

文章编号: 2095-2163(2021)04-0104-06

中图分类号: TP183

文献标志码: A

基于像素偏转模型和机器学习的室外图像天空像素检测

孟祥环, 罗素云

(上海工程技术大学 机械与汽车工程学院, 上海 201620)

摘 要: 图像中的天空区域对于基于视觉的地面机器人导航具有重要意义, 为了识别图像中的天空部分, 本文提出了一种基于像素偏转模型的 BP 神经网络天空识别方法。首先, 制作天空图像集和非天空图像集, 天空图像集由各种天气情况下的天空提取而成, 非天空图像集由非天空的景物构成, 主要包括建筑、汽车、树木、植物等; 其次, 使用提出的像素偏转模型提取天空图像集和非天空图像集的像素特征并进行处理, 对天空像素点和非天空像素点进行标注, 利用 BP 神经网络对像素特征进行训练, 得到权重文件; 最后, 使用得到的权重文件进行天空的识别。为了更好的说明本文算法和模型的优越性, 使用本文算法与 Otsu 算法、YeHu 算法、Graph-cut 算法和 Mask-Rcnn 算法模型进行了比较, 并设计了两组组对比实验, 第一组实验进行识别效果的主观评价, 第二组实验利用 CamVid 数据集的天空类进行算法精度的定量分析。

关键词: 天空识别; 像素偏转模型; BP 神经网络; Mask-Rcnn

Pixel detection in outdoor image sky based on pixel deflection model and machine learning

MENG Xianghuan, LUO Suyun

(College of Mechanical and Automotive Engineering, Shanghai University of Engineering Science School, Shanghai 201620, China)

[Abstract] The sky region in the image is of great significance for vision-based ground robot navigation. In order to identify the sky part in the image, this paper proposes a BP neural network sky recognition method based on pixel deflection model. Firstly, sky image sets and non-sky image sets were made. Sky image sets were extracted from the sky under various weather conditions. Non-sky image sets were composed of non-sky objects, including buildings, cars, trees, plants, etc. Then, the proposed pixel deflection model is used to extract and process the pixel features of sky image set and non-sky image set, and the sky pixel points and non-sky pixel points are labeled. Then, the BP neural network is used to train the pixel features and obtain the weight file. Finally, the weight file is used to identify the sky. In order to better illustrate the advantages of the algorithm and model in this paper, the algorithm in this paper was compared with Otsu algorithm, YeHu algorithm, Graph-cut algorithm and Mask-RCNN algorithm. In addition, two groups of comparative experiments were designed. The first group of experiments was used for subjective evaluation of recognition effect, and the second group of experiments was used for quantitative analysis of algorithm accuracy using sky class of CamVid data set.

[Key words] Sky recognition; Pixel deflection model; BP neural network; Mask-RCNN

0 引 言

图像中的天空像素检测一直以来都是计算机视觉技术的挑战, 并且具有广泛的应用, 例如: 自主车辆或无人机导航、实时天气分类、图像编辑、天空替换和场景解析。在城市气候研究方面也发挥着重要作用, 如: 使用鱼眼摄影计算天空视角因子(SVF), 用于城市的气象监测^[1]。

在计算机视觉的相关研究中, 天空检测技术大体上可以分为以下几种, 一种是遍历图片中的所有像素点, 逐像素地寻找与天空相关的像素点。这种

方法侧重于寻找与天空相关联的单个像素, 如采用一种基于物理的方法, 利用图片中的颜色变化来识别天空^[2]; 基于颜色值和每个颜色通道的二维多项式生成天空像素概率图, 进而进行天空的识别^[3]; 另一种方法则侧重于寻找天空和地面边界, 利用改进的能量函数优化和图像的梯度信息来优化提取天空边界线^[4], 然而实际的城市场景中, 天空部分容易被建筑物、旗帜或其他障碍物隔开, 造成天空区域不连续, 使天空区域的识别变得十分困难^[5]。随着深度学习的快速发展, 多层卷积网络被广泛的应用于各种场景之中, 它通过使用训练好的语义分割卷积神经网络(CNN)模型, 对室外图像中的不同类别

作者简介: 孟祥环(1994-), 男, 硕士研究生, 主要研究方向: 机器视觉和图像处理; 罗素云(1975-), 女, 博士, 副教授, 主要研究方向: 智能车辆环境感知及控制。

