

基于 Transformer 的复合材料多源图像实例分割网络

柯岩¹, 傅云², 周玮珠¹, 朱伟东^{1*}

(1. 浙江大学机械工程学院, 浙江杭州 310027;
2. 浙江西子势必锐航空工业有限公司, 浙江杭州 310018)

摘要: 为提高复合材料铺放质量, 辅助现场人员快速对缺陷进行检测, 提出一种基于 Transformer 的复合材料多源图像实时实例分割网络 Trans-Yolact, 用来对复合材料缺陷进行检测、分类、分割。在 Yolact 网络框架基础上, 针对复合材料缺陷特点, 从空间域与通道域两个维度, 增强网络对复合材料缺陷的检测能力。在空间域上, 常规卷积核具有空间尺度的局限性, 对狭长形、大尺寸缺陷的检测效果不佳。因此, 采用 CNN+Transformer 架构的 BoTNet 作为基础主干网络; 同时将 Transformer 引入 Yolact 网络的 FPN 结构中, 增强网络从非局部空间中获取信息的能力。在通道域上, 采用红外与可见光联立的检测方式, 并改进主干网络浅层结构, 将其分为可见光通道、红外通道、混合通道, 混合通道中引入通道域注意力机制, 进一步增强网络对红外与可见光图像的综合判断能力。实验结果表明: 改进后 Trans-Yolact 对复合材料缺陷 mAP 为 88.0%, 较基准 Yolact 网络提高 5.5%, 缺丝、扭转等狭长形缺陷 AP 提高 15.2%、5.1%, 包含部分大尺寸缺陷的异物类缺陷 AP 提高 9.1%。最终对 Trans-Yolact 网络进行结构化剪枝, 剪枝后较基准 Yolact 网络减少 26.5% 的计算量 (FLOPs)、减少 44.7% 的参数量; 检测帧数提高 58%, 达到 57.67 fps。并在大型龙门复合材料自动铺放设备上在线测试, 可以满足生产过程最大铺放速度 1.2 m/s 下, 复合材料缺陷的实时检测分割。

关键词: 缺陷检测; 多光谱融合; 纤维自动铺放; 深度学习; 实例分割

中图分类号: V261.97; TB332; TP183 文献标志码: A DOI: 10.3788/IRLA20220338

Transformer-based multi-source images instance segmentation network for composite materials

Ke Yan¹, Fu Yun², Zhou Weizhu¹, Zhu Weidong^{1*}

(1. School of Mechanical Engineering, Zhejiang University, Hangzhou 310027, China;
2. Xizi Spirit Aerospace Industry (Zhejiang) Ltd, Hangzhou 310018, China)

Abstract: In order to improve the quality of automatic fiber placement and assist on-site personnel to quickly detect defects, this paper proposes a real-time instance segmentation network named Trans-Yolact, which is based on Transformer. The Trans-Yolact is used to detect, classify and segment multi-spectrum images of composite material defects. Based on Yolact, aiming at the characteristics of composite material defects, Trans-Yolact's detection ability of composite material defects is enhanced from the two dimensions of space domain and channel domain. In the spatial domain, the convolution kernels have the limitation of spatial scale. The detection of narrow, long, large-size defects is not effective. Therefore, this paper adopts the BoTNet of the CNN+Transformer architecture as backbone; at the same time, the Transformer is introduced into the FPN structure of

收稿日期: 2022-06-20; 修订日期: 2022-10-20

基金项目: 浙江省尖兵“领雁”研发攻关计划 (2022C01134)

作者简介: 柯岩, 男, 硕士生, 主要从事机器视觉、深度学习等方面的研究。

导师(通讯作者)简介: 朱伟东, 男, 教授, 博士, 主要从事飞机装配、复合材料制造、机器视觉等方面的研究。

the Yolact network to enhance the network's ability to obtain information from non-local spaces. In the channel domain, the infrared and visible simultaneous detection method is adopted, and the shallow structure of the backbone is improved, which is divided into visible channel, infrared channel, and mixed channel. Channel domain attention mechanism is introduced in mixed channel. Enhance the comprehensive judgment ability of the network for infrared and visible images. The results show that the mAP of Trans-Yolact for composite defect detection is 88.0%, which is 5.5% higher than Yolact network, and the AP of narrow defects such as miss and twist are increased by 15.2% and 5.1%. The AP of foreign defects including some large-scale defects is increased by 9.1%. Finally, the Trans-Yolact network is pruned. After pruning, the amount of floating-point operations per second (FLOPs) and parameters are reduced by 26.5% and 44.7% compared with Yolact network. The number of detection frames is increased by 58%, reaching 57.67 fps. And the online test is carried out on the large-scale gantry composite material automatic laying equipment, which can meet the real-time detection and segmentation of composite material defects under the maximum laying speed of 1.2 m/s in the production process.

Key words: defects detection; multi-spectrum fusion; automatic fiber placement; deep learning; instance segmentation

0 引言

碳纤维复合材料因质量轻、强度高、耐高温,被广泛应用于航空航天领域^[1]。随着使用量的增加,越来越多的复合材料构件采用自动纤维铺放技术生产(Automated Fiber Placement, AFP)。受铺放路径、压辊压力、预热温度等影响,复合材料表面会出现缺丝、间隙、重叠、褶皱、扭转、架桥、气泡、异物等缺陷^[2]。起初采用人工方式对缺陷进行检测,但检测耗时过长(约占生产时间 32%^[3])。为缩短检测时间,各铺放团队将检测系统前置,开发出复合材料在线检测系统。

按检测传感器种类划分,可分为可见光检测与热红外检测。基于可见光图像的检测系统,如文立伟等^[4]将可见光相机与 UMAC 结合,可实现间隙、重叠类缺陷的检测与铺放头的闭环控制。魏天舒等^[5]使用图像分割算法提取缺陷边缘,并对缺陷进行分类。由于复合材料可见光图像的对比度过低,因此仅对与背景亮度差异较大的缺陷,检测效果较好。

基于热红外图像的检测系统是利用复合材料铺放过程中需要热激励源对其加热,不同种类的缺陷由于热传导方式、热传导率的不同,在红外图像中与正常铺放表面差异明显。最早由波音公司将热红外相机集成进 AFP 系统中^[6]。Denkena 等^[7-8]根据红外图像中温度的极值点,设置阈值分割缺陷所在区域,可以检测出扭转、间隙、重叠、架桥、异物类缺陷。Chen 等^[9]设计了一种可用于智能决策、参数优化与质量追溯的基于红外视觉的复合材料缺陷检测系统。王璇等^[10]设计了一种基于红外图像的复合材料

表面检测网络 AFP-CenterNet, 可以实现在无 GPU 加速的情况下以 4.2 fps 的检测速度对间隙、缺丝、扭转、气泡、起皱、异物进行检测。Gregory 等^[11]采用信息重建与视频分析,对缺陷定位与评估。Juarez 等^[12]采用机器学习的方法,利用热红外相机对重叠、间隙、褶皱、脱粘、扭转类缺陷进行检测,并发现借助红外热成像可以对铺放层间粘性减弱缺陷进行检测。最近, NASA^[13]提出了一种基于红外热成像的复合材料缺陷检测系统 ISTIS, 借助红外热图像,实时检测复合材料铺放过程中产生的重叠与间隙缺陷。并将该系统部署在 NASA Langley 的 AFP 系统中进行在线测试,可以检测出实际尺寸在 0.762 mm 内的缺陷。上述研究证明,热传导方式、热传导率突变的缺陷在基于红外图像的检测系统中检测效果更好。

在上述研究的基础上,康硕等^[14]首先建立一种红外与可见光联立的复合材料检测系统。分别将红外图像与可见光图像输入两个 CSP-DarkNet 网络中,对得到的特征图采用改进的特征金字塔网络结构进行多尺度预测。结果较单光谱检测 mAP 提高 6.3%,证实采用红外与可见光联立的方式可以进一步提升复合材料的检测效果。

从检测结果来看,复合材料检测系统由单一的目标检测向可获取缺陷轮廓的实例分割网络发展。传统的目标检测网络仅能获取缺陷的尺寸位置信息,无法获取缺陷轮廓、面积等细化信息。为了更好辅助现场人员识别缺陷、保证产品的质量溯源, Sacco 等^[15-16]提出了一种基于深度学习的缺陷分割算法。在线获

取可见光图像,对缺丝、间隙、重叠、扭转、褶皱类缺陷进行检测,并对图像进行语义分割,记录缺陷轮廓。将网络嵌入 AFP 铺层表面缺陷检测系统 ACSIS 中,检测准确率可达 75% 以上。

