



# 智能系统学报

CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

## 基于迁移学习的移动机器人单帧图像坡度检测算法

辛菁, 杜柯楠, 王媛媛, 刘丁

引用本文:

辛菁, 杜柯楠, 王媛媛, 等. 基于迁移学习的移动机器人单帧图像坡度检测算法[J]. *智能系统学报*, 2021, 16(1): 81-91.

XIN Jing, DU Kenan, Wang Yuanyuan, et al. Single frame image slope detection algorithm for mobile robots based on transfer learning[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(1): 81-91.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202009009>

## 您可能感兴趣的其他文章

### 基于LiDAR/INS的野外移动机器人组合导航方法

Integrated navigation approach for the field mobile robot based on LiDAR/INS

智能系统学报. 2020, 15(4): 804-810 <https://dx.doi.org/10.11992/tis.202008026>

### 多移动机器人协同搬运技术综述

Technologies for cooperative transportation by multiple mobile robots

智能系统学报. 2019, 14(1): 20-27 <https://dx.doi.org/10.11992/tis.201801038>

### 基于Kinect的改进移动机器人视觉SLAM

Improved V-SLAM for mobile robots based on Kinect

智能系统学报. 2018, 13(5): 734-740 <https://dx.doi.org/10.11992/tis.201705018>

### 基于图优化的移动机器人视觉SLAM

Visual-SLAM for mobile robot based on graph optimization

智能系统学报. 2018, 13(2): 290-295 <https://dx.doi.org/10.11992/tis.201612004>

### 移动机器人全覆盖信度函数路径规划算法

Complete-coverage path planning algorithm of mobile robot based on belief function

智能系统学报. 2018, 13(2): 314-321 <https://dx.doi.org/10.11992/tis.201610006>

### 一种基于P稳定局部敏感哈希算法的相似人脸检索系统设计

A similar-face-image-retrieval system design based on a P-stable locality-sensitive Hashing algorithm

智能系统学报. 2017, 12(3): 392-396 <https://dx.doi.org/10.11992/tis.201607005>

微信公众平台



关注微信公众号, 获取更多资讯信息

DOI: 10.11992/tis.202009009

# 基于迁移学习的移动机器人单帧图像坡度检测算法

辛菁, 杜柯楠, 王媛媛, 刘丁

(西安理工大学自动化与信息工程学院, 陕西 西安 710048)

**摘要:** 针对未知环境下移动机器人平稳上坡控制对坡度感知精度的要求, 本文提出了一种基于迁移学习的移动机器人单帧图像坡度检测算法。利用室内图像标准数据集训练深度卷积神经网络——全连接超像素池化网络 (deep convolutional neural field-fully connected superpixel pooling network, DCNF-FCSP) 并获得室内单帧图像深度估计网络模型; 将 DCNF-FCSP 模型中前 5 个图像特征提取层的网络参数迁移至室外图像深度估计网络中; 固定室外图像深度估计网络中图像特征提取部分的网络参数, 利用室外图像数据集对剩余 5 层的网络参数进行训练, 从而得到室外单帧图像深度估计网络; 将其应用在移动机器人坡度检测中, 根据单帧斜坡图像估计出斜坡角度。标准数据集和实际场景的深度估计和坡度检测实验表明: 本文所提出的基于迁移学习的移动机器人单帧图像坡度检测算法能够仅根据移动机器人车载相机采集的单帧斜坡 RGB 图像就可估计出精确的斜坡角度, 满足移动机器人在未知环境中对坡度感知精度的要求。

**关键词:** 未知环境; 移动机器人; 坡度检测; 室外; 单帧图像; 深度估计; 迁移学习; 深度卷积网络

**中图分类号:** TP391   **文献标志码:** A   **文章编号:** 1673-4785(2021)01-0081-11

中文引用格式: 辛菁, 杜柯楠, 王媛媛, 等. 基于迁移学习的移动机器人单帧图像坡度检测算法 [J]. 智能系统学报, 2021, 16(1): 81-91.

英文引用格式: XIN Jing, DU Kenan, Wang Yuanyuan, et al. Single frame image slope detection algorithm for mobile robots based on transfer learning[J]. CAAI transactions on intelligent systems, 2021, 16(1): 81-91.

## Single frame image slope detection algorithm for mobile robots based on transfer learning

XIN Jing, DU Kenan, Wang Yuanyuan, LIU Ding

(School of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048, China)

**Abstract:** To meet the requirement of slope perception accuracy for stable uphill control of mobile robots in an unknown environment, a single frame image slope detection algorithm for mobile robots is proposed in this paper based on transfer learning. First, the deep convolutional neural field-fully connected superpixel-pooling network (DCNF-FCSP) is trained using a standard indoor image dataset, and the depth estimation network model of indoor single frame images is obtained. Second, the network parameters of the first five image feature extraction layers in the DCNF-FCSP model are transferred to the outdoor image depth estimation network. Then, the network parameters of the image feature extraction part in the outdoor image depth estimation network are fixed, and the network parameters of the remaining five layers are trained using the outdoor image dataset; thus the outdoor single frame image depth estimation network is obtained. Finally, it is applied to the slope detection of a mobile robot, and the slope angle is estimated according to the single frame slope image. The depth estimation and slope detection experiments on a standard dataset and in actual scenes show that the proposed algorithm can estimate the accurate slope angle according to only a single frame slope RGB image captured by the vehicle-mounted camera of a mobile robot. The proposed algorithm meets the requirements of the slope perception accuracy of a mobile robot in an unknown environment.

**Keywords:** unknown environment; mobile robot; slope detection; outdoor; single image; depth estimation; transfer learning; deep convolutional network

收稿日期: 2020-09-01.

基金项目: 国家自然科学基金项目 (61873200, 61833013, U20A20225).

通信作者: 刘丁. E-mail: liud@xaut.edu.cn.

未知环境地形的自主感知是移动机器人自主导航的基础和根本保证, 也是移动机器人进行其他行为决策的前提。坡度检测作为移动机器人自

主感知未知环境地形的重要组成部分,其检测精度影响移动机器人在爬坡时的行为策略和方向,进而影响整个导航系统的性能。因此,坡度检测精度对移动机器人自主导航性能十分重要<sup>[1]</sup>。

惯性测量单元 (inertial measurement units IMU, IMUs)、倾斜传感器、激光传感器、超声波传感器和视觉传感器是坡度检测系统常用的传感器。与 IMUs 和倾斜传感器不同<sup>[2-3]</sup>,激光传感器<sup>[4-5]</sup>、超声波传感器<sup>[6]</sup>和视觉传感器在检测坡度中不需要移动机器人接触或攀爬斜坡,因而被更多地应用在移动机器人坡度检测与环境地形感知中。近年来,随着计算机视觉技术,尤其是单帧图像深度估计技术的快速发展,以及图像本身所包含的丰富语义信息,使得视觉传感器在机器人环境智能感知中得到了广泛的应用。基于此,本文提出了一种基于迁移学习的移动机器人单帧图像坡度检测算法,不需要移动机器人接触或攀爬斜坡,仅根据单帧斜坡 RGB 图像就能估计出深度信息进而检测出斜坡的角度,同时所提算法大大减少了单帧图像深度估计网络训练时对室外图像样本数量的要求,提高了网络的训练效率和估计精度,显著提高了移动机器人在未知环境中估计斜坡角度的精确度,为移动机器人自主导航提供了更加可靠、准确的依据。

