

DOI:10.19651/j.cnki.emt.2209683

基于单目深度估计和校准参数的距离测算方法

余萍 胡旭欣

(华北电力大学保定 071003)

摘要: 为了提升有监督学习的单目深度估计网络对于实际场景测距任务的准确性和适用性,提出了一种基于单目深度估计和校准参数的距离测算方法。首先通过引入多元注意力模块和优化设计网络结构,构建了一种融合全局上下文和空间注意力机制的网络(GSNet),然后制定校准参数以建立场景的预测距离与实际距离的比例关系,从而获得校准后的距离值。实验证明,融合网络 GSNet 和校准参数可以有效减小单目深度估计方法在实际测算距离的误差。相比于使用单目深度估计直接预测距离信息,本文方法测算距离的平均绝对误差小于 0.15 m,平均相对误差小于 10%,具有很好的可行性和准确性。

关键词: 单目深度估计;卷积神经网络;全局上下文模块;空间注意力机制

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 520.2

Distance measurement method based on monocular depth estimation and calibration parameters

Yu Ping Hu Xuxin

(North China Electric Power University, Baoding 071003, China)

Abstract: To improve the accuracy and applicability of the monocular depth estimation network with supervised learning for actual scene ranging tasks, a distance calculation method based on monocular depth estimation and calibration mechanism is proposed. Firstly, by introducing multivariate attention blocks and optimizing the design network structure, a network integrating global context and spatial attention mechanism (GSNet) is constructed. Then calibration parameters are formulated to establish the proportional relationship between the predicted distance of the scene and the actual distance of the scene, to obtain the calibrated distance value. Experimental results show that the fusion network GSNet and calibration mechanism can effectively reduce the error of the monocular depth estimation method in the actual measured distance. Compared with other methods, the average absolute error is less than 0.15 m, and the average relative error of the measured distance in this method is less than 10%, which has good feasibility and accuracy.

Keywords: monocular depth estimation; convolutional neural network; global context block; spatial attention mechanism

0 引言

作为一种被动获取距离信息的测距方法,单目深度估计可以从一张图片或一段视频中预测得出对应场景与摄像头之间的距离信息(深度)。相比于其他测距方法,单目深度估计具有更高的经济效益和更为广阔的应用前景,但由于方法的精准度和适用度还难以满足实际需要,目前主要应用于辅助完成场景理解/重建、智能机器人和无人驾驶等多种视觉任务。因此,为了完成要求较高的实际测距任务,还需要不断地改进方法,进一步提升方法在实际场景中的预测精度。

现今最先进的研究工作主要集中于使用卷积神经网络(CNN)模型的有监督学习方法。Eigen等^[1]首次构建了含有两种不同尺度的卷积神经网络,从而高质量地完成了深度估计任务。Li等^[2]将深层卷积神经网络(DCNN)用于单目深度估计,同时使用连续条件随机场(CRF)提升算法性能。Laina等^[3]使用优化的深层卷积神经网络 ResNet 构建单目深度估计网络,并引入 berHu 函数作为训练损失函数。Gan等^[4]充分发挥多种特征的效用,将绝对特征、相对特征和竖向特征组合到端到端的 CNN 网络中。Yin等^[5]在损失函数中引入 3D 空间的几何约束以提高深度估计的预测精度。Wang等^[6]在 Yin 等方法的基础上提出

收稿日期:2022-04-20

ForeSeE 的网络架构。该架构可以在不影响背景物体深度精度的前提下,更好地预测前景物体的深度信息。Lee 等^[7]设计了一种局部平面指导层(LPG),该层可以在网络的解码阶段定位和指导图像的几何关系,最终生成高质量的深度图。Aich 等^[8]首次在单目深度估计网络中加入了双向注意力模块,该模块可以通过多个阶段的注意力权重收集特征图的全局上下文信息。刘杰平等^[9]设计了一种以通道注意力模块为主体的增强模块,该模块可以解决一部分深度图中物体边界消失问题。

在单目深度估计的实际应用中,丁晓凤等^[10]结合深度估计和目标检测技术提出了一种在实际中估计建筑物体尺寸的算法。周东兴等^[11]在截取的单目视频图像上进行像素点匹配,通过伪双目测距方法获得了较为精准的距离信息。张刘等^[12]通过构建神经网络补偿测距误差,提升了多光照条件下对接孔位置的测量精度。刘安旭等^[13]通过原卷积空间传播网络中增加下采样过程并改进训练损失函数,提升了稀疏的真实深度图与密集的真实深度图融合方法的预测效果。

