

## 焊缝图像中结构光条纹的检测与分割

张世宽<sup>1,2,3,4</sup>, 吴清潇<sup>1,2,3\*</sup>, 林智远<sup>1,2,3,4</sup><sup>1</sup>中国科学院光电信息处理重点实验室, 辽宁 沈阳 110016;<sup>2</sup>中国科学院沈阳自动化研究所, 辽宁 沈阳 110016;<sup>3</sup>中国科学院机器人与智能制造创新研究院, 辽宁 沈阳 110169;<sup>4</sup>中国科学院大学, 北京 100049

**摘要** 为了在复杂噪声环境下从焊缝图像中精确地提取结构光条纹, 构建了语义分割与目标检测相结合的深度学习模型用于焊缝图像的检测。为了提高模型的检测速度, 在语义分割分支中, 通过添加并行下采样模块及缩减卷积核数量的策略对模型进行了优化, 并使该分支与目标检测分支的特征提取部分共享权重。针对焊缝图像中结构光条纹与背景像素比例失衡而导致模型分割结果偏向负样本的问题, 在损失函数中添加 Dice 系数来对模型进行修正。经实验验证, 该方法在保证实时性的基础上, 以较高的精度实现了结构光条纹的检测。

**关键词** 机器视觉; 结构光条纹; 语义分割; 目标检测; Dice 系数

中图分类号 TG409; TP391

文献标志码 A

doi: 10.3788/AOS202141.0515002

## Detection and Segmentation of Structured Light Stripe in Weld Image

Zhang Shikuan<sup>1,2,3,4</sup>, Wu Qingxiao<sup>1,2,3\*</sup>, Lin Zhiyuan<sup>1,2,3,4</sup><sup>1</sup>Key Laboratory of Opto-Electronic Information Processing, Chinese Academy of Sciences, Shenyang, Liaoning 110016, China;<sup>2</sup>Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, Liaoning 110016, China;<sup>3</sup>Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang, Liaoning 110169, China;<sup>4</sup>University of Chinese Academy of Sciences, Beijing 100049, China

**Abstract** In order to accurately extract structured light stripes from weld images in the complex noise environment, we proposed a deep learning model combining semantic segmentation with object detection to detect the weld images. In the semantic segmentation branch, the model was optimized by adding parallel downsampling modules and reducing the number of convolution kernels to increase the detection speed, and the feature extraction parts of this branch and the object detection branch shared the weights. Aiming at the problem that the proportion unbalance of structured light stripes and background pixels in the weld images caused the model segmentation results to be biased towards negative samples, we introduced a Dice coefficient into the loss function to correct the model. The experimental results show that the proposed method can achieve the extraction of structured light stripes with high accuracy on the basis of ensuring real-time performance.

**Key words** machine vision; structured light stripe; semantic segmentation; object detection; Dice coefficient

**OCIS codes** 150.0155; 100.4996; 100.3008

收稿日期: 2020-09-14; 修回日期: 2020-10-08; 录用日期: 2020-11-02

基金项目: 国家自然科学基金 (U1713216)

\* E-mail: wuqingxiao@sia.cn

## 1 引言

在面向焊缝跟踪的自动焊接机器人系统中,智能测量传感器是其核心技术,主要通过测距式传感器、电磁式传感器以及视觉传感器<sup>[1]</sup>等方式对焊缝进行测量。近年来,机器视觉等行业的快速发展使得基于主动视觉的激光视觉传感器<sup>[2]</sup>在机器人自动焊接系统中越来越重要。但是视觉传感器采集到的图像中往往包含大量的烟气、飞溅线以及弧光等噪声<sup>[3]</sup>,这些噪声会对结构光条纹检测产生很大的影响。

目前,很多方法已用于焊缝跟踪过程中采集获得的结构光条纹的特征识别。根据激光条纹的特点,可以通过 Randon 变换及模糊增强的方法获得较少噪声的激光条纹图像<sup>[4]</sup>;或者对自适应阈值分割后的图像利用形态学操作滤除孤点来获得较平整的激光条纹<sup>[5]</sup>;也可以通过视觉注意的方法获取焊接区域的清晰图像<sup>[6]</sup>。

在应对噪声变化多样的焊接过程时,传统方法的识别效果会有所下降,进而影响后续的焊缝特征提取过程。因此,近年来研究学者们逐渐提出了一些基于神经网络的算法来检测焊缝图像。在提取激光条纹中心线时,可以利用 BP(back propagation)神经网络来对均值滤波后的图像进行检测<sup>[7]</sup>;也可以通过全卷积神经网络对图像进行像素级预测,进而识别结构光条纹<sup>[8]</sup>;在焊缝定位时,可以利用深层卷积神经网络及相关滤波器所提取的特征来表征结构光条纹,之后经过非极大值抑制策略实现焊缝定位<sup>[9]</sup>。

本文面向复杂噪声环境下焊缝图像中结构光条纹提取困难的问题,提出了利用语义分割与目标检测两个分支相结合的深度学习模型用于焊缝图像的检测。语义分割分支基于 U-Net<sup>[10]</sup>的思想在复杂噪声环境中提取结构光条纹,由于焊缝图像分辨率较大(1280 pixel×1024 pixel),在语义分割分支中产生的计算量较大,因此在该分支中通过引入并行下采样模块及缩减卷积核数量的策略对 U-Net 网络进行了改进,从而降低了模型计算量。另外,针对焊缝图像的特点,在该分支的损失函数中引入了 Dice 指标来解决焊缝图像中的样本失衡问题,以提升分割精度。目标检测分支基于 YOLO<sup>[11]</sup>系列的思想采用框回归策略来提取焊缝图像中结构光条纹所在的位置,融合两个分支时采用权重共享策略来共享特征提取部分的卷积层权重,以进一步节省推

