

基于深度神经网络的中国移动营业厅人流量统计研究

陈乐, 丁戈, 肖忠良

(中国移动信息技术有限公司, 北京 102200)

摘要: 文章提出深度神经网络的方法, 将分治策略引入到人流量统计问题中。这个方法以 VGG16 作为特征编码网络, UNet 作为解码网络, 将人流量统计变成一个可拆解的任务。在统计图片上人数时, 总可以把图片分解成多个子区域, 使得每个区域的人流量计数都是在之前训练集上见过的人数类别。然后再把每个子区域上面的人数加起来, 就是统计结果。在对广州某个中国移动营业厅的人流量统计的实验中, 预测的速度能达到实时监测的速度和高精确度。

关键词: 分治策略; 人流量统计; 深度神经网络; VGG16; UNet

中图分类号: TP18

文献标识码: A

文章编号: 2096-4706 (2021) 24-0084-05

Research on People Flow Statistics of China Mobile Business Hall Based on Deep Neural Network

CHEN Le, DING Ge, XIAO Zhongliang

(GMCC, Beijing 102200, China)

Abstract: In this paper, the method of deep neural network is proposed, and the divide and conquer strategy is introduced into the problem of people flow statistics. This method takes VGG16 as the feature coding network and UNet as decoding network, which turns the people flow statistics into a detachable task. When counting the number of people on the pictures, we can always decompose the picture into multiple sub-regions, so that the people flow count in each area is the number of people seen on the previous training set. Then add up the number of people in each sub-region, which is the result of statistics. In the experiment of people flow statistics in a China mobile business hall in Guangzhou, the predicted speed can reach real-time monitoring speed, and has high accuracy.

Keywords: divide and conquer strategy; people flow statistics; deep neural network; VGG16; UNet

0 引言

截止 2021 年 6 月底, 中国移动客户达到 9.46 亿户, 线下营业厅遍布中国的每个大大小小的城市。5G 套餐客户达到 2.51 亿户, 稳居三大运营商之首^[1]。面对如此庞大的客户群体, 营业厅员工在多用户咨询和办理业务繁忙的时候, 难免会出现服务不周到, 工作效率下降的情况。而且在新冠疫情的影响下, 每个营业厅的人数也要得到相应的控制。而且客户能得到精确的每个营业厅的人数, 能帮助他们选择人数相对较少的营业厅办理业务。在提高中国移动的员工服务质量和精准控制营业厅人数来更好的进行防疫工作的压力之下, 本文章提出了一种基于神经网络和分治策略的精准和实时的人流量统计的方法。首先, 该方法需要先训练一个可以计数 0 到 10 个人的深度神经网络模型。然后再利用判别器来判断该子区域是否需要继续进行切割, 若子区域的人数大于 10 人的时候需要继续切割。最后合并不同的子区域的结果^[2]。该预测网络的整体结构是以 VGG16 作为特征编码器的主干网络, 参考 UNet 作为解码器, 一个计数分类器, 一个划分判别器。网络结构的各个部分各司其职, 相互作用, 提高了模型的预测速度和准确率。

1 人流量统计算法

1.1 人流量统计算法的结构图

算法的整体结构如图 1 所示, 由一个 VGG16 特征编码器, 一个 UNet 型的解码器, 一个计数分类器和一个划分判别器构成的。左边的 Conv3, Conv4 和 Conv5 是参考 VGG16 的网络进行的下采样卷积操作。中间有格子划分的 F_2 和 F_1 是通过上采样得到的。 F_1 是通过 F_0 上采样的得到的, F_2 是通过 F_1 上采样得到的, 它们的步长都是 2。这个部分是参考 UNet 型的解码器得到的。 F_1 和 F_2 划分出来的区域就是子区域。这时划分判别器就需要判断是否要继续进行上采样, 让子区域划分得更加小来满足子区域人数小于 10 的要求。 W_1 和 W_2 的值就只有 (0, 1), 来判断是否需要继续进行“错误! 未找到索引项。”上采样。最后再把特征图中的每个子区域让训练好的计数分类器来判断里面有多少人, 是属于 (0, 10) 中的哪一个类别, 得到 C_0 , C_1 和 C_2 ^[3]。