通过对比多种方法的优缺点,本文将基于局部平面指导层的单目深度估计网络作为基础网络框架,并引入全局上下文和空间注意力机制以进一步提升该网络的预测性能。由于训练数据集的图像特征和表示范围均与实际场景有所区别,若直接使用有监督学习的方法在实际场景中进行测距,其测算距离会有很大的误差。对于该问题,本文通过设定校准参数对深度图中的估计距离进行校准,以减小方法在实际距离的测算误差。

1 融合全局上下文和空间注意力机制的单目深度估计网络

1.1 融合全局上下文和空间注意力机制的组合模块

对于基于局部平面指导层的单目深度估计网络,本文认为在网络中提取密集的全局上下文信息将有助于提高结果的预测精度。在对全局上下文关系建模的研究中,Cao 等^[14]提出的全局上下文模块是一种简化的非局部块,该模块可以在少量增加计算资源的同时对全局长距离的依赖关系进行建模。同时,由于全局上下文模块的特征权重图表示为一种通道之间的相互作用关系,所以全局上下文模块也可以被认为是一种特殊的通道注意力模块。

Woo 等^[15]的研究证明了在多种图像处理任务中联合使用空间注意力模块和通道注意力模块比单独使用一种模块的效果更好。因此,本文在采用全局上下文模块的同时,将空间注意力模块一同引入单目深度估计网络。根据全局上下文和空间注意力两个模块的结构特点,本文设计了一种融合全局上下文和空间注意力机制的组合模块(GC+SA blocks)。该组合模块的结构如图 1 所示。

正如图 1 所示,组合模块由全局上下文模块和空间注意力模块两个模块组成,因此该模块也同时涵盖了两种功

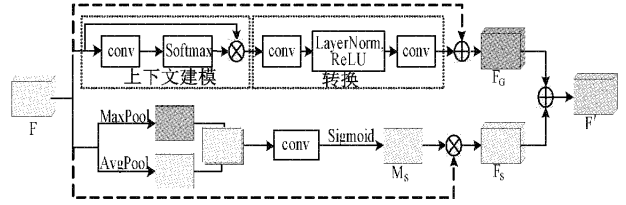


图 1 融合全局上下文和空间注意力机制的组合模块 (GC+SA blocks) 结构

能:一是对特征图像的全局长距离依赖关系进行建模,生成含有全局上下文信息的特征图;二是通过空间注意力图,有效提取和利用空间维度上的重要特征。

在组合模块中,输入特征图 F 首先分别进入全局上下文模块和空间注意力模块,输出特征图 F_G 和特征图 F_S ,然后再对特征图 F_G 和 F_S 进行求和操作,将对位置上的特征进行聚合,最后输出特征图 F' 。其执行过程用公式表示为:

$$z_i = x_i + W_2 \text{ReLU}(\text{LN}(W_1 \sum_{j=1}^{N_s} \frac{e^{w_k r_j}}{\sum_{m=1}^{N_s} e^{w_k r_m}} x_j)) +$$

$$[x_i \times \text{Sigmoid}(W^{3 \times 3}([\text{AvgPool}(F); \text{MaxPool}(F)]))] \quad (1)$$

式中: x 为特征图 F 上的特征值; i, j 和 m 表示为特征图 F 上的不同点位; N_s 为特征图 F 的最大点位数; W_k 为上下文关系建模部分中卷积操作的权重矩阵; W_1 为特征转换部分中第一个卷积操作的权重矩阵; W_2 为特征转换部分中第二个卷积操作的权重矩阵; $W^{3 \times 3}$ 表示为空间注意力模块中 3×3 卷积的权重矩阵。

1.2 融合网络的网络结构

本文将基于局部平面指导层的卷积神经网络作为研究的主体网络,并进一步优化其网络结构,使融合多元注意力机制的组合模块可以激发全部潜力,最大限度地提升网络性能。本文假设输入图像的分辨率为 S ,原网络解码器有 3 个不同的尺度阶段,即分辨率分别为 $S/8, S/4$ 和 $S/2$ 的解码阶段。通过观察和分析原网络的组成结构,本文提出在原网络解码器的多个不同尺度阶段放置融合多元注意力机制的组合模块,使组合模块可以提取不同尺度的有效特征。