通讯作者: 孟祥环 Email: 374019969@qq.com

收稿日期: 2020-11-23

(如天空、建筑物、树木和汽车)进行分类,例如 SegNet 和模型框架,但是它并不是专门为天空像素识别而设计的,这种方法需要大量的数据集和高性能的计算机做支撑。

不同的天空识别方法间进行横向对比是很困难的。首先,在大多数的研究报告中并没有准确度指标;其次,很多论文使用不同的指标和基准数据集进行报告。如:有报告得出 90.4% 的蓝天像素正确检测和 13% 的错误检测的结论^[2];有论文得到了鱼眼图像中标记的 SVF 均方根误差 (RMSE) 为 0.06^[6];有论文得到 80% 的验证图像的准确度为 0.90, 75% 的图像的准确度为 0.95^[7];有论文使用 SegNet, 得到了天空像素识别的准确度为 0.96^[8]。

考虑到天空检测的复杂性,本文提出了像素偏转模型,利用该模型可以在广泛的照明和天气条件下对图片中的天空部分进行识别,相比以往算法,具有自适应性。本文在 Camvid 语义分割数据上测试一些现有的天空像素识别方法,并证明了这种新的自适应算法在识别天空时具有更高的精确度。

1 像素偏转模型

计算机储存图像时是以像素为基本单位的。为了更好的描述 BRG 三通道图像像素灰度值的特点,本文提出了像素偏转模型如图 1 所示,通过像素偏转角可以简单快速的找到亮度值差异比较大的像素点,这些像素点多代表天空、雪、河流等白色场景物体。

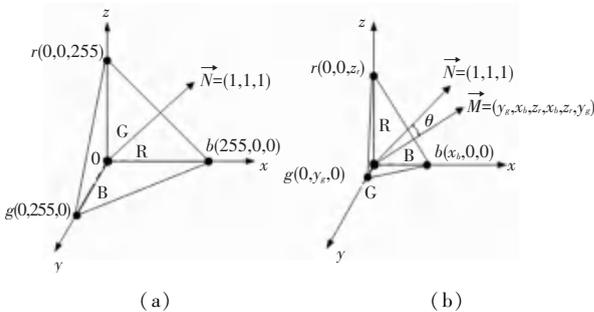


图 1 像素偏转模型

Fig. 1 Pixel deflection model

BRG 三通道图像单个通道的最大灰度值为 255,图 1 将 3 个通道的灰度值放在三维坐标系中表示,即 X 轴正方向代表 B 通道的像素值;Y 轴正方向代表 G 通道的像素值;Z 轴正方向代表 R 通道的像素值。则像素点的三维坐标可以表示为 $b(x_b, 0, 0)$, $g(0, y_g, 0)$, $r(0, 0, z_r)$ 。

3 个通道的最大灰度值点在三维空间构成了一

个三角形,如图 1(a)。这个三角形所在平面的法向量为 $\vec{N} = (1, 1, 1)$, 定义此向量为三通道像素平面标准法向量,所有其它像素点构成的三角形都应该在空间体 $o-bgr$ 内部,其构成的平面的法向量为 $\vec{M} = (z_r y_g, z_r x_b, y_g x_b)$, 如图 1(b) 所示。具体的公式推导过程如下:

这个三角形所在平面的法向量为 \vec{N} , 此向量即为三通道像素平面标准法向量,所有其他像素点构成的三角形都应该在空间体 o 内部,其构成的平面的法向量为 \vec{M} 。

设空间中有点的坐标为: $p_1 = (x_1, y_1, z_1)$, $p_2 = (x_2, y_2, z_2)$, $p_3 = (x_3, y_3, z_3)$ 。则该平面内有向量 $\vec{p}_{12} = (x_2 - x_1, y_2 - y_1, z_2 - z_1)$, $\vec{p}_{13} = (x_3 - x_1, y_3 - y_1, z_3 - z_1)$ 。平面的法向量为式(1):

$$\vec{N} = \vec{p}_{12} \times \vec{p}_{13} = \begin{vmatrix} \vec{i}_{(b)} & \vec{j}_{(g)} & \vec{k}_{(r)} \\ x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ x_3 - x_1 & y_3 - y_1 & z_3 - z_1 \end{vmatrix} = a\vec{i} + b\vec{j} + c\vec{k}. \quad (1)$$

其中: a, b, c 为式(2):

$$\begin{aligned} a &= (y_2 - y_1)(z_3 - z_1) - (y_3 - y_1)(z_2 - z_1), \\ b &= (z_2 - z_1)(x_3 - x_1) - (z_3 - z_1)(x_2 - x_1), \\ c &= (x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1). \end{aligned} \quad (2)$$

