

doi:10.13756/j.gtxyj.2020.02.002

光通信系统与网络技术

基于深度学习的屏幕通信定位跟踪算法研究

刘文楷,王海瑞,武梦龙

(北方工业大学 信息学院,北京 100144)

摘要:屏幕通信是以动态条码为信息载体,以可见光为媒介的近场通信方式,因其抗干扰能力强、不占用频谱资源和部署简单等特点,具有广泛的应用场景和极大的发展潜力。文章提出了一种基于深度学习方法的屏幕通信定位跟踪算法,采用快速区域卷积神经网络(Faster R-CNN)算法对信息区域进行处理,使光学摄像头在不依赖传统定位寻像图形的情况下智能定位携信区域,进而提高单帧携信量;采用卢卡斯-卡纳德(LK)光流法对抖动引入的帧间信息区域位置变化进行估计,提升了连续帧处理速率从而提升了系统通信速率。实验结果表明,Faster R-CNN算法的平均精度均值(mAP)达到了90.91%,同时引入LK光流法提升系统的处理效率,相比于仅采用Faster R-CNN的算法处理时间缩短59.5%以上,解决了终端系统处理能力瓶颈问题。

关键词:屏幕通信;深度学习;信息条码;定位跟踪算法

中图分类号:O436

文献标志码:A

文章编号:1005-8788(2020)02-0007-04

Investigation on Screen Communication Location and Tracking Algorithm based on Faster R-CNN

LIU Wen-kai, WANG Hai-rui, WU Meng-long

(School of Information, North China University of Technology, Beijing 100144, China)

Abstract:Screen communication is a near-field communication mode using dynamic bar code as the information carrier and visible light as the media. Due to its strong anti-interference ability, spectrum resources free, and simple deployment, it has a wide range of application scenarios and great development potential. We have proposed a screen communication location tracking algorithm based on the method of deep learning. The information area is processed by Faster Region with Convolutional Neural Networks Feature (Faster R-CNN) algorithm which makes the optical camera can intelligently locate the carrier area without relying on the traditional positioning finder pattern, resulting in the improvement of the single frame carrying. The Lucas-Kanade (LK) optical flow is used for position change estimation of the subsequent frame information region introduced by the jitter, which improves the continuous frame processing rate and the system communication rate. Experimental results show that the mean Average Precision (mAP) of the Faster R-CNN is reached up to 90.91% while the LK optical flow was introduced to improve the processing efficiency of the system, and overcome the bottleneck of the processing power of the terminal system. The processing time was shortened by 59.5% compared with the Faster R-CNN algorithm.

Key words: screen communication; deep learning; information bar code; position tracking algorithm

0 引言

随着硬件计算能力的快速增长,深度学习相关技术取得了长足的进步^[1-4],其被应用于图像处理领域解决了很多疑难问题,这使将深度学习应用于可见光屏幕通信成为可能。近年来国内外可见光屏幕通信研究主要集中于通信链路、多种形式的编码以及非实时的解析方式等方面,彩色和灰度等编码方式虽然提高了信息携带量,但增加了接收单元的识别解析难度以及降低了屏幕通信系统的可靠性^[5-8]。当前可见光屏幕通信主流传输方案采用寻找定位图形的方法来定位信息区域,而定位图形的大小也在一定程度上影响了带宽^[8-9]。同时,在实际

应用中收发两端产生晃动会导致信息无法正确接收,所以大多实验均固定收发两端以减少图像失真^[10]。以上问题导致可见光屏幕通信的实际应用效果并不理想。

针对当前可见光屏幕通信系统可靠性不高、通信速率较低和单帧携信量受限等问题,本文采用快速区域卷积神经网络(Faster Region with Convolutional Neural Networks, Faster R-CNN)算法对信息区域进行处理,使光学摄像头在不依赖传统定位寻像图形的情况下智能定位携信区域,进而提高单帧携信量,同时采用卢卡斯-卡纳德(Lucas-Kanade, LK)光流法对抖动引入的帧间信息区域位置变化进行估计,提高了连续帧处理速率从而提高了

