

复杂环境下一种基于 SiamMask 的时空预测移动目标跟踪算法

周珂¹⁾, 张浩博²⁾✉, 付冬梅²⁾, 赵志毅¹⁾, 曾惠²⁾

1) 北京科技大学高等工程师学院, 北京 100083 2) 北京科技大学自动化学院, 北京 100083

✉通信作者, E-mail: cocofay126@126.com

摘要 随着无人工厂、智能安监等技术在制造业领域的深入应用, 以视觉识别预警系统为代表的复杂环境下动态识别技术成为智能工业领域的重要研究内容之一。在本文所述的工业级视觉识别预警系统中, 操作人员头发区域由于其具有移动形态非规则性、运动无规律性的特点, 在动态图像中的实时分割较为困难。针对此问题, 提出一种基于 SiamMask 模型的时空预测移动目标跟踪算法。该算法将基于 PyTorch 深度学习框架的 SiamMask 单目标跟踪算法与 ROI 检测及 STC 时空上下文预测算法相融合, 根据目标时空关系的在线学习, 预测新的目标位置并对 SiamMask 模型进行算法校正, 实现视频序列中的目标快速识别。实验结果表明, 所提出的算法能够克服环境干扰、目标遮挡对跟踪效果的影响, 将目标跟踪误识别率降低至 0.156%。该算法计算时间成本为每秒 30 帧, 比改进前的 SiamMask 模型帧率每秒提高 3.2 帧, 算法效率提高 11.94%。该算法达到视觉识别预警系统准确性、实时性的要求, 对移动目标识别算法模型的复杂环境应用具有借鉴意义。

关键词 深度学习; 复杂环境; 移动目标识别; SiamMask; STC

分类号 TG142.71

Design and implementation of multi-feature fusion moving target detection algorithms in a complex environment based on SiamMask

ZHOU Ke¹⁾, ZHANG Hao-bo²⁾✉, FU Dong-mei²⁾, ZHAO Zhi-yi¹⁾, ZENG Hui²⁾

1) School of Advanced Engineering, University of Science and Technology Beijing, Beijing 100083, China

2) School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China

✉ Corresponding author, E-mail: cocofay126@126.com

ABSTRACT Moving target recognition in a complex environment is recently an important research direction in the field of image recognition. The current research focus is how to track moving objects online in complex scenes to meet the real-time and reliability requirements of image tracking and subsequent processing. With the in-depth application of unmanned factory, intelligent safety supervision and other technologies in the field of manufacturing industry, dynamic recognition technology in the complex environment represented by a visual recognition warning system has become an important research in the field of intelligent industry, and the detection requirements of high reliability and real-time for mobile target detection have been identified. In the industrial level vision recognition warning system described in this paper, the hair area of operators was difficult to be segmented in real time because of its irregular movement. To solve this problem, a space-time predictive moving target tracking algorithm was proposed based on the SiamMask model. This algorithm combined the SiamMask single target tracking algorithm based on the PyTorch deep learning framework with ROI detection and STC spatiotemporal context prediction algorithm. According to the online learning of the spatiotemporal relationship of the target, it predicted the new target location and corrected the algorithm of the SiamMask model to

收稿日期: 2019-06-06

基金项目: 国家自然科学基金资助项目 (61375010); 北京科技大学基本科研业务费资助项目 (FRF-OT-18-020SY)

realize the fast recognition of the target in the video sequence. The experimental results show that the proposed algorithm can overcome the influence of environmental interference and target occlusion on the tracking effect, reducing the target tracking error recognition rate to 0.156%. The computational time cost is 30 frames per second, which is 3.2 frames per second greater than the frame rate of the improved SiamMask model and 11.94% greater efficiency than that of the original SiamMask model. The algorithm meets the requirements of accuracy and real-time performance of the visual recognition and early warning system, and has reference significance for the application of the moving target recognition algorithm model in a complex environment.

KEY WORDS deep learning; complex environment; moving target recognition; SiamMask; STC

视频图像的跟踪及特征分割, 是视频应用中的一项基本任务, 如自动监视^[1], 车辆导航, 视频标签, 人机交互^[2] 和行为识别等。如何在使用场景中, 对视频流在线执行目标跟踪, 满足图像跟踪及后续处理的实时性、可靠性需求, 是目前图像识别领域的重要研究方向。

2016年, Bertinetto 等^[3] 提出 SiamFC 算法, 利用 CNN 的全卷积 (Fully-convolutional) 神经网络进行实时图像的跟踪处理, 克服了传统深度学习在实时视频图像处理中计算量大、效率过低的缺陷。2018年, Li 等^[4] 结合了 SiameseFC 和 Faster R-CNN 中的 RPN (Region proposal network), 使得实时图像的提取速度大大加快, 同时也达到较高精度效果, 其目标跟踪算法的代码实现精度在 VOT2018 排名第一, 鲁棒性大大增强。2019年, 中科院 Wang 等^[5] 提出新算法 SiamMask, 实时进行目标跟踪和半监督视频对象分割, 采用基于数百万视频帧上离线训练的全卷积网络 Siamese, 缩小任意目标跟踪与 VOS 之间的差距, 在视频跟踪任务上达到最优性能, 并且在视频目标分割上取得了当前最快速度。