理时间。基于上述方法,构建了 Tiny DT-Net(Tiny Dual Task Network)网络,通过该网络可以从焊缝图像中精确快速地分割出结构光条纹并提取出其位置,从而为后续的结构光条纹中心线以及焊缝特征点提取工作提供参考。

## 2 时间复杂度分析

深度学习模型涉及大量的浮点数运算,运算量直接决定了模型的前向推理时间。为了达到实时性,必须降低计算量,减少浮点数运算(FLOPs)<sup>[12]</sup>。卷积是深度学习模型中最主要的操作,本研究只对卷积操作的时间复杂度进行分析。

一次卷积的计算复杂度与输入输出通道数、输出特征图的高和宽以及卷积核大小相关,即

$$T_{\text{Time}} \sim O(C_{\text{in}} \cdot C_{\text{out}} \cdot H \cdot W \cdot K^2), \quad (1)$$

式中: $T_{\text{Time}}$ 为计算消耗时间; $O(\cdot)$ 为时间复杂度; $C_{\text{in}}$ 、 $C_{\text{out}}$ 为输入、输出通道数; $H$ 、 $W$ 为输出特征图的高和宽; $K$ 为卷积核的大小。在卷积操作中,卷积核大小通常为 $1 \times 1$ , $3 \times 3$ , $5 \times 5$ ,不宜过多裁剪,因此,本文选择从另外两个角度降低模型复杂度:1)降低特征图分辨率大小;2)减少输入输出通道数。

## 3 网络架构

模型框架主要包括两个分支,语义分割分支(Branch-seg)和目标检测分支(Branch-dec),如图 1 所示。

### 3.1 语义分割分支

语义分割分支的作用在于从包含大量烟气、飞溅线以及弧光等噪声的焊缝图像中精确地提取出结构光条纹。如图 1 中的 Branch-seg 所示,该分支采用的是基于 U-Net 的分割架构,其中特征提取模块包含 4 次下采样,同时在上采样过程中,利用跳跃连接对下采样与上采样阶段的特征图进行融合<sup>[10]</sup>,将高层语义信息与低层细节信息融合起来,进而获得更加丰富的图像信息。

语义分割模型不仅需要图像进行编码得到深层特征图,还需要对特征图进行解码来获取原图像信息<sup>[13]</sup>。因此,语义分割模型需要的计算量较大,在处理分辨率较高的焊缝图像时难以满足实时性要求。根据第 2 节的描述,降低图像分辨率及减少输入输出通道数可以降低模型的计算量。因此,为了加快模型推理速度,在该分支中可依据上述两点对 U-Net 进行改进。

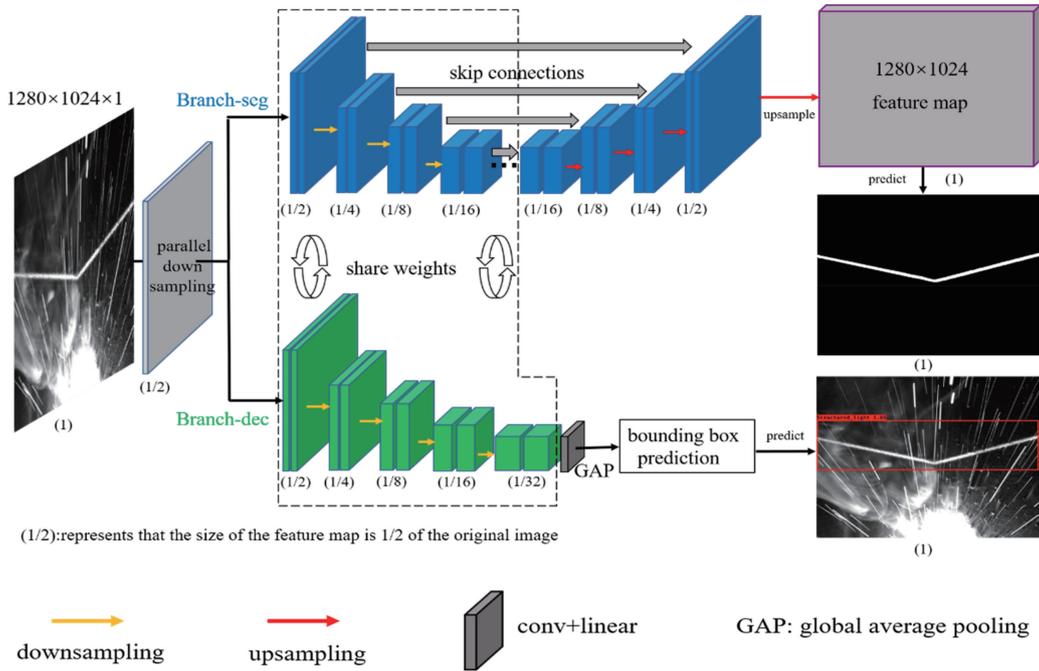


图 1 网络架构

Fig. 1 Network architecture

首先,在图像输入阶段引入并行下采样模块<sup>[14]</sup>来降低图像分辨率。焊缝图像分辨率大小为 1280 pixel×1024 pixel,针对焊缝图像而言,除了结构光条纹之外,原始图像中包含较多的冗余信息,因此可以采用下采样模块来降低图像分辨率,去除部分冗余信息,以获得更有效的信息表示,设计时采用池化下采样与步长为 2 的卷积并行方式来缩减图像大小,同时可以提取原图像的特征,如图 2 所示。其次,为了减少输入输出通道数,将 U-Net 架构编码器中卷积核的数量缩减为原来的 1/4,得到卷积核数量为 (16, 16), (32, 32), (64, 64), (128, 128), (256, 256) 的特征编码器。在译码器部分同样采用轻量型网络架构来恢复焊缝图像中结构光条纹的信息,并采用跳跃连接来丰富译码器部分的特征信息。在网络初期由于通过并行下采样模块提前进行了一次下采样,因此在分割网络的最后还需进行一次上采样,通过双线性插值上采样再经过一次卷积得到

