2023 IEEE 9th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)

2023 IEEE 9th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS) | 979-8-3503-1293-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/BigDataSecurity-HPSC-IDS58521.2023.00013

A Survey of Weakly-supervised Semantic Segmentation

1st Kaiyin Zhu
Department of Computer Science and Engineering
Central South University
ChangSha, China
zhukaiyin@csu.edu.cn

2nd Neal N. Xiong
Department of Computer, Mathematical and Physical Sciences
Sul Ross State University
Alpine, USA
nxiong@cs.gsu.edu

1st Mingming Lu
Department of Computer Science and Engineering
Central South University
ChangSha, China
mingminglu@csu.edu.cn

***Abstract*—Weakly-supervised image semantic segmentation is a popular technology in computer vision and deep learning today. The main goal of weakly-supervised semantic segmentation is to train a model by images with only coarse or sparse annotations. Specifically, it assigns a label to each pixel through coarse label refinement or sparse label propagation, etc. The existing semantic segmentation has a wide range of applications, which includes pedestrian detection, autonomous driving, medical image segmentation, etc. However, fully-supervised semantic segmentation requires pixel-level annotation, which is expensive in manpower and time, and more and more works have focused on weakly-supervised semantic segmentation in recent years. Thus, this paper provides a review of weakly supervised semantic segmentation. Firstly, this paper summarizes the state-of-the-art research results of weakly-supervised semantic segmentation. Secondly, the widely-used datasets and semantic segmentation models are introduced. Finally, this paper analyzes the existing problems and future development directions in the field of weakly-supervised semantic segmentation.**

***Index Terms*—Semantic segmentation, Sparse annotations, Weakly-supervised**

I. Introduction

The main goal of semantic segmentation is to divide an image into multiple parts by semantic information, and each part has its own semantic label. Specifically, the pixels belonging to the same object have the same label due to assigning one label to each pixel.

Semantic segmentation has a wide range of application scenarios. On the one hand, it can be implemented for traffic field, such as pedestrian detection [1], automatic driving [2] and smart city [3], as it can provide vehicles with scene information such as streets and pedestrians through semantic segmentation. On the other hand, Semantic segmentation is also used wildly in the medical field [4]. As Medical image segmentation [5], including abdominal organ segmentation, cardiac internal structure segmentation, etc., provides support for follow-up disease monitoring and other works [6]. All of these applications are based on semantic segmentation technology.

Recently, deep neural networks have been applied to semantic segmentation. With its excellent deep feature mining ability and fitting ability, the fully supervised semantic segmentation models have been greatly improved. The Fully Convolutional Networks(FCN) proposed by Long et al. [7] has a 20% improvement compared to traditional non-deep learning methods. The U-Net [8] proposed by Ronneberger et al. provides a basic encoder-decoder framework for subsequent semantic segmentation models. The DeepLab models [9]–[12] mark that fully supervised semantic segmentation based on deep learning has achieved very good practical results.

The emergence of weakly-supervised semantic segmentation is to solve the relieve of the high annotation cost of fully supervised semantic segmentation. The training of fully supervised semantic segmentation models requires a large number of pixel-level annotations, which is a huge labor and time-consuming process. Despite the unsupervised semantic segmentation can relieve this problem to some extend, unsupervised semantic segmentation remains difficult to achieve good practical results due to the lack of information such as the position and edge of objects. As a compromise between these two methods, weak supervision has interested many researchers, and it also has got great breakthroughs in recent years.

Weakly supervised semantic segmentation is mainly used in scenarios where it is difficult to obtain annotations. For example, in the field of medical image segmentation, data labeling has high requirements for professional knowledge, which greatly increases the cost of data labeling, so many works use methods based on weak supervision [51], [52]. And in the field of remote sensing image segmentation, image data usually has high dimensions, high resolution and complex spatial structure, which increases the difficulty and cost of labeling data. The method of using weak supervision [53], [54] can not only save costs, but also use some prior knowledge to

Corresponding author: zhukaiyin@csu.edu.cn

979-8-3503-1293-5/23/$31.00 ©2023 IEEE

DOI 10.1109/BigDataSecurity-HPSC-IDS58521.2023.00013

improve the segmentation effect.

