Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/33/20230242

A review on current progress of semantic segmentation

Jinqi Liu

Tiangong University. No. 399 BinShuiXi Road, XiQing District, Tianjin 300387, China

15613469426@163.com

Abstract. Semantic segmentation, as an important task in the field of computer vision, has wide applications in image analysis and scene analysis. These application domains include autonomous driving, medical image analysis, image identification, and intelligent video surveillance. However, it faces many challenges due to the complex image structures and some confusing relationships between objects. This paper aims to provide an overview of key concepts in the field of semantic segmentation, including datasets and annotations, data augmentation, some relevant algorithms and models, and loss functions. By introducing and analyzing these concepts, we can gain a comprehensive understanding of the research progress and future directions in semantic segmentation. This paper also provides research advancements in the field of semantic segmentation. Through the introduction and analysis of these different concepts, we gain a deeper understanding of the current state and challenges of semantic segmentation. With the continuous development of deep learning techniques, we can expect semantic segmentation to have broader applications in the fields of computer vision and artificial intelligence.

Keywords: semantic segmentation, deep learning, computer vision, image annotation, data augmentation, loss function.

1. Introduction

With the continuous development of advanced computer vision technologies, semantic segmentation has become one of the key tasks in certain data analysis and understanding. Its goal is to accurately classify regions in an image of pixel level [1] into different semantic categories. Semantic segmentation plays a crucial role in many application areas, such as autonomous driving, medical image analysis, image identification and intelligent video surveillance. With the advancement of deep learning techniques, especially the widespread use of Convolutional Neural Networks (CNN), significant progress has been made in semantic segmentation. Deep learning techniques, particularly the application of CNN, have brought about tremendous breakthroughs in semantic segmentation compared to traditional image segmentation methods. Traditional methods typically rely on manually designed feature extractors and complex image processing algorithms. Even with semi-automatic methods applied, there are still many inconveniences. On the other hand, deep learning methods can learn feature representations from a large amount of data, enabling more accurate and efficient semantic segmentation. However, despite the new hope brought by deep learning in semantic segmentation, there are still challenges to be addressed. Firstly, choosing an appropriate CNN architecture is crucial for the success of semantic segmentation. It can be challenging to find an approach that works well across all datasets and tasks due to varying network structures and their corresponding performances. Secondly, labelling the semantic segmentation dataset is often costly, and the accuracy of the annotations can affect the model. Additionally, issues such as class imbalance, fuzzy boundaries, and variations in target sizes are new challenges to semantic segmentation tasks. Among all these challenges, various loss functions exhibit distinct advantages and disadvantages in different scenarios.

The purpose of this paper is to provide a comprehensive overview of the latest research progress in the field of semantic segmentation and to summarize the latest methods and key technologies to help readers to understand the advanced information of current semantic segmentation research. Through this review, readers can analyze the key issues of semantic segmentation from different perspectives and gain a clearer understanding of the advantages and disadvantages of different methods. In addition, this paper will provide a comprehensive reference for both the academic and industrial communities in semantic segmentation research, to let them can choose appropriate methods for practical applications. Besides, we will also point out the problems and challenges existing in current research and analyze them further. The research content and discussions in this paper will be of great significance in promoting the development and application of the field of semantic segmentation.

2. Database and annotations

Database and annotations are the foundation of semantic segmentation since they determine the quality and diversity of the training data. Semantic segmentation requires large-scale databases which include pixel-wise annotations. They serve as the base of training deep learning models. Commonly used datasets include PASCAL VOC, Cityscapes, ADE20K and COCO, which consist of diverse scenes and object categories. The process of annotating these datasets can be a great waste of time-consuming and laborious. For example, manual annotation involves human experts carefully outlining object boundaries. It may assign semantic identifiers to text passages[2]. In contrast, semi-automated techniques like scribble-based annotation can only accelerate the process but take a lot of time still and sometimes will make mistakes as Mithun Kumar Kar et al. concluded that the primary challenges persist due to the variability of patient data and the limited availability of large datasets in medical imaging[3].