最后的特征图,并对其进行分割预测。

经过上述两点改进,在语义分割分支得到了一个轻量型的 U-Net 网络架构。在表 1 中,显示了改进后的 U-Net 架构与原 U-Net 网络的参数量以及计算量大小。可以看到,改进后的 U-Net 网络与原始 U-Net 相比,参数量缩小了近 88.6%,计算量缩小了近 97.6%,大大提高了语义分割分支的预测速度。

表 1 网络模型对比  
Table 1 Model comparison

Model	Parameter	FLOPs
U-Net	$1.32 \times 10^7$	$9.68 \times 10^{11}$
Modified U-Net	$1.50 \times 10^6$	$2.36 \times 10^{10}$

### 3.2 目标检测分支

在进行语义分割的同时,在模型中添加了目标检测分支来提取结构光条纹的位置。如图 3 所示,在所检测的焊缝图像中大致包含 3 种位置的结构光条纹。

在该分支的检测过程中,将焊缝图像中结构光条纹视为检测目标,在一张图片中只需检测出一个目标即可。因此,基于 YOLO-v2<sup>[15]</sup> 中划分单元格及 anchor 的思想,可以将最后提取的结构光条纹特征图视为一个 1×1 的单元格,如图 4 所示。在该单元格中利用 Sigmoid 函数来预测 bounding-box 中心点相对于单元格左上角的偏移值,则有

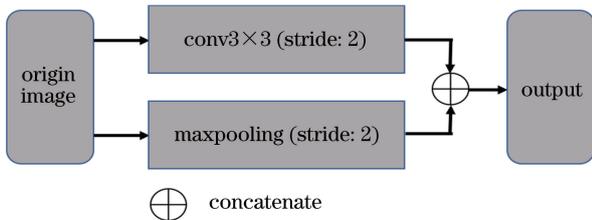


图 2 并行下采样模块

Fig. 2 Parallel downsampling module



图 3 焊缝结构光条纹的图像。(a)图像 1;(b)图像 2;(c)图像 3

Fig. 3 Weld structured light stripe images. (a) Image 1; (b) image 2; (c) image 3

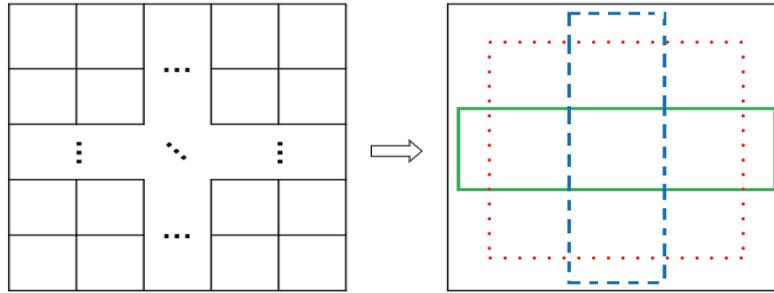


图 4 检测分支预测过程示意图

Fig. 4 Schematic diagram of prediction process of detection branch

$$b_x = \sigma(t_x) + c_x, \quad (2)$$

$$b_y = \sigma(t_y) + c_y, \quad (3)$$

式中: $t_x, t_y$  为网络的预测结果; $x, y$  为坐标方向; $b_x, b_y$  为边界框中心点的坐标; $\sigma(t_x), \sigma(t_y)$  为中心点的偏移值; $c_x, c_y$  为单元格左上角的坐标值。

同时根据图 3 所示的不同位置的结构光条纹,在该单元格中生成 3 个 anchor,anchor 的大小经过归一化后分别为(0.99219,0.29297),(0.21875,0.99609),(0.70313,0.87891),通过生成的 anchor 来修正包含结构光条纹的 bounding-box 的高宽,则有

$$b_h = p_h \exp t_h, \quad (4)$$

$$b_w = p_w \exp t_w, \quad (5)$$

式中: $b_h, b_w$  为最终边界框的高、宽; $p_h, p_w$  为 anchor 的高、宽; $t_h, t_w$  为网络预测的结果。

最后,在检测过程中还应预测每个 bounding-box 的置信度,选取置信度最高的 bounding-box 来作为结构光条纹检测的最终结果。

在设计网络时,检测分支依次经过特征提取部分、一个线性激活的卷积层、一个 GAP<sup>[16]</sup>层(全局平均池化)后产生了最后的预测结果。GAP 层的作用就是将最后生成的特征图划分成一个  $1 \times 1$  的单元格,并在该单元格上预测 3 个 bounding-box 的坐标偏移值、高宽以及置信度。因为只有结构光条纹一种目标,所以不用预测类别。目标检测分支最后的预测结果是一个形状为  $(N, 1, 1, 3 \times 5)$  的张量,