将点 b, g, r 代入平面的法向量方程得出像素平面的法向量,式(3):

$$\vec{M} = (z_r y_g, z_r x_b, y_g x_b). \quad (3)$$

像素平面法向量和标准像素法向量的夹角为式(4):

$$\theta = \arccos \left(\frac{\vec{M} \cdot \vec{N}}{|\vec{M}| \times |\vec{N}|} \right). \quad (4)$$

θ 即为像素偏转角, \arccos 为反余弦函数。像素偏转角可以反应某一点像素在 3 个通道的分布规律。当像素偏转角比较大时,3 个通道的像素值偏差较大。当像素偏转角较小,接近于 0 时,说明此像素点 3 个通道的灰度值十分接近。

2 图片中天空部分的识别

2.1 数据集的制作及特征提取

神经网络的训练需要大量的数据作为支撑,为了使本文的天空识别算法得到良好的识别效果,制

作了一个包含 3 231 张照片的数据集,如图 2 所示。这个数据集里的图片来源于 Corel1k 数据集,SVT 谷歌数据集, CityStreet 城市街景数据集和实景拍摄。数据集中包含了丰富的场景,如:城市街道场景、高速行驶场景、空旷自然风光、雨天阴天特殊场景和郊区林荫道等。将其天空部分和非天空部分进行提取,组合成天空和非天空提取图片(分辨率为 $3\ 096 \times 5\ 504$),用于天空特征的提取和训练。

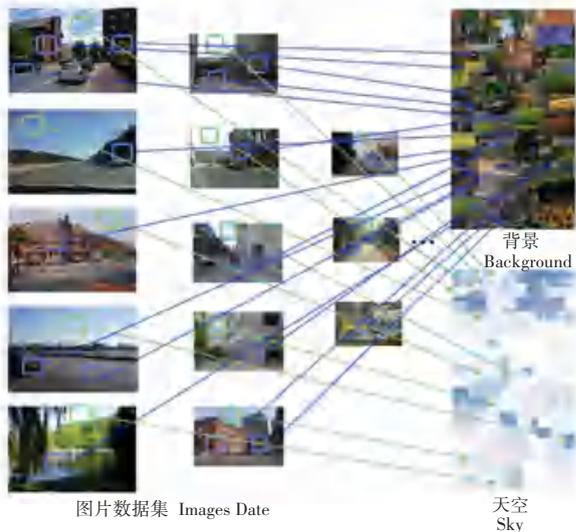


图 2 特征训练提取数据照片

Fig. 2 Image of feature training data extraction

待训练的特征经过提取,特征包括三通道的像素值,像素偏转角和像素点灰度最小值,经过筛选得到 319 552 条数据,部分数据见表 1,并将天空部分像素标签设置为 10,非天空部分标签设置为 0。

表 1 训练过程中所用到的特征

Tab. 1 Features used in the training process

像素点	B	G	R	$V(x)$	θ	训练目标值
1	82	32	7	7	43	0
2	111	38	19	19	31	0
3	93	1	4	1	45	0
4	132	36	38	36	24	0
5	170	179	188	170	2	10
6	154	181	226	154	9	10
7	153	190	232	153	10	10
8	167	189	238	167	8	10

2.2 BP 神经网络模型的建立

BP 神经网络模型是一种基于误差反向传播算法的多层前馈网络,是目前建立变量与因变量之间关系最广泛的神经网络模型之一,适用于本文通过特征参数预测能见度,其基本结构由输入层、隐藏

层、输出层组成。

其主要的公式如式(5)~式(10):

$$Y_j = f\left(\sum_{i=1}^m w_{ij}^1 X_i + q^1\right), \quad (5)$$

$$Z_k = f\left(\sum_{j=1}^n w_{jk}^2 Y_j + q^2\right), \quad (6)$$

$$f = \text{Sigmoid}(x) = \frac{1}{1 + e^{-x}}, \quad (7)$$

$$L = \frac{1}{k} \sum_{i=1}^k (\bar{Z}_k - Z_k)^2, \quad (8)$$

$$w_{ij}^1 = w_{ij} + \Delta w_{ij} = w_{ij} - \eta \frac{\partial L}{\partial w_{ij}}, \quad (9)$$

$$w_{jk}^2 = w_{jk} + \Delta w_{jk} = w_{jk} - \eta \frac{\partial L}{\partial w_{jk}}. \quad (10)$$

