

# 基于语义分割与深度估计的行车环境实时解析

林坤辉, 陈雨人

(同济大学 道路与交通工程教育部重点实验室, 上海 201800)

**摘要:** 道路行车环境的实时解析是智能驾驶的关键技术, 尽管神经网络在实现语义分割和深度估计上能取得不错的精度, 但由于模型参数多、计算量大等问题, 导致难以实现实时计算; 针对该问题, 提出了一个轻量化、高效的特征提取模块和一个综合考虑语义信息和深度信息的特征解码模块, 在一个网络中同时完成语义分割和深度估计两个任务; 在 CityScapes 数据集中, 语义分割预测结果的 mIOU 为 65.0%、深度估计结果的误差为 0.21, 并且在单个 GPU 上推断速度达到了 65FPS, 满足实时性要求。

**关键词:** 深度估计; 语义分割; 神经网络

## Real-time Semantic Segmentation and Depth Estimation Method for Driving Environment

Lin Kunhui, Chen Yuren

(Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai 201800, China)

**Abstract:** The real-time analysis of the road driving environment is the key technology of intelligent driving. Although the neural network can achieve good precision in semantic segmentation and depth estimation, it is difficult to realize real-time calculation due to problems such as many model parameters and large calculation amount. Aiming at this problem, this paper proposes a lightweight and efficient feature extraction module and a feature decoding module that comprehensively considers semantic information and depth information, and simultaneously performs two tasks of semantic segmentation and depth estimation in a network. In the CityScapes dataset, the mIOU of the semantic segmentation prediction result is 65.0%, the error of the depth estimation result is 0.21, and the inference speed reaches 65 FPS on a single GPU, meeting the real-time requirements.

**Keywords:** depth estimation; semantic segmentation; neural network

## 0 引言

道路行车环境的感知与解析是车辆辅助驾驶与自动驾驶的关键技术, 同时也是车辆进行判断和决策的基础。在目前实现道路行车环境感知的众多技术中, 计算机视觉技术凭借其设备安装简单、操作方便的优点, 一直以来都是研究的热点。它通过车辆前方摄像头获取的图像信息, 利用设计的算法对图像中的场景进行语义分割和深度估计, 实现行车环境的解析。

不过由于场景复杂、光线阴影变化大、物体遮挡等原因, 使得语义分割和深度估计这两个基础的计算机视觉任务极富挑战。近年来随着深度学习的迅速发展, 使得关于两个任务的研究取得了极大的进展。目前已经有不少学者提出实现语义分割和实现深度估计的神经网络, 并且能达到良好的精度, 但由于其网络结构过于庞大, 导致模型计算量过大, 无法达到实时计算和落地使用。针对该问题,

不少研究者通过精简网络结构, 减小网络的深度和宽度, 希望能达到实时计算的目的。但由于网络结构的精简, 导致了精度的下降, 因此在实时计算和精度这两方面的权衡, 成为了时下研究的难点。另外在道路场景的解析中, 语义信息和深度信息缺一不可, 目前大部分工作仍将二者当做两个独立的任务分别处理, 导致在实时性上再打折扣。

考虑到语义信息和深度信息具有一定相关性, 即相同的语义要素具有相近的深度信息, 本文提出了一个轻量级的网络模型, 同时完成语义分割和深度估计两个任务, 并且实现端到端的训练与推测。在保持一定精度的情况下, 实现实时计算 (30 fps 以上)。本文的主要工作如下:

1) 设计了一个轻量化、高效的特征提取模块。在保持模型精度的情况下, 大大减少了模型的参数量, 提高了模型计算速度;

2) 设计了一个基于多尺度卷积和注意力机制的解码模块, 用于捕捉特征间的语义信息或深度信息;

