2018 3rd International Conference on Advanced Robotics and Mechatronics (ICARM)

A Convolutional Neural Network Based on Double-tower Structure for Underwater Terrain Classification

Wei Liu, Rongxin Cui*,Yang Li, Shuqiang Liu
Research & Development Institute of Northwestern Polytechnical University in Shenzhen
Northwestern Polytechnical University
Shenzhen 518057, P.R. China
Email:* r.cui@nwpu.edu.cn

***Abstract*—Terrain classification plays a critical role in all robot systems especially in unknown environments. In recent years, researchers have proposed various algorithms to improve the efficiency and accuracy of terrain classification. Nevertheless, these methods still have some deficiencies in classification efficiency. In this paper, a double-tower convolutional neural network has been designed to implement end-to-end underwater terrain classification. The matched sonar image and visual image constitute an image pair, which is obtained at the same time by the sonar sensor and the visual sensor of the robot or underwater vehicle. The corresponding image pairs are set to be the input of the convolutional neural network, and the output of the network is the classification of the terrain. Then, terrain features from sonar and visual images are simultaneously applied to achieve terrain classification. Therefore, an end-to-end convolutional neural network with a classification function has been established in this paper.**

Keywords—sonar image, visual image, terrain classification, convolutional neural network

I. INTRODUCTION

In recent years, intelligent mobile robots have become a frontier research hotspot. Intelligent mobile robots involve the knowledge of multiple disciplines such as environmental awareness, information processing, control engineering, and are currently one of the most active areas for the development of science and technology.

At present, many researchers have put forward many good robust classification algorithms in the terrain classification of robots. Christie [1] proposed a system to achieve online real-time terrain classification for the legged robots, by using the sound signal generated during the movement of the robot. A multiclass Support Vector Machine (SVM) was utilized to extract the 32-dimensional feature vector from the acoustic data recorded by the microphone mounted on the robot as the input of classification. 7 types of information of different terrains have been utilized to train the SVM so as to realize real-time terrain classification. A novel terrain classification method based on Random Forest has been presented by Zhang [2] et al. This method extracted a large number of candidate features including color, texture, and geometric features at the first step. The experimental results showed that the feature selection method based on the random forest can effectively extract feature subsets highly correlated with the terrain, with higher classification accuracy and faster classification speed. Wu et al.[3] showed a terrain classification algorithm for small robots. The algorithm used the dynamic ground pressure data from a pressure sensor collected by a robot bipedal wheel to train a SVM to implement a terrain classifier. The results showed that pressure sensor data, combined with information such as motor torque and robot gait, was sufficient to distinguish among hard, slippery, grassy, and granular terrain types.

The rapid development of artificial intelligence technology provides new algorithm theories and opportunities for robots to achieve terrain classification. For examples, Hadsell et al. [4] and Muller et al. [5] combined supervised and unsupervised learning in deep learning networks. An effective algorithm suitable for terrain classification by terrestrial robots was proposed, and the classification effectiveness of the algorithm was excellent. Manjanna et al.[6] used unsupervised clustering method to classify the terrain based on the signal characteristics of multiple sensors in the walking behavior of the robot, and realized the robot's real-time perception of the terrain and assisted gait planning and optimization. Àngel et al. and Suger et al. [7-8] designed a semi-supervised learning algorithm that classifies multiple terrains with fine-grained terrain with the same degree of roughness by learning part of the marked data. It makes the autonomy of the robot in terrain classification get higher and higher. Otte [9] et al. explored the use of texture-based image descriptors. In this method, features are extracted in the form of a local three-valued model in a catastrophic image sequence and structural classification is performed using a Recurrent Neural Network (RNN). Abhinav [10] et al. proposed an end-to-end terrain classification model on the base of convolutional neural networks and Long-Short Term Memory, which makes full use of spatial information and time information. This model is currently widely applied for terrain classification of terrestrial robots.

Unlike intelligent robots in the terrestrial environment or restricted to indoor activities, underwater or amphibious intelligent robots need to be active in an unknown underwater environment. The intrinsic unstructured environment and the nature of the transmission medium make the acquired underwater images unclear, resulting in the lack of obvious features in the terrain area of interest. It also brings up the challenge to achieve terrain classification.