To better understand weakly supervised semantic segmentation, this study first introduces the commonly used datasets in the experiment of weakly supervised semantic segmentation, to provide some comparative standards for the effect evaluation of different works. Then we briefly introduces the classic fully supervised semantic segmentation model, since many weakly supervised methods are improvements on the classic fully supervised method. Next we sorted out four types of weakly supervised semantic segmentation models from the perspective of annotation methods, namely image-level label-based methods, point label-based methods, scribble label-based methods and bounding-box label-based methods. They are arranged in ascending order of annotation cost. Finally, we analyze the existing problems in the field of weakly supervised semantic segmentation and propose feasible solutions.

The remainder of this article is organized. Section 2 introduces some commonly used datasets for weakly-supervised semantic segmentation. Section 3 introduces the classic fully supervised and state-of-the-art weakly supervised semantic segmentation models in the order of full, image-level, point, scribble, and box-bounding annotations. Finally in section 4, this paper points out the existing problems in weakly-supervised segmentation and analyzes the future-research directions from the perspective of solutions.

II. Datasets for semantic segmentation

This section summarizes some common semantic segmentation datasets.

Pascal VOC 2012 [13] is the most commonly used dataset for semantic segmentation currently. It is also commonly used as an evaluation criterion for models in this field. Since most of works related to semantic segmentation use Pascal VOC 2012 as the benchmark dataset, the experimental results presented in this paper are also based on this dataset. The dataset contains 20 categories with the themes of people, animals, traffic, and indoor scene. The training set contains 2913 images , which contain 6929 objects. The test set contains 1452 images.

Pascal Context [14] is an extension of the Pascal VOC 2010 dataset.It annotated the entire training set in Pascal VOC 2010 at the pixel level, which includes 540 categories of annotations and 10103 images. Semantic boundaries dataset (SBD) [15] is an extension to Pascal VOC 2011, which annotates all unlabeled images. and contains 11355 fully annotated images as training set. Due to its huge data volume, SBD is currently replacing Pascal VOC as the most commonly used dataset in the field of semantic segmentation.

ScribbleSup [16] is a scribble annotation on the Pascal VOC dataset made by Lin 's team. It is widely used in scribble-based weakly-supervised semantic segmentation training.

Cityscapes [17] is a large-scale dataset containing 19 categories of road traffic scenes. It contains 3,475 finely annotated images and more than 20,000 coarsely annotated images. KITTI [18] contains pictures of traffic scenes captured by various types of sensors. Many subsequent works have annotated some of the images in this dataset.

Microsoft COCO [19] is a large-scale image recognition dataset, containing more than 80 categories. 82,783 training images, 40,504 validation images, and more than 80,000 test images are provided. In the field of semantic segmentation, it is generally used to pre-train models.

III. Semantic segmentation models

This section first introduces the fully supervised semantic segmentation models. Although the focus of this paper is on weak supervision, most weakly-supervised models generate pseudo-labels through some technical means, and then use the classic fully supervised segmentation model for training. Then the state-of-the-art weakly-supervised segmentation models based on image-level, point, scribble and bounding-box annotations are summarized in ascending order of annotation cost.

A. Fully supervised semantic segmentation models

The boom of fully supervised semantic segmentation in the field of deep learning started with the proposal of FCN [7] and U-Net [8] models in 2015. To cope with the limitations of convolutional networks in restoring image details, Long et al. proposed a Fully Convolutional Networks(FCN) [7] . The main idea is to convert the fully connected layer into convolutional layers, and perform upsampling through deconvolution to obtain features with the same size as the input image. This network structure has become the basic model for subsequent semantic segmentation [20]. The U-Net model [8] proposed by Ronneberger et al. consists of a contraction path for extracting image information and an expansion path for precise positioning, which is the encoder-decoder structure used in many subsequent works [21], [22].