3) 根据前两点设计的编码和解码模块, 搭建了一个端到端实现语义分割和深度估计的神经网络, 实现单个网络同时解决两个任务, 并且达到实时性的要求。

## 1 相关工作

语义分割是对图像中每一个像素进行分类的任务。文

收稿日期: 2019-10-11; 修回日期: 2019-10-23。

作者简介: 林坤辉(1995-), 男, 福建漳州人, 硕士研究生, 主要从事行车环境图像分割方向的研究。

陈雨人(1966-), 男, 江苏盐城人, 博士生导师, 教授, 主要从事道路安全与环境方向的研究。

献 [1] 开创性地提出了全卷积神经网络实现图像的语义分割, 抛弃了后处理的步骤, 使得语义分割结束了多阶段处理的时代, 实现了语义分割端到端的训练与预测。在此基础上, 文献 [2-3] 等网络提出了高效的深度卷积编码器-解码器框架, 并通过编码器和解码器中部分网络结构层的直接连接, 一方面使得网络的训练更加容易、提高网络的训练效率, 另一方面使得解码器能够获得低阶的物体特征细节, 从而改善物体边缘的预测效果, 整体上提高了预测的精度。文献 [4-5] 等提出的空洞卷积和金字塔池化, 增大了特征提取过程的感受野, 考虑了图像中的上下文关系, 进一步提高了预测精度。尽管上述网络能取得良好的分割效果, 却是建立在结构复杂、模型庞大的神经网络基础上, 无法满足实时性计算的需求。文献 [6-7] 等通过减少网络层数、减少特征图通道数目、精简网络结构等措施, 在牺牲一定精度的情况下, 将预测的速度提高到了 30FPS 以上。本文通过设计了一个高效的特征提取结构, 减少了模型的计算量, 在取得与文献 [7] 相近的精度下, 将预测速度提高到了 60FPS 以上。

深度估计也是一个逐像素考虑的任务, 旨在估计图像中每一个像素到观测位置的距离。文献 [8] 将深度估计作为一个回归问题, 首先提出使用卷积神经网络估算图像深度。得益于 CNN 强大的特征提取能力, 针对单张 RGB 图像的深度估计研究涌现了诸多成果, 相比于传统的方法, 在精度上取得了长足的进展。在监督学习方面有文献 [9-11] 等, 无监督学习方面有文献 [12-13] 等, 半监督学习方面有文献 [14]。本文提出的模型与上述模型存在较大差异, 一是在模型结构上做了精简优化, 使得深度的预测速度能达到实时的效果; 二是模型不仅完成深度估计, 同时也完成了语义分割任务, 二者在模型训练时, 能相互借鉴各自的信息, 使得模型精度更高, 泛化能力更强。

深度估计与语义分割的结合, 一定程度上是基于二者任务特点的相似性和物体深度与语义的相关性, 使得二者结合既能减少计算量, 又能相互增益、提高精度。在深度估计与语义分割结合方面, 文献 [15] 提出了一个递归处理语义信息和深度信息的网络结构, 尽管能充分利用深度信息和语义信息之间的关系, 但网络在深层时大量使用了大卷积核的卷积运算, 导致计算量爆炸增长网络参数量更是达到 350 M, 导致无法达到实时计算的要求。在处理语义分割与深度估计两项任务的关系方面, 本文提出的网络结构中, 语义分割和深度估计任务共享了大部分的权重, 使得模型能够捕捉二者之间的相关性; 同时对于深度估计和语义分割, 又有各自的注意力机制模块, 使得网络能学习各自任务的差异性。在网络计算量和实时性方面, 本文提

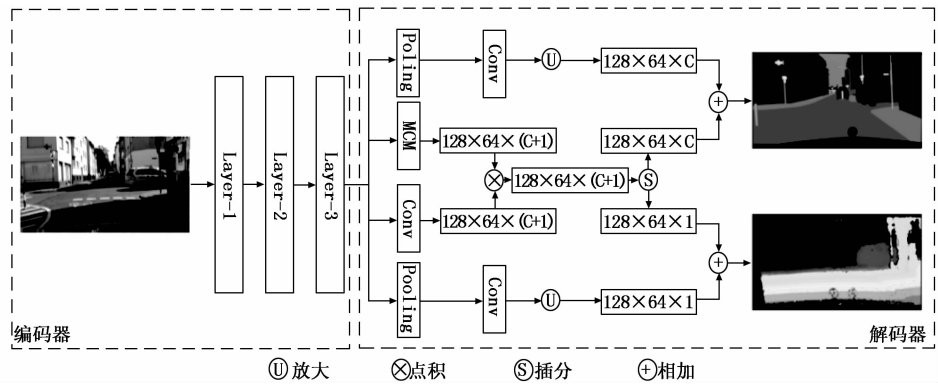


图 1 网络结构图

出的网络结构能在保持一定精度的情况下达到 65FPS 的速度, 模型参数仅为 1.2 M。

## 2 模型结构

### 2.1 总体结构

模型的总体是一个编码器-解码器的结构, 编码器用于提取图像的特征, 解码器用于对提取到的特征进行处理, 预测像素点所属的类别和深度。网络的总体结构如图 1 所示, 输入的图片依次经过编码器和解码器, 然后得到深度预测和语义分割预测的结果。

