Al-Kitab Journal for Pure Sciences (2026); 10(1):01-14.
https://doi.org/10.32441/kjps.10.1.p1

Al-Kitab Journal for Pure Sciences
ISSN: 2617-1260 (print), 2617-8141(online)
https://isnra.net/index.php/kjps

Semantic Segmentation of image Using Deep Learning: Review

Mohammad Kamal Ahmed[1*], Asim Majeed Murshid[2]

[1,2]**Department of Computer Science, college of Computer Science and Information Technology, University of Kirkuk, Iraq.**

[*]Corresponding Author: **stch21m011@uokirkuk.edu.iq**

Citation: ***Mohammad K. A., Murshid A. M.*** **Semantic Segmentation of image Using Deep Learning:Review**. Al-Kitab J. Pure Sci. [Internet]. 2025 Jan 12; 10(1):01-14 DOI:
https://doi.org/10.32441/kjps.10.01.p1

Keywords: Semantic Segmentation,U-Net Architecture, Convolutional Neural Networks, Deep Learning, Artificial Intelligence.

Article History

Received	20 Dec. 2024
Accepted	12 Jan. 2025
Available online	01 May. 2026

Abstract:

Semantic segmentation is considered as one of the most important and challenging problem in the field of computer vision which aims at assigning a class label to each pixel in an image which leads to sophisticated scene understanding. This task has been extensively used in various application areas including the self-driving cars, medical diagnosis, and environmental monitoring. Semantic segmentation has come a long way since its early detection algorithms based on features extractions to state-of-the-art deep learning methods .

This paper aims at presenting the evolution of the semantic segmentation, and specifically, how the deep learning has changed the field. The conventional approaches consisting of edge detection and histogram analysis offered a basic level of understanding but were constrained by the use of hand crafted features. Deep learning, however, is capable of learning features and has produced very promising results across numerous tasks. Some important architectures that have set the benchmark in the field include Fully Convolutional Networks (FCNs), U-Net, and DeepLab which have used convolutional layers, encoder-decoder architecture, and atrous convolutions for improving the accuracy of the segmentation. The article also reviews some of the publicly available datasets which include Cityscapes, PASCAL VOC and ISIC 2017 which are widely used to assess the performance of the segmentation models. These datasets differ in their complexity, resolution, and the application domain that they cover which makes the problems that they present to researchers diverse .Also, we compare the traditional and deep learning based feature extraction methods and present the characteristics of each method, their advantages, and disadvantages, and areas of application. This survey aims at assisting researchers and practitioners by presenting the current best practice in the form of state-of-the-art methodologies, discussing the potential of application of such methodologies in the real world, and identifying the directions for further research.Therefore, despite the advancement of deep learning in the area of semantic segmentation, there are still numerous issues which need to be addressed in the future, including efficiency, scalability, and domain specific issues. This all-encompassing review paper is hoped to be beneficial to those wishing to gain more knowledge on the current trends as well as find a way to contribute to the field of semantic segmentation in the future.

Keywords: **Semantic Segmentation, U-Net Architecture, Convolutional Neural, Artificial Intelligence, deep learning**.

التجزئة الدلالية للصورة باستخدام التعلم العميق: مراجعة

محمد كمال احمد , عاصم مجيد مرشد

stch21m011@uokirkuk.edu.iq

الخلاصة:

