A terrain segmentation method based on pyramid scene parsing-mobile network for outdoor robots

Botao Zhang¹,², Tao Hong², Rong Xiong¹ and Sergey A Chepinskiy³

Abstract
Terrain segmentation is of great significance to robot navigation, cognition, and map building. However, the existing vision-based methods are challenging to meet the high-accuracy and real-time performance. A terrain segmentation method with a novel lightweight pyramid scene parsing mobile network is proposed for terrain segmentation in robot navigation. It combines the feature extraction structure of MobileNet and the encoding path of pyramid scene parsing network. The depthwise separable convolution, the spatial pyramid pooling, and the feature fusion are employed to reduce the onboard computing time of pyramid scene parsing mobile network. A unique data set called Hangzhou Dianzi University Terrain Dataset is constructed for terrain segmentation, which contains more than 4000 images from 10 different scenes. The data set was collected from a robot’s perspective to make it more suitable for robotic applications. Experimental results show that the proposed method has high-accuracy and real-time performance on the onboard computer. Moreover, its real-time performance is better than most state-of-the-art methods for terrain segmentation.

Keywords
Mobile robot, terrain segmentation, visual perception, navigation, spatial pyramid pooling

Date received: 24 May 2021; accepted: 08 September 2021

Topic: AI in Robotics; Human–Robot/Machine Interaction
Topic Editor: Alessandro Di Nuovo
Associate Editor: Genci Capi

Introduction
The identification of feasible regions is crucial for robot navigation and path planning. Sensors such as laser, vision, and sonar are often used to perceive feasible regions, based on which various maps are constructed. Most existing methods can only divide regions into areas whether robots can travel through or not. However, when robots move across different terrains, as shown in Figure 1, their speeds and energy consumption vary dramatically. In addition, weather can also have some impact. For example, rainfalls will cause puddles on the road and increase passing risks, as shown in Figure 1(a). Humans and most mammals are more strategic in identifying specific terrains, such as grass, sand, and pebbles. Although recent studies have begun to shed light on identifying some outlined feasible areas, such as highways, concrete, and asphalt roads, they are still far from satisfactory. Terrain segmentation and recognition are

¹ State Key Laboratory of Industrial Control and Technology, Zhejiang University, Hangzhou, China
² School of Automation, Hangzhou Dianzi University, Hangzhou, China
³ Faculty of Control Systems and Robotics, ITMO University, St. Petersburg, Russia

Corresponding author:
Botao Zhang, State Key Laboratory of Industrial Control and Technology, Zhejiang University, Hangzhou 310058, China; School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China.
Email: billow@hdu.edu.cn
crucial for mobile robots to optimize navigation and path planning, especially for robots working in unstructured environments, such as agricultural robots, search, and rescue robots.

We have been worked on the identification and segmentation of outdoor terrains since 2015. As shown in Figure 2, we attempt to make a more detailed division of the terrains and provide support for the multiterrain map building and multiterrain-based navigation for outdoor robots. In 2020, a low-risk terrain recognition strategy was proposed to classify the outdoor terrains in Zhang et al.\textsuperscript{1} to make up for the deficiency of manual feature extraction and high misjudgment rate in risky terrains. In Wang et al.,\textsuperscript{2} a visual terrain classification method based on a convolutional neural network (CNN) and support vector machine was designed, focusing on terrain segmentation and its application to autonomous navigation of an outdoor robot.

In the navigation based on multiple terrains, a robot needs to automatically segment different terrains, including cement ground, grass, and puddles, to know its current surroundings quickly, plan appropriate paths, or make reasonable decisions.\textsuperscript{3–5} Yet, many challenges exist in terrain segmentation, as the outdoor environment is usually unstructured, uncertain, and sometimes interfered with by light, weather, pedestrians, vehicles, and other disturbances.\textsuperscript{6,7} Therefore, it is crucial to improve the accuracy and real-time performance of terrain segmentation.