其中:输入层用字母 X 定义,下标用 i 表示, X_i 为输入节点;隐藏层用字母 Y 定义,下标用 j 表示, Y_j 为隐藏层节点;输出层用字母 Z 定义,下标用 k 表示, Z_k 为输出节点;式中 w_{ij}^1 为输入层到隐藏层的权重, w_{jk}^2 为隐藏层到输出层的权重, f 为激活函数; q_1 为隐藏层各个节点阈值, q_2 为输出层各个节点阈值; L 为误差函数, \bar{Z}_k 为数据中实际值, Z_k 为输出节点; $\eta \in (0, 1)$ 为学习率,负号表示梯度下降的方向, $\frac{\partial L}{\partial w_{ij}}, \frac{\partial L}{\partial w_{jk}}$ 为梯度下降值。

输入层即有 5 个输入量,分别为 X_1 、 X_2 、 X_3 、 X_4 、 X_5 ,分别代表的是 BGR 三通道的像素值, $V(x)$ 三通道中最小的像素值和像素偏转角 θ 。五个特征参数,即输入层为 5 个节点。

隐藏层的节点没有明确的标准,通常是通过公式来确定。在实际问题中,隐含层节点数的选择首先是参考公式(11)来确定节点数的大概范围,然后用试凑法确定最佳的节点数。

$$j < \sqrt{(i - k)} + a. \quad (11)$$

其中, i 为输入层节点数; j 为隐含层节点数; k 为输出层节点数; a 为 $0 \sim 10$ 之间的常数。

对神经网络进行多次仿真训练,发现隐含层节点数为 10 时,能够达到最优的预测效果。

2.3 模型的训练及评估

基于图片的特征参数及对应的训练目标值作为神经网络的输入和输出,随机选取 70% (223 686 个样本) 的数据进行网络训练, 15% (47 933 个样本) 的数据进行神经网络模型的验证, 15% (47 933 个样本) 的数据进行模型的测试评估。通过 BP 神经网

络预测值的回归分析如图 3 所示,可以看出数据的训练结果优秀,多元统计系数 R^2 为 0.971 22,预测

值与实际值的一致性较好。因此,基于天空偏转模型的 BP 神经网络天空识别准确可靠。

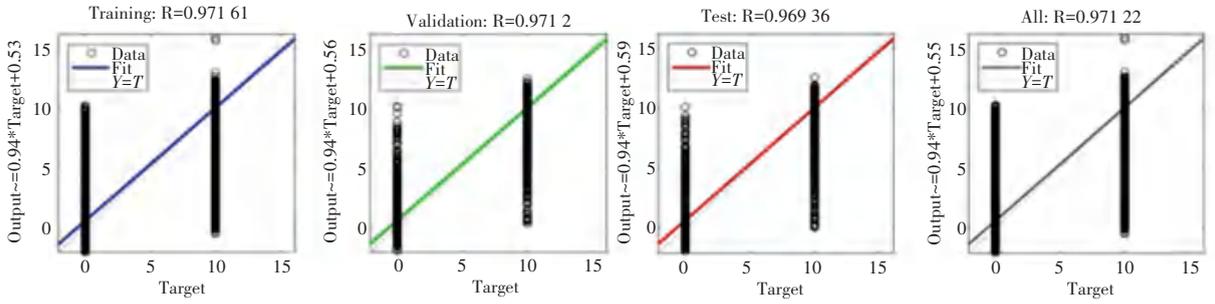


图 3 BP 神经网络的回归分析

Fig. 3 Regression analysis of BP neural network

2.4 误识别像素点的处理

图片中会有小部分的明亮干扰点,这些干扰点极易被误识别为天空像素。这些干扰点的主要来源有:白色建筑、斑马线、反光玻璃、大理石雕刻、白色广告牌,白色车辆和穿白色衣服的行人等。为了准

确的识别出天空,去除个别像素点的干扰,对错误识别的像素点进行剔除。首先,经过 BP 神经网络模型得到二分类图片(天空和非天空);其次,对二分类图片进行连通域计算;最后,舍弃连通域面积小于 10 的像素点。误识别像素剔除效果如图 4 所示。