In the demand of the underwater intelligent robot for the underwater terrain classification, many researchers have been studying underwater terrain classification and laid a foundation for underwater intelligent robots to perform activities in an unknown underwater environment. Rao [11] et al. used an auto-encoder network to extract features from visual and remote sensing data, and proposed a deep learning network joint distribution method between underwater sonar images and visual images. Then underwater terrain classification is performed from the perspective of semantic information. Their research shows that the process of learning multiple types of data improves the classification accuracy. Compared to the scenarios in terrestrial environments, there are currently limited studies on underwater terrain classification algorithms. The classification of underwater terrain has important significance for the development of robots.

In this paper we analyze typical algorithms for classification of terrestrial with considering good performance of the deep learning model in image classification. A new network structure is proposed for terrain classification in underwater environments. This paper mainly describes the process of establishing the data set in the section III. The section IV introduces the convolutional neural network model of the double tower structure we built. Finally, the section V summarizes this paper and discusses the advantages and disadvantages of the proposed model.

The main contributions of this article can be summarized as follow.

- We summarize the traditional terrain classification methods and discuss the research characteristics that the deep learning method has shown in terrain classification.
- We propose a deep learning method to achieve end-to-end terrain classification. A double-tower network structure is established to address the marine terrain classification difficulties.
- We present the future of a better terrain classification based on deep learning methods.

II. RELATED WORK

At present, more and more researchers have realized the excellent performance of artificial intelligence technology in the classification of underwater terrain, and began to use the deep learning method to achieve underwater terrain classification.

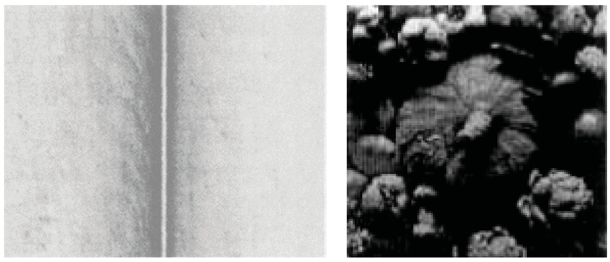
Considering the complex underwater environment, the way to obtain underwater terrain information is also different from the case in terrestrial environment. In general, the underwater terrain information can be obtained through a variety of sensors to get a variety of terrain data such as sonar, underwater cameras with sufficient resolution, touch sensors, and so on. Liu[12] proposed information contained in many images that can be integrated in the same scene as a composite image. Based on this idea, the pixel-level image fusion is recognized as an important method in many fields such as medical imaging, digital photography, and remote sensing. Our paper considers pixel-level image fusion through terrain information from multiple sensors and this method is of great significance for terrain classification. At the same time, some researchers have proposed the advantages of deep learning algorithms in image fusion. In the past three years, researches in various fields based on deep learning has become a hotspot, such as for digital photography (multi-focus image fusion, multi-exposure image fusion) [13-16], multimodal imaging (medical image fusion, infrared / visible light image fusion) [17-19], and a variety of image fusion methods basing on deep learning are proposed. These studies have shown that the deep learning network has the ability of image fusion, meanwhile, combining the advantages of the deep learning in image classification [20-21]. we establish a deep learning model basing on convolutional neural networks to realize underwater terrain classification based on the fusion of the sonar images and visual images.

III. DATASET

We focus on the terrain classification requirements of mobile robots when they are working in underwater environments. Sonar images from robot's sonar sensors and visual images from underwater cameras are used as basic data for terrain classification. The dataset is mainly composed of sonar images that have been labeled by experts, and visual images that match them. The terrain images contained in the dataset all show a one-to-one correspondence, that is to say, the sonar image correspond to the visual image and both images represent the same terrain scene at the same time, then the experts set the terrain classification labels of the images.

A. Sonar image and visual image

Side scan sonar is an important tool for accomplishing many tasks in the underwater environment, such as searching, detecting, etc. The sound waves have been launched and reflected by the sonar to form a sonar image to discover underwater objects. The received signal is transmitted via a streamer to the display unit on the deck. The display unit shows high-resolution sonar images of the bottom of the lake or other underwater objects, so we can obtain sonar images[22]（As shown in Figure 1(a)） of the underwater terrain by installing side scan sonars on the robot.



