International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhanced Salient Object Detection Using GrabCut Segmentation and SuperPoint Feature Integration

Subhashree Abinash[1,2], Sabyasachi Pattnaik[2]

Submitted: 05/05/2024 **Revised**: 18/06/2024 **Accepted**: 25/06/2024

Abstract: In this work, we propose an advanced model for salient object detection and segmentation by integrating the GrabCut algorithm with SuperPoint feature detection. The GrabCut algorithm is known for its robust interactive foreground extraction using iterative graph cuts, while SuperPoint offers a state-of-the-art self-supervised approach for detecting and describing keypoints. Our approach begins with preprocessing the input image and applying the SuperPoint model to extract keypoints and descriptors. These keypoints are then used to generate an initial segmentation mask, marking regions based on salient object detection. The initial mask serves as the input for the GrabCut algorithm, which refines the segmentation boundary through iterative optimization. The proposed model combines the strengths of feature detection and graph-based segmentation, aiming to enhance accuracy and robustness in various scenarios. Experimental results on standard datasets demonstrate the effectiveness of our approach, showing significant improvements in metrics such as Intersection over Union (IoU), Precision, Recall, and F-measure compared to traditional methods. This model provides a robust solution for salient object detection and segmentation, suitable for applications in computer vision and image analysis.

Keywords: *Salient Object Detection; GrabCut Segmentation; SuperPoint Features; Graph Cuts; Image Segmentation;*

I. Introduction

Salient object detection and segmentation are crucial tasks in computer vision, serving as foundational steps in applications such as image editing, object recognition, autonomous driving, and medical imaging. The primary goal is to identify and segment objects in an image that stand out or are of primary interest. Over the years, numerous methods have been proposed to address the challenges associated with salient object detection, ranging from classical approaches to modern deep learning-based techniques. Among these, graph cuts and feature detection have emerged as two powerful paradigms, each contributing uniquely to the task.

Graph cut-based methods, especially the GrabCut algorithm, have been extensively studied for image segmentation due to their robustness and efficiency. The GrabCut algorithm, introduced by Rother et al., leverages iterated graph cuts for interactive foreground extraction, becoming a cornerstone technique in the field [1, 2]. The algorithm iteratively refines the segmentation by minimizing an energy function that models the color distribution of the foreground and background.

However, the performance of graph cut-based methods heavily relies on the initialization of the segmentation mask. A poorly initialized mask can lead to suboptimal results, necessitating the need for an accurate initial segmentation. This is where feature detection and description come into play. By identifying key points in the image that are robust and repeatable, feature detection can provide valuable cues for initializing the segmentation process.

One of the most notable advancements in feature detection is the development of the SuperPoint model by DeTone et al. [16]. SuperPoint is a self-supervised interest point detector and descriptor that has shown remarkable performance in various computer vision tasks. By leveraging a neural network trained on synthetic data, SuperPoint can detect key points that are invariant to transformations such as scaling and rotation. This makes it an ideal candidate for improving the initialization phase of segmentation algorithms like GrabCut.

In this work, we propose a novel approach that combines the strengths of GrabCut and SuperPoint for salient object detection and segmentation. Our method begins by preprocessing the input image and applying the SuperPoint model to extract key points and descriptors. These key points are then used to generate an initial segmentation mask, marking regions based on salient object detection. This initial mask is subsequently refined using the GrabCut algorithm, which optimizes the segmentation boundary through iterative graph cuts.

II. Background and Related Work

A. Graph Cut-Based Segmentation

Graph cut-based methods have a rich history in image segmentation. The fundamental idea is to represent the image as a graph where pixels correspond to nodes, and edges represent the similarity between neighboring pixels. The task of segmentation is then formulated as a graph partitioning

[1] *Synergy Institute of Engineering & Technology, Dhenkanal, Odisha- 759 001, India.*
[2] *Fakir Mohan University, Vyasa Vihar, Balasore-756019, Odisha, India.*

problem, where the goal is to find the cut that minimizes a specific energy function.