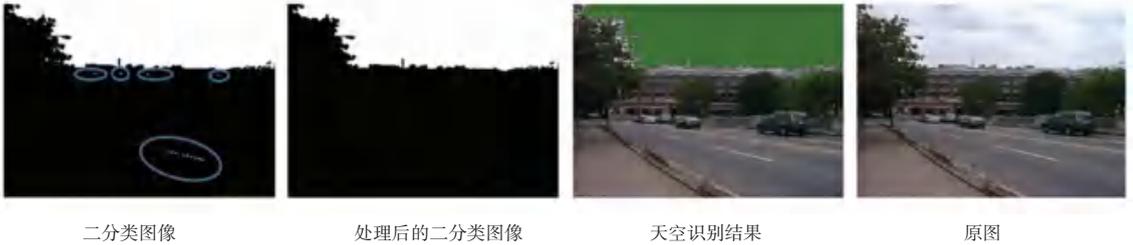


图 4 误识别像素剔除效果

Fig. 4 Error recognition pixel culling effect

通过图 4 中的二分类图片可以看出由于白色建筑和斑马线的存在,造成了较多天空区域像素点的错误识别。在经过本文算法处理后的二分类图像中,可以看出误识别的像素点被很好的剔除了,并准确的识别到了天空部分。

2.5 算法流程

本文的算法流程如图 5 所示。首先对待检测的图像 I 进行遍历,得到初步的二分类图像 I_1 ;计算二分类图像的连通域,删除面积小于 10 的连通域,得到精确的二分类图像 I_2 ;利用公式 (12) 得到最终的识别结果。

$$R = I \times (I_2 \times \frac{1}{255}), \quad (12)$$

其中: I 为待检测图像; I_2 为精确的二分类图像; R 为识别结果。

3 实验结果与分析

3.1 Camvid 数据集的使用

Camvid 是第一个带有语义标签的视频集合数据集,提供了 ground truth 标签,将像素与 32 个语义种类相关联。该数据集从行驶汽车的视角进行捕捉,解决了算法之间进行定量评估时实验数据集的不统一的问题。Camvid 的 32 个语义种类中包括天空,数据集中天空部分标注的像素为: $R = 128, G = 128, B = 128$,因此可以对不同算法的天空识别准确率进行定量评估。

3.2 算法的评价指标

在进行语义分割结果评价时,常常将预测出来的结果分为 4 个部分: true positive, false positive, true negative, false negative, 其中 negative 就是指非物体



图 5 算法的流程图

Fig. 5 Flow chart of the algorithm

标签的部分(可以直接理解为背景),positive 就是指有标签的部分,如图 6 所示。prediction 图被分成 4 个部分,其中大块的白色斜线标记的是 true negative (TN,预测中真实的背景部分),红色线部分标记是 false negative (FN,预测中被预测为背景,但实际上并不是背景的部分),蓝色的斜线是 false positive (FP,预测中分割为某标签的部分,但是实际上并不是该标签所属的部分),中间荧光黄色块就是 true positive (TP,预测的某标签部分,符合真值)。

IoU 指标就是交并比,在语义分割中一直作为标准度量被使用。交并比不仅仅在语义分割中使用,在目标检测等方向也是常用的指标之一,如图 7 所示。其定义为式(13):

$$IoU = \frac{target \cap prediction}{target \cup prediction} = \frac{a \cap b}{a \cup b}, \quad (13)$$

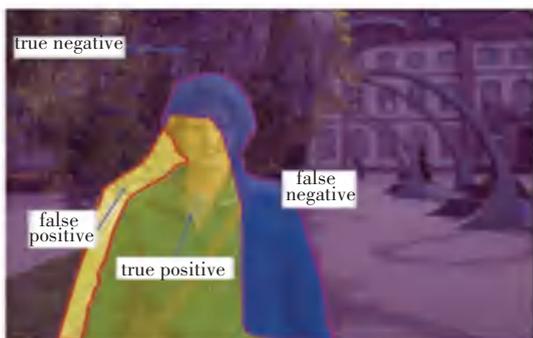


图 6 预测结果

Fig. 6 Prediction results



图 7 语义分割的评价指标——IoU

Fig. 7 Evaluation index of semantic segmentation -- IoU

3.3 算法主观评价

将本文算法与 Otsu 算法、YeHu 算法、Graph-cut 算法和 Mask-Rcnn 算法进行比较。对比实验基

于 ubuntu 系统,Python 语言和 Camvid 数据集实施,识别结果如图 8 所示。



图 8 算法主观评价

Fig. 8 Subjective evaluation of the algorithm

从图 8 可以看出 Otsu 和 Yehu 算法的识别效果最差,graph-cut 和 Mask Rcnm 的识别效果较好,本文的识别效果最为出色。Otsu 和 Yehu 算法存在天空错误识别的问题,具体表现在白色物体的错误识