TABLE I
Comparison of state-of-the-art fully supervised semantic segmentation models on Pascal VOC 2012 test dataset. In the Sup. column, "F" stands for fully annotation-based models. The MIoU column is the performance of models on Pascal VOC 2012 test dataset by default.

Method	*Sup.*	*Year*	*MIoU*
FCN [7]	F	2015	62.2%
U-Net [8]	F	2015	-
DeepLabv1 [9]	F	2016	71.6%
DeepLabv2 [10]	F	2017	79.7%
DeepLabv3 [11]	F	2017	86.9%
DeepLabv3+ [12]	F	2018	89.0%
SegNeXt [23]	F	2022	90.6%

The DeepLab series has become the most widely used semantic segmentation model due to its excellent performance. DeepLab v1 [9] is actually an FCN model. The breakthrough is that it adds a conditional random field at the end of the model to make the edge of the classification result more refined. DeepLabv2 [10] and Deeplabv3 [11] propose Atrous Spacial Pyramid Pooling(ASPP) and add several layers of atrous convolution on the basis of v1, which makes the segmentation more robust by controlling the sampling rate and the field of

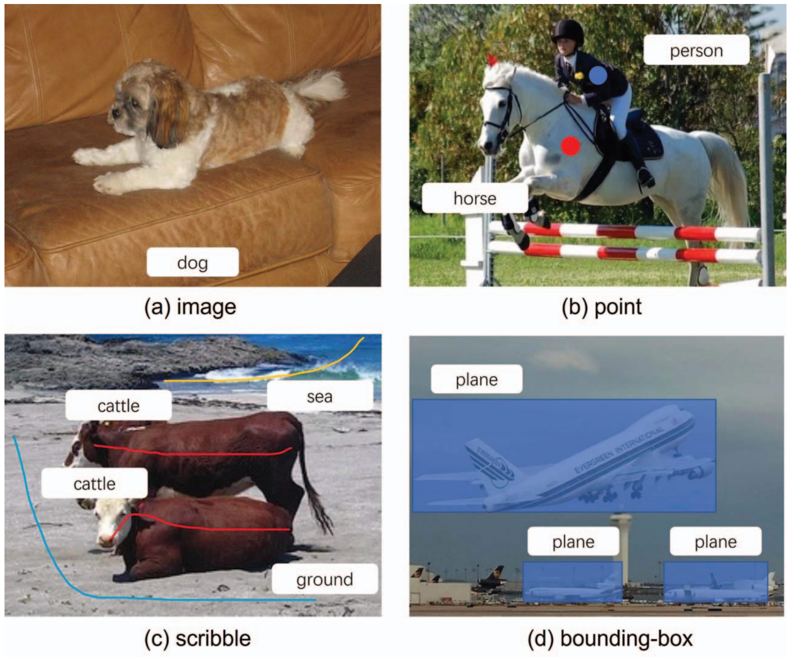

Fig. 1. Four Weakly-supervised Labeling Methods.

view of the filter. DeepLabv3+ [12] combines the ASPP module with the encoder-decoder structure. TABLE I compares classic fully supervised semantic segmentation models with current state-of-the-art models.

B. Weakly-supervised semantic segmentation models

Weakly supervised semantic segmentation is divided into the following four categories in terms of labeling complexity, as shown in Fig. 1.

The first type is image-level labels. According to the number of objects in the image, one or more labels should be given. It has the lowest labeling cost, but does not give the boundaries or locations of objects. The second type of point is to mark a point on the subject of each object in the picture, which is a very sparse annotation. The third type of scribble draws a scribble on the main body of each object in the image to represent the location and approximate area of the object. The fourth type of bounding-box marks a rectangular box for each object to indicate the boundary and position of the object. This labeling method has a slightly higher overhead, but it gives edge information and eliminates a lot of background interference.