由于复合材料缺丝、扭转类缺陷长宽比在 12~4 之间,属于狭长形缺陷。而常规卷积核具有空间尺度的局限性,很难从图像全局的角度进行综合判断,因此对上述缺陷的检测是复合材料检测的难点。为了解决卷积核只能聚合局部空间信息的劣势,目前计算机领域主流方法是将 Transformer 变换引入 CNN 网络中^[17],增强网络对非局部空间信息的整合。在目标检测 (DETR^[18]、Deformable-DETR^[19])、图像分割 (SETR^[20]) 任务中取得亮眼的成效。将引入 Transformer 的网络进一步细分,可以分为 Pure-Transformer 架构 (如 ViT^[21]),和 CNN+Transformer 架构 (如 BoTNet^[22])。Pure-Transformer 架构需要在大型训练集中进行预训练,再运用在特定数据集中微调模型,否则难以收敛;同时缺乏对于偏置的归纳能力。而 CNN+Transformer 架构是将 Transformer 核心机制 (自注意力机制) 变形为模块,插入 CNN 架构中,其兼具 CNN 平移不变性的优点以及 Transformer 获取全局信息的能力。

文中设计了一种基于 Transformer 的复合材料多源图像实时实例分割网络 Trans-Yolact(Transformer Yolact)。选择检测效果最优的红外与可见光联立检测方式,在检测网络中并联图像分割模块,实时对复合材料缺陷进行检测、分类、分割。针对复合材料缺陷特点,空间域上引入 Transformer 变换,通道域上改进主干网络信息提取方式,进一步优化网络。第 1 节介绍了复合材料图像数据集的拍摄方式,红外与可见光图像的矫正、配准方式,并分析复合材料缺陷特点。第 2 节根据缺陷特点与基准 Yolact 网络特点,有针对性的提出改进方法,并详细阐述 Transformer 增强网络获取图像全局信息的数学原理。第 3 节从网络收敛的稳定性、检测的准确性对网络进行验证。并在此基础上,对网络进行剪枝优化,部署在复合材料实时检测系统中,进行在线验证。

1 图像采集与分析

1.1 图像采集

文中主要讨论复合材料常见的缺丝、扭转、异

物、褶皱、架桥、气泡 6 类缺陷。在生产过程中采集缺陷,效率较低,无法满足深度学习所需的数据量。为此我们搭建复合材料图像采集平台。如图 1 所示,采集平台由红外相机、可见光相机、加热灯、可见光源等组成。所用面阵红外相机型号为 Xenics Gobi+ 640,相机分辨率为 640×480,镜头焦距为 10 mm,可以采集 8~14 μm 波段的热红外图像。面阵可见光相机型号为海康 MV-CA013-21 UM,相机分辨率为 1280×1024,镜头为 MVL-HF0828 M-6 MPE,镜头焦距为 8 mm。

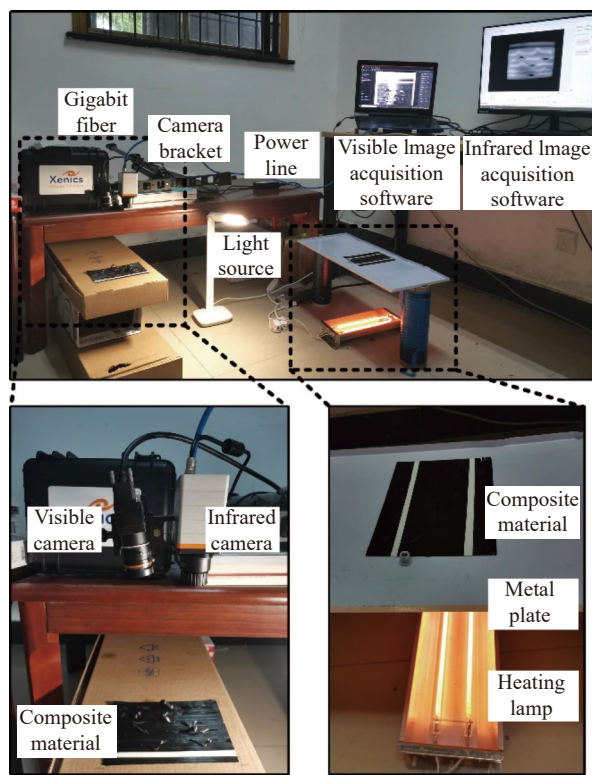


图 1 复合材料缺陷图像采集平台

Fig.1 Acquisition platform of composite material defect image

实际铺放过程中,需要用加热灯对基板均匀加热至 90 °C,而后使用压辊将待铺放的复合材料压在基板表面。在拍摄图像时模拟上述铺放过程,使用加热灯加热金属薄板并保持其温度在 80~110 °C,将复合材料铺放在金属薄板上,模拟热量由基板传导至复合材料中。铺放若干秒后,将复合材料移动到拍摄区域,使用红外和可见光相机采集缺陷图像。

1.2 图像矫正与配准

因拍摄距离较近,红外图像的畸变十分严重。需

要对红外相机进行标定,而后利用相机内参对红外图像畸变进行矫正。文中采用氧化铝标定板,在红外相机标定过程中,黑色块吸收红外热辐射更多,在红外图像中更亮;白色块对光反射能力更强,在红外图像中更暗。但由于标定板经哑光处理,白色块仍能吸收一定的红外热辐射。采用张氏标定法^[23]处理时,存在部分特征点无法准确识别,见图 2(a)。将锡箔贴在白色块上见图 2(b),增加白色块对红外辐射的反射能力,得到标定图像可以准确识别全部特征点见图 2(c)。通过特征点求解红外相机内参数,对红外图像进行畸变矫正,结果见图 2(d)。

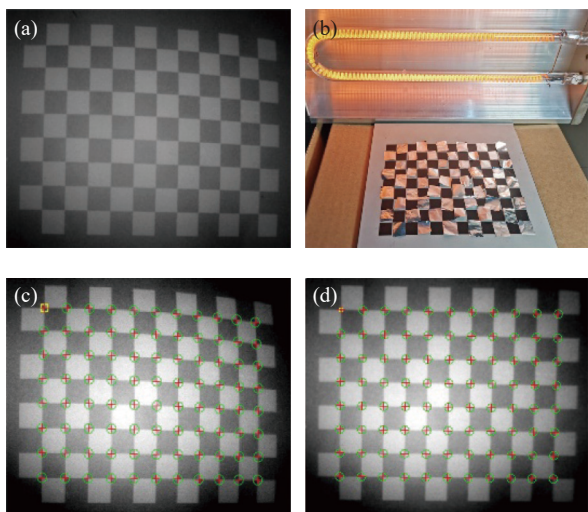


图 2 (a) 原红外图像; (b) 增强标定板反射; (c) 红外图像畸变矫正前; (d) 红外图像畸变矫正后

Fig.2 (a) Infrared image; (b) Enhance the reflection of the calibration plate; (c) Infrared image before distortion correction; (d) Infrared image after distortion correction

矫正红外图像后,由于红外相机与可见光相机光心与光轴不同,导致图像存在差异,见图 3(a),需要对图像进行配准。因图像采集过程中,红外相机与可见光相机的相对位置保持不变,且标定板所在平面与复合材料铺放平面重合,所以借助图像的仿形变换可以实现图像的快速配准。借助标定板的四个角点,红外图像中坐标为 $\{(x_{IR}^1, y_{IR}^1), \dots, (x_{IR}^4, y_{IR}^4)\}$,可见光图像中坐标为 $\{(x_{VI}^1, y_{VI}^1), \dots, (x_{VI}^4, y_{VI}^1)\}$,通过公式 (1) 可以求出仿形变换矩阵 M 。后续所有可见光图像均通过上述仿形变换矩阵配准至红外相机视角,见图 3(b)。

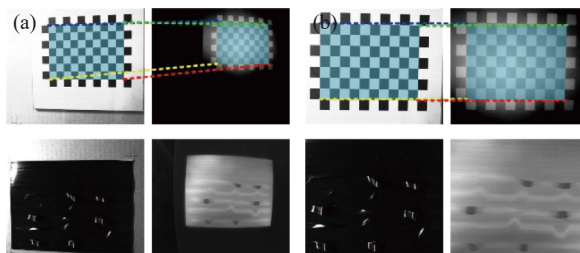


图 3 (a) 仿形变换前; (b) 仿形变换后

Fig.3 (a) Before transformation; (b) After transformation

$$\begin{bmatrix} x_{IR} \\ y_{IR} \\ 1 \end{bmatrix} = M \cdot \begin{bmatrix} x_{VI} \\ y_{VI} \\ 1 \end{bmatrix} = \begin{bmatrix} \gamma^1 & \gamma^2 & d_x \\ \gamma^3 & \gamma^4 & d_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_{VI} \\ y_{VI} \\ 1 \end{bmatrix} \quad (1)$$