### 2.2 编码器

编码器部分共包含三个子层, 每个子层由一个降采样单元和若干个改进的残差单元组成, 各子层的组成分布如表 1 所示。随着层数的加深, 改进的残差单元模块增加, 输入的特征图尺寸减小、通道数目增加。

表 1 编码器结构图

层	单元	个数	尺寸
Layer 1	降采样单元	1	512×256×32
	改进的残差单元	2	512×256×32
Layer 2	降采样单元	1	256×128×64
	改进的残差单元	4	256×128×64
Layer 3	降采样单元	1	128×64×128
	改进的残差单元	8	128×64×128

降采样单元。降采样单元设置的目的在于减小特征图尺寸同时增大通道的数目, 从而达到扩大感受野和生成高阶特征的目的。文中所使用的降采样单元包含两个分支, 一个是卷积核大小为 3×3、步长为 2 的卷积操作, 另一个是池化大小为 2×2、步长为 2 的最大池化操作。这两个操作都能将特征图的大小缩小到原来的一半, 达到降采样单元的目的之一——减小特征图尺寸。通过级联两个分支的输出, 整个降采样单元输出的通道数则变成输入的两倍, 实现了通道数的增加。

改进的残差单元。本文针对文献 [16] 中提出的残差单元改进思路在于减少单元中的参数量和计算量, 从而达到减少模型参数量、加速计算的目的。对于输入通道数为  $N_m$ 、输出通道数为  $N_{out}$ 、输出特征图尺寸为  $h \cdot w$ 、卷积核

大小为  $f_h \cdot f_w$  的卷积操作来说, 参数量大小  $N_p$  为:

$$N_p = N_m \cdot f_h \cdot f_w \cdot N_{out} \quad (1)$$

计算量大小  $N_o$  为:

$$N_o = N_m \cdot f_h \cdot f_w \cdot N_{out} \cdot h \cdot w \quad (1)$$

从式 (1)、(2) 中可以发现, 影响模型参数量和计算量的因素有通道数和卷积核的尺寸, 且成正相关。因此, 本文针对通道数和卷积核尺寸, 对残差单元进行了改进, 改进后的残差单元如图 2 所示。首先将输入在通道维度上一分为二, 分别进入两个不同的卷积分支, 实现输入通道  $N_m$  和输出通道的减小  $N_{out}$  的减小。其次, 将卷积分支上  $3 \times 3$  的卷积拆分成  $3 \times 1$  和  $1 \times 3$  卷积核, 减小了卷积核的大小。最后, 级联两个分支的输出, 恢复了通道数, 并与原始输入直接相加, 维持残差结构。由于通道拆分会导致不同分支之间的通道无法进行特征的组合, 因此在单元最后增加一个通道的重组, 重新分布通道的顺序, 保证通道间特征的交流。假设残差单元的输入通道和输出通道数相同, 则改进前, 残差单元的参数量大小为  $N_m \times 3 \times 3 \times N_{out} \times 2 = 18N_m^2$ , 计算量大小为  $N_m \times 3 \times 3 \times N_{out} \times 2 \times h \times w = 18h\tau w N_m^2$ , 而改进后的残差单元的参数量大小为  $\frac{N_m}{2} \times 1 \times 3 \times \frac{N_{out}}{2} \times 4 \times 2 = 6N_m^2$ , 计算量大小为  $\frac{N_m}{2} \times 1 \times 3 \times \frac{N_{out}}{2} \times 4 \times 2 \times h \times w = 6h\tau w N_m^2$ 。通过比较可以知道, 计算量和参数量均变为原来的  $1/3$ , 从而提高了模型的计算速度。