收稿日期:2019-08-14

基金项目:国家自然科学基金资助项目(61675138)

作者简介:刘文楷(1968—),男,河北保定人。博士,教授,主要研究方向为光电集成与无线光通信。

通信作者:武梦龙,副教授。E-mail: wumenglong@126.com

系统通信速率。实验结果表明，Faster R-CNN 平均精度均值 (Mean Average Precision, mAP) 达到了 90.91%，同时引入 LK 光流法提升系统的处理效率，克服了终端系统处理能力瓶颈，相较于仅采用 Faster R-CNN 算法，处理时间缩短了 59.5% 以上。

1 可见光屏幕通信系统概述

本文涉及可见光屏幕通信系统中接收单元对 Faster R-CNN 算法和 LK 光流法的引入，其中 Faster R-CNN 算法的引入使接收单元可不依赖于定位寻像图形对信息区域进行定位，同时可设计占用较少或不包含定位图形的条码以达到提升单帧携信量的目的；跟踪算法可减少系统计算开销，解决了 Faster R-CNN 算法处理连续帧图像所需时间较长的问题。

本文所涉及的可见光屏幕通信系统框架如图 1 所示。系统分为发送单元与接收单元，发送单元首先将数据根据任意编码规则生成对应条码并通过屏幕显示；接收单元通过光学摄像头捕获信号，将生成的视频进行帧拆分，采用 Faster R-CNN 算法确定图像中信息区域的位置，并采用跟踪算法对后续帧信息区域的位置进行跟踪，最后进行解码和写入文件。

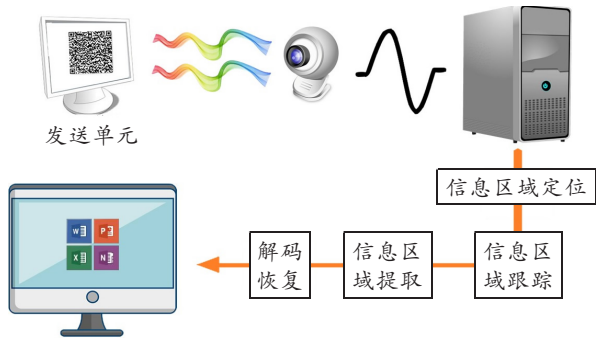


图 1 可见光屏幕通信系统框架图

2 屏幕通信定位跟踪算法

2.1 基于 Faster R-CNN 的条码定位算法

现有可见光屏幕通信系统接收方案大多依靠定位寻像图形进行定位，如图 2 所示，此类图形占据条码面积，导致条码可携带信息量受到了限制，从而限制了系统性能，且接收单元需要与之对应的解码器进行解码，普适性较低。通过对神经网络进行训练，使接收单元可不依赖于定位寻像图形自适应地识别多种类型条码并进行定位，并可通过设计包

含较少或不含定位图形的条码以达到提升单帧携信量的目的，根据条码类型的不同，携信量提升最高可达 30%。

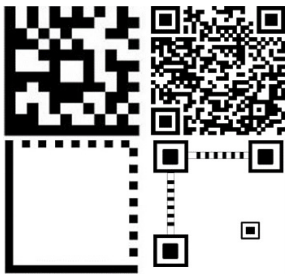


图 2 常见条码及定位寻像图形

为提高可见光屏幕通信系统单帧的信息携带量以及适用性，接收单元引入了 Faster R-CNN 算法，通过训练卷积神经网络使接收单元学习棋盘格图像特征，从而达到在尽可能去除非有效信息的定位点和校正点后可以智能提取出棋盘格图像的要求。Faster R-CNN 算法的主要结构是区域建议网络 (Region Proposal Network, RPN) 和特征提取网络，如图 3 所示。