تعتبر التجزئة الدلالية واحدة من أهم المشاكل وأكثرها تحديًا في مجال الرؤية الحاسوبية والتي تهدف إلى تعيين تسمية فئة لكل بكسل في الصورة مما يؤدي إلى فهم متطور للمشهد. تم استخدام هذه المهمة على نطاق واسع في مجالات تطبيقية مختلفة بما في ذلك السيارات ذاتية القيادة والتشخيص الطبي ومراقبة البيئة. لقد قطعت التجزئة الدلالية شوطًا طويلاً منذ خوارزميات الكشف المبكرة القائمة على استخراج الميزات إلى أحدث أساليب التعلم العميق. تهدف هذه الورقة إلى تقديم تطور التجزئة الدلالية، وعلى وجه التحديد، كيف غيّر التعلم العميق هذا المجال. قدمت الأساليب التقليدية المكونة من اكتشاف الحافة وتحليل الهيستوجرام مستوى أساسيًا من الفهم ولكنها كانت مقيدة باستخدام ميزات مصنوعة يدويًا. ومع ذلك، فإن التعلم العميق قادر على تعلم الميزات وقد أنتج نتائج واعدة للغاية عبر العديد من المهام. تتضمن بعض البنيات المهمة التي وضعت معيارًا في هذا المجال الشبكات التلافيفية الكاملة (FCNs) وU-Net وDeepLab التي استخدمت طبقات التلافيفية، وبنية الترميز وفك التشفير، والالتواءات غير المتجانسة لتحسين دقة التجزئة. تستعرض المقالة أيضًا بعض مجموعات البيانات المتاحة للجمهور والتي تتضمن Cityscapes وPASCAL VOC وISIC 2017 والتي تُستخدم على نطاق واسع لتقييم أداء نماذج التجزئة. تختلف مجموعات البيانات هذه في تعقيدها ودقتها ومجال التطبيق الذي تغطيه مما يجعل المشكلات التي تقدمها للباحثين

متنوعة. كما نقارن بين طرق استخراج الميزات التقليدية والتعلم العميق ونقدم خصائص كل طريقة ومزاياها وعيوبها ومجالات التطبيق. يهدف هذا الاستطلاع إلى مساعدة الباحثين والممارسين من خلال تقديم أفضل الممارسات الحالية في شكل منهجيات حديثة، ومناقشة إمكانات تطبيق مثل هذه المنهجيات في العالم الحقيقي، وتحديد الاتجاهات لمزيد من البحث. لذلك، وعلى الرغم من التقدم الذي أحرزته تقنيات التعلم العميق في مجال التجزئة الدلالية، لا تزال هناك العديد من القضايا التي تحتاج إلى معالجة في المستقبل، بما في ذلك الكفاءة وقابلية التوسع والقضايا الخاصة بالمجال. ونأمل أن تكون ورقة المراجعة الشاملة هذه مفيدة لأولئك الذين يرغبون في اكتساب المزيد من المعرفة حول الاتجاهات الحالية وكذلك إيجاد طريقة للمساهمة في مجال التجزئة الدلالية في المستقبل.

الكلمات المفتاحية (التجزئة الدلالية، معمارية U-Net، الشبكات العصبية التلافيفية، الذكاء الاصطناعي، التعلم العميق)