上述算法在已有开源数据库的实时识别中均有出色表现, 但在应用场景里, 由于识别背景的复杂性及遮挡性、目标特征的差异性、环境影响的随机性等因素, 会存在一定的目标误识别现象^[6-7]。针对如何在识别准确度要求较高的应用场景中 (如生产安全智能预警系统中的移动目标实时图像识别), 实现高效、快速的实时图像识别, 开展了本文研究内容。提出一种基于 SiamMask 的时空预测移动目标跟踪算法, 克服环境干扰、目标遮挡对跟踪效果的影响, 实现视频序列中的目标快速识别, 对移动目标识别算法模型的复杂环境应用具有借鉴意义。

1 基于 SiamMask 的时空预测移动目标跟踪算法设计

1.1 应用场景提出

机床加工在工业生产中应用极为广泛, 实际

生产操作环节里, 由于人的安全防护不符合规范, 操作人员的头发往往是导致严重安全事故的隐患之一。如女性操作车床时, 若未扎好头发或未戴帽子, 散落的发梢极易被卷进高速旋转的车床部件中, 带来严重的人身安全事故。随着数字化管理手段的普及, 操作环节的全时监控已在生产过程中得到普遍应用, 预警系统可通过监控视频的目标识别, 达到对危险因素监测、预警的目的。然而头发由于移动中的非规则性、运动无规律性, 使得头发区域在运动图像中的分割较为困难。常见的动态图像识别 Camshift 算法以颜色直方图为特征对目标进行跟踪^[8-9], 对刚性目标的跟踪具有较强的鲁棒性。但当目标过于复杂、目标被遮挡或做加速运动时, 很容易发生目标跟丢的情况。目前国内外头发区域分割多集中在二维图像中^[10-12], 对移动目标的头发区域分割尚无工业实现的先例。

本文的动态识别检测以北京科技大学工程训练中心金工实习车床加工环节为样本, 其结果需满足生产环节的安全操作预警, 要求较高的识别准确性及算法效率。因此提出了一种基于 PyTorch 深度学习框架, 将 SiamMask 单目标跟踪算法与 ROI (Region of interest) 检测及 STC (Spatio-temporal context learning) 时空上下文目标跟踪算法融合的方法^[13], 满足实际场景中对人员头发区域动态分割识别的任务需求。

1.2 基于 SiamMask 模型的面部动态检测实现

动态视频检测的重要技术难题就是如何更快、更有效, 本文基于 2019 年 5 月 PySOT 目标跟踪库^[14] 的最新成果 SiamMask 模型进行面部动态检测^[15]。SiamMask 是当前视频图像目标跟踪模型中, 最为出色的快速跟踪和分割方法。该算法基于 PyTorch 深度学习框架, 实时进行目标跟踪和半监督视频对象分割, 通过增加目标分割的损失, 改进了全卷积网络 Siamese 跟踪方法的离线训练过程。经过训练, SiamMask 完全依赖于初始化单个边界框并在线操作, 产生类别未知的目标分割 Mask 和实时每秒 35 帧的旋转边界框。在 VOT-

2018、DAVIS-2016 和 DAVIS-2017 上均有极为出色表现。

1.2.1 算法移植测试优势

如图 1 所示, Mask 生成的网络采用的两层 1*1 卷积, 分别是 256 通道和 63*63 通道数, 并采用了 Refinement module(微调模块), 这是一种不同分辨率的特征融合策略。在 SiamFC 和 SiamRPN 基础上分别加上目标分割分支网络, 形成两分支和三分支的 SiamMask 网络。图 1 左边是共享权值的孪生网络, 在对模板(上分支网络)和搜索区域(下分支网络输入)特征提取后, 进行卷积生成与模板最相似的相应区域。图中, *d 表示深度相关权重叠加, Row 表示响应窗口, Mask 表示视频目标物体, Box 为预测盒, Score 为预测得分。f_θ、h_φ、b_σ、s_φ、p_ω 分别为 CNN、目标物体、预测盒、预测得分、响应窗口预测得分的处理函数^[5]。

SiamMask 模型通过大量的深度学习网络离线训练, 将在线学习工作量节约至最小状态, 整个模型通过初始帧跟踪目标设定即可完成整段视频的

目标跟踪处理, 具有非常强的实时性、便捷性、计算低成本性。考虑应用场景对识别快速性的要求, 采用 SiamMask 对图像进行初步处理。为更好的测试算法可用性, 算法处理环境为工业环境中易于实现的单 GPU 处理板。

图 2 为 SiamMask 模型在工业系统视频中的测试效果, 对不同发型及尺寸视频目标的头部追踪, 该模块表现出较高的识别效率, 平均识别速度可达每秒 26.8 帧。

1.2.2 算法移植测试误识别现象

尽管算法识别效率高, 但在环境嘈杂时, 会出现跟踪误识别情况。如图 3(a)、(b)所示, 当追踪对象周围出现深色服装、肉色物体等干扰源时, 算法产生误识别现象; 图 3(c)为人脸被头发遮挡后, 该模型失去跟踪目标。此外, 在目标出画面之外再进来的情况下, 继续跟踪效果不佳。