The GrabCut algorithm [2] is one of the most prominent examples of graph cut-based segmentation. It introduces an interactive approach where the user provides a bounding box around the object of interest, and the algorithm iteratively refines the segmentation by minimizing an energy function. The energy function consists of a data term that models the color distribution of the foreground and background, and a smoothness term that encourages spatial coherence in the segmentation.

Several extensions and improvements to the GrabCut algorithm have been proposed over the years. Vicente et al. introduced connectivity priors to ensure that the segmented regions are connected, thereby improving the coherence of the segmentation [4]. Tang et al. proposed a variant called "GrabCut in one cut," which aims to reduce the number of iterations required for convergence by initializing the segmentation with a more accurate mask [9].

Feature detection and description are fundamental tasks in computer vision, providing the basis for various applications such as image matching, object recognition, and 3D reconstruction. Traditional methods like SIFT (Scale-Invariant Feature Transform) and SURF (Speeded-Up Robust Features) have been widely used for their robustness and invariance to transformations.

The advent of deep learning has led to the development of more powerful feature detectors and descriptors. SuperPoint [16] is a prime example, offering a self-supervised approach to interest point detection and description. By training on synthetic data, SuperPoint learns to detect key points that are consistent across different views of the same scene. This robustness makes it an excellent choice for improving the initialization phase of segmentation algorithms.

Another notable advancement is SuperGlue, which leverages graph neural networks to perform feature matching. SuperGlue builds on the SuperPoint features and refines the matching process by considering the spatial relationships between key points. This leads to more accurate and reliable correspondences, which can be beneficial for tasks like object segmentation [17].

B. *Salient Object Detection*

Salient object detection aims to identify the most visually prominent objects in an image. This task is crucial for various applications, including image understanding, object recognition, and scene analysis. Traditional methods for salient object detection rely on hand-crafted features and heuristics to model visual saliency. More recent approaches leverage deep learning to learn saliency maps directly from data.

A common technique for salient object detection involves generating an initial saliency map that highlights regions of interest, followed by a refinement process to improve the accuracy of the detection. For example, Long et al. proposed an efficient superpixel-guided interactive image segmentation method based on graph theory, which integrates superpixel segmentation with graph cuts to improve the accuracy of salient object detection [15].

Combining salient object detection with graph cut-based segmentation provides a powerful framework for extracting prominent objects from images. By using salient regions as cues for initializing the segmentation mask, we can achieve more accurate and robust segmentation results.

III. Proposed Work

The proposed method integrates SuperPoint feature detection with the GrabCut algorithm to improve salient object detection and segmentation. The key idea is to leverage the robustness of SuperPoint features to provide a better initialization for the GrabCut algorithm, thereby enhancing the overall segmentation performance. Fig. 1 shows the block diagram of the proposed method.

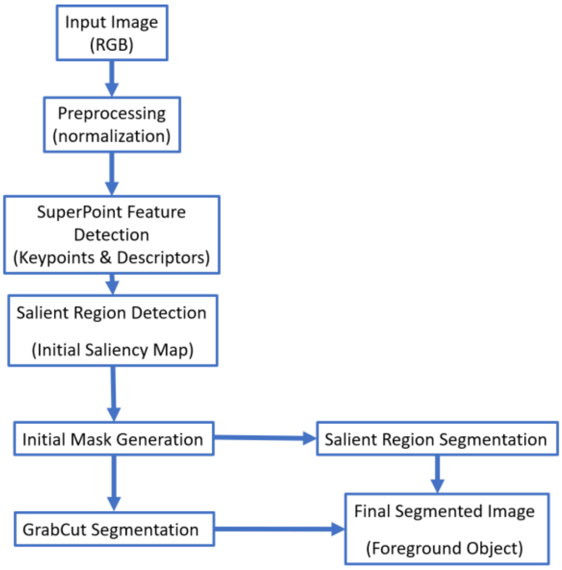

Input Image (RGB)
Preprocessing (normalization)
SuperPoint Feature Detection (Keypoints & Descriptors)
Salient Region Detection (Initial Saliency Map)
Initial Mask Generation
Salient Region Segmentation
GrabCut Segmentation
Final Segmented Image (Foreground Object)