根据数量和位置的不同,组合模块可以有七种不同的放置方法。其中,在解码器分辨率为 $S/8$ 和 $S/4$ 尺度上放置组合模块(GC+SA blocks)的方法为最优的模块放置方法。本文使用这种最优方法和组合模块共同构建了融合全局上下文和空间注意力机制的单目深度估计网络(GSNet)。GSNet 的网络结构如图 2 所示。

1.3 训练损失函数

尺度不变误差是单目深度估计常用的一种训练损失函数,其表示为:

$$D(g) = \frac{1}{N} \sum_i g_i^2 - \frac{\lambda}{N^2} (\sum_i g_i)^2 \quad (2)$$

式中: $g_i = \log \bar{d}_i - \log d_i$, d_i 为像素点 i 的真实深度, \bar{d}_i 为

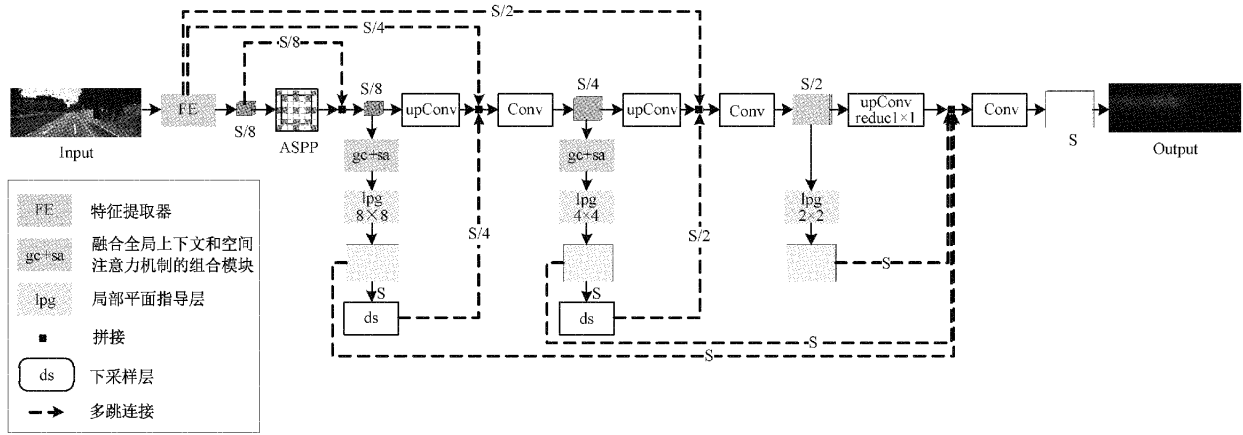


图 2 融合网络 GSNet 的网络结构

像素点 i 的估计深度; λ 设置为 0.5; N 代表有效且真实的深度值的像素数量。

因为设置更高的 λ 将强制使误差方差更小, 所以本文将 λ 设置为 0.85。同时, 适当的扩展损耗函数的范围可提高收敛速度和最终训练的结果。因此, 本文使用的训练损失函数 L 定义为:

$$L = \alpha \sqrt{D(g)} \quad (3)$$

式中: α 是一个常量, 在所有实验中都设置为 10。

1.4 实验评价指标

单目深度估计常用的七项评价指标定义如下:

尺度不变误差 (SIlog):

$$\frac{1}{N} \sum_i (\log \bar{d}_i - \log d_i)^2 - \frac{1}{N^2} \left(\sum_i \log \bar{d}_i - \log d_i \right)^2$$

$$\delta_k: \% \text{ of } \bar{d}_i, s. t. \max\left(\frac{\bar{d}_i}{d_i}, \frac{d_i}{\bar{d}_i}\right) = \delta_k < thr$$