式中: (x_{VI}, y_{VI}) 为平面一点在可见光图像中的坐标; (x_{IR}, y_{IR}) 为该点在红外图像中的坐标; d_x 、 d_y 为图像间在 x 、 y 轴上的平移量; γ^n 为表征图像间旋转、放缩、斜切的参数。

1.3 复合材料缺陷分析

按上述方法采集到共计 1634 个不同的复合材料缺陷图像,其中包含扭转 (428)、褶皱 (350)、架桥 (359)、气泡 (273)、缺丝 (110)、异物 (114)。可见光图像配准、红外图像矫正后如图 4(a)、(b),对缺陷采用 coco 数据集格式进行实例分割标注见图 4(c)。

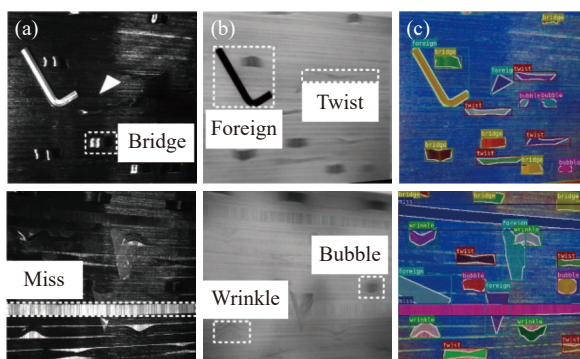


图 4 (a) 可见光图像; (b) 红外图像; (c) 缺陷标签

Fig.4 (a) Visible image; (b) Infrared image; (c) Defect label

观察红外图像和可见光图像,总结复合材料缺陷特点如下: (1) 大多数缺陷在红外或可见光图像中均可识别,但少数缺陷则只能在特定图像中进行识别。如扭转和部分缺丝仅能在红外图像中识别;某些异物由于导热良好,其表面温度与铺放表面温度一致,仅能在可见光图像中识别。(2) 综合红外与可见光图像信息,可以更有效的对某些种缺陷进行识别。例如架桥类、气泡,复合材料脱离铺放表面,迎光面上会出现

反光带,在可见光图像中更亮,同时因脱离基板温度更低,在红外图像中更暗。(3) 缺丝与扭转类缺陷,属于狭长形缺陷;部分异物类缺陷,属于尺寸较大形缺陷。上述超出 CNN 卷积核感受野范围,常规的卷积难以获取全部信息。

2 网络设计与改进

2.1 Trans-Yolact 网络总体设计

实例分割网络通常是串联目标检测网络与语义

分割网络,即首先对图像进行目标检测,再对预测框内图像进行分割。典型的如 Mask-RCNN,这种方法检测时间较长 (>100 ms)、检测帧数较低 (<8 fps),无法应用于实时检测领域。Yolact^[24] 网络则首先采用并行设计,在目标检测网络中并行接入掩膜分支如图 5 所示,将实例分割任务分解为两个并行的子分支,目标检测分支与掩膜分支。从而缩短检测时间 (<60 ms)、提升检测帧数 (>30 fps),满足实时检测的要求。

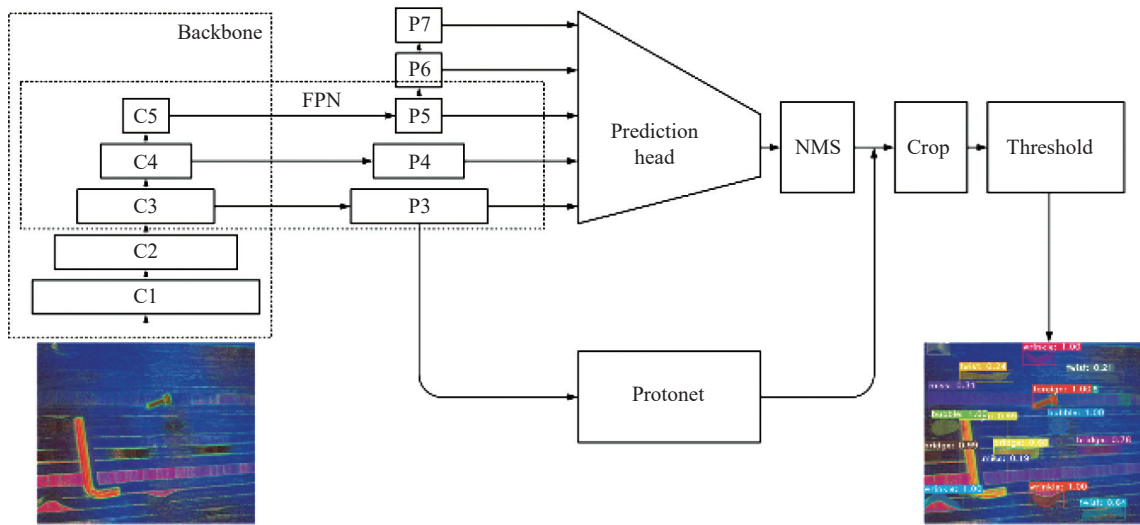


图 5 Yolact 网络结构

Fig.5 Network structure of Yolact

在复合材料在线检测中,铺放设备的最大铺放速度约为 0.8~1.2 m/s。而单张图像所能检测的范围约为 80~100 mm。计算可知网络需要至少 15 fps 的检测帧数,才能保证不漏检缺陷。因此,文中选择目标检测分支与语义分割分支并联的 Yolact 网络为基础框架,保证网络检测帧数满足要求。

文中在 Yolact 网络框架基础上,从以下 3 个方面改进,使之能够更好的检测复合材料缺陷。(1) 针对复合材料缺丝、扭转等狭长形缺陷,常规的卷积核具有空间尺度的局限性,在主干网络深层引入 CNN+Transformer 模块,增加主干网络对非局部信息的整合能力;(2) 针对复合材料架桥、气泡、异物等缺陷,需要综合红外与可见光图像信息进行判断。改进主干网络浅层,将其分为可见光通道、红外通道、混合通道,并在混合通道中引入基于通道域的注意力机制 (SE 模块),增强混合通道对红外与可见光图像综合判

断能力;(3) 针对 Yolact 网络采用图像金字塔类型的 FPN 结构,不同深度特征层间无法进行语义沟通,将其改进为基于 Transformer 的 FPN 结构,增强不同深度特征层、不同空间域间信息的沟通。

Trans-Yolact 网络结构如图 6 所示,由主干网络、FPN 网络、目标检测分支与掩膜分支组成。将红外图像 IR 与可见光图像 VI 变换至 $\mathbb{R}^{1 \times 550 \times 550}$ 输入主干网络图 6(b) 中,主干网络由浅至深提取出一系列特征图像 (C1、C2、C3、C4、C5)。之后将 C3、C4、C5 特征层输入基于 Transformer 的 FPN 结构中,见图 6(c),生成预测图像 (P3、P4、P5、P6、P7)。将提取的全部预测图像输入目标检测分支,见图 6(d),目标检测分支在预测物体种类、候选框偏移量的基础上,增加对掩膜权重值 $W_{Mask} \{\alpha, \beta, \gamma, \dots, \delta\} \in \mathbb{R}^{32}$ 的预测。对得到的候选框进行非极大值抑制筛选 (Non-maximum suppression, NMS),得到目标检测输出。掩膜分支与目标检测

分支并行, 见图 6(e), 将主干网络的浅层特征图像 C3 输入其中。掩膜分支通过全卷积网络 (Fully convolutional network, FCN) 生成 32 张分辨率高的非局部原型掩膜 $\text{Mask} \{ \text{Mask}^1, \text{Mask}^2, \dots, \text{Mask}^{32} \} \in \mathbb{R}^{32 \times 138 \times 138}$, 将原型掩膜与目标检测分支得到的掩膜权重值进行线性组合 (见公式 (2)), 得到图像的语义分割输出