Vision sensors are often used to collect information of terrains.\textsuperscript{8} Color space and geometric shape conversion methods are widely used in scenes with fewer changes and sharper contrasts.\textsuperscript{9–12} However, in complex outdoor environments, images collected by vision sensors often contain multiple types of terrains with similar characteristics. The segmentation of complex scenes usually requires pixel-level classification and has attracted the attention of many researchers. In Zhang et al.,\textsuperscript{13} the proposed approach extracted color features in multiple color spaces and combined texture features and geometric features to describe terrains. Then, feature selection and training based on the random forest algorithm are employed for terrain classification. In Liu et al.,\textsuperscript{14} the method combined the watershed algorithm and graph segmentation to solve the oversegmentation problem of images and obtained more precise boundaries of several terrains. By fusing the color histogram and local binary pattern features, the extreme learning machine was used to classify different regions. The feature extraction and screening process of most traditional terrain segmentation methods are cumbersome, with the representation of features limiting their performance, being unsuitable for the segmentation of complex geomorphic scenes.

Deep learning-based methods have developed rapidly in recent years and are widely used in semantic segmentation. In Long et al.,\textsuperscript{15} the proposed method applied $1 \times 1$ convolution kernel to replace the fully connected layer of CNN and then introduced the (fully convolutional network (FCN), by which realized the end-to-end semantic segmentation. Based on FCN, the SegNet was proposed in Badri-narayanan et al.,\textsuperscript{16} in which the pixel-level classification was made, and the characteristic is that the decoder performs a nonlinear upsampling of the input feature map.
According to the max-pooling index transmitted from the corresponding encoder. In Li et al.,17 the U-Net was applied to the terrain segmentation of aerial photography, and the cascaded dilated convolution module was used at the bottom layer to collect multiscale features. It could reduce the influence of interference and improve accuracy. In recent years, deep learning has been widely used in the field of robotics and autonomous driving. Researchers have proposed many novel methods for the detection and segmentation of driveable areas of vehicles and obstacles detection. In Frickenstein et al.,18 binary binarized driveable area detection network was proposed based on structured urban environments. It used binary weights and activations on encoders, bottlenecks, and decoders to reduce the amount of calculation and the storage space of the model. In addition, binary dilated convolution was used to strengthen the ability to extract features and improve segmentation accuracy.

Although vision-based deep learning methods can simplify the process of terrain segmentation, there are still some problems, for instance,

1. Most previous segmentation networks have complex structures and lack particular optimization for outdoor robots. A high-precision segmentation network will involve a tremendous amount of calculation and slow down the speed of robots. How to balance the segmentation accuracy and speed is challenging.

2. In terrain segmentation, only a few public data sets are available. A segmentation network with high accuracy relies on a data set with various terrains.19 It is very tedious and challenging to construct a diverse and influential terrain data set for terrain segmentation.

In this article, a terrain segmentation method based on pyramid scene parsing mobile network (Psp-MobileNet) is presented for outdoor robots. The main contributions of this article can be summarized as follows:

- A terrain segmentation network, Psp-MobileNet, was designed and applied to terrain segmentation in unstructured environments. The network structure is optimized to reduce the onboard computing time of segmentation based on depthwise separable convolution, spatial pyramid pooling, feature fusion, and so on. Experimental results show that the
The proposed Psp-MobileNet provides a good balance between segmentation accuracy and real-time performance.

- A practical and annotated terrain segmentation data set was built in this study. To make samples consistent with the real scene of robot navigation, more than 4000 terrain images are collected by a mobile robot and then the unstructured terrain data set called Hangzhou Dianzi University Terrain Data set is generated after artificial screening and image annotation.

### A terrain segmentation method based on Psp-MobileNet

Pyramid scene parsing network (PspNet), proposed in 2017, is good at mining global context information. It can generate a set of feature maps of different sizes through a pyramid pooling module. It strengthens the network detection of the relationship between the global and the local area of the picture. ResNet is used to extract features for PspNet, which introduces a large number of parameters into the network. Therefore, for feature extraction, this study chose MobileNetV1, a lightweight neural network proposed by Google in 2017. It replaces the standard convolutional layer in VGG (Visual Geometry Group) with depthwise separable convolution, significantly reducing the number of parameters.

#### The network structure of Psp-MobileNet

The proposed Psp-MobileNet inherits the structure of PspNet, including encoding and decoding paths, as shown in Figure 3. Some adjustments are made to the network structure to align with the terrain segmentation requirements of outdoor robots. The specific network parameters are presented in Table 1.