将 SiamMask 模型对 7 段测试视频的追踪效果进行统计, 得到表 1 所列情况。

结合图 3 和表 1 数据可以看出, SiamMask 网

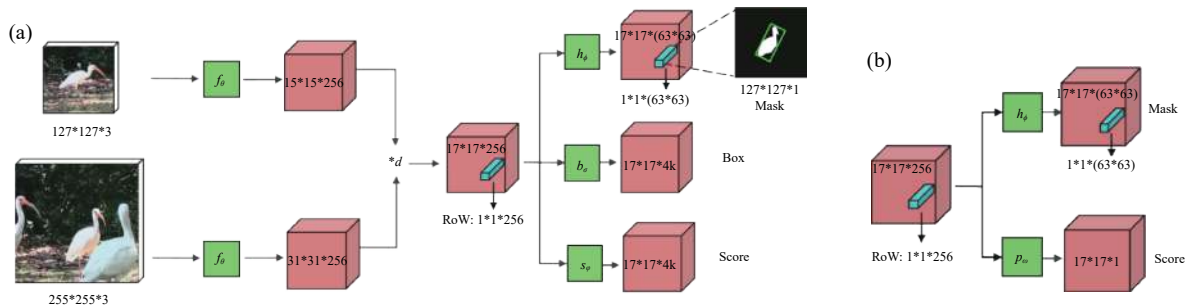


图 1 SiamMask 模型算法流程图^[5]: (a)三分支变型架构; (b)二分支变型架构核心

Fig.1 SiamMask model algorithmic flow chart^[5]: (a) three-branch variant architecture; (b) two-branch variant head



图 2 SiamMask 模型面部检测效果: (a)束发头部跟踪; (b)长发头部跟踪 I; (c)长发头部跟踪 II

Fig.2 SiamMask model face detection effect: (a) bundle head tracking; (b) long hair head tracking I; (c) long hair head tracking II



图 3 SiamMask 模型测试误识别现象: (a)深色干扰源误识别; (b)肉色干扰误识别; (c)头发遮挡误识别

Fig.3 SiamMask model test misrecognition phenomenon: (a) misidentification of dark interference sources; (b) misidentification of flesh color interference; (c) misidentification of hair occlusion

表 1 SiamMask 模型目标跟踪效果统计

Table 1 Statistics of target tracking effect of the SiamMask model

Video No.	Frame number of false detection	Analysis on the causes of false inspection	Total frames	Failure rate/%
1	0	Little change in this movement	361	0
2	87	Misidentified as dark cloth	288	30.21
3	98	Part of the face is blocked by the hair	192	51.04
4	674	Initialization offset, screen will pop up in recognition	1380	48.84
5	131	The target moves out of the screen slightly and the recognition is lost	240	54.58
6	753	Large proportion of face selection in initialization area	1360	55.37
7	0	Accurate initialization and small action range	241	0

络模型会在以下情况产生误识别现象:

1) 初始化区域选定的面部比例较大时, 效果存在较多误识别现象.

2) 干扰项出现时若产生误识别现象, 在干扰因素消失时, 无法及时恢复目标跟踪. 在本文的对象视频中表现为, 目标移出视频后再返回则不再跟踪. 由图 3 所示, 目标误识别后, 即便目标仍在同一区域, 也难以校正.

1.3 SiamMask 模型的修正处理

针对 SiamMask 模型的应用不足, 提出算法修正. 对原始工业视频做 ROI 提取, 并以 STC 时空上下文跟踪法^[16-17]对视频序列进行匹配校正, 从而减少工业环境中的目标漏检、误检率, 实现复杂环境中头发区域的精准检测. 算法框架图如图 4 所示. 图中, P_k 为第 k 帧图像, P'_k 为 STC 预测的第 k 帧图像, Φ 为判断函数.

修正 1: ROI 区域提取. 如上文所述, SiamMask 模型初始 Mask 对视频后序识别的误差具有积累性, 因此在算法修正时首先对视频进行了识别目标与非相关性背景的分离预处理^[18].

修正 2: STC 匹配校正. STC 模型是一种考量了目标与背景相关性的算法理念, 具有较高的目标跟踪精度与速度. 本文基于 STC 算法, 结合时间

与空间信息, 根据对头发目标时空关系的在线学习预测新的目标位置, 并与 SiamMask 模型的跟踪结果进行匹配校正.

1.3.1 工业视频 ROI 提取

本文所使用的视频来自于实际车工场地监控镜头录制的彩色视频, 分辨率为 1920×1080 , 帧率为每秒 24 帧, 总帧数截取 3427 帧. 从图 5(a) 可以看出, 视频画面中包含多个机床工位, 往来人员较多, 这给单个目标的跟踪、检测带来一定干扰. 面对实际情况, 针对单个车床目标, 需要对获得的视频进行 ROI 提取操作. 其中包括: 视频 ROI 区域设定、全图运动检测两部分.

(1) ROI 区域设定.