## 1 单帧图像深度估计方法

图像的深度估计是指对构成二维图像的每个像素点对应的深度值进行估计,其中深度值指的是相机所拍摄图像中每个像素点所对应的真实场景中 3D 物体到相机光心所在平面的垂直距离。准确有效的深度信息能提升图像分类、目标识别、语义分割等诸多计算机视觉任务的结果<sup>[7]</sup>。随着近年来 3D 技术的广泛应用和现有普通 3D 传感器精度受限的影响,从 2D 图像深度估计信息来获取环境的 3D 结构成为计算机视觉研究领域的热点问题。对于单帧图像,由于难以获取目标场景的视差信息,只能通过图像的自身特征及先验知识来完成深度估计,因而具有较大的挑战性。目前大多数关于图像深度估计的研究都基于多帧图像进行,典型的方法是采用立体视觉技术。整个过程类似于人眼的双目视觉成像机理,这种技术相对比较成熟,但该类方法获取环境信息成本较高<sup>[8]</sup>,还需要图像中具有较明显的辅助信息,且重建结果易受到物体遮挡和对应关系匹配误差的影响。因此,使用简洁的单帧图像深度估计获取环境信息成为三维重建研究中一项重要

的研究内容<sup>[9]</sup>。然而由于室外环境复杂多变,且干扰因素较多,根据单帧图像推断室外环境信息一直是三维重建中的一个挑战性问题。

目前市场上已存在多种用于获取室外场景深度信息的传感器,如深度相机、激光雷达等。其中,激光雷达是无人驾驶领域最常用的距离测量传感器,具有高精度、便于携带且抗干扰等优点,然而其结构复杂且成本较高,难以广泛应用于工业。传统的单帧图像深度估计方法通常依赖光学几何成像原理等,其难点在于几何分析模型很难从单帧图像中获取足够用于推测场景三维信息的相关特征。

随着深度学习在图像感知领域的快速发展,研究者们已经将多种深度学习模型成功应用到图像处理中,并根据图像特征信息来对图像的深度信息进行估计。本质上,基于深度学习的单帧图像深度估计方法的核心是构造并训练一个将二维 RGB 图像转换成深度图像的深度网络<sup>[10]</sup>。常用的深度学习模型是卷积神经网络 (convolutional neural network, CNN),已提出一系列基于 CNN 的单帧图像深度估计算法<sup>[10-19]</sup>。Eigen 等<sup>[11]</sup>提出一种多尺度 CNN 来实现单帧图像深度估计,该方法通过粗尺度 (coarse) 与精尺度 (fine-scale) 2 个方面分别考虑图像的全局以及局部信息,相较于直接应用 CNN 网络估计深度值的精确度有了较为明显的提升,其不足之处是图像深度估计效率较低。为此, Liu 等<sup>[12]</sup>提出深度卷积神经网络——全连接超像素池化网络 (deep convolutional neural field-fully connected superpixel pooling network, DCNF-FCSP) 模型,将卷积神经网络与条件随机场 (conditional random fields, CRF) 结合在一起对单帧图像深度估计网络进行优化,将图像像素点的深度估计转换为像素块的深度估计,在不影响深度估计精度的前提下提高了估计速度。单帧图像深度估计的性能依赖于大量和多样性的图像训练数据集<sup>[13]</sup>。然而,由于室外图像深度估计缺乏足够的室外图像数据集作为支撑,深度估计结果较为不理想。随后,研究者从不角度提出了进一步提高估计精度的方法,如采用多个深度网络<sup>[14-17]</sup>, Hu 等<sup>[14]</sup>采用图像深度估计基础网络得到单帧图像粗略的深度估计结果,再通过精细网络对深度估计结果进行优化,并在损失函数中加入对象边缘的深度梯度来优化对象边界的深度估计结果,提高了估计精度,但该方法的图像深度估计效率比较低,不能满足环境三维重建对实时性的要求。Kuznetsov 等<sup>[15]</sup>采用半监督学习方式对图像特征进行学习,以减少训练深度卷积神经网络所

需要的训练样本。Fu 等<sup>[16]</sup> 基于空间和深度分辨率的折衷提出了一种回归分类级联网络。梁煜等<sup>[17]</sup> 将全卷积网络、挤压-激励网络 (squeeze-and-excitation networks, SENet)、残差网络 (residual neural network, ResNet) 三者结合提出了一种多网络构成的编码器解码器结构, 提高了单帧图像深度估计的精度。此外, 还有研究者通过设计新的损失函数<sup>[10, 18-19]</sup>, 在保证深度估计精度的同时, 提高了估计的实时性。

综上, 已有的单幅图像深度估计方法可以获得比较精确的深度估计结果, 然而其需要大量的图像数据集, 仍不能很好地解决室外单幅图像深度估计中训练数据有限且深度估计精度不理想的问题, 基于此, 受迁移学习在计算机视觉、机器学习等领域的成功应用, 本文提出了一种基于迁移学习的室外单帧图像深度估计方法, 旨在能够在节省训练样本的同时提高深度估计结果的精确度, 在此基础上进一步提出了一种改进的移动机器人坡度检测算法, 能够根据移动机器人车载相机采集的单帧斜坡 RGB 图像就可估计出精确的斜坡角度。

本文的主要贡献如下:

1) 提出了一种基于迁移学习的室外单帧图像深度估计方法, 不仅在室外图像标准数据集上取得了精确的深度估计结果, 而且对于实际拍摄的室外单帧图像也能进行准确的深度估计;

2) 将基于迁移学习的室外单帧图像深度估计方法与基于视觉的机器人坡度检测方法相结合, 提出了一种改进的移动机器人单帧图像坡度检测算法, 能够仅根据移动机器人车载相机采集的单帧斜坡 RGB 图像就可估计出精确的斜坡角度, 满足移动机器人在未知环境中对坡度感知精度的要求。

## 2 基于迁移学习的移动机器人单帧图像坡度检测算法原理

### 2.1 算法整体框架

本文提出的基于迁移学习的移动机器人单帧图像坡度检测算法整体如图 1 所示, 主要由斜坡边缘提取、斜坡边缘像素点坐标转换和斜坡图像深度估计 3 个部分组成。首先, 对 Kinect 传感器采集到的斜坡 RGB 图像进行滤波和边缘提取, 以获得斜坡图像上下边缘像素点的坐标; 其次, 将斜坡图像边缘像素点坐标由图像像素坐标系转换到相机坐标系中, 得到边缘像素点的坐标; 接着, 将斜坡图像边缘像素点坐标由图像像素坐标系转换到相机坐标系中, 得到斜坡真实长度  $l$ ; 最后,

通过基于迁移学习的室外单帧图像深度估计网络对斜坡图像进行深度估计, 得到斜坡上下边缘像素点的深度值, 作差得到斜坡投影长度  $a$ , 再通过反余弦函数便可得到斜坡角度  $\theta = \arccos(a/l)$ 。其中, 如何根据单帧斜坡 RGB 图像估计出斜坡上下边缘像素点的深度值, 是获得精确的斜坡角度估计值的关键, 本文中该值可通过所设计的深度估计网络获得。下面详细介绍本文所提出的基于迁移学习的室外单帧图像深度估计算法基本原理。

