Improved simultaneous localization and mapping algorithm combined with semantic segmentation model

Xuerong Cui1, Shengjie Xue2, Juan Li2, Shibao Li1, Jianhang Liu2 and Haihua Chen1

Abstract
In the past decades, emerging technologies such as unmanned driving and indoor navigation have developed rapidly, and simultaneous localization and mapping has played unparalleled roles as core technologies. However, dynamic objects in complex environments will affect the positioning accuracy. In order to reduce the influence of dynamic objects, this article proposes an improved simultaneous localization and mapping algorithm combined with semantic segmentation model. First, in the pre-processing stage, in order to reduce the influence of dynamic features, fully convolutional network model is used to find the dynamic object, and then the output image is masked and fused to obtain the final image without dynamic object features. Second, in the feature-processing stage, three parts are improved to reduce the computing complexity, which are extracting, matching, and eliminating mismatching feature points. Experiments show that the absolute trajectory accuracy in high dynamic scene is improved by 48.58% on average. Meanwhile, the average processing time is also reduced by 21.84%.

Keywords
Dynamic scene, semantic segmentation, simultaneous localization and mapping, visual positioning

Date received: 9 January 2021; accepted: 30 March 2021

Handling Editor: Lyudmila Mihaylova

Introduction
In the field of vision processing,1 using monocular and binocular cameras to obtain image information and achieve accurate positioning has always been a research hotspot. In recent years, visual SLAM (Simultaneous Localization and Mapping) has a wide range of applications, such as unmanned driving, urban planning, geographic surveying, and so on. However, the unmanned landing of vehicles still needs further technical innovation. It is not only the state estimation of a few vehicles2 but also the integration of the data link of the whole intelligent vehicle network.3,4 There are many excellent and advanced visual SLAM schemes, such as ORB-SLAM2,5 VINS-MONO6 LSD-SLAM (Large-Scale Direct Monocular SLAM),7 and so on. However, their
theoretical backgrounds are based on a static environment and the pose solving method and map building are based on geometric information. However, for a dynamic environment, visual SLAM still lacks a good solution, because the proposed mathematical geometry methods cannot deal with the fuzziness caused by moving objects. For complex environments, the robustness of the visual SLAM scheme needs to be improved. At the same time, feature point extraction and descriptor calculation in visual SLAM take up a lot of computing time, so for a real-time SLAM system, a more efficient and accurate algorithm is essential when extracting, matching, and eliminating mismatching feature points.

At present, the localization theory of visual SLAM is mainly divided into two categories. The first category is called feature point method, which extracts pixel-level feature points from pictures, and calculates geometric relations to obtain rotation matrix and translation information, and then obtains the final localization result (ORB-SLAM and ORB-SLAM). The second category is the direct method, which calculates the minimum photometric error to get the optimized rotation matrix and translation information Semi-Direct Monocular Visual Odometry (SVO) and LSD-SLAM. However, if the surrounding environment is dynamic, the feature matching relationship in the visual SLAM system will be affected by different magnitudes, which will eventually lead to the decline of positioning accuracy. At present, the visual SLAM system has been effectively applied in urban traffic, and some progress has been made in solving the offset imposed by dynamic environment. For example, inertial measurement unit (IMU) including accelerometer, gyroscope, and magnetometer can provide multiple degrees of freedom. IMU can solve the scale problem of initialization and provide better initial pose. Combining IMU shows higher accuracy and robustness (VINS-MONO) in dynamic complex environment. In the field of deep learning, after the image is segmented with semantic segmentation network, the dynamic objects are restored to static background, which also effectively solves the influence of dynamic features, in DynaSLAM. However, it is still a problem whether the background repaired by DynaSLAM system completely fits the original image.