This section will introduce the research results of weakly-supervised semantic segmentation in recent years from these four directions.

1) Segmentation algorithm based on image-level labels: The mainstream method of weak supervision for image labeling is to first use the classic classification model to generate image heatmaps, class activation maps, etc., as the initial coarse labels for semantic segmentation, and then gradually refine the labels through various methods to obtain the image's location and edge information.

The Class Activation Map (CAM) [24] proposed by Zhou et al. provides a technical means for generating coarse labels for many subsequent works [49] [50]. It locates the image regions that can be used for distinction by projecting the output weights obtained by global draw pooling in the convolutional neural network onto the feature map, and obtains the class activation map. This positioning technique is also widely used in subsequent weak supervision segmentation. The IRNet proposed by Ahn,Jiwoon et al. [25] learns the center bias and boundary of objects based on CAM, and obtains instance regions according to the random walk algorithm. Yao et al. [26] introduced a self-attention mechanism to generate pseudo-labels.

Puzzle-CAM [27] proposed by Sanghyun Jo et al. improves the CAM mechanism so that it no longer only focuses on discriminative areas, but on the most integrated areas in the object, shifting the focus of CAM from classification problems to segmentation problems. The RecurSeed [28] they proposed subsequently reduces the false-detection phenomenon of small-scale objects and non-detection of large-scale objects through recursive iteration. Many image-level label-based segmentation methods are based on CAM to generate pseudo-labels. The CAM method is proposed based on classification tasks and cannot be well applied to segmentation tasks. In order to solve this problem, Bircanoglu et al. [29] proposed a framework that employs an iterative approach in a modified encoder-decoder-based segmentation model, which simultaneously supports classification and segmentation tasks. The results of the above methods will be compared in II.

TABLE II
COMPARISON OF STATE-OF-THE-ART IMAGE-LEVEL WEAKLY-SUPERVISED SEMANTIC SEGMENTATION MODELS ON PASCAL VOC 2012 TEST DATASET. IN THE SUP. COLUMN, "I" STANDS FOR IMAGE-LEVEL ANNOTATION-BASED MODELS.

Method	*Sup.*	*Year*	*MIoU*
CAM [24]	I	2015	-
IRNet [25]	I	2019	64.8%
SGAN [26]	I	2019	67.2%
Puzzle-CAM [27]	I	2021	72.2%
L2G [31]	I	2022	73%
PPC [30]	I	2022	73.5%
RS+EPM [28]	I	2022	73.6%
ISIM [29]	I	2022	74.98%

2) Segmentation algorithm based on point: Point annotation is inspired by the way of locating objects in human life. Humans are accustomed to pointing a point with a finger to indicate the approximate location of an object. Point-based research idea has attracted wide attention in many fields [33], [34], but it has not been well studied in semantic segmentation. Amy Bearman et al. [35] used point supervision for semantic segmentation for the first time and achieved 12.9% MIoU higher than the current image-level segmentation. R. Austin McEver et al. [36] proposed a method to generate pseudo-labels using a combination of class activation maps and label propagation.

3) Segmentation algorithm based on scribble: The mainstream method used in scribble annotations is to propagate sparse pixel annotations through technical means such as conditional random fields to generate pseudo-annotations. Inspired by interactive commercial software, Lin et al. [16] used scribble annotation in weakly-supervised semantic segmentation for the first time. They made scribble annotation for several datasets of Pascal VOC. Tang et al. [37] proposed a normalized cut loss to optimize the training process. Ke et al. [39] used the metric learning method to shorten the distance between the same semantic pixels and increase the distance between different semantic pixels, to map each pixel of the image into the feature space containing semantics. This method can achieve better results on datasets with any annotation form.

Liang et al. [40] proposed a tree energy loss, which uses a minimum spanning tree algorithm to obtain semantic information between pixels so that single-stage training can achieve better results than previous work. Zhang et al. [42] proposed a new regularization loss and used a vision transformer as the backbone network to extract feature relationships at different levels. In the research work in the field of semantic segmentation in recent years, the transformer model has received more and more attention [45].

