黄平等^[29]则通过构造基于改进相对熵与均值比的融合差异图提出了无监督的变化检测方法,并将其用于洪灾前后的变化趋势估计。

可以发现,变化检测方法正好可以用于印刷缺陷检测问题,因为印刷缺陷在本质上就是检测图像相较参考图像的变化之处。因此, Tang 等^[10]基于变化检测的思路,以 VGG-tiny (Visual Geometry Group Network-Tiny)^[30]作骨架网络提取检测图像和参考图像的特征图,通过特征图差值进行信息融合后,再使用分组金字塔池化 (Group Pyramid Pooling, GPP) 完成印刷电路板缺陷检测。 Tang 等^[10]分别测试了使用最大池化和平均池化两种方法的 GPP 模型,结果表明,最大分组金字塔池化 (Max-Pooling Group Pyramid Pooling, MP-GPP) 的性能更好。相似的, Haik 等^[11]针对动态数据打印提出了两种缺陷检测模型,一种是在像素空间进行信息融合的伪彩色单次检测器 (Pseudo-color Single Shot Detector, Pseudo-SSD),另一种是在语义空间进行信息融合的变化检测单次检测器 (Change-Detection Single Shot Detector, CD-SSD)。据我们所知, CD-SSD 是目前性能最优的变化检测方法之一。

和上述同样基于变化检测思路设计的印刷缺陷检测模型相比,本文提出的 CoNet 有以下特点:

1) 在骨架网络部分, CoNet 采用一对只包括三次下采样的轻量化网络提取图像对的中层语义特征,因为工业检测问题并不过分依赖高层语义特征,该设计可以更好地平衡检测精度和速度。

2) 头部网络 MsCDM 采用了非对称的双通路特征金字塔结构, 该结构通过增加一个运算量极低的次通路, 将检测

特征图通过“类残差结构”传递给各个尺度的输出模块, 在几乎不降低速度的条件下提升检测精度。

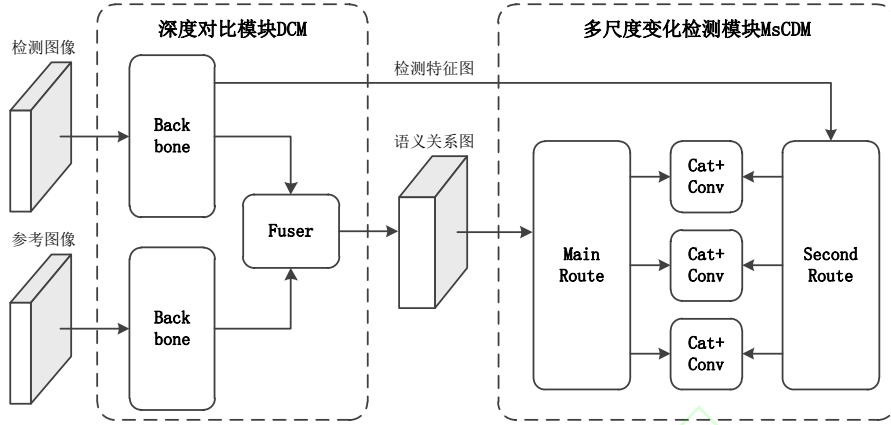


图2 CoNet 模型概览

Fig. 2 Overview of CoNet model

2 深度对比网络 CoNet

本文采用变化检测的思路, 将孪生卷积神经网络与非对称的双通路特征金字塔结构结合, 提出了一个端到端的印刷缺陷检测模型——CoNet。如图 2 所示, 该模型以检测图像和参考图像组成的图像对为输入, 首先通过深度对比模块 DCM (Deep Comparison Module) 得到检测图像与参考图像的语义关系图, 然后利用 MsCDM 模块对语义关系图进行后处理, 最终输出大、中、小三个尺度的预测结果, 其中包括置信度、边界框与缺陷类别三部分预测值。最后, 本文采用基于 GIoU (Generalized Intersection over Union) [31] 的多目标损失函数训练模型, 引导模型更好地学习图像对比的能力。

2.1 深度对比模块 DCM

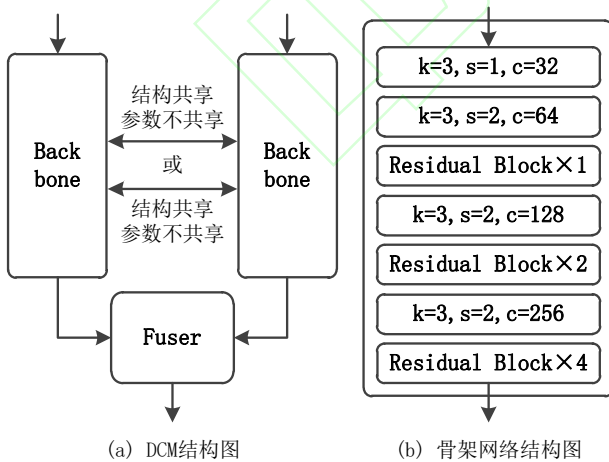


图3 DCM 模块示意图

Fig. 3 Schematic diagram of DCM module

如图 3(a)所示, DCM 主要由两个步骤组成: 1) 将检测图像与参考图像组成的图像对映射到语义空间; 2) 通过特征融合算子在语义空间中挖掘检测图像与参考图像之间的语义

关系。可以说, DCM 就是 CoNet 的核心所在, 因为本文的核心思想就是期望将模板匹配方法中的对比思想和深度学习中的语义特征结合, 通过在语义空间中进行图像对比增强模型鲁棒性。