2.1. Annotations are of importance

The selection and annotation accuracy of the dataset forms the foundation of semantic segmentation research. High-quality annotations are vital to achieving accurate and convincing semantic segmentation models. The need for compatibility among dataset taxonomies is a problem that hinders advancements in unified semantic segmentation[4]. There are some situations that rare object classes cannot be ignored and are hard to have a consistent naming[5]. The accuracy and consistency of annotations are crucial for semantic segmentation research. It is necessary to address issues such as dataset category incompatibility, reduce subjectivity, and ensure consistency. In addition, annotations are also related to data augmentation since labels describe transformations[6].

2.2. Introduction of some datasets

Bolei Zhou et al., in their essay [5], mainly introduced scene paring. Scene parsing is a more strict form of scene labelling, where scene labelling involves dividing the entire image into regions and assigning these parts labels, sometimes even using rough approximations for labelling large areas. On the other hand, semantic segmentation focuses not on the entire image but only on the objects. So scene paring which can recognize and segment objects in an image is a key problem in computer vision and also semantic segmentation. In this writing, they analyzed the ADE20K dataset. Compared to other databases, for example, COCO, Imagenet, and Pascal-Context/Part dataset, the ADE20K dataset has ADE20K dataset. It even has a larger average number of object classes per image and in terms of images and object instances. Besides, a much broader set of object classes makes ADE20K have more part classes, showing the high annotation complexity of this database. Thus, they gave two ways to apply scene paring: automatic image content removal and scene synthesis.

John Lambert et al. mentioned in their writing[4] that the task of semantic segmentation is to divide an image into regions like objects, background and chaos. Nowadays, the accuracy of it on certain datasets is better and better. However, different datasets have different taxonomies, so they presented MSeg, a composite dataset that includes a large-scale annotation and bases on several datasets: COCO, ADE20K, Mapillary, IDD, BDD, Cityscapes, and SUN RGB-D.

2.3. Data augmentation

In many machine learning scenarios, even if some suitable datasets are chosen for a project, there may still be insufficient data to train high-quality models. To solve this problem, data augmentation can be used to enlarge the amount of available training data. Data augmentation is a very broad concept, including many certain ways to transfer training data and applying in the entire context of machine learning[6]. Data Augmentation is a solution for the problem of limited data, which comprises a set of techniques aimed at increasing the size and quality of training datasets, ultimately enabling the construction of improved Deep Learning models[7]. Besides, this technology is often used when it is necessary to increase a model's generalization capabilities. For the limited size of the training set, there is usually a fundamental issue that overfitting hampers our ability to achieve perfect generalization of models, both in terms of fitting well to observed data during training and to unseen data during testing[8]. Besides, deep convolutional neural networks heavily rely on large datasets to avoid overfitting, but numerous application domains lack access to large volumes of data[7]. Just to solve this issue, Alex Hernández-García and Peter König turned data augmentation into a better regularization, which plays a crucial role in machine learning[9]. They figured the improvement in generalization achieved through explicit regularization could be attained solely through data augmentation. Sometimes far too little data in research is a terrible problem called the "Big Data Wall"[10]. This wall can be broken by some projects focused on the use of practical, robust, scalable and easy-to-implement data augmentation pre-processing techniques. It is precisely because data augmentation plays a crucial role when there is insufficient data that many researchers have studied extensively. Shervin Minaee et al. presented an extensive review[11] of over 150 deep learning-based models for text classification. The main content of their task is to thoroughly examine these models' technical contributions, identify similarities, and highlight their strengths. In the aspect of language modelling, to cross the "Big Data Wall", Claude Coulombe did some research on related techniques and ensured that these APIs are robust, scalable and user-friendly. Minaee figured that huge GAFAMs (Google, Amazon, Facebook, Apple, Microsoft) were able to maintain their dominant position over smaller companies because they have access to such large amounts of data. Therefore, we can conclude that the development of more effective data augmentation techniques is beneficial for promoting fair competition across various industries.

