

# 融合目标检测和人体关键点检测的 铁路司机行为识别

姚巍巍, 张洁

(西南交通大学 机械工程学院, 成都 610031)

**摘要:** 随着我国经济的快速发展, 铁路运输在交通运输的地位愈为重要, 在传统人工监管无力应对铁路司机安全监督的情况下, 使用机器实现自动实时司机行为识别早已成为了一项极有意义的工作; 为实现随车部署、实时进行铁路司机行为识别的目的, 基于目标框检测算法实现目标检测和关键点检测的融合, 搭建了一种可以同时检测司机人体关键点和手机的神经网络; 经过网络运行输出人体姿态后, 通过分析人体各关节角度和人体关键点与手机目标的位置关系等后处理对 6 类司机行为进行了分类识别, 并通过 TensorRT 框架对模型进行了模型推理速度的加速和体积上的压缩; 实验表明, 该模型在嵌入式设备 TX2 上推理速度为 25 ms, 可以达到较好检测效果下实时运行的目标; 实现了实时进行铁路司机行为识别的目的。

**关键词:** 深度学习; 目标检测; 关键点检测; 半精度加速

## Railway Driver Behavior Recognition Based on Fusion Object Detection and Person Keypoints Detection

Yao Weiwei, Zhang Jie

(School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China)

**Abstract:** With the rapid development of China's economy, the role of railway transportation in transportation becomes more important. In the case that traditional manual supervision is unable to cope with the safety supervision of railway drivers, using machines to realize automatic real-time driver behavior recognition has already become a very meaningful task. In order to realize the real-time railway driver behavior recognition on the embedded device, a neural network that can simultaneously detect key points of the driver's human body and mobile phones is constructed based on object detection and person keypoints detection. Six types of driver behavior are identified by post processing of analyze the relationship between the joint angles of the human body and the key points of the human and the target of the mobile phone. And the model was accelerated and compressed for operation on embedded devices through TensorRT framework. Experiments show that the inference time of model is 25 ms on the embedded device TX2, which can achieve the goal of better accuracy and real-time operation. The purpose of real-time identification of railway driver behavior was achieved.

**Keywords:** deep learning; object detection; person keypoints detection; half-precision acceleration

### 0 引言

随着我国经济高速全面发展, 铁路运输能力的快速提高, 铁路运输在交通运输的地位愈为重要。确保铁路安全平稳成为铁路运输部门的一项重要工作。在众多的铁路安全事故中, 人始终是影响铁路行车安全的重要因素。文献[1]总结了 1990~2009 年 10 年中重大列车事故原因, 其中人为因素占总事故数目的 64.76%。因此, 对铁路系统中的人加强监管就显得极为重要, 尤其是直接关系到整车安全的机车驾驶员。

当前, 我国铁路管理部门在原有安全监测设备基础上

重新设计, 构建了机车车载系统、地面专家系统一体化的机车安全防护体系(6A 系统)<sup>[2]</sup>。这套系统在每辆机车的机车驾驶室后方安装了摄像头以加强对驾驶员的监督。当列车开始行驶后, 摄像头会开启并记录驾驶室人员视频信息。当列车停止运行后, 铁路管理人员会以人工浏览察看视频的方式进行分析, 用于司机的绩效评定和事故后责任的分析。可见, 当前的驾驶室视频系统只能做到对驾驶员事先的心理提醒, 无法对驾驶员在机车运行过程中不当行为进行实时提醒。随着深度学习在图像处理方面的发展应用, 使用计算机实现自动司机行为识别已经成为了可能。搭建一种车载智能监控系统, 实现实时司机行为识别, 如检测出问题便及时将图片传回地面铁路管理人员, 便成为一项极有意义的工作。

由此, 本文融合目标检测和关键点检测两种深度学习算法, 创新实现了一种可以同时检测司机人体关键点和手机的神经网络, 通过分析人体关节角度和人体关键点与手机目标的位置关系对司机行为进行了分类识别, 并利用半