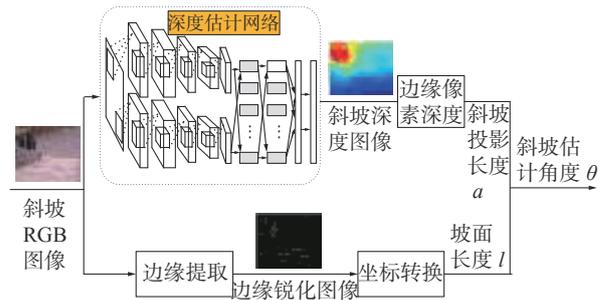


图 1 坡度检测算法

Fig. 1 General framework of slope detection algorithm

### 2.2 基于迁移学习的室外单帧图像深度估计算法基本原理

本文提出的基于迁移学习的室外单帧图像深度估计算法如图 2 所示, 主要由室内图像深度估计和参数迁移 2 部分组成。首先, 采用室内图像语义分割数据集 (NYU depth dataset V2, NYU-D V2) 对室内图像深度估计网络进行训练, 得到 DCNF-FCSP 室内图像深度估计网络模型; 然后, 将该网络模型中前 5 个图像特征提取层的网络参数迁移至室外图像深度估计网络中; 最后, 固定迁移至室外图像深度估计网络的图像特征提取网络参数, 利用 Make 3D 室外图像数据集对剩余的 2 个卷积层和 3 个全连接层的网络参数进行训练, 从而得到室外单帧图像深度估计网络。

下面详细介绍基于迁移学习的室外图像深度估计方法的具体实现原理, 包括室内图像深度估计和基于迁移学习的单帧图像深度估计。

#### 2.2.1 室内图像深度估计

本文借鉴文献 [12] 提出的室内图像深度估计方法对图像深度估计网络进行初始化。首先, 采用简单线性迭代聚类 (simple linear iterative clustering, SLIC) 超像素分割方法将室内 RGB 图像划分为多个超像素块; 然后, 通过全卷积网络 (full convolutional network, FCN) 对室内 RGB 图像进行特征提取, 在此基础上, 将图像像素点的深度估计问题转化为对图像超像素块的深度估计问题, 并采用超像素池化来获得图像超像素块的特征信

息;最后,采用条件随机场 (conditional random fields, CRF) 参数优化的形式构建损失函数对网络

进行优化,进而得到更加准确的室内图像深度估计结果。

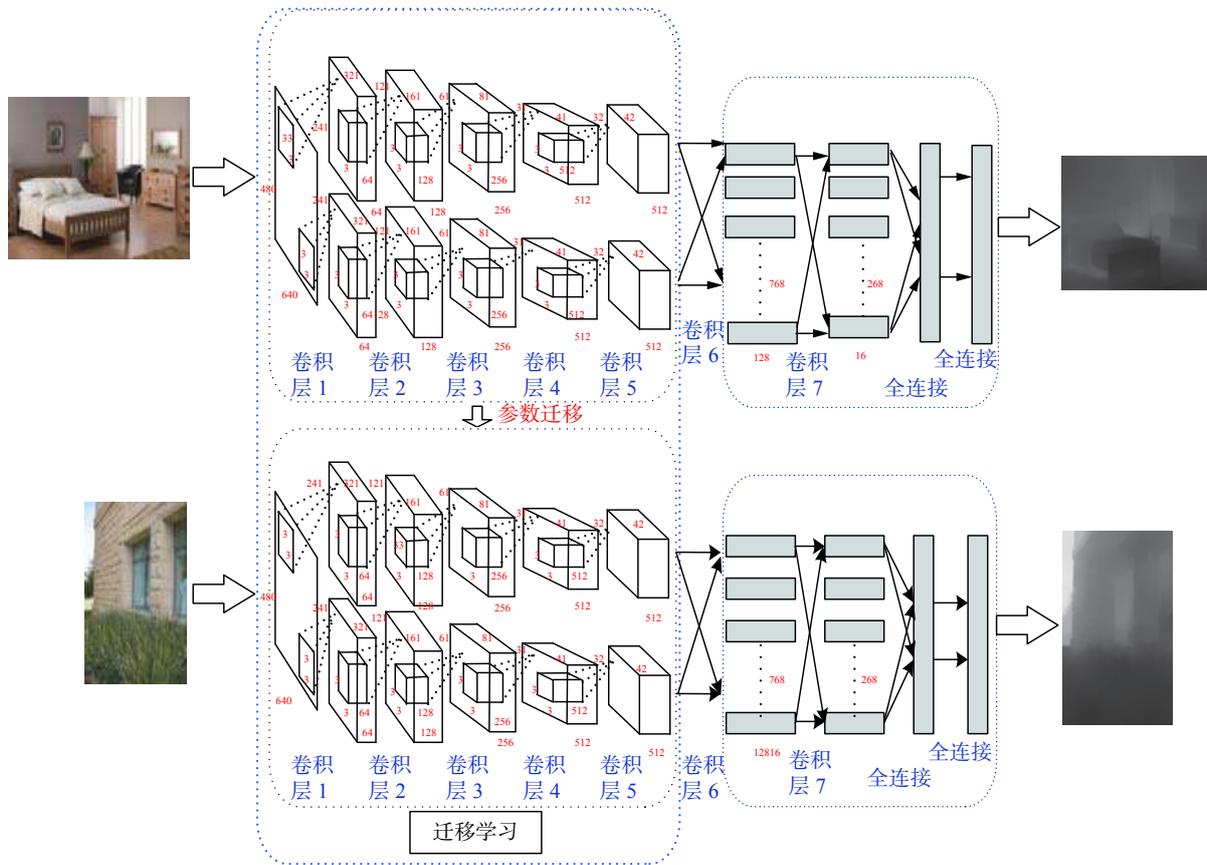


图 2 基于迁移学习的室外单帧图像深度估计算法

Fig. 2 General framework of outdoor single image depth estimation algorithm based on transfer learning

假设输入 RGB 图像为  $x = [x_1 \ x_2 \ \dots \ x_n]^T$ , 向量  $z = [z_1 \ z_2 \ \dots \ z_n]^T$  是  $n$  个超像素块的预测深度值, 其对应的真实深度值为  $y = [y_1 \ y_2 \ \dots \ y_n]^T$ 。

CRF 是在给定随机变量  $x$  的条件下, 随机变量  $y$  的马尔可夫随机场  $P_r(y|x)$ , 其表达式为

$$P_r(y|x) = \frac{1}{Z(x)} \exp(-E(y, x)) \quad (1)$$

式中:  $E(\cdot)$  为能量函数, 用于表示图像超像素块与超像素块之间的关系;  $Z(\cdot)$  为配分函数, 可将输入图像深度预测问题转化为求解后验概率最大的值<sup>[12]</sup>:

$$y^* = \arg \max_y P_r(y|x) \quad (2)$$

### 2.2.2 基于迁移学习的室外单帧图像深度估计

传统的机器学习通常需要大量的训练样本, 然而现实生活中, 因受数据获取的限制, 往往难以满足。因此, 本文采用迁移学习来解决训练室外单帧图像深度估计网络所需要的图像数据集有限的问题。由于 CNN 具有良好的层次结构, 而且对于 CNN 来说前面几层所学习到的特征都是通用特征, 随着网络层次的加深, 后面的网络层更加偏