1- Introduction

Semantic segmentation is one of the most important and widely used task in computer vision which involves assigning of a specific label for each pixel in an image so as to provide a detailed understanding of the scene. This capability is very important for numerous applications such as autonomous driving, medical imaging, environmental monitoring, and augmented reality. Since the development of semantic segmentation techniques, the field has grown from the use of traditional hand crafted feature based approaches to the current state of the art deep learning approaches. In the self-driving domain, semantic segmentation is very important as it assists in localizing lanes, obstacles, as well as pedestrians to enhance safety of the car(1). In medical context, it is utilized for the identification of the anatomic structures or pathological regions, which may facilitate the diagnosis and treatment decision(2). Likewise, in the environmental monitoring, it can be employed to classify land cover which may in turn facilitate better management of resources and Disaster response(3). These examples show how semantic segmentation can contribute to the development of new technologies in different fields. Previous semantic segmentation methods include traditional approaches like edge detection, region growing and clustering as the basic approaches. Although these methods offered the initial knowledge, they failed to deal with the complicated patterns, light variations, as well as the objects of interest with complex shapes (4). Deep learning has brought a shift in the field by allowing the network to learn features and hierarchical representation on its own. Some of the architectures that have set new benchmarks in the field include Fully Convolutional Networks (FCNs) (5), U-Net (6) ,and DeepLab (7), among others. Research in semantic segmentation has been greatly advanced by the availability of data sets that have been made public. Extension of benchmark datasets such as Cityscapes, PASCAL VOC, ADE20K, ISIC are used for evaluation of models. For instance, Cityscapes specifically target the urban environment to facilitate research on autonomous driving while ISIC contains dermoscopic images with annotations to support advancement of medical imaging (8). These datasets are used for evaluating the algorithms and also to see where exactly the algorithms require enhancement. Current research focuses on the development of context sensitive segmentation models. Some techniques like Atrous Spatial Pyramid Pooling (ASPP) in DeepLab (7) and Transformer-based architetures (9) have been proved effective in learning the contextual information and long-range dependency thus improving the accuracy of segmentation. Furthermore, the pre-trained models and the transfer learning and fine-tuning of these models have been established as the standard approaches to learn from large datasets to particular domains efficiently (10). However, there are still some challenges in semantic segmentation despite these developments. Some of the challenges that are still present include; high computational costs, annotation challenges and domain adaptation problems. Lightweight architectures and semi-supervised learning approaches are currently under a lot of research to solve such issues (11). Also, another important aspect that is being paid much attention to is how to achieve real-time performance of the model without sacrificing the accuracy for applications such as robotics and self-driving cars(12). Comparing to the previous evolution of semantic segmentation, this survey identifies and analyzes the most significant methods, their applications, and further research directions.

Table1: A list of the most important abbreviations:

Abbreviations	Full Form	Abbreviations	Full Form
AI	Artificial Intelligence	SIFT	Scale-Invariant Feature Transform
CNN	Convolutional Neural Network	ViT	Vision Transformer
FCN	Fully Convolutional Network	mIoU	Mean Intersection over Union
DL	Deep Learning	GLCM	Gray-Level Co-occurrence Matrix
IoU	Intersection over Union	HOG	Histogram of Oriented Gradients
ISIC	International Skin Imaging Collaboration	ResNet	Residual Network
ASPP	Atrous Spatial Pyramid Pooling	CNNs	Convolutional Neural Networks

2- DATASETS

The datasets are crucial for the growth of semantic segmentation field as they present various situations and benchmarks to assess the models. They differ in their size, the resolution of the images, the number of classes, and the area of application. Table 2 presents 10 most popular available datasets which are commonly used in urban scene understanding, medical imaging, and general object segmentation.

Table 2: Overview of publicly accessible datasets

Dataset	Application Domain	Resolution	Images/Size	Classes	Year	Ref.
Cityscapes	Urban scenes	2048×1024	5,000 images	30	2016	(13)
PASCAL VOC	Object recognition	Variable	11,530 images	20	2012	(14)
ADE20K	General scenes	Variable	27,574 images	150	2017	(15)
COCO	Generaobject detection	Variable	328,000 images	80	2017	(16)
ISIC 2017	Medical imaging	1024 × 1024	2,750 images	2	2018	(17)
CamVid	Urban scenes	960×720	701 images	32	2008	(18)
Mapillary Vistas	Street-level imagery	Variable	25,000 images	66	2017	(19)
KITTI	Autonomous driving	1242×375	15,000 frames	34	2015	(20)
DRIVE	Retinal imaging	565×584	40 images	2	2004	(21)
LIP	Human parsing	Variable	50,462 images	20	2018	(22)

3-LITERATURE REVIEW

Semantic segmentation is one of the most important areas in the field of computer vision and has been greatly enhanced by the advancement in deep learning (DL) techniques. Several research works have been conducted to find the best DL approach that would enhance the segmentation accuracy, efficiency, and applicability of the models in different fields. This section discusses in detail the key research works done between the years 2020 to 2024 in descending order of years.