(a) (b)

Fig.1. Image Example

Through installing a high-resolution underwater camera to obtain visual image data (As shown in Figure 1(b)) [23], compared with ordinary optical images, underwater visual images have disadvantages such as unclear contrast and severe noise, which causes difficulty in terrain identification and classification. In order to match the sonar image with the visual image, it is necessary to ensure that the visual image captured by the camera and the sonar image are the same topography of the seabed environment.

B. Image enhancement

The purpose of image enhancement is to highlight the advantages of sonar images and visual images, and make full use of its advantages to achieve the terrain classification in the later process.

1) Underwater visual image enhancement

Underwater visual images are different from ordinary optical images due to the features of water absorbs and scatters light, mainly caused by limited illumination range and other uncertainties, and low contrast in underwater environments.

Based on the requirements of extracting image features in this paper, image spatial enhancement techniques are used to enhance the image information. Common image spatial enhancement techniques include logarithmic transformation, contrast stretching, histogram equalization, and sharpening [24]. Histogram equalization is a commonly used spatial enhancement method. It mainly considers the statistical grouping of the values of each pixel of the image, but does not consider the position information, and it is easy to enhance noise and image details together. In order to avoid the interference caused by the noise information enhancement, we need to denoise before the histogram equalization.

The underwater image noise information mainly contains Gaussian white noise and water particle scattering noise. In order to ensure the quality of the image after histogram equalization, we need to design a low-pass filter and homomorphism filter to eliminate Gaussian white noise and water particle scattering noise, respectively. The specific parameters of the filter need to be adjusted and set according to the experimental results.

2) Side scan sonar image enhancement

The side scan sonar image is disturbed by the relatively serious noises such as marine environment noise and equipment noise during the formation process, and the uneven echo intensity caused by suspended particles and asymmetry structure in the ocean, resulting in the side scan sonar images have more complex gray levels and lower image contrast. Therefore, the purpose of sonar image enhancement is mainly to achieve image contrast enhancement based on noise removal.

At present, the denoising methods of side scan sonar images mainly include spatial domain filtering and multi-scale transforms filtering. We choose the spatial domain filtering method to achieve the purpose of denoising by directly transforming the image pixel gray scale. The side scan sonar data is generally 16 bits (i.e., the gray value range is [0,65535]), and the gray distribution is more concentrated, mainly in the low gray area. The purpose of our image enhancement is to highlight the target in the image, and at the same time, the image features of the target should be kept obvious. Therefore, we perform filtering on the side scan sonar image firstly. The commonly used filtering methods are based on the wavelet function filter processing, such as Haar Wavelet function, Daubechies wavelet function, Coifelets, and so on. Wavelet processing has good effect on the filtering of side scan sonar images, and shows good ability in both edge retention and denoising [25]. In order to increase the anomaly of the target and the background in the image and improve the recognition efficiency of the target, the reduction of the color gradation is adopted, and the threshold is set so that the sonar image has a lower limit gray and an upper limit gray that are significantly different from the normally continuously varying gray. The high-gray pixels with low probability of occurrence and the low-gray pixels with low recognition efficiency are represented by two special gray values.

3) Dataset establishment

Each training input should consist of two image patches, one patch is from the sonar terrain image from the side scan sonar, and the other patch is from the topographic image captured by the camera. Assuming that the image of the underwater environment used to create the data set includes 10 types of terrain. Which needs to be emphasized is that they all come from image-enhanced terrain images:

$$\langle P^S_{n\times n}(\mathbf{p}), P^V_{n\times n}(\mathbf{p})\rangle$$

Where $P^S(\mathbf{p})$ is an image patch from the $\mathbf{p}$ position in the sonar image P^S, and $P^V(\mathbf{p})$ is a topographic image patch from the same position in the visual image P^V. That is to say, for the sonar image and the visual image describing the seafloor environment, P^S and P^V are two images that describe the same seafloor terrain. $P^S(\mathbf{p})$ can be selected according to topographical features. For example, the topographic feature in a terrain image is represented as sandy ground, then the center position coordinate $\mathbf{p}$ can be selected at this position, and an n*n image patch can be selected as $P^S_{n\times n}(\mathbf{p})$, and meanwhile look for an image patch of the same size corresponding to the position in the corresponding P^V image, denoted as $P^V_{n\times n}(\mathbf{p})$, then mark the terrain image which has a feature is "sand" and its label is [0 0 0 1 0 0 0 0 0 0]. The position belongs to its classification is marked as 1, and the others are marked as 0.