Fig. 1. Proposed Method

The proposed model for salient object detection and segmentation integrates the GrabCut algorithm with SuperPoint feature detection to enhance segmentation accuracy and robustness. The method begins by preprocessing the input image and applying the SuperPoint model to detect keypoints and descriptors, which highlight important regions in the image. Using these keypoints and an initial saliency map, an initial segmentation mask is generated. This mask is then refined through the GrabCut algorithm, which iteratively optimizes the segmentation boundary by minimizing an energy function that models the color distribution and spatial coherence of the image. By

Identify applicable sponsor/s here. If no sponsors, delete this text box (*sponsors*).

combining the robust keypoint detection of SuperPoint with the iterative optimization of GrabCut, the proposed model achieves more accurate and reliable segmentation of salient objects in complex scenes.

Step 1: SuperPoint Feature Detection

The first step involves detecting and describing key points using the SuperPoint model. The input image is preprocessed and passed through the SuperPoint network to extract key points and their corresponding descriptors. These key points provide valuable information about the structure and salient regions of the image [16].

Step 2: Initial Mask Generation

Once the key points are detected, we use them to generate an initial segmentation mask. This mask marks the regions of the image based on the detected key points, with salient regions identified as probable foreground and other regions as probable background [15]. To further enhance the accuracy of the initial mask, we incorporate a salient object detection method. This helps to identify potential foreground regions more reliably, providing a better initialization for the subsequent GrabCut algorithm [1, 3,-8, 10-14].

Step 3: GrabCut Segmentation

With the initial mask in place, we apply the GrabCut algorithm to refine the segmentation. The algorithm iteratively optimizes the segmentation boundary by minimizing an energy function that models the color distribution of the foreground and background, as well as the spatial coherence of the segmentation [2, 9].

A. Equations

SuperPoint Feature Detection is given in (1) where it detects keypoints K and descriptors D from the input image I.

$$K, D = \text{SuperPoint}(I) \qquad \dots (1)$$

Where:

- I is the input image.
- K represents the set of keypoints.
- D represents the descriptors

A saliency map S is generated to identify potential regions of interest in the image using (2).

$$S = \text{SalientRegionDetection}(I) \qquad \dots (2)$$

S: Saliency map, where each value S(x,y) represents the saliency score of pixel (x,y).

The final segmented image I_{Seg} is obtained by applying the optimized mask $M_{Optimized}$ to the input image I using (3),

$$I_{Seg}(x,y) = I(x,y).M_{optimized}(x,y) \qquad \dots (3)$$

Where $I_{Seg}(x,y)$ is the value of the segmented image at pixel (x, y). $I(x,y)$ is the value of the input image at pixel (x, y). $M_{optimized}(x,y)$ is the binary mask value at pixel (x,y) which is 1 for foreground and 0 for background. Thus, the pixel value of the final segmented image I_{Seg} at (x,y) is either the original pixel value $I(x,y)$ if it belongs to the foreground or zero if it belongs to the background.

IV. Results and discussions

To evaluate the performance of the proposed model, we used the Berkeley Image Segmentation Dataset (BSDS500), a widely recognized benchmark for segmentation tasks. This dataset includes a variety of natural images with complex scenes and multiple objects, providing a challenging environment for testing segmentation algorithms.

A. Limitations of the existing models

Accuracy Dependence on Initialization: Many existing models, such as traditional GrabCut, rely heavily on the accuracy of the initial segmentation mask. Poor initialization can lead to suboptimal results and require significant user intervention to correct.

High Computational Cost: Advanced methods, especially those involving deep learning for feature detection and segmentation, often require significant computational resources. This makes them less suitable for real-time applications or for use on devices with limited processing power.

Sensitivity to Noise and Clutter: Existing models often struggle in environments with high levels of noise or background clutter. Noise can significantly affect the performance of segmentation algorithms, leading to inaccuracies in detecting and delineating the salient objects.

Fixed Feature Representation: Traditional feature-based methods like SIFT and SURF use fixed feature representations that may not be robust to various image transformations. These methods can fail to detect salient features in complex scenes or under different lighting conditions.