4) Segmentation algorithm based on bounding-box: The bounding-box annotation uses the bounding-box label as the coarse label of the object to train the model, and gradually optimizes the edges.

TABLE III
COMPARISON OF STATE-OF-THE-ART POINT LEVEL, SCRIBBLE LEVEL, BOUNDING-BOX LEVEL WEAKLY-SUPERVISED SEMANTIC SEGMENTATION MODELS ON PASCAL VOC 2012 TEST DATASET.IN THE SUP. COLUMN, "P" STANDS FOR POINT ANNOTATION-BASED MODELS, "S" FOR SCRIBBLE AND "B" FOR BOUNDING-BOX. THE MIOU COLUMN IS THE PERFORMANCE OF MODELS ON PASCAL VOC 2012 TEST DATASET BY DEFAULT, AND THE (VAL) TAG INDICATES THAT THE RESULTS ARE BASED ON PASCAL VOC 2012 VAL DATASET.

Method	*Sup.*	*Year*	*MIoU*
What's the Point [35]	P	2016	46.1%
PCAM [36]	P	2020	70.5%
SPML [39]	P	2021	73.2%
TEL [40]	P	2022	74.2%
ScribbleSup [16]	S	2016	63.1%(val)
Normalized cut [37]	S	2018	74.5%
MBI [38]	S	2021	75.1%(val)
BPG [41]	S	2019	76.0%
SPML [39]	S	2021	76.1%
TEL [40]	S	2022	77.3%
GDC [32]	S	2021	80.34%(val)
DFR [42]	S	2021	82.9%
BCM [46]	B	2019	70.2%(val)
Box2Seg [43]	B	2020	76.4%(val)
SPML [39]	B	2021	74.7%
A2GNN [44]	B	2022	75.2%

Song et al. [46] proposed a box-driven class-wise masking model(BCM) to remove regions within the bounding box that are irrelevant to objects, and trained with an adaptive loss guided by the fill rate within the bounding box for each class. Lee et al. [47] employed an object detector to identify subjects within the bounding box, and used the results as pseudo-labels for fully supervised training.

Kim et al. [48] proposed background-aware pooling to separate foreground and background information in bounding boxes, and used the noise-aware loss to remove noise in the results. Zhang et al. [44] proposed an affinity attention graph neural network, which first generates a confidence seed, calculates the semantic relationship between pixels through the affinity in the graph network, propagates the seed to unlabeled pixels, and uses the bounding-box annotation as a subsequent constraint. Methods based on point, scribble and bounding-box annotations will be given in III.

IV. DISCUSSION

Through the summary of works in the field of weakly supervised semantic segmentation, we have the following findings: (1) The fully supervised semantic segmentation method has not made much breakthrough in recent years since DeepLabv3+ was proposed in 2018, and the experimental results were raised to 89% MIoU. The possible reason is that the pixel-level labeling is very expensive, so the amount of data in the dataset has reached a bottleneck, making it difficult to significantly improve the results from the perspective of the model. (2) Although there is still a gap in performance between semantic segmentation methods based on weakly supervised annotation and methods based on full supervision, weakly supervised methods have developed rapidly in recent years. Among them, image-level annotation is the most mainstream research field, and the performance of the ISIM method proposed in 2022 has reached 74.98% MIoU. (3) The scribble-based weakly supervised segmentation method performs well considering the labeling cost and performance. On the one hand, it has little annotation overhead; on the other hand, in the DFR methods proposed in 2022, the scribble-based method has already achieved 82.9% MIoU. This result is relatively close to the effect of full supervision.

To sum up, great progress has been made in the field of weakly-supervised semantic segmentation in recent years, but there are still some problems.