重于学习特定的特征。所以, 本文选用迁移学习将室内图像深度估计网络模型中提取通用特征的网络参数进行迁移, 以减少训练室外图像深度估计网络所需的数据量, 同时节省了网络训练时间。

迁移学习主要针对源任务与目标任务之间的知识迁移, 通过将源任务上训练好的模型权重直接应用于目标任务中, 从而有效利用域间的知识迁移来解决目标任务标记数据不足的问题。通过迁移源任务中预先训练的模型权重, 来显著减少目标任务模型中需要学习的参数数量, 从而实现利用较小规模的样本就可以进行快速有效的模型训练<sup>[20]</sup>。针对不同的研究领域, CNN 迁移过程中所选取的迁移对象也不相同。本文选用归纳式迁移学习中的参数迁移方法将室内图像深度估计网络参数迁移至室外图像深度估计网络中, 以完成室外图像深度估计任务, 如图 3 所示。

假设源域与目标域共享模型参数, 网络由 7 个卷积层以及 3 个全连接层组成, 其中前 5 层是基于原始图像进行特征提取, 从第 6 个卷积层

开始引入超像素块的概念, 所以可以认为前 5 层卷积层所提取到的特征在源域与目标域中共用,

而网络后 2 个卷积层和 3 个全连接层主要是针对目标任务进行精细计算。

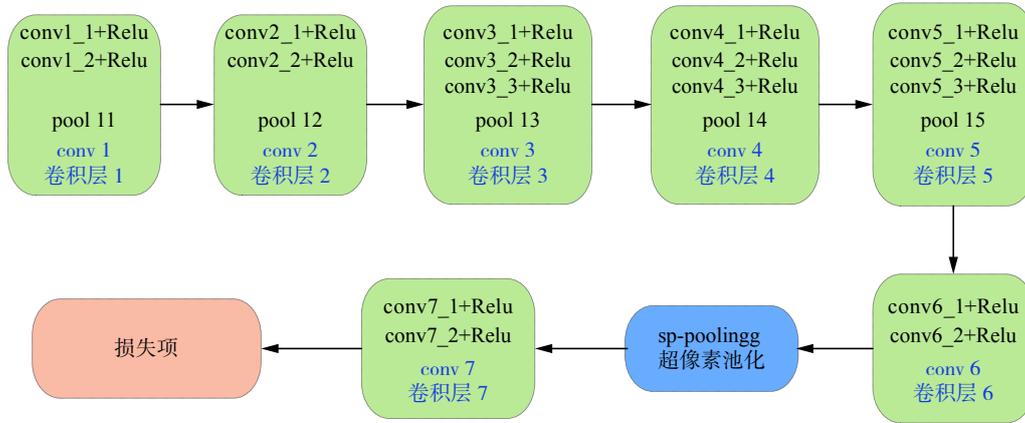


图 3 室外图像深度估计特征提取部分网络结构

Fig. 3 Network structure of feature extraction for outdoor image depth estimation

本文将 DCNF-FCSP 室内图像深度估计网络中前 5 层图像特征提取的网络参数迁移至室外图像深度估计网络中以完成室外图像的特征提取, 如图 4 所示。

室内图像深度估计数据集种类较多, 且数据量充足, 因而提取到的图像特征较为丰富, 可以弥补由于室外图像数据集不足而导致特征提取不完备的缺陷。将室外图像提取特征作为其后网络

层的输入, 结合室外图像深度估计标准数据集 Make 3D 对后面的网络层进行训练并更新参数, 来获取最终的图像深度的估计值。采用 CRF 参数优化的形式构建室外图像深度估计网络的损失函数<sup>[12, 21]</sup>, 能够在考虑预测深度值与实际深度值之间关系的基础上, 利用输入图像邻域超像素块之间的相似性信息来对图像深度估计值进行平滑处理, 进而得到输入图像更加准确的像素点深度估计值。

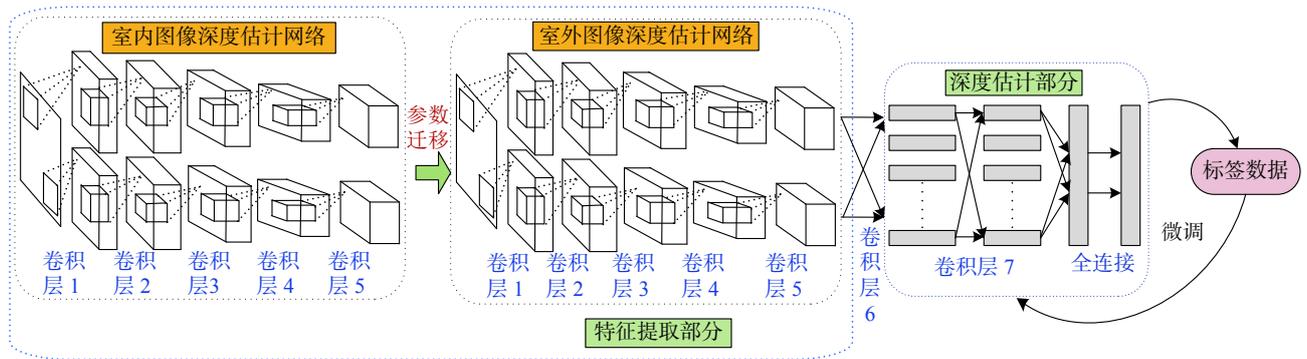


图 4 基于迁移学习的深度估计网络结构

Fig. 4 Network structure of depth estimation based on transfer learning

### 3 实验研究

本部分进行了 2 组实验分别用来验证本文所提出的室外单帧图像深度估计算法和移动机器人单帧图像坡度检测算法的有效性。

#### 3.1 基于迁移学习的室外单帧图像深度估计实验

本实验的目的是为了验证所提出的室外图像深度估计方法的有效性与优越性。采用室外图像深度估计标准数据集 Make 3D 对网络模型进行训练与测试, 并通过实际拍摄的室外图像对深度估

计网络进行测试。最后采用 3 种图像深度估计评估标准对所提出的室外图像深度估计方法的性能进行评估。

本文采用的 3 种图像深度估计评估标准定义为

1) 平均相对误差 (mean relative error, MRE), 主要反映估计值的可信程度, 计算公式为

$$MRE = \frac{1}{N} \sum_{i=1}^N \frac{|d_p^{gt} - d_p|}{d_p^{gt}} \quad (3)$$

2) 均方根误差 (root-mean-square error, RMSE)

是一种定量的衡量方式, 主要反映估计值与真实值之间的偏差, 计算公式为

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (d_p^{gt} - d_p)^2} \quad (4)$$