Above all, the steps of building a database are: (1) selecting a terrain image patch $P^S_{n\times n}(\mathbf{p})$ having terrain features in the sonar image P^S; (2) selecting the terrain image patch $P^V_{n\times n}(\mathbf{p})$ having the same terrain feature, the same location and the same size in the visual image P^V; and (3) a classification label is set for the formed image pair. Repeat the above steps, the dataset is formed by image pairs with classification labels, which consist of the sonar terrain image from the side scan sonar and the visual terrain image from the underwater camera. Each sample in the dataset can be described as$\langle (P^s(\mathbf{p}), P^v(\mathbf{p})), \mathrm{C}\rangle$, where C represents the terrain classification.

IV. TECHNICAL APPROACH

Considering the excellent performance of deep learning in image fusion and classification, this paper proposes A Convolutional Neural Network Based on Double-tower Structure. The network is an end-to-end deep learning network to achieve terrain classification which can simultaneously extract effective information from sonar images and visual images. The advantage of this algorithm model is that it simultaneously learns terrain features from sonar images and visual images, and on this basis it extracts and fuses features through the neural network layer. It does not require specialized or complicated mathematical operations to extract features from the image and make full use of the effective information of the sonar and visual images. This section mainly illustrates the network structure and describes the training methods.

A. Network architecture

The structure of our proposed network is shown in Figure 2. The structure contains 9 network layers. There is a branch structure in the network, namely the C1 layer to the C5 layer of the network. The C1 and C3 network layers are convolutional network layers. The kernel of the C1 network layer is set according to the type of the input image. If the input images are gray level images, the 32*4*4*1 kernels are set, that is to say, the number of the kernel is 32, and 4*4 refers to the size of the kernel and performs convolution training in one channel. If the input images are color images, the 32*4*4*3 kernels are set, and performs convolution training is achieved in 3 channels. The three channels refer to R, G, and B channel, respectively. Correspondingly, the kernel of the C3 layer is set to 64*4*4*1 or 64*4*4*3. Both C2 and C4 are pooling network layers with sigmoid function as the activation function, and their pooling layer template size is set to 2*2. The sigmoid function is used as the activation function to activate the network layer in the pooling network layer training process. Then perform the pooling operation. The C5 layer is a fully connected layer with 6400 neurons set. The purpose of the C1 to C5 network layers is to learn the topographical features contained in the input images P^S and P^V.