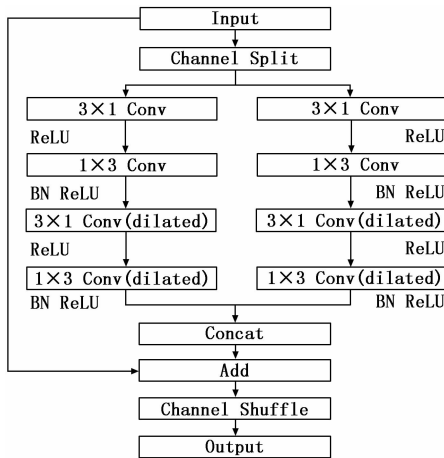


图 2 改进的残差单元结构图

另外在改进的残差单元左右两个分支中, 会各自使用两个空洞卷积。空洞卷积相比于普通的卷积操作而言, 能获得更大的感受野, 提取得到的特征更具有全局性。对于语义分割和深度估计这种细粒度的任务而言, 对单一像素的预测很大程度上可以借鉴于周围的像素点, 因此当感受野较大时, 提取得到的特征更有利于对像素点的语义和深度作出更准确的预测。

### 2.3 解码器

解码器的结构如图 1 所示, 共包含两大部分。第一部分是中间两个分支, 用于捕捉语义信息与深度信息的共同

点。这两个分支分别是多尺度卷积模块 (Multi-scale Convolution Module) 分支和普通的卷积运算分支。两个分支输出的通道个数均为  $C+1$ , 其中  $C$  个通道为语义通道, 1 个通道为深度通道。由于深度特征和语义特征在一定程度上具有很大的相似性, 一般而言具有相同语义的像素也具有相近的深度值。对于车、行人等交通参与者, 其本身的深度和周围像素点之间会发生突变, 而这变化可以通过语义的边缘捕捉到。因此, 这两部分通道在分支内计算时, 相互融合, 从而达到捕捉语义信息和深度信息共同点的目的。两个分支的输出通过点积运算合并后, 再将通道拆分成两部分, 分别是包含  $C$  个通道的语义部分和包含 1 个通道的深度部分, 再参与后续的计算。

多尺度卷积模块的结构如图 3 所示。编码器的输出依次进行了三个不同尺度卷积运算, 分别是  $7 \times 7$ ,  $5 \times 5$ ,  $3 \times 3$ 。不同尺寸的卷积感受野大小不一样, 尺度大的卷积可捕捉距离较远像素间的语义和深度信息, 尺度小的卷积则可捕捉距离较近像素的语义和深度信息, 综合不同大小的卷积核, 就能解析不同尺度的特征。不同尺寸卷积提取到的信息, 均通过  $1 \times 1$  的卷积操作进行特征的整合, 而后小尺寸卷积分支的输出通过上采样与大尺度卷积分支输出相加, 将不同尺度卷积提取到的信息融合, 最终输出结果。总共有两次融合、三个上采样操作, 得到和输入相同特征图大小和通道数的输出。

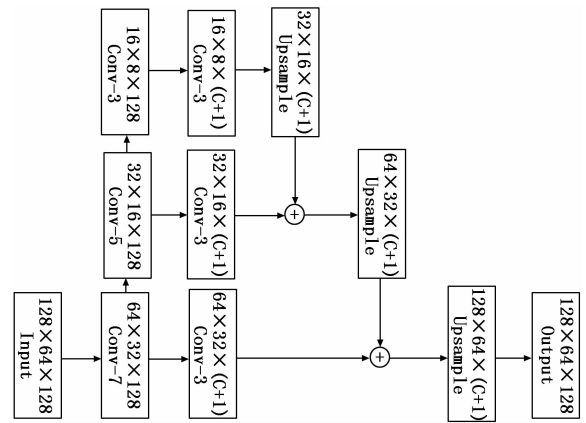


图 3 多尺度卷积模块结构图

第二部分是旁路的两个分支, 用于捕捉语义和深度各自独特的信息。分支上的操作依次是全局平均池化、 $1 \times 1$  的卷积和上采样。语义分支输出了特征图大小不变、通道数为  $C$  的语义信息, 深度分支输出了特征图大小不变、通道数为 1 的深度信息。