#### Encoding path

Considering the computational cost and the limited capabilities of onboard computers of outdoor mobile robots, the lightweight network MobileNetV1 is chosen for feature extraction. Although MobileNetV2 is relatively newer, it performed poorly in our tests (detailed test results of Psp-MobileNetV2 are shown in the “Settings for training and testing” section). The depthwise separable convolution used by the MobileNetV1 divides the convolution process into two steps.

**Depthwise convolution.** Each convolution kernel of depthwise convolution is responsible for one input channel. That is, one of them is convolved with only one convolution kernel. Therefore, depthwise convolution is usually applied to two-dimensional structures, and the number of its outputs is the same as the inputs. As shown in Figure 4, where the size of inputs is $H \times W \times 3$, depthwise convolution will use three convolution kernels of $k \times k (k \geq 2)$ to convolve with the input to extract three feature layers of $h \times w \times 3$. The parameters are determined by the following equation

$$N_{\text{depthwise}} = k \times k \times 3$$

(1)

**Pointwise convolution.** The depthwise convolution cannot extract features between channels. Therefore, it is necessary to extract the depth information of the input layer and adjust the number of channels through pointwise convolution. The operation of pointwise convolution is the same as
traditional convolution, except that its convolution kernel size is $1 \times 1 \times C$, where $C$ is the number of channels of the input layer.

As shown in Figure 5, when the feature map output by the depthwise convolution is $h \times w \times 3$, the number of output channels required by the depthwise separable convolution is four. Pointwise convolution will use four convolution kernels of $1 \times 1 \times 3$ to convolve with the feature map, and outputs are four feature layers of $h \times w \times 4$. The parameters are determined by the following equation

$$N_{\text{pointwise}} = 1 \times 1 \times 3 \times 4 = 12$$  \hspace{1cm} (2)

By the above convolutions, the total parameters of the depthwise separable convolution are $N_{\text{DP}}$ as shown in the following equation

$$N_{\text{DP}} = 3k^2 + 12$$  \hspace{1cm} (3)

whereas the total parameters of traditional convolution with the same situation are $12k^2$ as shown in the following equation

$$N_{\text{conv}} = k \times k \times 3 \times 4 = 12k^2$$  \hspace{1cm} (4)

In the case of $k \geq 2$, depthwise separable convolution reduces the parameters and speeds up the feature extraction. In the encoding path, to enhance the feature extraction, five depthwise separable convolutions are used in the fourth feature layer to ensure that the final feature layer is accurate.

### Decoding path

The characteristic of the decoding path is pooling the generated feature map with four different sizes. It can improve the segmentation effect of targets of various sizes in samples and avoid losing context relationships. The $1 \times 1$ pooling layer uses global pooling on the feature map to obtain global features in each channel. A more extensive size pooling layer usually has a stronger ability to get trim details. To prevent possible repetition after pooling, the size of the pooling layer is set to alternate odd and even numbers. To make up for the loss of crucial information caused by pooling and improve the accuracy and generalization ability, a feature map is combined with four pooling layers for feature fusion. Standard convolution is used to integrate the contextual information of the fused feature

| Input | Size | Operator | c | n | s | Output |
|-------|------|----------|---|---|---|--------|
| Input image | $576^2 \times 3$ | Standard conv2D | 32 | 1 | 2 | fl |
| f1 | $288^2 \times 32$ | Depthwise separable conv | 128 | 1 | 2 | f1 |
| f2 | $144^2 \times 128$ | Depthwise separable conv | 256 | 1 | 1 | f2 |
| f3 | $72^2 \times 512$ | Depthwise separable conv | 512 | 5 | 1 | f3 |
| f4 | $36^2 \times 512$ | AveragePooling2D | 4 | 1 | Pool[f3] |
| pool[3] | $36^2 \times 512$ | Standard conv2D | 512 | 1 | Pool[f3] |
| ConcatMap | $36^2 \times 512$ | Concatenate | 1 | ConcatMap |
| FinalMap | $288^2 \times N$ | softmax | $N$ | FinalMap |
| FinalMap | $288^2 \times N$ | return | $N$ | Net output |

$c$: the number of channels output by the convolution; $n$: the number of repetitions; $s$: the convolution strides; $N$: the number of categories classified.

Figure 4. The process of depthwise convolution.

Figure 5. The process of pointwise convolution.
maps. After resizing and upsampling the feature map, the number of channels is adjusted to the number of categories. Each channel corresponds to a segmentation category, and each pixel value in the channel represents the weight of the pixel belonging to the corresponding category. And then, the softmax function is used to normalize the weight to the probability value.