There are also many ways to achieve data augmentation, for example, geometric transformations, neural style transfers, interpolation of images, random partial deletions, generative adversarial network (GAN) data generation, acoustic transformations of the input data, interfering with vocal tract length and adding noise[7]. Shichao Zhang et al. figured that in some situations, traditional data augmentation methods (cropping, rotating, and flipping) could not significantly improve the accuracy of detection and segmentation[12]. So they introduced a model called OFA-Net (One For All Network) that aims to integrate object detection and semantic segmentation tasks. This model is capable of fusing features from both detection and segmentation data. Frans P. Boogaard et al. found that the visibility of nodes significantly improves when multiple viewpoints in 2D are utilized[13]. However, during the tests, the count of nodes was consistently underestimated because of the limitation of 2D. Therefore, they expressed their anticipation for the application of 3D techniques. Research conducted by Martin Hahner et al. explored different global augmentation techniques and local augmentation techniques[14]. Both types of data augmentation led to enhanced performance. These findings indicate the potential for effectively applying and adapting these techniques to other advanced 3D object detection methods. Based on their discovery, Bolai Xin et al. explored a novel method[15] in this field, which was implemented on PointNet++ for 3D semantic segmentation of point clouds from three kinds of tomato plants. The point clouds were successfully segmented into different biological structures.

2.4. Algorithms and models of semantic segmentation

Algorithms and models are the core part of semantic segmentation. They all have good feature extraction abilities. These algorithms and models can be mainly divided into two parts: those based on image features and those based on deep learning.

2.5. Algorithms and models based on image features

Mithun Kumar Kar et al. turned image segmentation into a technique that involves clustering different regions of an image into distinct object classes[3]. They believed it is a crucial aspect of computer vision applications, as it directly impacts various critical tasks including image analysis, feature calculation, object detection, and classification. Since people are more focused on pixel-level segmentation[16] rather than localized segmentation of an image, semantic image segmentation provides multi-level representations of an image based on object classes by assigning class labels to each pixel.

This group of algorithms and models include those based on color and texture features, as well as those based on edge and texture features. These methods mainly rely on traditional computer vision techniques and feature extraction methods. All of them can catch semantic information to some extent, but, especially in complex scenes, they often suffer from the limitation of the ability to represent limited features.

2.6. Algorithms and models based on image color and texture features

Using algorithms and models based on image color and texture features is one of the traditional methods of image segmentation. To achieve pixel-level semantic segmentation, the fundamental idea of it is to analyze and model the color and texture information in the image. Nowadays, commonly used color feature algorithms include color histograms, color clustering, and color space transformations.

Dengsheng Zhang in his paper described that he believed the feature of the most significant of images is color[17]. Color feature algorithms are powerful tools for image retrieval and recognition. In addition, the component of color histogram is explained in detail. He defined three ways to create color histograms: component histogram, indexed colour histogram, and dominant color histogram. Furthermore, he calculated the color coherence vector (CCV) as it incorporated spatial information into color histograms.

As for color clustering, Dengsheng Zhang also described K-means clustering, as one of the most classic segmentation methods[17]. And others also did some research in the same way. Leon A. Gatys et al. defined color clustering as a way to reduce the number of colors in the image while maintaining its visual quality[18]. In this study, the authors acknowledge a possible limitation of the original method, which involves the transfer of colors from the original image. This process has the potential to modify the appearance of the image in unexpected ways. Their main work of them is to compare two ways: color histogram matching and luminance-only transfer. They finally figure style transfer only in the luminance channel is better. After CAI Leizhen and LEUNG On Yin turned to the parameterized complexity of Coloured Clustering, and give FPT algorithms, where parameter k is for both the number of stable edges and the number of unstable edges[19]. In February 2023 Leon Kellerhals et al. provided more detailed information about color clustering [20]. To extend the former study, they gave an algorithm, where k represents the number of edges to be selected and n corresponds to the number of vertices.