具体 ROI 提取步骤中有两个关键点: ROI 区域起点位置、ROI 区域的分辨率. 应用先验知识, 可以确定 ROI 区域左上角起点像素位置 $A(x,y)$, ROI 区域宽度为 l_w , 高度为 l_h , 如式 (1) 所示:

$$\begin{aligned}
 x &= l'_w \times 0.268 \\
 y &= l'_h \times 0.502 \\
 l_w &= l'_w \times 0.426 \\
 l_h &= l'_h \times 0.497
 \end{aligned} \tag{1}$$

式中, l_w 为原画面宽度, l'_h 为原画面高度.

通过自动获取原始视频的信息, 按照式 (1) 进

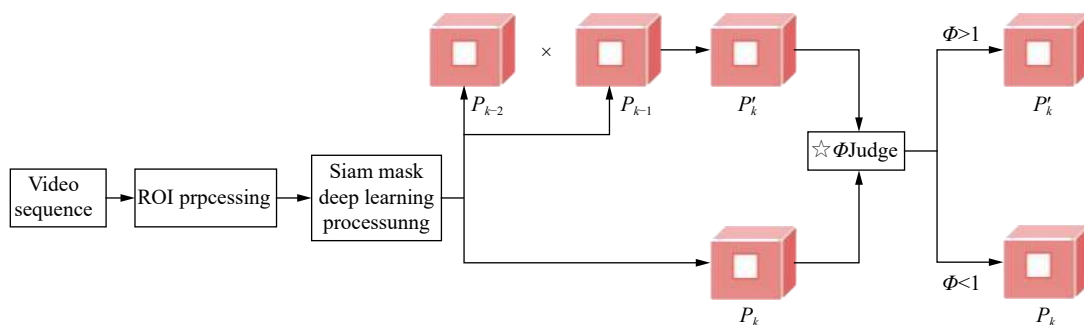


图 4 基于 SiamMask 模型的时空预测移动目标跟踪算法框架图

Fig.4 Framework of spatiotemporal prediction moving target tracking algorithms based on the SiamMask Model



图5 车工监控视频 ROI 提取结果。(a)原始画面;(b)ROI 提取画面

Fig.5 ROI extraction result of a locomotive monitoring video: (a) original picture; (b) ROI extraction picture

行相对比例的计算, 处理结果如图 5(b)所示, 裁切得到的视频宽度为 816 个像素、高度为 536 个像素、ROI 左上角起点坐标像素为(516, 541)。

(2)ROI 全图运动检测及自动提取。

步骤 1: 设 $I_k(z)$ 为像素点 $z(x_z, y_z)$ 灰度值, 对输入视频相邻的三帧图像, 进行像素点灰度值变化计算。

步骤 2: 判断相邻帧之间灰度变化, 对差值标

志赋值。在训练中选取灰度差阈值 $\Delta I_T=30$ 。

步骤 3: 判断两个差值标志位逻辑结果同时为 1 时, 代表三帧中有运动目标存在。

步骤 4: 对于输入的序列图像, 按照时序统计每三帧之间的灰度变化像素点的个数, 当运动目标出现时, 灰度变化像素点个数成跳变趋势, 从该时刻进行 ROI 区域的提取, 即按照 ROI 区域先验设定尺寸进行裁切。如图 6 所示。

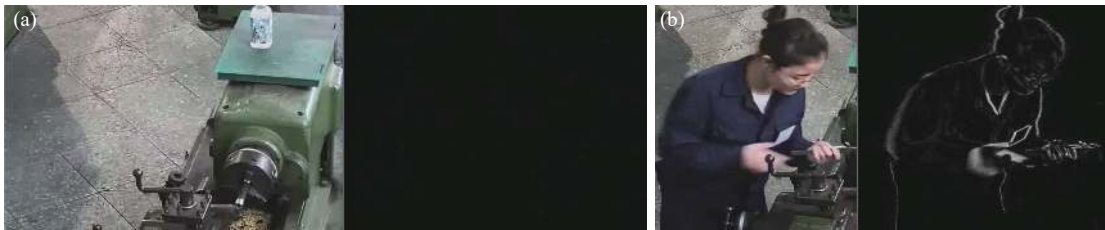


图6 运动图像检测灰度图。(a)无运动目标时的原图/灰度图;(b)运动目标出现时的原图/灰度图

Fig.6 Gray level image of moving image detection: (a) original image / gray level image without moving object; (b) original image / gray level image when moving object appears

1.3.2 STC 匹配校正

为了解决复杂环境、头发遮挡等对头部跟踪的影响, 在深度学习框架后的滤波跟踪框架中融合了图像空间上下文信息, 以实现目标变化的自适应调整。STC 图像空间上下文算法通过贝叶斯框架统计目标及其背景的相关性并进行置信图计算, 根据置信图的似然概率最大位置预测新的目标位置。

根据当前帧确定上下文特征集合: $T^c = \{c(z) = (I(z), z) | z \in \Omega_c(t^*)\}$, 其中, $t^*(x_t, y_t)$ 是目标区域中心, $I(z)$ 是目标区域像素 z 的灰度值, $\Omega_c(t^*)$ 是