First, the limitations of dataset scenarios. Due to the huge annotation cost of semantic segmentation, most of the datasets are based on mainstream research fields such as road traffic and medical imaging, lacking rich practical scenarios, and the application fields and application effects are also limited. Second, the size of the model is too large. Many research results use multi-stage training and design a large-scale deep learning model, which requires high equipment and computing power. Third, the training and inference speed is slow. It is difficult to meet the real-time requirements in practical applications.

In response to the above problems, researchers can make an effort to the following perspectives. Firstly, for the dataset scene limitations, we can utilize the domain adaptation to extend the semantic segmentation models to more application scenarios. Secondly, for the model scale and inference speed, we can consider how to improve the loss function, and take single-stage and lightweight models.

V. Conclusion

This paper summarizes and analyzes the most advanced models in the field of weakly supervised semantic segmentation in the past five years, considering both the annotation cost and performance. In order to deepen the understanding of weakly supervised segmentation work, we also introduce commonly used datasets in this field and classic fully supervised semantic segmentation models. Finally, by summarizing and analyzing the research literature, we put forward our opinions on the problems and development directions in the field of weakly supervised semantic segmentation.

Acknowledgment

This work was partially supported by the National Natural Science Foundation of China under Grant No. U20A20182 and 62177019.

References

[1] Chunxue Wu, Bobo Ju, Yan Wu, Xiao Lin, Naixue Xiong, Guangquan Xu, Hongyan Li, and Xuefeng Liang. Uav autonomous target search based on deep reinforcement learning in complex disaster scene. *IEEE Access*, 7:117227–117245, 2019.

[2] M YANC. Review on semantic segmentation of road scenes. *Laser & Optoelectronics Progress*, 58(12):36–58, 2021.

[3] Prabhat Kumar, Randhir Kumar, Gautam Srivastava, Govind P Gupta, Rakesh Tripathi, Thippa Reddy Gadekallu, and Neal N Xiong. Ppsf: a privacy-preserving and secure framework using blockchain-based machine-learning for iot-driven smart cities. *IEEE Transactions on Network Science and Engineering*, 8(3):2326–2341, 2021.

[4] Yongbin Gao, Xuehao Xiang, Naixue Xiong, Bo Huang, Hyo Jong Lee, Rad Alrifai, Xiaoyan Jiang, and Zhijun Fang. Human action monitoring for healthcare based on deep learning. *Ieee Access*, 6:52277–52285, 2018.

[5] Mohammad Hesam Hesamian, Wenjing Jia, Xiangjian He, and Paul Kennedy. Deep learning techniques for medical image segmentation: achievements and challenges. *Journal of digital imaging*, 32(4):582–596, 2019.

[6] Chunxue Wu, Chong Luo, Naixue Xiong, Wei Zhang, and Tai-Hoon Kim. A greedy deep learning method for medical disease analysis. *IEEE Access*, 6:20021–20030, 2018.

[7] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[9] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*, 2014.

[10] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.

[11] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.

[12] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.

[13] Mark Everingham and John Winn. The pascal visual object classes challenge 2012 (voc2012) development kit. *Pattern Anal. Stat. Model. Comput. Learn., Tech. Rep*, 2007:1–45, 2012.

[14] Roozbeh Mottaghi, Xianjie Chen, Xiaobai Liu, Nam-Gyu Cho, Seong-Whan Lee, Sanja Fidler, Raquel Urtasun, and Alan Yuille. The role of context for object detection and semantic segmentation in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 891–898, 2014.

[15] Alberto Garcia-Garcia, Sergio Orts-Escolano, Sergiu Oprea, Victor Villena, and Jose Garcia-Rodriguez. A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv:1704.06857*, 2017.

[16] Di Lin, Jifeng Dai, Jiaya Jia, Kaiming He, and Jian Sun. Scribblesup: Scribble-supervised convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3159–3167, 2016.

[17] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016.

[18] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013.