3) 平均对数误差 (lg), 主要是为了更便捷发现数据之间的关系, 所以对数据转换也是一种解决方法, 计算公式为

$$lg = \frac{1}{N} \sum_{i=1}^N |\lg d_p^{gt} - \lg d_p| \quad (5)$$

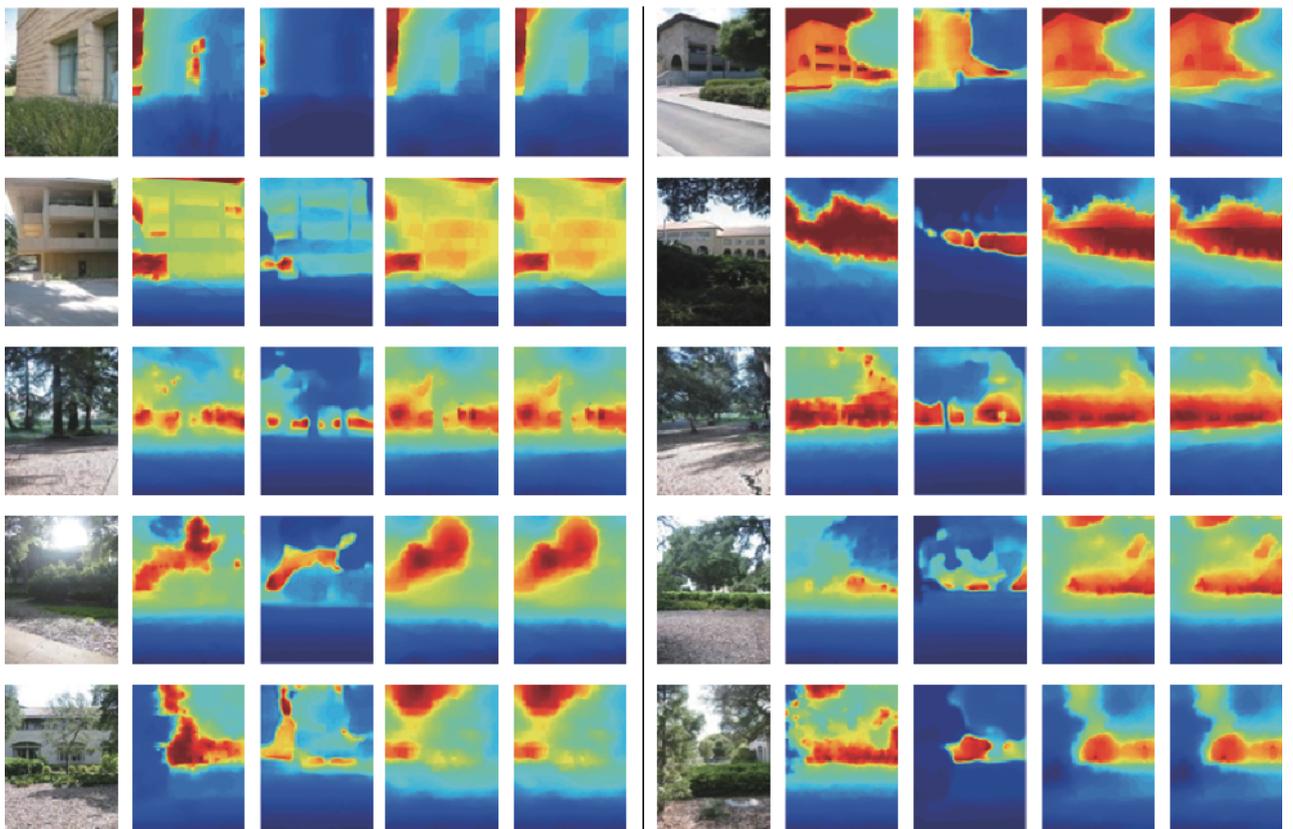
式中:  $N$  是输入图像的像素总数;  $d_p^{gt}$  是像素点的真实深度;  $d_p$  是像素点的预测深度。

实验 1 基于 Make 3D 标准数据集的深度估计

本实验采用 <http://make3d.cs.cornell.edu/data.html> 网站上提供的 Make 3D 标准数据集<sup>[22]</sup>进行了室外单帧图像深度估计实验。Make 3D 数据集是目前最主流的室外图像标准数据集, 包含 534 对 RGB-D 图像对 (包括 RGB 图像和其对应的深度

图像), 其中, 400 对图像对用于训练, 134 对图像用于测试。每帧图像的深度值均由激光雷达进行采集, 相较于 Kinect 相机采集的深度信息, 该测距仪可以得到室外图像更加精确的深度信息, 而且测距范围更大, 与普通的深度传感器相比还具有分辨率高、抗干扰能力强等特点。由于 Make 3D 图像中深度层次区分不明显, 所以采用颜色热图表示其深度估计结果。本文从该数据集包含的 134 个测试样本图像中根据光照以及环境复杂度的不同选取了 10 帧图像进行深度估计算法验证, 深度估计结果如图 5 所示。

由图 5 可以看出, 本文所提出的室外图像深度估计方法可以准确地估计出室外单帧图像的深度值。为了进一步验证本文所提出方法的有效性, 结合图像深度估计性能评价标准对估计结果进行量化分析, 并与 FCRN (fully convolutional residual networks, FCRN)<sup>[19]</sup> 和 DCNF-FCSP<sup>[12]</sup> 深度估计方法进行比较, 结果如表 1 所示。



(a) RGB 图像 1 (b) 标准的深度热图 (c) FCRN 估计结果<sup>[19]</sup> (d) DCNF-FCSP 估计结果<sup>[12]</sup> (e) 本文算法估计结果 (f) RGB 图像 2 (g) 标准的深度热图 (h) FCRN 估计结果<sup>[19]</sup> (i) DCNF-FCSP 估计结果<sup>[12]</sup> (j) 本文算法估计结果

图 5 Make 3D 数据集的深度估计结果 (蓝色: 近; 红色: 远)

Fig. 5 Results of depth estimation on Make 3D dataset (blue: close; red: far)

从表 1 可以看出本文所提出的基于迁移学习的室外单帧图像深度估计方法在 3 个性能评价指标上均优于 FCRN 深度估计方法, 在平均相对误

差和均方根误差上均优于 DCNF-FCSP 深度估计方法, 而且在平均相对误差这一评价指标上具有较为明显的优势。

表 1 Make 3D 数据集深度估计结果对比  
Table 1 Comparison of depth estimation results for Make 3D dataset

| 场景 | FCRN <sup>[19]</sup> |        |         | DCNF-FCSP <sup>[12]</sup> |               |                | 本文方法(迭代1 5000次) |               |                |
|----|----------------------|--------|---------|---------------------------|---------------|----------------|-----------------|---------------|----------------|
|    | MRE                  | lg     | RMSE    | MRE                       | lg            | RMSE           | MRE             | lg            | RMSE           |
| 1  | 0.1024               | 0.0631 | 14.2517 | <b>0.1530</b>             | 0.0734        | <b>7.3304</b>  | 0.1532          | 0.0754        | 7.6390         |
| 2  | 0.5710               | 0.4357 | 28.8680 | <b>0.1449</b>             | <b>0.0695</b> | <b>10.0483</b> | 0.1451          | 0.0716        | 10.3562        |
| 3  | 0.6145               | 0.4645 | 21.0095 | 0.1836                    | 0.0769        | <b>7.9044</b>  | <b>0.1684</b>   | <b>0.0725</b> | 8.0804         |
| 4  | 0.6843               | 0.4123 | 37.3963 | 1.1952                    | 0.3877        | <b>29.3387</b> | <b>1.1560</b>   | <b>0.3845</b> | 29.6222        |
| 5  | 0.4973               | 0.3352 | 15.3152 | 0.3625                    | <b>0.1465</b> | 12.3683        | <b>0.3562</b>   | 0.1496        | <b>12.2847</b> |
| 6  | 0.5127               | 0.3897 | 26.0401 | 0.6315                    | <b>0.2087</b> | 21.6887        | <b>0.6128</b>   | 0.2101        | <b>21.1751</b> |
| 7  | 0.5259               | 0.3736 | 20.2887 | 0.5640                    | <b>0.2790</b> | 21.0787        | <b>0.5559</b>   | 0.2821        | <b>20.6893</b> |
| 8  | 0.7211               | 0.4315 | 23.5747 | <b>0.3659</b>             | <b>0.2379</b> | <b>16.1994</b> | 0.3705          | 0.2445        | 16.3725        |
| 9  | 0.5435               | 0.3821 | 23.0167 | 0.6329                    | 0.1674        | 9.2323         | <b>0.5901</b>   | <b>0.1597</b> | <b>8.5783</b>  |
| 10 | 0.6701               | 0.3653 | 23.7896 | 1.1973                    | 0.2840        | 21.6687        | <b>1.1469</b>   | <b>0.2818</b> | <b>21.2793</b> |
| 均值 | 0.5442               | 0.3653 | 23.3550 | 0.5431                    | <b>0.1931</b> | 15.6858        | <b>0.5255</b>   | 0.1932        | <b>15.6077</b> |