In recent years, the Deep Convolutional Neural Network (DCNN) has developed rapidly, showing a dominant position in many visual recognition tasks and emerging excellent network structures such as Resnet. Similarly, image semantic segmentation is also an important part of image processing and machine vision technology, and an important branch of the artificial intelligence (AI) field. Many excellent semantic segmentation networks have developed better architectures based on fully convolutional network (FCN), such as SegNet, U-Net, Mask-RCNN, and PspNet. Semantic information obtained by fusing semantic segmentation can not only be introduced into BA (Bundle Adjustment) and function optimization but also be used to build semantic maps to provide more accurate judgment for positioning. In the aspect of seeking the combination of semantic segmentation network and SLAM, there have been many excellent schemes such as DynaSLAM, which uses multi-view geometry method and deep learning method to jointly detect dynamic objects and finally repairs the key frames affected by dynamic objects to generate static maps. However, its drawbacks are also obvious. If there is no dynamic scene information of the current frame in other frames, it will lead to repair cracks in the results, which will affect the accuracy. Similarly, DS-SLAM also proposed the combination of semantic segmentation network and mobile consistency checking method, which can reduce the influence of dynamic objects. However, the actual influence is deviated, and there are still many undetected feature points in the actual moving object area. In this article, semantic segmentation network is used to remove dynamic objects, which directly avoids the problems of repairing cracks in images and incomplete removal of dynamic feature points.

Videos and images in dynamic environment are studied to eliminate the influence of dynamic objects. TUM data set is used to verify that the influence of dynamic objects and the existence of dynamic feature points in the original image are eliminated, which further solves the problem that ORB-SLAM cannot effectively deal with dynamic feature points. This fundamentally eliminates the influence of dynamic objects, and can provide robust static environment tracking for camera visual positioning.

The main contributions of this article are as follows:

- The method of image semantic segmentation is introduced. By fusing the semantic image and the original image of data set, moving objects and the existence of dynamic feature points in the original image are eliminated, which further solves the problem that ORB-SLAM cannot effectively deal with dynamic feature points. This fundamentally eliminates the influence of dynamic objects, and can provide robust static environment tracking for camera visual positioning.
- In order to reduce the difficulty of initialization, this article proposes an improved FAST feature point extraction algorithm, which combines the idea of machine learning and reduces the influence of image noise without sacrificing the speed of extracting feature points.
- At the same time, facing the feature point matching in different environments, this article sets the threshold in the system. In the case of multiple feature points in complex environment, FLANN (Fast Nearest Neighbor Algorithm)
is used to improve the matching rate. In the case of a simple environment with few feature points, the violent matching algorithm is still used to pursue higher accuracy. Different algorithm processing in different situations improves the time complexity of the system.

In view of mismatching, an improved RANSAC\(^24\) (Random Sample Consensus) algorithm is proposed, which prioritizes and dynamically adjusts iteration times. On the premise of retaining the randomness of RANSAC algorithm, the problem of randomness of iteration times of calculation parameters is avoided, and the final result is trained more effectively and quickly, which improves the speed and ensures the accuracy.

**System introduction**

In visual SLAM scheme, ORB-SLAM2, as the most mainstream SLAM system, has shown predominant performance in real-time positioning speed and accuracy, as well as loop detection and repositioning. Meanwhile, the ORB feature points are proposed by ORB-SLAM2, which not only improves the feature extraction rate but also improves the efficiency of feature matching. In this article, ORB-SLAM2 is adopted as the overall SLAM framework, and a series of improvements is proposed to optimize the algorithm of feature point module of ORB-SLAM2 system while combining the semantic segmentation model (Figure 1).

**Dynamic object elimination model based on semantic segmentation**

Semantic segmentation is to classify every pixel in an image and determine the category, such as background, person, car, and bed. Since the FCN structure came into being, FCN has become the basic framework of semantic segmentation, and many excellent frameworks are based on the improvement of FCN. In this article, in order to achieve semantic segmentation, the FCN is used to train on the MIT ADE20K\(^25\) data set and carry out pixel-level semantic segmentation. The images obtained by the FCN model can distinguish people, tables, chairs, and other objects and label them with different colors. Because of the particularity of TUM data set, this article takes people as dynamic objects, and eliminates the pixel information of people to eliminate its influence on system accuracy (Figure 2).

The dynamic object elimination model works as follows:

- Selecting indoor pictures in MIT ADE20K data set to train FCN network and generate semantic segmentation model.
- Using FCN to segment the image data set and obtaining semantic segmentation images.
- Using OpenCV to add masks to dynamic objects in semantic segmentation images.
- Performing binary not-operation on each pixel value corresponding to the semantic mask image and the original image to obtain a processed dynamic object rejection image.
The final image is loaded into the improved ORB-SLAM2 system to run, and the final measurement accuracy is obtained.