2.1.1 孪生骨架网络

与 ResNet[32]等卷积神经网络不同, DCM 中的骨架网络是一个只包括三次下采样的轻量化卷积神经网络, 共 18 个卷积层。之所以采用浅层网络, 主要出于两点考虑:

1) 工业缺陷检测问题对模型的计算复杂度要求较高, 需要保证实时性, 在采用孪生结构的情况下, 网络过深会成倍地增加计算量, 因此 DCM 采用轻量化网络。

2) 本质上, 深度对比就是对检测图像与参考图像的对应局部进行特征提取与相似度计算, 而局部区域的大小则是模型的感受野。感受野过小会导致模型无法获取足够的语义信息, 过大则会令不相似区域的语义特征相似度过高。因此 DCM 采用只有三次下采样的浅层网络控制感受野。

如图 3(b)所示, 骨架网络由卷积层和残差模块组成。卷积层包括卷积、批归一化与激活三个步骤, 参数 k 、 s 、 c 分别表示卷积核大小、步长、卷积核个数, 步长为 2 时, 卷积层具有下采样的功能。残差模块由两个卷积核大小分别为 1 和 3 的卷积层组成, 并采用跳跃结构将输入值短接到第二个卷积层的激活函数前。具体来说, 卷积层和残差模块的运算过程可以分别记为式(1)和式(2), 其中, 带泄露修正线性单元 (Leaky Rectified Linear Units, LeakyReLU) 的参数 α 为 0.1。

$$CONV(X) = LeakyReLU(BN(Conv(X))) \quad (1)$$

$$ResBlock(X) = LeakyReLU(X + CONV_{3 \times 3}(CONV_{1 \times 1}(X))) \quad (2)$$

为了同时处理输入图像对, CoNet 采用孪生的网络结构, 使用一对结构完全相同的骨架网络分别处理检测图像和参考图像, 从而保证图像对的特征图维度相同。但是, 因为卷积神经网络是有参数模型, 在参数不同时, 两个网络结构完全

相同的模型在本质上依旧是两个不同的映射函数。因此,如图3(a)所示,本文提出了以下两种不同的方案。

1) 结构相同,参数共享。该方案的核心目标就是,通过相同的非线性映射函数 E ,将原始图像从像素空间投影到语义空间,令投影后的图像具有可度量的语义相似性。可度量的语义相似性是指,原图的局部区域经过映射函数 E 后,被投影为语义空间中的一个点,该方案希望投影点的距离就是局部区域之间的语义相似性。

2) 结构相同,参数不共享。相较于参数共享方案,该方案的缺点在于,两个参数不同的模型本质上是两个不同的非线性映射函数 $E1$ 、 $E2$,未训练的模型所得投影点的相似性关系无法得到保证。但是,该方案通过减少约束条件扩大了训练时的搜索空间,令模型拥有更强的表征能力,只是训练难度有所增加。因此,两种方案各有优劣。

2.1.2 特征融合算子

以孪生骨架网络输出的语义特征图 F_{det} 和 F_{ref} 为输入,特征融合算子的目标就是整合二者的信息,并挖掘它们之间的相关性。这种相关性可以是显式的相似性,也可以是隐式

的不可解释的相关性。因此,针对特征融合算子,本文也提出了两种方案:一种是引入先验信息设计的语义差分算子;另一种是基于卷积层设计的卷积融合算子。

1) 语义差分算子。由2.1.1节可知,原始图像经过孪生骨架网络投影后具有可度量的语义相似性,当局部区域语义相似时,投影点距离足够近,反之投影点距离足够远。基于这一先验信息,语义差分算子可以如式(3)定义,输出结果越接近0.5表示语义相似度越高,越接近0或1则表示相似度越低。

$$SubFuser(F_{det}, F_{ref}) = \sigma(F_{det} - F_{ref}) \quad (3)$$

2) 卷积融合算子。不同于语义差分算子,该算子不进行显式的语义相关性挖掘,而是在特征通道维度拼接图像对的语义特征图后,使用两个卷积核大小分别为1和3的卷积层进行特征融合。该算子虽然可解释性不足,但得益于卷积层参数的可学习性,经过大量数据训练之后,其表征能力相对更强。

$$ConvFuser(F_{det}, F_{ref}) = CONV_{3 \times 3} (CONV_{1 \times 1} (Concat(F_{det}, F_{ref}))) \quad (4)$$

