

DOI: 10. 19650/j.cnki.cjsi.J2006152

大视场大规模目标精确检测算法应用研究^{*}

张 堃¹, 姜朋朋¹, 华 亮¹, 费敏锐², 周挥宇³

(1. 南通大学电气工程学院 南通 226007; 2 上海大学机电工程与自动化学院 上海市电站自动化技术重点实验室 上海 210053; 3. 英国莱斯特大学信息学部 莱斯特 LE1 7RH)

摘 要:提出了一种基于深度学习的大视场大规模目标精确检测算法。以香烟滤棒为检测对象,研究具有强粘连、强阴影遮挡、强畸变、低对比度的大规模目标精确识别问题。提出一种 SAAU-net 模型,在 U 型网络基础上引入具有空间感知的自注意力增强模块和焦点损失函数有效地解决了精确语义分割问题。在精确识别环节中,根据圆相切几何规律设计了结构元匹配算法定位圆心并利用隐马尔可夫链优化搜索方向最终实现目标精确快速识别。在实验部分,分别对比了语义分割和目标识别环节中几种主流的算法。并在工业现场进行了 50 000 盒滤棒应用测试。试验结果表明,所提方法在语义分割、目标识别以及工业应用测试中都展现出强大竞争力,目标识别精度达到 99.95%。

关键词:大视场;大规模目标;自注意力增强;U 型网络;焦点损失函数;隐马尔科夫模型

中图分类号:TP273 TH89 **文献标识码:**A **国家标准学科分类代码:**510.80

Investigation of accurate detection of large-scale objects under a wide field of view

Zhang Kun¹, Jiang Pengpeng¹, Hua Liang¹, Fei Minrui², Zhou Huiyu³

(1.School of Electrical Engineering, Nantong University, Nantong 226007, China; 2.School of Mechatronic Engineering and Automation, Shanghai University, Shanghai Key Laboratory of Power Station Automation Technology, Shanghai 210053, China; 3.Department of Informatics, University of Leicester, LE1 7RH, UK)

Abstract: A new method is proposed to achieve large-scale cigarette filters detection under a wide field of view. It is expected to obtain accurate detection under the situation of overlapping, sheltering, distortion and low contrast. Firstly, by adding the spatially aware based the self-attention argument module and a Focal Loss function, an improved U-net model named as SAAU-net is proposed, which can achieve highly accurate semantic segmentation. After receiving a correct segmentation mask, the circle center of the cigarette filters is determined by object detection. Based on the circle theorem, the structural element matching is used to detect the circle centers. Hidden Markov model (HMM) is employed for direction searching. Experiments performed in simulation and industry application environments (with 5000 boxes) verify that the accuracy of the proposed approach can achieve 99.95%. Results also show that the proposed method has strong robustness to adapt different challenging environments.

Keywords: wide field of view; large scale objects; self-attention argument; U-net; focal loss; hidden Markov model

0 引 言

工业流水线常需要对复杂环境下大规模目标进行精确识别。以香烟过滤嘴厂为例,每台机器每分钟生产 12

盒 4 000 支装/盒烟滤嘴棒。正常情况下,滤棒堆垛装盒合格在 3 990~4 010 根之间波动。但由于机械设备精度误差,经常会发生堆垛装盒失误的现象。传统滤棒装盒检测的方法为人工检测,通过人工粗略计数来估计装盒是否正确。为了保证出厂率,一般人工计数的装盒会计

收稿日期:2020-03-04 Received Date:2020-03-04

^{*} 基金项目:国家自然科学基金重点项目(61633016)、国家自然科学基金面上项目(61872425)、江苏省六大人才高峰项目(XNY-039)、江苏省高等学校自然科学研究重大项目(19KJA350002)资助

数为3 950根,通过损失50根滤棒净利润来提高出厂合格率。

本文主要研究如何将机器视觉技术引入到滤棒堆垛装盒在线检测,其所面临的关键问题主要有:1)当出现低分辨率大视场、低对比度、阴影遮挡、边缘畸变和目标强粘连时,大规模目标如何精确分割;2)如何找出一种策略能实现大规模类圆堆垛目标精确目标识别。

经过前期研究,课题组提出的基于稀疏特征的小视场高速检测算法^[1]已经能针对低分辨率、低对比度目标进行精确识别。此外,也针对滤棒装盒检测提出了一种融合传统机器视觉和统计学方法^[2]。该方法能将检测精度提高到 ± 10 根。但是针对一些阴影遮挡非常严重,装盒规则化不强的情况,检测精度会浮动到 ± 20 根。