其中, $N$  为图片数量,3 是 anchor 的个数,5 包含 bounding-box 的中心点坐标偏移值( $t_x, t_y$ )、高宽( $t_h, t_w$ )以及置信度。

如图 1 中的 Branch-dec 分支及虚线部分所示,该分支与 Branch-seg 分支在特征提取部分共享权重,不会产生额外的特征提取过程,节省了推理时间,所以该分支的检测速度也比较快。

通过上述 Branch-seg 分支的分割与 Branch-dec 分支的检测,可以快速精确地从焊缝图像中分割结构光条纹并提取出其位置。

## 4 基于焊缝图像的损失函数

### 4.1 语义分割分支的损失函数

语义分割模型的本质仍然是一个分类任务,在对焊缝图像中的结构光条纹进行检测时,只涉及到条纹及背景两类,因此应选择二值交叉熵损失函数,函数表达式为

$$L_{\text{bce}}(Y^{\text{pred}}, Y^{\text{gt}}) = -\frac{1}{N} \sum_{b=1}^N [Y_b^{\text{gt}} \log(Y_b^{\text{pred}}) + (1 - Y_b^{\text{gt}}) \log(1 - Y_b^{\text{pred}})], \quad (6)$$

式中: $L_{\text{bce}}(\cdot)$  为二值交叉熵损失函数; $N$  为每次迭代过程中样本总数; $Y_b^{\text{gt}}$  为第  $b$  个样本的真实值; $Y_b^{\text{pred}}$  为其预测值。

图 5 是焊缝图像及其分割标注图,从图中可以

发现,结构光条纹在图像中所占比例过小,正样本数量远远小于负样本数量,从而导致类别失衡。因此,在训练过程中,图像中大量负类样本( $Y_b^{\text{gt}}=0$ )的损

失值会对损失函数产生主要影响,使得网络在梯度更新过程中的权重变得更倾向于预测焊缝图像中的背景点,从而影响条纹线的预测精度。

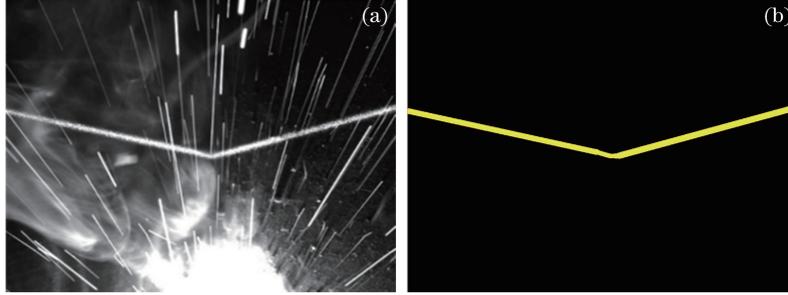


图 5 焊缝图像。(a)原始图像;(b)分割标注图

Fig. 5 Weld images. (a) Original image; (b) annotated image of segmentation

为了解决焊缝图像正负样本类别失衡而导致模型预测结果偏移的问题,引入 Dice 指标( $D$ )<sup>[17]</sup>来修正损失函数,在训练过程中使用其相反数  $D_{\text{loss}}$  来修正,即

$$D = -D_{\text{loss}} = \frac{2N_{A \cap B}}{N_A + N_B}, \quad (7)$$

式中: $N_A$  和  $N_B$  分别为真实值  $A$ 、预测值  $B$  中正类样本的个数; $N_{A \cap B}$  为  $A$  与  $B$  交叉部分正类样本的个数,交叉部分的正样本个数越多,Dice 指标的值越大, $D_{\text{loss}}$  越小,且预测值与标签值的相似度越高,所以该指标主要关注的是焊缝图像中的正类样本在重叠区域的贡献,与占焊缝图像较大部分的背景像素无关。因此在损失函数中添加此指标后,网络在训练过程中可以较好地关注正类样本,从而解决焊缝图像中的类别失衡问题。在语义分割分支中最终采用的损失函数为

$$L_{\text{bce\_Dice}}(Y^{\text{pred}}, Y^{\text{gt}}) = \frac{1}{2} \cdot L_{\text{bce}}(Y^{\text{pred}}, Y^{\text{gt}}) - \frac{1}{N} \sum_{b=1}^N \frac{2Y_b^{\text{pred}} \cdot Y_b^{\text{gt}}}{Y_b^{\text{pred}} + Y_b^{\text{gt}}}. \quad (8)$$

#### 4.2 目标检测分支的损失函数

在检测分支中所采用的损失函数  $L_{\text{dec}}$  由边界框定位损失及置信度损失组成,表达式为

$$L_{\text{dec}} = \lambda_{\text{coord}} \sum_{i=0}^{s^2} \sum_{j=0}^B 1_{ij}^{\text{obj}} \sum_{r \in (x, y, w, h)} (t_{\text{truth}}^{(r)} - b_{\text{bbox}, i}^{(r)})^2 + \lambda_{\text{obj}} \sum_{i=0}^{s^2} \sum_{j=0}^B 1_{ij}^{\text{obj}} (c_i - \hat{c}_i)^2 + \lambda_{\text{noobj}} \sum_{i=0}^{s^2} \sum_{j=0}^B 1_{ij}^{\text{noobj}} (c_i - \hat{c}_i)^2, \quad (9)$$

式中: $s$  为单元格; $B$  为边界框; $t_{\text{truth}}^{(r)}$  为实际的边界框的中心点坐标偏移值及高宽; $b_{\text{bbox}, i}^{(r)}$  为预测的边界框的中心点坐标偏移值及高宽; $c_i$  为边界框的实际置信度; $\hat{c}_i$  为预测的当前边界框的置信度; $\lambda_{\text{coord}}$ ,