**收稿日期:** 2019-11-20; **修回日期:** 2019-12-09。

**基金项目:** 国家自然科学基金项目(51775449)。

**作者简介:** 姚巍巍(1994-), 男, 山东泰安人, 硕士研究生, 主要从事机器视觉方向的研究。

**通讯作者:** 张洁(1975-), 女, 重庆人, 博士, 教授, 博士生导师, 主要从事智能控制、机器视觉方向的研究。

精度加速和层级融合的方法对模型进行了加速和压缩, 最终将其成功的部署在嵌入式边缘计算设备上, 实现了较好检测效果下的实时运行。

## 1 数据采集

本文将需要检测的司机行为分为以下几类: 1) 离岗: 驾驶室内未检测到司机; 2) 使用手机: 司机使用手机并处于打电话或者玩手机的状态; 3) 进行手势: 司机做出右手向前方示意或者高举右手握拳的规定行为; 4) 站立: 司机处于站立状态; 5) 坐: 司机处于正常坐下状态; 6) 瞌睡: 司机趴在驾驶台上或后仰在椅子上。

为能够确保数据的多样性, 我们采集多组列车驾驶室原始视频数据进行抽帧提取训练图片以及测试图片, 标记其中出现司机人体关键点和手机目标框。针对司机人体关键点, 根据实际需求, 采用和 AI Challenger 数据集<sup>[3]</sup>相近的标注结构, 标注每个人的目标框和人体的 14 个关节, 分别为头部、脖子、左右肩、左右肘、左右手腕、左右髋、左右膝和左右踝部, 对于可见、可知不可见和不可知不可见的人体关键点分别以数字“0”、“1”、“2”进行区分, 标注如图 1 所示。



图 1 人体关键点标注

我们共提取了 2 750 张不同司机姿态的图片, 共标注了 4 379 个人体对象和 1 139 个手机目标。其中共有主司机目标 3 015 个, 副司机目标 1 299 个, 其他人员 65 个。司机站立姿态 1 315 个, 正常坐姿 2 734 个, 瞌睡 330 个。司机进行手势对象 1 590 个。并将其按 10:1 的比例随机划分为训练集和测试集。

## 2 搭建模型

根据需求, 本文中的模型需要同时检测出人体关键点和手机目标框才可以进行司机是否使用手机的判别。传统意义上这是两类截然不同的图像识别任务, 需要分别运行人体关键点检测网络和目标检测网络完成识别。然而同时运行两种网络并保持输出结果的实时性, 无疑需求强大的计算能力。在每台机车上部署高性能主机耗价昂贵, 而机车也未必能满足计算机运行所需的稳定的环境和电力需求。而选择远程部署实时分析的方法, 当前我国的网络建设也很难满足实时传输机车视频的需求。因此, 本文融合目标检测和关键点检测两种深度学习算法, 基于目标框检测实现了一种可以同时检测司机人体关键点和手机目标的神经网络。

### 2.1 模型融合

目标检测算法可以分为单阶段和双阶段两类, 单阶段目标检测通过神经网络提取图片特征, 输出长宽为原始图片  $n$  分之一的特征图, 特征图每个点回归预测固定个数的目标框的中心点、长宽、置信度和类别, 通过非最大抑制算法<sup>[4]</sup>得到最终的目标框和目标类别, 如 Yolo<sup>[5]</sup> 和 SSD<sup>[6]</sup> 网络。人体关键点检测算法根据先检测关键点再连接成人体还是先检测出人体再检测每个关键点可以分为自底而上和自顶而下两种方式, 其都是通过神经网络对人体关键点位置进行直接预测。自顶而下一般具有相对较高的准确率, 但是处理速度较低, 而且当出现多人时, 处理速度会出现明显的波动, 不利于实时检测, 近些年代表性的算法为 Alphapose<sup>[8]</sup>。自底而上的处理速度较快, 不受人数多少的影响, 其经过神经网络处理后分别得到两张特征图, 分别为人体关键点的置信度和用作指示关键点的指向作为肢体连接的概率, 通过最大匹配的方法连接各个个体关键点组成人体姿势。近些年代表性的算法为 Openpose<sup>[7]</sup>。