**Feature point extraction: improved FAST algorithm**

As the SLAM system requires real-time performance, it is a key element to quickly extract effective feature points for image matching, and then obtain the required location. ORB-SLAM2 proposed a scheme of combining FAST feature points with binary robust independent elementary features (BRIEF) descriptors, which can meet the real-time requirements. The FAST feature point method proposes that each FAST corner is judged by selecting 16 pixels around it through a circle with a radius of 3, and each pixel is subtracted from the central pixel value. If there are \( N \) continuous pixel values that satisfy the difference value greater than the set threshold, the central point is judged as FAST corner (Figure 3).

Since it is necessary to traverse all the pixels in the neighborhood circle of the center point, and because there are \( n \) consecutive pixels which has low efficiency and long-term initialization. Therefore, combined with the idea of feature selection in machine learning, an improved FAST algorithm is proposed. At first, this article selects pixel point \( p \) whose brightness is \( I_p \). Then, the 16 points on the neighborhood circle with radius of
3 are selected with \( p \) as the center, and all points are divided into singular points and even points by labels. Finally, the statistics \( \Delta \) is obtained by the formula as follows

\[
\Delta = \sum_{i=1}^{n} |(I_{\text{odd}} - I_p)^2 - (I_{\text{even}} - I_p)^2|
\]  

(1)

where \( I_{\text{odd}} \) are the singular points and \( I_{\text{even}} \) are the even points. The final statistics are compared with the preset threshold \( Q \), and if it is greater than the threshold, it is judged as an FAST corner point. The improved FAST feature extraction method is simple and fast, which can guarantee the accuracy and improve the initialization successful probability.

**Feature point matching: introducing the FLANN algorithm**

ORB-SLAM2 uses violent matching algorithm for feature point matching between reference frame and current frame during monocular initialization. Each feature point of the reference frame will match each feature point of the current frame, and then the best matching point will be found with the shortest distance between different feature point descriptors. Although the positioning accuracy is the best, the computing complexity is the highest. Therefore, the FLANN algorithm is introduced to realize efficient matching. There are many index parameter structures of FLANN, such as the random \( k-d \) tree and preferential search \( k \)-means tree. The method of Multi-Probe LSH\(^2^7\) (Locality Sensitive Hashing) is selected in this article. The core idea is to use the probe sequence to detect multiple hash buckets containing adjacent points. If the neighboring points of the feature points are not detected in the bucket, the neighboring points have a high probability in the adjacent bucket. Multi-Probe LSH increases the probability of finding neighboring points. LSH algorithm in FLANN library is used to match the descriptor distance between reference frame and current frame to get a matched pair. Then, it will traverse the distance \( x \) of the matched pair to obtain the maximum distance \( X_{\text{max}} \) and the minimum distance \( X_{\text{min}} \). Finally, if the distance of the matched pair is less than two times of the minimum distance, that is, \( X < 2 \times X_{\text{min}} \), it will be eliminated and the rest pairs will be established as the correct matched pairs (Figure 4).

The flow of improved feature point matching algorithm is as follows:

1. The number of feature points of the current frame is checked, and if it is less than the set threshold value \( Q \), go to step 2, otherwise go to step 3.
2. Due to the small number of frames, it is suitable for the violent search algorithm to find the optimal solution and obtain higher accuracy in a simple environment.
3. Due to the large number of current frames, the FLANN algorithm is best suited in a complex environment to meet the requirements of computing time with suboptimal precision.

The improved feature point matching algorithm can effectively decrease the matching time and the visual SLAM system can complete positioning with high accuracy and speed in all kinds of environments.