由目标矩形确定的局部上下文区域的图像灰度与位置的统计建模, $c(z)$ 为置信图函数^[19]。如图 7 所示, 对灰度级图像进行阈值化处理, 得到的二值化图像可进一步排除背景冗余信息。

设 SiamMask 模型第 k 帧追踪目标图像为 P_k , 空间上下文预测第 k 帧目标图像为 P'_k 。设计判断函数 Φ 进行 P_k 与 P'_k 中目标位置的相似度比对, 根据比对结果进行更新校正, 形成新的目标跟踪结果。

步骤 1: 计算 $k-1$ 帧 ROI 区域置信图:

$$c_{k-1}(t) = \sum_{c(z) \in T^c} P(t|c(z), o)P(c(z)|o) \quad (2)$$



图7 图像灰度及阈值化处理。(a)原始视频图像;(b)灰度化处理结果;(c)阈值化处理结果

Fig.7 Gray level and threshold processing of image: (a) original video image; (b) grayscale processing results; (c) threshold processing results

式中, o 表示所跟踪目标, P 为上下文先验模型.

$$P(c(z)|o) = I_{k-1}(z)\omega_\sigma(z-t^*) = h_{k-1}^{sc}(t-z) \quad (3)$$

式中, $\omega_\sigma(z) = a \cdot e^{-z^2/\sigma^2}$ 是一个权重函数, 其中, a 是归一化参数, 取值为 $[0,1]$, σ 是一个尺度参数, σ^2 为高斯函数方差. $h^{sc}(t-z)$ 是时空上下文模型, 是傅里叶变换后的频率域计算变形.

步骤 2: 计算 $k-1$ 帧 $\Omega_c(t^*)$ 上下文区域的空间上下文模型:

$$h_{k-1}^{sc}(t) = F^{-1}\left(\frac{F(b \cdot e^{-|t-t^*|^\alpha})^\beta}{F(I(t)\omega_\sigma(t-t^*))}\right) \quad (4)$$

式中, b 是归一化参数, α 是尺度参数, β 是目标形状参数.

步骤 3: 更新空间上下文 $h^{sc}(x)$ 模型:

$$H_2^{sc}(t) = h_1^{sc}(t) \quad (5)$$

$$H_k^{sc}(t) = (1-\rho)H_{k-1}^{sc}(t) + \rho h_{k-1}^{sc}(t) \quad (6)$$

式中, ρ 为模型更新的学习率.

步骤 4: 在第 k 帧计算上下文先验模型及置信图:

$$P(c(z)|o) = I_k(z)\omega_\sigma(z-t_{k-1}^*) \quad (7)$$

$$c_k(t) = F^{-1}(F(H_k^{sc}(t)) \odot I_k(t)\omega_\sigma(z-t_{k-1}^*)) \quad (8)$$

步骤 5: 将第 k 帧得到的置信图极值点作为目标在 k 帧的位置输出:

$$t_k^* = \arg \max_{x \in \Omega(t^*)} c_k(t) \quad (9)$$

将 t_k^* 与 SiamMask 模型该帧目标位置 t_k 对比. 设计判断函数 $\Phi = |(t_k^* - t_k)/(t_{k-1} - t_{k-2})|$.

若 $\Phi > 1$, 则第 k 帧目标图像输出为 P_k' , 第 k 帧的目标位置输出更新为 t_k^* ; 若 $\Phi < 1$, 则追踪目标图像为 P_k , 第 k 帧的目标位置输出更新为 t_k .

2 算法研究结果及分析

2.1 算法训练

与 SiamMask^[15] 一样, 使用 Warmup 预热学习率的方式, 在 Davis2017^[20] 数据集上进行训练, 并在机加工操作数据视频集中进行头部识别测试. 其中, 训练输入为图像尺寸 127 像素 \times 127 像素, 255 像素 \times 255 像素, 数据增强采用随机抖动与搜索策略, 网络初始化权重采用 SiamMask 模型在 ImageNet-1k Datasets 上预训练的权重, 梯度下降策略采用 SGD. 训练时在开始进行的 5 轮训练中, 先将学习率从 10^{-3} 线性增长到 5×10^{-3} , 之后通过 15 轮训练将学习率以对数规律降为 5×10^{-4} .

图 8 为本文算法在训练及测试集中的准确率、损失率曲线图. 可以看到, 在训练集上, 15 轮次左右训练后, 准确率及损失率均有效收敛. 在测试集上, 由于数据特征更为集中, 呈现了更快速的收敛性.