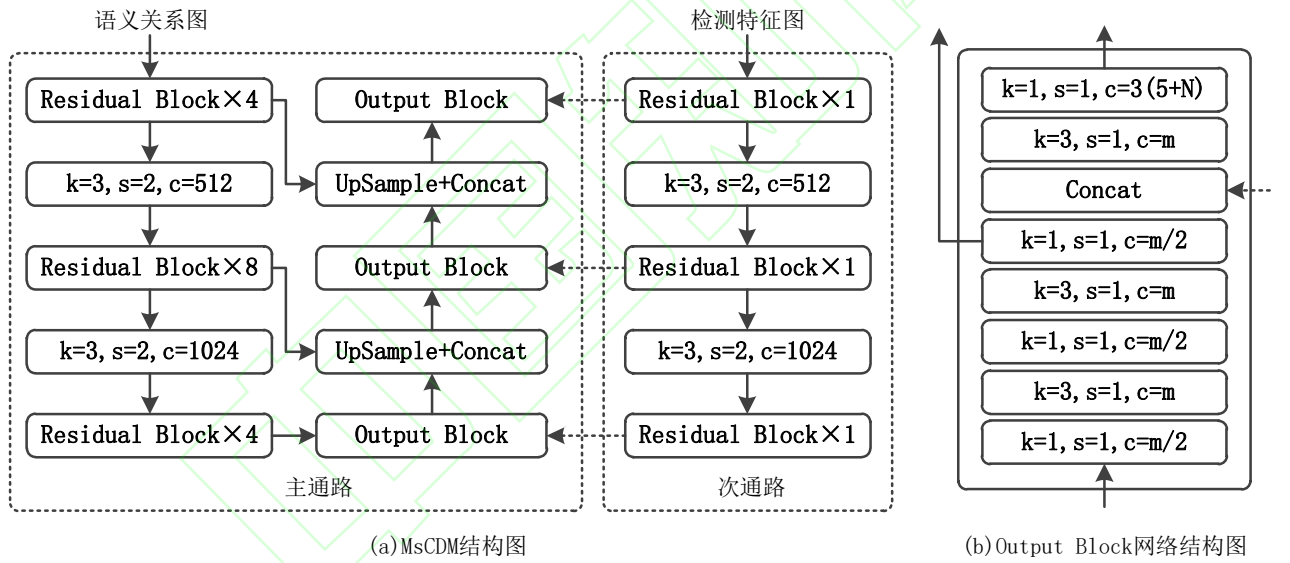


图4 MsCDM 模块示意图

Fig. 4 Schematic diagram of MsCDM module

2.2 多尺度变化检测模块 MsCDM

本文采用非对称的双通路特征金字塔结构,设计了一种多尺度变化检测模块 MsCDM,该模块由两个独立的信息通路组成:语义关系图通路、检测特征图通路。其中,语义关系图通路是必须存在的,该通路以深度对比模块输出的语义关系图 F_{diff} 为输入,并对其中的差异信息进行判定、定位与分类,因此是检测模块的主通路。而检测特征图通路是可选的次通路,该通路以检测图像的语义特征图作为输入,通过残差模块和步长为2的卷积层,将特征图变换后作为辅助信息传输至三种尺度的输出模块中,用于提升检测性能。

2.2.1 MsCDM 网络结构

如图4(a)所示,MsCDM的主通路由两个下采样过程和两个上采样过程组成,形成一个完整的特征金字塔结构。金字塔的下采样分支和上采样分支之间,信息的交互通过跳跃连接和特征通道拼接完成。以 F_{diff} 为输入,该通路首先在下采样分支上得到三种尺度的中间值 F_s 、 F_m 、 F_l ,分辨率分别为原图的1/8、1/16、1/32;然后,从 F_l 开始,模型先通过输出模块得到大尺度预测结果 O_l ,再用缩放因子为2的双线性插值层处理中间值,并将上采样后的值与 F_m 拼接得到下一尺度的输入值;最后,循环上一步骤得到中、小尺度的预测结果 O_m 、 O_s 。

与主通路不同,检测特征图通路只包括一个轻量级的下采样分支,由图 4(a)可知,该分支只包括三个残差模块和两个步长为 2 的卷积层。因为该通路的作用只是将检测特征图作为辅助信息用于输出模块,所以,在尽量少的增加运算量的前提下,此处采用了与主通路不同的轻量级网络。

连接以上两个通路的模块是输出模块 Output Block。如图 4(b)所示,输出模块由若干参数不同的卷积层堆叠而成,主通路输入值经过 5 个卷积层之后,同时进入两个分支:一个作为中间值被输入到上采样模块,另一个与次通路的检测特征图拼接,并用于预测结果计算。根据输出模块的网络结构可知,次通路的检测特征图只参与预测结果计算,而不会进入主通路的特征金字塔中。

2.2.2 预测结果解码

由图 4(b)的输出层参数可知,MsCDM 的输出值的通道数都是 $3(5+N)$,假设原图的分辨率为 (H, W) ,那么 CoNet 预测结果的维度分别是:

$$\begin{cases} O_s \in \mathbb{R}^{H/8 \times W/8 \times 3(5+N)} \\ O_m \in \mathbb{R}^{H/16 \times W/16 \times 3(5+N)} \\ O_l \in \mathbb{R}^{H/32 \times W/32 \times 3(5+N)} \end{cases} \quad (5)$$

以小尺度预测结果 O_s 为例,如图 5 所示,该结果分辨率为原图的 $1/8$,一共包括 $H/8 \times W/8$ 个长为 $3(5+N)$ 的向量,每个向量又分为三部分,每部分对应一组预测结果。具体来说,一组预测结果共包括 $5+N$ 位数,分别代表三种预测值:

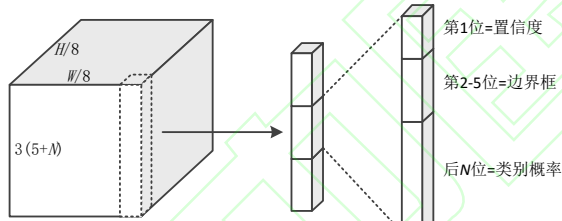


图5 解码过程示意图

Fig. 5 Schematic diagram of decoding process

1) 置信度: 预测结果的第 1 位用于表示对应位置是否存在缺陷。也就是说,将原图划分为 $H/8 \times W/8$ 的网格后,使用模型预测值表示对应网格中存在缺陷的置信度。如式(6)所示,解码过程就是使用 Sigmoid 函数将预测结果的第 1 位映射到 $(0, 1)$ 区间。

$$p_{\text{conf}} = \frac{1}{1 + e^{-o_1}} \quad (6)$$

2) 边界框中心点: 预测结果的第 2、3 位用于表示缺陷边界框中心点相对网格中心点的偏移量。假设网格中心点为 (x_c, y_c) ,解码过程可以记为式(7)。首先用系数为 0.5 的双曲正切函数将预测值映射到 $(-0.5, 0.5)$ 区间,与基准中心点相加后,再乘以当前尺度的下采样倍数 s 。因此边界框中心点最多只会相较网格中心点向任意方向偏移半个网格,不会落在当前网格之外。

$$\begin{cases} p_x = s(x_c + \frac{e^{o_2} - e^{-o_2}}{2(e^{o_2} + e^{-o_2})}) \\ p_y = s(y_c + \frac{e^{o_3} - e^{-o_3}}{2(e^{o_3} + e^{-o_3})}) \end{cases} \quad (7)$$