1.2 VGG16 的特征编码器

图 2 是原先的 VGG16 的网络结构层图。由图结构可知, 总共有 13 个卷积层和 3 个全连接层。其中的池化部分都是用了 max pooling, 激活函数是用了 ReLU 函数。卷积核的大小都是 3×3 的^[4]。文章人流量统计算法中的结构, 就是去掉了 VGG16 中所有全连接层, Conv5 就是上图结构中除了全连接层中的最后的输出特征图, 也就是原图中的输入图

片下采样 32 倍得到的输出特征图。Conv5 的结果直接通过计数分类器, 预测出 C_0 。 C_0 就是当前整体图片的预测输出结果, 也相当于由普通的 VGG16 网络预测得到的结果。

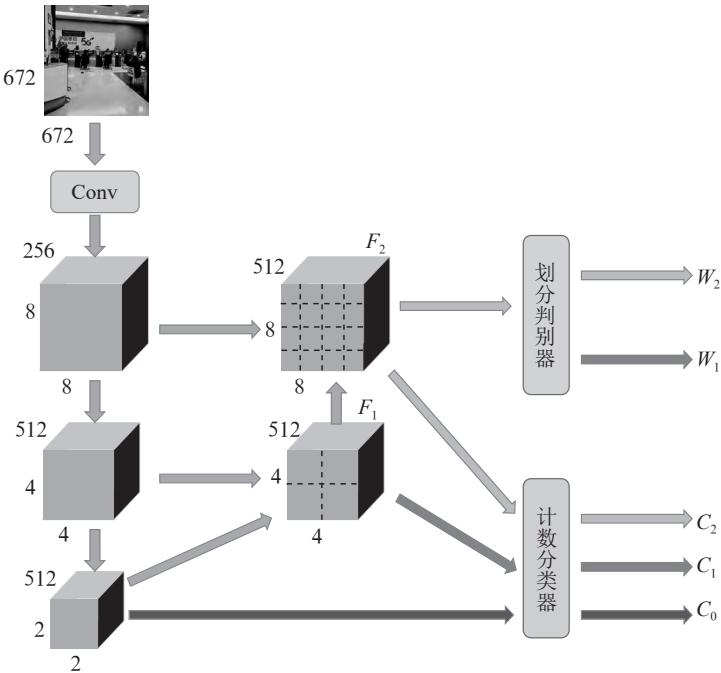


图 1 人流量统计算法的结构图

卷积网络的整体结构					
A	A-LRN	B	C	D	E
11- 隐藏层	11- 隐藏层	13- 隐藏层	13- 隐藏层	16- 隐藏层	19- 隐藏层
输入图片 (672*672)					
Conv3-64	Conv3-64 LRN	Conv3-64 Conv3-64	Conv3-64 Conv3-64	Conv3-64 Conv3-64	Conv3-64 Conv3-64
最大池化层					
Conv3-128	Conv3-128	Conv3-128 Conv3-128	Conv3-128 Conv3-128	Conv3-128 Conv3-128	Conv3-128 Conv3-128
最大池化层					
Conv3-256 Conv3-256	Conv3-256 Conv3-256	Conv3-256 Conv3-256	Conv3-256 Conv3-256 Conv1-256	Conv3-256 Conv3-256 Conv3-256	Conv3-256 Conv3-256 Conv3-256 Conv3-256
最大池化层					
Conv3-512 Conv3-512	Conv3-512 Conv3-512	Conv3-512 Conv3-512	Conv3-512 Conv3-512 Conv1-512	Conv3-512 Conv3-512 Conv3-512	Conv3-512 Conv3-512 Conv3-512 Conv3-512
最大池化层					
Conv3-512 Conv3-512	Conv3-512 Conv3-512	Conv3-512 Conv3-512	Conv3-512 Conv3-512 Conv1-512	Conv3-512 Conv3-512 Conv3-512	Conv3-512 Conv3-512 Conv3-512 Conv3-512
最大池化层					
4096- 全连接层					
4096- 全连接层					
1000- 全连接层					
Soft-max 输出函数					