实验 2 实际拍摄的室外图像深度估计

本实验的目的是通过对实际拍摄的室外图像深度估计, 来进一步验证所提方法在真实场景中的有效性和优越性。实验中使用博世 (BOSCH) DEL40 手持激光测距仪进行实际深度值测量, 随

机选取图像中一个矩形区域的 4 个顶点对图像深度估计质量进行评估, 深度估计结果如图 6 所示。由图 6 可以看出, 本文所提出的深度估计方法对于实际拍摄的室外单帧图像在光照均匀且场景简单的环境下可以实现较好的深度估计。

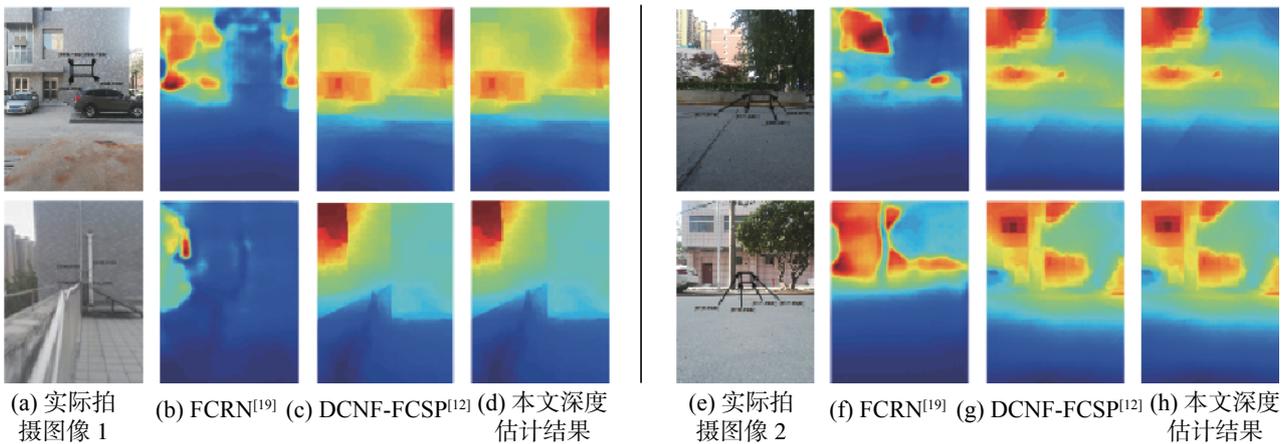


图 6 实际拍摄室外图像深度估计结果 (蓝色: 近; 红色: 远)

Fig. 6 The actual outdoor image depth estimation result (blue: close; red: far)

表 2 进一步给出了本文所提出的室外单帧图像深度估计方法与主流的 FCRN 深度估计方法<sup>[19]</sup>、DCNF-FCSP 深度估计方法<sup>[12]</sup>的量化估计结果, 其中深度的真实值由激光测距仪获得。由表 2 可以看出, 本文所提出的深度估计方法对于光照均

匀且场景简单的室外单帧图像能够进行准确的深度估计。通过 3 种方法在室外图像中任意选取的矩形区域顶点的深度估计结果与真实距离的对比可以看出, 本文所提出的深度估计方法对于实际拍摄的室外单帧图像深度估计的精度更高。

表 2 实际拍摄室外图像中所选矩形区域的深度估计结果比较

Table 2 Comparison of depth estimation results of selected regions in actual outdoor images

| 矩形1                         | (185, 220) | (185, 286) | (220, 220) | (220, 286) | 平均误差          |
|-----------------------------|------------|------------|------------|------------|---------------|
| 真实值/m                       | 16.7       | 16.7       | 16.7       | 16.7       | —             |
| FCRN/m <sup>[19]</sup>      | 13.2373    | 9.9595     | 14.0511    | 10.9642    | 4.6470        |
| DCNF-FCSP/m <sup>[12]</sup> | 16.8084    | 18.3771    | 19.4635    | 20.6221    | 2.1178        |
| 本文结果/m                      | 16.5392    | 18.0835    | 19.1507    | 20.2916    | <b>1.8967</b> |
| 矩形2                         | (279, 241) | (279, 308) | (304, 241) | (304, 308) | 平均误差          |
| 真实值/m                       | 12.2       | 12.2       | 12.2       | 12.2       | —             |
| FCRN/m <sup>[19]</sup>      | 16.6765    | 15.9572    | 13.9033    | 13.6180    | 2.8387        |
| DCNF-FCSP/m <sup>[12]</sup> | 12.7644    | 12.4089    | 12.4916    | 12.1964    | 0.2671        |
| 本文结果/m                      | 12.5389    | 12.1918    | 12.2688    | 11.9806    | <b>0.1588</b> |
| 矩形3                         | (310, 250) | (310, 270) | (330, 250) | (330, 270) | 平均误差          |
| 真实值/m                       | 7.5        | 7.5        | 7.5        | 7.5        | —             |
| FCRN/m <sup>[19]</sup>      | 4.6831     | 4.1835     | 4.4692     | 4.0478     | 3.1541        |
| DCNF-FCSP/m <sup>[12]</sup> | 8.9247     | 8.9263     | 8.4528     | 8.5408     | 1.2111        |
| 本文结果/m                      | 8.7965     | 8.7980     | 8.3372     | 8.4227     | <b>1.0886</b> |
| 矩形4                         | (236, 212) | (236, 247) | (266, 212) | (266, 247) | 平均误差          |
| 真实值/m                       | 13.5       | 13.5       | 13.5       | 13.5       | —             |
| FCRN/m <sup>[19]</sup>      | 22.1212    | 21.2585    | 19.5015    | 18.4432    | 6.8311        |
| DCNF-FCSP/m <sup>[12]</sup> | 13.6837    | 13.3962    | 13.9285    | 12.0558    | 0.5400        |
| 本文结果/m                      | 13.4732    | 13.1908    | 13.7447    | 13.8707    | <b>0.2378</b> |

表 3 给出了在式 (3)~(5) 所定义的 3 种深度估计性能评价标准下上述 3 种深度估计方法的性能比较结果。从表 3 中的对比结果可以看出, 与 FCRN 和 DCNF-FCSP 方法相比, 本文所提出的室外单帧图像深度估计方法性能更优。

