文章编号　1004-924X(2023)17-2611-15

零部件光学影像精准定位的轻量化深度学习网络

牛小明[1]，曾　理[1*]，杨　飞[2]，何光辉[1]
(1. 重庆大学 数学与统计学院，重庆 401331；
2. 长春长光辰谱科技有限公司，吉林 长春 130000)

摘要：光学影像精准定位是提高工业生产效率和质量的重要环节。传统图像处理定位方法由于光照、噪声等环境因素的影响，在复杂场景下定位精度低、易受干扰；而经典深度学习网络虽然在自然场景目标检测、工业安检、抓取、缺陷检测等得到了广泛应用，但是其海量数据的训练需求、复杂系统的深度学习大模型、检测框的冗余及不精确等问题，导致它不能直接应用于工业零部件像素级精准定位。针对以上问题，构建了一种零部件光学影像像素级精准定位的轻量化深度学习网络方法。网络总体选用Encoder-Decoder架构，Encoder使用三级bottleneck级联，在降低特征提取参变量的同时充分提升了网络的非线性；Encoder与Decoder对应特征层实施融合拼接，促使Encoder在上采样卷积后可以获得更多的高分辨率信息，进而更完备地重建出原始图像细节信息；最后，利用加权的Hausdorff距离构建了Decoder输出层与定位坐标点的关系。实验结果表明：轻量化深度学习定位网络模型参数为57.4 kB，定位精度小于等于5 pixel的识别率大于等于99.5%，基本满足工业零部件定位精度高、准确率高和抗干扰能力强等要求。

关　键　词：机器视觉；光学影像；深度学习；精准定位；轻量化
中图分类号：TP391.4　　**文献标识码**：A　　**doi**：10.37188/OPE.20233117.2611

Lightweight deep learning network for accurate localization of optical image components

NIU Xiaoming[1], ZENG Li[1*], YANG Fei[2], HE Guanghui[1]

(1. *College of Mathematics and Statistics, Chongqing University, Chongqing* 401331, *China*;
2. *Chang Chun Champion Optics Co., Ltd., Changchun* 130000, *China*)
* *Corresponding author, E-mail: drlizeng@cqu.edu.cn*

Abstract: Precise optical image localization is crucial for improving industrial production efficiency and quality. Traditional image processing and localization methods have low accuracy and are vulnerable to environmental factors such as lighting and noise in complex scenes. Although classical deep learning networks have been widely applied in natural-scene object detection, industrial inspection, grasping, defect detection, and other areas, directly applying pixel-level precise localization to industrial components is still challenging owing to the requirements of massive data training, complex deep learning models, and redundant and imprecise detection boxes. To address these issues, this paper proposes a lightweight deep learning network approach for pixel-level accurate localization of component optical images. The overall design of the network adopts an Encoder - Decoder architecture. The Encoder incorporates a three-level bottle-

收稿日期：2023-06-05；**修订日期**：2023-06-19.
基金项目：国家自然科学基金资助项目(No. 62076043)；国家重点研发计划资助项目(No. 2020YFB2007001)

neck cascade to reduce the parameter complexity of feature extraction while enhancing the network's nonlinearity. The Encoder and Decoder perform feature layer fusion and concatenation, enabling the Encoder to obtain more high-resolution information after upsampling convolution and to reconstruct the original image details more comprehensively. Finally, the weighted Hausdorff distance is utilized to establish the relationship between the Decoder′s output layer and the localization coordinates. Experimental results demonstrate that the lightweight deep learning localization network model has a parameter size of 57.4 kB, and the recognition rate for localization accuracy less than or equal to 5 pixels is greater than or equal to 99.5%. Thus, the proposed approach satisfies the requirements of high localization accuracy, high precision, and strong anti-interference capabilities for industrial component localization.

Key words: machine vision; optical image; deep learning; precise localization; lightweight

1 引　言

机器视觉定位[1]是一种基于光学摄像头或其他传感器获取物体位置和姿态信息,并结合算法进行数据处理,最终实现对目标物体精确定位和跟踪的技术。该技术在工业4.0中发挥着不可或缺的作用。在智能制造中,视觉定位技术可以用于产品或零部件的定位[2]、检测和识别,实现产线流水线的自动化和智能化,进而大幅提升生产效能,降低人工成本。在机器人自动化装配中[3],视觉定位技术可以帮助机器人定位和识别零部件,实现自动化装配。在智能物流中,视觉定位技术可以用于物品的识别[4]和追踪,提高物流效率和准确性;在智能仓储中,视觉定位技术可以用于物品的分类和存储,提高仓储效率和管理水平。

机器视觉定位技术通常包括以下几个步骤:图像采集、图像预处理、特征提取、目标匹配和坐标转换(图像的像素坐标转换为机器人的空间物理坐标)。在流水线自动安装时,机器视觉定位不精准对微型芯片、高灵敏零部件、易碎且价格昂贵的屏幕等产品的影响可能会更加严重。具体影响包括:(1)损坏产品,机器视觉定位不准确会导致机器在操作时未能正确识别产品位置,可能会对其进行错误的处理或施加过大的力量,从而导致产品损坏;(2)降低产品质量,机器视觉定位不准确可能会导致产品的安装不够精准或位置不正确,进而导致产品质量下滑,功能异常或者寿命缩减;(3)增加生产成本,由于机器视觉定位不准确,机器需要进行额外的处理或人工干预,从而增加生产成本;(4)安全隐患,机器视觉定位不准确还可能会导致产品被放置在不正确的位置,从而产生安全隐患。对于微型芯片和高灵敏零部件等产品,错误的操作可能会导致电路短路等严重问题。因此,在智能制造生产过程中,机器视觉的定位准确性和精度起到至为关键的作用。

随着人工智能的发展,深度学习方法在自动驾驶、智能安防、工业制造和医学影像等领域得到了广泛的应用。AlexNet[5]是深度学习在计算机视觉领域的开端,它使用深度卷积神经网络(Convolutional Neural Network,CNN)对大规模图像数据集ImageNet进行分类,同时也在目标检测领域进行了尝试。2014年,Girshickt[6]等提出了R-CNN(Region based Convolutional Neural Network)算法,它使用区域建议法取出候选目标区域,然后对单个区域采用CNN进行特征提取,并利用支持向量机(SVM)实施分类,紧接着选用回归模型对目标框实施微调,R-CNN在PASCAL VOC数据集上获得了比较好的效果。SPP-Net[7]对R-CNN算法进行了改进,引入了空间金字塔池化(Spatial Pyramid Pooling,SPP),促使网络可以对任意尺寸的输入图像进行分类和检测;对比R-CNN,SPP-Net识别率更高且运行时间更短。Fast R-CNN[8]引入了ROI(Region of Interest Pooling)池化,使得网络可以在整张图像上进行前向传播,并对每个候选区域进行分类和定位;相比R-CNN和SPP-Net,Fast R-CNN更快、更准确。2015年,Ren[9]等提出了Faster R-CNN网络,它引入区域建议网络(Region Proposal Network,RPN)来生成候选目标框,然后再将这些框输入到Fast R-CNN中进行分类和定位。因其速度和准确率表现优秀,Faster R-CNN是目