Roman Starosolski conducted research on color space transformations[21]. In traditional lossless image compression methods, color space transformation is a crucial step aimed at reducing redundancy and improving compression efficiency. However, existing color space transformation techniques may suffer from inefficiency or significant distortion issues. The author proposes a novel, simple, and efficient color space transformation method for achieving lossless image compression. This method is based on linear transformations of color channels, which reduce redundancy while preserving the accuracy of image colors. Selecting appropriate transformation matrices, makes it possible to maintain image quality while reducing compression bit rates.

DOI: 10.54254/2755-2721/33/20230242

2.7. Algorithms and models based on boundary features

Segmentation methods based on boundary features identify boundary information by detecting changes in brightness, allowing for the description of object styles. These methods then group these boundary information points to determine the final segmentation results. Commonly used boundary detection algorithms include Canny boundary detection, Sobel operator, Laplacian operator, etc.

The Canny edge detection algorithm is a gradient-based method for detecting edges, which is capable of identifying strong edges in an image while exhibiting excellent noise robustness.

The traditional Canny algorithm requires manual threshold selection, which becomes challenging due to the complexity of images and noise interference. To address this issue, Keong-Hun Choi et al. proposed an adaptive threshold method[22] based on the Actor-Critic algorithm[23]. This method introduces an Actor-network and a Critic network, where the Actor-network generates candidate thresholds and the Critic network evaluates the performance of these thresholds. By employing a reinforcement learning framework, the Actor-Critic algorithm gradually adjusts the thresholds and automatically learns the optimal threshold during the training process to adapt to the characteristics and noise levels of different images, thereby enhancing the performance and accuracy of Canny boundary detection. Then about the application of this similar method in related fields, Yuan Chao et al. introduced an adaptive threshold selection mechanism and proposed a boundary location and identification method for electronic components based on an improved Canny algorithm[24]. As they mentioned, four directions (horizontal, vertical, and diagonal) calculated the gradient magnitude. Moreover, it has to be mentioned, segmentation predictions are more likely to make mistakes near the boundary[25]. To further enhance segmentation by using boundary information, Yuhui Yuan et al. mainly used two steps: localizing the pixels of the boundary and determining the corresponding interior pixel for each boundary pixel[26]. They establish the correspondence by training a model to determine a direction pointing from the boundary pixel towards an interior pixel. And the team of Towaki Takikawa made their model specifically emphasize the processing of boundary-related information[27], leading to an architecture that is highly effective in producing sharper predictions around object boundaries. Besides, Shubhankar Borse et al. proposed a novel semantic segmentation approach[28] based on boundary-aware loss, which effectively learns to estimate the degree of parameter transformation between boundaries and object boundaries. Jie Du et al. proposed Boundary-Sensitive loss (BS-loss)[29], which can automatically prioritize challenging boundaries for segmentation, such as thin structures and blurred boundaries. This allows for the generation of more precise object boundaries.

3. Some operators

Three methods for detecting edges in images are the Sobel operator, the Laplacian operator, and the Gaussian operator. These methods use convolutional operations to compute the difference in grayscale values of pixels and apply Gaussian smoothing to enhance the edges.

In traditional quantum image edge detection methods, there are issues of low accuracy and high computational complexity. The study[30] of Wenjie Liu et al. proposed a new edge detection method, Quantum Sobel edge detection (QSED), based on NEQR (a novel enhanced quantum representation)[31]. This method is for quantum images using an eight-direction Sobel operator. The method utilizes the eight-direction Sobel operator to calculate the edge responses in the quantum image, enabling accurate edge localization and identification while effectively reducing noise interference in quantum images.