3) 边界框大小: 预测结果的第 4、5 位用于表示缺陷边界框的宽、高。本文为每个尺度预设了三种宽高比的锚框,每个锚框的面积与当前网格的面积相等。假设锚框的宽、高为 w_a, h_a ,如式(8)所示,边界框大小的解码过程就是将指数函数激活的预测值作为系数对锚框的宽、高进行缩放。

$$\begin{cases} p_w = w_a \cdot e^{o_4} \\ p_h = h_a \cdot e^{o_5} \end{cases} \quad (8)$$

4) 缺陷类别: 预测结果中剩余的 N 位分别表示缺陷属于某种类别的概率,其解码过程就是使用 Softmax 函数激活预测结果后 N 位,令它们的和为 1。

$$p_c = \frac{e^{o_c}}{\sum_{i=1}^N e^{o_{5+i}}} \quad (9)$$

此外,得到预测边界框后,还需要使用非极大值抑制算法 (Non-Maximum Suppression, NMS) 进行边界框去重。

2.3 损失函数

与预测结果解码相同,模型训练时的损失函数同样分为三部分:置信度、边界框与缺陷类别。三者之间的关系为:

$$L = L_{\text{conf}} + g_{\text{conf}} L_{\text{bbox}} + g_{\text{conf}} L_{\text{cls}} \quad (10)$$

其中, g_{conf} 表示置信度真值,当锚框为阴性时取值为 0,此时,边界框和缺陷类别的损失函数不参与模型训练;反之则在训练时同时计算三部分损失函数。

置信度和缺陷类别预测是分类问题,因此本文采用交叉熵函数作为这两部分的损失函数,计算方法如下:

$$L_{\text{conf}}(p_{\text{conf}}, g_{\text{conf}}) = -g_{\text{conf}} \ln(p_{\text{conf}}) - (1 - g_{\text{conf}}) \ln(1 - p_{\text{conf}}) \quad (11)$$

$$L_{\text{cls}}(p_c, g_c) = -\sum_{c=1}^N g_c \ln(p_c) \quad (12)$$

边界框的中心点与大小预测则是回归问题,在该部分,本文采用了基于 GIoU 设计的损失函数。GIoU 的计算方法如式(13)所示, A 和 B 表示两个边界框, C 表示二者的最小闭包。由式(13)可知,GIoU 的取值范围是 $(-1, 1)$,当两个边界框的距离无限远时,其取值为 -1;当二者重合时,取值则为 1。因此,GIoU 损失函数可以记为式(14),当预测框与真实框重合时,损失为 0。

$$GIoU(A, B) = IoU(A, B) - \frac{C - A \cup B}{C} \quad (13)$$

$$L_{\text{bbox}}(p_{\text{bbox}}, g_{\text{bbox}}) = 1 - GIoU(p_{\text{bbox}}, g_{\text{bbox}}) \quad (14)$$

3 实验与结果分析

3.1 数据集与评估指标

本文在两个数据集上进行了实验,一个是公开的印刷电路板缺陷数据集 DeepPCB,另一个是本文收集的真实工业场景下的立金缺陷数据集。

DeepPCB 是一个包括 6 种缺陷的数据集,其通过图像对的形式组织数据,并采用边界框的方式标注缺陷,因此适用于测试本文提出的 CoNet 模型。该数据集共 3000 张分辨率为 640×640 的图像,正、负样本各 1500 张,二者一一对应后组成 1500 个图像对。为了与现有印刷缺陷检测方法对比,本文采用 Tang 等^[10]使用的方式,将数据集划分为训练集和测试集两部分,前者 1000 个图像对,后者 500 个图像对。

立金缺陷数据集是一个真实的银行卡表面印刷缺陷数据集,共包括 6 个印刷区域,1384 张图片,缺陷类别包括 5 种,分别是:残缺、偏移、漏印、多印、划痕。与 DeepPCB 不同,立金缺陷数据集将每个印刷区域的所有图像划分为正、负样本集,而非图像对。同一区域的正、负样本集中的所有图像可以交叉组合。由于立金缺陷数据集的负样本较少,为了得到更可靠的实验结果,本文在数据集划分时采用五折交叉验证的方式,将各区域的正、负样本划分为 5 份后,每次实验选 4 份作为训练集,剩余 1 份则作为测试集。如此重复 5 次,最终结果取 5 次实验的均值。

本文将目标检测任务中常用的 mAP 作为评估指标,当检测边界框与标注边界框的 IoU (Intersection over Union) 大于 0.5 且类别相同时,该检测边界框为阳性。

3.2 实验细节

本文所有实验的运行环境相同,包括: Xeon(R) Platinum 8163 2.5GHz CPU $\times 1$; Tesla P100-16G GPU $\times 1$ 。