当 thr 分别等于 1.25, 1.25² 和 1.25³ 时, k 分别等于 1, 2 和 3。

绝对相对误差 (Abs Rel): $\frac{1}{N} \sum_{\bar{d} \in N} |\bar{d} - d|$

平方相对误差 (Sq Rel): $\frac{1}{N} \sum_{\bar{d} \in N} \|\bar{d} - d\|^2 / d$

均方根误差 (RMSE): $\sqrt{\frac{1}{N} \sum_{\bar{d} \in N} \|\bar{d} - d\|^2}$

均方根对数误差 (RMSElog):

$$\sqrt{\frac{1}{N} \sum_{\bar{d} \in N} \|\log \bar{d} - \log d\|^2}$$

$$\log 10: \frac{1}{N} \sum_{\bar{d} \in N} \|\log_{10} \bar{d} - \log_{10} d\|^2$$

因为单目深度估计的评价指标较多, 不同的改进算法有可能使某些指标提升, 而某些指标下降。因此本文制定了两种综合性的评价指标: 精度和 (Accuracy Sum) 及误差和 (Error Sum), 以此综合评估算法的改进效果。

本文将精度和 (Accuracy Sum) 定义为三种阈值精度

的和值, 表示为:

$$\text{Accuracy Sum} = \delta_1 + \delta_2 + \delta_3 \quad (4)$$

本文将误差和 (Error Sum) 定义为各数据集常用的除阈值精度之外的相关误差评价指标的和值, 其中 KITTI 深度数据集的误差和 (Error Sum) 表示为:

$$\text{Error Sum}_{\text{KITTI}} = \text{Abs Rel} + \text{Sq Rel} + \text{RMSE} + \text{RMSElog} \quad (5)$$

将 NYU Depth V2 数据集的误差和 (Error Sum) 表示为:

$$\text{Error Sum}_{\text{NYU}} = \text{Abs Rel} + \text{RMSE} + \log 10 \quad (6)$$

1.5 融合网络 GSNet 与其他单目深度估计网络的实验结果

本文将融合全局上下文和空间注意力机制的组合模块与最优的网络结构结合起来, 设计完成了融合全局上下文和空间注意力机制的单目深度估计网络 (GSNet), 并将融合网络 GSNet 与其他主流方法的网络进行比较。多种网络在 KITTI 深度数据集上的实验结果如表 1 所示, 在 NYU Depth V2 数据集上的实验结果如表 2 所示。

由表 1 可以看出, GSNet 的所有评价指标均为表中的最优数值。相比于 Lee 等提出的网络, GSNet 的 5 个主要评价指标 δ_1 、 δ_3 、Sq Rel、RMSE 和 RMSElog 均有所改善, GSNet 的精度和 (Accuracy Sum) 在可提升的范围内提高了 3.77%, 误差和 (Error Sum) 在可降低的范围内降低了 1.68%。

在表 2 中, 相比于 Lee 等提出的网络, GSNet 评价指标中的 δ_2 和 δ_3 指标有明显的提升, Abs Rel 指标有小幅改善, GSNet 的精度和 (Accuracy Sum) 在可提升的范围内提高了 2.1%, 误差和 (Error Sum) 保持不变。

除了通过评价指标评估预测结果, 本文也从深度图的角度判断网络的预测性能。在 KITTI 深度测试集上, 测试图如图 3(a) 所示, Lee 等人提出的单目深度估计网络和 GSNet 网络经过灰度渲染后的深度图分别如图 3(b) 和图 3(c) 所示。

表1 多种单目深度估计网络在KITTI深度数据集上的实验结果

方法	编码器	$\delta_1 \uparrow$	$\delta_2 \uparrow$	$\delta_3 \uparrow$	Accuracy sum \uparrow	Abs Rel \downarrow	Sq Rel \downarrow	RMSE \downarrow	RMSE $\log \downarrow$	Error Sum \downarrow
文献[1]	—	0.702	0.898	0.967	2.567	0.203	1.548	6.307	0.282	8.34
文献[4]	ResNet50	0.890	0.964	0.985	2.839	0.098	0.666	3.933	0.173	4.87
文献[5]	ResNeXt101	0.938	0.990	0.998	2.926	0.072	—	3.258	0.117	—
文献[7]	ResNeXt101	0.956	0.993	0.998	2.947	0.059	0.241	2.756	0.096	3.152
GSNet	ResNeXt101	0.957	0.993	0.999	2.949	0.059	0.236	2.710	0.094	3.099