**Mismatch elimination: improved RANSAC algorithm**

After the extraction and matching of feature points, a series of well-matched feature point pairs can be obtained. However, there will be mismatching between feature points in the obtained matched pairs, which means that, non-corresponding feature points will be detected as matching ones. If the mismatching point pairs are not eliminated, the positioning accuracy will be affected. In ORB-SLAM2 framework, the RANSAC algorithm is used to eliminate the mismatching. Its principle is to fit the model through a set of random assumed local points, then substitute other points into the model to check whether it is a local point and evaluate the advantages and disadvantages. After a certain number of iterations, the best model is evaluated, which is suitable for all local points. Because of its complete randomness, the model evaluated by
RANSAC algorithm is very robust. However, the number of iterations needs to be set artificially, which brings different influences. The high number of iterations will consume calculation time, while the model trained by the low number of iterations cannot check local points effectively. To solve the said problems, this article proposes an improved RANSAC algorithm which prioritizes and dynamically adjusts iteration times (Figure 5).

First, the maximum iteration times $MAX$ is set, and eight feature point pairs are randomly extracted at each iteration for later calculation. At the same time, the ratio sum $C$ of the optimal distance to the suboptimal distance is calculated for the eight feature point pairs extracted at each iteration

$$C = \sum_{i=1}^{8} \frac{X_{\text{best}}}{X_{\text{minor}}}$$  \hspace{1cm} (2)

where $X_{\text{best}}$ is the optimal distance and $X_{\text{minor}}$ is the suboptimal distance. Second, the optimal pair of feature points is placed in the last iteration by sorting.

![Flowchart of Improved RANSAC Algorithm](image-url)
Finally, the iteration times are dynamically adjusted. While ensuring that the iteration times are reduced orderly, the feature point pair calculated in the next round will be better and the obtained model will be more accurate. In this way, the influence of excessive iteration times on the system efficiency will be effectively avoided.

The algorithm flows of dynamically adjusting iteration times are as follows:

1. Introducing statistics $\text{ep} = (\text{goodcounts}/\text{counts})^2$, learning rate $p = 0.01$, point = 8, and initial iteration times $M = \text{MAX}$. Among them, goodcounts are the number of feature point pairs detected as local points in the current model, and counts are the number of feature point pairs matched as a whole.

2. Set the threshold of statistics as 0.5. And when the statistic ep is less than 0.5

$$rnum = \log\left(\sqrt{\frac{1}{\text{ep}^\text{point}}}\right)$$

when the statistic ep is more than 0.5

$$rnum = \log\left(\sqrt{\frac{1}{\text{ep}^\text{point}}^2}\right)$$

3. If $rnum > 0$ or $num \leq rnum \times \text{MAX}$, the iteration times will not be changed and the iterations are still performed in sequence. On the contrary, the maximum number of iterations is updated as

$$\text{max\_iters} = \left\lceil \frac{num}{rnum} \right\rceil$$

where $num = \log(1-p)$. If the score of the current iteration exceeds the threshold or if it has reached the final iteration, then the model with the highest score is selected to eliminate the mismatch. With the proposed RANSAC algorithm, the model can be found more efficiently and accurately, the mismatched pairs can be eliminated, and the original randomness of the algorithm can be maintained.

**Experimental results**

In order to verify the feasibility and effectiveness of the proposed algorithm, this article uses the dynamic object sequence of the TUM data set. Compared with other static scenes, the dynamic scene sequence of the TUM data set can be divided into low dynamic set and high dynamic set. These data sets can effectively check the robustness of the SLAM system with dynamic objects and can also be used to distinguish maps and find changes in scenes. In this section, all experiments were conducted on a computer with Intel i7 CPU, GTX960M GPU, and 8 GB memory. In addition, the test results of the original ORB-SLAM2 system in the TUM data set set are compared in accuracy and speed, so as to quantify the advantages of the improved system. At the same time, in order to avoid the contingency of the experimental data, the average value of the results of five runs is taken for each series.

**Evaluation on TUM data set**

In this section, result comparison will be showed between the proposed algorithm and the ORB-SLAM2 system. The used high dynamic TUM data set includes the rgbd_dataset-freiburg3_walking_halfsphere, rgbd_dataset_freiburg3_walking_XYZ, and rgbd_dataset_freiburg3_walking_rpy. The used low dynamic TUM data set includes the rgbd_dataset_freiburg3_sitting_halfsphere and the rgbd_dataset_freiburg3_sitting_XYZ. Both two data sets moving directions are the same.

The final quantitative results are shown in Table 1. This article gives the RMSE (root mean square error), mean error, and median error of the measurement results and shows the robustness and accuracy compared with ORB-SLAM2 test results.