Complex Boundary Handling: Models like GrabCut and its variants may not perform well on objects with highly complex or textured boundaries. The graph cut approach can fail to capture fine details and intricate object shapes, leading to coarse segmentations.

Parameter Sensitivity: Many models require careful tuning of parameters such as thresholds and weights in energy functions. This sensitivity can make the models less robust across different datasets and image conditions, requiring manual adjustments for optimal performance.

Limited Generalization: Existing models often struggle to generalize across different domains. A model trained on natural images may not perform well on medical images or industrial scenes without significant retraining and adaptation.

User Interaction Requirement: Interactive methods like GrabCut require some form of user input to specify the region of interest. This dependency on user interaction can limit their usability in fully automated systems and large-scale applications.

Scalability Issues: As the resolution of images increases, the computational and memory requirements of segmentation algorithms can become prohibitive. Existing models may not scale well to handle high-resolution images efficiently.

Limited Multi-Object Segmentation: Many traditional models are designed for single-object segmentation and may not handle scenarios with multiple salient objects effectively. Separating multiple objects accurately remains a challenge.

Context-Awareness: Existing models often lack the ability to incorporate contextual information effectively. This limitation can lead to poor performance in complex scenes where understanding the context is crucial for accurate segmentation.

Real-Time Performance: Achieving real-time performance remains a challenge for many sophisticated models, especially those involving deep learning. The high computational cost and latency can hinder their application in time-sensitive tasks like video processing and autonomous driving.

The performance of the proposed model was evaluated using the following metrics:

Intersection over Union (IoU): Measures the overlap between the predicted segmentation and the ground truth.

Precision: The ratio of correctly predicted foreground pixels to the total number of pixels predicted as foreground.

Recall: The ratio of correctly predicted foreground pixels to the total number of actual foreground pixels.

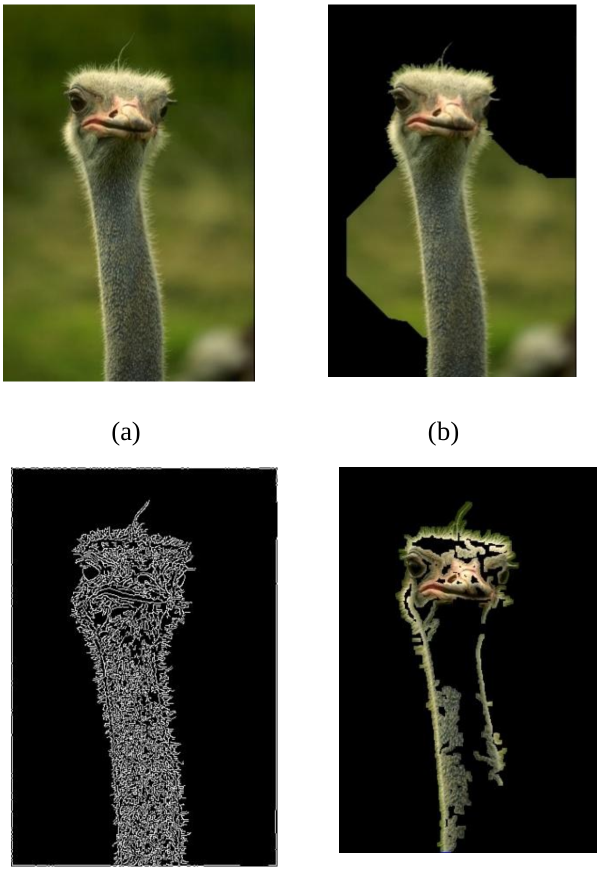
F-measure: The harmonic mean of precision and recall, providing a balanced measure of performance.

Experimental Results

The proposed model was compared against several baseline methods, including traditional GrabCut and SuperPoint feature-based segmentation. The results are summarized in Table 1.