Suzhen Yuan et al. presented a fast Laplacian of Gaussian (LoG) edge detection algorithm[32] for quantum images. The algorithm obtains a discrete Laplacian of Gaussian (LoG) mask by discretizing the continuous Gaussian operator. It consists of Gaussian filtering, zero crossing operation, and threshold detection. These techniques enable the algorithm to perform edge detection in a shorter amount of time while maintaining high accuracy.

DOI: 10.54254/2755-2721/33/20230242

4. Algorithms and models based on deep learning

Unlike traditional methods, deep learning-based semantic segmentation algorithms and models utilize deep neural networks for end-to-end training and learning, enabling them to better capture the semantic information of images. Typical deep learning models include Convolutional Neural Networks (CNNs), Fully Convolutional Networks (FCNs), U-Net, and SegNet, etc. These models leverage the learning of local and global features in images to achieve pixel-level semantic segmentation, resulting in more accurate segmentation results in complex scenes. Additionally, deep learning-based methods exhibit strong adaptability and generalization capabilities, allowing them to adapt to different scenarios and data characteristics. Some scientists like Mithun Kumar Kar et al. believe with the increasing availability of datasets and graphical processing units, deep learning-based semantic image segmentation techniques have become more accurate[3].

George Papandreou et al. described proposed a solution for processing weakly labelled data in Convolutional Neural Networks (CNN) and the combination of well-labelled data and improperly labelled data[33]. The combination of a Deep Convolutional Neural Network (DCNN)[34] and a fully connected Conditional Random Field (CRF)[35] was used to get high-resolution segmentations. They developed novel online Expectation-Maximization (EM) methods[33] that can be used for training in both weakly-supervised and semi-supervised settings, utilizing image-level or bounding box annotation. They demonstrated that the integration of weak or strong annotations from multiple datasets leads to additional enhancements. However, there are still some problems. Towaki Takikawa et al. introduced a new two-stream CNN architecture called Gated-SCNN (GSCNN)[27] that changed the situation that Deep CNNs handle various types of information related to recognition in image data, such as color, shape, and texture information, all together.

E. Shelhamer et al., in their article[36], introduced Fully Convolutional Networks (FCNs), a semantic segmentation approach based on convolutional neural networks (CNNs). The authors transform the CNN architecture into a fully convolutional form, enabling it to process input images of arbitrary sizes and produce dense prediction maps of the same dimensions. The paper explores the application and performance of FCNs in semantic segmentation tasks and discusses improvement strategies for network architecture, such as skip connections. The fully connected layers of the network accept images of any size and generate outputs of corresponding spatial dimensions.

Olaf Ronneberger et al. proposed a network called U-Net[37] along with a training strategy that relies heavily on data augmentation to efficiently provide annotated samples. The background lies in the field of biomedical research, where accurately identifying and segmenting structures and tissues in images is a crucial task. The main challenge in biomedical tasks is the difficulty of acquiring thousands of images for training. This paper builds upon the foundation of fully convolutional layers and modifies them to handle a range of training images and produce more precise segmentations, such as accurately segmenting target structures in biomedical images, including cells, organs, or pathological regions.

Vijay Badrinarayanan et al. introduced SegNet[1], a deep convolutional neural network architecture designed specifically for pixel-wise image segmentation. SegNet follows an encoder-decoder architecture, where the encoder extracts high-level features from the input image, and the decoder follows by a pixel-wise classification layer and reconstructs the segmentation map from the encoded features.

5. Loss function

Selecting an appropriate loss function is crucial when designing complex deep-learning architectures for image segmentation[16], as it greatly influences the learning process of the algorithm. When dealing with complex objectives like segmentation, it is not feasible to determine a single universal loss function. In most cases, the choice of loss function depends on the specific properties of the dataset used for training. Table 1 below shows the 4 categories[38] Jun Ma divided the loss function into.

Table 1. Categories of se,amtic segmentation loss functions[38].