**HDU-terrain data set**

To make the network more adaptable to workspaces for outdoor mobile robots, we constructed the HDU-Terrain Dataset. As shown in Figure 6, a Scout MINI robot equipped with a monocular camera was used to collect samples with a resolution of 640 × 480 at 0.5 s/frame from 10 different scenes. Four thousand and seven original terrain images were retained after manual elimination of repetitive ones. Afterward, the Labelme semantic annotation software was employed to conduct semantic annotation. Samples were divided into 12 semantic categories: grass, concrete, masonry, mud, woodland, bumpy road, tree, obstacle, person, ditch, water, and background.

As shown in Figure 7, we have made a more detailed division of outdoor terrains compared with Zhang et al.\(^1\) and Wang et al.\(^2\). Based on the characteristics of common working spaces for outdoor robots, in HDU-Terrain Dataset, the feasible terrains were divided into grass, woodland, mud, cement, masonry, and bumpy road. Bumpy road includes gravel roads and uneven hard roads, which have a high traffic cost for wheeled robots. Unfeasible terrains are divided into trees, obstacles, persons, ditches, and water, where obstacles include rocks, railings, vehicles, street lamps, and other objects. The HDU-Terrain Dataset is publicly available at https://hduraslab.github.io/pages/research.html.

Figure 8 shows the overall percentage of semantic categories of the entire data set. It can be seen from Figure 8 that there is always a tree in almost every picture. Mud, person, and ditch occupy relatively small percentages. This proportion provides a helpful reference for the comparison in the “Ablation Study for MobileNet” section.

HDU-Terrain Dataset contains 10 sub-data sets. Figure 9 shows its specific composition and semantic categories distribution. The building12_trail sub-data set was gathered in sunny weather, and images were suffered from strong light interference. During the photographing of the yueya_trail sub-data set, the robot moved quickly, resulting in many blurred frames. In addition, many images in the data set contain small and irregular targets, such as various obstacles and persons.

**Experiments of terrain segmentation**

**Settings for training and testing**

The HDU-Terrain Dataset is randomly divided into a training set, a validation set, and a test set. The images accounted for 81%, 9%, and 10% of the total data set, respectively. All models are trained through the training set on a PC with NVIDIA RTX2070 (Santa Clara, CA, USA) Super GPU (Graphic Processing Unit). We used the...
“poly” learning rate policy.\textsuperscript{22} The current learning rate during the training process is shown in the following equation
\[ \text{lr} = \text{base\_lr} \times \left(1 - \frac{\text{iter}}{\text{max\_iter}}\right)^{\text{power}} \] (5)
where base\_lr is the initial learning rate, iter represents the current number of iterations, and max\_iter represents the number of iterations. We set base\_lr to 0.01 and power to 0.9. Due to the limited video memory of the GPU, the batch size is set to 4. A total of 50 epochs are trained, and 812 steps are trained in each epoch, which is equivalent to 40 K iterations. Using the Adam optimizer, the weight decay is set as 0.001. Considering the time cost of training, transfer learning is introduced into the network.

Five semantic segmentation metrics are used to evaluate the models: the number of model parameters, segmentation speed (frame/s), mean intersection over union (MIoU), mean pixel accuracy (MPA), and pixel accuracy (PA). PA and intersection over union (IoU) are derived from the confusion matrix and they are generally accepted metrics for semantic segmentation.\textsuperscript{15} The computation of PA is shown in the following equation
\[ \text{PA} = \frac{\sum_{s=0}^{n} p_{ss}}{\sum_{s=0}^{n} \sum_{t=0}^{n} p_{st}} \] (6)
where \( n \) represents the number of categories to be divided. In this article, 0–11 represents 12 semantic categories in Figure 7, respectively. \( p_{st} \) is the number of \( s \)-class pixels predicted to be \( t \)-class in images. If \( s = t \), the pixel points are classified correctly and \( p_{ss} \) is the number of correctly classified pixels. MIoU and MPA are calculated by the following equations\textsuperscript{23}
\[ \text{MIoU} = \frac{1}{n + 1} \sum_{s=0}^{n} \left( \sum_{t=0}^{n} \frac{p_{ss}}{p_{st}} + \sum_{t=0}^{n} p_{ts} - p_{ss} \right) \] (7)
The model test is divided into two parts. The test set in HDU-Terrain Dataset is employed to test the accuracy, and the Jetson AGX Xavier (Santa Clara, CA, USA) is used to test the real-time performance. Jetson is an embedded system designed by NVIDIA. It has strong image processing ability. Therefore, we apply it to the terrain segmentation for robots.