Pose proposal network<sup>[9]</sup> 借鉴单阶段目标检测的思想, 将人体姿态估计转换为目标检测问题, 即将每个人体关键点标注的人体关节视为一个目标框, 从图像中直接回归出包含人体关键点的目标框。将目标框的中心点视为要预测的人体关键点, 通过计算肢体连接概率对关键点进行连接, 得到人体姿态。相关实验<sup>[9]</sup>表明, Pose proposal network 在 mpii、coco 等主流数据集上取得了和 openpose 等以直接预测关键点的人体姿态估计算法不相上下的成绩, 而模型的计算量较小, 运行的速度大为提高, 适于计算受限情况下的部署。

由此, 我们可以通过目标框识别将人体关键点和手机识别统一起来, 通过一个主干网络统一提取图片中的特征信息, 分别得到  $H \times W$  大小的包含人体关键点、手机目标框的特征图 B 和只包含肢体关系的特征图 C, 如式 (1) 所示:

$$B = \{P_k^i, M^i\}_{i \in \{1, \dots, H \times W\}, k \in \{1, \dots, K\}}$$

$$C = \{\{p(k_i, k_{i+1}, x, x + \Delta x)\}^i\}_{i \in \{1, \dots, H \times W\}, (k_i, k_{i+1}) \in L} \quad (1)$$

式中,  $P_k^i$  代表特征图 B 第  $i$  个点包含的第  $k$  类人体特征点,  $M^i$  代表特征图 B 第  $i$  个点包含的手机目标。 $\{p(k_i, k_{i+1}, x, x + \Delta x)\}^i$  代表特征图 C 第  $i$  个点中  $k_i, k_{i+1}$  关键点在规定范围  $\Delta x$  内构成肢体的概率, 我们设定  $k_i, k_{i+1}$  间距小于  $\Delta x$  才可以构成肢体,  $L$  代表所有的人体肢体。 $P_k^i, M^i$  组成如式 (2) 所示:

$$P_k^i = \{conf_k^i, o_{x,k}^i, o_{y,k}^i, \omega_k^i, h_k^i\}$$

$$M^i = \{conf^i, o_x^i, o_y^i, \omega^i, h^i\} \quad (2)$$

式中,  $conf$  为  $P_k^i, M^i$  目标框的置信度, 当  $conf$  大于阈值时, 认为该点存在应识别的人体关键点或手机对象,  $o_x, o_y$  为该点目标框中心点位置,  $\omega, h$  为该点目标框的长和宽。对于手机目标, 应回归的每个目标框的长宽  $\omega, h$  已标注在训练集中, 而对于人体关节并没有事先标注, 为方便训练, 人体关节目标框的  $\omega, h$  我们统一预设为此人体关节所在的人体框对角线长度的 0.2 倍。

特征图 B 通过 *conf* 确定目标框, 并通过  $o_x$ 、 $o_y$ 、 $w$ 、 $h$  得到每个目标框的位置大小后, 通过非最大值抑制去除重复的目标框, 将人体关键点代入肢体关系特征图 C 中进行二分图匹配连接, 区分得到不同的人类个体, 最后输出人体姿势和手机目标, 如图 2 所示。

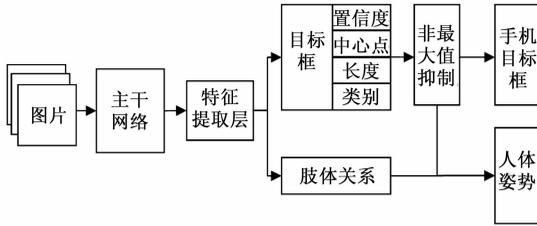


图 2 融合模型设计

### 2.2 模型压缩加速

为了满足模型在嵌入式设备上运行的需求, 本文通过使用 TensorRT 框架<sup>[10]</sup>完成对模型的压缩提速。TensorRT 是英伟达公司为了满足深度学习的高速推理需求所推出的只用于部署推理的神经网络框架。其通过实现模型简化和降低模型运行计算复杂度对模型推理速度进行加速和对模型体积压缩, 使嵌入式设备也可以满足实时运行神经网络的需求。