(Lee, MyeongSeok, 2022) presented ShuffleNetV2+, a light-weight model for real-time semantic segmentation on edge devices. The model provided the competitive performance on the Cityscapes dataset in terms of the mIoU score of 89.4 while consuming much less computational power as compared to other existing approaches(23)

(Zhao et al., 2020) introduced a new framework called PointRend for enhancing the segmentation of boundaries with the help of point-wise features. This approach set new records of performance on COCO and ADE20K datasets, especially in the task of the boundary refinement(24) **(Dosovitskiy et al., 2020)** introduced Vision Transformers (ViT), a new architecture that used global attention to learn the relationships between different parts of an image. This transformer-based approach provided reasonable segmentation performance on

PASCAL VOC and Cityscapes datasets, indicating the shift from convolutional neural networks (25). **(Xie et al., 2021)** presented SegFormer, a new and straightforward transformer-based architecture for semantic segmentation. Due to the proper balance between the computational cost and the accuracy of the segmentation, SegFormer provided remarkable results on ADE20K and COCO datasets and was identified as a viable solution for task that require limited resources (26). **(Esteva et al., 2021**) focused on medical imaging where U-Net and its variants were used for the skin lesion segmentation. Implementing the approach on the ISIC 2017 dataset, the method gave a high segmentation score of 95% thus proving the utility of deep learning in medical context (27). **(He, Xin, 2022)** the performance of the transformer-based methods was enhanced by Cheng et al. with Swin Transformers. With the help of Swin Transformers, they used a nested feature extraction approach and got the best results on large scale datasets such as ADE20K with efficient computational cost (28). **(Bragagnolo, L.2021)** Deep learning has demonstrated potential in automating landslide mapping. A study employing the U-Net architecture on Landsat-8 imagery of Nepal attained recall, precision, and F1-scores of 0.74, 0.61, and 0.67, respectively. This illustrates U-Net's capability for developing and updating landslide scar databases, thereby improving landslide susceptibility mapping and risk evaluation(29). **(Ishihara, 2021)** presented a novel multi-task learning authors approach incorporated for segmentation improving tasks the with segmentation other performance tasks in such autonomous as driving. object In detection this and work, observed the a boost in performance on challenging categories including road signs and pedestrians on the Cityscapes dataset (30). **(Su, Daobilige, 2021)** developed a novel dataset for agricultural robotics which made it possible to state the segmentation tasks in farmlands. The benchmark dataset allowed the development of a new hybrid transformer-CNN model that provided 92% mIoU and opened new opportunities for the precision agriculture branch (31). **(Grill et al., 2020)** introduced BYOL (Bootstrap Your Own Latent), an SIML that greatly reduced the need for labelled data. Pretraining of segmentation models with the help of BYOL enhanced the generalization of the models especially in the settings where a large number of labeled data are not available (32). **(Mao, Wei, 2022)** presented a semi-supervised approach for semantic segmentation and set new records on Cityscapes and COCO datasets. The framework proved that it is possible to get almost comparable results with the models trained on a large number of labeled images using only several hundred labeled samples (33). **(Mirikharaji, 2023**) revised the ISIC dataset used for dermatological image segmentation with high-quality annotations collected from multiple sources.This enriched the dataset which in turn helped enhance the model training for skin lesion detection thus adding value to the medical field (34).