$\text{Seg} \in \mathbb{R}^{1 \times 138 \times 138}$ 。最终将分割图像放缩至源图像大小, 并与目标检测结果融合, 得到对某一个缺陷的实例分割输出。

$$\text{Seg} = \sum_{i=1}^{32} W_{\text{Mask}}^i \cdot \text{Mask}^i \quad (2)$$

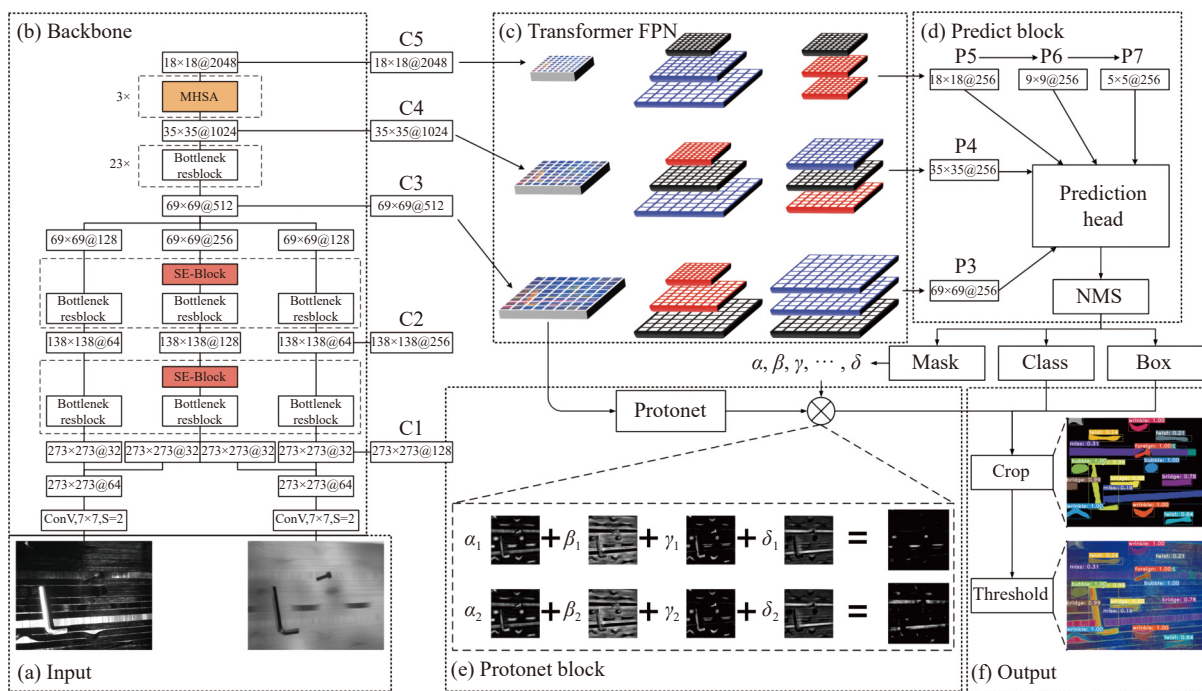


图 6 Trans-Yolact 网络结构

Fig.6 Network structure of Trans-Yolact

2.2 主干网络深层引入 Transformer

Trans-Yolact 主干网络采用基于 CNN+ Transformer 架构的 BoTNet 网络, 增加主干网络从高能特征图像中获取全局语义信息的能力。核心是将主干网络最深层 C5 块组, 替换为多头自注意力机制模块 (Multi-Head Self-Attention, MHSA), 如图 7 所示。

MHSA 模块可视为输入矩阵 $X \in \mathbb{R}^{d \times W \times H}$ 到输出矩阵 $Z \in \mathbb{R}^{d \times W \times H}$ 的映射。其中 d 是矩阵 X 的通道数, W 是图像的宽度, H 是图像的高度。

$$F_{\text{MHSA}} : X \rightarrow Z (Z \in \mathbb{R}^{d \times W \times H}) \quad (3)$$

对输入矩阵 X 进行卷积核为 W^Q, W^K, W^V 的 1×1 的卷积。将卷积结果宽度维度 W 与高度维度 H 上信息整合为一维向量, 得到查询矩阵、键矩阵和值矩阵 Q, K, V 。

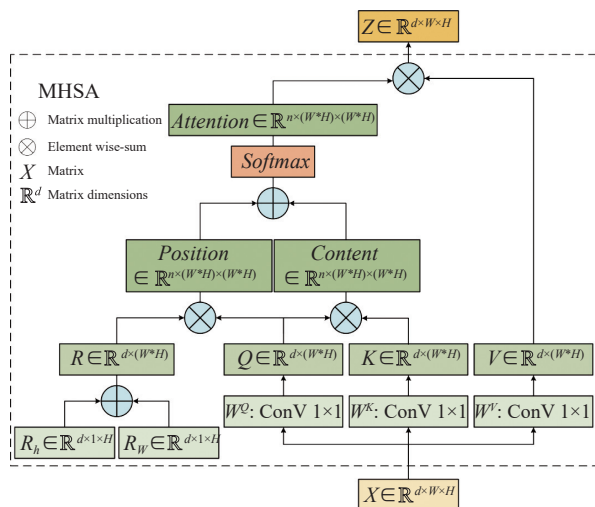


图 7 MHSA 结构

Fig.7 Multi-Head Self-Attention structure

$$\begin{cases} Q = \text{view}(W^Q * X)^{d \times (W \cdot H)} \\ K = \text{view}(W^K * X)^{d \times (W \cdot H)} \\ V = \text{view}(W^V * X)^{d \times (W \cdot H)} \end{cases} \quad \begin{matrix} Q, K, V \in \mathbb{R}^{d \times (W \cdot H)} \\ W^Q, W^K, W^V \in \mathbb{R}^{d \times 1 \times 1} \end{matrix} \quad (4)$$

接下来,对 Q 、 K 、 V 进行编码,分为内容编码(Content)与位置编码(Position)。内容编码建立起键矩阵 K 和查询矩阵 Q 的内容关系矩阵 Content , 维度为 $\mathbb{R}^{n \times (W \cdot H) \times (W \cdot H)}$, 其中 n 代表多头注意力机制的头数。借助内容编码,图像中某点会与图像全局的任意一点产生关联,因此可以增加网络从图像全局获取信息的能力。位置编码引入宽度维度权重参数 $R_w \in \mathbb{R}^{d \times W \times 1}$ 与高度维度权重参数 $R_h \in \mathbb{R}^{d \times 1 \times H}$, 扩展相加后得到空间域键矩阵 $R \in \mathbb{R}^{d \times W \times H}$ 。位置编码建立起空间域键矩阵 R 与查询矩阵 Q 的位置关系矩阵 Position , 可以增强 Transformer 保留矩阵空间域信息的能力。

$$\begin{aligned} \text{Content} &= K^T \times Q \in \mathbb{R}^{n \times (W \cdot H) \times (W \cdot H)} \\ \text{Position} &= R^T \times Q \in \mathbb{R}^{n \times (W \cdot H) \times (W \cdot H)} \\ R &= R_h R_w \in \mathbb{R}^{d \times 1 \times H} \mathbb{R}^{d \times W \times 1} \in \mathbb{R}^{d \times W \times H} \end{aligned} \quad (5)$$

将上文得到的 Content 与 Position 加和, 并进行归一化操作, 得到自注意力矩阵 Att 。将值矩阵 V 与自注意力矩阵 Att 进行矩阵相乘, 可以将值矩阵 V 映射为输出矩阵, 完成 Transformer 的解码操作。

$$\text{Att} = \text{softmax}(\text{Content} + \text{Position}) \in \mathbb{R}^{n \times (W \cdot H) \times (W \cdot H)} \quad (6)$$

$$Z = \text{view}(V \times \text{Att})^{d \times W \times H} \in \mathbb{R}^{d \times W \times H} \quad (7)$$

2.3 改进主干网络浅层

复合材料缺陷有时需要综合红外与可见光图像中的信息进行判断, 因此对主干网络通道域进行改进。在主干网络 C1 模块中分别对红外与可见光图像 $\text{IR}, \text{VI} \in \mathbb{R}^{1 \times 550 \times 550}$ 进行卷积, 得到 $\mathbb{R}^{64 \times 273 \times 273}$ 。从卷积结果中分别提取 $\mathbb{R}^{32 \times 273 \times 273}$ 矩阵拼接在一起形成新矩阵 Mix_{C1} 。采用先卷积、后沿通道方向拼接的方法, 可以利用卷积核的感受野消除部分红外与可见光图像配准误差。C1 模块输出为 $\text{IR}_{\text{C1}}, \text{VI}_{\text{C1}} \in \mathbb{R}^{32 \times 273 \times 273}, \text{Mix}_{\text{C1}} \in \mathbb{R}^{64 \times 273 \times 273}$, 划分为红外通道、可见光通道、融合通道。三种通道分别侧重主干网络从红外、可见光、综合红外与可见光中获取信息。在 C2、C3 模块中, 红外通道与可见光通道采用普通的 ResNet 模块。而融合通道需要综合红外图像和可见光图像中的信息, 因此在 ResNet 模块后采用基于通道域的注意力机制 (SE 模