近年,深度学习算法在大数据和卷积神经网络支撑下,在工业现场的应用越来越多^[3-4]。尤其针对复杂目标分割问题有较好的解决方案。全卷积网络(fully convolutional networks, FCN)^[5]毫无疑问是语义分割领域的经典之作。在FCN出现之前,传统的卷积神经网络(convolutional neural networks, CNN)分割是将像素周围一个小区域作为CNN输入,做训练和预测,这样低效且不准确(忽略整体信息)。U-net^[6]采用了完全不同的特征融合方式:拼接。U-net采用将特征在channel维度拼接在一起,相比较FCN网络形成了更“厚”的特征,实现了网络对图像特征的多尺度特征识别。以上两种常见的图像分割模型核心模块是进行卷积操作获得不同维度特征。但是随时网络层加深以及原始图像存在特征难以捕捉的情况时,卷积学习的特征对模型贡献有限。近年,Attention注意力机制^[7]通过模拟人类特有的大脑信号处理机制在深度学习中的应用越来越广泛^[8-9]。Attention和U-net的融合是一种新的尝试^[10],文献[11]在U-net解码阶段,通过生成权重矩阵来进行特征增强,对于边缘不清晰的目标有较好的效果,但是Attention机制只部署在了U-net的Decode解码阶段,在Encode编码阶段无法生成新的特征。Self-Attention^[7]是Attention注意力机制新的发展方向,并作为替代卷积的手段在语义识别中脱颖而出。此外,基于轮廓分割思路的目标识别方法,如分水岭算法^[12]、水平集算法^[13]虽然能针对粘连目标有较好的分割效果,但针对复杂工业条件下单独通过识别边缘来提高识别精度,其效果仍然不佳。

针对以上问题及技术发展趋势,本文利用深度学习进行精确语义分割,得到效果较好的目标分割Mask图,为后续精确检测提供了基础支撑。提出一种SAAU-net模型实现了强粘连、强阴影遮挡、低对比度密集目标的精确语义分割,并通过改进的Focal Loss损失函数解决了目标、背景比例失调问题。在获得目标Mask基础上,本文放弃了边缘检测的识别思路,设计了结构元定位圆心和

基于隐马尔可夫的优化搜索算法,通过确定圆心来解决粘连目标和图像畸变带来的识别问题。最终系统的综合误差为 ± 2 根,整体精度达到99.95%,且系统检测速度可稳定在5 s/盒以内,满足实际工业现场要求。

1 基于SAAU-net的精确分割研究

1.1 Self-Attention Argument 机制

U-net的编码-解码网络结构在浅层网络时能够学习到合理的特征表示,但随着网络层数加深,编码-解码类模型将难以学到维度广且紧凑的特征表示^[14]。此外,网络模型中卷积运算有一个显著缺陷,即在局部近邻区域工作因而忽略全局信息。Self-Attention机制的出现很好地解决了这样的问题,并在语言理解领域已经得到了很好的应用。Self-Attention机制中, Q (Query), K (Key), V (Value)3个矩阵均来自同一输入,首先计算 Q 与 K 之间的点乘,并进行尺度缩放。再利用Softmax操作将其结果归一化为概率分布,然后再乘以矩阵 V 就得到权重求和的表示。该操作可以表示为式(1):

$$\text{Attention}(Q, K, V) = \text{soft max} \left(\frac{QK^T}{\sqrt{D_k}} \right) V \quad (1)$$

语言理解是一维向量特征表达,所以Self-Attention是否适用在2维图像一直是研究热点。自注意力满足交换律如式(2)所示。

$$\text{MHA}(\pi(X)) = \pi(\text{MHA}(X)) \quad (2)$$

式中: π 表示对于向量位置的任意置换。这使得对于模拟高度结构化的数据(例如图像)而言,不是很有效(二维图像数据并不满足这样的关系)。2019年,Google Brain提出了一种Attention Argument机制^[15],解决了Self-Attention无法在二维图像上存储位置信息的弊端,并在Resnet^[16]等模型上测试其识别性能比单独使用卷积要好。本文受其启发,U-net的编码-解码网络架构,和Self-Attention之前在语言理解领域常用网络模型类似。探索将Attention Argument机制移植到U-net网络中的可行性。为了加强深层网络中紧凑特征的学习能力,本文设计的SAAU-net网络在第3层到第7层网络部署了Self-Attention Argument模块,其他层保留传统卷积操作,其网络结构如图1所示。

在Self-Attention基础上,通过独立添加相对宽(W)和相对高(H)实现二维相关自注意力。对于像素 $i = (i_x, i_y)$ 关于像素 $j = (j_x, j_y)$ 的二维Attention计算方式如式(3)所示:

$$l_{i,j} = \frac{q_i^T}{\sqrt{D_k}} (k_j + r_{j_x-i_x}^W + r_{j_y-i_y}^H) \quad (3)$$

式中: q_i 表示对于 i 的查询矢量,也就是 Q 的第 i 行,也就

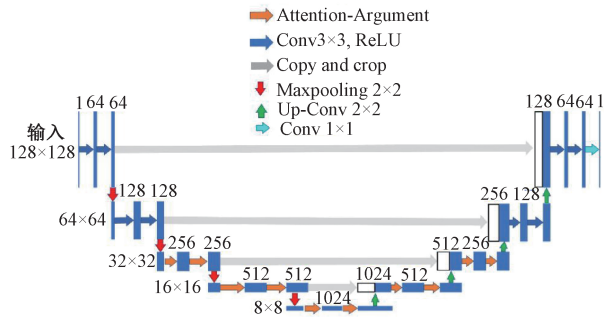


图1 SAAU-net 网络结构

Fig.1 Structure diagram of SAAU-net

是长为 D_k 的矢量; k_j 表示对于 j 的查询矢量,也就是 K 的第 j 行,也就是长为 D_k 的矢量; $r_{j_x-i_x}^w$ 和 $r_{j_y-i_y}^h$ 表示对于相对宽度 $j_x - i_x$ 和相对高度 $j_y - i_y$ 学习到的嵌入表示,各自均为 D_k 长度的矢量,对应于 r 组成的参数矩阵 r^w 和 r^h 是 $(2W-1, D_k)$ 和 $(2H-1, D_k)$ 。输出的包含位置信息的注意力权重矩阵如式(4)所示:

$$Attention = softmax\left(\frac{QK^T + S_H^{rel} + S_W^{rel}}{\sqrt{D_k}}\right) V \quad (4)$$