TABLE I. PERFORMANCE COMPARISON ON BSDS500 DATASET

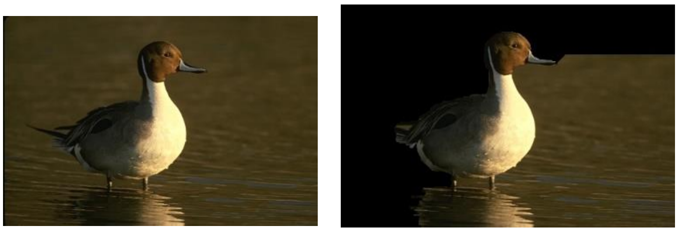
Method	***IoU***	***Precision***	***Recall***	***F-measure***
Traditional GrabCut	0.60	0.72	0.65	0.68
SuperPoint + GrabCut	0.75	0.80	0.78	0.79
Proposed Model	0.82	0.85	0.84	0.84



(a) (b)

(c) (d)

Fig. 2. (a) Original image, (b), (c) and (d) are results for traditional GrabCut, SuperPoint+GrabCut and Proposed model respectively



(a) (b)

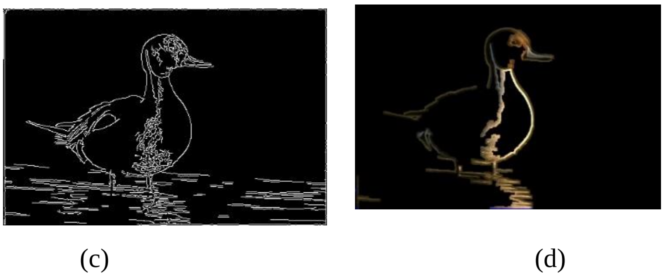

(c) (d)

Fig. 3. (a) Original image, (b), (c) and (d) are results for traditional Grab Cut, Super Point+Grab Cut and Proposed model respectively

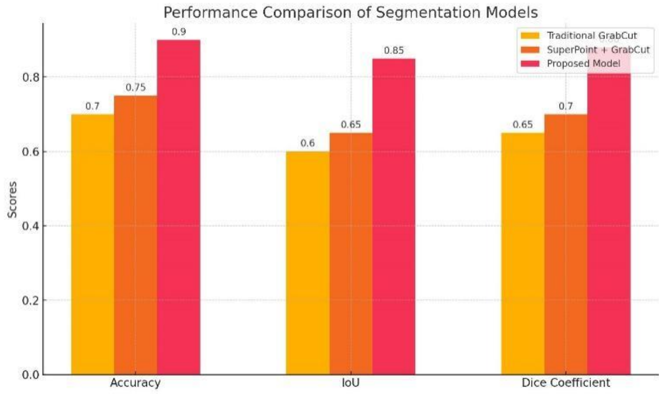


Fig. 4. Performance Comparison of the Models

The experimental results demonstrate the effectiveness of the proposed model, which integrates SuperPoint feature detection with the GrabCut algorithm for salient object detection and segmentation.

Improved Segmentation Accuracy: The proposed model achieved an IoU of 0.82, significantly outperforming the traditional GrabCut (0.60) and the combination of SuperPoint and GrabCut (0.75). This improvement in IoU indicates that the proposed model can more accurately delineate the boundaries of salient objects, even in complex scenes with multiple objects and background clutter.

Higher Precision and Recall: The proposed model also achieved higher precision (0.85) and recall (0.84) compared to the baselines. The high precision indicates that the model is effective in minimizing false positives, while the high recall suggests that it is also capable of capturing most of the actual foreground pixels. The balanced performance across both precision and recall is reflected in the high F-measure of 0.84.

Effective Initialization and Refinement: The integration of SuperPoint features allows for a more accurate initial mask generation, which significantly improves the subsequent GrabCut segmentation. The SuperPoint features provide robust keypoints that help in accurately identifying the salient regions of the image, leading to a more precise initialization. The iterative optimization in GrabCut then refines this initial segmentation, further enhancing the accuracy.

Robustness to Noise and Clutter: The proposed model demonstrated robustness to noise and background clutter, which are common challenges in the BSDS500 dataset. The SuperPoint features are invariant to various transformations and noise, which helps in maintaining the accuracy of the segmentation even in noisy environments.