别,均有较大面积的非天空区域被错误识别;Graph-cut 算法对天空部分的识别优于 Otsu 和 Yehu 算法,但是对天空细节的识别不够精确,建筑的误识别也比较严重;Mask Rcnm 算法识别的准确度较高,基本

没有错误识别,但是天空和背景的边界不够清晰,过于简单;本文算法的处理结果优于本文提到的其他算法,能够在准确识别图片中天空部分的同时排除白色背景的干扰,能够准确找到天空和背景的边界,尤其是天空和树木间的复杂边界部分,本文算法相比于其他算法结果优秀。

3.4 算法的定量分析

为了定量的分析算法的识别结果,本文对不同算法在 Camvid 数据集上的识别结果进行了统计,结果见表 2。从表 2 可以看到,5 种算法均能将图片中

的天空部分识别出来,但是不同算法对非天空部分的错误识别差别较大;Otsu 算法、YeHu 算法、Graph-cut 算法的 IoU 值较低,这主要是因为 3 种算法错误识别的天空像素点比较多;Mask-Rcnn 算法相比于其他算法,错误识别的天空部分面积最小,但是由于该算法识别到的天空像素总量相对较少,IoU 的值并没有很高;本文算法正确识别到的天空像素最接近 Camvid 数据集里原有天空像素面积,算法的天空识别效果较为出色,IoU 值最高,达到了 91.7%。

表 2 算法的定量分析

Tab. 2 Quantitative analysis of the algorithm

算法	识别到的天空像素	正确识别的天空像素	错误识别的天空像素	IoU
Camvid	4 146 423	4 146 423	0	100.0%
本文	4 315 324	4 048 547	264 737	91.7%
Otsu	7 150 967	4 140 762	3 010 205	57.9%
Yehu	8 248 612	4 165 289	4 083 323	50.6%
Graph-cut	11 976 871	4 129 972	7 846 899	34.4%
Mask-Rcnn	3 646 423	3 442 760	103 663	84.5%

4 结束语

本文提出了一种基于 BP 神经网络的天空区域检测算法,并创新性的提出了天空偏转模型,利用该模型成功的描述了天空像素点的特征。本文整理了多场景条件下的天空数据集,收集了不同天气情况下的天空像素特征,并成功用于 BP 神经网络进行训练,得到了很好的识别结果;将本文算法与 Otsu 算法、YeHu 算法、Graph-cut 算法和 Mask-Rcnn 算法进行主观评价和定量分析,从主观评价可以看出本文算法在对天空边缘细节的处理和排除背景中白色物体的干扰方面,相比与其他算法表现突出;从算法的定量分析可以看出本文算法的交并比领先于其他算法,达到了 91.7%。

参考文献

[1] GRIMMOND C S B, POTTER S K, ZUTTER H N, et al. Rapid methods to estimate sky-view factors applied to urban areas[J]. International Journal of Climatology, 2001.

[2] LUO J, ETZ S P. A physical model-based approach to detecting sky in photographic images[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2002, 11(3):201-12.

[3] GALLAGHER A C, LUO J, HAO W. Improved blue sky detection using polynomial model fit[C]// International Conference on Image Processing. IEEE, 2004.

[4] SHEN Y, WANG Q. Sky Region Detection in a Single Image for Autonomous Ground Robot Navigation[J]. INTERNATIONAL JOURNAL OF ADVANCED ROBOTIC SYSTEMS, 2013, 10(10):362.

[5] UEMURA T, HUIMIN L U, KIM H, et al. Transfer-learning deep convolutional neural network for classification of polyp candidates on CT colonography[J]. 2017.

[6] CHAPMAN L, THORNES J E. Real-Time Sky-View Factor Calculation and Approximation[J]. Journal of atmospheric and oceanic technology, 2004, 21(5):730-741.

[7] SCHMITT F, PRIESE L. Sky Detection in CSC-segmented Color Images.[C]// Visapp - Fourth International Conference on Computer Vision Theory & Applications. DBLP, 2009.

[8] LIANG J, GONG J, SUN J, et al. Automatic sky view factor estimation from street view photographs—a big data approach[J]. Remote Sensing, 2017, 9(5): 411.