(Kumar, P, 2023) added new annotations to the COCO dataset to make it suitable for semantic segmentation. The enlarged dataset offered various challenges which led to the advancement of the latest segmentation models (35). **(Liu, Yanyan., 2024)** incorporated an attention mechanism into the DeepLab framework so that it could handle the details of the images. This modification led to an enhancement of performance in on 2022 benchmarks Wu PASCAL et VOC al. and ADE20K worked (36). **(Sun, Yanwen, 2021)** A novel lightweight deep neural network has been introduced for the real-time semantic segmentation of surgical instruments in Robot-Assisted Minimally Invasive Surgery (RMIS). The model employs Ghost modules integrated with MobileNetV3 in a two-stage methodology to maintain high accuracy while minimizing computational costs. The validation results indicate that the model excels in real-time segmentation relative to existing methods, rendering it appropriate for surgical intelligence applications(37). **(Niu, Ruigang., 2022)** Is challenging to resolve the semantic segmentation within the global VHR context. pertaining to the air to address this issue, a novel framework for images, termed Hybrid Multiple Attention Network (HMANet), was proposed. It incorporates Class Augmented Attention (CAA) and Region Shuffle Attention (RSA) modules to enhance category-level relationships while simultaneously decreasing computational expenses. Evaluations on the ISPRS Vaihingen, Potsdam, and iSAID datasets indicate that HMANet surpasses existing methods in both accuracy and efficiency.(38) **(Subasi, 2024)** introduced a dataset and a framework for 3D Semantic Segmentation in the field of healthcare with emphasis on MRI scans. The method used in this paper was a 3D CNN based approach which outperformed other methods in localizing brain tumors on the BraTS dataset(39).

(Lai, Xin, 2021) explored semi-supervised learning for semantic segmentation with the help of consistency training and data augmentation. Their framework produced impressive results on Cityscapes with only a fraction of the labeled data, thus highlighting the efficiency and scalability of the approach(40)

4-FEATURE EXTRACTION METHODS

Feature extraction is very important in semantic segmentation since it defines the quality of information that is used to distinguish between pixels and assign them to certain categories. It will therefore improve the models' ability to comprehend spatial, contextual and semantic information. This section gives an overview of the conventional feature extraction techniques and the state-of-the-art approaches which are based on deep learning.

4-1 TRADITIONAL METHODS

The conventional approaches used feature engineering where features were hand crafted and had to be fine tuned. These methods were useful in simple applications and when dealing with simple patterns and variations in real world images.

- **Edge Detection**: The Sobel operator, Canny edge detector, and Laplacian of Gaussian were the most used methods for the extraction of edge information from the images. These methods offered fundamental structural information but had no semantic understanding(41)
- **Histogram of Oriented Gradients (HOG)**: HOG was employed to describe the shape and texture of the object in the region of interest by computing the distribution of gradient orientations. This was especially useful in identifying rigid objects but had a difficulty in identifying complex textures(42)
- **Scale-Invariant Feature Transform (SIFT)**: SIFT identified interest points and their corresponding descriptors which allowed for matching across scales and rotations. Although very precise, it was rather slow and therefore not suitable for real-time applications(43)
- **Gray-Level Co-occurrence Matrix (GLCM)**: GLCM is a method of computing texture features by calculating the relationship between the gray levels of pixels. It was widely employed in medical imaging but it was not capable of dealing with more challenging environments(44)

Although these traditional methods laid the ground work for the first segmentation models, their stiffness and the fact that they could not be updated readily to changing conditions made them inadequate for use in real life scenarios.

4-2 Deep Learning Methods

Deep learning has changed the feature extraction process by providing a way for the models to learn the hierarchical features from the raw data. Convolutional Neural Networks (CNNs) form the basis of these methods, offering efficient and non-data specific approaches to feature extraction.

4-2-1 Fully Convolutional Networks (FCNs)

Fully Convolutional Networks were proposed by Long et al. in 2015 (45)where they eliminated fully connected layers and replaced them with convolutional layers to enable dense prediction at the pixel level. The FCNs obtain multi-scale features through a series of convolutional layers and such feature extraction is critical for current state-of-the-art segmentation models(46),(47)

4-2-2 U-Net Architecture :

U-Net was proposed by (Ronneberger et al., 2015) (48) for biomedical image segmentation. It has an encoder-decoder architecture with skip connections that help in concatenation of lowlevel spatial information from the encoder with high level semantic information from the decoder which enhances the overall segmentation performance

4-2-3 DeepLab Series :

The DeepLab series employed several features in the feature extraction process:

- **Atrous Convolutions (Dilated Convolutions)**: These convolutions increase the receptive field without increasing the number of parameters, enabling the model to capture multi-scale contextual information (49).
- **Atrous Spatial Pyramid Pooling (ASPP)**: In ASPP, the spatial pyramid pooling layer uses atrous convolutions with varied dilation rates for capturing multi-scale information which helps in segmentation of objects of varying sizes (50).