块^[25]增强整合不同通道间内容的能力, 如图 6(b) 所示。同时考虑 SE 模块串联接入网络中, 会增加网络正向传播时间、降低网络的检测帧数。而主干网络的浅层更接近红外与可见光源图像, 浅层的特征图能更好的反应红外与可见光的差异性。因此融合通道仅在主干网络的浅层 (C2、C3 块组) 添加 SE 模块。最终将 $\text{IR}_{\text{C3}}, \text{VI}_{\text{C3}} \in \mathbb{R}^{128 \times 69 \times 69}, \text{Mix}_{\text{C3}} \in \mathbb{R}^{256 \times 69 \times 69}$ 拼接在一起, 作为 C3 模块输出 $\mathbb{R}^{512 \times 69 \times 69}$ 送入 C4 模块中。

2.4 基于 Transformer 的 FPN 结构

Yolact 网络采用图像金字塔类型的 FPN 结构, 这种结构的预测分支直接对特征图像进行预测, 虽可缩短 FPN 传播时间, 但缺少不同深度特征图像间的语义沟通。将 Transformer 变换引入 FPN 结构中, 如图 8 所示, 不仅可以增强不同深度特征层间的沟通, 同时也可以提高特征图像空间域的信息交流^[26]。具体的, 基于 Transformer 的 FPN 结构可分为基于自身的 Transformer 变换 (Self-Transformer, ST)、深层特征增强浅层特征的 Transformer 变换 (Grounding Transformer, GT)、浅层特征增强深层特征的 Transformer 变换 (Rendering Transformer, RT)。

ST 变换如图 9(a) 所示, 其是对输入矩阵 $X \in \mathbb{R}^{d \times W \times H}$ 自身的 Transformer 变换, 由于其变换方式与 MHSA 模块类似, 在此不多赘述。

GT 变换是高维特征矩阵 $X^c \in \mathbb{R}^{d_c \times W_c \times H_c}$ 与低维特征矩阵 $X^f \in \mathbb{R}^{d_f \times W_f \times H_f}$ 到输出矩阵 $Z^f \in \mathbb{R}^{d_f \times W_f \times H_f}$ 的映射, 如图 9(b) 所示, 其中输出矩阵尺寸与低维特征矩阵尺寸相同。GT 变换是高维特征矩阵指导低维特征矩阵变换为输出矩阵, 其目的是增加低维特征图像中高维语义信息。因此, 对高维特征矩阵 X^c 进行卷积核权重为 W^Q, W^K 的 1×1 卷积, 得到查询矩阵 Q^c 与键矩阵 K^c ; 对低维特征矩阵 X^f 进行卷积核权重为 W^V 的 1×1 卷积, 得到值矩阵 V^f 。

$$F_{\text{GT}} : X^c, X^f \rightarrow Z^f \left(X^c \in \mathbb{R}^{d_c \times W_c \times H_c}; X^f, Z^f \in \mathbb{R}^{d_f \times W_f \times H_f} \right) \quad (8)$$

$$\begin{cases} Q^c = \text{view}(W^Q * X^c)^{d_c \times (W_c \cdot H_c)} \\ K^c = \text{view}(W^K * X^c)^{d_c \times (W_c \cdot H_c)} \\ V^f = \text{view}(W^V * X^f)^{d_f \times (W_f \cdot H_f)} \\ (W^Q, W^K \in \mathbb{R}^{d_c \times 1 \times 1}, W^V \in \mathbb{R}^{d_f \times 1 \times 1}) \end{cases} \quad (9)$$

GT 变换的编码操作是利用查询矩阵 Q^c 与键矩阵 K^c 计算得到注意力矩阵 Att 。由于值矩阵 $V^f \in \mathbb{R}^{d_f \times (W_f \cdot H_f)}$ 与注意力矩阵 $\text{Att} \in \mathbb{R}^{(W_c \cdot H_c) \times (W_c \cdot H_c)}$ 维度不同, 无法直接进

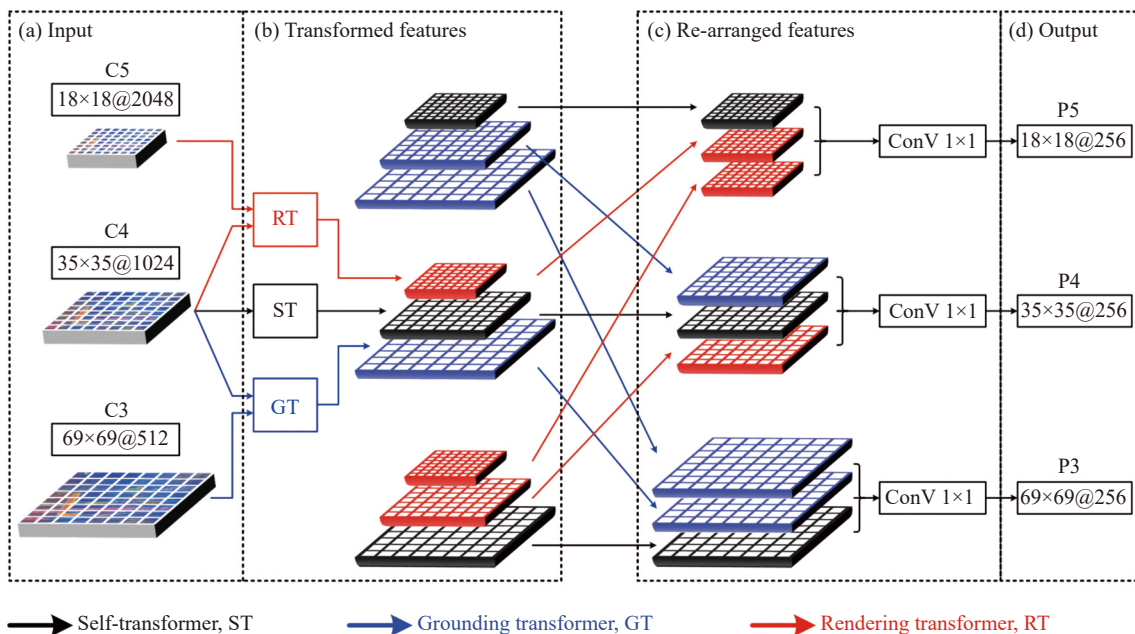


图 8 引入 Transformer 的 FPN 结构

Fig.8 FPN structure based on Transformer

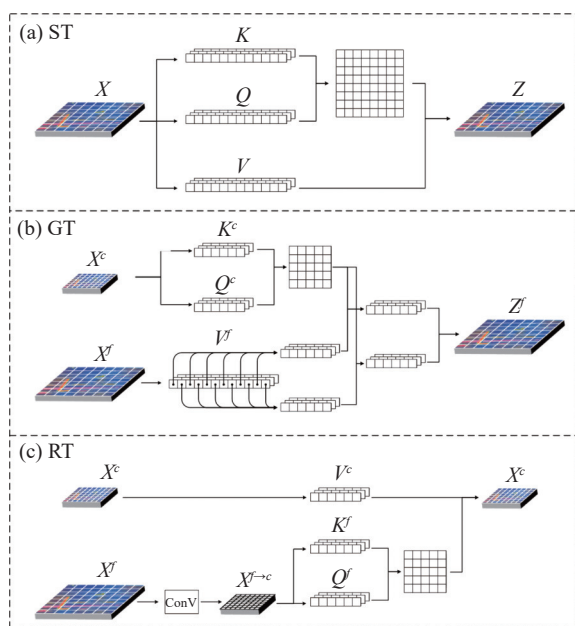


图 9 (a) ST 变换; (b) GT 变换; (c) RT 变换

Fig.9 (a) Self-Transformer; (b) Grounding Transformer; (c) Rendering Transformer

行矩阵乘积。因此在 Transformer 解码操作前, 需要将值矩阵 V^f 拆解为若干子矩阵 $\{V_1^f, V_2^f, \dots, V_n^f \in \mathbb{R}^{d_f \times (W_c \cdot H_c)}\}$, 之后将每个子矩阵分别与 Att 相乘得到子矩阵的解码输出 $\{Z_1^f, Z_2^f, \dots, Z_n^f \in \mathbb{R}^{d_f \times (W_c \cdot H_c)}\}$ 。最终将上述输出拼接为 $\mathbb{R}^{d_f \times W_f \times H_f}$ 尺寸, 完成 Transformer 解码操作, 得到 GT 变换的输出 Z^f 。