表 3 实际拍摄室外图像深度估计结果对比

Table 3 Comparison of depth estimation results of actual outdoor images.

| 算法                        | MSE           | lg<br>(越小越好)  | RMSE          |
|---------------------------|---------------|---------------|---------------|
| FCRN <sup>[19]</sup>      | 0.3301        | 0.1624        | 2.3220        |
| DCNF-FCSP <sup>[12]</sup> | 0.0763        | 0.0357        | 0.6282        |
| 本文方法                      | <b>0.0622</b> | <b>0.0294</b> | <b>0.5514</b> |

综上所述, 本文所提出的基于迁移学习的室外单帧图像深度估计方法在实际拍摄的室外单帧图像上也取得了较高的深度估计精度, 是一种有效的室外单帧图像深度估计方法。

### 3.2 移动机器人坡度检测实验

本实验的目的是通过 2 组具有不同倾斜角的

斜坡在不同视角下的坡度检测实验, 验证所提出的改进的移动机器人单帧图像坡度检测算法的有效性和可行性。实验中的斜坡图像采集装置主要包括 TurtleBot 移动机器人和 Kinect 深度相机, 如图 7 所示, 安装在移动机器人上的 Kinect 相机可以从不同的视角获取各种角度的斜坡图像。



图 7 斜坡图像采集装置  
Fig. 7 Slope image acquisition device

实验场地如图 8(a)、(e) 所示。对 2 个视野开阔, 便于观察且倾角不同的斜坡分别从 5 个不同视角进行了坡度检测实验。分别采用绝对误差、相对误差和误差均值 3 个指标对所提出的移动机器人单帧图像坡度检测算法估计出的斜坡角度的精确度进行评估。室外不同视角下的斜坡

边缘提取和深度估计的实验结果分别如图 8 (b)~(d)、(f)~(h) 所示, 其量化估计结果如表 4 所示。由图 8 和表 4 结果所示, 将本文所提出的基于迁移学习的室外单帧图像深度估计方法应用于

移动机器人坡度检测中, 对于不同倾角和不同视角下的斜坡均能得到更加准确的坡度检测结果, 大大提高了移动机器人在未知环境中感知坡度的精确度。

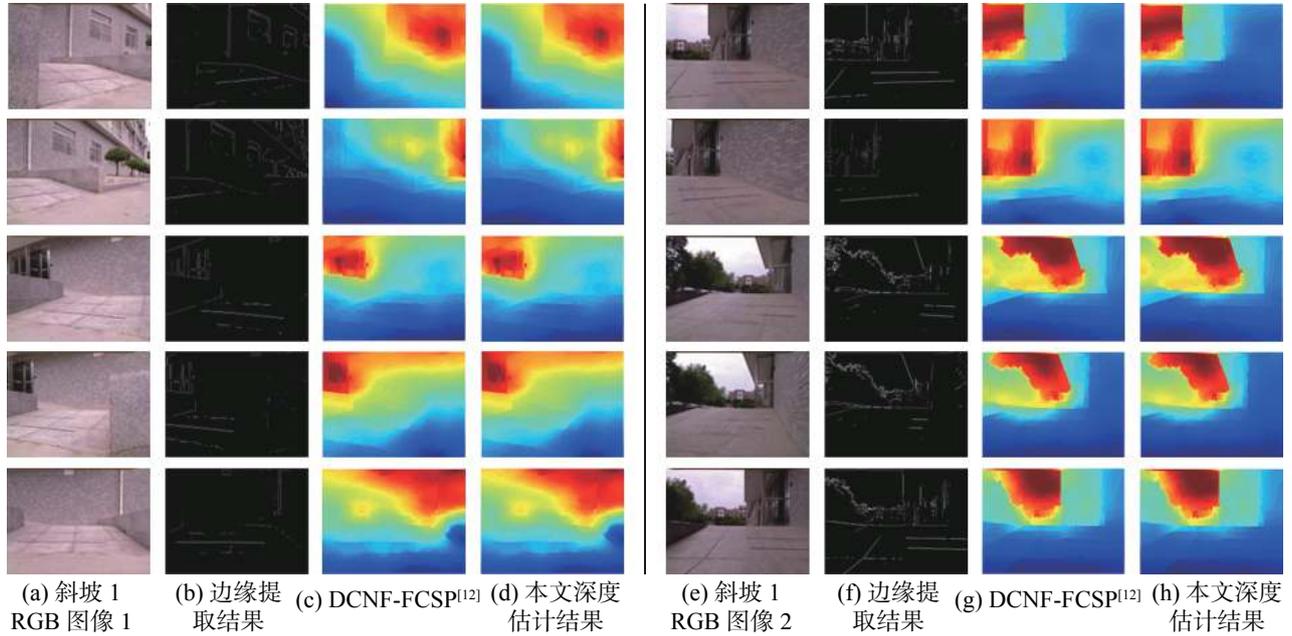


图 8 室外斜坡边缘提取结果与深度估计结果 (蓝色: 近; 红色: 远)

Fig. 8 Outdoor slope edge extraction results and depth estimation results (blue: close; red: far)

表 4 室外斜坡角度估计结果对比

Table 4 Comparison of outdoor slope angle estimation results

| 场景   | 真实角度/(°) | 基于DCNF-FCSP <sup>[12]</sup> |          |        | 基于本文方法   |               |              |
|------|----------|-----------------------------|----------|--------|----------|---------------|--------------|
|      |          | 预测角度/(°)                    | 绝对误差/(°) | 相对误差/% | 预测角度/(°) | 绝对误差/(°)      | 相对误差/%       |
| 1    | —        | 10.5920                     | 2.6695   | 20.13  | 10.9247  | 2.3368        | 17.62        |
| 2    | —        | 18.1446                     | 4.8831   | 36.82  | 15.2941  | 2.0326        | 15.33        |
| 3    | 13.2615  | 10.0462                     | 3.2153   | 24.25  | 10.8929  | 2.3686        | 17.86        |
| 4    | —        | 15.1603                     | 1.8988   | 14.32  | 11.4789  | 1.7826        | 13.44        |
| 5    | —        | 16.4518                     | 3.1903   | 24.06  | 16.0326  | 2.7711        | 20.90        |
| 6    | —        | 13.5225                     | 3.6843   | 37.45  | 10.9825  | 1.1443        | 14.67        |
| 7    | —        | 12.1817                     | 2.3435   | 23.82  | 11.6227  | 1.7845        | 18.14        |
| 8    | 9.8382   | 12.2018                     | 2.3636   | 24.02  | 11.8937  | 2.0555        | 20.89        |
| 9    | —        | 12.3618                     | 2.5236   | 25.65  | 11.8244  | 1.9862        | 20.19        |
| 10   | —        | 12.8103                     | 2.9721   | 30.21  | 11.9170  | 2.0788        | 21.12        |
| 误差均值 | —        | —                           | 2.9744   | 26.07  | —        | <b>2.0341</b> | <b>18.02</b> |