图 2 VGG16 的网络分层图

1.3 UNet 的解码器

图 3 就是 UNet 网络的整个网络结构图, 这个网络的左半部分是由两个 3×3 的卷积层和两个最大池化层, 步长为 2 组成的, 激活函数是 ReLU 函数^[5]。左边结构每进行一次下采样, 通道数都会乘以 2。这个网络的右半部分是由一个 2×2 的上采样卷积, 激活函数是 ReLU 函数。右边的网络通道, 是由左边对应的特征图通道和经过上采样得到的

特征图通道叠加的, 最后得到的特征图, 有一半包含了左边对应特征图的信息。因为这个池化的方式是最大池化, 所以一定程度上进行下采样, 都会损失了很多图片中的信息。这里的通道融合的方式, 能将浅层低维度的特征图和深层高维度的特征图的信息进行融合, 让最终的特征图中蕴含着更加丰富的图片信息, 让图像的预测更加的全面和准确。

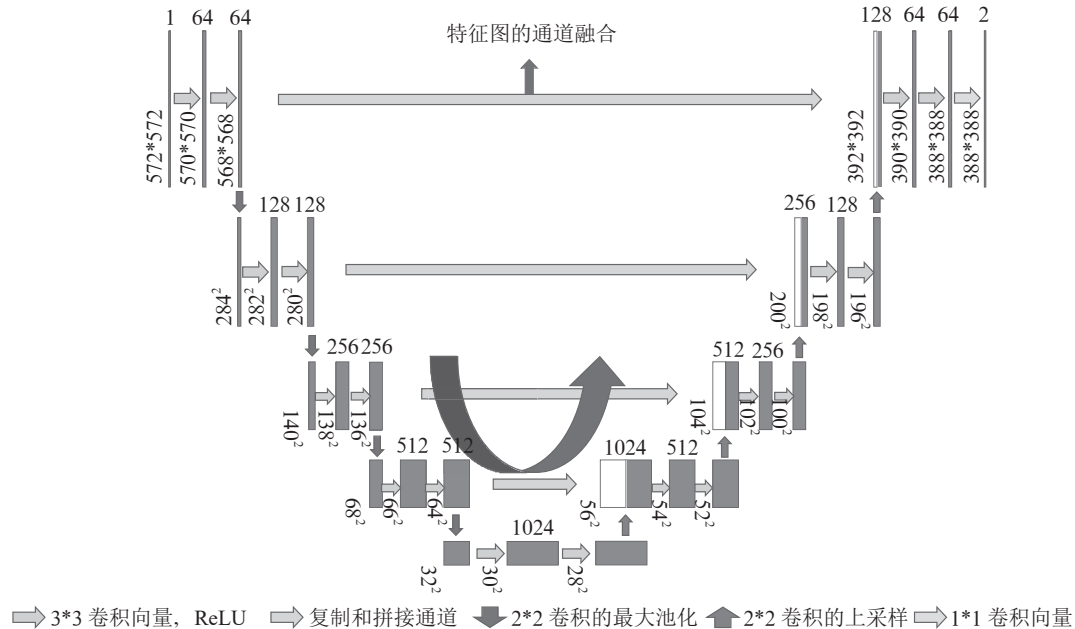


图 3 UNet 网络结构图

1.4 划分判别器和计数分类器

划分判别器是用来判断经过特征通道融合的特征图是否需要继续进行上采样操作。图中 F_1 的特征图就是把 64×64 像素的原图分为了四个子区间, 每个子区间对应着 16×16 , 但是它的感受野是整张原图。如果这个子区间的人数大于 10 的话, 则划分判别器会让该特征图继续进行步长为 2 的上采样操作, 直到子区间人数小于 10 为止^[6]。划分判别器由一个 2×2 的平均池化层 (步长为 2), 一个 512 个通道的 1×1 卷积和一个 1 个 1×1 输出卷积构成的, 最后通过一个 sigmoid 函数输出 0 或者 1。