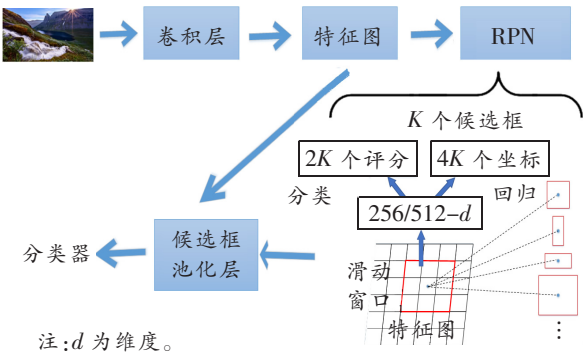


图 3 Faster R-CNN 算法结构图

首先将图像输入卷积神经网络进行特征提取，用 RPN 生成候选框，将候选框映射到特征图上，通过候选框池化层使每个候选框生成固定尺寸的特征图，将特征图进行全连接操作，之后利用分类函数 Softmax 层进行分类，再利用回归函数 Smooth L1 完成边框回归任务，从而提升预测位置精度。

RPN 是一个基于完全卷积网络的深度学习网络，用于生成高质量的候选框，它与检测网络共享整个图形的卷积特征，解决了选择性搜索 (Selective Search) 的速度问题，极大地提高了对象检测的效果。在速度与精度上可以初步满足可见光屏幕通信系统的需要。

RPN 采用一个 $n \times n$ 的滑动窗口在卷积特征图上进行滑动并生成 512 维特征向量，然后分别输入到分类层和回归层以获得分类信息和位置信息。

在每个滑动窗口位置，同时预测多个候选框，其中每个位置的最大可能建议窗口数为 K ，所以回归层输出 $4K$ 个向量来编码 K 个候选框坐标 $(x, y, \text{width}, \text{height})$ (中心点坐标及候选框的宽高)，分类层输出 $2K$ 个向量来判别每个候选框是否是目标。锚点位于滑动窗口中，默认使用 3 个尺度和 3 个横纵比，在每个滑动位置产生 $K=9$ 个锚点，对于 $W \times H$ 的卷积特征图，共有 $W \times H \times K$ 个锚点， W 和 H 分别为卷积图的宽和高。

滑动窗口和锚点的组合可以基本把条码可能出现的区域全部覆盖，从而达到较为精确的定位效果，满足了在去除传统定位点的同时要求准确定位的需求。

Faster R-CNN 的损失函数由式(1) [1] 给出：

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{\text{cls}}} \sum_i L_{\text{cls}}(p_i, p_i^*) + \lambda \frac{1}{N_{\text{reg}}} \sum_i p_i^* L_{\text{reg}}(t_i, t_i^*), \quad (1)$$

式中： i 为锚点的索引； p_i 为该锚点是目标的概率，若锚点为正则 $p_i=1$ ，若锚点为负则 $p_i=0$ ； N_{cls} 为小批量样本数量； $L_{\text{cls}}(p_i, p_i^*)$ 为两个类别的对数损失； p_i^* 为真实标签，若锚点为正则 $p_i^*=1$ ，锚点为负则 $p_i^*=0$ ； λ 为平衡参数； N_{reg} 为锚点的数量； t_i 为预测框的坐标向量； t_i^* 为与标签样本对应的坐标向量； $L_{\text{reg}}(t_i, t_i^*)$ 为回归损失， $L_{\text{reg}}(t_i, t_i^*) = R(t_i, t_i^*)$

且 $R(x) = \begin{cases} 0.5x^2 & |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases}$ ， R 为回归函数的 loss 函数。

以不含定位图形的普通棋盘格图形为例，使用 Faster R-CNN 算法可以较准确地定位单帧图片中棋盘格所在区域，因棋盘格相对背景特征明显，所以识别提取效果较好，效果图如图 4 所示。图中棋盘格所在区域为信息区域，其他为背景区域，红色边框表示找到的信息区域为定义的 code 类的概率为 1.0。