其中,两个 S 都是 $\in R^{HW \times HW}$ 的矩阵,表示沿着宽高维度的相对位置。Self-Attention Argument 模块工作机制如图2所示。Attention Argument 输入分成3个部分,第1部分为上层网络得到的 Feature Maps,其张量形状为 (H, W, Fin) ,

W, Fin),为了方便表达本文所有的张量形状均简化为高、宽、深度。 W 和 H 分别为 Feature Map 宽和高, Fin 为输入特征维度(深度)。第2和第3部分分别为和高度和宽度相关的测量矩阵,其张量形状为 $(2H+1) \times D_k$, $(2W+1) \times D_k$ 。Feature Maps 通过一个标准卷积生成一个 KQV 张量,其张量形状为 $(H, W, 2D_k + D_v)$ 。再将其按照某个特定维度进行切割,形成3个张量:“ Q (Query Tensor)”、“ K (Key Tensor)”和“ V (Value Tensor)”。 Q 和 K, V 的张量形状中宽和高和 KQV 张量保持一致,张量深度分别 D_k, D_k 和 D_v 。然后分别将 Q 张量和 K 张量平坦化成 $Flatten_Q \in R^{HW \times D_k}$, $Flatten_K \in R^{HW \times D_k}$ 。进一步对2个矩阵进行转置和内积计算后 reshape 成注意力矩阵 $\in R^{HW \times HW}$ 。同时,张量 Q 分别和高度相关测量矩阵以及宽度相关测量矩阵进行内积后再分别 reshape 成宽度编码矩阵和高度编码矩阵,且保证矩阵尺寸和注意力矩阵一致。3个矩阵相加后,得到的包含位置信息的注意力权重矩阵。张量 V 再和注意力权重矩阵进行内积生成 attention map $\in R^{H \times W \times D_v}$ 。进一步,通过求和得到对某个像素点进行编码时,相邻所有像素点对该像素的相关性编码。接下来,采用 1×1 卷积核对所有得到的编码特征进行组合拼接。最后将输入项通过一个标准 3×3 卷积得到的 feature map 和拼接好的编码特征组合成一个新的 feature map 作为输出结果,以代替常规卷积操作,由于该机制均为矩阵操作,所以在计算复杂度方面仅有少量增加。

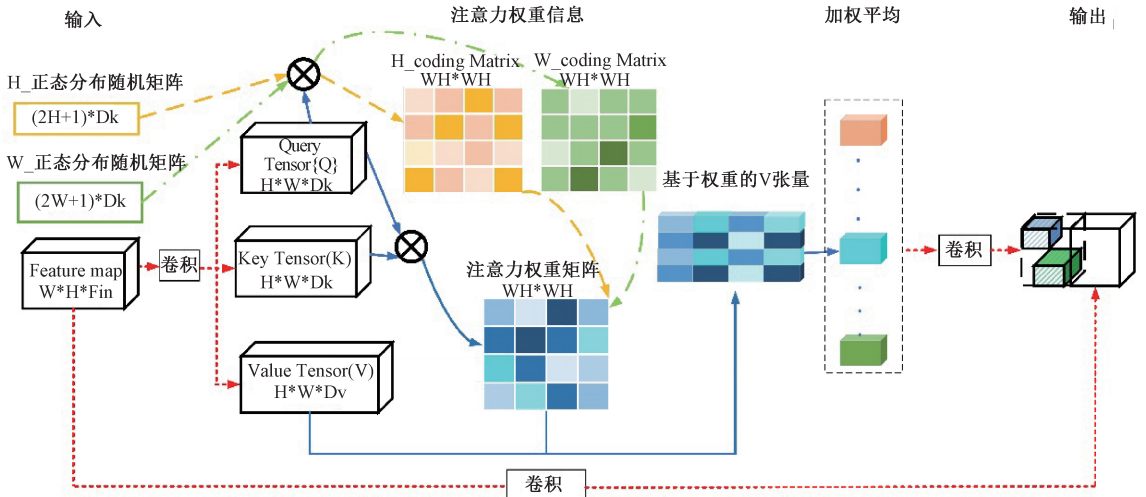


图2 Self-Attention Argument 机制流程

Fig.2 Diagram of Self-Attention Argument process

1.2 Focal Loss 损失函数

传统 U-net 中使用的损失函数为交叉熵 cross entropy loss,但是在本项目中现场实际图片中滤棒目标区域占总体区域的 80% 以上。使用交叉熵的话,对于难以判断的区域,模型就会出现欠拟合的现象。

Focal Loss 主要解决正负样本比例严重失衡的问题^[17]。该损失函数降低了大量重复样本在训练中所占的权重。通过在交叉熵损失函数基础上加了一个因子,其中 γ 使得减少易分类样本的损失。使得模型更关注于困难、错分的样本。加入平衡因子 α ,用来平衡正负样本