表2 多种单目深度估计网络在NYU Depth V2数据集上的实验结果

方法	$\delta_1 \uparrow$	$\delta_2 \uparrow$	$\delta_3 \uparrow$	Accuracy Sum \uparrow	Abs Rel \downarrow	RMSE \downarrow	$\log_{10} \downarrow$	Error Sum \downarrow
文献[1]	0.769	0.950	0.988	2.707	0.158	0.641	—	—
文献[5]	0.875	0.976	0.994	2.845	0.108	0.416	0.048	0.572
文献[7](DenseNet161)	0.885	0.978	0.994	2.857	0.110	0.392	0.047	0.549
GSNet(DenseNet161)	0.885	0.980	0.995	2.860	0.109	0.393	0.047	0.549

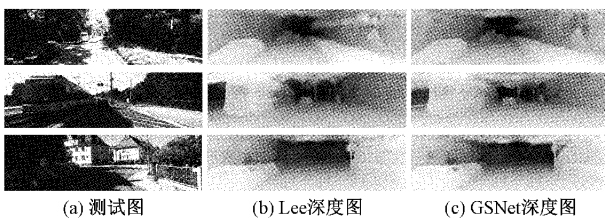


图3 KITTI测试图和两个网络经过灰度渲染后的深度图

在NYU Depth V2测试集上,测试图如图4(a)所示, Lee等提出的单目深度估计网络和GSNet网络经过灰度渲染后的深度图分别如图4(b)和(c)所示。

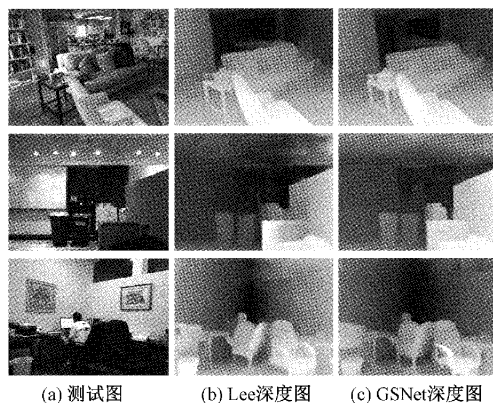


图4 NYU测试图和两个网络经过灰度渲染后的深度图

通过图3和4可以看出,相比于Lee等人的深度结果,GSNet深度图上的物体轮廓更为流畅,近景细节更加清晰,表面一致性和连贯性更好,但也可以看出GSNet容易忽略连通区域内具有相似特征的物体。

从GSNet在两个数据集上的实验结果可知,本文提出的组合模块和优化的网络结构具有有效性,融合网络

GSNet的性能超过了原网络的性能表现。深度图的对比结果也证明了GSNet具有更为优秀的深度估计性能,可以进一步提升深度估计结果的清晰度和准确度。

2 基于单目深度估计和校准参数的距离测算方法

2.1 实际距离测算方法

在实际应用方面,本文根据电力工程测距的需要,提出一种可视化监管功能的构想,即在周围有带电设备的场景下,安全监管人员可以通过传回的图像数据判断施工人员是否违规靠近了带电设备或者已接近与带电设备的安全距离。

由于深度数据集表示的距离范围通常是固定的,但是在实际现场拍摄到的照片可以表示的距离范围是不一定的。对于这种情况,再加上数据集本身的局限性,直接由单目深度估计网络预测的距离信息会存在较大误差,需要找到一种方法尽量去校准估计距离值以减小误差。

在对单目深度估计网络的研究过程中,可以发现网络生成深度信息的过程是先通过多层卷积神经网络用Sigmoid函数得到图中每个像素点位上的深度权重,再将数据集可以表示的最大深度值乘以图上各点的权重值,从而生成各点对应场景的距离信息。因此,根据深度权重和最大深度值的乘法关系,可以在实际应用中安排一个已知实际测量距离的参考点,然后通过实际测量距离和估计距离值计算校准参数 k ,再利用该参数校准其他点的估计值。该校准参数定义为:

$$k = \frac{d_0}{\bar{d}_0} \quad (7)$$

式中: k 为校准参数; d_0 表示为参考点对应的实际距离值;