4-2-4 Transformers for Feature Extraction :

Transformers have been proven to be effective for semantic segmentation. ViT and SegFormer have been shown to learn global context and handle long-range relationships and outperform CNN baselines on many tasks (51), (52).

4-2-5 Hybrid Models :

The hybrid models incorporate both CNNs and transformers for a particular task. For instance, Swin Transformers implement a hierarchical structure that learns stacked features as well as global attention, thus producing the state-of-the-art performance on various benchmarks(53).

4-2-6 Pretrained Models and Transfer Learning :

Some of the pre-trained networks such as ResNet, VGG and EfficientNet are frequently applied as feature extractors. These models can be fine-tuned on smaller data sets all while leveraging the knowledge from these models in a process called transfer learning this reduces the training time and the computational cost (54).

CONCLUSION

The concept of image segmentation as one of the key problems of computer vision has developed considerably over time starting from the handcrafted feature-based approaches through to the deep learning-based solutions. The conventional approaches to image segmentation such as edge detection, region-based segmentation, and clustering offered initial knowledge but had a limited capacity to deal with the challenges that come with diverse data sets. Although these methods were computationally light, they lacked the capabilities of providing high accuracy and flexibility in areas of application. The use of deep learning has brought about a change in the field by making the models able to learn features and hierarchical representation of the input data. Some of the architectures such as Fully Convolutional Networks (FCNs), U-Net, and DeepLab has become the new standards and has produced impressive results in various applications including medical imaging, self-driving cars, and environmental surveillance. Furthermore, the usage of transformers and other advanced models including hybrid models has also been used to improve the segmentation results by understanding both local and global contexts. The available benchmarks, Cityscapes, PASCAL ADE20K, VOC, and ISIC have been very helpful in the advancement of the field since they provide reference data for models. However, some issues are still there, such as the efficiency issue, the real-time processing issue, and the domain adaptation issue for dealing with different application contexts. Also, the reliance on large datasets of labeled data is still a major challenge hence there is growing interest in semi-supervised and unsupervised the learning. Field of image Therefore, segmentation, it the can field be remains stated dynamic that, with although the deep current learning research has efforts significantly aimed advanced at reducing computational costs, improving model's robustness and studying the latest model structures. Some of the future prospects for the development of the field are further development of multi-modal data, semi-supervised learning, models. And the lightweight further development of the image segmentation will enable these innovations to find its use in more areas of application across various industries and scientific fields.

REFRENCES:

1. Badue C, Guidolini R, Carneiro RV, Azevedo P, Cardoso VB, Forechi A, et al. Self-driving cars: A survey. Expert Syst Appl. 2021 Mar;165:113816.

2. Esteva A, Chou K, Yeung S, Naik N, Madani A, Mottaghi A, et al. Deep learning-enabled medical computer vision. NPJ Digit Med. 2021 Jan 8;4(1):5.

3. Li X, Ding J, Liu J, Ge X, Zhang J. Digital Mapping of Soil Organic Carbon Using Sentinel Series Data: A Case Study of the Ebinur Lake Watershed in Xinjiang. Remote Sens (Basel). 2021 Feb 19;13(4):769.

4. Su Z, Liu W, Yu Z, Hu D, Liao Q, Tian Q, et al. Pixel Difference Networks for Efficient Edge Detection. 2021 Aug 16;

5. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. p. 3431–40.

6. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In 2015. p. 234–41.

7. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Trans Pattern Anal Mach Intell. 2018 Apr 1;40(4):834–48.

8. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data. 2018 Aug 14;5(1):180161.

9. Dosovitskiy A. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:201011929. 2020;

10. Kora P, Ooi CP, Faust O, Raghavendra U, Gudigar A, Chan WY, et al. Transfer learning techniques for medical image analysis: A review. Biocybern Biomed Eng. 2022 Jan;42(1):79–107.

11. Yao Z, Wang L. Multi-pathway feature integration network for salient object detection. Neurocomputing. 2021 Oct;461:462–78.

12. Wan S, Ding S, Chen C. Edge computing enabled video segmentation for real-time traffic monitoring in internet of vehicles. Pattern Recognit. 2022 Jan;121:108146.

13. Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, Benenson R, et al. The Cityscapes Dataset for Semantic Urban Scene Understanding. 2016 Apr 6;

14. Everingham M, Eslami SMA, Van Gool L, Williams CKI, Winn J, Zisserman A. The Pascal Visual Object Classes Challenge: A Retrospective. Int J Comput Vis. 2015 Jan 25;111(1):98–136.

15. Zhou B, Zhao H, Puig X, Xiao T, Fidler S, Barriuso A, et al. Semantic Understanding of Scenes through the ADE20K Dataset. 2016 Aug 18;

16. Lin TY, Maire M, Belongie S, Bourdev L, Girshick R, Hays J, et al. Microsoft COCO: Common Objects in Context. 2014 May 1;

17. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data. 2018 Aug 14;5(1):180161.

18. Brostow GJ, Fauqueur J, Cipolla R. Semantic object classes in video: A high-definition ground truth database. Pattern Recognit Lett. 2009 Jan;30(2):88–97.

19. Neuhold G, Ollmann T, Rota Bulo S, Kontschieder P. The mapillary vistas dataset for semantic understanding of street scenes. In: Proceedings of the IEEE international conference on computer vision. 2017. p. 4990–9.

20. Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? The KITTI vision benchmark suite. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2012. p. 3354–61.

21. Staal J, Abramoff MD, Niemeijer M, Viergever MA, van Ginneken B. Ridge-Based Vessel Segmentation in Color Images of the Retina. IEEE Trans Med Imaging. 2004 Apr;23(4):501–9.

22. Gong K, Liang X, Zhang D, Shen X, Lin L. Look into Person: Self-supervised Structure-sensitive Learning and A New Benchmark for Human Parsing. 2017 Mar 15;

23. Lee M, Kim M, Jeong CY. Real-time semantic segmentation on edge devices: A performance comparison of segmentation models. In: 2022 13th International Conference on Information and Communication Technology Convergence (ICTC). IEEE; 2022. p. 383–8.

24. Kirillov A, Wu Y, He K, Girshick R. Pointrend: Image segmentation as rendering. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. p. 9799–808.

25. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. 2020 Oct 22;

26. Xie E, Wang W, Yu Z, Anandkumar A, Alvarez JM, Luo P. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. 2021 May 31;

27. Esteva A, Chou K, Yeung S, Naik N, Madani A, Mottaghi A, et al. Deep learning-enabled medical computer vision. NPJ Digit Med. 2021 Jan 8;4(1):5.

28. He X, Zhou Y, Zhao J, Zhang D, Yao R, Xue Y. Swin Transformer Embedding UNet for Remote Sensing Image Semantic Segmentation. IEEE Transactions on Geoscience and Remote Sensing. 2022;60:1–15.

29. Bragagnolo L, Rezende LR, da Silva RV, Grzybowski JMV. Convolutional neural networks applied to semantic segmentation of landslide scars. Catena (Amst). 2021 Jun;201:105189.

30. Ishihara K, Kanervisto A, Miura J, Hautamaki V. Multi-task learning with attention for end-to-end autonomous driving. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021. p. 2902–11.