$\lambda_{\text{obj}}$ ,  $\lambda_{\text{noobj}}$  为各项损失的权重; $i$  为所在的单元格; $j$  为所选的边界框; $1_{ij}^{\text{obj}}$  和  $1_{ij}^{\text{noobj}}$  为指示函数,当前边界框负责预测目标时, $1_{ij}^{\text{obj}}$  为 1,否则为 0,当前边界框不负责预测目标时, $1_{ij}^{\text{noobj}}$  为 1,否则为 0。式中第一项代表 bounding-box 中心点坐标偏移值及高宽的定位损失,第二项与第三项分别用来计算判别为结构光条纹目标的 bounding-box 的置信度损失( $c_i=1$ )以及判别为非结构光条纹目标的 bounding-box 的置信度损失( $c_i=0$ )。

## 5 基于焊缝图像的模型训练

### 5.1 数据获取及增强

深度卷积神经网络的权重参数是在采集数据的基础上通过多次迭代训练而获得,因此,获取焊缝图像数据集并对其进行标注是进行模型训练及测试的前提。在焊缝跟踪过程中,传感器所采集的图像包括未起弧时较干净的图像以及中间过程中含有各种噪声干扰的图像,因此在训练模型时需要准备不同类型的图像,以提高模型对各种焊缝图像的识别性能,提升其泛化能力。实验过程中采集了 516 张  $1280 \text{ pixel} \times 1024 \text{ pixel}$  大小的焊缝图像,并通过以下三种方式对采集到的焊缝图像进行数据扩增:a) 竖直方向平移 40 pixel;b) 对图像进行  $30^\circ$ 、 $45^\circ$ 、 $90^\circ$  旋转;c) 水平垂直翻转。之后通过标注软件来标注每幅图像相应的标签图。数据扩增之后最终得到 1012 张焊缝图像,扩增后的部分数据图像如图 3 所示。

### 5.2 模型训练

基于上述方法获得的数据,对设计的网络模型进行训练。模型是基于 Keras 以及 Tensorflow 框架实现的<sup>[18]</sup>,其余部分通过 python3.5 来实现。

训练以及预测时的实验配置为: Ubuntu 16. 04 操作系统, NVIDIA GTX1080 显卡, 8G 显存, Intel i7-6700 CPU。训练过程中, 以 656 张图像作为训练集、164 张作为验证集、192 张作为测试集来进行模型训练及测试。训练过程中的训练参数设置如下: 训练迭代次数为 16400 次, 学习率设置为 0. 0001, 采用 Adam 优化器进行优化, 由于显存较小, 单次迭代过程中参与训练的图片数量 batchsize 设置为 2。

## 6 实验分析与对比

### 6.1 网络性能测试

训练完成之后, 对测试集中的焊缝图像进行预测, 记录模型推理时间, 并通过标签图像计算预测精度。在语义分割场景任务中, 一般选择交并比 (IOU) 作为模型的精度评价指标, 表达式为

$$I = \frac{A_{\text{area}} \cap B_{\text{area}}}{A_{\text{area}} \cup B_{\text{area}}}, \quad (10)$$

式中:  $A_{\text{area}} \cap B_{\text{area}}$  为  $A$  与  $B$  的交集区域的面积;  $A_{\text{area}} \cup B_{\text{area}}$  为  $A$  与  $B$  的并集区域的面积。

表 2 各分支预测速度及精度

Table 2 Prediction speed and accuracy of each branch

Model	$I_{\text{cls0}} / \%$	$I_{\text{cls1}} / \%$	Mean IOU / %	AP / %	Time / ms	Frame rate / (frame·s <sup>-1</sup> )
Branch-seg	99.74	87.81	93.78	—	40	25
Branch-dec	—	—	—	97.40	17	59
Total-result	99.74	87.81	93.78	97.40	42	24

由表 2 可得, 当 Branch-seg 与 Branch-dec 分支单独运行时, 分割模型的检测速度为 40 ms/张,  $I_{\text{cls1}}$  为 87.81%, 检测模型的检测速度为 17 ms/张, AP 值为 97.40%。当两个分支的特征提取部分通过权重共享策略来同时进行推理时, 总体模型的检测速度为

所有类的平均交并比可通过混淆矩阵计算得来, 即

$$M_{\text{iou}} = \frac{1}{k+1} \sum_{i=0}^k \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}} + N_{\text{FN}}}, \quad (11)$$

式中:  $M_{\text{iou}}$  为交并比;  $k$  为类别数;  $N_{\text{TP}}$  为样本实际值为 1, 预测值为 1;  $N_{\text{FP}}$  为实际值为 0, 预测值为 1;  $N_{\text{FN}}$  为实际值为 1, 预测值为 0。

在目标检测场景任务中, 选择  $\text{mAP}^{[19]}$  值作为检测的评价指标, mAP 指所有数据中全部类别的 AP (平均精度) 平均值。

实验中对网络的 Branch-seg 语义分割分支与 Branch-dec 目标检测分支单独运行时的预测速度及精度进行了测试, 如表 2 所示。在 Branch-seg 分支的检测中, 预测结果分为背景和结构光条纹两类,  $I_{\text{cls0}}$  表示背景的分类比,  $I_{\text{cls1}}$  表示结构光条纹的分类比。此任务中, 分割的重点是结构光条纹, 因此在实验中以  $I_{\text{cls1}}$  作为主要精度指标来衡量该分支的精度。在 Branch-dec 分支中, 检测目标只包含结构光条纹一种类别, 因此本研究采用 AP 值作为该分支的精度指标。