解码器最后融合两部分分支的输出, 即综合考虑了语义和深度的相关信息和各自提取的独特信息, 分别输出了模型的语义预测结果和深度预测结果。

### 2.4 损失函数

对于深度估计, 我们使用文献 [17] 中提出的损失函数, 定义为:

$$L_i^D = \begin{cases} |d_i|, & |d_i| \leq c \\ \frac{d_i^2 + c^2}{2c}, & |d_i| > c \end{cases} \quad (3)$$

式中,  $d_i$  是像素  $i$  预测值和实际值的差值,  $c$  为所有差值最大值的  $1/5$ , 即  $c = \max(d_i) / 5$ 。该损失函数对于差值较大的像素点, 具有较大的梯度, 使得模型在训练的时候, 对于预测结果较差的部分, 权重更新幅度更大, 更有利于模型的收敛, 加快了模型的训练。

对于语义分割, 使用的是交叉熵损失函数, 定义为:

$$L_i^S = - \sum_j y_j^i \log \hat{Y}_j^i \quad (4)$$

式中,  $N$  为语义类别总数,  $y^i$  是像素点  $i$  的类别真实值, 假设其所属类别为  $c$ , 当  $j=c$  时,  $y_j^i=1$ , 当  $j \neq c$  时,  $y_j^i=0$ 。 $\hat{Y}_j^i$  是像素点属于  $j$  类的预测概率值。对于总体的损失函数, 为深度估计损失和语义分割损失引入权重系数  $a_1$ 、 $a_2$ , 平衡二者影响, 表示如下:

$$L = \frac{1}{N} \sum_{i=1}^N (a_1 L_i^D + a_2 L_i^S) \quad (5)$$

### 3 实验结果与分析

#### 3.1 实验设置

数据集本文实验所使用的数据集为 Cityscapes 数据集<sup>[18]</sup>, 包含了欧洲 50 个城市在不同季节、不同天气条件下的街道场景。其中语义标注数据包括 19 个类别, 深度标注数据为视差值。整个数据集共包含 5 000 张图像, 训练集为 2 975 张, 验证集为 500 张, 测试集为 1 525 张。

评价指标语义分割结果采用的评价指标为平均交并比 (mIoU, mean intersection-over-union), 计算 19 个类别的  $IoU$  的平均值, 其中  $IoU$  的反映了预测区域和实际区域的重叠程度, 是实际区域和预测区域的交集比上二者的并集计算方法如下:

$$IoU = \frac{TP}{TP + FP + FN}$$

式中,  $TP$ 、 $FP$ 、 $FN$  分别表示实际为真预测也为真、实际为假预测为真、和实际为假预测为假的像素点的个数。

深度估计的评价指标采用以下指标 ( $\tilde{x}_i$  为预测值)。

$$\text{平均相对误差 (rel)}: \frac{1}{n} \sum_i \frac{|\tilde{x}_i - x_i|}{x_i};$$

$$\text{均方根误差 (rms)}: \sqrt{\frac{1}{n} \sum_i (\tilde{x}_i - x_i)^2};$$

对数空间均方根误差 (rms (log)):

$$\sqrt{\frac{1}{n} \sum_i (\log \tilde{x}_i - \log x_i)^2}$$

不同阈值下的准确率 ( $\delta$ ):  $\delta = \max\left(\frac{\tilde{x}_i}{x_i}, \frac{x_i}{\tilde{x}_i}\right)$ , 满足不同  $\delta$  取值时,  $\tilde{x}_i$  的占比。该指标反映了预测值和实际值的比值在一定范围内所占的比例, 能直观反映在不同准确度要求下预测的精度。通常情况下  $\delta=1.25, 1.25^2, 1.25^3$ 。

#### 3.2 实验结果分析

定量分析如表 2 所示, 在语义分割效果方面, 本文选取了几个在语义分割任务中具有代表性的网络作为对比, 如 SegNet、ENet、PSPNet 和 ICNet。相比于 SegNet, 本文提出的网络不仅具有更高的分割精度, 并且在计算速率上提升了四倍; 相比于 ENet, 本文提出的网络在同样达到高实时性的基础上, 在语义分割精度上提升了 10.7%; 相比于 ICNet, 语义分割的精度相近, 但参数量更少, 并在实时性上实现了翻倍; 尽管精度不及 PSPNet, 但在实时计算性能上提升了 83 倍。