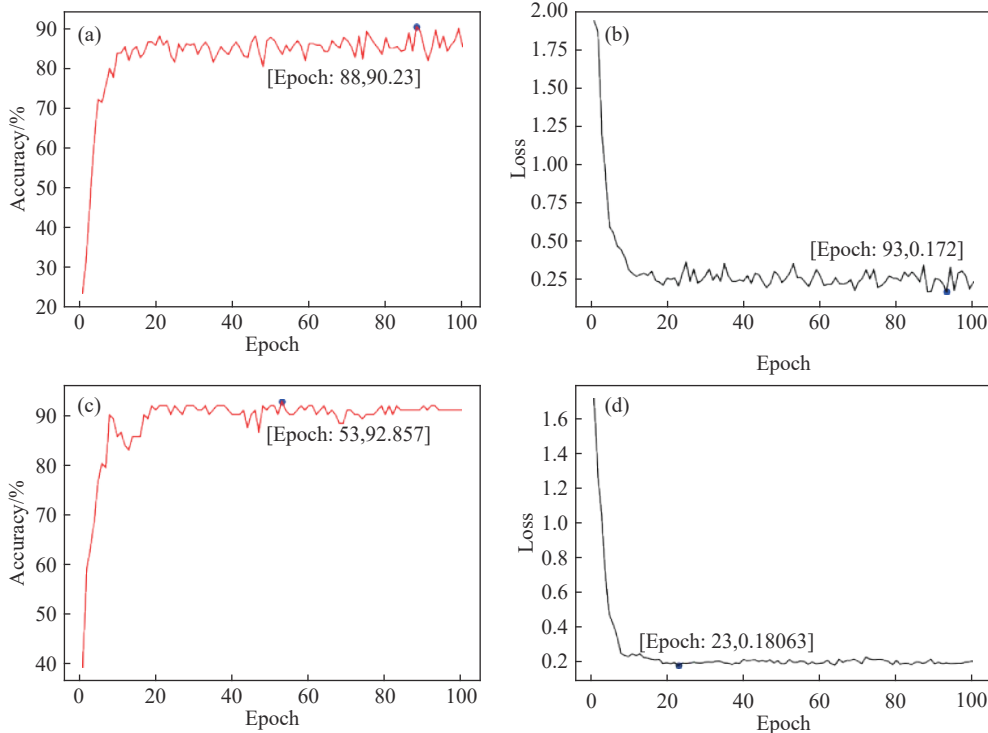


图 8 算法训练/测试准确率及损失率曲线. (a) 训练集准确率曲线图; (b) 训练集损失率曲线图; (c) 测试集准确率曲线图; (d) 测试集损失率曲线图
Fig.8 Algorithm training/test accuracy and loss rate curve: (a) training set accuracy curve; (b) training set loss curve; (c) test set accuracy curve; (d) test set loss curve

2.2 算法应用效果测试

2.2.1 测试指标设计

本文采用图像检测中常用的标准性能度量参数 IoU (Intersection over union) 进行检测判断. 当检测到的目标位置与真实目标位置相交的重叠部分为 0 时为失败, 即 $\text{IoU}=0$ 认定为跟踪失败. 通过式 (10) 统计第 i 段视频误检帧所占的比例.

$$Q_i = \frac{1}{M_s} \sum_{k=1}^{M_s} \Phi_k \quad (10)$$

式中, M_s 为视频总帧数; Φ_k 为判断因子, 当该帧中 $\text{IoU}=0$ 时, $\Phi_k = 1$, 其余情况 $\Phi_k = 0$.

$$\text{设计统计指标: } Q = \frac{1}{M} \sum_{i=1}^M Q_i \quad (11)$$

式中, M 为视频个数. Q 为所有视频片段的错误率的均值.

2.2.2 测试数据对比

为保证对比测试的初始区域一致性, SiamMask 模型与本文所提算法均采用被测视频的初始帧人脸区域分割结果. 通过对 7 段测试视频进行统计, 得到前文表 1 及下文表 2 的数值.

从表 1 中视频的测试数据可计算得知 $Q=34.29\%$, 即算法误识别的平均可能性为 34.29%. 从表 2 数据可计算得知 $Q=0.156\%$, 本文提出的算法误识别率有较大降低.

2.3 算法时间成本

由于本文算法必须满足智能安全识别系统实时、高效、可靠的要求, 因此在实际使用环境中测试其综合时间成本.

为了检验效果, 用外接矩形框对目标跟踪轮廓边界进行了包围, 如图 9(a) 所示. 可以看到, 经过算法校正, 背景中的噪声轮廓得到了大幅度去除, 保留得到的头部轮廓是大面积连通域的外轮廓. 对于头发轮廓的边界点坐标以向量的形式进行存储, 即获取到视频序列中人员头发的位置坐标, 为后续头发目标是否进入危险区的判定提供基础.

如图 9(a) 中所示, 较大蓝色矩形框代表二级危险区, 较小绿色矩形框代表一级危险区. 对于危险区的位置信息记录如下, 二级危险区的具体设计为: 左上角顶点像素位置坐标为 (418, 274), 宽度为 164 个像素, 高度为 147 个像素. 一级危险区的具体设计为: 左上角顶点像素位置坐标为 (471, 311), 宽度为 107 个像素, 高度为 93 个像素. 在图 9(b) 中设计了四种报警级别的区域范围, 分别是预警 II 级、预警 I 级、危险、非常危险.

对于两组坐标进行比对判别, 判断头发目标进入危险区的程度以触发不同类别的报警. 得到结果如图 10 所示, 图 10(b) 左上角为具体文字报警内容.