本身的比例不均。

$$L_{focal-loss} = \begin{cases} -\alpha(1-y')^{\gamma}\log y' \\ -(1-\alpha)y'^{\gamma}\log(1-y') \end{cases} \quad (5)$$

为了解决 batch 中正负比例范围在 75%~90% 波动情况。本文使用自适应整定 Focal Loss 参数策略。本文将 γ 值设为 2。对于 α 值,本文设计对每一个 batch 中统计正、负样本数。根据:

$$\alpha = \begin{cases} \alpha_P = \frac{|P| + |N|}{|P|}, & y = 1 \\ \alpha_N = \frac{|P| + |N|}{|N|}, & y = 0 \end{cases} \quad (6)$$

式中: P 定义为一批标签中“1”的总数;而 N 则是“0”的总数; y 表示标签的每个元素。基于上述分析,SAAU-net 的损失函数为式(5)和(6)。

2 基于隐马尔科夫的搜索策略研究

利用基于 SAAU-net 深度学习模型,可以获得较好的目标 Mask,但仍会出现少量分割不佳的情况。本文设计了一种融合类圆相切规律和隐马尔科夫推理的快速圆心搜索策略,规避了边缘识别依赖分割效果的弊端,转而从通过识别圆心反向进行精确计数。

2.1 结构元定位圆心

多个滤棒堆叠满足“圆相切”几何规律,即每个滤棒最多被 6 个滤棒包围,且近似以 60°/滤棒分布。只要确定一个滤棒的圆心即可推断出相邻滤棒的位置。

但是根据 SAAU-net 获得的 Mask 的结果,所获得的滤棒并非标准圆,呈现为不规则多边形及边缘锯齿状,如图 3 所示。传统基于中垂线找圆心的方法不适用,本文提出一种结构元定位圆心的方法,该方法可以降低 4 000 余个不规则圆心定位的误差累计。

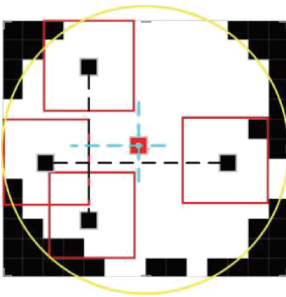


图3 结构元匹配

Fig.3 Structural element matching

圆心定位算法流程如下:

- 1) 在图像中央 50×50 感兴趣区域按匹配寻找第 1 个结构元位置。该结构元满足所有像素均为 255。
- 2) 对已检测到的结构元按照 0°, 90°, 180° 和 270°

4 个方向进行逐像素移动,当满足结构元出现 0 像素时停止移动。即确定将结构元移动到边缘位置。

3) 通过对 90°/270° 垂直方向和 0°/180° 水平方向上结构元的中心连线取中垂线,其交点可近似确定为滤棒中心。

2.2 基于隐马尔科夫的圆相切搜索策略

根据上文可知本方法获得是近似圆心,并且滤棒收到挤压和空间约束,并不完全满足“圆相切”的规律。在一个圆心基础上进行定步长和定角度搜索,误差累计会越来越大,最终会极大影响计数识别精度。因此,本文设计了一种基于隐马尔科夫模型(hidden Markov model, HMM)推理的圆相切搜索策略。

根据圆相切几何关系,当确定了一个圆心后,相邻的 6 个圆会出现在方向 $O_1:30^\circ, O_2:90^\circ, O_3:150^\circ, O_4:210^\circ, O_5:270^\circ, O_6:330^\circ$ 范围内。只需要确定合适的搜索步长 d ,即可搜索到 6 个方向上相邻圆内的结构元。根据相切原理,步长 d 具有一个可变范围: d_{min} 为历史滤棒半径累计平均值 r_{avg} , d_{max} 为 $3r_{avg}$ 。为了使 6 个方向上的圆心定位更准确,本文提出基于高斯分布来观测搜索步长 d 大小。

考虑到结构元确定的圆心为近似圆心,且滤棒受到挤压造成形变导致获得的也是近似的圆相切模型。如果能够预测出更优化的搜索步长 d 与搜索角度 O ,即可及时释放误差累计,提高整体检测精度。因此,采用 HMM 对搜索步长 d 和搜索角度 O 进行优化预测。以下称搜索步长缩放比例与搜索角度构成的叠加状态为比例-角度状态。

将搜索步长缩放比例与搜索角度作为 HMM 的隐藏链。从圆心出发,向 6 个方向到边缘的距离 x_i ,其中 $i=1:6$ 记为 HMM 的观测链。

比例-角度隐藏链状态表示如式(7)所示:

$$S = \{S_1, S_2, \dots, S_m\}, m = 9 \quad (7)$$

式中:缩放比例范围定为 $\pm 10\%$,其代表着基于 x 来缩放搜索步长 d 。搜索角度 O 也同时满足 $\pm 15^\circ$ 阈值范围,如图 4 所示。即有缩放比例-角度状态序列 $S, S_1 = -10\%, -15^\circ; S_2 = -10\%, 15^\circ; \dots; S_9 = 10\%, 15^\circ$ 。