42 ms/张, 推理时间少于两个分支的总时间, 具有一定的实时性。同时由表 2 还可得出, 模型中主要是语义分割分支花费的时间较多, 因此在 3.1 节中对语义分割分支进行模型计算量的优化是有必要的。

网络的实际检测效果如图 6 所示, 对于语义分

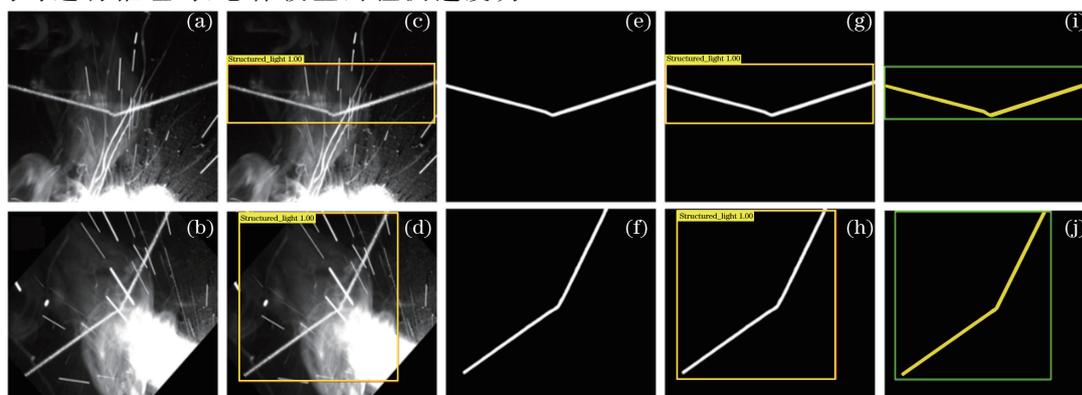


图 6 各分支预测结果。(a)(b)原始图像;(c)(d)检测分支的结果;(e)(f)分割分支的结果;(g)(h)总结果;(i)(j)真实标签

Fig. 6 Prediction results for each branch. (a)(b) Original images; (c)(d) results of Branch-dec; (e)(f) results of Branch-seg; (g)(h) total results; (i)(j) ground truth

割部分,可以在焊缝图像包含复杂噪声的环境下精确地分割结构光条纹;对于目标检测部分,能够精确地提取结构光条纹的位置;两个分支的检测精度都较高。

## 6.2 不同损失函数性能表现

实验过程中,在语义分割分支分别采用  $L_{\text{bce}}(Y^{\text{pred}}, Y^{\text{gt}})$  与  $L_{\text{bce\_Dice}}(Y^{\text{pred}}, Y^{\text{gt}})$  损失函数对网络进行训练,并在此基础上进行测试,通过测试精度来评价两种损失函数在面对焊缝图像检测时分割性能的优劣性。实验结果如表 3 所示,在面对结构光条纹分割过程中的正负类别失衡问题时,  $L_{\text{bce}}(Y^{\text{pred}}, Y^{\text{gt}})$  损失函数表现性能一般,  $I_{\text{cls1}}$  为 86.74%, 而采用  $L_{\text{bce\_Dice}}(Y^{\text{pred}}, Y^{\text{gt}})$  损失函数时的精度为 87.81%, 提高了 1.07%, 这证明该损失函数更适用于结构光条纹的检测过程,较好地解决了结构光条纹与背景像素比例失衡的问题。

表 4 不同网络性能对比

Table 4 Performance comparison of different networks

Model	$I_{\text{cls0}}/\%$	$I_{\text{cls1}}/\%$	Mean IOU /%	AP /%	Time /ms	Frame rate / (frame · s <sup>-1</sup> )
DT-Net	99.75	88.01	93.88	97.91	248	4
Tiny DT-Net	99.74	87.81	93.78	97.40	42	24
DT-ENet	99.72	86.73	93.22	97.40	135	7
DT-MobileNet_UNet	99.71	86.28	92.99	95.56	74	13

由表 4 可得,改进后的模型与使用 U-Net 网络作为语义分割分支的模型相比,检测速度大幅提升,分割及检测精度基本没有下降,  $I_{\text{cls1}}$  为 87.81%, AP 值为 97.40%, 并且与另外两种算法相比,其分割、检测精度以及实时性都较好;使用 U-Net 作为语义分割分支的模型,其分割及检测精度虽然较高,但是检测速度较慢,无法满足实时性要求,实用性较差。

为了更直观地体现改进后的网络性能,在图 7 中展示了多幅复杂噪声环境下的焊缝图像,同时给出了表 4 所示的 4 种方法的检测效果。图 7 中第 1 列是不同类型的焊缝图像,其余各列依次为 DT-Net、Tiny DT-Net、DT-ENet、DT-MobileNet\_UNet 的预测结果。对比图 7 中的第 2 列(DT-Net)与第 3 列(Tiny DT-Net)可以看到,在面对复杂噪声环境下的焊缝图像检测任务时, Tiny DT-Net 在缩小模型计算量、提高实时性的基础上,较好地完成了结构光条纹的检测,达到了与 DT-Net 相近的提取效果。对于图 7 中第 4 列(DT-ENet)与第 5 列(DT-MobileNet\_UNet)的预测结果,当焊缝图像如图 7(a)、图 7(d)中存在较多的弧光干扰时,