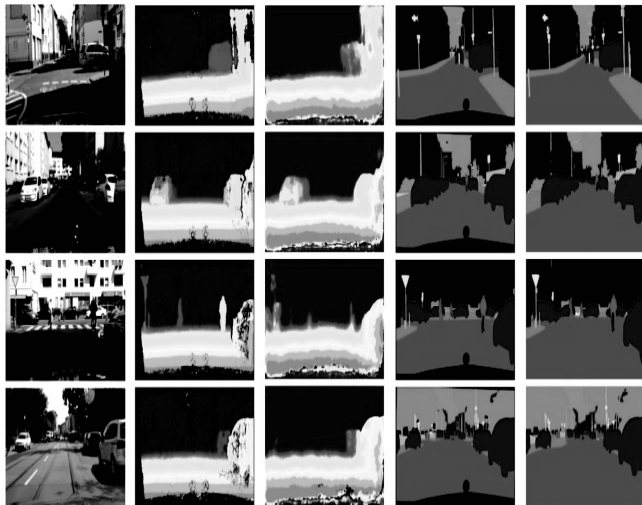
在深度估计方面, 各项评价指标均好于 DepthNet。在深度估计与语义分割同时完成方面, 本文提出的网络在语义分割和深度估计精度上均优于 HybridNet, 且能达到实时的效果。

定性分析图 4 是 CityScapes 测试集图片深度估计和语义分割预测结果的可视化。对比深度估计的真实值和预测值, 可以发现算法能很好地捕捉场景中出现的行人、车辆等使空间距离分布发生突变的物体, 这得益于网络中深度分割和语义分割共享的解码模块, 使得在深度信息预测时, 得以借鉴语义信息。对比语义分割的真实值和预测值, 可以发现网络能将路、天空、建筑、绿化等环境要素和行人、车辆、非机动车、交通标志等交通要素完整且清晰地分割出来和 ENet, 并且在成块的分割结果中不会出现其他错误的分类。

整体而言, 提出的网络在深度估计和语义分割任务中能达到良好的精度, 并且能满足高实时性的要求, 说明改进残差模块在参数量降低的情况下仍有很高的特征提取能力, 整个编码器模块具有较高的编码效率; 同时也说明了解码器能很好地解析出编码器提取的特征。特别是在语义任务中, 无论是大物体还是小物体, 都具有良好的分割精度, 说明解码器中的多尺度模块能有效解析大小不同物体的特征, 完成预测。

表 2 实验结果对比表

方法	语义分割 mIoU	深度估计						模型大小	
		$\delta=1.25^1$	$\delta=1.25^2$	$\delta=1.25^3$	ARD/%	RMSE	RMSE-log	参数量	实时性(FPS)
SegNet <sup>[19]</sup>	57	×	×	×	×	×	×	—	16.7
ENet <sup>[6]</sup>	58.3	×	×	×	×	×	×	—	76.9
PSPNet <sup>[4]</sup>	81.2	×	×	×	×	×	×	—	0.78
ICNet <sup>[7]</sup>	69.5	×	×	×	×	×	×	2.1M	30.3
DepthNet <sup>[20]</sup>	×	0.605	0.819	0.915	0.23	12.35	0.434	—	—
HybridNet A2 <sup>[21]</sup>	66.6	0.597	0.822	0.929	0.24	12.09	0.434	—	—
ours	69.0	0.713	0.887	0.94	0.21	11.05	0.423	1.2M	65



(a) 原始图像 (b) 深度真实值 (c) 深度预测值 (d) 语义真实值 (e) 语义预测值

图 4 深度估计与语义分割结果

## 4 结束语

道路行车环境的实时解析是智能驾驶的关键技术, 随着关于神经网络研究的迅速发展, 在实现单目图像的语义分割和深度估计上已经能实现一定的精度, 但仍存在模型参数多、计算量大、难以实时计算等问题, 导致无法真正落地使用。针对该问题, 本研究提出了一个轻量化、高效的特征提取模块和一个综合考虑语义信息和深度信息的特征解码模块, 在一个网络中同时完成语义分割和深度估计两个任务。在 CityScapes 数据集中, 语义分割预测结果的 mIOU 为 65.0%、深度估计结果的误差为 0.21, 并且在单个 GPU 上推断速度达到了 65 fps, 满足实时性要求。