RT 变换是低维特征矩阵 $X^f \in \mathbb{R}^{d_f \times W_f \times H_f}$ 与高维特征矩阵 $X^c \in \mathbb{R}^{d_c \times W_c \times H_c}$ 到输出矩阵 $Z^c \in \mathbb{R}^{d_c \times W_c \times H_c}$ 的映射, 如图 9(c) 所示, RT 变换是低维特征矩阵增强高维特征矩阵, 目的是提升小目标的判别能力。因此需要从低维特征矩阵 X^f 中获取查询矩阵 Q^f 与键矩阵 K^f , 利用 Q^f 与 K^f 进行 Transformer 的编码操作。再从高维特征矩阵 X^c 中获取值矩阵 V^c , 值矩阵与注意力矩阵进行矩阵相乘, 完成 Transformer 解码, 得到 RT 变换的输出 Z^c 。

由于低维特征图像尺寸大于高维特征图像, 如果直接使用权重为 $W^Q, W^K \in \mathbb{R}^{d_f \times 1 \times 1}$ 的 1×1 卷积得到的查询矩阵 Q^f 与键矩阵 K^f 会造成编码冗余。因此在卷积前, 增加一层卷积层, 将 X^f 转化为尺寸与 X^c 相同的矩阵。

$$F_{RT} : X^f, X^c \rightarrow Z^c \quad (10)$$

$$(X^f \in \mathbb{R}^{d_f \times W_f \times H_f}; X^c, Z^c \in \mathbb{R}^{d_c \times W_c \times H_c})$$

$$\begin{cases} Q^f = \text{view}(W^Q * \text{ConV}(X^f))^{d_c \times (W_c \cdot H_c)} \\ K^f = \text{view}(W^K * \text{ConV}(X^f))^{d_c \times (W_c \cdot H_c)} \\ V^c = \text{view}(W^V * X^c)^{d_c \times (W_c \cdot H_c)} \\ W^Q, W^K, W^V \in \mathbb{R}^{d_c \times 1 \times 1} \end{cases} \quad (11)$$

3 实验

3.1 实验训练细节

实验环境如下: 操作系统采用 Windows10 操作系

统,深度学习框架采用 Pytorch1.9.1 框架, CUDA11.2, 网络在训练与测试时采用两张型号为 NVIDIA GeForce RTX 2080 的 GPU, CPU 型号为 Intel Core i7-9700 K, 64 G 内存。

按 8 : 2 的比例随机将数据集划分为训练集与验证集, 训练集中包含扭转 (317)、褶皱 (284)、架桥 (263)、气泡 (201)、缺丝 (84)、异物 (84); 验证集中包含扭转 (111)、褶皱 (66)、架桥 (96)、气泡 (72)、缺丝 (26)、异物 (30)。输入网络前, 使用随机翻转、随机改变对比度、随机改变亮度的操作, 防止网络过拟合。模型训练周期为 300 个 epoch, 使用随机梯度下降法, 初始学习率为 1×10^{-3} , 动量设置为 0.9, 学习率衰减因子 5×10^{-4} , 学习率分别在网络训练至 150 轮、200 轮、250 轮衰减至原来的 1/10。

在训练过程中 Trans-Yolact 网络损失随迭代轮数的变化如图 10 所示, 模型在 100 轮内, 总损失值与各部分损失值下降显著; 迭代 200~300 轮, 网络基本稳定, 总损失值变化不超过 ± 0.5 。可以证明采用文中方法引入 Transformer 后, 网络可以在迭代较少轮数后稳定收敛。

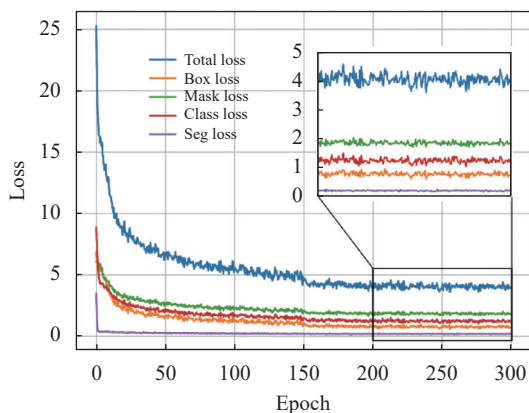


图 10 Trans-Yolact 网络损失

Fig.10 Loss of Trans-Yolact

3.2 实验结果分析

实验结果分析主要从红外与可见光联立检测的有效性、网络结构改进的有效性两方面进行论证。选择交并比为 0.5 时准确度 AP50 作为评价指标, 实验 P-R 曲线如图 11 所示, 实验数据见表 1。在未对网络结构改进的情况下, 分别测试了可见光检测、红外检测、红外与可见光联立检测三种方法。在基准 Yolact

网络上, 仅用可见光图像检测的 mAP 为 75.7%, 仅用红外图像检测的 mAP 为 78.8%, 采用红外与可见光图像联立检测的 mAP 为 82.5%, 较可见光图像检测与红外图像检测提升 6.8%、3.7%。进一步观察 PR 曲线可以发现, 对于扭转类, 采用可见光检测效果较差; 架桥、气泡类, 采用红外与可见光联立检测对准确度的提升较为明显。这与分析缺陷时所得结论相互印证, 证明采用红外与可见光图像联立检测相比单红外或可见光图像检测更为有效。

对网络结构改进的有效性进行论证, 文中所提出的 Trans-Yolact 的 mAP 为 88.0%, 相较红外与可见光图像联立检测的基准 Yolact 网络提升 5.5%。按缺陷类别进行分析, 除最易检测的架桥类缺陷外, 其余种类缺陷在 Trans-Yolact 网络中的检测准确度均高于 Yolact 网络。尤其对于缺丝类、扭转类等狭长形缺陷, 引入 Transformer 变换的 Trans-Yolact 网络查准率较基准网络提升 15.2%、5.1%。类似的, 异物类由于存在部分尺寸较大、红外与可见光图像差异明显的缺陷, Trans-Yolact 网络查准率较基准网络提升 9.1%。证明文中改进网络结构的方式可以有效提升对复合材料缺陷的检测效果。

为了进一步验证引入 Transformer 变换和改进主干网络浅层结构对网络的影响。对文中提出的 Trans-Yolact 网络进行消融实验, 在 Yolact 网络基础上仅增加 Transformer 变换 (对应 2.2 节与 2.4 节) 和仅改进主干网络浅层结构 (对应 2.3 节) 进行实验。仅增加 Transformer 变换的网络, 对缺丝类、扭转类缺陷分别提升 12.1%、3.5%; 对存在部分尺寸较大的异物类缺陷检测准确度提升 4.9%, 从正面证明引入 Transformer 变换可以显著提高网络从非局部空间中获取信息的能力、增强对狭长形、大尺寸缺陷的检测能力。而对于长宽近似相等的、尺寸较小的气泡类、架桥类缺陷, 其检测准确度变化为 -4.3%、-0.7%, 从反面证明 Transformer 变换对于局部小尺寸缺陷检测准确度提升不大, 甚至为负作用。而仅改进主干网络浅层, 对褶皱类、气泡类、异物类缺陷检测的准确度提升 2.7%、2.4%、5.4%, 证明对于需要综合红外与可见光信息进行判断的缺陷, 从通道域对主干网络浅层的改进是有效的。但对于同样需要综合红外与可见光信息判断的架桥类缺陷, 其检测准确度为 93.3%, 较