表 3 不同损失函数的性能对比

Table 3 Performance comparison of different loss functions

Loss	$I_{\text{cls0}}/\%$	$I_{\text{cls1}}/\%$	Mean IOU /%
$L_{\text{bce}}(Y^{\text{pred}}, Y^{\text{gt}})$	99.72	86.74	93.23
$L_{\text{bce\_Dice}}(Y^{\text{pred}}, Y^{\text{gt}})$	99.74	87.81	93.78

## 6.3 不同网络性能测试

如 3.1 节所述,在 Branch-seg 分割分支中,以 U-Net 为基础,设计了改进后的轻量化语义分割模型,同时在 Branch-dec 检测分支中,与 Branch-seg 分支共享特征提取部分的权重,从而加快了网络整体的推理速度。在这一节中将改进后的模型(Tiny DT-Net)与分别使用原始 U-Net 网络、MobileNet\_UNet 以及 ENet 作为语义分割分支的模型(分别称之为 DT-Net、DT-MobileNet\_UNet、DT-ENet)进行对比,测试其精度及速度指标,实验结果如表 4 所示。

图 7(m)、图 7(q)、图 7(p)、图 7(t)的分割结果均出现了断裂缺失现象,导致预测精度降低,且在图 7(m)中还检测出了噪声点;另外由图 7(s)中 DT-MobileNet\_UNet 预测结果的细节放大部分来看,其模型分割结果不够精细,在条纹线的相交处存在欠分割现象。

从图 7(b)中可以发现,当结构光条纹的一部分被弧光完全遮挡时,4 种方法的预测结果均有缺陷,但是 DT-Net 与 Tiny DT-Net 的预测缺陷相对较小,精度较高,因此在结构光条纹的整体架构已经分割出来的基础上,在后续的焊缝位置提取过程中可以消除该影响。

## 7 结 论

为了提高复杂噪声环境下焊缝跟踪的可靠性,利用深度学习中语义分割及目标检测相结合的模型,对焊缝跟踪过程中的结构光条纹进行检测。另外针对语义分割模型计算量过大的问题,基于 U-Net 的思想设计了轻量化分割网络,提高了检测速度。最后为了解决焊缝图像中正负类别失衡的问

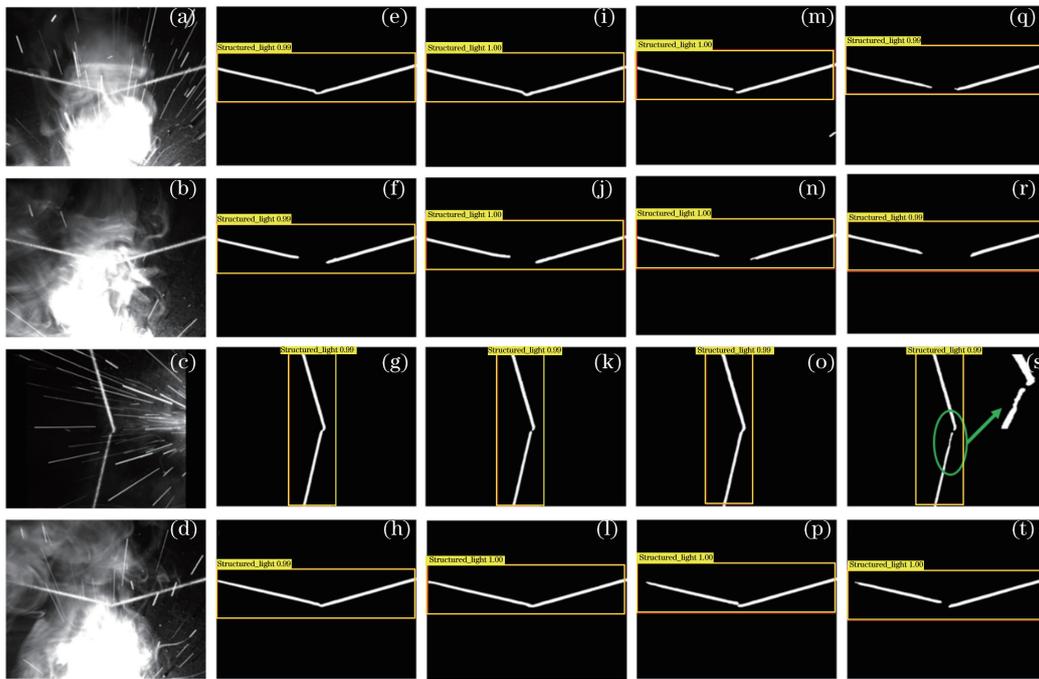


图 7 不同网络预测效果图。(a)(b)(c)(d)原始图像;(e)(f)(g)(h) DT-Net 检测结果;(i)(j)(k)(l) Tiny DT-Net 检测结果;(m)(n)(o)(p) DT-ENet 检测结果;(q)(r)(s)(t) DT-MobileNet\_UNet 检测结果

Fig. 7 Recognition results of different networks. (a)(b)(c)(d) Original images; (e)(f)(g)(h) test results of DT-Net; (i)(j)(k)(l) test results of Tiny DT-Net; (m)(n)(o)(p) test results of DT-ENet; (q)(r)(s)(t) test results of DT-MobileNet\_UNet

题,在语义分割分支的损失函数中添加 Dice 指标来修正网络对焊缝图像中背景像素的倾向性预测问题,从而提高预测精度。实验结果表明,通过上述方法可以在含有复杂噪声的焊缝图像中,以较高的精度检测出结构光条纹,同时具有一定的实时性。