Categories	Loss Function
Distribution-based Loss	CE, WCE, DPCE, TopK loss, Focal loss
Compound Loss	ELL, Combo
Region-based Loss	SS, IoU/Jaccard, Dice, Tversky, GD, pGD, Asym., Focal Tversky
Boundary-based Loss	HD Loss, Boundary Loss

Towaki Takikawa et al. used a new loss function[27] that exploited the duality between the tasks of semantic segmentation and semantic boundary prediction. Shubhankar Borse et al. believed that cross-entropy loss did not take into consideration the spatial distance between pixels and the target boundaries[28]. As a result, it is unable to effectively measure localized spatial changes, such as translation, rotation, or scaling, between predicted and target boundaries. Thus they introduce a boundary distance-based loss function called InverseForm. Sarmad F. Ismael et al. introduced a new loss function[39] that combines weighted focal loss with Jaccard loss. This loss function achieves better performance compared to traditional cross-entropy and focal loss in terms of mean intersection over union (mIoU) and overall accuracy. Tsung-Yi Lin et al. proposed a new loss function called Focal Loss[40] to solve the class imbalance. This loss function is an improvement over the cross entropy loss and allows the model to focus more on complex negative examples. And then, the team of Xiang Li proposed Generalized Focal Loss (GFocal)[41], which extends the concept of Focal Loss from its discrete form to a continuous version. This modification enables successful optimization of the loss function. In terms of computer-aided diagnosis and treatment planning, the team of Jie Du introduced a boundary-sensitive loss function[29], which can do well in segmenting hard regions and enhances the overall performance of segmentation in medical images.

6. Conclusion

This article provides a comprehensive review of the research progress in the field of semantic segmentation in recent years, covering the key aspects such as relevant databases and their annotations, the application of related algorithms and models, data augmentation techniques, and loss functions research. Through analysis and discussion of these various aspects, we have drawn the following conclusions:

1.We should attach great importance to databases and annotations in semantic segmentation research. Accurate and consistent annotations contribute to obtaining better semantic segmentation models. However, challenges remain, such as dataset incompatibility and annotation consistency. Future research can focus on solving these problems and exploring more efficient annotation techniques and methods, providing more effective support for the development of the field of semantic segmentation.

2.Data augmentation techniques play a crucial role in addressing the issue of insufficient data by increasing the size and quality of the training datasets. These methods improve the model's generalization ability and performance, dealing with the problem of overfitting caused by limited training data. With the continuous development of technology, we can expect more techniques, such as AI, to be applied in data augmentation.

3.The related algorithms and models have powerful performance and generalization abilities in semantic segmentation tasks. From early network models to the latest innovative methods, the continuous evolution of these methods has been driving the development of the field of semantic segmentation. However, there are still many unresolved issues, such as finding a balance between network depth and complexity, optimizing model parameters, and so on.

4.Choosing the appropriate loss function is crucial for segmentation projects. Different loss functions may have different effects on different situations and datasets. Therefore, researchers need to select suitable loss functions based on specific situations and continuously explore new loss functions to improve model performance and generalization ability.

This research has analyzed current challenges and issues by thoroughly exploring research progress in semantic segmentation. This research will provide researchers and practitioners with an improved understanding of the latest advancements and critical technologies for semantic segmentation tasks. Additionally, this research proposes future research directions to promote the development of the field. This research anticipates that better semantic segmentation methods will emerge and be applied in broader applications, providing more value and significance to human society.