通过 HMM 计算,每个 x_i 对应一个概率最大状态序列 S_m ,并通过极大似然法计算概率。滤棒各比例-角度初始概率如式(8)所示:

$$\psi = \{\psi_1, \psi_2, \dots, \psi_m\}, m = 9 \quad (8)$$

式中: ψ_i 表示状态 S_i 的初始概率。

经过统计分析可以得出任意的状态 S_i 出现的次数为 N_i ,经过一步转移的匹配的状态 S_i 到另一状态 S_j ,则状态转移概率如式(9)所示:

$$a_{ij} = \frac{\text{状态转移 } S_i \rightarrow S_j \text{ 次数}}{\text{状态 } S_i \text{ 出现的次数}} = \frac{n_j}{N_i} \quad (9)$$

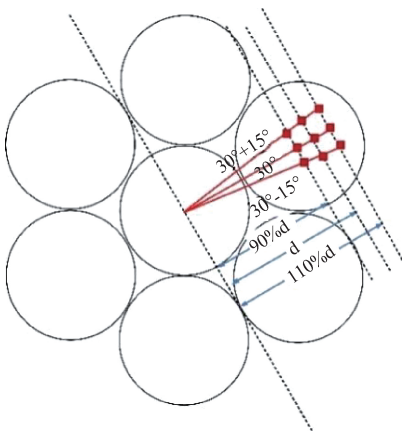


图 4 比例-角度状态

Fig.4 Ratio-angle status

将所有滤棒匹配比例-角度状态 S_i 的一步转移概率以矩阵的形式列出,则会得到一个 9×9 的一步转移概率矩阵:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{19} \\ a_{21} & a_{22} & \cdots & a_{29} \\ \vdots & \vdots & \ddots & \vdots \\ a_{91} & a_{92} & \cdots & a_{99} \end{bmatrix} \quad (10)$$

由此可以确定 HMM,其状态转移过程如图 5 所示。

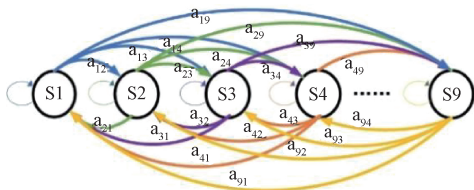


图 5 隐马尔科夫链状态转移

Fig.5 HMM status transition

实验发现,状态 S_i 和 x_i 服从高斯分布。状态 S_i 所对应的 x_i 个数记为 N'_i 因为 x_i 与 S_i 存在一一对应的关系,所以 $N_i = N'_i$ 。 x_i 的高斯密度函数的均值 μ_i 和方差 σ_i^2 记为:

$$\mu_i = \frac{1}{N'_i} \sum_{t=1}^{N'_i} x_i(t), 1 \leq i \leq m, 1 \leq t \leq N'_i \quad (11)$$

$$\sigma_i^2 = \frac{1}{N'_i - 1} \sum_{t=1}^{N'_i} (x_i(t) - \mu_i)^2 \quad (12)$$

由于不同方向状态 S_i 对应的距离 x 计算得到的均值 μ_i 和方差 σ_i^2 不同,所以在同一距离 x_k 对应不同比例-角度 S_i 发生概率也不同。给定距离 x_k ,其发生概率为:

$$b_i(x_k) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x_k - \mu_i)^2}{2\sigma_i^2}} \quad (13)$$

根据 Viterbi 算法,只需要使式 (14) 概率最大即可实现 x_i 对应的比例-角度 S_i 概率预测。

$$P(X, Y | \lambda) = \pi_i b_{X_1 Y_1} a_{X_1 Y_2} b_{X_2 Y_2} \cdots a_{X_n Y_n} \quad (14)$$

3 实验与讨论

本文实验使用的硬件环境为相机:300 万像素 CCD 工业相机;光源:贴片式条形 LED 灯带;CPU: E5- 2620 2. 10 GHz;GPU: TITAN XP 12 G;内存 64 G。软件环境是 Python3. 6, TensorFlow 1. 9. 0, VS 2015。

3.1 图像获取及数据库建立

所有试验图像数据为现场采集,现场图像处理硬件布置如图 6 所示。实际滤棒堆垛装盒图像如图 7 所示。

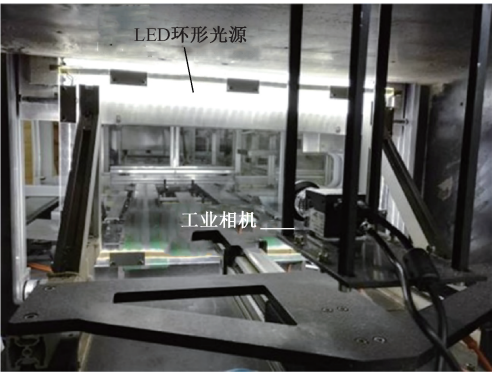
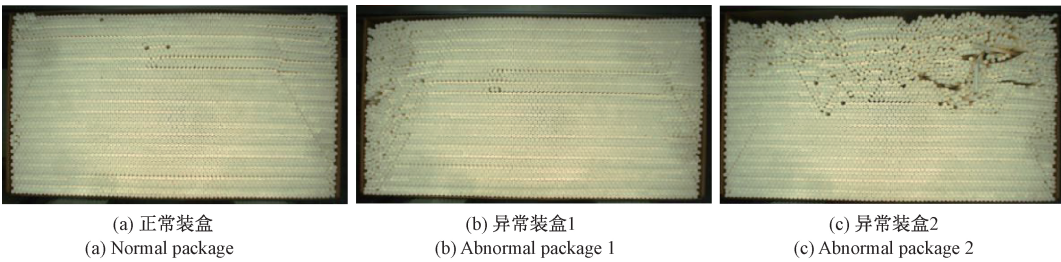