\bar{d}_0 表示为参考点在深度图中的估计距离值。

在得到相关点位的距离值后,为了衡量方法的测距精度,使用多个评价指标包括绝对误差、相对误差、平均绝对误差和平均相对误差对方法进行评估。

绝对误差是指图中像素点对应的实际距离值与通过方法测算的距离值的差值的绝对值,定义为:

$$e_{AE} = |l - \tilde{l}| \quad (8)$$

式中: l 为对应场景点位与相机成像中心的实际距离值; \tilde{l} 为通过方法测算得到的距离值。

相对误差是指图中像素点对应的实际距离值与通过方法测算的距离值的绝对误差占实际距离值的百分比,定义为:

$$e_{RE} = \frac{|l - \tilde{l}|}{l} \times 100\% \quad (9)$$

平均绝对误差是求图中所有测距点位上的绝对误差的平均值,定义为:

$$e_{MAE} = \frac{1}{N} \sum_{i \in N} |l_i - \tilde{l}_i| \quad (10)$$

式中: N 为图中测距点位的数量; l_i 为场景点位 i 与相机成像中心的实际距离值; \tilde{l}_i 为通过方法测算得到的场景点位 i 与相机成像中心的距离值。

平均相对误差是求图上所有测距点位的相对误差的平均值,定义为:

$$e_{MRE} = \frac{1}{N} \sum_{i \in N} \frac{|l_i - \tilde{l}_i|}{l_i} \times 100\% \quad (11)$$

综上所述,基于单目深度估计和多元注意力机制的距离测算方法的实现流程如图 5 所示,具体步骤为:

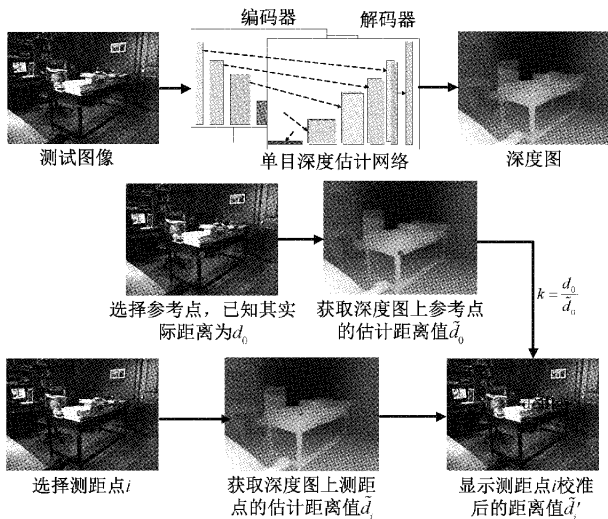


图 5 实际测算方法流程

1) 在室内数据集中训练融合网络 GSNet,在训练好的网络上测试实景照片,并获取对应的深度图。

2) 在终端中打开照片,在照片中确定一个参考点,该

点可以选用现实中较好测量的就近点位,测量该点与摄像头的实际距离值。然后通过深度图获取该点位的估计距离值,再通过式(7)计算校准参数 k 。

3) 在图片中继续选择测距点,将对应点位从深度图上获取到的估计距离值乘以校准参数,计算校准后的距离值,并在图片的相应点位上显示校准后的距离值。

4) 计算各点位距离测算的绝对误差、相对误差、平均绝对误差和平均相对误差,评估方法在实际场景中的测距性能。

2.2 实际距离测算实验

为了验证本文提出的距离测算方法是否可以提高实际测距的精度,首先在室内场地使用手机拍摄了多张带有 10 个测距图标的测试图像,用激光测距仪测量了各个图标与拍摄位置之间的距离,并进行了记录。

然后将测试图像分别使用无校准参数、有校准参数未使用融合网络以及有校准参数使用融合网络的距离测算方法进行了实验。实拍测试图像如图 6(a)所示,对应的深度图如图 6(b)所示,距离测算方法生成的结果图像如图 6(c)所示。