为了与其他性能优秀的变化检测方法进行对比,本文不仅实现了 CoNet 模型,还复现了 Tang 等^[10]提出的 MP-GPP 模型和 Haik 等^[11]提出的 CD-SSD 模型,并在两个数据集上进行了对比实验。与本文提出的 CoNet 相似,MP-GPP 和 CD-SSD 都采用变化检测思路进行印刷缺陷检测,其中 MP-GPP 是 DeepPCB 数据集的基线模型,而 CD-SSD 则是基于变化检测思路解决印刷缺陷检测问题的最新方法。同时,本文方法还与模板匹配方法和 Faster R-CNN 进行了对比,二者在 DeepPCB 数据集上的实验结果来自于 Tang 等^[10]公开的论文,而在立金缺陷数据集上的实验结果则是本文的真实测试值。其中,模板匹配方法是本团队已商业化的方案,因此可以保证其精度是有意义的;Faster R-CNN 的实验则基于开源的 MMDetection^[33]完成。

对于 DeepPCB 数据集,本文将输入图像的分辨率设置为 640×640 ,并将训练批次大小设置为 4;对于立金缺陷数据集,本文首先采用长边切分与短边填充结合的方式,将原

图转化为 320×320 的图像块,然后将训练批次设置为 8。训练时,这两个数据集都会进行数据增广,并且每个图像对中的两张图像通过的数据增广是一致的。

所有实验都用 Adam 优化器训练,初始学习率为 $3e-4$,并且采用热启动和余弦衰减策略动态调整学习率,终止学习率为 $1e-6$ 。每次实验总共训练 50 轮,前 2 轮学习率处于热启动阶段,后 48 轮学习率再进入余弦衰减阶段。最终用于测试的模型为第 50 轮保存的模型。

3.3 结果分析

首先,本文在 DeepPCB 数据集上评估 CoNet 模型,并与其他方法进行对比,所有结果如表 1 所示。

CoNet 模型的平均精度为 99.1%,相较于现有的四种方法都取得了更好的检测结果。与模板匹配方法相比,CoNet 的平均精度 mAP 提高了 9.8 个百分点,在性能上全面领先。与 Faster R-CNN 相比,CoNet 则提高了 1.5 个百分点,尤其在 Open 和 Short 类型上,CoNet 分别高出了 2.2 和 3.3 个百分点。与同类型的变化检测方法 MP-GPP 和 CD-SSD 相比,CoNet 依旧取得了更好的检测结果,其平均精度在超过了 99% 的情况下相较 MP-GPP 和 CD-SSD 分别提升了 0.5 个百分点和 0.7 个百分点,并且在各种类型缺陷上的精度都等于或者高于这两种方法。综上可知,相较于现有的各类缺陷检测方法,本文提出的 CoNet 模型具有更优的性能。

然后,考虑到 DeepPCB 数据集的难度较小,实验结果可能不足以反映它们的真实性能。本文又在更复杂的立金缺陷数据集上进行了类似的评估,实验结果如表 2 所示。

CoNet 的平均精度为 69.8%,相较两种思路类似的方法 MP-GPP 和 CD-SSD,分别提升了 3.5 个百分点和 2.4 个百分点;相较传统方法和 Faster R-CNN,则分别提升了 12 个百分点和 5.3 个百分点。除了 CD-SSD 得到了比 MP-GPP 更高的平均精度,表 2 的整体结果与 DeepPCB 数据集上的结果基本一致。值得注意的是,在立金缺陷数据集上,如果只输入检测图像训练 Faster R-CNN,会出现不收敛的情况;但如果将检测图像和参考图像进行堆叠,组成 6 通道的伪图像,则训练收敛,表 2 中的 Faster R-CNN*就是通过该方法训练所得。分析训练数据发现,立金缺陷数据集中存在较多的内容漏印和偏移的缺陷,这可能就是 Faster R-CNN 在正常训练时不收敛的原因,同时也与前文提出的“内容相关性”缺陷的概念相吻合。根据 Daudt 等^[27]提出的变化检测方法可知,堆叠检测图像和参考图像在本质上就是一种图像对信息融合的方式,因此可以在一定程度上解决“内容相关性”问题。

此外,通过比较各个方法在五折交叉验证中的最大和最小平均精度差值发现,CoNet 具有更稳定的性能表现,其差值为 1.2 个百分点,而 MP-GPP 和 CD-SSD 则分别是 2.3 个百分点和 1.7 个百分点。这表明,CoNet 不仅精度更高,而且对训练数据的自适应性也更强。

表1 在 DeepPCB 数据集上的检测结果

Tab. 1 Detection results on DeepPCB dataset

方法	平均精度/%	不同缺陷类型的精度/%					
		Open	Short	Mousebite	Spur	Copper	Pin-hole
模板匹配	89.3	88.2	87.6	90.3	88.9	91.5	89.2
Faster R-CNN	97.6	96.8	95.4	97.9	98.7	97.4	99.5
MP-GPP	98.6	98.5	98.5	99.1	98.2	98.5	99.4
CD-SSD	98.4	97.7	97.9	98.4	98.1	99.1	99.2
CoNet	99.1	99.0	98.7	99.1	99.2	99.3	99.5

注 CoNet = 孪生骨架网络参数共享+差分融合算子+双通路结构。

表2 在立金缺陷数据集上的检测结果

Tab. 2 Detection results on Lijin defect dataset

方法	平均精度/%	五折交叉验证的精度/%				
		Flod-1	Flod-2	Flod-3	Flod-4	Flod5
模板匹配	57.8	55.9	57.3	60.1	54.1 (-3.7)	60.3 (+2.5)
Faster R-CNN*	64.5	64.9	65.1 (+0.6)	63.8 (-0.8)	64.7	64.1
MP-GPP	66.3	67.4 (+1.1)	66.4	65.1 (-1.2)	65.8	66.8
CD-SSD	67.4	68.2 (+0.8)	67.7	67.1	66.5 (-0.9)	67.6
CoNet	69.8	69.5	70.2	69.7	69.3 (-0.5)	70.5 (+0.7)