## 参考文献:

- [1] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern recognition [C]. 2015: 3431 - 3440.
- [2] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation [A]. MICCAI [C]. 2015.
- [3] Pohlen T, Hermans A, Mathias M, et al. Full-resolution residual networks for semantic segmentation in street scenes [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. 2017: 4151 - 4160.
- [4] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network [A]. Proceedings of the IEEE conference on computer vision and pattern recognition [C]. 2017: 2881 - 2890.
- [5] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and Fully Connected CRFs [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 40 (4):

834 - 848.

- [6] Paszke A, Chaurasia A, Kim S, et al. Enet: a deep neural network architecture for real-time semantic segmentation [J]. arXiv preprint arXiv: 1606. 02147, 2016.
- [7] Zhao H, Qi X, Shen X, et al. Icnet for real-time semantic segmentation on high-resolution images [A]. Proceedings of the European Conference on Computer Vision (ECCV) [C]. 2018: 405 - 420.
- [8] Dharmasiri T, Spek A, Drummond T. Joint prediction of depths, normal and surface curvature from RGB images using CNNs [A]. 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) [C]. IEEE, 2017: 1505 - 1512.
- [9] Eigen D, Puhrsch C, Fergus R. Depth map prediction from a single image using a multi-scale deep network [A]. Advances in neural information processing systems [C]. 2014: 2366 - 2374.
- [10] Laina I, Rupprecht C, Belagiannis V, et al. Deeper depth prediction with fully convolutional residual networks [A]. 2016 Fourth international conference on 3D vision (3DV) [C]. IEEE, 2016: 239 - 248.
- [11] Liu F, Shen C, Lin G. Deep convolutional neural fields for depth estimation from a single image [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. 2015: 5162 - 5170.
- [12] Garg R, BG V K, Carneiro G, et al. Unsupervised CNN for single view depth estimation: Geometry to the rescue [A]. European Conference on Computer Vision [C]. Springer, Cham, 2016: 740 - 756.
- [13] Godard C, Mac Aodha O, Brostow G J. Unsupervised monocular depth estimation with left-right consistency [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. 2017: 270 - 279.
- [14] Kuznetsov Y, Stuckler J, Leibe B. Semi-supervised deep learning for monocular depth map prediction [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. 2017: 6647 - 6655.
- [15] Zhang Z, Cui Z, Xu C, et al. Joint task-recursive learning for semantic segmentation and depth estimation [A]. Proceedings of the European Conference on Computer Vision (ECCV) [C]. 2018: 235 - 251.
- [16] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [A]. Proceedings of the IEEE conference on computer vision and pattern recognition [C]. 2016: 770 - 778.
- [17] Laina I, Rupprecht C, Belagiannis V, et al. Deeper depth prediction with fully convolutional residual networks [A]. 2016 Fourth international conference on 3D vision (3DV) [C]. IEEE, 2016: 239 - 248.
- [18] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset

for semantic urban scene understanding [A]. Proceedings of the IEEE conference on computer vision and pattern recognition [C]. 2016: 3213 - 3223.

[19] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder - decoder architecture for image segmentation [J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39 (12): 2481 - 2495.

[20] Ivanecky B J. Depth estimation by convolutional neural networks [D]. Brno University of Technology, 2016.

[21] Lin X, Sánchez-Escobedo D, Casas J R, et al. Depth estimation and semantic segmentation from a single RGB image using a hybrid convolutional neural network [J]. Sensors, 2019, 19 (8): 1795.