Ablation study for MobileNet

MobileNet is used to build encoding paths because of its lightweight characteristic, but there are some differences in the features extracted by different MobileNets. The performance of several terrain segmentation networks composed of varying encoding paths is tested in this section.

1. Psp-MobileNet: The proposed and recommended network.
2. Psp-MobileNet-F5: Compared with Psp-MobileNet, we carry out one more depthwise separable convolution in the encoding path and extract a total of five layers of feature maps. The structure of the decoding path is PspNet. This test aims to verify that more feature extraction layers do not improve the algorithm’s performance.
3. Psp-MobileNet-F3: Compared with Psp-MobileNet, we carry out a less depthwise separable convolution in the encoding path and extract a total of three layers of feature maps. The structure of the decoding path is also PspNet. The purpose of this test is to verify that fewer feature extraction layers can reduce the accuracy.
4. Psp-MobileNetV2: MobileNetV2\(^{24}\) is employed to extract features on the encoding path, and the structure of the decoding path is also PspNet. This comparative test is to verify whether MobileNetV2 or MobileNetV1 is more suitable to combine with PspNet.

As presented in Table 2, Psp-MobileNet has the highest segmentation accuracy. Although there are differences among the parameters of the four networks, the memory occupied by the corresponding model does not affect the regular operation of the mobile robots. In addition, the four networks have the same segmentation speed. We believe that the MobileNet with four feature extraction layers is the most suitable for the proposed Psp-MobileNet.

Comparison of terrain segmentation

To evaluate the performance of the proposed network, several segmentation methods are designed and compared as follows.

1. Seg-MobileNet: The semantic segmentation network SegNet\(^{16}\) and MobileNet are used to design this segmentation network.
2. U-MobileNet: The semantic segmentation network U-Net\(^{25}\) and MobileNet are used to design this segmentation network.
3. Deep-MobileNet: The semantic segmentation network DeepLabV3\(^+\)\(^{26}\) and MobileNet are used to design this segmentation network.
4. Psp-MobileNet: The proposed network.

To ensure the reliability of the comparison, the training settings of the four networks are the same.

As presented in Table 3, compared with the other three methods, the proposed Psp-MobileNet has the highest segmentation accuracy. Its MIoU and MPA are 4.8% and 3.0%
higher. Psp-MobileNet also maintains the fastest segmentation speed, which is twice as fast as others.

It can be seen from Figures 10 and 11 that all the methods perform well on large areas of grass, masonry, bumpy road, woodland, and tree. There is no significant difference in the accuracy for these categories. But for concrete, ditch, mud, obstacle, and person, Psp-MobileNet has apparent advantages, especially for the person. As presented in Tables 4 and 5, IoU and PA of Psp-MobileNet for the person are 1.5 times and 1.3 times
that of other methods, respectively. It proves that Psp-MobileNet has an excellent adaptive ability for small targets or low proportion targets. In addition, the segmentation accuracy of Psp-MobileNet for each category is almost the highest among all these methods. It shows that Psp-MobileNet has no significant weakness in the terrain segmentation of HDU-Terrain Dataset in segmentation accuracy. However, it is about twice as fast as the other methods, as presented in Table 3. The real-time performance for segmentation is critical for a mobile robot.

### Table 4. The IoU of all categories in the Psp-MobileNet, the Seg-MobileNet, the U-MobileNet, and the Deep-MobileNet.

| Network          | Grass | Concrete | Masonry | Bumpy road | Woodland | Tree | Ditch | Water | Mud | Obstacle | Person |
|------------------|-------|----------|---------|------------|----------|------|-------|-------|-----|----------|--------|
| Seg-MobileNet    | 89.40 | 79.61    | 94.38   | 92.98      | 92.37    | 90.88| 70.87 | 86.57 | 65.23| 64.35    | 30.60  |
| U-MobileNet      | 89.77 | 82.69    | 94.55   | 92.72      | 92.37    | 90.84| 65.33 | 83.56 | 64.02| 64.48    | 27.89  |
| Deep-MobileNet   | 90.38 | 84.91    | 94.89   | 92.20      | 92.33    | 90.56| 66.53 | 88.30 | 63.91| 61.78    | 28.19  |
| Psp-MobileNet    | 91.43 | 85.98    | 95.33   | 93.09      | 93.36    | 90.88| 76.80 | 89.42 | 67.67| 68.14    | 43.95  |

Psp-MobileNet: pyramid scene parsing-mobile network; IoU: intersection over union.