References

[1] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 12, pp. 2481-2495, 2017, doi: 10.1109/tpami.2016.2644615.
[2] A. Hinze, R. Heese, A. Schlegel, and A. Paschke, "Manual semantic annotations: User evaluation of interface and interaction designs," Journal of Web Semantics, vol. 58, p. 100516, 2019/10/01/ 2019, doi: https://doi.org/10.1016/j.websem.2019.100516.
[3] M. K. Kar, M. K. Nath, and D. R. Neog, "A Review on Progress in Semantic Image Segmentation and Its Application to Medical Images," SN Computer Science, vol. 2, no. 5, 2021, doi: 10.1007/s42979-021-00784-5.
[4] J. Lambert, Z. Liu, O. Sener, J. Hays, and V. Koltun, "MSeg: A composite dataset for multi-domain semantic segmentation," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 2879-2888.
[5] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, "Scene Parsing through ADE20K Dataset," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 21-26 July 2017 2017, pp. 5122-5130, doi: 10.1109/CVPR.2017.544.
[6] M. Bayer, M.-A. Kaufhold, and C. Reuter, "A Survey on Data Augmentation for Text Classification," ACM Computing Surveys, vol. 55, no. 7, pp. 1-39, 2022, doi: 10.1145/3544558.
[7] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," Journal of Big Data, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.
[8] Y. Xue, "An Overview of Overfitting and its Solutions," Journal of Physics: Conference Series, vol. 1168, no. 2, p. 022022, feb 2019, doi: 10.1088/1742-6596/1168/2/022022.
[9] A. Hernández-García and P. König, "Data augmentation instead of explicit regularization," ArXiv, vol. abs/1806.03852, 2018.
[10] C. Coulombe, "Text Data Augmentation Made Simple By Leveraging NLP Cloud APIs," ArXiv, vol. abs/1812.04718, 2018.
[11] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. A. Chenaghlu, and J. Gao, "Deep Learning--based Text Classification," ACM Computing Surveys (CSUR), vol. 54, pp. 1 - 40, 2020.
[12] S. Zhang, Z. Zhang, L. Sun, and W. Qin, "One For All: A Mutual Enhancement Method for Object Detection and Semantic Segmentation," Applied Sciences, vol. 10, no. 1, p. 13, 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/1/13.
[13] P. B. Frans, S. A. H. R. Kamiel, and W. K. Gert, "Robust node detection and tracking in fruit-vegetable crops using deep learning and multi-view imaging," Biosystems Engineering, vol. 192, pp. 117-132, 2020, doi: https://doi.org/10.1016/j.biosystemseng.2020.01.023.
[14] H. Martin, D. Dengxin, L. Alexander, and G. Luc Van, "Quantifying Data Augmentation for LiDAR based 3D Object Detection," ArXiv, vol. abs/2004.01643, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:214794957.
[15] B. Xin, J. Sun, H. Bartholomeus, and G. Kootstra, "3D data-augmentation methods for semantic segmentation of tomato plant parts," (in English), Frontiers in Plant Science, Original Research vol. 14, 2023-June-12 2023, doi: 10.3389/fpls.2023.1045545.
[16] S. Jadon, "A survey of loss functions for semantic segmentation," 2020.