前最常用的目标检测算法之一。2016年，Redmon[10]等提出YOLO（You Only Look Once）算法，YOLO是一种One-Stage的实时目标检测算法，它将输入图像分成栅格，然后对每个栅格预测出多个目标框和类别概率，因此推理速度非常快。SSD[11]（Single Shot Multi-Box Detector）也是一种One-Stage的目标检测算法，它使用多尺度卷积特征图来检测不同尺度的目标，并在每个特征图位置上同时预测多个目标框和类别概率；SSD在速度和准确率方面都有较好的表现。2017年，Lin[12]等提出RetinaNet网络，RetinaNet是一种基于Focal Loss的目标检测算法。Focal Loss通过缩小易分类样本的权重，加强难分类样本的权重，解决了目标检测中正负样本不均衡的问题。相比较SSD和YOLO，RetinaNet的性能相对更好。Mask R-CNN[13]是对Faster R-CNN的改进，引入了Mask Head进行实例分割，它能够同时检测物体的边界框和生成物体的掩模，是目前应用较广的实例分割算法之一。2018年，Cai[14]等提出Cascade R-CNN算法，它引入级联结构进一步提升目标检测准确率，每个级联层都是一个独立的分类器，用来筛选出更准确的目标框。2019年，Zhou[15]等提出基于关键点检测的ExtremeNet网络，它利用网络模型检测出目标物体上下左右方向的4个极值点和一个中心点，并利用几何关系找到有效点组，一组极点代表一个检测结果，该模型有着较快的检测速度。Li[16]等提出了多尺度目标检测网络TridentNet，它采用空洞卷积并设计了3种不同尺寸感受野的并行分支，可以处理不同尺寸的物体；训练时使用3个分支，测试时使用1个分支，有效解决了目标尺寸变化的问题。Zhu[17]等提出了可以根据目标尺寸自由选择特征层的FSAF（Feature Selective Anchor-Free），针对每一个特征层，计算其损失值，使用损失值最小的特征层当作最好的特征层进行检测，该模型较好地解决了目标尺寸变化的问题。2020年，Tan[18]等提出了EfficientDet网络，它引入BiFPN（Bi-directional Feature Pyramid Network）进行多尺度特征融合，同时使用Compound Scaling进行模型结构优化，EfficientDet在速度和准确率方面都有很好的表现。YOLOv4[19]是在YOLOv3的基础上进行技术升级和迭代，包括利用SAM（Spatial Attention Module）来提升感受野响应，使用SPP（Spatial Pyramid Pooling）来增加网络对目标的表达能力等，在COCO数据集上的mAP_{50}为43.5%。Sparse R-CNN[20]采用一种稀疏注意力机制，目的是减少冗余计算和加速推理，在COCO数据集上，Sparse R-CNN在mAP_{50}方面比Faster R-CNN高2.5%，速度比它快5倍以上。Beal[21]等提出了ViT-FRCNN网络，它使用Transformer替换骨干网络，借助注意力机制对图像全局特征进行编码，基于Transformer的ViT-FRCNN网络在标准数据集上获得了出色的性能。另外，Qiu[22]等提出了BoderDet网络，它使用边界对齐自适应地进行边界特征提取，并将边界对齐操作封装成对齐模块集成到FCOS（Fully Convolutional One-Stage）网络，使用高效的边界特征提高了重叠目标的检测精度。Yang[23]等提出了基于边界圆的无锚框检测方法CircleNet，只需要学习边界圆的半径就可以实现目标检测，与边界框相比，具有优越的检测性能和较好的旋转一致性。2021年，Chen[24]等提出RepPoints v2算法，它基于Anchor-Free的思想，利用学习目标的重心和特征点进行检测；相比较RepPoints v1，RepPoints v2，它对网络结构、特征提取和损失函数等实施了改进，进而提升了检测精度和速度；在COCO数据集上测试，RepPoints v2的mAP_{50}比RepPoints v1高1.6%。Yao[25]等提出了一种端到端的检测算法Efficient DETR，该算法利用密集先验知识初始化检测网络，仅用3个编码器和1个解码器就达到了较高的检测精度，而且提高了收敛速度。Lang[26]提出了无锚框检测网络DAFNe，结合中心点和角点间距预测边界框，并将中心度感知函数扩展到任意四边形，从而提高目标定位精度，对于旋转目标效果更好。Ge[27]等提出了基于无锚机制的检测方法YOLOX，它利用解耦头将分类和回归任务进行解耦，改善模型收敛速度的同时提高了检测性能。2022年，Yu[28]等使用三段级联设计完善了目标检测和重识别，实现每个阶段注意力结构紧密特征交叉，进而使网络从粗到细进行目标特征学习，能够更加清晰地区分目标和背景特征。Cheng[29]等提出了AOPG（Anchor-Free Oriented Proposal Generator）网络，它将特征图映射

到图像上，把位于真实框中心区域的顶点作为正样本，构建新的区域标签分配模块，缓解了正样本所占比例小的问题，并且使用对齐卷积消除了特征和旋转框之间的不对齐。Huang[30]等提出了一种无锚框自适应标签分配策略，能够从热力图中获取任意方向和形状的特征，自适应调整高斯权重来适配不同的目标特征，使用联合优化损失完善非对齐优化任务，使检测速度和检测精度得到大幅提升。

深度学习目标检测算法目前已初步应用在工业机器视觉中，比如制造过程中的产品缺陷检测、工厂的异常行为检测、工厂的安全隐患监测，指导机器人进行自动化加工、装配等。然而，将它直接应用于零部件精准定位时存在一定的局限性：(1)经典的深度学习算法需要海量训练数据，通常需要大量的标注数据，这对于工业零部件视觉精准定位数据采集来说是一个挑战；(2)模型参数量大、对硬件资源要求高，经典的深度学习算法模型参数量偏大，如YOLOv3模型参数量为162 MB，大模型意味着对硬件资源的占用也相对较多，从而需要大量的计算资源来进行训练和推理，这对于一些小型或中小型企业来说是一个负担；(3)稳定性问题，工业环境中极易存在噪声和干扰，在复杂的光线和噪声环境条件下，深度学习算法可能会出现性能下降或无法工作的情况；(4)工业零部件视觉定位精度要求高，针对零部件定位或插件、精密部件的螺孔定位等，有的定位精度要求在毫米量级，有的定位精度要求在百分之一毫米量级，在机器人、相机作用距离和相机分辨率固定的情况下对图像的定位算法精度要求极高(2～5 pixel)，而经典的深度学习算法会出现检测框冗余及不精确，可能会导致其不能直接应用于工业零部件像素级精准定位。针对以上问题，本文构建了一种工业零部件精准定位的轻量化深度学习网络(Industry Light Weight Localization Network，ILWLNet)。网络整体结构采用Encoder-Decoder架构，Encoder采用多级bottleneck模块[31]，内部融入非对称卷积和空洞卷积，可以有效降低特征提取参数，增大感受野；Decoder中的上采样卷积同样融入和非对称卷积和空洞卷积，恢复图像的同时进一步降低模型参数；Encoder与Decoder对应特征层实施融合拼接，促使Encoder在上采样卷积时可以获得更多的高分辨率信息，进而更完备地重建出原始图像细节信息。最后，利用加权的Hausdorff距离构建了Decoder输出层与定位坐标点的关系。该轻量化深度学习定位网络具有定位精度高、准确率高和抗干扰能力强等特性，基本满足工业零部件精准定位的需求。

2　原　理

2.1　零部件光学影像定位系统的硬件构成

图1为零部件光学影像定位系统的硬件组成。它主要由照明子系统、图像采集子系统、机械运动子系统、机器人以及计算机组成。照明子系统由光源和光源控制器构成；图像采集子系统由镜头、CCD相机和图像采集卡等构成；机械运动子系统由传送带、卡槽、支撑座和伺服运动系统等构成。

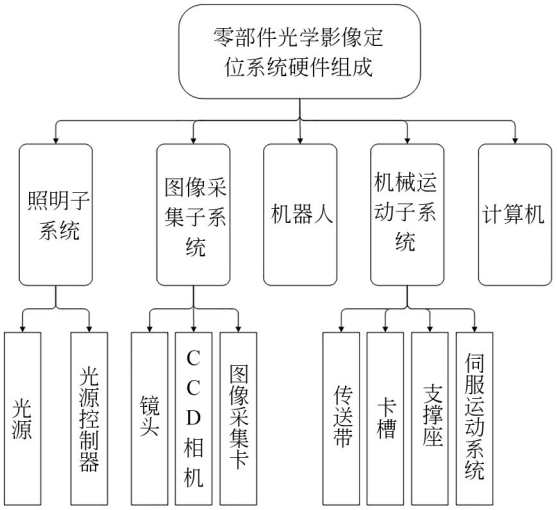


图1　定位系统硬件组成

Fig. 1　Hardware components of localization system

2.2　工作原理

精准定位系统的基本工作原理如下：由运动控制系统对物件进行运动控制，配合卡槽位完成对物件的粗定位；光学子系统负责打光、由成像系统获取图像数据，经过图像采集卡进行A/D转换，转换成数字信号、送入至计算机进行零部件的像素级精准定位；将定位后的精准像素坐标经过坐标转换为机器人的空间物理坐标，控制机器