在将来的工作中,将继续研究在保证算法实时性的基础上,对模型架构进一步优化,以提高检测精度,使模型在强噪声环境下的结构光条纹检测结果更理想。

### 参 考 文 献

- [1] Guo J C, Zhu Z M, Sun B W. A multifunctional monocular visual sensor based on combined laser structured lights [J]. Transactions of the China Welding Institution, 2019, 40(10): 1-7, 161.  
郭吉昌, 朱志明, 孙博文. 基于组合激光结构光的多功能视觉传感器[J]. 焊接学报, 2019, 40(10): 1-7, 161.
- [2] Zhang B, Chang S, Wang J, et al. Feature points extraction of laser vision weld seam based on genetic algorithm[J]. Chinese Journal of Lasers, 2019, 46(1): 0102001.  
张斌, 常森, 王桔, 等. 基于遗传算法的激光视觉焊缝特征点提取 [J]. 中国激光, 2019, 46(1): 0102001.
- [3] Yan Z G, Xu D, Li Y, et al. Automated robotic welding based on teaching and visual correction [J]. Robot, 2007, 29(3): 256-260.  
鄢治国, 徐德, 李原, 等. 基于示教与视觉纠偏的机器人自动焊接方法研究 [J]. 机器人, 2007, 29(3): 256-260.
- [4] Liu X W, Chen X M, Liu C Y. Image processing in welding seam tracking with structure light based on Radon transform and fuzzy-enhancement [J]. Transactions of the China Welding Institution, 2017, 38(2): 19-22, 1-2.  
刘习文, 陈显明, 刘超英. 基于拉东变换和模糊增强的结构光焊缝跟踪图像处理 [J]. 焊接学报, 2017, 38(2): 19-22, 1-2.
- [5] Huo P, Li J M, Zeng H L, et al. The real-time image processing V-shaped seam method based on a structure light [J]. Electric Welding Machine, 2014, 44(12): 144-148.  
霍平, 李军萌, 曾翰林, 等. 一种基于结构光的 V 型焊缝实时图像处理方法 [J]. 电焊机, 2014, 44(12): 144-148.
- [6] Chen H Y, Ren Y F, Wang Y N, et al. Clear image acquisition of V-shaped welding area based on visual attention [J]. Transactions of the China Welding Institution, 2018, 39(9): 19-24, 35, 129-130.

- 陈海永, 任亚非, 王亚男, 等. 基于视觉注意的 V 形焊接区域清晰图像获取 [J]. 焊接学报, 2018, 39(9): 19-24, 35, 129-130.
- [7] Li Y H, Liu P, Zhou J B, et al. Center extraction of structured light stripe based on back propagation neural network [J]. Acta Optica Sinica, 2019, 39(12): 1212005.
- 李玥华, 刘朋, 周京博, 等. 基于 BP 神经网络的结构光光条中心提取 [J]. 光学学报, 2019, 39(12): 1212005.
- [8] Zhang Y S, Yang G W, Wang Q Q, et al. Weld feature extraction based on fully convolutional networks [J]. Chinese Journal of Lasers, 2019, 46(3): 0302002.
- 张永帅, 杨国威, 王琦琦, 等. 基于全卷积神经网络的焊缝特征提取 [J]. 中国激光, 2019, 46(3): 0302002.
- [9] Zou Y B, Zhou W L, Chen X Z. Research of laser vision seam detection and tracking system based on depth hierarchical feature [J]. Chinese Journal of Lasers, 2017, 44(4): 0402009.
- 邹焱飏, 周卫林, 陈向志. 基于深度分层特征的激光视觉焊缝检测与跟踪系统研究 [J]. 中国激光, 2017, 44(4): 0402009.
- [10] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation [M] // Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015: 234-241.
- [11] 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. New York: IEEE Press, 2016: 779-788.
- [12] Howard A G, Zhu M L, Chen B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications [EB/OL]. (2017-04-17) [2020-08-29]. <https://arxiv.org/abs/1704.048>.
- [13] Zhao H S, Qi X J, Shen X Y, et al. ICNet for real-time semantic segmentation on high-resolution images [EB/OL]. (2017-04-27) [2020-08-29]. <https://arxiv.org/abs/1704.08545>.
- [14] Paszke A, Chaurasia A, Kim S, et al. Enet: a deep neural network architecture for real-time semantic segmentation [EB/OL]. (2016-06-07) [2020-08-29]. <https://arxiv.org/abs/1606.02147>.
- [15] Redmon J, Farhadi A. YOLO9000: better, faster, stronger [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.
- [16] Lin M, Chen Q, Yan S C. Network in network [EB/OL]. (2014-03-04) [2020-08-29]. <https://arxiv.org/abs/1312.4400>.
- [17] Milletari F, Navab N, Ahmadi S A. V-net: fully convolutional neural networks for volumetric medical image segmentation [C] // 2016 Fourth International Conference on 3D Vision (3DV), October 25-28, 2016, Stanford, CA, USA. New York: IEEE Press, 2016: 565-571.
- [18] Abadi M, Barham P, Chen J, et al. TensorFlow: a system for large-scale machine learning [C] // 2016 12th USENIX Conference on Operating Systems Design and Implementation, November 2-4, 2016, Savannah, GA, USA. New York: USENIX, 2016: 265-283.
- [19] Lin Z Y, Wu Q X, Fu S F, et al. Dual-NMS: A method for autonomously removing false detection boxes from aerial image object detection results [J]. Sensors (Basel, Switzerland), 2019, 19(21): E4691.