TensorRT 的模型简化通过整合卷积层 (convolution layer) 和批归一化层 (batch normalization layer) 为一层实现。神经网络中, 卷积层运算可以表示为:

$$x^{i+1} = W * x^i \quad (3)$$

其中:  $x^i$  代表特征图输入,  $x^{i+1}$  代表特征图输出,  $W$  代表卷积核参数,  $*$  代表卷积运算。而批归一化层可以表示为:

$$x_{bn}^{i+1} = \gamma \frac{x^{i+1} - \mu}{\sigma} + \beta \quad (4)$$

其中:  $\mu$  代表事先学习中保存的训练集  $x^{i+1}$  的均值,  $\sigma$  代表  $x^{i+1}$  的标准差,  $\gamma, \beta$  代表两个调节因子。可以将 (3) 和 (4) 融合为:

$$x_{bn}^{i+1} = \gamma \frac{W * x^i - \mu}{\sigma} + \beta = \frac{\gamma}{\sigma} W * x^i - \frac{\gamma}{\sigma} \mu + \beta \quad (5)$$

式 (5) 将批归一化层的参数  $\sigma$ 、 $\gamma$  乘入卷积核参数中, 将  $-\frac{\gamma}{\sigma} \mu + \beta$  作为卷积层的偏差, 以此完成对模型参数的简化。

TensorRT 降低模型运行计算复杂度通过降低模型运行时计算的字节位数实现。神经网络运行时默认采用单精度 32 位浮点数进行计算, 而 TensorRT 可以将模型更换为半精度 16 位浮点数乃至 8 位整数进行推算并保存模型参数, 调用设备的 GPU 实现模型提速。相关实验表明, 当模型采用 16 位浮点数计算时, 模型运行速度有近乎原有速度数倍的提升, 而模型的精度几乎没有变化。

### 2.3 行为判别

当完成人体骨架和手机的检测后, 模型输出识别出的人体各个关节点和手机的目标框。我们可以通过分析人体各部分的角度关系得到司机的姿态和动作, 并通过检测得

到的手机框与人体手部框和人体头部框的关系可以得到司机是否接电话或者使用手机。

如图 3 所示, 本文通过人体各关节点坐标得到人体肢体的向量, 通过余弦公式得到肢体之间的角度。本文计算右大臂和躯干角度  $\alpha$ 、右大臂与右小臂角度  $\beta$  对司机是否正进行手势进行二判别; 计算左右大腿和人体躯干的角度  $\delta$ 、 $\gamma$  判别司机处于站立还是坐着; 计算人体躯干与水平面角度  $\epsilon$ 、头部与躯干角度  $\zeta$  判别司机是否瞌睡; 如未在规定范围内检测到人体, 则判断为离岗。当检测结果中出现手机目标框时, 则计算手机目标框与各个人的手部、头部的目标框的重叠情况, 当手机目标框中心点落于手部、头部的目标框内或手部、头部的目标框中心点落于手机目标框内时, 判定其处于使用手机的状态。整体的后处理流程如图 4 所示。