[19] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

[20] Feng Xia, Ruonan Hao, Jie Li, Naixue Xiong, Laurence T Yang, and Yan Zhang. Adaptive gts allocation in ieee 802.15. 4 for real-time wireless sensor networks. *Journal of Systems Architecture*, 59(10):1231–1242, 2013.

[21] Hongju Cheng, Zhe Xie, Yushi Shi, and Naixue Xiong. Multi-step data prediction in wireless sensor networks based on one-dimensional cnn and bidirectional lstm. *IEEE Access*, 7:117883–117896, 2019.
[22] Yonglei Yao, Naixue Xiong, Jong Hyuk Park, Li Ma, and Jingfa Liu. Privacy-preserving max/min query in two-tiered wireless sensor networks. *Computers & Mathematics with Applications*, 65(9):1318–1325, 2013.
[23] Meng-Hao Guo, Cheng-Ze Lu, Qibin Hou, Zhengning Liu, Ming-Ming Cheng, and Shi-Min Hu. Segnext: Rethinking convolutional attention design for semantic segmentation. *arXiv preprint arXiv:2209.08575*, 2022.
[24] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2921–2929, 2016.
[25] Jiwoon Ahn, Sunghyun Cho, and Suha Kwak. Weakly supervised learning of instance segmentation with inter-pixel relations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2209–2218, 2019.
[26] Qi Yao and Xiaojin Gong. Saliency guided self-attention network for weakly and semi-supervised semantic segmentation. *IEEE Access*, 8:14413–14423, 2020.
[27] Sanghyun Jo and In-Jae Yu. Puzzle-cam: Improved localization via matching partial and full features. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 639–643. IEEE, 2021.
[28] Sang Hyun Jo, In Jae Yu, and Kyung-Su Kim. Recurseed and certainmix for weakly supervised semantic segmentation. *arXiv preprint arXiv:2204.06754*, 2022.
[29] Cenk Bircanoglu and Nafiz Arica. Isim: Iterative self-improved model for weakly supervised segmentation. *arXiv preprint arXiv:2211.12455*, 2022.
[30] Ye Du, Zehua Fu, Qingjie Liu, and Yunhong Wang. Weakly supervised semantic segmentation by pixel-to-prototype contrast. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4320–4329, 2022.
[31] Peng-Tao Jiang, Yuqi Yang, Qibin Hou, and Yunchao Wei. L2g: A simple local-to-global knowledge transfer framework for weakly supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16886–16896, 2022.
[32] Xin Sun, Changrui Chen, Xiaorui Wang, Junyu Dong, Huiyu Zhou, and Sheng Chen. Gaussian dynamic convolution for efficient single-image segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(5):2937–2948, 2021.
[33] Anmin Fu, Xianglong Zhang, Naixue Xiong, Yansong Gao, Huaqun Wang, and Jing Zhang. Vfl: a verifiable federated learning with privacy-preserving for big data in industrial iot. *IEEE Transactions on Industrial Informatics*, 2020.
[34] Jin Zhao, Jifeng Huang, and Naixue Xiong. An effective exponential-based trust and reputation evaluation system in wireless sensor networks. *IEEE Access*, 7:33859–33869, 2019.
[35] Amy Bearman, Olga Russakovsky, Vittorio Ferrari, and Li Fei-Fei. What's the point: Semantic segmentation with point supervision. In *European conference on computer vision*, pages 549–565. Springer, 2016.
[36] R Austin McEver and BS Manjunath. Pcams: Weakly supervised semantic segmentation using point supervision. *arXiv preprint arXiv:2007.05615*, 2020.
[37] Meng Tang, Abdelaziz Djelouah, Federico Perazzi, Yuri Boykov, and Christopher Schroers. Normalized cut loss for weakly-supervised cnn segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1818–1827, 2018.
[38] Fabio Cermelli, Massimiliano Mancini, Samuel Rota Bulo, Elisa Ricci, and Barbara Caputo. Modeling the background for incremental and weakly-supervised semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):10099–10113, 2021.