注 CoNet = 孪生骨架网络参数共享+差分融合算子+双通路结构; (+/-*) 表示当前精度相比该方法平均精度的偏移量。

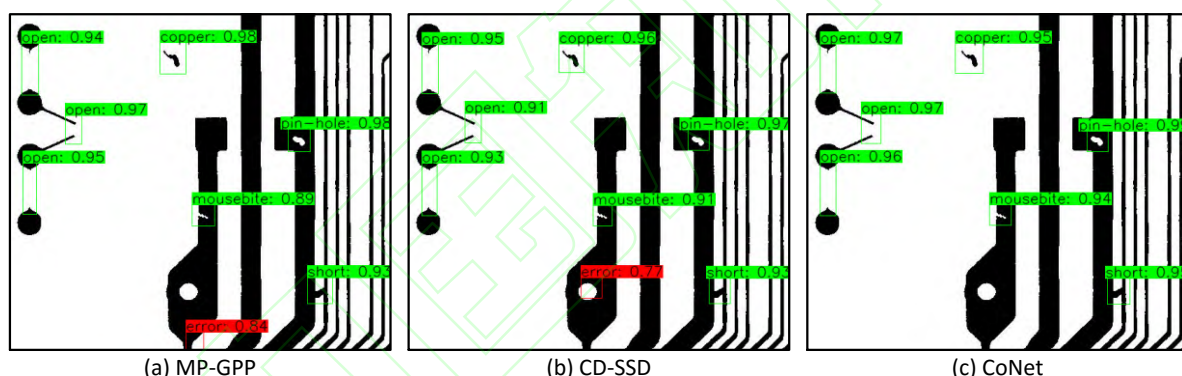


图6 三种变化检测方法在 DeepPCB 数据集上的检测结果可视化

Fig. 6 Visualization of detection results of three change detection methods on DeepPCB dataset

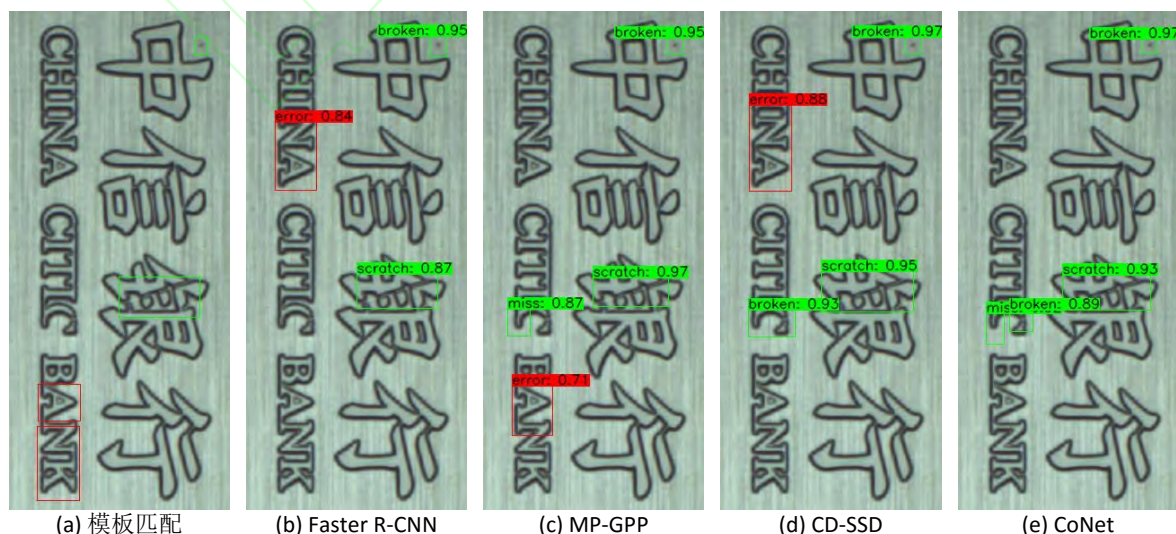


图7 表 2 中的五种方法在立金缺陷数据集上的检测结果可视化

Fig. 7 Visualization of detection results of the five methods in Table 2 on the Lijin defect dataset

除了定量分析 CoNet 的检测精度,本文还通过可视化方法更直观地比较了对比实验中的五种方法。图 6 是三种变化检测方法在 DeepPCB 数据集上的检测结果,相较 MP-GPP 和 CD-SSD,本文提出的 CoNet 在没有漏检的情况下,还避免了印刷内容边缘处的误检。类似的,图 7 是立金缺陷数据集上的可视化结果。相较图 7(a)和图 7(b)两个非变化检测方法的结果,图 7(c)~图 7(e)的漏检更少,这说明变化检测方法对印刷缺陷检测任务是有效的。然后,相较剩余两种变化检测方法,本文提出的 CoNet 进一步提高了检出率,并且依旧没有出现误检。该结论与表 1 和表 2 的结果基本一致。

3.4 消融实验与复杂度分析

本节首先通过消融实验对 CoNet 的两个主要模块及其不同方案进行分析,然后再对 CoNet 的参数量、浮点运算次数和单次检测的耗时进行分析。

3.4.1 针对 DCM 中各组件的消融实验

由 2.1.1 和 2.1.2 两小节可知,孪生骨架网络和特征融合算子分别存在两种不同的方案,在 MsCDM 采用双通路结构时,将这两个组件的可行方案交叉组合,然后在 DeepPCB 数据集和立金缺陷数据集上进行对比,结果如表 3 所示。