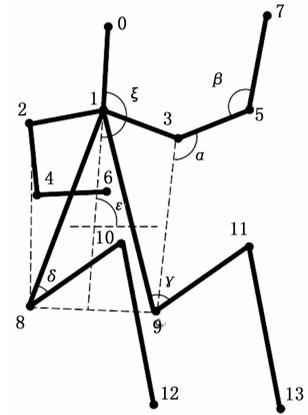


图 3 人体骨架示意图

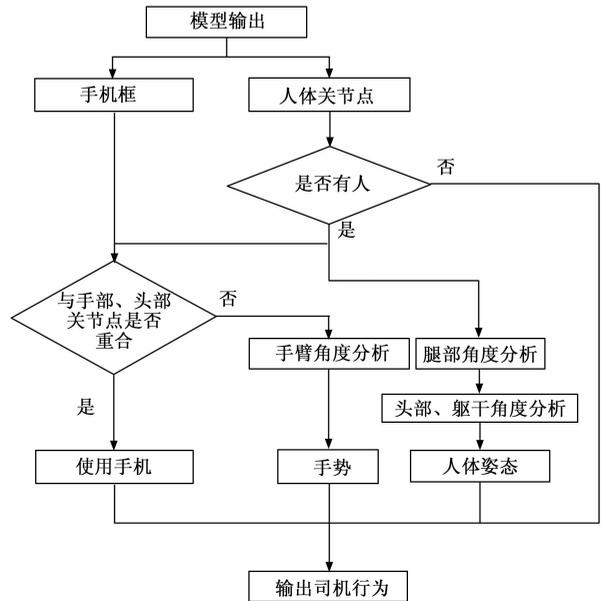


图 4 后处理流程图

## 3 实验与结果分析

本文使用了 Pytorch 深度学习框架完成融合模型的搭建。基于嵌入式设备计算量限制的需求, 主干网络选择有

较好精度表现并且计算量相对较少的 Densent<sup>[11]</sup>, 特征提取层由两层 3×3、一层 1×1 的卷积层和相应的批归一化层、relu 非线性激活层组成, 损失函数选择均方误差函数 (mean square error, MSE)。训练 GPU 为 RTX2070 8G, 系统为 Ubuntu16.04, python 环境为 python3.6.8。训练设置图片输入大小为 288×288, 初始学习率为 1e-3, 按 1e-3 \* (1-当前循环次数/总循环次数) 衰减, 每批次输入图片张数设为 32, 共循环训练集 1 000 次。为使模型不对训练集产生过拟合现象、提高样本多样性和图片质量, 使用随机旋转、左右翻转、随机裁剪、随机改变图片明暗度、饱和度和等方式在训练过程中扩增训练集。

本文选择的部署嵌入式平台为较为廉价、适于工业部署的 NVIDIA Jetson TX2, 如图 5 所示。NVIDIA Jetson TX2 是英伟达公司推出的嵌入式高性能开发板, 其采用 NVIDIA Pascal 架构, 配备 256 个 CUDA 核心, CPU 和 GPU 共用内存为 8 GB, 可进行 16 位浮点数计算。

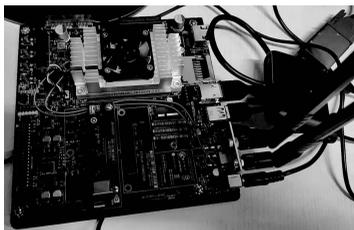


图 5 NVIDIA Jetson TX2

本文采用 OKS (目标关键点相似度, Object Keypoint Similarity)<sup>[12]</sup> 作为每张图片人体关键点识别好坏的衡量标准, 其计算公式如下:

$$OKS_p = \gamma \frac{\sum_i \exp\{-d_{pi}^2/2S_p^2\delta_i^2\}\delta(v_{pi} = 1)}{\sum_i \delta(v_{pi} = 1)} \quad (6)$$

式中,  $p$  表示真实目标中当前检测的人的代码,  $i$  代表被标记的人的每一个关节,  $d_{pi}$  表示真实目标中每个人和预测的每个人的关键点的欧氏距离,  $S_p$  表示当前人的尺度因子, 等于人所占面积的平方根, 表示人体每个关键点的归一化因子, 这个值反映了当前关键点的识别的难易程度,  $v_{pi}$  代表第  $i$  个点是否可见。

统计测试集所有图片的 OKS 以计算 MAP (mean average precision, 均值平均精度) 作为人体关键点的衡量指标。对于手机目标, 本文则采用目标检测中常用的 IOU (intersection over union, 交并比) 作为识别好坏的衡量标准, 同样统计测试集所有图片的 MAP 作为手机目标框的衡量指标。