[39] Tsung-Wei Ke, Jyh-Jing Hwang, and Stella X Yu. Universal weakly supervised segmentation by pixel-to-segment contrastive learning. *arXiv preprint arXiv:2105.00957*, 2021.
[40] Zhiyuan Liang, Tiancai Wang, Xiangyu Zhang, Jian Sun, and Jianbing Shen. Tree energy loss: Towards sparsely annotated semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16907–16916, 2022.
[41] Bin Wang, Guojun Qi, Sheng Tang, Tianzhu Zhang, Yunchao Wei, Linghui Li, and Yongdong Zhang. Boundary perception guidance: A scribble-supervised semantic segmentation approach. In *IJCAI International joint conference on artificial intelligence*, 2019.
[42] Bingfeng Zhang, Jimin Xiao, and Yao Zhao. Dynamic feature regularized loss for weakly supervised semantic segmentation. *arXiv preprint arXiv:2108.01296*, 2021.
[43] Viveka Kulharia, Siddhartha Chandra, Amit Agrawal, Philip Torr, and Ambrish Tyagi. Box2seg: Attention weighted loss and discriminative feature learning for weakly supervised segmentation. In *European Conference on Computer Vision*, pages 290–308. Springer, 2020.
[44] Bingfeng Zhang, Jimin Xiao, Jianbo Jiao, Yunchao Wei, and Yao Zhao. Affinity attention graph neural network for weakly supervised semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
[45] Wei Zhang, Shiwei Zhu, Jian Tang, and Naixue Xiong. A novel trust management scheme based on dempster–shafer evidence theory for malicious nodes detection in wireless sensor networks. *The Journal of Supercomputing*, 74(4):1779–1801, 2018.
[46] Chunfeng Song, Yan Huang, Wanli Ouyang, and Liang Wang. Box-driven class-wise region masking and filling rate guided loss for weakly supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3136–3145, 2019.
[47] Jungbeom Lee, Jihun Yi, Chaehun Shin, and Sungroh Yoon. Bbam: Bounding box attribution map for weakly supervised semantic and instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2643–2652, 2021.
[48] Youngmin Oh, Beomjun Kim, and Bumsub Ham. Background-aware pooling and noise-aware loss for weakly-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6913–6922, 2021.
[49] Yewang Chen, Lida Zhou, Songwen Pei, Zhiwen Yu, Yi Chen, Xin Liu, Jixiang Du, and Naixue Xiong. Knn-block dbscan: Fast clustering for large-scale data. *IEEE transactions on systems, man, and cybernetics: systems*, 51(6):3939–3953, 2019.
[50] Shaobo Huang, Zhiwen Zeng, Kaoru Ota, Mianxiong Dong, Tian Wang, and Neal N Xiong. An intelligent collaboration trust interconnections system for mobile information control in ubiquitous 5g networks. *IEEE transactions on network science and engineering*, 8(1):347–365, 2020.
[51] Xiangde Luo, Minhao Hu, Wenjun Liao, Shuwei Zhai, Tao Song, Guotai Wang, and Shaoting Zhang. Scribble-supervised medical image segmentation via dual-branch network and dynamically mixed pseudo labels supervision. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part I*, pages 528–538. Springer, 2022.
[52] Zhang Chen, Zhiqiang Tian, Jihua Zhu, Ce Li, and Shaoyi Du. C-cam: Causal cam for weakly supervised semantic segmentation on medical image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11676–11685, 2022.
[53] Yongxiu Zhou, Honghui Wang, Ronghao Yang, Guangle Yao, Qiang Xu, and Xiaojuan Zhang. A novel weakly supervised remote sensing landslide semantic segmentation method: Combining cam and cyclegan algorithms. *Remote Sensing*, 14(15):3650, 2022.
[54] Yinxia Cao and Xin Huang. A coarse-to-fine weakly supervised learning method for green plastic cover segmentation using high-resolution remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 188:157–176, 2022.