人进行精准打孔、插件等，如图 2 所示。

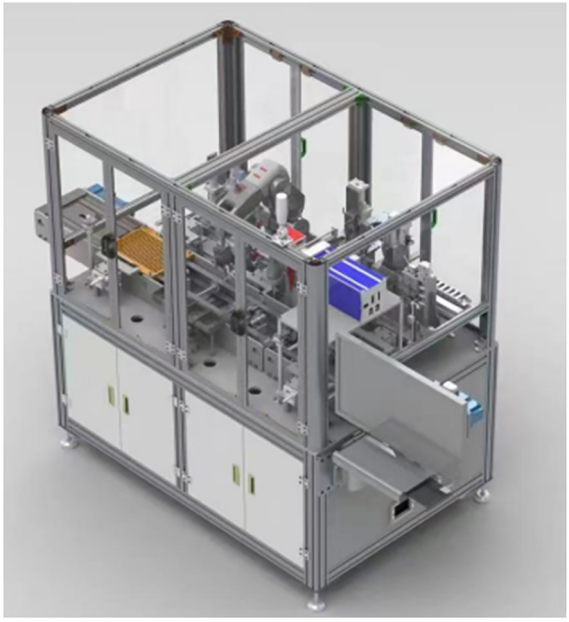


图 2　零部件光学影像定位系统

Fig. 2　Optical image localization system for industry component

3　轻量化深度学习定位网络

3.1　网络架构

ILWLNet 网络结构如图 3 所示。网络总体设计选用 Encoder-Decoder 架构，Encoder 和 Decoder 分别由三级 Down_Block 和三级 Up_Block 级联而成。每级 Down_Block 由 4 个配置不同的 bottleneck 模块串联构成；每级 Up_Block 对输入数据进行上采样卷积后与对应的 Down_Block 输出数据进行 padding，并送入至 conv_async 模块，促使 Encoder 在进行上采样卷积时可以获得更多的高分辨率信息，进而更完整地重建出原始图像的细节信息。目标点的个数由 Encoder 输出特征图与 Decoder 输出语义层经过全连接、拼接、全连接回归获得。最后，利用加权的 Hausdorff 距离建立 Decoder 输出语义层与定位坐标点的关系，并结合回归的目标点数量偏差构建最终的损失函数形成闭环训练；推理阶段，Decoder 输出语义层经过 Otsu 分割即可得到最终的零部件精准定位坐标。

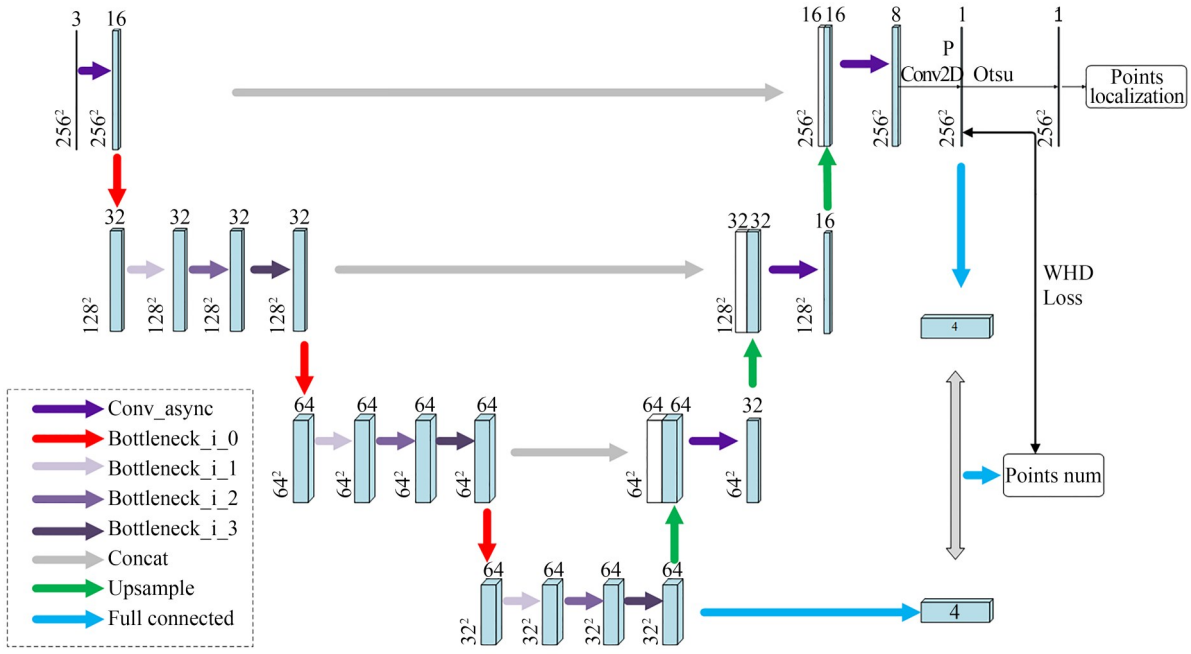


图 3　轻量化深度学习定位网络结构

Fig. 3　Lightweight deep learning localization network architecture

ILWLNet 网络算法流程如图 4 所示，包括 Encoder、Decoder、目标点数量回归和目标点位置回归。

Encoder 流程如下：输入的图像首先归一化成 3×256×256，经过 conv_async 输出 16×256×256 的特征图，记为 X_1，其中 conv_async 内部进行

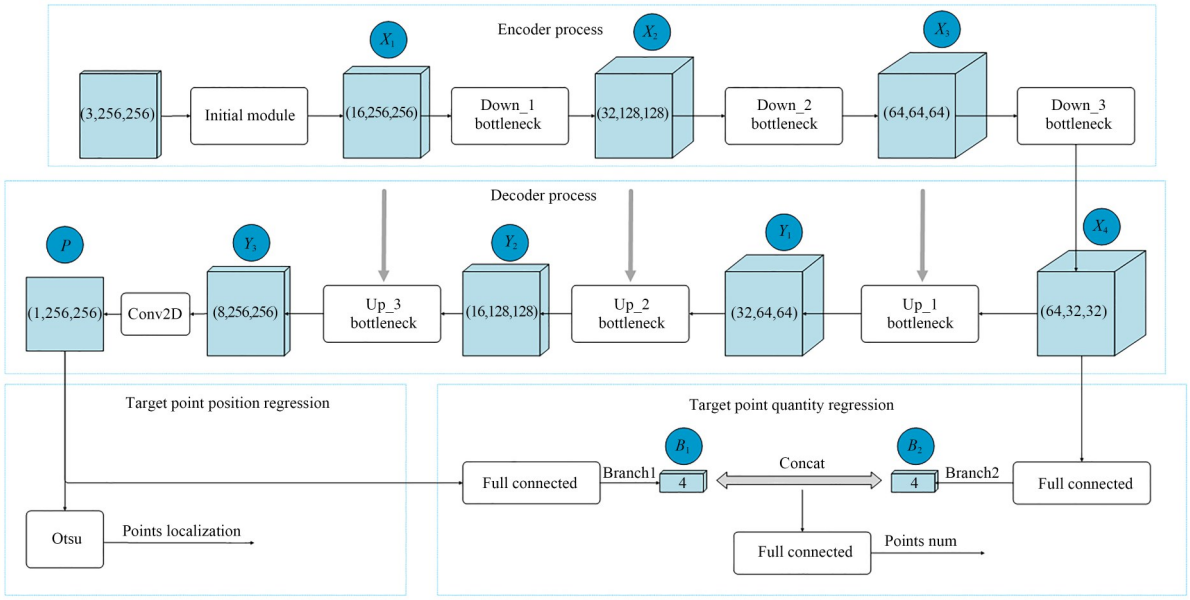


图 4　轻量化深度学习定位网络算法流程

Fig. 4　Flowchart lightweight deep learning localization network algorithm

了 1×1 的投影卷积和 3×1，1×3 的非对称卷积，得到初级特征图的同时，降低了直接卷积的运算量和模型参数；其次，经过 Down_Block_1 模块得到 32×128×128 的特征图，实现下采样和进一步的深度特征提取，记为 X_2，其中 Down_Block_1 内部包含 bottleneck_1_0，bottleneck_1_1，bottleneck_1_2 和 bottleneck_1_3 四个模块，具备降采样、低卷积运算量和拓展感受野的能力，多级串联也提升了网络整体的非线性拟合能力；再次经过 Down_Block_2 模块得到 64×64×64 的特征图，记为 X_3，其中 Down_Block_2 内部包含 bottleneck_2_0，bottleneck_2_1，bottleneck_2_2 和 bottleneck_2_3 四个模块，功能同 Down_Block_1；最后，经过 Down_Block_3 模块得到 64×32×32 的特征图，记为 X_4，其中 Down_Block_3 内部包含 bottleneck_3_0，bottleneck_3_1，bottleneck_3_2 和 bottleneck_3_3 四个模块，功能同 Down_Block_1。