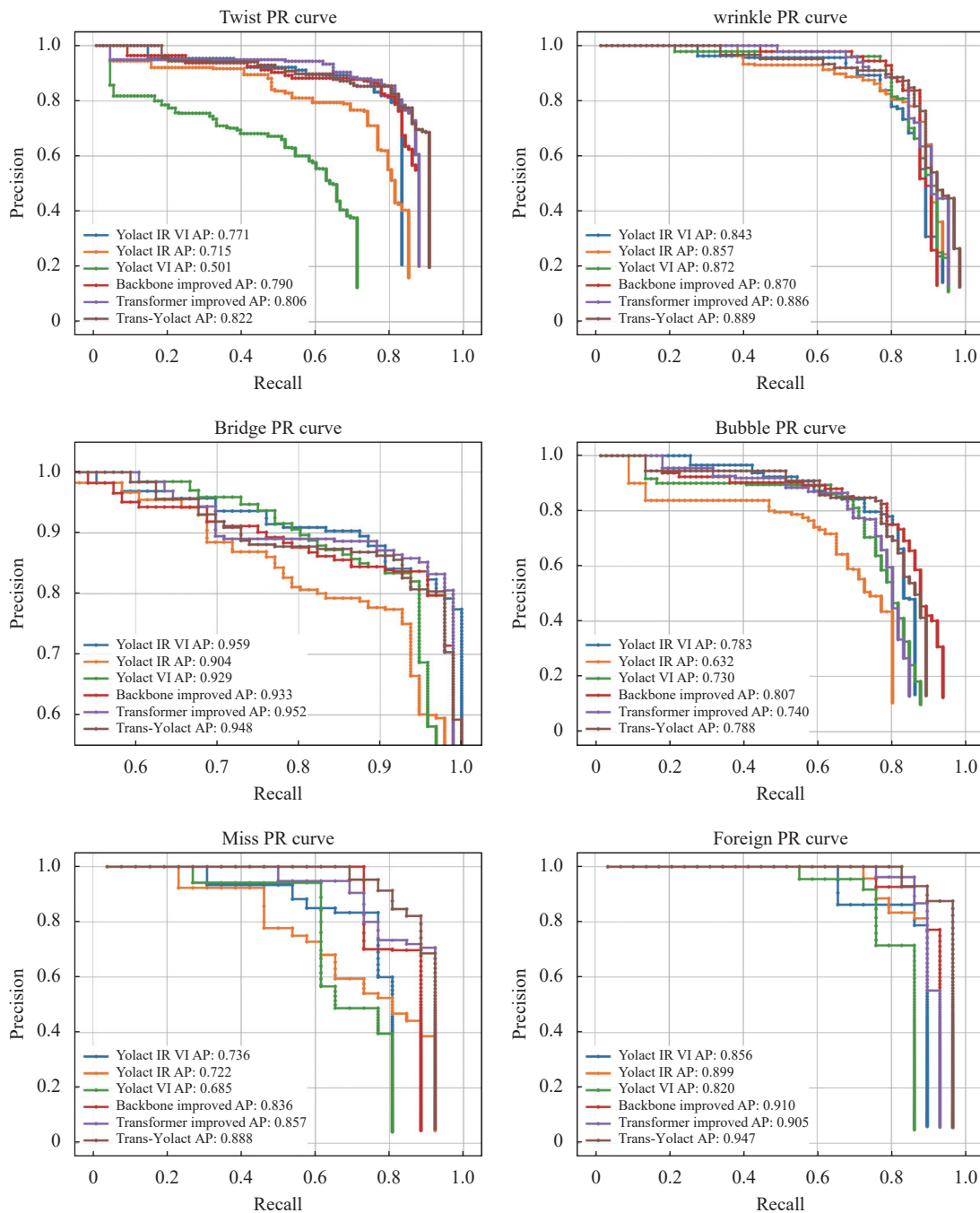


图 11 六种复合材料缺陷的 PR 测试曲线

Fig.11 PR test curves of six composite defects

基准网络变化-2.6%，其可能原因是架桥类缺陷检测较为容易，从红外与可见光图像检测准确度达 95.9%，通道域改进对其影响不大。综上所述，文中所提出的 Trans-Yolact 网络增强从空间域中获取信息的能力，提升对狭长形、大尺寸缺陷的检测准确度；也增强了网络整合不同通道域信息的能力，提升网络从多源图像检测缺陷的准确度。

图 12 是 Yolact 网络与 Trans-Yolact 网络的效果对比，可以看出引入 Transformer 后，第一、四行图像中检测出未检测到的缺丝缺陷、第三行提升缺丝缺陷的检测置信度、第二行图像中检测出未检测到的扭转、气泡类缺陷、第四行提升尺寸相对较大异物缺陷的检测置信度，证明 Trans-Yolact 网络显著增强从非局部空间中获取信息的能力。

表 1 六种复合材料缺陷的 AP
Tab.1 AP of six composite defects

	Twist	Wrinkle	Bridge	Bubble	Miss	Foreign	Total
Yolact IR VI	0.771	0.843	0.959	0.783	0.736	0.856	0.825
Yolact IR	0.715	0.857	0.904	0.632	0.722	0.899	0.788
Yolact VI	0.501	0.872	0.929	0.730	0.685	0.820	0.756
Backbone improved	0.790	0.870	0.933	0.807	0.836	0.910	0.857
Transformer improved	0.806	0.886	0.952	0.740	0.857	0.905	0.858
Trans-Yolact	0.822	0.889	0.948	0.788	0.888	0.947	0.880

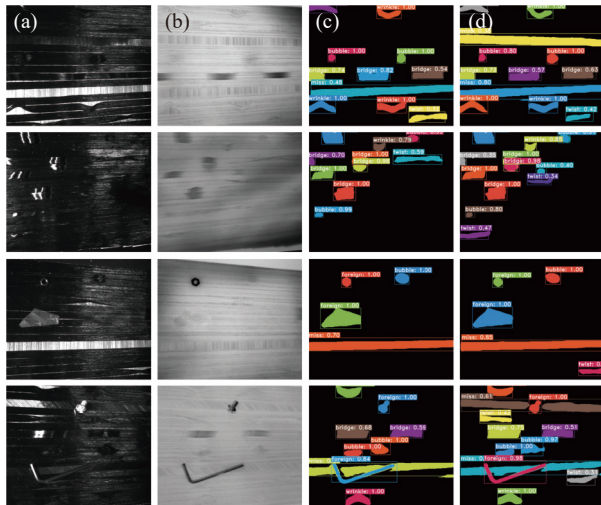


图 12 (a) 可见光图像; (b) 红外图像; (c) Yolact 检测结果; (d) Trans-Yolact 检测结果

Fig.12 (a) Visible image; (b) Infrared image; (c) Yolact detection results; (d) Trans-Yolact detection results

3.3 网络剪枝优化

对 Trans-Yolact 与 Yolact 进行结构化剪枝, 获得一系列网络。并在复合材料缺陷数据集中对上述网络进行评估。网络结构与其对应结果如表 2 所示。Yolact 和 Trans-Yolact 最大检测帧数与网络准确度的关系, 如图 13 所示。图中圆半径代表模型所占内存大小。

在对网络剪枝优化后, Trans-Yolact 系列中检测速度最快的 V 模型, 其检测最大帧数比 Yolact 系列中检测速度最快的 III 模型高出 10 fps, 达到近 60 fps; Trans-Yolact 系列中均衡检测精度与检测速度的 IV 模型, 相比 Yolact 系列中 I 模型, 网络的计算量 (FLOPs) 减少 26.5%、参数量减少 44.7%、内存空间占用减少 44.0%, 同时检测准确度提高 3.55%, 检测帧数提高 58%, 达到 57.67 fps, 远超复合材料在线检测要求。可以证明, 基于 Transformer 的 Trans-Yolact, 其在复合材

表 2 剪枝优化的网络结构与测试数据

Tab.2 Prune-optimized network structure and test data

	Trans-Yolact					Yolact		
	I	II	III	IV	V	I	II	III
C1	1	1	1	1	1	1	1	1
C2(SE)	3(√)	3(√)	3(√)	3(√)	3(√)	3(×)	3(×)	3(×)
C3(SE)	4(√)	4(√)	4(√)	4(√)	4(√)	4(×)	4(×)	4(×)
C4	23	18	12	8	4	23	12	4
C5(MHSA)	3(√)	3(√)	3(√)	3(√)	3(√)	3(×)	3(×)	3(×)
FLOPs	79.79 G	71.95 G	63.74 G	58.27 G	50.11 G	79.28 G	64.22 G	53.27 G
Params	44.69 M	38.62 M	31.92 M	27.45 M	22.59 M	49.62 M	37.34 M	28.40 M
File space	176.6 M	152.8 M	126.5 M	108.9 M	89.8 M	194.5 M	146.3 M	111.2 M
mAP	88.03	87.63	86.69	86.03	83.50	82.48	82.10	80.50
FPS	37.72	45.27	55.68	57.67	59.03	36.53	42.40	48.25

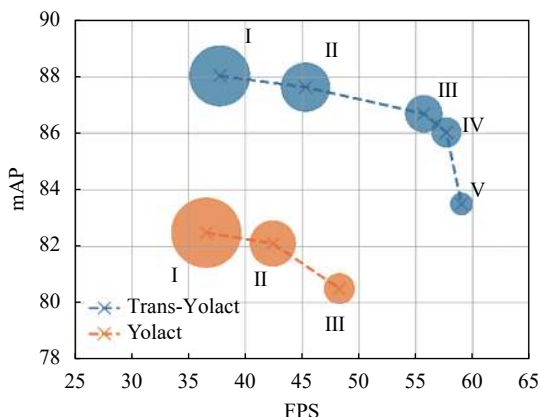


图 13 Yolact 与 Trans-Yolact 在不同网络深度层下的实验

Fig.13 Performance of Yolact and Trans-Yolact under different network depth layers

料缺陷检测领域的表现能力优于基准 Yolact 网络。

3.4 在线部署与验证

最终将剪枝优化后的 Trans-Yolact 系列 IV 网络集成在复合材料实时检测软件中, 在大型龙门复合材料自动铺放设备上上进行在线检测测试, 如图 14 所示。