计数分类器是一个 21 个类别的训练好的分类器, 区间范围是 0 到 10, 间隔是 0.5。该分类器是 Resnet50, 通过对 21 个类别的 2 000 张图片进行训练得到的具有权重的模型。如图 2, F_1 特征图所示, 里面有 4 个区间, 这 4 个子区间分别作为输入, 输入给计数分类器进行预测。最后将四个子区间的预测结果相加, 向下取整即为最终的输出结果。

1.5 人流量统计算法的损失函数

这是人流量统计算法的损失函数: $L = \sum_{i=0}^N L_C^i + L_R^N$, 由两个部分组合而成。左边部分是多个交叉熵损失函数, 是用来监督在对子区域人数统计的不同级别的分类问题。右边部分是用来监督最终划分输出的函数。该网络可以支持多任务网络的训练方式, 所以才有一个关于 N 次的求和^[7]。

2 模型算法实验

2.1 中国移动营业厅采集数据

历时两个星期, 在广州天府路中国移动营业厅用手机采集图片。采集方式是在营业时间段, 每隔半个小时在营业厅的各个角度采集若干张含有人群信息的图片。然后总共采集了大概 200 张左右的有效图片, 人数范围在 10 ~ 70 人左右。最后在网上开源的 ShanghaiTech 的数据集中融入了 100 张左右的图片, 组成了最后 400 张左右的数据集, 部分图片如图 4 所示。



图 4 中国移动营业厅数据集

2.2 生成人群热力图的标签方式

人群流量的标签方式一般有两种,一种是对人头进行检测框的标记,另外一种是直接对图片生成热力图,如图5所示。对于采集到的数据集,需要标记它们的真实标签即图片中的真实人数。标签方式是生成人群热力图,然后人群流量统计就变成了对热力图的积分计算,甚至可以计算每平方米的人数和聚集程度,来判断相应需求^[8]。原图上的标注方式是,在人头所在的像素点上标记一个红点,然后再根据红点生成相应的人群热力图^[9]。



图5 数据集图片的标注方式

2.3 训练网络

本文采用深度神经网络的方法来预测中国移动营业厅的真实实时人数。采用将原图区域进行划分成若干个子区域,再在子区域进行预测人数的方式来预测真实人数。该方法训练网络的步骤如下:步骤一:数据预处理,对采集好的图片进行降噪和除去黑白噪点的操作。使得图片更加的平滑,过滤掉无效信息的图片。步骤二:数据增强,因为采集的数据不是特别的充足,所以对图片进行随机裁剪和翻转的操作增强数据集的数量。步骤三:根据标记好的红点,生成每一张图片的人群密度热力图。步骤四:模型训练,把数据集按照8:1:1的比例,将数据集分为训练集、验证集和测试集。随机初始化网络中的所有参数, batch size 设置为64,把每一批的数据集喂给模型进行正向传播和反向传播进行参数的更新。训练次数设置为100,最后选取训练了至少50次的损失函数值最小的模型为最佳的训练模型。

3 模型预测结果分析

3.1 中国移动营业厅数据集的标签人数分布

从经历半个月左右的时间在中国移动营业厅收集到的数据,图6可见,营业厅大部分的人流数量集中在20人到60人。有时候比较繁忙的时候能达到100人左右。由于这种人数,所以训练分类网络的时候,把子区域的人群计数分类定为了0-10人的范围。

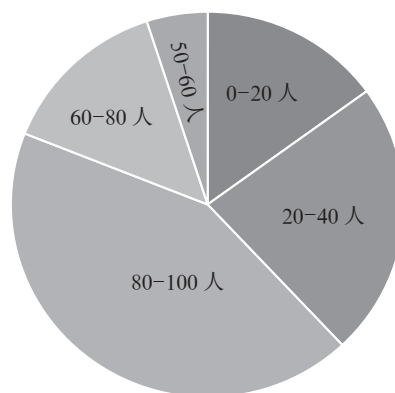


图6 数据集人数分布

3.2 损失函数和预测效果评估

人流量统计模型的损失函数由分类的交叉熵损失和回归函数共同组成,所以选用最小二乘法的梯度下降训练方法。损失函数一般是用来测试训练模型的时候真实值和预测值的差异的,一般来说损失函数的值越低,模型的预测效果越好。但是一般来说会选择验证集中,真实值和预测值最小的模型作为最终的模型参数输出^[10]。图7就是训练模型时候的损失函数和准确率。