图 6 和表 1 分别展示了模型的训练曲线和融合模型在测试集上的 MAP。人体关键点和手机框识别均处于较为良好的识别水平, 满足进行人体识别后处理的要求。

表 1 模型检测精度

人体关键点 MAP	0.773 7
手机目标框 MAP	0.866 1

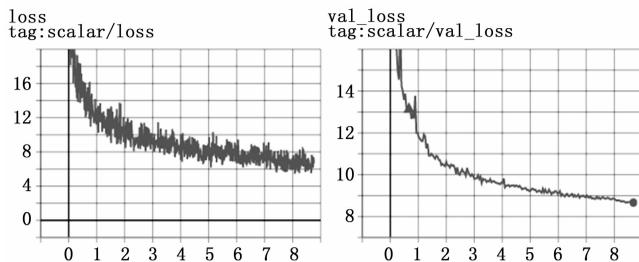


图 6 训练损失和测试集损失曲线

随后, 搭建 TensorRT 加速模型和压缩模型体积。由于模型采用 Pytorch 搭建, 无法直接转换为 TensorRT 可读取的参数格式, 先将 Pytorch 模型转换为 ONNX (Open Neural Network Exchange) 开放式格式的模型, 再输入 TensorRT 程序读入。表 2 展示了 TensorRT 加速前后模型在精度、速度和体积上的区别, 表中各项都是在 Jetson TX2 中测得。经过 TensorRT 加速后, 模型在人体关键点上 MAP 下降了 0.000 7, 而在手机目标框上上升了 0.000 2, 模型精度基本没有太大变化。而加速后, 模型的推理时间较低到了两位数范围以内, 模型运行速度加速比为 8.08, 基本满足实时运行的要求。同时, 模型的体积降低了 36.8%, 节省了嵌入式设备的存储空间。图 7 展示了模型在测试集上的检测效果, 模型展现了较好的鲁棒性, 在驾驶室内光照情况有巨大区别的情况下依然可以较为精准地识别出人体姿态。

表 2 加速前后模型对比

模型	Pytorch	TensorRT
人体关键点 MAP	0.773 7	0.773 0
手机目标框 MAP	0.866 1	0.866 2
推理时间/ms	202	25
模型大小/MB	58.2	36.8

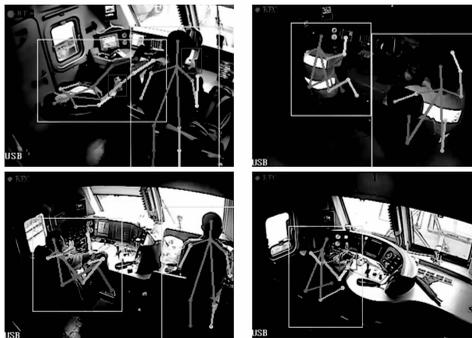


图 7 模型检测效果图

表 3 统计了经过第 2.3 节所描述的后处理后, 模型在测试集的精确率 (Precision) 和召回率 (Recall)。模型在大多数行为检测中表现较好, 其中在使用手机的行为上检测表现最好, 精确率达到 92.31%。模型在检测司机离岗的情况最差, 精确率为 62.12%。

表 3 后处理检测效果 %

司机行为	精确率	召回率
使用电话	92.31	85.71
做手势	74.47	74.83
站立	67.23	74.77
坐	85.37	73.22
瞌睡	87.10	73.97
离岗	62.12	58.57

### 4 结束语

在这项工作中, 本文成功融合了目标检测和关键点检测两种深度学习算法, 创新实现了一种可以同时检测司机人体关键点和手机的神经网络, 通过分析人体关节角度和人体关键点与手机目标框位置对司机行为进行了分类识别。并对算法模型进行了加速和压缩, 最终将其成功地部署在嵌入式边缘计算设备上, 模型推理速度为 25 ms, 可以达到实时运行的目标。另一方面, 模型在各类别行为的检测效果不一, 其中离岗行为识别的检测效果不佳, 模型训练的样本量较少, 基于各关节角度的后处理方法仍有较大的改善空间, 需进一步研究。

#### 参考文献:

[1] Evans A W. Fatal train accidents on Europe’s railways; 1980 – 2009 [J]. Accident; analysis and prevention, 2011, 43 (1): 391 – 401.

[2] 刘 峰, 申宇燕, 张瑞芳, 等. 机车车载安全防护系统应用研究 [J]. 铁路技术创新, 2015 (2): 17 – 21.

[3] Wu J, Zheng H, Zhao B, et al. AI challenger: a large – scale dataset for going deeper in image understanding [EB/OL]. arXiv preprint arXiv: 1711.06475.

[4] Neubeck, Alexander, Gool L V. Efficient non – maximum suppression [A]. Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition [C]. IEEE, 2007: 258 – 265.

[5] Yuan Q, Cong G, Ma Z Y, et al. Time – aware point – of – interest recommendation [A]. International AGM SIGIR Conference on Research and Development in Information Retrieval [C]. ACM, 2013: 363 – 372.