图 6 现场硬件布置

Fig.6 Assembly line at field



(a) 正常装盒
(a) Normal package

(b) 异常装盒1
(b) Abnormal package 1

(c) 异常装盒2
(c) Abnormal package 2

图 7 实际滤棒堆垛装盒
Fig.7 Cigarette filter rod package

带状 LED 光源布置在一面透光玻璃上,装盒机在完成滤棒堆垛装盒后,通过传动带移动到透光玻璃。进一步通过碰撞力,玻璃将整盒滤棒堆垛整齐。由于滤棒为纤维原料,长时间工作后,镜面上会有附着一些粘性纤维,

导致堆垛出现轻微异常,即容易引发图 8(a)、(d)所示的问题。LED 光源为环形布置,导致四周和中心光照强度不同,即容易引发图 8(b)、(c)所示的问题。系统整体检测流程如图 9 所示。

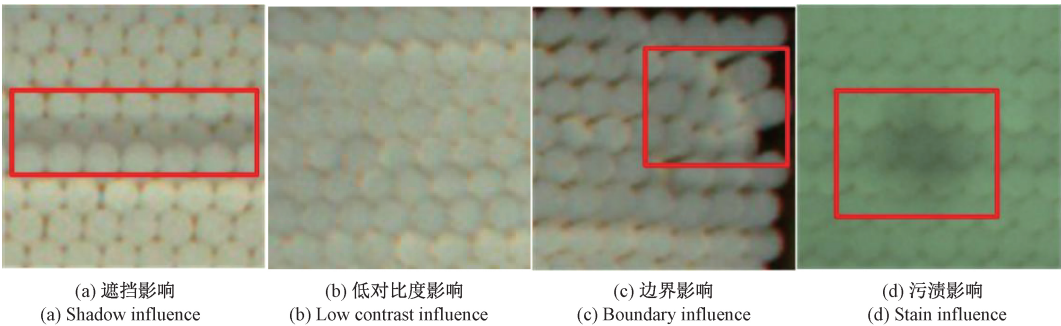


图 8 正常装盒中检测难点分析
Fig.8 Difficulty analysis under qualified package inspection

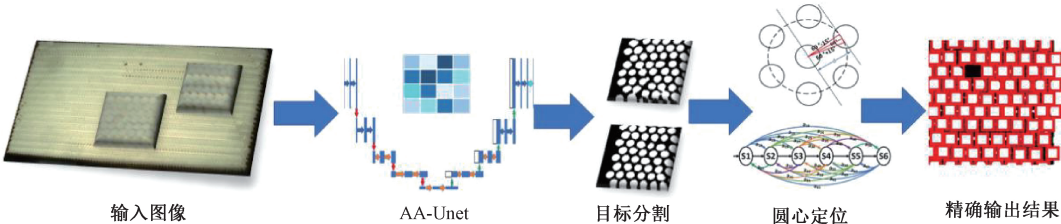


图 9 整体检测流程
Fig.9 Diagram of the whole detection process

本次实验共采集了 3 000 幅全尺寸图像,图像规格为 1 920×1 440 pixel, JPG 格式。通过随机选择,建立一个 1 500 幅图像的测试集,该测试集仅用于评估目的。剩余 1 500 幅图像 80%作为训练,20%作为验证。在每张全尺寸图像上随机选择 100 个 128×128 pixel 作为训练和验证样本,累计采集 12 000 个训练样本、3 000 个验证样本和 15 000 个测试样本。

3.2 语义分割效果比较

本文的模型与当前的主流分割模型 U-net 和 FCN 做了比较,其中 U-net 分别采用基于 Focal Loss 和 Cross Entropy 损失函数以及 U-net 变种网络 Attention Gated U-net。分割准确率和敏感率如表 1 所示,本文提出的方法均优于比较方法。图 10 为实际分割效果对比。当采用 Cross Entropy 损失函数时,模型对正负样本严重失调辨识能力比较弱。而 Focal Loss 对模型辨识有很强的作用。此外,U-net 较 FCN 能在低对比度,强粘连的情况下,分割结果更好。本文的 SAAU-net 较 Attention Gated U-net 在下采样同步增加 Attention 机制,识别精度又较其进一步提高。

表 1 语义分割模型性能对比		
Table 1 Semantic segmentation performance (%)		
模型类型	准确率	敏感度
U-net_FL	84.3	81.8
U-net_CEL	78.3	74.3
FCN_FL	80.7	78.6
AGU-net_FL	88.5	84.9
SAAU-net_FL	91.6	89.6