Decoder 流程如下：首先，特征图 X_4 和 X_3 作为参变量进入至 Up_Block_1 模块，得到 32×64×64 的语义图 Y_1，其中 Up_Block_1 内部实现 X_4 上采样并与 X_3 特征拼接、拼接后的特征图进入 conv_async 模块，Up_Block_1 具备上采样、低卷积运算和多特征融合能力，促使 Encoder 在进行上采样卷积时可以获得更多的高分辨率信息；其次，语义图 Y_1 和特征图 X_2 作为参变量进入 Up_Block_2 模块，得到 16×128×128 的语义图 Y_2，功能实现同 Up_Block_1；最后，语义图 Y_2 和特征图 X_1 作为参变量进入至 Up_Block_3 模块，得到 8×256×256 的语义图 Y_3。

目标点数量回归：特征图 X_4 与语义图 P 分别经过全连接得到 Branch2 的 B_2 和 Branch1 的 B_1，经过特征拼接及全连接回归得到最终的定位目标点数量。

目标点位置回归：Decoder 得到的最终语义图 P 可以体现每个坐标点的激活概率值，但是并不能直接返回预测目标点坐标；通过 3.3 节构建的加权 Hausdorff 距离损失函数，将预测点坐标与语义图 P 进行关联，再次融合目标点数量预测误差，利用 3.4 节构建的最终损失函数进行闭环训练。模型闭环训练收敛且达到指定误差后，推理阶段将语义图 P 经过 Otsu 分割即可得到最终的目标点位置。

3.2　模块结构

图 5 介绍了 ILWLNet 网络的通用网络模块和外层网络模块。通用网络模块：conv_async 包含 1×1 投影卷积、3×1 与 1×3 的非对称卷积（内置空洞卷积），实现浅层特征提取同时降低了直接卷积的运算量。外层网络模块：Down_Block_i 包含 bottleneck_i_0，bottleneck_i_1，bottleneck_i_2 和 bottleneck_i_3 四部分，具备降采样、低卷积

运算量和拓展感受野的能力，多级串联可提升网络整体的非线性拟合能力；Up_Block_i内部包含上采样、特征拼接和conv_async模块，具备特征尺寸扩张、低卷积运算和多特征融合能力。

图6介绍了ILWLNet网络的内层网络模块。bottleneck_i_0：左链路包括MaxPooling2D和Padding模块，右链路包含2×2卷积（步长为2）、3×3卷积、1×1扩张卷积和Dropout2d，两条链路特征相加并经过PReLU输出，各个卷积模块后面都追加BatchNorm和PReLU模块用于提升非线性能力和降低过拟合风险。bottleneck_i_1：右链路包括1×1投影卷积、3×3卷积、1×1扩张卷积和Dropout2d，直通的左链路与右链路特征相加并经过PReLU输出；各个卷积模块后面都追加BatchNorm和PReLU。bottleneck_i_2：右边链路包括1×1投影卷积、3×3空洞卷积、1×1扩张卷积和Dropout2d，直通的左链路与右边链路特征相加并经过PReLU输出；各个卷积模块后面都追加BatchNorm和PReLU。bottleneck_i_3：右链路包括1×1投影卷积、5×1和1×5的非对称卷积、1×1扩张卷积和Dropout2d，直通的左链路与右链路特征相加并经过PReLU输出；各个卷积模块后面都追加BatchNorm和PReLU。

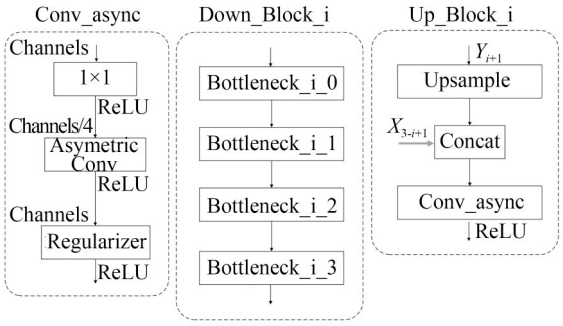
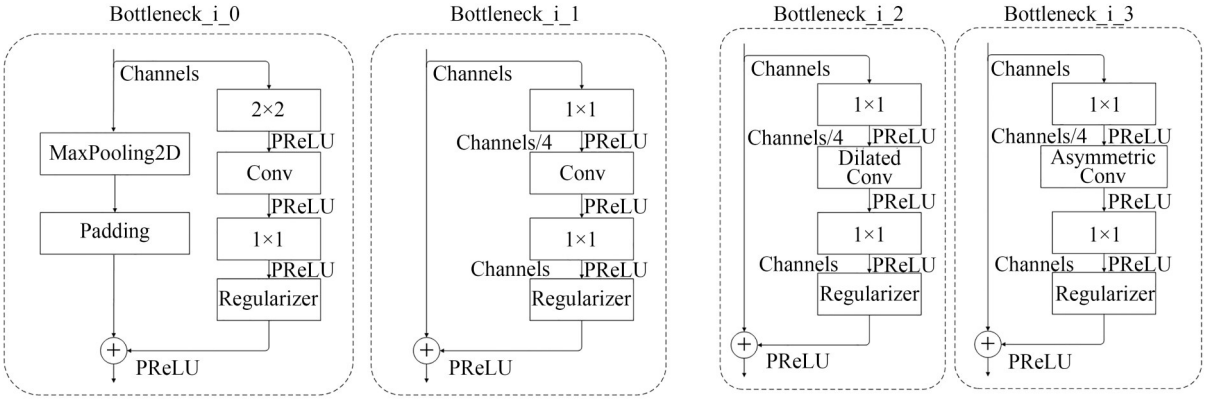


图5　外层网络模块结构

Fig. 5　Structure of outer network module



(a) Bottleneck_i_0和bottleneck_i_1模块
(a) Bottleneck_i_0 and bottleneck_i_1 module

(b) Bottleneck_i_2和bottleneck_i_3模块
(b) Bottleneck_i_2 and bottleneck_i_3 module

图6　内层网络模块结构

Fig. 6　Structure of inner network modules

3.3　加权Hausdorff距离

本文的损失函数构建来源于Hausdorff距离[32]。X, Y是两个无序的点集，$d(x,y)$表示两个点集X, Y之间的距离，其中，$x\in X, y\in Y$，本文采用欧氏距离。X, Y拥有的点的数量可以不同$\Omega\subset \mathbf{R}^2$表示包含所有可能点的空间，则集合$X\subset \mathbf{R}$与集合$Y\subset \mathbf{R}$的Hausdorff距离定义为：

$$d_{\mathrm{H}}(x,y)=\max(d(X,Y),d(Y,X)),\quad(1)$$

$$d(X,Y)=\sup_{x\in X}\inf_{y\in Y}d(x,y),\quad(2)$$

$$d(Y,X)=\sup_{y\in Y}\inf_{x\in X}d(x,y),\quad(3)$$

其中：

$$\mathrm{d}(x,y)\leqslant d_{\max}=\max_{x\in\Omega,y\in\Omega}d(x,y).\quad(4)$$

Hausdorff距离的最大短板是对轮廓上的点的距离计算相对敏感[33]。为了优化这个问题，通常采用加权的Hausdorff距离，如下：

$$d_{\mathrm{AH}}(x,y)=d_{\mathrm{A}}(X,Y)+d_{\mathrm{A}}(Y,X),\quad(5)$$

$$d_{\mathrm{A}}(X,Y)=\frac{1}{|X|}\sum_{x\in X}\min_{y\in Y}d(x,y),\quad(6)$$

$$d_{\mathrm{A}}(Y,X)=\frac{1}{|Y|}\sum_{y\in Y}\min_{x\in X}d(x,y), \quad (7)$$

其中：$|X|$,$|Y|$分别为集合X,Y点的数量，在本文中，Y表示图像目标定位点坐标标签集合，X表示预测的图像目标定位点坐标集合。

语义图P可以得到每个点的激活概率值，但是并不能返回预测点的坐标。为了建立语义结果与坐标点最终的联系，采用加权Hausdorff距离（Weighted Hausdorff Distance，WHD）函数进行构建，即：

$$d_{\mathrm{WH}}(x,y)=d_{\mathrm{W}}(X,Y)+d_{\mathrm{W}}(Y,X), \quad (8)$$

$$d_{\mathrm{W}}(X,Y)=\frac{1}{S+\varepsilon}\sum_{x\in\Omega}p_x\min_{y\in Y}d(x,y), \quad (9)$$

$$d_{\mathrm{W}}(Y,X)=\frac{1}{|Y|}\sum_{y\in Y}M_{\alpha}{}_{x\in\Omega}d_p(x,y), \quad (10)$$

$$d_{\mathrm{p}}(x,y)=p_x d(x,y)+(1-p_x)d_{\max}, \quad (11)$$

$$S=\sum_{x\in\Omega}p_x, \quad (12)$$

$$M_{\alpha}{}_{a\in\mathrm{A}}[f(\lambda)]=\left(\frac{1}{|A|}M_{a\in\mathrm{A}}f^{\alpha}(\lambda)\right)^{\frac{1}{\alpha}}. \quad (13)$$