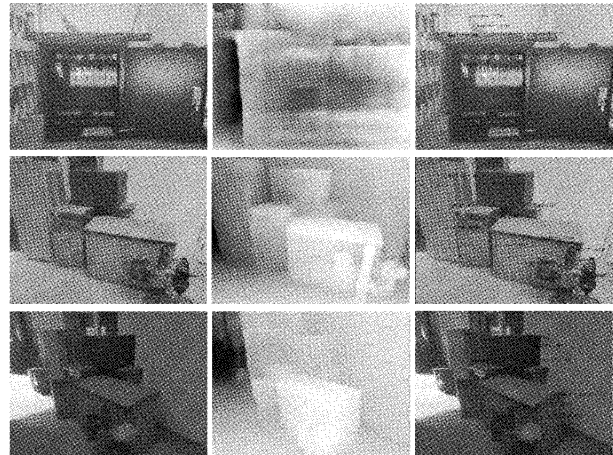


图 6 测试图像、深度图与距离测算方法生成的图像

根据实验图像,将图中各测距图标的测算距离和用激光测距仪测量的实际距离列入表中,并计算方法的评价指标。以第一张实拍测试图为例,10 个测距点位的实际测量距离值和通过三种方法得到的测算距离值如表 3 所示。其中,方法 1 为无校准参数方法,方法 2 为未使用融合网络的有校准参数方法,方法 3 为使用融合网络 GSNet 的有校准参数方法。

根据表 3 中的实验数据,以第一张实验图为例,计算多个方法的评价指标,其结果如表 4 所示。

根据多种距离测算方法的实验数据,比较无校准参数方法、未使用融合网络的有校准参数方法和使用融合网络的有校准参数方法在测试图像 10 个测距点位上的平均误差,其平均绝对误差和平均相对误差分别如图 7 和 8 所示。

表3 实拍测试图像的距离测算结果 m

测距点	测量值	方法1	方法2	方法3
		测算值	测算值	测算值
1	1.46	2.53	1.46	1.46
2	1.53	2.3	1.34	1.34
3	1.43	2.55	1.36	1.47
4	1.42	2.65	1.64	1.53
5	1.49	2.36	1.78	1.39
6	1.63	2.79	1.64	1.62
7	1.58	2.81	1.74	1.63
8	1.51	2.21	1.59	1.26
9	1.72	2.56	1.41	1.47
10	1.7	2.8	1.53	1.57

表4 多种距离测算方法的绝对误差和相对误差

测距点	绝对误差/m			相对误差/%		
	方法1	方法2	方法3	方法1	方法2	方法3
1	1.07	0	0	73.29	0	0
2	0.77	0.19	0.19	50.33	12.42	12.42
3	1.12	0.07	0.04	78.32	4.90	2.80
4	1.23	0.22	0.11	86.62	15.49	7.75
5	0.87	0.29	0.1	58.39	19.46	6.71
6	1.16	0.01	0.01	71.17	0.61	0.61
7	1.23	0.16	0.05	77.85	10.13	3.16
8	0.7	0.08	0.25	46.36	5.30	16.56
9	0.84	0.31	0.25	48.84	18.02	14.53
10	1.10	0.17	0.13	64.71	10.00	7.65

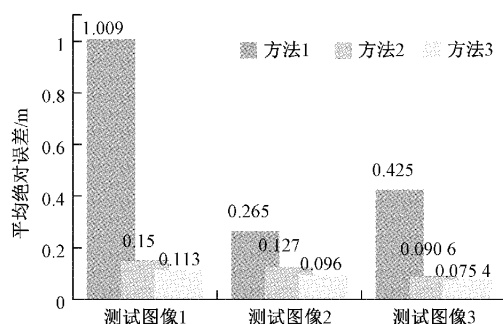


图7 3种方法在测试图像上的平均绝对误差

由以上实验结果可知,在测试图像中,有校准参数方法的误差均明显小于无校准参数方法。在有校准参数方法中,使用融合网络的方法在多个点位上的误差略低于未使用融合网络的方法。相比于使用单目深度估计直接预测距离信息,使用融合网络的有校准方法测算距离的平均绝对误差小于0.15 m,平均相对误差小于10%,具有很好的可行性和准确性。

由此可以证明,校准参数和融合网络对于基于单目深

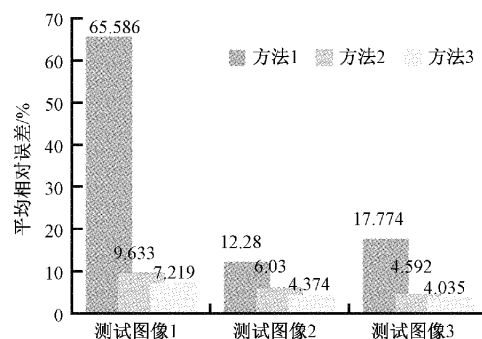


图8 三种方法在测试图像上的平均相对误差

度估计的距离测算方法是有效的,校准参数可以大幅减小测距结果的误差,提升方法的测算精度。在测算误差大大降低后,该距离测算方法可以应用于测量佩戴摄像头的人员与带电物体的距离,在添加报警程序后,可以在人员接近安全距离的阈值时发出告警,从而防止事故的发生。