3.3 目标定位精度比较

由于语义分割精度还有近 10% 的误差,所以目标定位算法是进行最终精确计数的关键。本文对比了 3 种方法,其中 2 种为经典的 Haar 和 HoG 算法,此外还对比了课题组前期研究成果^[2]。对比效果如图 11 所示。在出现分割效果不佳的情况下,Haar 和 HoG 算法都出现了较严重的识别误差,文献[2]的方法由于加入了统计推理能减小误差,但是针对阴影遮挡情况,由于分割出的滤棒已经严重失真,导致该区域完全不能被识别到。本文提出算法由于采用圆心定位策略,所以边缘误差对目标识别影响很小。

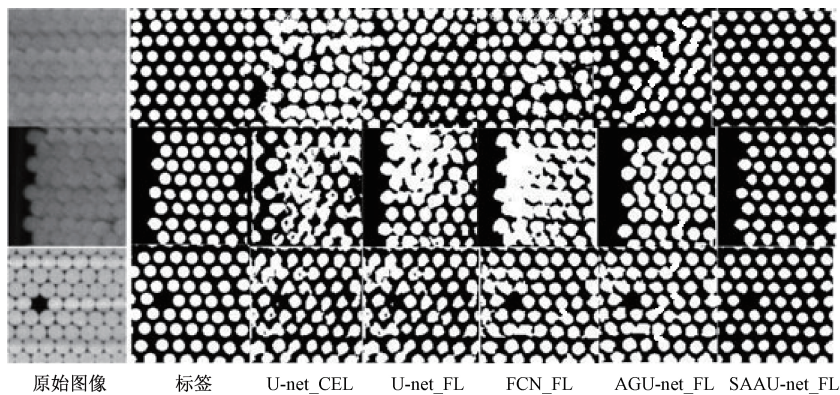


图 10 可视化结果对比

Fig.10 Visualize segmentation performance comparison

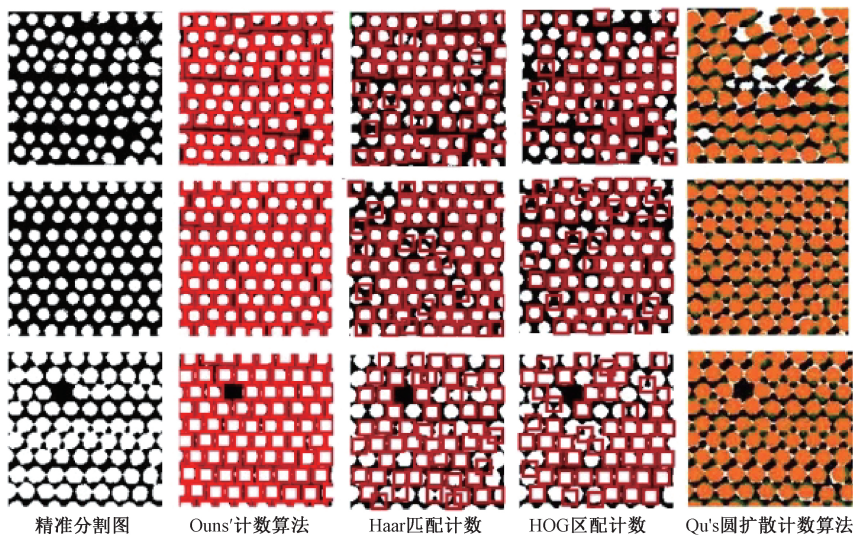


图 11 目标识别算法对比效果

Fig.11 Diagram of object detection comparison

3.4 工业性能对比分析

为了衡量在工业现场的实际性能,针对两台装盒机同时进行 50 000 盒共 11 个批次的计数精度对比,两台设备分别运行文献[2]和本文方法。对比结果如表 2 所示,冷机启动时,前 3 个批次,文献[2]方法基本稳定,当批次越来越大后,室内灯光对图像影响越来越大。玻璃上附着的静电纤维也提高了识别难度。机器长时间运行,装盒质量波动加大,滤棒遮挡,滤棒变形的情况越来越多,文献[2]方法越来越不稳定。本文采用 SAAU-net 模型对变形和粘连较严重的滤棒目标有较好的抑制作用并可以准确地确定各滤棒的中心点,对滤棒的计数具有较高的精度。

表 2 不同批次运行计数效果对比
Table 2 Cigarette filters counting comparison under different batches

批次 ×10 ³	文献 2 方法		本文方法	
	误检率/%	误检数	误检率/%	误检数
0~5	0.407	1~3	0.02	1~1
5~10	0.410	1~5	0.018	1~1
10~15	0.434	1~5	0.020	1~1
15~20	0.519	1~7	0.024 1	1~2
20~25	0.562	1~6	0.028 2	1~2
25~30	0.602	1~10	0.029 3	1~2
30~35	0.631	1~6	0.029 5	1~2
35~40	0.687	1~8	0.029 8	1~2
40~45	0.782	1~11	0.030 7	1~2
40~45	0.806	1~7	0.032 0	1~2
45~50	0.854	1~9	0.032 9	1~2

4 结 论

本文以香烟滤棒堆垛装盒检测为例,研究了大视场大规模目标精确识别问题。针对低对比度、阴影遮挡、大视场畸变和强粘连情况,设计了基于 SAAU-net 深度学习模型进行目标分割,并通过改进的 Focal Loss 解决了正负样本严重失衡的问题。在获得良好分割结果基础上,根据类圆堆垛的几何特征,分别设计了结构元圆心定位和隐马尔科夫圆相切搜索策略,解决了大规模目标精确识别问题。试验表明,由于检测方法鲁棒性强,在不同的工况情况下都能精确识别,检测精度达到 99.95%。并且通过优化搜索算法,整体系统检测速度不高于 5 s/盒,满足企业实际需求。本研究可以推广到其他大视场大规模类圆精确检测系统中。