表 2 基于 SiamMask 模型的时空预测算法目标跟踪效果统计

Table 2 Statistics of the target tracking effect of the spatiotemporal prediction algorithms based on the SiamMask model

Video No.	Frame number of false detection	Analysis on the causes of false inspection	Total frames	Failure rate/%
1	0	Little change in this movement	361	0
2	0	Misidentified as dark cloth	288	0
3	1	Part of the face is blocked by the hair	192	0.52
4	2	Initialization offset, screen will pop up in recognition	1380	0.15
5	1	The target moves out of the screen slightly and the recognition is lost	240	0.42
6	0	Large proportion of face selection in initialization area	1360	0
7	0	Accurate initialization and small action range	241	0

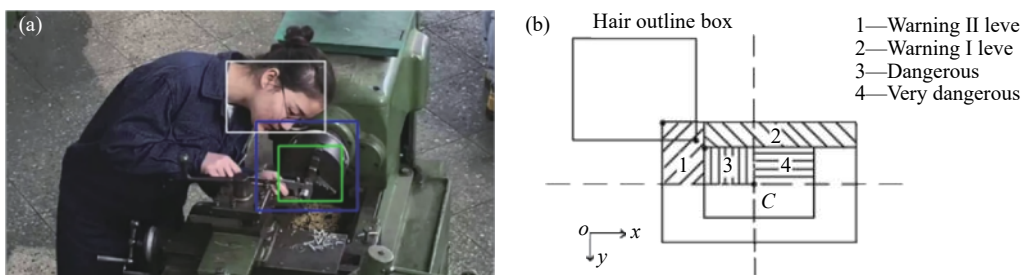


图 9 固定危险区划分示意图. (a) 视频危险区划分图; (b) 危险级别划分示意图

Fig.9 Fixed danger zone division diagram: (a) video dangerous zone division map; (b) diagram of hazard classification

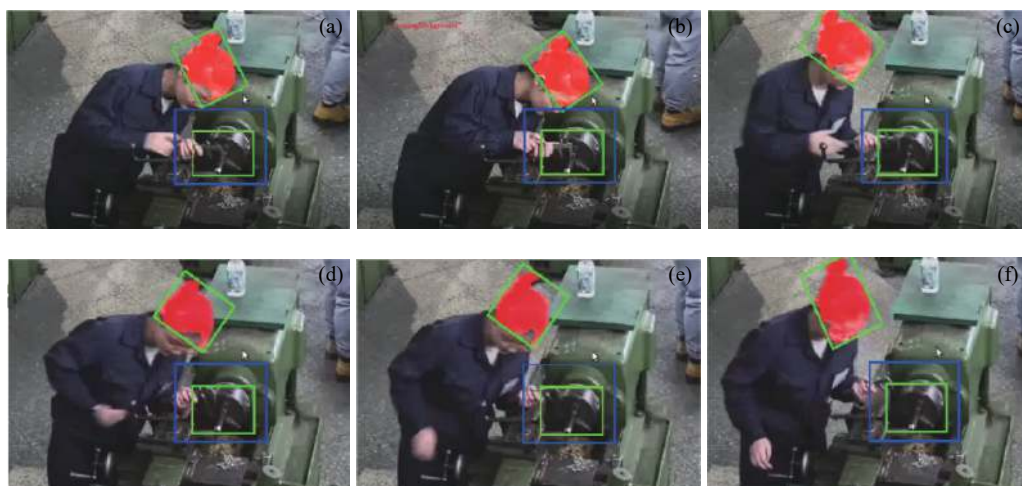


图 10 头发目标跟踪报警结果. (a)第 30 帧; (b)第 35 帧; (c)第 60 帧; (d)第 70 帧; (e)第 75 帧; (f)第 80 帧

Fig.10 Hair target tracking alarm results: (a) frame 30; (b) frame 35; (c) frame 60; (d) frame 70; (e) frame 75; (f) frame 80

对于算法的处理时间进行统计. 利用 `cvGetTickCount` 函数得到从操作系统启动到当前所经历的时钟周期数, 利用 `cvGetTickFrequency` 函数返回每秒的计时周期数 N . 则程序运行时间 T 可以通过式 (12) 计算得到.

$$\tau = \frac{n_r - n_s}{N} \quad (12)$$

式中, n_r 为当前时钟周期数, n_s 为开始时钟周期数.

测试视频的基本信息为: 宽度 816 像素, 高度 536 像素, 总帧数 3426 帧, 帧率为每秒 30 帧, 图像处理速度为每秒 30 帧. 按照式 (12) 计算, 处理时间为 0.137 ms.

算法的计算效率比改进前的 SiamMask 模型图像处理速度每秒提高 3.2 帧, 算法效率提高约 11.94%, 可以满足智能安全预警系统的实时快速性要求.

3 结论

(1) 基于 PyTorch 深度学习框架的 SiamMask 模型单目标跟踪算法, 引入 ROI 检测及 STC 时空上下文目标跟踪算法, 结合时间与空间信息, 根据对头发目标时空关系的在线学习预测新的目标位置, 对 SiamMask 模型进行了算法校正.