(上接第 232 页)

其中每个网络都有一个主控站用来汇集网络节点采集到的数据, 主控站安装有 4G 无线数据终端, 用于将数据发送到服务器。移动设备中安装 APP 系统调用服务器数据。

### 4.2 系统应用测试

分别对系统登录功能、现场巡检功能、视频监控功能、消防管理功能和通信管理功能进行应用测试, 测试结果如表 1 所示。

表 1 系统功能测试结果

测试内容	预期结果	实际结果
正确登录	登录成功	登录成功
错误登录	无法登录	无法登录
查看 GPS 位置	显示 30 个位置	显示 28 个位置
查看现场视频	显示现场视频	显示现场视频
查看消防系统	显示消防信息	显示消防信息
建立点位	建立成功	建立成功
绑定设备	绑定成功	绑定成功

通过对 APP 系统功能的测试, 系统的各项操作结果与预期结果基本一致, 在 GPS 定位中缺少两个 GPS 信号, 其他功能都实现了测试目标, 系统功能具有一定的完整性。

对系统性能进行测试, 使用华为 P30 512GB 内存手机持续使用 APP 6 小时、12 小时、18 小时和 24 小时, CPU 占用率、图片处理器每秒刷新的帧数和内存使用结果如表 2 所示。

表 2 系统性能测试结果

时间	CPU 占用率/%	图像帧数/s	内存使用率/%
6 小时	36	60	48
12 小时	52	42	65
18 小时	65	30	72
24 小时	74	22	88

通过系统性能的测试可以看出系统性能压力较大, 在 12 小时范围内系统还可以正常操作, 到 18 小时、24 小时系统会出现卡顿、机器发热的问题, 在满足 24 小时远程管理上所设计的 APP 与实际需求还存在一定差距。

## 5 结束语

此次设计研究的系统实现了物对物、物对人的信息共

享, 结合项目基建现场施工全过程布置网络节点, 并安装各类设备, 实现对工程项目的标准化、扁平化和集约化管理, 通过及时全面的掌控现场施工人员的活动信息, 将现场施工的进度、质量与安全等置于受控状态, 对于基建工程精准化程度和现场安全管控水平的全面提升具有非常积极的现实意义和理论价值。此外, 要是实现基建现场智能化、自动化管理还需要进行大量的数据分析, 为施工决策提供指导, 在接下来的研究中还需着重对基建现场采集到的数据进行深度挖掘, 降低基建现场管理风险, 并且在系统使用性能方面还需要进一步的优化。

### 参考文献:

[1] 李超. 电力基建工程项目施工现场管理问题研究 [J]. 电子制作, 2014 (21): 243 - 243.

[2] 梁雪青. 基于互联网+的配网基建现场作业移动应用研究与建设 [J]. 现代计算机, 2017 (22): 28 - 31.

[3] 姜唯, 屈从军, 孙红燕, 等. 基建项目管理信息系统实用化考核探索研究 [J]. 中国管理信息化, 2017, 20 (14): 57 - 59.

[4] 周琴. 电力基建管理信息系统深化应用探讨 [J]. 重庆电力高等专科学校学报, 2017, 22 (3): 54 - 56.

[5] 王文娟, 刘美文. 基建档案的动态系统管理 [J]. 环球市场信息导报, 2017 (21): 111 - 111.

[6] 朴春慧, 赵静, 李建斌. 基于 UML 的工程资料管理系统分析与设计 [J]. 微计算机信息, 2008, 24 (6): 19 - 21.

[7] 孙晔. 电力基建工程 ERP 系统 [A]. 2015 电力行业信息化年会论文集 [C]. 北京: 人民邮电出版社, 2015: 143 - 145.

[8] 侯冬薇, 王美红, 陈晓云, 等. 基建信息化平台的发展探讨 [J]. 中国管理信息化, 2018, 21 (3): 153 - 155.

[9] 杨子怡. 信息系统三大能力与软件三层架构模式分析 [J]. 信息通信, 2018 (3): 143 - 144.

[10] 陈俊浩. 如何利用 3G4G 无线网 GPS 定位技术实现通信工程现场监理 [J]. 通讯世界, 2018 (2): 61 - 62.

[11] 欧阳跃军. 基于 ZigBee 的无线组网技术研究 [J]. 电信技术, 2018, 535 (S1): 37 - 39.

[12] 郭晓玲. 移动通信 4G 技术的发展与展望 [J]. 科技展望, 2015, 25 (25): 111 - 112.

[13] 王素青, 吴超. 基于 CC2530 的环境监测系统的设计与实现 [J]. 计算机测量与控制, 2015, 23 (8): 2650 - 2653.

[14] 张蓝春. 基于 Android 平台架构应用程序开发研究 [J]. 电脑编程技巧与维护, 2017 (6): 52 - 53.