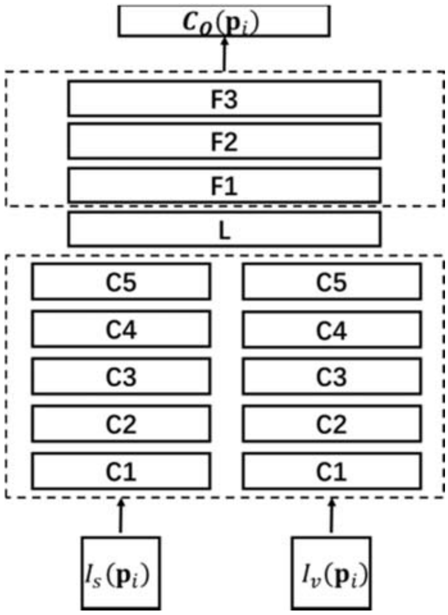


Fig.2. The network structure proposed by this paper

The L network layer is a fully connected network layer. The terrain features learned in the C5 layer will be connected at the L network layer. That is to say, the L network layer includes all features obtained from the sonar image patch $P^S(\mathbf{p})$ and the visual image patch $P^V(\mathbf{p})$, which includes all terrain features in $\mathbf{p}$ position. For all the features in $\mathbf{p}$ position， $\mathbf{p}$ is a vector (x, y) representing the center position of the image block, the number of neural units in the L layer is 10,000. Based on these features, terrain classification will be achieved through F1 to F3 networks. The F1 layer also is a fully connected network layer, and the number of set neurons is 8192. The F2 network layer is a convolutional network layer. The kernel of the network layer also needs to be set according to the input image. If the input image is gray level image, 64*2*2*1 kernels are set, where 2*2 refers to the size of the kernel and performs convolution training is implemented in 1 channel. If the input image is a color image, it needs to pass convolution training through R, G, B three channels and 64*2*2*3 kernels are set. The F3 network layer is a fully-connected network layer with a softmax function. It has 4096 neurons. In the process of training in the fully connected network layer, the softmax function is used as an activation function to activate the network layer, and full connection training is performed. The number of neurons in the output layer are equal to the terrain classification summary in the dataset. That is, if the dataset contains 10 types of terrain, the neurons of the output network layer sets 10, and each neuron outputs 0 or 1, forming a 10 dimensional vector corresponding to the class of the input image patch. the position of the "1" in the vector indicate the terrain type of the input image patches.

B. Network training

In this paper, based on the convolutional neural network with double tower structure, the network is trained using the back-propagation algorithm to determine the network parameters. It should be emphasized that the branch part of the network structure can share weights. The premise of weight sharing is that the image types of the input image pair are consistent, and both are gray scale images or are color images. The back-propagation algorithm can be divided into four different parts: forward propagation, loss function, back propagation, update weights.

Firstly, Extracting samples from the data set and sliding them in step size **s** in the sonar image patch $P^S(\mathbf{p})$ and the visual image patch $P^V(\mathbf{p})$ with a rectangular window of size $k{\times}k$ to obtain a sonar image patch $I_s(\mathbf{p}_i)$ and a visual image patch$I_v(\mathbf{p}_i)$, where $\mathbf{p}_i$ indicates the center coordinate of the *i*-th image patch (x_i, y_i) and *k*=64, *s*=8 are set (As shown in Figure 3).

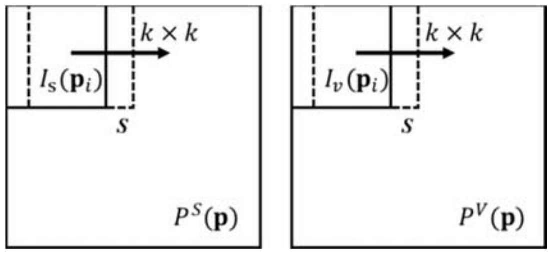


Fig.3. The diagram of the input image

Secondly, paired image patches $(I_s(\mathbf{p}_i), I_v(\mathbf{p}_i))$composed of matched visual topographic image patches and sonar terrain image patches in the sample data are input to the network for propagation. Computing the terrain class $C_O(\mathbf{p}_i)$ of the image patch actually output through network training, which is the forward propagation process:

$$C_O(\mathbf{p}_i) = f\left(\left(I_s(\mathbf{p}_i), I_v(\mathbf{p}_i)\right), \widehat{\boldsymbol{\theta}}\right) \tag{1}$$

Where *f*(·) is the function that reflects the learning of the convolutional neural network, and $\boldsymbol{\theta}$ is the weight parameters of the network. The weight parameters are randomly initialized and then updated in the calculation process of the network. The initialized weight parameters may be such random values [0.3, 0.2, 0.1, 0, 0.2 …]. The network cannot extract the accurate feature through the initialized weight value, so it is unable to give any reasonable conclusion and it is impossible to determine which category the image belongs to. Therefore, the loss function in the backward propagation is needed to update the network weight value so as to find the desired weight parameters. There are many definitions of loss function. In this paper, the most classic and commonly used Mean Square Error (MSE) function is selected:

$$\mathrm{MSE} = \frac{1}{2m}\sum_{i=1}^{m}(y_o - y_t)^2 \tag{2}$$

Here, y_t is the real classification value of the image $C(\mathbf{p}_i)$, y_o is the classification value of the image trained by the network, m presents the number of times the network forwardly propagates data, that is, once it is calculated once more, it adds one to the value of m until the network converges and no longer calculated data. And then the mean square error is used to calculate the loss value. Since the weight values of the network are all initialized randomly, the value of this loss will be large when the network is initially trained. Our ultimate goal is to let the network predictions to be the same or the closest to the actual values. In order to achieve this goal, we need to reduce the loss value as much as possible. The smaller the loss value, the closer the prediction results are, which shows that the established network structure is reasonable and the network parameters are closer to optimal. In the process of obtaining the optimal network parameters, we need to find the weight value that can reduce the network loss by constantly adjusting the network weight value.