表3 DCM 采用不同方案的检测结果

Tab. 3 Detection results of different DCM

骨架网络	融合算子	不同数据集上的平均精度 /%	
		DeepPCB	立金缺陷
√	SubFuser	99.1	69.8
√	ConvFuser	98.8	68.9
×	SubFuser	98.6	68.1
×	ConvFuser	98.8	69.1

注: √和×分别表示参数共享和参数不共享。

首先,在 DeepPCB 数据集上,四种组合方案中的平均精度最高值为 99.1%,在参数共享和语义差分算子组合时取得。当该组合的骨架网络改为参数不共享时,平均精度出现了 0.5 个百分点的下降。当采用卷积融合算子时,不论骨架网络是否参数共享,平均精度都是 98.8%。

然后,在立金缺陷数据集上的实验也取得了类似的结果。当使用语义差分算子时,骨架网络参数共享与不共享两种情况下的平均精度分别为 69.8%和 68.1%,后者相对下降了 1.7 个百分点。当使用卷积融合算子时,两种骨架网络方案的平均精度接近,只相差 0.2 个百分点。

根据两种融合算子在不同情况下的性能表现可以得出结论:差分融合算子需要与参数共享的孪生骨架网络组合才能取得更好的效果,而卷积融合算子则可以适应不同类型的骨架网络。合理的解释是,差分融合算子基于强先验信息设计,其希望原始图像间的语义相似性可以转化为语义空间中投影点间的距离,因此参数共享的方案更加合适。而卷积融合算

子因为其可学习的特性,对输入数据的自适应能力更强,所以不会过分依赖骨架网络输出的语义特征。虽然在表 3 中卷积融合算子的性能略低于差分融合算子,但这并不完全表示卷积融合算子就不可取,因为可学习的模型往往需要更多的数据训练,而本文使用的数据集规模都较小。

3.4.2 针对 MsCDM 中双通路结构的消融实验

本文在 DCM 采用骨架网络参数共享与语义差分算子组合的前提下,分别测试了 MsCDM 采用单通路结构与双通路结构时的性能,结果如表 4 所示。相较只有主通路的单通路结构,增加次通路的双通路结构在 DeepPCB 数据集上的平均精度提升了 1.2 个百分点,在立金缺陷数据集上则提升了 2.1 个百分点。由实验结果可知,双通路结构是有效的。

表4 不同 MsCDM 的检测结果

Tab. 4 Detection results of different MsCDM

MsCDM 的结构		不同数据集上的平均精度 /%	
主通路	次通路	DeepPCB	立金缺陷
√		97.9	67.7
√	√	99.1	69.8

注: √表示 MsCDM 中使用对应通路。

3.4.3 CoNet 复杂度分析

作为工业检测任务的一种,印刷缺陷检测问题需要算法具有一定的实时性。本小节对 CoNet 的复杂度进行分析,结果如表 5 所示。实验时输入图像的分辨率为 640×640,硬件设备为: Xeon(R) Platinum 8163 2.5GHz CPU ×1; Tesla P100-16G GPU ×1。

表5 采用不同结构的 MsCDM 时 CoNet 的复杂度

Tab. 5 Complexity of CoNet by using MsCDM of different structures

方法	参数量	运算量	耗时/ms
A	49.8×10^6	96.0×10^9	32.3
A+次通路	62.6×10^6	106.1×10^9	35.7

注: A=孪生骨架网络参数共享+差分融合算子+主通路。

首先,在 MsCDM 采用单通路结构时,CoNet 的参数量约为 49.8×10^6 ,浮点运算次数约为 96×10^9 ,在指定设备上完成前向运算和 NMS 的总耗时约为 32.3ms(千次实验均值)。增加次通路后,参数量和运算量分别增加了 12.8×10^6 和 10.1×10^9 ,耗时则增加了 3.4ms。可见,增加次通路并不会过分影响检测速度,CoNet 最终的耗时为 35.7ms,具有较好的实时性。

4 结语

本文将变化检测用于印刷缺陷检测问题, 将对比思想与语义特征结合, 提出了一种通用的印刷缺陷检测模型 CoNet。具体来说, CoNet 首先通过深度对比模块 DCM 挖掘图像对的语义关系图, 然后再利用双通道的多尺度变化检测模块 MsCDM 在语义关系图上定位并识别印刷缺陷。得益于 DCM 强大的表征能力与 MsCDM 利用双通道结构引入的更多有效信息, CoNet 最终在 DeepPCB 数据集和立金缺陷数据集上都取得了优于现有基于深度学习的变化检测方法的性能。实验结果表明, 在语义空间进行图像对比并用于解决印刷缺陷检测任务的思路是可行的, 但依旧存在可优化的部分。后续的研究可以从两方面展开: 一方面是进一步提高 CoNet 在弱小缺陷上的检测性能; 另一方面是引入弱监督、自监督等方法, 解决基于深度学习的变化检测方法需要大量有标注数据训练的问题。