### Table 5. The PA of all categories in the Psp-MobileNet, the Seg-MobileNet, the U-MobileNet, and the Deep-MobileNet.

| Network          | Grass | Concrete | Masonry | Bumpy road | Woodland | Tree | Ditch | Water | Mud | Obstacle | Person |
|------------------|-------|----------|---------|------------|----------|------|-------|-------|-----|----------|--------|
| Seg-MobileNet    | 93.89 | 87.15    | 96.22   | 95.42      | 94.67    | 95.68| 80.80 | 98.51 | 69.57| 78.66    | 45.98  |
| U-MobileNet      | 92.67 | 84.54    | 96.27   | 95.82      | 95.14    | 96.41| 78.28 | 98.39 | 69.15| 82.13    | 44.21  |
| Deep-MobileNet   | 93.32 | 86.73    | 96.75   | 95.66      | 95.47    | 95.41| 74.83 | 95.86 | 68.84| 78.55    | 41.67  |
| Psp-MobileNet    | 94.67 | 87.44    | 97.14   | 95.74      | 95.00    | 96.20| 85.08 | 97.73 | 70.60| 82.40    | 57.59  |

Psp-MobileNet: pyramid scene parsing-mobile network; PA: pixel accuracy.

Figure 12. The segmentation results of the Psp-MobileNet, the Seg-MobileNet, the U-MobileNet and the Deep-MobileNet. (a) Image, (b) Ground truth, (c) Seg-MobileNet, (d) U-MobileNet, (e) Deep-MobileNet, and (f) Psp-MobileNet. Psp-MobileNet: pyramid scene parsing-mobile network.
That is why we believed that Psp-MobileNet is more suitable for landscape segmentation.

Figure 12 shows the segmentation results of the above four methods. There are many unstructured terrains in outdoor environments, which is very challenging. SegMobileNet and U-MobileNet are prone to segment terrains mistakenly in the corner areas of images. Psp-MobileNet could not only segment complex small targets such as persons and obstacles but also accurately describe the edges of large-area terrains. Its false positives and false negatives areas are significantly smaller than other methods.

Conclusions
This article presented a terrain segmentation method for field robots based on Psp-MobileNet by combining PspNet and MobileNet. Compared with several state-of-the-art methods, it has a good balance between accuracy and real-time performance. We also established a more realistic and annotated terrain segmentation data set by sampling terrains from a robot’s perspective. This study contributes to a more detailed division of the terrains and can support the multiterrain map building and multiterrain-based navigation for mobile robots. Experimental results show that the proposed terrain segmentation method based on Psp-MobileNet has excellent real-time performance and maintains acceptable accuracy on an onboard computer often used by robots.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work is supported by the Key Research and Development Project of Zhejiang Province (Grant No. 2019C04018) and the Science and Technology Project of Zhejiang Province (Grant No. 2019C01043).

ORCID iDs
Botao Zhang https://orcid.org/0000-0002-7826-3121
Sergey A Chepinskiy https://orcid.org/0000-0002-0193-8688