This paper uses the Stochastic gradient descent (SGD) algorithm to achieve the optimization of network parameters. The objective function of the algorithm is shown in (3), namely, the network parameter $\boldsymbol{\theta}$ needs to be computed to makes the output vector of the network be the closest to the corresponding category vector of the input images. The SGD method completes a calculation each time when a sample is input. After repeated iterations, back propagates the error of the output vector of the network and the label vector, and adjusts the weight parameter of the network, and stops iterating until the error value is minimum. The network weights will no longer change, indicating network convergence.

$$\min_{\theta} \frac{1}{2m} \sum_{i=1}^{m} (C_O(\mathbf{p}_i) - C(\mathbf{p}_i))^2 \tag{3}$$

During the training process, it is necessary to set a descent rate or training rate. When the training rate is set larger, it means that the faster the network converges, and the faster the weight parameters is updated. But when the set value is too large, it may jump over it when it reaches the lowest point quickly, resulting in failure to obtain the optimal weight value. When the set training rate is too small, it will lead to too long training time to make the network convergence. Therefore, setting the training rate needs to be adjusted according to actual conditions.

V. Conclusions

In recent years, image classification technology based on deep learning has achieved rapid development. Our paper reviews the development of terrain classification technology in robotics and the progress of deep learning in the field of classification. Based on the superior performance of deep learning in the field of classification, a new network structure for underwater terrain classification is proposed.

The network proposed in this paper has disadvantages and difficulties while there are many advantages, and the disadvantages or difficulties can be summarized as:

- The input of deep learning network proposed in this paper is an image pair composed of a sonar image and a visual image. This image pair requires two images to describe the same scene at the same time, and the matching degree of the image is required to be strict.
- The network structure has a branch structure, so it is necessary to ensure that the input image blocks are all gray level images or all color images, that is, to ensure the structural symmetry of the branch structure. If one of the input image blocks is a color image and the other is a gray image, the branch structure is asymmetric, and the training difficulty of the network will increase, which has also become a major limitation of the network structure.
- When training deep learning networks, image pairs with labels are required to form a data set, and a large number of image pairs are required in the data set. It takes more manpower and time to establish such a data set, which is more difficult. This is also the reason why our paper builds a network structure based on theory only and does not mention specific experimental results. We have started to work on the establishment of a specific datasets.

In short, based on the good performance of deep learning in the field of image classification, it has broad development prospects in the field of terrain classification, and will further promote the intelligent development of robots in environmental awareness.

Acknowledgments

This work was supported by the Shenzhen Municipal Science and Technology Innovation Committee JCYJ20170817145216803 and the National Natural Science Foundation of China under grants nos. 61472325, 61733014, 51579210.

References

[1] J. Christie, N. Kottege. Acoustics based terrain classification for legged robots[C] IEEE International Conference on Robotics and Automation. IEEE, 2016.

[2] H. Zhang, X. Dai, F. Sun, et al. Terrain classification in field environment based on Random Forest for the mobile robot[C]// Control Conference. IEEE, 2016:6074-6079.

[3] X. A. Wu, T. M. Huh, R. Mukherjee, et al. Integrated Ground Reaction Force Sensing and Terrain Classification for Small Legged Robots[J]. IEEE Robotics & Automation Letters, 2016, 1(2):1125-1132.