3 结 论

单目深度估计在实际应用中存在两项问题,一是预测精度较低,难以满足高精度的测距需求;二是适用性较低,训练集与实际图像难以精准匹配。本文通过构建融合全局上下文和空间注意力机制的网络(GSNet)进一步提升了单目深度估计网络的预测性能,然后通过设置参数校准了深度图像表示的距离范围,进一步降低了实际场景的测算误差。

在KITTI深度数据集和NYU Depth V2数据集中,GSNet网络的评价指标更为优秀,由此证明了GSNet具有更好的深度预测性能。在实际场景的测试中,使用校准参数和融合网络GSNet的距离测算方法的精度也是最佳的,由此证明了本文提出的距离测算方法的有效性。但同时也必须指出,本文提出的距离测算方法受到网络性能、数据集和点位选择等多重因素的影响,具有一些不稳定因素,仍然需要继续探索和研究。下一步将继续改进单目深度估计的网络结构以提升深度估计的预测性能,并改进图像点位的选择方法,剔除一些异常点位,以进一步减小方法的测算误差。

参考文献

- [1] EIGEN D, PUHRSCHE C, FERGUS R. Depth map prediction from a single image using a multi-scale deep network[C]. Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal: MIT Press, 2014: 2366-2374.
- [2] LI B, SHEN C, DAI Y, et al. Depth and surface normal estimation from monocular images using regression on deep features and hierarchical CRFs[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015:

- 1119-1127.
- [3] LAINA I, RUPPRECHT C, BELAGIANNIS V, et al. Deeper depth prediction with fully convolutional residual networks [C]. 2016 Fourth International Conference on 3D Vision (3DV), Stanford: IEEE, 2016: 239-248.
- [4] GAN Y, XU X, SUN W, et al. Monocular depth estimation with affinity, vertical pooling, and label enhancement [C]. Proceedings of the European Conference on Computer Vision (ECCV), Munich: Springer, 2018: 224-239.
- [5] YIN W, LIU Y, SHEN C, et al. Enforcing geometric constraints of virtual normal for depth prediction[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul: IEEE, 2019: 5684-5693.
- [6] WANG X, YIN W, KOND T, et al. Task-aware monocular depth estimation for 3D object detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12257-12264.
- [7] LEE J H, HAN M, KO D W, et al. From big to small: Multi-scale local planar guidance for monocular depth estimation [DB/OL]. (2019-07-24) [2022-05-11]. <https://arxiv.org/abs/1907.10326>.
- [8] AICH S, VIANNEY J M U, ISLAM M A, et al. Bidirectional attention network for monocular depth estimation[C]. 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an: IEEE, 2021: 11746-11752.
- [9] 刘杰平, 温峻文, 梁亚玲. 基于多尺度注意力导向网络的单目图像深度估计[J]. 华南理工大学学报:自然科学版, 2020, 48(12):52-62.
- [10] 丁晓凤, 熊赅晖, 林祖轩. 基于参照物和深度估计的建筑物体尺度估算[J]. 现代计算机, 2021(10): 99-107.
- [11] 周东兴, 韩东升. 基于单目视频的伪双目动态测距方法[J]. 电子测量技术, 2021, 44(3): 38-44.
- [12] 张刘, 吴广栋, 吴坤洋, 等. 双波段、双通道对接系统测距误差补偿方法研究[J]. 电子测量与仪器学报, 2021, 35(4): 73-81.
- [13] 刘安旭, 黎向锋, 刘晋川, 等. 改进卷积空间传播网络的单目图像深度估计[J]. 电子测量技术, 2021, 44(23): 78-85.
- [14] CAO Y, XU J, LIN S, et al. GCNet: Non-local networks meet squeeze-excitation networks and beyond [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Seoul: IEEE, 2019: 1971-1980.
- [15] WOO S, PARK J, LEE J, et al. CBAM: Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision(ECCV), Munich: Springer, 2018: 3-19.

作者简介

余萍, 副教授, 主要研究方向为电力系统通信、电力物联网和图像处理等。

E-mail: well_yp@sina.com

胡旭欣, 硕士研究生, 主要研究方向为单目深度估计。