Table 1 shows that the accuracy of the proposed algorithm has been improved to a certain extent regardless of whether it is a high dynamic data set or a low dynamic data set. In terms of absolute trajectory error,
the RMSE is improved by 48.58% on average. This is because that the large number of dynamic objects is removed, and then the accuracy is improved, especially in complex dynamic environment. However, the average improvement of RMSE on low dynamic data set is only 25.22%, which also proves that ORB-SLAM2 algorithm is reasonable for static environments. Therefore, the proposed algorithm can improve the result very little on static environments (Figures 6–8).

**Figure 6.** APE comparison between proposed algorithm and ORB-SLAM2 in fr3/walking halfsphere data set (a) proposed algorithm and (b) ORB-SLAM2.

**Figure 7.** RPE comparison between proposed algorithm and ORB-SLAM2 in fr3/walking halfsphere data set (a) proposed algorithm and (b) ORB-SLAM2.

**Time evaluation**

This article not only combines semantic segmentation networks but also improves the feature module of ORB-SLAM2 system, so it not only improves the accuracy but also further reduces the processing time. Table 2 shows that the improved ORB-SLAM2 algorithm displays excellent time superiority in both high and low dynamic sequences of the TUM data set with
Figure 8. Comparison of trajectory errors in yaw, pitch and roll between proposed algorithm and ORB-SLAM2 in fr3/walking halfsphere dataset.

Table 2. Average computing time.\(^1\)

| Data set             | ORB-SLAM2 algorithm | Proposed algorithm | Improvements |
|----------------------|---------------------|--------------------|--------------|
|                      | Mean\(^{10}\)       | Median\(^{10}\)    | Mean\(^{10}\) | Median\(^{10}\) | Mean\(^{10}\) | Median\(^{10}\) |
| walking_xyz          | 0.0424              | 0.0469             | 0.0299       | 0.0289       | 29.48%          | 38.38%          |
| walking_halfsphere   | 0.0332              | 0.0321             | 0.0301       | 0.0288       | 9.34%           | 10.28%          |
| walking_rpy          | 0.0381              | 0.0401             | 0.0281       | 0.0265       | 26.25%          | 33.91%          |
| sitting_halfsphere   | 0.0409              | 0.0403             | 0.0290       | 0.0281       | 29.10%          | 30.27%          |
| sitting_xyz          | 0.0376              | 0.0363             | 0.0320       | 0.0304       | 14.90%          | 16.25%          |
an improvement of 21.84% in average time and 25.82% in median time. It also shows that the proposed feature module improves not only the original high precision and high speed but also the robustness of the whole system.

Conclusion
This article proposes an improved SLAM algorithm combined with semantic segmentation model in dynamic environment. FCN is used to segment semantic images, and then dynamic objects are identified and eliminated, which improves the positioning accuracy of SLAM system in dynamic environment. At the same time, three parts of ORB-SLAM2 system are improved, which are extracting, matching, and eliminating mismatching feature points. The proposed algorithm can not only improve the initialization speed of the system but also improve the robustness for different situations. Finally, the improved ORB-SLAM2 algorithm has been well verified in the TUM data set. In the future, the key research direction is how to identify all kinds of dynamic objects in video, and eliminate the influence of their feature points according to the scene.

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 was supported by the National Natural Science Foundation of China (nos 61938204, and 61972417), the Key Research and Science Foundation of China (nos 61671482, 61902431, 18CX02136A). The Central Universities (nos 19CX05003A-9, 19CX05003A-10, 2019GGX101048, and the Fundamental Research Funds for the research, authorship, and/or publication of this article: This work was supported by the National Natural Science Foundation of China (nos 61938204, and 61972417), the Key Research and Science Foundation of China (nos 61671482, 61902431, 18CX02136A). The Central Universities (nos 19CX05003A-9, 19CX05003A-10, 2019GGX101048, and the Fundamental Research Funds for Shandong Province with no. 18CX02136A).