虽然大视场多目标精确识别已经具有逐步完善的理论并在实践中得到了部分应用场景的验证,但受制于数据误差、人为和环境因素等,目前该领域还不能实现理想的无人值守,如受光学镜头的限制,针对大视场场景下的图像边缘往往受畸变影响很大。今后将重点围绕基于领域知识的深度学习模型进行研究突破。完全无监督的深度学习技术已经发展到一定的瓶颈,融合领域知识和深度学习技术可以将深度学习的黑箱模式转变成一种可控模式。如针对边缘畸变的场景设计深度学习机制,有效的矫正边缘畸变对图像识别的影响。

参考文献

[1] 张堃, 华亮, 费敏锐, 等. 基于稀疏特征的小视场高速检测算法应用研究[J]. 仪器仪表学报, 2018, 39(12): 179-189.
ZHANG K, HUA L, FEI M R, et.al. Applied research on sparse feature representation for high-speed intelligent detection under a small field of view[J]. Chinese Journal of Scientific Instrument, 2018, 39(12): 179-189.

[2] QU H J, ZHANG K, ZHANG P J. Research on cigarette filter rod counting system based on machine vision[C]. Communications in Computer and Information Science, 2017, 761:513-523.

[3] 包俊, 叶波, 王晓东, 等. 基于 SSDAE 深度神经网络的钛板电涡流检测图像分类研究[J]. 仪器仪表学报, 2019, 40(4): 238-247.
BAO J, YE B, WANG X D, et.al. Research on eddy current detection image classification of titanium plate

based on SSDAE deep neural networks[J]. Chinese Journal of Scientific Instrument, 2019, 40(4): 238-247.

[4] LEE H, EUM S, KWON H. R-CNN: multi-expert R-CNN for object detection[J]. IEEE Transactions on Image Processing, 2020, 29:1030-1044.

[5] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.

[6] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]. Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS, 2015, 9351: 234-241.

[7] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]. Advances in Neural Information Processing Systems, 2017: 5998-6008.

[8] GUO H, TANG K Z, WANG W J. Two-level attention network with multi-grain ranking loss for vehicle re-identification[J]. IEEE Transactions on Image Processing, 2019, 28(9): 4328-4338.

[9] YU T, YU Z, TAO D. Compositional attention networks with two-stream fusion for video question answering[J]. IEEE Transactions on Image Processing, 2020, 29: 1204-1218.

[10] LYU Y, LI J, LIU S. Attention guided U-net with atrous convolution for accurate retinal vessels segmentation[J]. IEEE Access, 2020, 8:32826-32839.

[11] OKTAY O, FOLGOC M, LEE M H, et.al. Attention U-Net: Learning where to look for the pancreas[C]. MIDL 2018, Imaging with Deep Learn, 2018: arXiv: 1804.03999.

[12] ZHANG K, FEI M R, HUI Y Z. Using entropy based mean shift filter and modified watershed transform for grain segmentation[J]. International Journal of Automation and Computing, 2015, 12(2):199-207.

[13] LI C M, XU C Y, GUI C F. Level set evolution without re-initialization: A new variational formulation[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2005, 1:430-436.

[14] ZHANG Z, ZHANG Y, BAO J, et.al. Human activity recognition based on motion sensor using U-Net[J]. IEEE Access, 2019, 7: 75213-75226.

[15] BELLO I, ZOPH B, VASWANI A, et. al, Attention augmented convolutional networks [C]. IEEE International Conference on Computer Vision (ICCV), 2019;3285-3294.

[16] HE K, ZHANG S R, SUN J. Deep residual learning for image recognition [C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2016;770-778.

[17] LIN T Y, GOYAL P, HE K M, et.al, Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42 (2): 318-327.

作者简介



张堃,2005 年于江苏大学获得学士学位,2010 年于浙江大学获得硕士学位,2016 年于上海大学获工学博士学位。现为南通大学副教授。主要研究方向为过程控制、人工智能、机器视觉。
E-mail: zhangkun_nt@ 163.com

Zhang Kun received his B. Sc. degree from Jiangsu University in 2005, received his M. Sc. degree from Zhejiang University in 2010 and received his Ph. D. degree from Shanghai University in 2016. He is currently an associate professor at Nantong University. His main research interests include machine vision, artificial intelligence and process control.



华亮(通信作者),2001 年于南通工学院获得学士学位,2007 年和 2014 年于浙江工业大学分别获得硕士学位和博士学位,现为南通大学电气工程学院院长、教授,主要研究方向为机器人控制、图像处理。
E-mail: hualiang@ ntu.edu.cn

Hua Liang (Corresponding author) received his B. Sc. degree from Nantong Instituted Technology in 2001, and received his M. Sc. degree and Ph. D. degree both from Zhejiang University of Technology in 2007 and 2014, respectively. He is currently a professor at Nantong University. His main research interests include robot control, machine vision, pattern recognition.