[6] Ferenc G, Ye M, Lee W C. Location recommendation for out – of – town users in location – based social networks [A]. ACM International Conference on Information & Knowledge Management [C]. ACM, 2013 : 721 – 726.

[7] Lian D F, Zheng K, Ge Y, et al. GeoMF: Scalable location recommendation via joint geographical modeling and matrix factorization [J]. ACM Transactions on Information Systems, 2018, 36 (3): 1 – 29, 33.

[8] Rafal Z. Initialization of non negative matrix factorization with vertices of convex polytope [A]. Proceedings of the 11st International Conference on Artificial Intelligence and Soft Computing [C]. Zakopane, Poland, 2012: 448 – 455.

[9] Feng S S, Li X T, Zeng Y F, et al. Personalized ranking metric embedding for next new POI recommendation [A]. International

pression [A]. 18th International Conference on Pattern Recognition (ICPR’06) [C]. 2006, 3: 850 – 855.

[5] Redmon J, Divvala S K, Girshick R B, et al. You only look once: unified, real – time object detection [A]. Computer vision and pattern recognition [C]. IEEE: Piscataway, NJ, 2016: 779 – 788.

[6] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector [A]. European Conference on Computer Vision [C]. Springer: Berlin, 2016: 21 – 37.

[7] Cao Z, Simon T, Wei S E, et al. OpenPose: realtime multi – person 2D pose estimation using part affinity fields [A]. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition [C]. Piscataway, NJ: IEEE, 2017: 7291 – 7299.

[8] Fang H S, Xie S Q, Tai Y W, et al. RMPE: regional multi – person pose estimation [A]. Proceedings of the IEEE International Conference on Computer Vision [C]. 2017: 2334 – 2343.

[9] Sekii, Taiki. Pose proposal networks [A]. Proceedings of the European Conference on Computer Vision (ECCV) [C]. 2018: 342 – 357.

[10] NVIDIA. Nvidia TensorRT High Performance Deep Learning Inference for Production Deployment [EB/OL]. <https://developer.nvidia.com/tensorrt>.

[11] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks [A]. Proceedings of the 2017 IEEE conference on computer vision and pattern recognition [C]. Piscataway, NJ: IEEE, 2017: 4700 – 4708.

[12] Papandreou G, Zhu T, Kanazawa N, et al. Towards accurate multi – person pose estimation in the wild [A]. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition [C]. Piscataway, NJ: IEEE, 2017: 4903 – 4911.

[1] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks [A]. Proceedings of the 2017 IEEE conference on computer vision and pattern recognition [C]. Piscataway, NJ: IEEE, 2017: 4700 – 4708.

[2] Papandreou G, Zhu T, Kanazawa N, et al. Towards accurate multi – person pose estimation in the wild [A]. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition [C]. Piscataway, NJ: IEEE, 2017: 4903 – 4911.

[3] Wang Y, Wu Y, Zhang J, et al. Deep learning for point – of – interest recommendation [A]. Proceedings of the AAAI International Conference on Artificial Intelligence [C]. AAAI Press, 2015: 2069 – 2075.

[4] Liu Q, Wu S, Wang L, et al. Predicting the next location: a recurrent model with spatial and temporal contexts [A]. Thirtieth AAAI Conference on Artificial Intelligence [C]. AAAI Press, 2016: 194 – 200.

[5] Zhao P P, Zhu H F, Liu Y C, et al. Where to Go Next: A Spatio – temporal LSTM model for Next POI Recommendation [J]. arXiv preprint arXiv: 1806.06671, 2018.

[6] Fu M S, Hong Q, Moges D, et al. Attention based collaborative filtering [J]. Neurocomputing, 2018, 1311: 88 – 98.

[7] 徐振亮, 李艳焕, 闫 利, 等. 共线方程线性化的矩阵模型 [J]. 北京大学学报 (自然科学版), 2016, 52 (3): 403 – 408.

[8] Lian D F, Zhao C, Xie X, et al. GeoMF: Joint geographical modeling and matrix factorization for point – of – interest recommendation [A]. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [C]. 2014: 831 – 840.