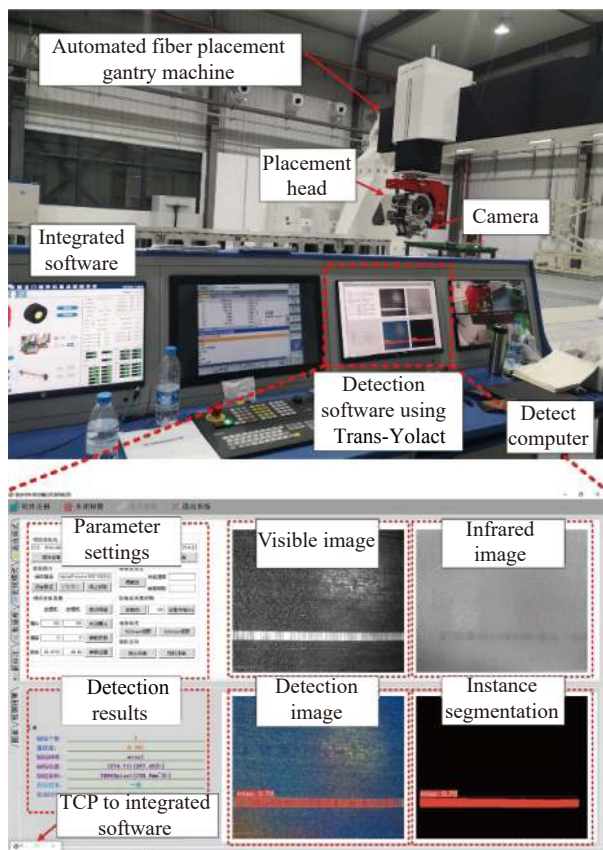


图 14 现场检测实验

Fig.14 Field detecting experiment

在线采集刚铺放的复合材料红外图像与可见光图像, 经网络检测得到缺陷的种类、面积、轮廓掩膜等信息, 利用 TCP 协议传输至上位机集成软件中。在线检测电脑 GPU 为单张 Quadro 4000, 检测帧数最快可达 45 fps, 可以满足最大铺放速度 1.2 m/s 下, 对最小 3 mm×3 mm 缺陷及单股复合材料丝束缺陷的实时检测。

4 结 论

文中提出了一种基于 Transformer 变换的复合材料红外与可见光联立的实例分割网络 Trans-Yolact。以实时实例分割网络 Yolact 为基础框架, 从空间域和通道域两个维度增强网络的信息整合能力。改进后 Trans-Yolact 网络对复合材料缺陷检测准确度可达 88.0%, 较基准提高 5.5%。并对网络进行结构化剪枝, 较基准 Yolact 网络减少 26.5% 的计算量 (FLOPs)、减少 44.7% 的参数量; 检测帧数提高 58%, 达到 57.67 fps。最后在大型龙门复合材料自动铺放设备上上进行在线测试, 满足生产过程中复合材料以最大铺放速度 1.2 m/s 下的实时检测。并得出以下结论:

(1) 空间域上引入 Transformer 可以显著增强网络从非局部空间中获取信息的能力。常规卷积核具有空间尺度的局限性, 文中选择 CNN+Transformer 结构的 BoTNet 作为基础主干网络、并采用基于 Transformer 的 FPN 结构。相较基准网络, Trans-Yolact 网络对狭长形的缺丝类、扭转类检测准确度提升 15.2%、5.1%, 对含有部分大尺寸的异物类缺陷检测准确度提升 9.1%。

(2) 通道域上采用红外与可见光联立的检测方式, 较单红外或单可见光图像 mAP 提高 3.7%、6.8%, 可以增强复合材料检测效果。在此基础上, 改进主干网络浅层结构, 并引入基于通道域的注意力机制 (SE 模块), 可以进一步提升检测准确度 3.2%, 增强混合通道对红外与可见光图像综合判断能力。

对复合材料缺陷进行实例分割, 在得到缺陷预测框的基础上, 可以获取缺陷面积、缺陷轮廓掩膜等信息, 更好地辅助现场人员对复合材料缺陷进行识别, 以及对复合材料构件进行产品溯源。但仍存在缺陷轮廓边缘识别模糊、铺放速度较慢时缺陷重复计数等问题, 需要后续持续优化。

参考文献:

- [1] Brüning J, Denkena B, Dittrich M A, et al. Machine learning approach for optimization of automated fiber placement processes [J]. *Procedia CIRP*, 2017, 66: 74-78.
- [2] Harik R, Saidy C, Williams S J, et al. Automated fiber placement defect identity cards: Cause, anticipation, existence, significance, and progression[R]. 2018.
- [3] Rudberg T, Nielson J, Henscheid M, et al. Improving AFP cell performance [J]. *SAE International Journal of Aerospace*, 2014, 7(2): 317.
- [4] Wen L W, Song Q H, Qin L H, et al. Defect detection and closed-loop control system for automated fiber placement forming components based on machine vision and UMAC [J]. *Acta Aeronautica et Astronautica Sinica*, 2015, 36(12): 3991-4000. (in Chinese)
- [5] Wei T S. Research on image detection method for defects of composite prepreg tapes[D]. Zibo: Shandong University of Technology, 2018. (in Chinese)
- [6] Ritter J A, Sjogren J F. Real-time infrared thermography inspection and control for automated composite material layout: US. Patent 7, 513, 964[P]. 2009-04-07.
- [7] Denkena B, Schmidt C, Völtzer K, et al. Thermographic online monitoring system for automated fiber Placement processes [J]. *Composites Part B: Engineering*, 2016, 97: 239-243.
- [8] Schmidt C, Denkena B, Völtzer K, et al. Thermal image-based monitoring for the automated fiber placement process [J]. *Procedia CIRP*, 2017, 62: 27-32.
- [9] Chen M, Jiang M, Liu X, et al. Intelligent inspection system based on infrared vision for automated fiber placement[C]//2018 IEEE International Conference on Mechatronics and Automation (ICMA). IEEE, 2018: 918-923.
- [10] Wang X, Kang S, Zhu W D. Defect detection of laminated surface in the automated fiber placement process based on improved CenterNet [J]. *Infrared and Laser Engineering*, 2021, 50(10): 20210011. (in Chinese)
- [11] Gregory E D, Juarez P D. In-situ thermography of automated fiber placement parts[C]//AIP Conference Proceedings, 2018, 1949(1): 060005.
- [12] Juarez P D, Gregory E D. In situ thermal inspection of automated fiber placement manufacturing[C]//AIP Conference Proceedings, 2019, 2102(1): 120005.
- [13] Juarez P D, Gregory E D. In situ thermal inspection of automated fiber placement for manufacturing induced defects [J]. *Composites Part B: Engineering*, 2021, 220: 109002.
- [14] Kang S, Ke Z Z, Wang X, et al. Detection method of defects in automatic fiber placement based on fusion of infrared and visible images [J]. *Acta Aeronautica et Astronautica Sinica*, 2022, 43(3): 556-568. (in Chinese)
- [15] Sacco C, Radwan A B, Anderson A, et al. Machine learning in composites manufacturing: A case study of automated fiber placement inspection [J]. *Composite Structures*, 2020, 250: 112514.
- [16] Sacco C. Machine learning methods for rapid inspection of automated fiber placement manufactured composite structures[D]. US: University of South Carolina, 2019: 57-68.
- [17] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C]//31st Annual Conference on Neural Information Processing Systems, 2017: 5999 - 6009.
- [18] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//European Conference on Computer Vision. Springer, 2020: 213-229.
- [19] Zhu X, Su W, Lu L, et al. Deformable detr: Deformable transformers for end-to-end object detection[EB/OL]. (2020-10-08)[2022-06-20]. <https://arxiv.org/abs/2010.04159>.
- [20] Zheng S, Lu J, Zhao H, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 6881-6890.
- [21] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is worth 16x16 words: Transformers for image recognition at scale[C]//International Conference on Learning Representations, 2020.
- [22] Srinivas A, Lin T Y, Parmar N, et al. Bottleneck transformers for visual recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 16519-16529.
- [23] Zhang Z. Flexible camera calibration by viewing a plane from unknown orientations[C]//Proceedings of the Seventh IEEE International Conference on Computer Vision. IEEE, 1999: 666-673.
- [24] Bolya D, Zhou C, Xiao F, et al. Yolact: Real-time instance segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9157-9166.
- [25] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [26] Zhang D, Zhang H, Tang J, et al. Feature pyramid transformer[C]//European Conference on Computer Vision. Springer, 2020: 323-339.