这里ε设定为10^{-6}，M_{α}为广义均值函数，$p_x\in[0,1]$为语义图P中每个坐标对应的概率值。

3.4 损失函数

为了训练整个ILWLNet，结合加权Hausdorff距离函数和回归的点的数量差，构建ILWLNet整体损失函数如下：

$$L(p,G)=d_{\mathrm{WH}}(p,G)+L_{\mathrm{reg}}(C-\hat{C}(p)), \quad (14)$$

其中：G为图像的标签包含目标点的坐标和目标数量，$C=|G|$,$\hat{C}(p)$为预测的目标点数量。

$$L_{\mathrm{reg}}(x)=\begin{cases}0.5x^2 & \text{for}\,|x|<1\\ |x|-0.5 & \text{for}\,|x|\geqslant 1\end{cases}. \quad (15)$$

$L_{\mathrm{reg}}(x)$为回归项，这里采用L_1平滑函数。

4 实验与结果分析

4.1 实验数据集与训练策略

为了充分验证ILWLNet的性能，本文选用3组数据集。数据集一：笔记本螺孔定位数据集，原始数据共358张，原图与中心点坐标融合后的示例图见图7的第一行。数据集二：笔记本螺孔定位数据集，原始数据共318张，原图与中心点坐标融合后的示例图见图7的第二行。数据集三：纱车机定位数据集，原始数据共529张，原图与中心点坐标融合后的示例图见图7的第三行。

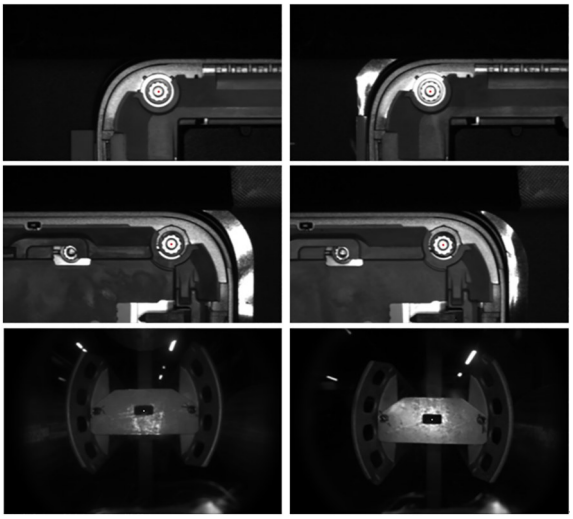


图7 数据集示例

Fig. 7 Dataset sample

ILWLNet训练和测试的模型均运行在Ubuntu16.04操作系统的工作站上。硬件的具体配置如下：CPU，Intel Xeon（R）CPU E5-1650 V4 3.6 GHz 12核；内存，32 GB，显卡，Nvidia GeForce 1080Ti；软件配置如下：CUDA，10.0；cuDNN，7.6.1.34；深度学习框架，PyTorch 1.0.0；Python，3.6版本。选择Loss值最小的模型作为单次测试验证的最优模型，实验模型超参数配置如表1所示。

表1 模型超参数设置

Tab. 1 Model super parameter setting

Super parameter	Value
Learning Rate	4×10^{-5}
Optimizer	SGD
Radius	5
Batch size	64
Momentum	0.9

4.2 评价指标

为了充分验证ILWLNet的性能，本文利用Precision，Root Mean Squared Error（RMSE）和Mean Average Hausdorff Distance（MAHD）进行衡量。其计算公式如下：

$$P=\frac{TP}{TP+FP},\tag{16}$$

$$RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^{N}|e_i|^2},\tag{17}$$

$$e_i=\hat{C}_i-C_i,\tag{18}$$

$$MAHD=\frac{1}{N}\sum_{n=1}^{N}d_{\mathrm{AH}}(x,y),\tag{19}$$

其中：TP（True Positive）表示预测正确，实际为真；FP（False Positive）表示预测错误，实际为真；FN表示预测错误，实际为假；TN表示预测正确，实际为假；N表示测试集的数量，C_i表示第i张图片中实际目标的个数，$\hat{C}_i$表示预测的第i张图片目标个数。

4.3　实验结果与分析

4.3.1　对比分析

在本次实验中，综合考虑不同配比的训练数据对模型定位性能的影响，对训练数据、验证数据和测试数据选用以下配比进行实验，Train∶Val∶Test分别设置为10%∶10%∶80%，20%∶10%∶70%，30%∶10%∶60%，40%∶10%∶50%，50%∶10%∶40%，60%∶10%∶30%，70%∶10%∶20%，80%∶10%∶10%，在图8和图9的Loss曲线图中分别对应loss_1～loss_9。使用ILWLNet算法训练，对8种不同配比的数据训练并将训练过程Loss值作可视化处理，以利于分析训练过程中损失函数Loss值的变化情况。

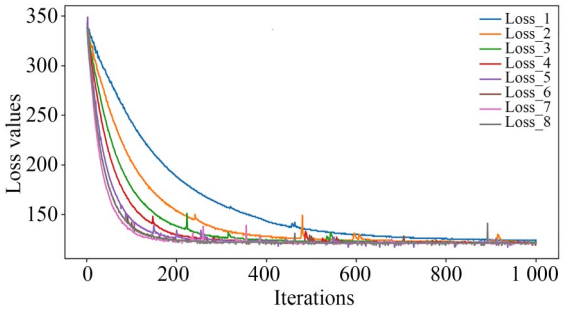


图8　数据一训练的loss曲线

Fig. 8　Trainning loss curves of dataset one

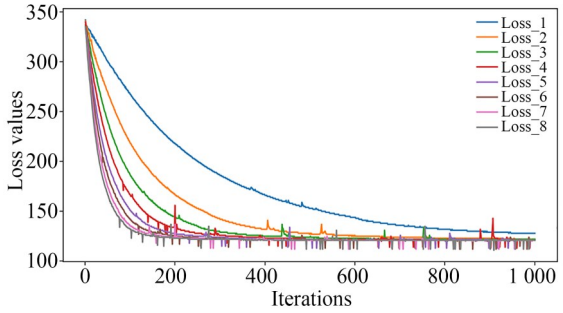


图9　数据二训练的loss曲线

Fig. 9　Trainning loss curves of dataset two

如图8和图9所示，两种数据集在训练中Loss值的整体变化趋势一致，随着训练数据量的增加，收敛速度越快。其中，训练数据大于等于40%，200次迭代后基本收敛，然后逐渐趋于稳定，但依然出现小幅震荡；训练数据大于等于20%且小于等于40%，600次迭代后基本收敛，然后逐渐趋于稳定；训练数据占比10%，1 000次训练后仍会收敛。结合表3和表4，虽然训练样本集在20%比例时，推理结果可以得到98%以上，但是为了达到定位精度小于等于5 pixel的识别率大于等于99.5%，训练样本比例建议大于等于50%。与深度学习所需要的海量数据相比，ILWLNet只需要150张左右相对较少的数据即可收敛到很好的效果，更适用于工业智能制造的实际应用场景。

本文以Precision，MSE和MAHD作为评价指标，计算不同配比的训练数据在1 000迭代后得到最优的收敛模型用于ILWLNet推理，推理计算得到测试集的Precision和RMSE值；其中，训练集、验证集和测试集数据三者无交集。判定为定位准确的参数条件为：推理得到的中心点坐标与标签的中心点坐标均方差误差小于等于5，结果如表2和表3所示。