References
1. Zhang Q, Zhang B, Lv Q, et al. Low-risk terrain classification based on convolutional neural network. Control Theory Appl 2020; 37(9): 1944–1950.
2. Wang W, Zhang B, Wu K, et al. A visual terrain classification method for mobile robots’ navigation based on convolutional neural network and support vector machine. Trans Inst Meas Control 2021. DOI:10.1177/0142331220987917.
3. Lenskiy AA and Lee J. Terrain images segmentation in infrared spectrum for autonomous robot navigation. In: International forum on strategic technology, Ulsan, Korea (South), 13–15 October 2010, pp. 33–38.
4. Das A, Giri R, Chatterjee A, et al. Self-navigation and positioning robot-SNAPBot. In: International conference on inventive computation technologies, Coimbatore, India, 26–28 February 2020, pp. 1–5.
5. Sun J, Meng Y, Tan J, et al. A vision-based perception framework for outdoor navigation tasks applicable to legged robots. In: Chinese automation congress, Jinan, China, 20–22 October 2017, pp. 2894–2899.
6. Schilling F, Chen X, Folkesson J, et al. Geometric and visual terrain classification for autonomous mobile navigation. In: IEEE/RSJ international conference on intelligent robots and systems, Vancouver, BC, Canada, 24–28 September 2017, pp. 2678–2684.
7. Wang T, Sung D, and Lin P. Terrain classification, navigation, and gait selection in a leg-wheel transformable robot by using environmental RGBD information. In: International automatic control conference, Taoyuan, 4–7 November 2018, pp. 1–1.
8. Suryamurthy V, Raghavan VS, Laurenzi A, et al. Terrain segmentation and roughness estimation using RGB data: path planning application on the CENTAURO robot. In: IEEE-RAS international conference on humanoid robots, Toronto, ON, Canada, 15–17 October 2019, pp. 1–8.
9. Chen M, Dai X, Lv Z, et al. Terrain classification in field environment with grid-based SURF feature. In: Chinese control conference, Nanjing, China, 28–30 July 2014, pp. 8253–8258.
10. Lenskiy AA and Lee J. Rugged terrain segmentation based on salient features. In: International conference on control, automation and systems, Gyeonggi-do, Korea (South), 27–30 October 2010, pp. 1737–1740.
11. Khan YN, Komma P, Bohllmann K, et al. Grid-based visual terrain classification for outdoor robots using local features. In: IEEE symposium on computational intelligence in vehicles and transportation systems proceedings, Paris, France, 11–15 April 2011, pp. 16–22.
12. Zou Y, Chen W, Xie L, et al. Comparison of different approaches to visual terrain classification for outdoor mobile robots. Pattern Recognit Lett 2014; 38: 54–62.
13. Zhang H, Dai X, Sun F, et al. Terrain classification in field environment based on random forest for the mobile robot. In: Chinese control conference, Chengdu, China, 27–29 July 2016, pp. 6074–6079.
14. Liu F, Ma X, Li X, et al. Terrain recognition for outdoor mobile robots. In: Chinese automation congress, Jinan, China, 20–22 October 2017, pp. 4257–4262.
15. Long J, Shellhamer E, and Darrell T. Fully convolutional networks for semantic segmentation. IEEE Trans Pattern Anal Mach Intell 2015; 37(4): 640–651.
16. Badrinarayanan V, Kendall A, and Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 2017; 39(12): 2481–2495.
17. Li X, Jiang Y, Peng H, et al. An aerial image segmentation approach based on enhanced multi-scale convolutional neural network. In: International Conference on Industrial Cyber Physical Systems, Taipei, 6–9 May 2019, pp. 47–52.
18. Frickenstein A, Vemparala M, Mayr J, et al. Binary DAD-Net: binarized driveable area detection network for autonomous driving. In: International Conference on Robotics and Automation, Paris, France, 31 May–31 August 2020, pp. 2295–2301.
19. Liyanage DC, Hudjakov R, and Tamre M. Hyperspectral imaging methods improve RGB image semantic segmentation of unstructured terrains. In: International Conference on Mechatronic Systems and Materials, Bialystok, Poland, 1–3 July 2020, pp. 1–5.
20. Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network. In: IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017, pp. 6230–6239.
21. Howard AG, Zhu M, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications. 2017, arXiv:1704.04861.
22. Chen L, Papandreou G, Kokkinos I, et al. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell 2018, 40(4): 834–848.
23. Garcia-Garcia A, Orts-Escolano S, Oprea SO, et al. A review on deep learning techniques applied to semantic segmentation. 2017, arXiv:1704.06857.
24. Sandler M, Howard A, Zhu M, et al. MobileNetV2: inverted residuals and linear bottlenecks. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018, pp. 4510–4520.
25. Ronneberger O, Fischer P, and Brox T. U-Net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015, pp. 234–241.
26. Chen LC, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: European Conference on Computer Vision, Munich, Germany, 8–14 September 2018, pp. 833–851.