(2) 提出的算法应用于复杂环境中头发目标视频图像识别. 研究表明: 所提算法能够克服环境干扰、目标遮挡对跟踪效果的影响, 将目标跟踪误识别率降低至 0.156%.

(3) 其计算时间成本为每秒 30 帧, 比改进前的 SiamMask 模型帧率每秒提高 3.2 帧, 算法效率提高约 11.94%, 可以满足智能安全预警系统的实时快速性要求.

参 考 文 献

- [1] Xing J L, Ai H Z, Lao S H. Multiple human tracking based on multi-view upper-body detection and discriminative learning // 2010 20th International Conference on Pattern Recognition. Istanbul, 2010: 1698
- [2] Liu L W, Xing J L, Ai H Z, et al. Hand posture recognition using finger geometric feature // Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012). Tsukuba, 2012: 565
- [3] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional siamese networks for object tracking // European Conference on Computer Vision. Cham: Springer, 2016: 850
- [4] Li B, Yan J J, Wu W, et al. High performance visual tracking with siamese region proposal network // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 8971
- [5] Wang Q, Zhang L, Bertinetto L, et al. Fast online object tracking and segmentation: a unifying approach // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 1328
- [6] Li J Y, Zhao Y K, Xue Z E, et al. A survey of model compression for deep neural networks. *Chin J Eng*, 2019, 41(10): 1229
(李江昀, 赵义凯, 薛卓尔, 等. 深度神经网络模型压缩综述. 工程科学学报, 2019, 41(10): 1229)
- [7] Shen L L, Hang N. No-reference image quality assessment using joint multiple edge detection. *Chin J Eng*, 2018, 40(8): 996
(沈丽丽, 杭宁. 联合多种边缘检测算子的无参考质量评价算法. 工程科学学报, 2018, 40(8): 996)
- [8] Wang X, Tang Z M. An improved camshaft-based particle filter algorithm for real-time target tracking. *J Image Graph*, 2010, 15(10): 1507
(王鑫, 王振民. 一种改进的基于 Camshift 的粒子滤波实时目标跟踪算法. 中国图像图形学报, 2010, 15(10): 1507)
- [9] Chu H X, Xie Z Y, Wang K J. An improved camshift target

- tracking algorithm based on joint color-texture histogram. *J Xi'an Jiaotong Univ*, 2018, 52(3): 145
(初红霞, 谢忠玉, 王科俊. 一种结合颜色纹理直方图的改进型 Camshift 目标跟踪算法. 西安交通大学学报, 2018, 52(3): 145)
- [10] Yacoob Y, Davis L S. Detection and analysis of hair. *IEEE Trans Pattern Anal Mach Intell*, 2006, 28(7): 1164
- [11] Fu W L, Hu F Q. Hair detection with mean shift. *Microcomput Appl*, 2010, 26(9): 62
(傅文林, 胡福乔. 基于 mean shift 的头发自动检测. 微型电脑应用, 2010, 26(9): 62)
- [12] Wang Z Y, Yang D L. Research on detection method of hair-occlusion in face recognition. *Microcomput Its Appl*, 2016, 35(2): 32
(王志一, 杨大利. 人脸识别中发型遮挡检测方法研究. 微型机与应用, 2016, 35(2): 32)
- [13] Ding C J, Yan B. Target tracking algorithm combines STC with CamShift. *Transduc Microsyst Technol*, 2018, 37(5): 108
(丁承君, 闫彬. 时空上下文与 CamShift 相结合的目标跟踪算法. 传感器与微系统, 2018, 37(5): 108)
- [14] Eriksson D, Bindel D, Shoemaker C. *Surrogate Optimization Toolbox Hithub* (2019-1-5)[2019-6-5]. <https://github.com/dme65/pySOT>
- [15] Qiang Wang, Li Zhang, Luca Bertinetto, et al. Torr. *SiamMask Hithub* (2019-5-5)[2019-6-5]. <https://github.com/foolwood/Siam-Mask>
- [16] Zhang L, Suganthan P N. Robust visual tracking via co-trained kernelized correlation filters. *Pattern Recognit*, 2017, 69: 82
- [17] Chen C, Wu Z X, Jiang Y G. Emotion in context: Deep semantic feature fusion for video emotion recognition // *Proceedings of the 24th ACM International Conference on Multimedia*. New York, 2016: 127
- [18] Huang Y, Huang X, Li S G, et al. Compensation method for a robot vision system with an occluded camera field. *Chin J Eng*, 2018, 40(3): 381
(黄煜, 黄翔, 李沈杲, 等. 存在视场丢失的机器视觉精度补偿方法. 工程科学学报, 2018, 40(3): 381)
- [19] Zhang J, Wang X, Fan H B. Spatio temporal context target tracking algorithm of self-adaption learning. *Comput Eng*, 2018, 44(6): 294
(张晶, 王旭, 范洪博. 自适应学习的时空上下文目标跟踪算法. 计算机工程, 2018, 44(6): 294)
- [20] Pont-Tuset J, Perazzi F, Caelles S, et al. *The 2017 Davis Challenge on Video Object Segmentation* (2017-4-3)[2019-6-5]. <https://arxiv.org/abs/1704.00675>