#### 4 结束语

本文提出了一种基于迁移学习的室外单帧图

像深度估计方法, 通过将室内图像深度估计网络参数迁移至室外图像深度估计网络中, 对室外单帧图像深度进行估计, 解决了室外图像数据集有

限的问题。实验结果表明,本文所提出的室外图像深度估计方法在 Make 3D 标准数据集和实际拍摄的室外单帧图像上都取得了较高的深度估计精度。在此基础上将其与基于视觉的移动机器人坡度检测相结合进一步提出了一种改进的移动机器人单帧图像坡度检测算法,实验结果表明移动机器人坡度检测精度得到了显著提高,检测精度满足移动机器人在未知环境中对坡度感知精度的要求。在今后的工作中,我们将进一步将本文所提出的单帧图像深度估计结果应用到图像语义分割中以提高图像语义分割网络的性能,最终完成未知环境下的高精度 3D 语义地图构建。

## 参考文献:

- [1] 朱江,王耀南,余洪山,等.未知环境下移动机器人自主感知斜坡地形方法[J].仪器仪表学报,2010,31(8):1916-1920.  
ZHU Jiang, WANG Yaonan, YU Hongshan, et al. Mobile robot autonomous perceiving slope terrain under unknown environment[J]. Chinese journal of scientific instrument, 2010, 31(8): 1916-1920.
- [2] YU Jinxia, CAI Zixing, DUAN Zhuohua. Dead reckoning of mobile robot in complex terrain based on proprioceptive sensors[C]//Proceedings of 2008 International Conference on Machine Learning and Cybernetics. Kunming, China, 2008: 1930-1935.
- [3] LI Zhibin, TSAGARAKIS N G, CALDWELL D G. Stabilizing humanoids on slopes using terrain inclination estimation[C]//Proceedings of 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. Tokyo, Japan, 2013: 4124-4129.
- [4] LU Jixin, KOBAYASHI Y, EMARU T, et al. Indoor slope and edge detection by using two-dimensional EKF-SLAM with orthogonal assumption[J]. International journal of advanced robotic systems, 2015, 12(4): 44.
- [5] HARA S, SHIMIZU T, KONISHI M, et al. Autonomous mobile robot for outdoor slope using 2D LiDAR with uniaxial gimbal mechanism[J]. *Journal of robotics and mechatronics*, 2020, 32(6): 1173-1182.
- [6] TAREEN S A K, KHAN H M. Novel slope detection and calculation techniques for mobile robots[C]//Proceedings of the 2nd IEEE International Conference on Robotics and Artificial Intelligence. Rawalpindi, Pakistan, 2016: 158-163.
- [7] 王泉德,张松涛.基于多尺度特征融合的单目图像深度估计[J].华中科技大学学报(自然科学版),2020,48(5):7-12.  
WANG Quande, ZHANG Songtao. Monocular depth estimation with multi-scale feature fusion[J]. Journal of Huazhong University of Science and Technology (natural science edition), 2020, 48(5): 7-12.
- [8] 毕天腾,刘越,翁冬冬,等.基于监督学习的单幅图像深度估计综述[J].计算机辅助设计与图形学学报,2018,30(8):1383-1393.  
BI Tianteng, LIU Yue, WENG Dongdong, et al. Survey on supervised learning based depth estimation from a single image[J]. Journal of computer-aided design and computer graphics, 2018, 30(8): 1383-1393.
- [9] 何磊,苏松志,李绍滋.单摄像头下基于样本学习的人体深度估计[J].智能系统学报,2014,9(2):161-167.  
HE Lei, SU Songzhi, LI Shaozi. Human depth estimation on the basis of the sample learning method under a single camera[J]. CAAI transactions on intelligent systems, 2014, 9(2): 161-167.
- [10] 孙蕴瀚,史金龙,孙正兴.利用自监督卷积网络估计单图像深度信息[J].计算机辅助设计与图形学学报,2020,32(4):643-651.  
SUN Yunhan, SHI Jinlong, SUN Zhengxing. Estimating depth from single image using unsupervised convolutional network[J]. Journal of computer-aided design and computer graphics, 2020, 32(4): 643-651.
- [11] 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, Quebec, Canada, 2014: 2366-2374.
- [12] LIU Fayao, SHEN Chunhua, LIN Guosheng, et al. Learning depth from single monocular images using deep convolutional neural fields[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 38(10): 2024-2039.
- [13] RANFTL R, LASINGER K, HAFNER D, et al. Towards robust monocular depth estimation: mixing datasets for zero-shot cross-dataset transfer[J]. IEEE transactions on pattern analysis and machine intelligence, 2020, 42(8): 1939-3539.
- [14] HU Junjie, OZAY M, ZHANG Yan, et al. Revisiting single image depth estimation: toward higher resolution maps with accurate object boundaries[J]. arXiv: 1803.0867, 2018.
- [15] KUZNIETSOV Y, STÜCKLER J, LEIBE B. Semi-super-

- vised deep learning for monocular depth map prediction[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: CVPR, 2017: 2215–2223.
- [16] Fu H, Gong M, Wang C, et al. A Compromise Principle in Deep Monocular Depth Estimation[J]. arXiv preprint arXiv: 1708.0826, 2018.
- [17] 梁煜, 张金铭, 张为. 一种改进的卷积神经网络的室内深度估计方法 [J]. 天津大学学报(自然科学与工程技术版), 2020, 53(8): 840–846.  
LIANG Yu, ZHANG Jinming, ZHANG Wei. An improved indoor depth estimation method using convolutional neural networks[J]. Journal of Tianjin University(science and technology), 2020, 53(8): 840–846.
- [18] GODARD C, MAC AODHA O, BROSTOW G J. Unsupervised monocular depth estimation with left-right consistency[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 6602–6611.
- [19] LAINA I, RUPPRECHT C, BELAGIANNIS V, et al. Deeper depth prediction with fully convolutional residual networks[C]//Proceedings of the 4th International Conference on 3D Vision. Stanford, CA, USA, 2016: 239–248.
- [20] 许敏, 史茨中, 葛洪伟, 等. 一种具有迁移学习能力的 RBF-NN 算法及其应用 [J]. 智能系统学报, 2018, 13(6): 959–966.  
XU Min, SHI Yingzhong, GE Hongwei, et al. A RBF-NN algorithm with transfer learning ability and its application[J]. CAAI transactions on intelligent systems, 2018, 13(6): 959–966.
- [21] 李耀宇, 王宏民, 张一帆, 等. 基于结构化深度学习的单目图像深度估计 [J]. 机器人, 2017, 39(6): 812–819.  
LI Yaoyu, WANG Hongmin, ZHANG Yifan, et al. Structured deep learning based depth estimation from a monocular image[J]. Robot, 2017, 39(6): 812–819.
- [22] SAXENA A, SUN Min, NG A Y. Make3D: learning 3D scene structure from a single still image[J]. IEEE transactions on pattern analysis and machine intelligence, 2009, 31(5): 824–840.

### 作者简介:



辛菁, 教授, 主要研究方向为机器人智能感知与自主运动, 主持国家自然科学基金项目 2 项、省部级基金项目多项。发表学术论文 40 余篇。



杜柯楠, 硕士研究生, 主要研究方向为语义地图构建。



刘丁, 教授, 主要研究方向为复杂系统建模与优化控制, 主持国家自然科学基金项目 3 项、国家科技重大专项及 863 项目多项, 获国家技术发明奖二等奖 1 项、省科技进步一等奖 3 项。发表学术论文 200 余篇, 出版专著 3 部。