参考文献

- [1] 陶显, 侯伟, 徐德. 基于深度学习的表面缺陷检测方法综述 [J]. 自动化学报, 2021, 47(5): 1017-1034. (TAO X, HOU W, XU D. A survey of surface defect detection methods based on deep learning. Acta Automatica Sinica, 2021, 47(5): 1017-1034).
- [2] ISHIMARU I, HATA S, HIROKARI M. Color-defect classification for printed-matter visual inspection system [C]// Proceedings of the 2002 4th World Congress on Intelligent Control and Automation (Cat. No. 02EX527). Piscataway: IEEE, 2002: 3261-3265.
- [3] LUO J, ZHANG Z. Automatic colour printing inspection by image processing [J]. Journal of Materials Processing Technology, 2003, 139(1/2/3): 373-378.
- [4] SHANKAR N G, RAVI N, ZHONG Z W. On-line defect detection in web offset printing [C]// Proceedings of the 2003 4th International Conference on Control and Automation Proceedings. Piscataway: IEEE, 2003: 794-798.
- [5] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [6] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. Nature, 2015, 521(7553): 436-444.
- [7] REDMON J, FARHADI A. YOLOv3: an incremental improvement [EB/OL]. [2021-10-27]. <https://arxiv.org/abs/1804.02767>.
- [8] REN S Q, HE K M, 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.
- [9] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 2961-2969.
- [10] TANG S L, HE F, HUANG X L, et al. Online PCB defect detector on a new PCB defect dataset [EB/OL]. [2021-10-27]. <https://arxiv.org/abs/1902.06197>.
- [11] HAIK O, PERRY O, CHEN E, et al. A novel inspection system for variable data printing using deep learning [C]// Proceedings of the 2020 IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway: IEEE, 2020: 3541-3550.
- [12] SHANKAR N G, RAVI N, ZHONG Z W. A real-time print-defect detection system for web offset printing [J]. Measurement, 2009, 42(5): 645-652.
- [13] SUN X G, CHEN B, ZHANG L. On-line print-defect detecting in an incremental subspace learning framework [J]. Sensor Review, 2010, 31(2): 138-143.
- [14] WANG Y P, XU S W, ZHU Z P, et al. Real-time defect detection method for printed images based on grayscale and gradient differences [J]. Journal of Engineering Science and Technology Review, 2018, 11(1): 180-188.
- [15] MALGE P, NADAF R. PCB defect detection, classification and localization using mathematical morphology and image processing tools [J]. International Journal of Computer Applications, 2014, 975: 8887.
- [16] TSAI D M, HSIEH C Y. Automated surface inspection for directional textures [J]. Image and Vision Computing, 1999, 18(1): 49-62.
- [17] KUMAR A. Computer-vision-based fabric defect detection: A survey [J]. IEEE Transactions on Industrial Electronics, 2008, 55(1): 348-363.
- [18] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 2117-2125.
- [19] 马佳良, 陈斌, 孙晓飞. 基于改进的 Faster R-CNN 的通用目标检测框架 [J]. 计算机应用, 2021, 41(9): 2712-2719. (MA J L, CHEN B, SUN X F. General object detection framework based on improved Faster R-CNN [J]. Journal of Computer Applications, 2021, 40(9): 2712-2719.)
- [20] YANG Z, LIU S, HU H, et al. Reppoints: Point set representation for object detection [C]// Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 9657-9666.
- [21] SUN P, ZHANG R, JIANG Y, et al. Sparse r-cnn: End-to-end object detection with learnable proposals [C]// Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 14454-14463.
- [22] LIONG S T, GAN Y S, HUANG Y C, et al. Automatic defect segmentation on leather with deep learning [EB/OL]. [2021-10-27]. <https://arxiv.org/abs/1903.12139>.
- [23] HE Y, SONG K C, MENG Q G, et al. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features [J]. IEEE Transactions on Instrumentation and Measurement, 2019, 69(4): 1493-1504.
- [24] 冯涛, 陈斌, 张跃飞. 基于改进的 Mask R-CNN 的染色体图像分割框架 [J]. 计算机应用, 2020, 40(11): 3332-3339. (FENG T, CHEN B, ZHANG Y F. Chromosome segmentation framework based on improved Mask R-CNN [J]. Journal of Computer Applications, 2020, 40(11): 3332-3339.)
- [25] ZHAO J J, GONG M G, LIU J, et al. Deep learning to classify difference image for image change detection [C]// Proceedings of the 2014 International Joint Conference on Neural Networks. Piscataway: IEEE, 2014: 411-417.
- [26] ZHAN Y, FU K, YAN M L, et al. Change detection based on deep siamese convolutional network for optical aerial images [J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(10): 1845-1849.
- [27] DAUDT R C, BERTR L S, ALEXANDRE B. Fully convolutional Siamese networks for change detection [C]// Proceedings of the 2018 25th IEEE International Conference on Image Processing. Piscataway: IEEE, 2018: 4063-4067.
- [28] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]// Proceedings of the 2015 International Conference on Medical Image Computing and Computer-assisted Intervention. Berlin: Springer, 2015: 234-241.
- [29] 黄平平, 段盈宏, 谭维贤, 等. 基于融合差异图的变化检测方法及其在洪灾中的应用 [J]. 雷达学报, 2021, 10(1): 143-158. (HUANG P P, DUAN Y H, TAN W X, et al. Change detection method based on

- fusion difference map in flood disaster [J]. Journal of Radars, 2021, 10(1): 143-158.)
- [30] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2021-10-27]. <https://arxiv.org/abs/1409.1556>.
- [31] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression [C]// Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 658-666.
- [32] HE K M, ZHANG X Y, RES S Q, et al. Deep residual learning for image recognition [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 770-778.
- [33] OpenMMLab. MMDetection [EB/OL]. [2020-09-12]. <https://github.com/open-mmlab/mmdetection>.

WANG Youxin, born in 1997, M. S. candidate. His research interests include industrial defects detection, object detection, multimodal video representation.

CHEN Bin, born in 1970, Ph. D., research fellow. His research interests include industrial detection, deep learning.