通过表3和表4可以看出，ILWLNet网络可以进行零部件的精准定位。当训练数据占比在20%以上时，定位误差小于等于5 pixel可以取得

表2　ILWLNet在数据集一中的定位测试结果

Tab. 2　ILWLNet localization test results on dataset one

(Train,Val,Test)	Precision/%	RMSE	MAHD	Model
(10%,10%,80%)	95.40	2.55	5.11	
(20%,10%,70%)	99.20	1.96	3.92	
(30%,10%,60%)	98.60	2.72	5.44	
(40%,10%,50%)	99.43	1.97	3.93	
(50%,10%,40%)	100	1.45	2.90	
(60%,10%,30%)	100	1.45	2.91	57.4k
(70%,10%,20%)	100	1.98	3.96	
(80%,10%,10%)	100	1.70	3.40	

表3　ILWLNet在数据集二中的定位测试结果

Tab. 3　ILWLNet localization test results on dataset two

(Train,Val,Test)	Precision/%	RMSE	MAHD	Model
(10%,10%,80%)	74.50	4.12	8.24	57.4k
(20%,10%,70%)	98.18	2.24	4.49	
(30%,10%,60%)	98.90	1.98	3.96	
(40%,10%,50%)	100	1.85	3.7	
(50%,10%,40%)	100	1.68	3.37	
(60%,10%,30%)	100	2.05	4.10	
(70%,10%,20%)	100	1.59	3.17	
(80%,10%,10%)	100	1.52	3.01	

表4　ILWLNet推理时间测试结果

Tab. 4　Test results of ILWLNet inference time(ms)

数量	GPU推理平均时间	CPU阈值分割平均时间	ILWLNet网络平均运行时间
100张	8.34	86.58	94.92

至少98%的准确率，平均准确率高于99%；当训练数据集在50%及其以上时，定位误差小于等于5 pixel的准确率可以达到100%。随着训练数据的增加，测试推理得到的RMSE和MAHD误差值呈整体变小的趋势。相比较经典的深度学习网络(YOLO3，162 MB)，ILWLNet模型的参数量非常少，只有57.4 kB。这是由于ILWLNet网络的Decoder和Encoder的层级较少，均为3层，利用Bottleneck技术代替直接的卷积方式且内部融入了非对称卷积和空洞卷积，因此模型参数量非常小。利用构建的加权Hausdorff距离将输出的语义概率图与定位坐标点进行关联并构建Loss函数进行闭环回归，ILWLNet能够用于零部件的精准定位。另外，ILWLNet网络采用Up_Block与Down_Block的对应特征层融合以及多级Bottleneck级联，不仅提升了整体网络的非线性而且较好地恢复了原始图像的细节信息，配合构建的加权Hausdorff距离，表明ILWLNet得到了非常理想的定位精度。由于构建的ILWLNet模型参数较小，仅需相对较少的样本训练即可取得比较好的结果，通过表2和表3的不同配比实验定位结果得到证实；因此，ILWLNet亦可适用于小样本训练，进而满足工业光学影像定位的小样本实际需求。

表4展示了ILWLNet推理时间测试结果，测试图片共100张；其中，ILWLNet网络前向推理平均时间为8.34 ms/frame，阈值分割的平均时间为86.58 ms/frame，ILWLNet的整体运行时间平均为94.92 ms/frame，满足工业光学影像精准定位的200 ms/frame需求。

4.3.2　测试集定位结果与分析

在闭环1 000次迭代训练过程中，最小loss值对应的模型记为ILWLNet在该次训练过程的最优模型。为了检验ILWLNet网络的实际定位效果，利用3组测试集数据进行验证。图10～图12分别显示了数据集一、数据集二和数据集三的定位结果。其中，第一行显示的是原始图、第二行显示的是语义概率图P经过归一化得到的结果、第三行显示的是语义概率图P经过Otsu阈值分割后的二值图、第四行显示的是原图加载预测中心点后的融合图。

从第二行可以看出，ILWLNet通过Encoder-Decoder架构可以还原出原图(图像复原，用于缺陷检测[34])及其语义信息(语义分割[35])，Decoder的输出语义概率图P呈现了完整的原图信息；另外，ILWLNet将预测语义概率图P与定位中心点坐标通过加权Hausdorff距离进行了有效关联，因此，概率图中的螺孔中心点区域灰度明显高于图片中的任何其他区域，从而可以通过Otsu自适应分割方法将中心点区域精准分割出来。第三行呈现了语义概率图P经过Otsu自适应分割后的结果，可以看出除中心点区域为白外，图片中的其他区域均为黑，因此得到的中心点坐标不会发散。

图10左一的螺孔内部受到光照(拍摄角度和光线)的影响，小部分区域呈现黑色；左二的外壳的最左边没有拍摄清楚，整体呈现黑色，与其他图相差较大。图11右一和右二的螺孔最里面白色内圈明显不圆(拍摄角度影响)。此类图片如果采用传统的圆拟合、模板匹配等方法，定位结果会大打折扣。仿真及实测结果表明，ILWLNet算法不仅可以对正常的零部件样本进行精准定位，而且对受到光线干扰、部分缺损的零部件图片的定位效果仍然很好。

图12为纱车机图像定位结果。此组图像受光照影响较为严重，四组图像均出现明显的反光，右二和右一的矩形中心孔周围受到了强光干

扰;另外,图片中的矩形孔并不唯一,利用传统模板匹配方法可能会出现误匹配,再叠加光照的影响,传统定位算法性能会受到严重影响。通过仿真及实测的结果可以看出,ILWLNet算法不仅适用于圆形的零部件定位,而且适用于其他样式的零部件精准定位,同时兼顾较好的抗干扰效果。