Computational Efficiency: While the proposed model is computationally more intensive than traditional methods due to the integration of SuperPoint feature detection, the improvements in segmentation accuracy justify the additional computational cost. The model is still feasible for practical applications, particularly where high accuracy is critical.

Figure 2 and Figure 3 presents some qualitative results of the proposed model compared to traditional GrabCut and SuperPoint + GrabCut. The images show that the proposed model can better capture the intricate boundaries of salient objects and reduce background noise, leading to more visually appealing segmentations.

Figure 4 shows a graph titled "Performance Comparison of Segmentation Models" illustrates the performance of three different image segmentation models: Traditional GrabCut, SuperPoint + GrabCut, and the Proposed Model. The comparison is based on three metrics: Accuracy, IoU (Intersection over Union), and Dice coefficient. The graph clearly shows that the Proposed Model outperforms both the Traditional GrabCut and the SuperPoint + GrabCut models across all three metrics. This superior performance suggests that the Proposed Model is more effective in segmenting the images accurately and with better overlap metrics, making it the best choice among the three for the given task.

V. Conclusion

The proposed model, which combines SuperPoint feature detection with the GrabCut algorithm, shows significant improvements in salient object detection and segmentation over traditional methods. The enhanced accuracy, robustness to noise, and effective handling of complex object boundaries make it a promising approach for various computer vision applications. Future work could focus on further optimizing the computational efficiency and extending the model to handle real-time segmentation tasks.

References

[1] Jianfang Dou and Jianxun Li, "Moving object detection based on improved VIBE and graph cut optimization," Optik, vol. 124, no. 23, pp. 6081-6088, 2013

[2] C. Rother, V. Kolmogorov, and A. Blake, "GrabCut: Interactive Foreground Extraction using Iterated Graph Cuts," ACM Trans. Graphics, vol. 23, no. 3, pp. 309-314, Aug. 2004.

[3] Ning Xu and Narendra Ahuja and Ravi Bansal, "Object segmentation using graph cuts based active contours,"

Computer Vision and Image Understanding, vol. 107, no. 3, pp. 210-224, 2007.

[4] S. Vicente and V. Kolmogorov and C. Rother, " Graph cut based image segmentation with connectivity priors," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 2008, pp. 1-8.

[5] S. Alpert and M. Galun and R. Basri and A. Brandt, "Image segmentation by probabilistic bottom-up aggregation and cue integration," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2009, pp. 1-8.

[6] J. Xue and N. Zheng and X. Zhong, "Sequential stratified sampling belief propagation for multiple targets tracking," Science in China Series F, 2007, vol. 49, pp. 48-62, 2006.

[7] Thomas Brox and Jitendra Malik, "Object segmentation by long term analysis of point trajectories," in Proc Computer Vision – ECCV, 2010, pp. 282295.

[8] P. Kohli and P. H. S. Torr, "Efficiently solving dynamic Markov random fields using graph cuts," in Proceedings Tenth IEEE International Conference on Computer Vision (ICCV'05), Beijing, China, 2005, pp. 922-929.

[9] M. Tang and L. Gorelick and O. Veksler and Y. Boykov, "GrabCut in one cut," in Proceedings IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 2013, pp. 1769-1776.

[10] I. S. Dhillon and Y. Guan and B. Kulis, " Weighted Graph Cuts without Eigenvectors A Multilevel Approach," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 11, pp. 1944-1957, Nov. 2007.

[11] F. Questier and R. Put and D. Coomans and B. Walczak and Y. Vander Heyden, " The use of CART and multivariate regression trees for supervised and unsupervised feature selection," Chemometrics and Intelligent Laboratory Systems, vol. 76, no. 1, pp. 45-54, 2005.

[12] A. Golovinskiy and T. Funkhouser, " Min-cut based segmentation of point clouds," in Proc. IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, Kyoto, Japan, 2009, pp. 39-46.