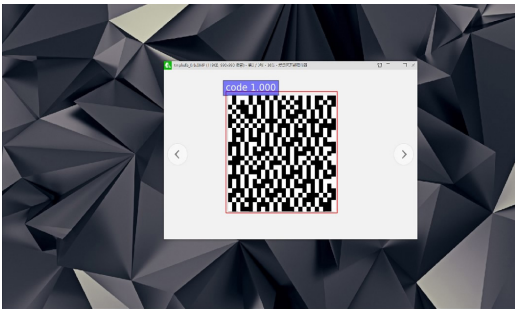


图 4 Faster R-CNN 算法定位棋盘格效果图

2.2 采用 LK 光流法的连续帧条码跟踪算法

可见光屏幕通信系统的处理能力由于硬件平台的不同而有所差异，由于采用 Faster R-CNN 算法对信息区域定位的计算量较大，现有的大多数设备处理能力无法满足实时处理的实际要求，因此引入 LK 光流法应对接收单元与发送单元之间产生的小运动导致的信息区域位置变化，减少计算开销并提升系统速率，克服系统性能瓶颈。

可见光屏幕通信系统应用场景符合 LK 光流法的 3 个条件：亮度恒定、运动相对缓慢和区域性一致，如式(2)所示：

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t), \quad (2)$$

式中： I 为坐标为 (x, y) 的点在 t 时刻的亮度，因亮度恒定，所以在变化 δx 、 δy 和 δt 后亮度相等。将式(2)泰勒展开：

$$I(x + \delta x, y + \delta y, t + \delta t) = I(x, y, t) + \frac{\partial I}{\partial x} \delta x + \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial t} \delta t, \quad (3)$$

化简即

$$\frac{\partial I}{\partial x} V_x + \frac{\partial I}{\partial y} V_y + \frac{\partial I}{\partial t} = 0, \quad (4)$$

即可得到某个点的速度矢量(式中， V_x 和 V_y 分别为该点横、纵坐标轴上的速度矢量)，因区域一致性，所以只需要几个点即可判断信息区域整体的运动情况。

使用 LK 光流法对帧中的信息区域位置进行跟踪，因其计算量小，所以速度相对较快，可达到 30 FPS，同时精度较高，相比于仅采用 Faster R-CNN 算法可大幅提升终端系统处理效率。同时为提升定位跟踪准确度，保证后续信息被正确恢复，可采用穿插定位跟踪的方法，即一帧由 Faster R-CNN 算法进行定位，后续几帧由 LK 光流法跟踪，再由 Faster R-CNN 算法进行位定。

3 实验结果与分析

本文对所设计方法进行了实验验证，实验所用设备操作系统为 Windows 10，中央处理器(Central Processing Unit, CPU)为 CORE i7 6700HQ，图形处理器(Graphics Processing Unit, GPU)为 GTX 1060 3G。

引入 Faster R-CNN 算法可较为精确地定位信息区域地位置，为解码器降低了计算量。本文使用 VGG16 网络进行特征提取，定义一个 code 类，通过将彩色条码与无关背景合成来产生训练图像，为

防止过拟合, 迭代 4 940 次, 其损失函数收敛曲线如图 5 所示。迭代 4 940 次, 训练损失最终降至 0.03, mAP 达到了 90.91%, 表明 Faster R-CNN 算法的引入使得接收单元的 mAP 达到了 90.91%, 可以保证在传输中去掉纠错以及校正棋盘格后的系统稳定性。

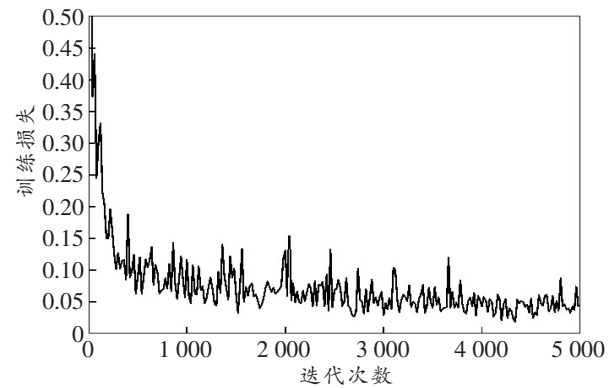


图 5 损失函数收敛曲线

使用 LK 光流法可以较为精确地对信息区域进行跟踪, 同时速度较快。图 6 所示为 LK 光流法运动轨迹图, 如图所示, 彩色的线条表示信息区域的运动路径, 在由 Faster R-CNN 算法精准定位的结果下, 可以对其进行跟踪, 而无需消耗大量的计算资源, 提升了系统的通信速率及稳定性。