[4] R. Hadsell, P. Sermanet, J. Ben, A. Erkan, M. Scoffier, K. Kavukcuoglu, U. Muller, and Y. LeCun. Learning long-range vision for autonomous off-road driving. Journal of Field Robotics, 26(2):120–144, Feb. 2009. ISSN 1556-4959

[5] U. A. Muller, L. D. Jackel, Y. LeCun, and B. Flepp. Real-time adaptive off-road vehicle navigation and terrain classification. Proc. SPIE, 8741:87410A–87410A–19, 2013

[6] S. Manjanna, G. Dudek. Autonomous gait selection for energy efficient walking. IEEE International Conference on Robotics and Automation. IEEE, 2015:5155-5162.

[7] S. N. Àngel, E. H. Teniente, M. Morta, et al. Terrain Classification in Complex Three-dimensional Outdoor Environments. Journal of Field Robotics, 2015, 32(1):42-60.

[8] B. Suger, B. Steder, W. Burgard. Traversability analysis for mobile robots in outdoor environments: A semi-supervised learning approach based on 3D-lidar data. IEEE International Conference on Robotics and Automation. IEEE, 2015:3941-3946.

[9] S. Otte, S. Laible, R. Hanten, et al. Robust visual terrain classification with recurrent neural networks. European symposium on artificial neural networks. Bruges, Belgium, 2015: pp.451-456.

[10] A. Valada, W. Burgard. Deep spatiotemporal models for robust proprioceptive terrain classification. International Journal of Robotics Research, 2017, 36(13-14):1521-1539.

[11] D. Rao, M. D. Deuge, N. Nourani–Vatani, et al. Multimodal learning and inference from visual and remotely sensed data[J]. International Journal of Robotics Research, 2017, 36(1):24-43.

[12] Y. Liu, X. Chen, Z. Wang, et al. Deep learning for pixel-level image fusion: Recent advances and future prospects[J]. Information Fusion, 2018, 42:158-173.

[13] Y. Liu, X. Chen, H. Peng, Z. Wang, Multi-focus image fusion with a deep convolutional neural network, Inf. Fusion 36 (2017) 191–207.

[14] B. Yang, J. Zhong, Y. Li, Z. Chen, Multi-focus image fusion and super-resolutionwith convolutional neural network, Int. J. Wavelets Multiresolut. Inf. Process. 15(4) (2017) 1750037:1–15.

[15] C. Du, S. Gao, Image segmentation-based multi-focus image fusion through multi-scale convolutional neural network, IEEE Access 5 (2017)15750–15761.

[16] N. Kalantari, R. Ramamoorthi, Deep high dynamic range imaging of dynamic scenes, ACM Trans. Graph. 36 (4) (2017) 144:1–12.

[17] J. Zhong, B. Yang, Y. Li, F. Zhong, Z. Chen, Image fusion and super-resolution withconvolutional neural network, Proceedings of Chinese Conference on Pattern Recognition, (2016), pp. 78–88.

[18] Y. Liu, X. Chen, R. Ward, Z. Wang, Image fusion with convolutional sparse representation, IEEE Signal Process. Lett.23(12)(2016)1882–1886.

[19] Y. Liu, X. Chen, J. Cheng, H. Peng, A medical image fusion method based on convolutional neural networks, Proceedings of 20th International Conference on Information Fusion, (2017), pp. 1–7.

[20] V. U. Sameer, R. Naskar, N. Musthyala, et al. Deep Learning Based Counter–Forensic Image Classification for Camera Model Identification[J]. 2017:52-64.

[21] L. Wang, J. Zhang, P. Liu, et al. Spectral–spatial multi-feature-based deep learning for hyperspectral remote sensing image classification[J]. Soft Computing, 2017, 21(1):213-221.

[22] S. Li, H. Teng, Y. Ling, et al. The Real-time Enhancement Technology of the Side Scan Sonar Image[J]. Journal of Applied Acoustics, 2006, 25(5):284-289.

[23] L. Wu. Research on Underwater Robot Vision Perception and Image Processing Methods[D]. Huazhong University of Science and Technology, 2005.

[24] P. Sahu, N. Gupta , N. Sharma. A Survey on Underwater Image Enhancement Techniques [J]. International Journal of Computer Applications, 2014, 87(13): 19-23.

[25] H. Li, H. Teng, H. Song, et al. Analysis of Image Filtering Effect of Side-scan Sonar Wavelet Function. [J]. Hydroaphic Surveying And Charting, 2009, 29(3):65-67.