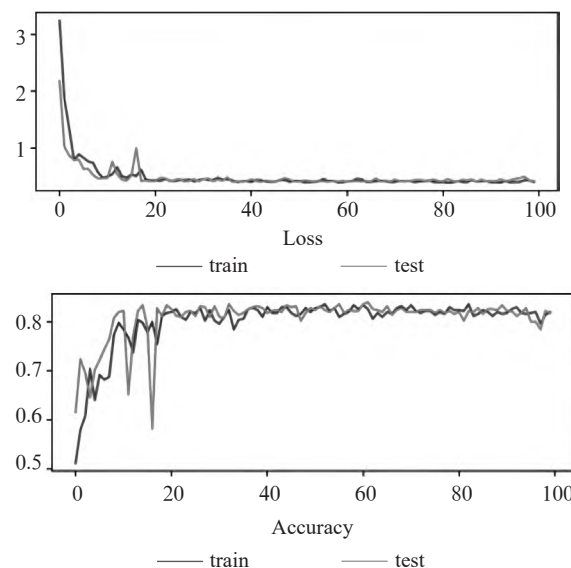


图7 训练模型的损失函数和准确率

在测试集中测试模型性能的方法是平均绝对误差,即真实值和预测值的差值的绝对值和的平均值。平均绝对误差的公式: $\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)$ ^[11]。在测试集上的结果是,平均绝对误差:0.27即每次预测只差了0.27个人数的统计。结果的准确率(误差少于1)为98.87%。从结果上来看,模型预测的准确度还是十分可观的。而且模型在一秒钟内可以预测大概30张到40张的图片,满足的实时监测的需求。

4 结论

本文提出一种基于分治策略,将原图分割成若干个子区域进行区域计数的人流量统计的深度神经网络模型。该网络模型参考了传统神经网络VGG16作为特征提取的主干网络

和 UNet 网络作为解码器。在测试和预测真实的中国移动营业厅时,模型的人数准确率基本保障不会出现错误而且能达到实时更新营业厅人数的效果。这很好的满足的让营业厅的营业员精准定位服务,而且能让客户根据每个营业厅的实时人数来选择人数更少的营业厅办理业务。在疫情管控营业厅的实时人数时,可以起到监督和监控的作用。

参考文献:

- [1] 中国移动. 营运数据 [EB/OL].[2021-09-25].https://www.chinamobileltd.com/sc/ir/operation_q.php.
- [2] SAM D B, SAJJAN N N, BABU R V, et al. Divide and Grow: Capturing Huge Diversity in Crowd Images with Incrementally Growing CNN [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE: Salt Lake City, 2018: 3618-3626.
- [3] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation [J/OL].arXiv: 1511.00561 [cs.CV]. (2015-11-02). <https://arxiv.org/abs/1511.00561>.
- [4] HE K M, ZHANG X Y, REN S Q, et al. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification [J/OL].arXiv: 1502.01852 [cs.CV]. (2015-02-06). <https://arxiv.org/abs/1502.01852>.
- [5] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition [J/OL].arXiv: 1409.1556 [cs.CV]. (2014-09-04). <https://arxiv.org/abs/1409.1556>.

[6] LEMPITSKY C, ZISSERMAN A. Learning To count objects in images [C]//NIPS' 10: Proceedings of the 23rd International Conference on Neural Information Processing Systems. Red Hook: Curran Associates, 2010: 1324-1332.

[7] LI R B, XIAN K, SHEN C H, et al. Deep attention-based classification network for robust depth prediction [J/OL].arXiv: 1807.03959 [cs.CV]. (2018-07-11). <https://arxiv.org/abs/1807.03959>.

[8] NIU Z X, ZHOU M, WANG L, et al. Ordinal Regression with Multiple Output CNN for Age Estimation [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 4920-4928.