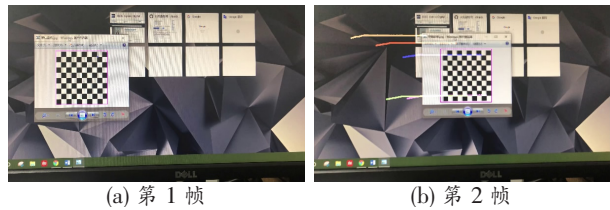


图 6 LK 光流法运动轨迹

经实验验证, 采用 LK 光流法进行位置估计每帧需要的时间约为 0.032 s。如图 7 所示, 相比于只采用 Faster R-CNN 算法, 引入 LK 光流法可以大幅提升系统的处理效率。仅采用 Faster R-CNN 算法处理 49 帧图像所消耗的时间为 5.089 s。为避免

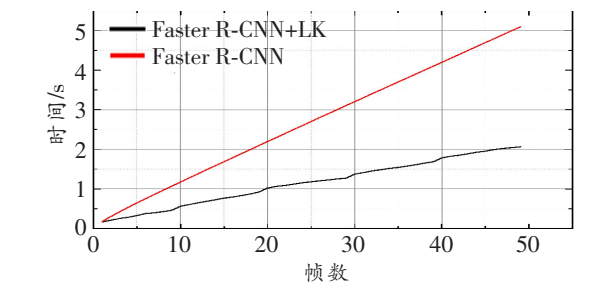


图 7 处理效率对比

误差叠加, 在本实验中, 每 10 帧使用一次 Faster R-CNN 算法确定当前帧的信息区域位置, 后续 9 帧采用 LK 光流法对信息区域进行跟踪, 处理 49 帧图像所消耗的时间为 2.063 s, 处理效率提升了 59.5%, 由此可得出结论: 采用 Faster R-CNN 算法与 LK 光流法相结合的处理方式可以有效应对接收单元处理性能瓶颈。

4 结束语

本文针对可见光屏幕通信系统的接收单元, 设计了引入 Faster R-CNN 算法和 LK 光流法的定位跟踪方法。实验结果表明, Faster R-CNN 算法的引入可使接收单元在不依赖定位寻像图形的情况下较为精确地定位信息区域, mAP 达到了 90.91%; 引入 LK 光流法提升了系统的处理效率, 有效克服了接收单元处理能力瓶颈, 处理速率可达 21 FPS。未来可通过引入结构更精准快速的深度学习框架以及设计更加高效的编码方案和包含较少定位图像的编码方式来进一步提高通信速率。

参考文献:

[1] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137—1149.

[2] 刘文楷, 徐一鸣, 武梦龙. 基于 CNN 的可见光屏幕通信识别与解析机制[J]. 光通信研究, 2018(6): 42—46.

[3] 张明江, 李红卫, 赵卫虎, 等. 深度学习在军用光缆线路无人机巡检中的应用[J]. 光通信研究, 2018(6): 57—61.

[4] 谢智歌, 王岳青, 窦勇, 等. 基于卷积—自动编码机的三维形状特征学习[J]. 计算机辅助设计与图形学学报, 2015, 27(11): 46—52.

[5] 姜珊, 胡贵军, 王海燕, 等. 基于彩色二维条码流的 Display-to-Camera 通信系统[J]. 电子学报, 2016, 44(4): 767—774.

[6] Yamsang P, Pongyart W, Vanichchanunt P. Multilevel Grey-Scale Visual Coding for Data Transmission from LCD Screens to Android Smartphone Cameras [C]//2017 8th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES). Chonburi, Thailand: IEEE, 2017: 1—4.

(下转第 60 页)