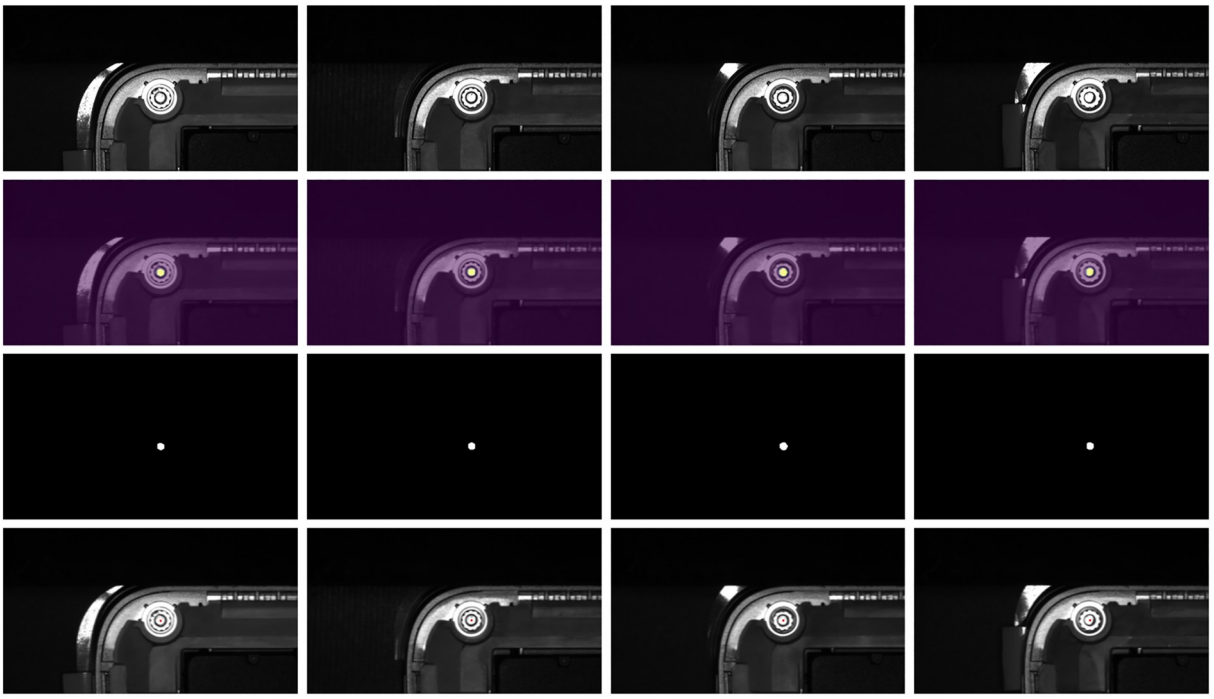


图 10　数据集一的测试定位结果

Fig. 10　Localization test results on dataset one

图 11　数据集二的测试定位结果

Fig. 11　Localization test results on dataset two

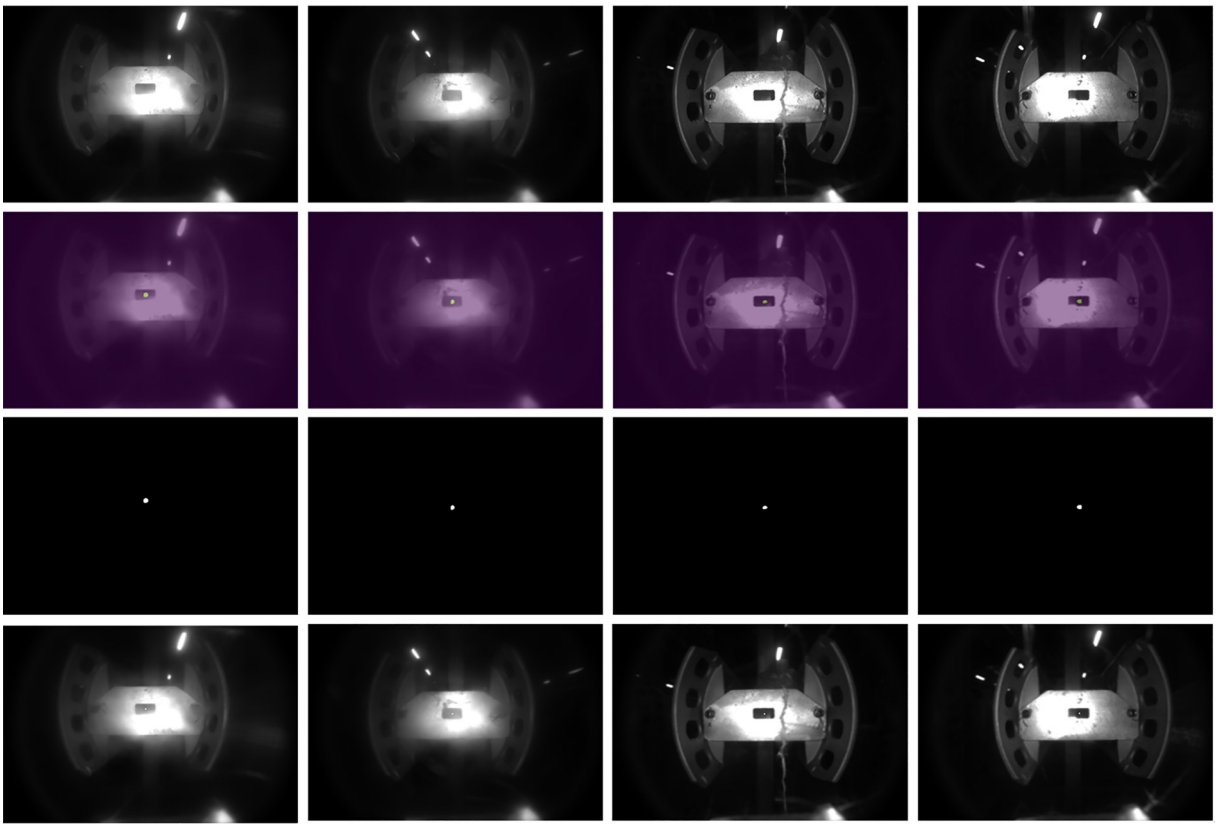


图 12　数据集三的测试定位结果

Fig. 12　Localization test results on dataset three

4.4　消融实验

为了验证 conv_async，Down_Block_1，Down_Block_2，Down_Block_3，Up_Block_1，Up_Block_2，Up_Block_3，特征拼接融合(Encoder 与 Decoder 对应层 padding 及融合，记为 Fusion)以及 WHD 对模型性能的影响，对 ILWLNet 进行消融研究，实验结果如表 6 所示。测试数据选用数据集一，其 Train∶Val∶Test 数据集比例选用 50%∶10%∶40%。

通过表 5 可以看出：不包含 WHD 模块，ILWLNet 算法不收敛，表明 WHD 模块可以有效关联输出语义概率图 P 与定位坐标点的关系。配置一：只配置 WHD 模块，其他模块中的卷积均采用经典的 3×3 卷积，此时模型大小为 85.1 kB，识别率为 87.94%，说明本文架构不采用多级 bottleneck 级联仍可以收敛，不过相比较 ILWL-

表 5　不同策略对模型性能的影响对比

Tab. 5　Comparison of impact of different strategies on model performance

conv_asyn	Down_Block_1	Down_Block_2	Down_Block_3	Up_Block_1	Up_Block_2	Up_Block_3	Fusion	WHD	Precision	Model
√	√	√	√	√	√	√	√		—	—
								√	87.94%	85.1k
							√	√	93.62%	109k
√	√	√	√				√	√	92.91%	92.2k
√				√	√	√	√	√	87.23%	75.7k
√	√	√	√	√	√	√		√	85.82%	54.8k
√	√	√	√	√	√	√	√	√	**100%**	**57.4k**

Net，模型参数量增加了 27.7 kB，定位准确率降低了 12.06%。配置二：只配置 Fusion 以及 WHD 模块，其他模块同配置一，模型大小为 109 kB，识别率为 93.62%；相比较配置一，通过 Encoder 与 Decoder 对应层的融合使得识别率提升约 5.68%，但是模型参数增加了接近 24 kB，表明相对较大的模型在充分训练后一定程度上可以取得相对较好的识别效果。配置三：配置 conv_async，Down_Block_1，Down_Block_2，Down_Block_3，Fusion 和 WHD 模块，相比较配置一准确率提升 4.97%，模型大小减小 7.1 kB，表明 bottleneck_i 模块可降低模型参数量，同时多级 bottleneck_i 模块级联增加了网络的非线性，可提升识别率。配置四：采用 conv_async，Up_Block_1，Up_Block_2，Up_Block_3，Fusion 和 WHD 模块，准确率基本不变，略微下降，模型参数减小 9.4 kB，表明 Up_Block 中的非对称卷积和空洞卷积起到了降低模型参数量的作用，但是没有与 Encoder 的对应层融合，对识别率提升没有太大影响。配置五：ILWLNet 算法中去掉 Fusion 模块，其他模块同配置一，模型参数量相比配置一降低了 30.3 kB，但是识别率却降低了 2% 左右，表明 Down_Block 的 bottleneck_i 模块以及 Up_Block 融合的 conv_async 模块可有效降低模型参数量，Fusion 模块对于网络整体性能的提升起到了很大的作用；通过配置一和配置二的比较，Fusion 模块提升 4.97% 的识别率（各模块内部卷积均为常用的 3×3 卷积），而相比较 ILWLNet，配置五的识别率却降低了 14.2% 左右，表明 conv_async、Down_Block、Up_Block 和 Fusion 模块的级联进一步提升了整体 ILWLNet 的非线性及识别性能。

5　结　论

本文根据现代工业光学影像定位精度高、占用资源小、抗干扰好、速度快的要求，构建了零部件视觉精准定位的轻量化深度学习网络，并介绍了零部件光学影像定位系统的硬件结构和工作原理。然后，阐述了轻量化深度学习定位网络的架构、模块结构、加权 Hausdorff 距离及其损失函数。实验结果表明：轻量化深度学习定位网络模型参数为 57.4k；通过实际产线数据仿真，训练数据集多于 150 张，定位精度小于等于 5 pixel 的识别率不小于 99.5%，基本满足工业零部件定位的精度高、准确率高和抗干扰能力强等要求。仿真测试表明，ILWLNet 已经在笔记本螺孔和纱车机实际工业产线数据中取得了较好的识别率和定位精度，可是后续将 ILWLNet 进行产线实际应用，需要考虑数据的泛化性，即需要更多不同种类的产线定位数据集进行验证以及上线测试。此外，ILWLNet 的推理时间大多消耗在 Encoder-Decoder 阶段后的 Otsu 自适应分割流程中，因此后续会对其进行进一步优化，提升推理速度。

参考文献：

[1] 高贯斌，牛锦鹏，刘飞，等. 基于各向异性误差相似度的六自由度机器人定位误差补偿[J]. 光学 精密工程，2022，30(16)：1955-1967.
GAO G B，NIU J P，LIU F，*et al*. Positioning error compensation of 6-DOF robots based on anisotropic error similarity[J]. *Opt. Precision Eng.*，2022，30(16)：1955-1967.（in Chinese）