ORCID iD
Xuerong Cui https://orcid.org/0000-0002-2326-9518

References
1. Zheng X and Cai Z. Privacy-preserved data sharing towards multiple parties in industrial IoTs. IEEE J Select Area Commun 2020; 38(5): 968–979.
2. Xiao Z, Xiao D, Havyarimana V, et al. Toward accurate vehicle state estimation under non-Gaussian noises. IEEE Internet Things J 2019; 6(6): 10652–10664.
3. Cai Z and Zheng X. A private and efficient mechanism for data uploading in smart cyber-physical systems. IEEE Trans Netw Sci Eng 2018; 7(2): 766–775.
4. Zheng X, Cai Z and Li Y. Data linkage in smart internet of things systems: a consideration from a privacy perspective. IEEE Commun Mag 2018; 56(9): 55–61.
5. Mur-Artal R and Tardós JD. Orbslam2: an open-source SLAM system for monocular, stereo, and RGB-D cameras. IEEE T Robot 2017; 33(5): 1255–1262.
6. Qin T, Li P and Shen S. VINS-Mono: a robust and versatile monocular visual-inertial state estimator. IEEE T Robot 2018; 34(4): 1004–1020.
7. Engel J, Schöps T and Cremers D. LSD-SLAM: large-scale direct monocular SLAM. Cham: Springer, 2014.
8. Georges HM, Wang D and Xiao Z. GNSS/low-cost MEMS-INS integration using variational adaptive cubature Kalman smoother and ensemble regularized ELM. Math Probl Eng 2015; 2015: 682907.
9. Mur-Artal R, Montiel JMMM and Tardos JD. ORB-SLAM: a versatile and accurate monocular SLAM system. IEEE T Robot 2015; 31(5): 1147–1163.
10. Forster C, Pizzoli M and Scaramuzza D. SVO: fast semidirect monocular visual odometry. In: 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–June 2014. New York: IEEE.
11. Cai Z, Zheng X and Yu J. A differential-private framework for urban traffic flows estimation via taxi companies. IEEE T Ind Inform 2019; 15(12): 6492–6499.
12. Xiao J, Xiao Z, Wang D, et al. Short-term traffic volume prediction by ensemble learning in concept drifting environments. Knowl Based Syst 2019; 164: 223–225.
13. Bescos B, Facil JM, Civera J, et al. DynaSLAM: tracking, mapping, and inpainting in dynamic scenes. IEEE Robot Autom Lett 2018; 3(4): 4076–4083.
14. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, 27–30 June 2016. New York: IEEE.
15. Long J, Shelhamer E and Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Boston, MA, 7–12 June 2015. New York: IEEE.
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. 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, 5–9 October 2015. New York: Springer.
18. He K, Gkioxari G, Dollar P, et al. Mask R-CNN. In: Proceedings of the IEEE international conference on computer vision, Venice, 22–29 October 2017. New York: IEEE.
19. Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, HI, 21–26 July 2017. New York: IEEE.
20. Yu C, Liu Z, Liu X, et al. DS-SLAM: a semantic visual SLAM towards dynamic environments. In: 2018 IEEE/RSJ international conference on intelligent robots and
systems, Madrid, 1–5 October 2018, pp.1168–1174. New York: IEEE.

21. Sturm J, Engelhard N, Endres F, et al. A benchmark for the evaluation of RGB-D SLAM systems. In: 2012 IEEE/RSJ international conference on intelligent robots and systems, Vilamoura-Algarve, 7–12 October 2012. New York: IEEE.

22. Rosten E and Drummond T. Machine learning for high-speed corner detection. In: European conference on computer vision, Graz, 7–13 May 2006. New York: Springer.

23. Muja M and Lowe D. FLANN—fast library for approximate nearest neighbors user manual. Vancouver, BC, Canada: Computer Science Department, University of British Columbia, 2009.

24. Fischler MA and Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Comm ACM 1981; 24(6): 381–395.

25. Zhou B, Zhao H, Puig X, et al. Semantic understanding of scenes through the ade20k dataset. Int J Comput Vision 2019; 127(3): 302–321.

26. Calonder M, Lepetit V, Strecha C, et al. Brief: binary robust independent elementary features. In: European conference on computer vision, Crete, 5–11 September 2010. New York: Springer.

27. Lv Q, Josephson W, Wang Z, et al. Multi-probe LSH: efficient indexing for high-dimensional similarity search. In: Proceedings of the 33rd international conference on Very large data bases, Vienna, 23–27 September 2007.