[9] GIRSHICK R. Fast R-CNN [C]//2015 IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE, 2015: 1440-1448.

[10] IDREES H, TAYYAB M, ATHREY K, et al. Composition Loss for Counting, Density Map Estimation and Localization in Dense Crowds [J/OL].arXiv: 1808.01050 [cs.CV]. (2018-08-02). <https://arxiv.org/abs/1808.01050>.

[11] MASKA M, ULMAN V, SVOBODA D, et al. A benchmark for comparison of cell tracking algorithms [J]. Bioinformatics, 2014, 30 (11): 1609-1617.

作者简介: 陈乐(1982—),男,汉族,重庆人,项目总监,硕士研究生,研究方向:AI0ps、业务支撑系统运营支撑;丁戈(1995—),男,汉族,广东广州人,项目经理,硕士研究生,研究方向:AI0ps 算法;肖忠良(1986—),男,汉族,广东广州人,项目经理,硕士研究生,研究方向:AI0ps、业务支撑系统运营支撑。

(上接 83 页)网络在一天(24 小时)中对于该节点的预测结果与真实结果较为接近,上文也提到该模型在测试集上的预测准确率为 90.84%,说明图卷积神经网络模型可以通过训练获得较好的预测性能。

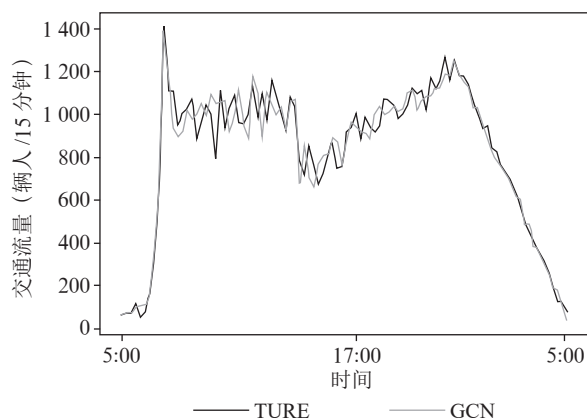


图3 预测结果对比

5 结论

随着 5G 网络和大数据技术的不断发展,商用车联网将迎来新一轮产业变革,建立智慧交通服务平台能够以高扩展性、高容错性、高效性、高可靠性完成交通流大数据的处理和分析,提升交通安全与拥堵主动调控能力。

本文基于目前城市交通中客观存在的问题,结合 5G-V2X 通信技术和大数据分析技术,设计开发了基于 5G-V2X 的智慧交通服务平台。设计平台的过程中综合考虑了数据采集、传输和存储中的难题,支持数据的实时采集、实时传输,可支持每秒 1 000 条数据的插入和亿条级数据的存储和管理;在交通智能化上,经过对算法模型的深入研究,采用图卷积神经网络作为交通流预测模型,交通流预测的准确率(拟合优度)高于 90%。

参考文献:

- [1] 胡勇剑. 智慧交通在智慧城市中的探索与应用 [J]. 交通与运输, 2020, 33 (S2): 190-193+199.
- [2] 郑文超, 贲伟, 汪德生. 智慧交通现状与发展 [J]. 指挥信息系统与技术, 2018, 9 (4): 8-16.
- [3] 李新洲. 车联网 LTE-V2X 与 5G-V2X (NR) 对比分析 [J]. 信息通信技术与政策, 2020 (7): 93-96.
- [4] 丁启枫, 杜昊, 吕玉琦. 5G-V2X 应用场景和通信需求研究 [J]. 数字通信世界, 2019 (2): 24-25.
- [5] 李志帅, 吕宜生, 熊刚. 基于图卷积神经网络和注意力机制的短时交通流量预测 [J]. 交通工程, 2019, 19 (4): 15-19+28.
- [6] 朱凯利, 朱海龙, 刘靖宇, 等. 基于图卷积神经网络的交通流量预测 [J]. 智能计算机与应用, 2019, 9 (6): 168-170+177.

作者简介: 刘剑(1975—),男,汉族,四川成都人,副教授,博士,研究方向:车联网、大数据、物流管理。