[2] 任同群，江海川，张建昆，等. 微小磁性零件装配设备自动标定及误差补偿[J]. 光学 精密工程，2023，31(2)：214-225.
REN T Q，JIANG H CH，ZHANG J K，*et al*. Automatic calibration and error compensation for micro magnetic parts assembly equipment[J]. *Opt. Precision Eng.*，2023，31(2)：214-225.（in Chinese）

[3] 陈平，雷学军，李灿，等. 3D位姿估计结合阻抗控制的沉孔装配[J]. 光学 精密工程，2022，30(22)：2889-2900.
CHEN P，LEI X J，LI C，*et al*. Assembly of countersunk-hole parts based on 3D pose estimation and impedance control[J]. *Opt. Precision Eng.*，2022，30(22)：2889-2900.（in Chinese）

[4] 潘睿志，林涛，李超，等. 基于深度学习的多尺寸汽车轮辋焊缝检测与定位系统研究[J]. 光学 精密工程，2023，31(8)：1174-1187.
PAN R ZH，LIN T，LI CH，*et al*. A research on multi size automobile rim weld detection and positioning system based on depth learning[J]. *Opt. Precision Eng.*，2023，31(8)：1174-1187.（in Chinese）

[5] KRIZHEVSKY A，SUTSKEVER I，HINTON G E. ImageNet classification with deep convolutional

neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90.

[6] GIRSHICK R, DONAHUE J, DARRELL T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. 2014 *IEEE Conference on Computer Vision and Pattern Recognition. June* 23-28, 2014, *Columbus*, *OH*, *USA*. IEEE, 2014: 580-587.

[7] HE K, ZHANG X, REN S, *et al*. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904-1916.

[8] GIRSHICK R. Fast R-CNN[C]. 2015 *IEEE International Conference on Computer Vision* (*ICCV*). *December* 7-13, 2015, *Santiago*, *Chile*. IEEE, 2016: 1440-1448.

[9] 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.

[10] REDMON J, DIVVALA S, GIRSHICK R, *et al*. You only look once: unified, real-time object detection[C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 27-30, 2016, *Las Vegas*, *NV*, *USA*. IEEE, 2016: 779-788.

[11] LIU W, ANGUELOV D, ERHAN D, *et al*. *SSD*: *Single Shot MultiBox Detector*[M]. Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.

[12] 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*, *Italy*. IEEE, 2017: 2980-2988.

[13] HE K M, GKIOXARI G, DOLLÁR P, *et al*. Mask R-CNN[C]. 2017 *IEEE International Conference on Computer Vision* (*ICCV*). *October* 22-29, 2017, *Venice*, *Italy*. IEEE, 2017: 2961-2969.

[14] CAI Z W, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection[C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition. June* 18-23, 2018, *Salt Lake City*, *UT*, *USA*. IEEE, 2018: 6154-6162.

[15] ZHOU X Y, ZHUO J C, KRÄHENBÜHL P. Bottom-up object detection by grouping extreme and center points[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 15-20, 2019, *Long Beach*, *CA*, *USA*. IEEE, 2020: 850-859.

[16] LI Y H, CHEN Y T, WANG N Y, *et al*. Scale-aware trident networks for object detection[C]. 2019 *IEEE/CVF International Conference on Computer Vision* (*ICCV*). *October* 27 - *November* 2, 2019, *Seoul*, *Korea* (*South*). IEEE, 2020: 6054-6063.

[17] ZHU C C, HE Y H, SAVVIDES M. Feature selective anchor-free module for single-shot object detection[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 15-20, 2019, *Long Beach*, *CA*, *USA*. IEEE, 2020: 840-849.

[18] TAN M X, PANG R M, LE Q V. EfficientDet: scalable and efficient object detection[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 13-19, 2020, *Seattle*, *WA*, *USA*. IEEE, 2020: 10781-10790.

[19] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection [EB/OL]. 2020: *arXiv*: 2004.10934. https://arxiv.org/abs/2004.10934.

[20] SUN P Z, ZHANG R F, JIANG Y, *et al*. Sparse R-CNN: end-to-end object detection with learnable proposals[C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 20-25, 2021, *Nashville*, *TN*, *USA*. IEEE, 2021: 14454-14463.

[21] BEAL J, KIM E, TZENG E, *et al*. Toward transformer-based object detection [EB/OL]. 2020: *arXiv*: 2012.09958. https://arxiv.org/abs/2012.09958.

[22] QIU H, MA Y, LI Z, *et al*. BorderDet: border feature for dense object detection[EB/OL]. 2020: *arXiv*: 2007.11056. https://arxiv.org/abs/2007.11056.

[23] YANG H, DENG R, LU Y, *et al*. CircleNet: anchor-free detection with circle representation[EB/OL]. 2020: *arXiv*: 2006.02474. https://arxiv.org/abs/2006.02474.

[24] CHEN Y H, ZHANG Z, CAO Y, *et al*. Rep-

Points v2: verification meets regression for object detection[C]. *Proceedings of the 34th International Conference on Neural Information Processing Systems. December 6 - 12, 2020, Vancouver, BC, Canada.* New York: ACM, 2020: 5621-5631.

[25] YAO Z, AI J, LI B, *et al*. Efficient DETR: improving end-to-end object detector with dense prior. [EB/OL]. 2021: *arXiv*: 2104.01318. https://arxiv.org/abs/2104.01318.

[26] LANG S, VENTOLA F, KERSTING K. DAFNe: a one-stage anchor-free approach for oriented object detection. [EB/OL]. 2021: *arXiv*: 2109.06148. https://arxiv.org/abs/2109.06148.

[27] GE Z, LIU S, WANG F, *et al*. YOLOX: exceeding YOLO series in 2021[EB/OL]. 2021: *arXiv*: 2107.08430. https://arxiv.org/abs/2107.08430.

[28] YU R, DU D W, LALONDE R, *et al*. Cascade transformers for end-to-end person search [C]. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 18-24, 2022, New Orleans, LA, USA*. IEEE, 2022: 7267-7276.

[29] CHENG G, WANG J B, LI K, *et al*. Anchor-free oriented proposal generator for object detection [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1-11.

[30] HUANG Z C, LI W, XIA X G, *et al*. A general Gaussian heatmap label assignment for arbitrary-oriented object detection[J]. *IEEE Transactions on Image Processing*, 2022, 31: 1895-1910.

[31] PASZKE A, CHAURASIA A, KIM S, *et al*. ENet: a deep neural network architecture for real-time semantic segmentation[EB/OL]. 2016: *arXiv*: 1606.02147. https://arxiv.org/abs/1606.02147.

[32] DUBUISSON M P, JAIN A K. A modified Hausdorff distance for object matching[C]. *Proceedings of 12th International Conference on Pattern Recognition. October 9-13, 1994, Jerusalem, Israel.* IEEE, 2002: 566-568.

[33] SCHUTZE O, ESQUIVEL X, LARA A, *et al*. Using the averaged Hausdorff distance as a performance measure in evolutionary multiobjective optimization[J]. *IEEE Transactions on Evolutionary Computation*, 2012, 16(4): 504-522.

[34] 乔健,陈能达,伍雁雄,等. 融合注意力机制的金属锅圆柱表面缺陷检测[J]. 光学 精密工程, 2023,31(3): 404-416.
QIAO J, CHEN N D, WU Y X, *et al*. Defect detection of cylindrical surface of metal pot combining attention mechanism [J]. *Opt. Precision Eng.*, 2023,31(3): 404-416. (in Chinese)

[35] 任凤雷,杨璐,周海波,等. 基于改进BiSeNet的实时图像语义分割[J]. 光学 精密工程,2023,31(8): 1217-1227.
REN F L, YANG L, ZHOU H B, *et al*. Real-time semantic segmentation based on improved BiSeNet[J]. *Opt. Precision Eng.*, 2023,31(8): 1217-1227. (in Chinese)

作者简介:

牛小明(1983—),男,博士研究生,高级工程师,2008年于重庆大学获得硕士学位,现为中国兵器装备集团自动化研究所部门技术总师,主要从事机器视觉目标检测、缺陷检测、OCR识别、光电目标识别及跟踪、语音识别、声纹识别、语义理解和机器人等领域的研究。E-mail: niuxiaoming1@163.comn

通讯作者:

曾　理(1959—),男,四川郫县人,博士,教授,博士生导师,1997年于重庆大学获得博士学位,主要从事工业CT重建、图像处理等方面的研究。E-mail: drlizeng@cqu.edu.cn