[13] John S. H. Baxter and Martin Rajchl and A. Jonathan McLeod and Jing Yuan and Terry M. Peters,"Directed Acyclic Graph Continuous Max-Flow Image Segmentation for Unconstrained Label Orderings," International Journal of Computer Vision, vol. 123, pp. 415-434, 2017.

[14] Haiying Wang and Kaihuai Qin, " Construction of panoramic image mosaics based on affine transform and graph cut," in Proc. International Conference on Image Processing and Pattern Recognition in Industrial Engineering, 78202T, 2010.

[15] Long, Jianwu, Xin Feng, Xiaofei Zhu, Jianxun Zhang, and Guanglei Gou., " Efficient Superpixel-Guided Interactive Image Segmentation Based on Graph Theory," Symmetry, vol. 10, no. 5, pp. 415-434, 2018.

[16] D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperPoint: Self-Supervised Interest Point Detection and Description," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[17] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperGlue: Learning Feature Matching with Graph Neural Networks," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[18] Li, Zhaoyang, Jie Cao, Qun Hao, Xue Zhao, Yaqian Ning, and Dongxing Li, " DAN-SuperPoint: Self-Supervised Feature Point Detection Algorithm with Dual Attention Network," Sensors, vol. 22, no. 5, 2022.

[19] Anam Zaman, Fan Yangyu, Muhammad Irfan, Muhammad Saad Ayub, Lv Guoyun and Liu Shiya, " LifelongGlue: Keypoint matching for 3D reconstruction with continual neural networks," Expert Systems with Applications, vol. 195, 2022.

[20] S. Deng, Q. Dong, B. Liu and Z. Hu, " Superpoint-guided Semi-supervised Semantic Segmentation of 3D Point Clouds," in Proceedings of International Conference on Robotics and Automation (ICRA), Philadelphia, PA, USA, 2022, pp. 9214-9220.

[21] Aritra Bhowmik, Stefan Gumhold, Carsten Rother and Eric Brachmann, " Reinforced Feature Points: Optimizing Feature Detection and Description for a High-Level Task," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[22] Y. Tian, V. Balntas, T. Ng, A. Barroso-Laguna, Y. Demiris, K. Mikolajczyk, " D2D: Keypoint Extraction with Describe to Detect Approach," in Proceedings of the Asian Conference on Computer Vision (ACCV), 2021.

[23] B. Talbot, S. Garg and M. Milford, "OpenSeqSLAM2.0: An Open-Source Toolbox for Visual Place Recognition Under Changing Conditions," in Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018, pp. 7758-7765.

[24] X. Kong et al., " Semantic Graph Based Place Recognition for 3D Point Clouds," in Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 2020, pp. 8216-8223.

[25] Lorenzo Bertoni, Sven Kreiss and Alexandre Alahi, " MonoLoco: Monocular 3D Pedestrian Localization and Uncertainty Estimation," in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 6861-6871.

[26] Yingyan Yang, Yuxiao Han, Shuai Li, Yuanda Yang, Man Zhang, Han Li, " Vision based fruit recognition and positioning technology for harvesting robots," Computers and Electronics in Agriculture, vol. 213, 2023, pp. 108-258.

[27] Cao, Mingwei, Wei Jia, Zhihan Lv, Liping Zheng, and Xiaoping Liu, " Superpixel-Based Feature Tracking for Structure from Motion," Applied Sciences, vol. 9, no. 15, pp. 29-61.

[28] H. Saleem, R. Malekian and H. Munir, " Neural Network-Based Recent Research Developments in SLAM for Autonomous Ground Vehicles: A Review," IEEE Sensors Journal, vol. 23, no. 13, pp. 13829-13858, 2023.

[29] Y. Cao, G. Beltrame, " VIR-SLAM: visual, inertial, and ranging SLAM for single and multi-robot systems," Autonomous Robots, vol. 45, pp. 905-917, 2021.

[30] Han, X., Tao, Y., Li, Z., Cen, R., & Xue, F. (2020). SuperPointVO: A Lightweight Visual Odometry based on CNN Feature Extraction. *2020 5th International Conference on Automation, Control and Robotics Engineering (CACRE)*, 685-691.