[17] D. Zhang, "Color Feature Extraction," in Fundamentals of Image Data Mining: Analysis, Features, Classification and Retrieval, D. Zhang Ed. Cham: Springer International Publishing, 2019, pp. 49-80.
[18] L. A. Gatys, M. Bethge, A. Hertzmann, and E. Shechtman, "Preserving color in neural artistic style transfer," arXiv preprint arXiv:1606.05897, 2016.
[19] C. Leizhen and L. On Yin, "Alternating Path and Coloured Clustering," ArXiv, vol. abs/1807.10531, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:51864778.
[20] L. Kellerhals, T. Koana, P. Kunz, and R. Niedermeier, "Parameterized Algorithms for Colored Clustering," arXiv preprint arXiv:2302.00630, 2023.
[21] R. Starosolski, "New simple and efficient color space transformations for lossless image compression," Journal of Visual Communication and Image Representation, vol. 25, 07/01 2014, doi: 10.1016/j.jvcir.2014.03.003.
[22] K.-H. Choi and J.-E. Ha, "An Adaptive Threshold for the Canny Edge Detection with Actor-Critic Algorithm," arXiv preprint arXiv:2209.08699, 2022.
[23] V. Mnih et al., "Asynchronous methods for deep reinforcement learning," presented at the Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48, New York, NY, USA, 2016.
[24] Y. Chao, F. Shi, W. Shan, and D. Liang, "Edge Location and Identification Method for Electronic Components based on Improved Canny Algorithm," International Journal of Circuits, Systems and Signal Processing, vol. 16, pp. 209-214, 2022.
[25] A. Kirillov, Y. Wu, K. He, and R. B. Girshick, "PointRend: Image Segmentation As Rendering," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9796-9805, 2019.
[26] Y. Yuan, J. Xie, X. Chen, and J. Wang, "SegFix: Model-Agnostic Boundary Refinement for Segmentation," ArXiv, vol. abs/2007.04269, 2020.
[27] T. Takikawa, D. Acuna, V. Jampani, and S. Fidler, "Gated-SCNN: Gated Shape CNNs for Semantic Segmentation," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 5228-5237, 2019.
[28] S. Borse, Y. Wang, Y. Zhang, and F. Porikli, "Inverseform: A loss function for structured boundary-aware segmentation," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 5901-5911.
[29] J. Du, K. Guan, P. Liu, Y. Li, and T. Wang, "Boundary-Sensitive Loss Function With Location Constraint for Hard Region Segmentation," IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 2, pp. 992-1003, 2023, doi: 10.1109/JBHI.2022.3222390.
[30] W. Liu and L. Wang, "Quantum image edge detection based on eight-direction Sobel operator for NEQR," Quantum Information Processing, vol. 21, no. 5, 2022, doi: 10.1007/s11128-022-03527-4.
[31] Y. Zhang, K. Lu, Y. Gao, and M. Wang, "NEQR: a novel enhanced quantum representation of digital images," Quantum Information Processing, vol. 12, no. 8, pp. 2833-2860, 2013, doi: 10.1007/s11128-013-0567-z.
[32] S. Yuan, S. E. Venegas-Andraca, C. Zhu, Y. Wang, X. Mao, and Y. Luo, "Fast Laplacian of Gaussian Edge Detection Algorithm for Quantum Images," presented at the 2019 IEEE International Conferences on Ubiquitous Computing & Communications (IUCC) and Data Science and Computational Intelligence (DSCI) and Smart Computing, Networking and Services (SmartCNS), 2019.
[33] G. Papandreou, L.-C. Chen, K. P. Murphy, and A. L. Yuille, "Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1742-1750.
[34] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected

CRFs," IEEE Trans Pattern Anal Mach Intell, vol. 40, no. 4, pp. 834-848, Apr 2018, doi: 10.1109/TPAMI.2017.2699184.
[35] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected CRFs with Gaussian edge potentials," presented at the Proceedings of the 24th International Conference on Neural Information Processing Systems, Granada, Spain, 2011.
[36] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," IEEE Trans Pattern Anal Mach Intell, vol. 39, no. 4, pp. 640-651, Apr 2017, doi: 10.1109/TPAMI.2016.2572683.
[37] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, (Lecture Notes in Computer Science, 2015, ch. Chapter 28, pp. 234-241.
[38] M. Jun, "Segmentation Loss Odyssey," ArXiv, vol. abs/2005.13449, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:218900753.
[39] S. F. Ismael, E. Aptoula, and K. Kayabol, "A Joint Semantic Segmentation Loss Function for Imbalanced Datasets," 2022.
[40] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 2, pp. 318-327, 2020, doi: 10.1109/TPAMI.2018.2858826.
[41] X. Li, C. Lv, W. Wang, G. Li, L. Yang, and J. Yang, "Generalized Focal Loss: Towards Efficient Representation Learning for Dense Object Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 3, pp. 3139-3153, 2023, doi: 10.1109/TPAMI.2022.3180392.