31. Su D, Kong H, Qiao Y, Sukkarieh S. Data augmentation for deep learning based semantic segmentation and crop-weed classification in agricultural robotics. Comput Electron Agric. 2021 Nov;190:106418.

32. Grill JB, Strub F, Altché F, Tallec C, Richemond PH, Buchatskaya E, et al. Bootstrap your own latent: A new approach to self-supervised Learning. 2020 Jun 13;

33. Mao W, Liu M, Salzmann M. Weakly-supervised action transition learning for stochastic human motion prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. p. 8151–60.

34. Mirikharaji Z, Abhishek K, Bissoto A, Barata C, Avila S, Valle E, et al. A survey on deep learning for skin lesion segmentation. Med Image Anal. 2023;88:102863.

35. Kumar P, Kumar V. Exploring the Frontier of Object Detection: A Deep Dive into YOLOv8 and the COCO Dataset. In: 2023 IEEE International Conference on Computer Vision and Machine Intelligence (CVMI). 2023. p. 1–6.

36. Liu Y, Bai X, Wang J, Li G, Li J, Lv Z. Image semantic segmentation approach based on DeepLabV3 plus network with an attention mechanism. Eng Appl Artif Intell. 2024 Jan;127:107260.

37. Sun Y, Pan B, Fu Y. Lightweight Deep Neural Network for Real-Time Instrument Semantic Segmentation in Robot Assisted Minimally Invasive Surgery. IEEE Robot Autom Lett. 2021 Apr;6(2):3870–7.

38. Niu R, Sun X, Tian Y, Diao W, Chen K, Fu K. Hybrid Multiple Attention Network for Semantic Segmentation in Aerial Images. IEEE Transactions on Geoscience and Remote Sensing. 2022;60:1–18.

39. Subasi A. Artificial intelligence for 3D medical image analysis. In: Applications of Artificial Intelligence in Healthcare and Biomedicine. Elsevier; 2024. p. 357–75.

40. Lai X, Tian Z, Jiang L, Liu S, Zhao H, Wang L, et al. Semi-supervised semantic segmentation with directional context-aware consistency. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021. p. 1205–14.

41. Rahman MdA, Amin MFI, Hamada M. Edge Detection Technique by Histogram Processing with Canny Edge Detector. In: 202020 3rd IEEE International Conference on Knowledge Innovation and Invention (ICKII). IEEE; 2020. p. 128–31.

42. Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE; p. 886–93.

43. Lowe DG. Distinctive Image Features from Scale-Invariant Keypoints. Int J Comput Vis. 2004 Nov;60(2):91–110.

44. Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. IEEE Trans Syst Man Cybern. 1973 Nov;SMC-3(6):610–21.

45. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. p. 3431–40.

46. Abdulrahman Safar A, Salih DM, Murshid AM. Pattern recognition using the multi-layer perceptron (MLP) for medical disease: A survey. Int J Nonlinear Anal Appl [Internet]. 2023;14:2008–6822. Available from: http://dx.doi.org/10.22075/ijnaa.2022.7114

47. Ahmed MS, Fakhrudeen AM. COVID-19IraqKirkukDataset: Development and evaluation of an Iraqi dataset for COVID-19 classification based on deep learning. International Journal of Nonlinear Analysis and Applications [Internet]. 2023;14(1):2507–18. Available from: https://ijnaa.semnan.ac.ir/article_7317.html

48. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In 2015. p. 234–41.

49. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Trans Pattern Anal Mach Intell. 2018 Apr 1;40(4):834–48.

50. Lin T. Focal Loss for Dense Object Detection. arXiv preprint arXiv:170802002. 2017;

51. Dosovitskiy A. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:201011929. 2020;

52. Xie E, Wang W, Yu Z, Anandkumar A, Alvarez JM, Luo P. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. 2021 May 31;

53. Rambhatla SS, Chellappa R, Shrivastava A. The pursuit of knowledge: Discovering and localizing novel categories using dual memory. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021. p. 9153–63.

54. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 770–8.