Semantic Image Segmentation with Scantly Annotated Data

Subjects: Computer Science, Artificial Intelligence

Contributor: Adithi D. Chakravarthy , Dilanga Abeyrathna , Mahadevan Subramaniam , Parvathi Chundi , Venkataramana Gadhamshetty

Semantic image segmentation is the task of assigning to each pixel the class of its enclosing object or region as its label, thereby creating a segmentation mask. The success of deep networks for the semantic segmentation of images is limited by the availability of annotated training data. The manual annotation of images for segmentation is a tedious and time-consuming task that often requires sophisticated users with significant domain expertise to create high-quality annotations over hundreds of images.

image processing   image segmentation   machine vision   neural networks   semi-supervised learning

1. Introduction

Semantic image segmentation is the task of assigning to each pixel the class of its enclosing object or region as its label, thereby creating a segmentation mask. Due to its wide applicability, this task has received extensive attention from experts in several areas, such as autonomous driving, robot navigation, scene understanding, and medical imaging. Owing to its huge success, deep learning has become the de-facto choice for semantic image segmentation. Recent approaches have used convolutional neural networks (CNNs) [1][2] and fully convolutional networks (FCNs) [3][4][5] for this task and achieved promising results. Several recent surveys [6][7][8][9][10][11] describe the successes of semantic image segmentation and directions for future research.

Typically, large volumes of labeled data are needed to train deep CNNs for image analysis tasks, such as classification, object detection, and semantic image segmentation. This is especially so for semantic image segmentation, where each pixel in each training image has to be labeled or annotated in order to infer the labels of the individual pixels of a given test image. The availability of densely annotated images in sufficient numbers is problematic, particularly in domains such as material science, engineering, and medicine, where annotating images is time consuming and requires significant user expertise. For instance, while reading retinal images to identify unhealthy areas, it is common for graders (with ophthalmology training) to discuss each image at length to carefully resolve several confounding and subtle image attributes [12][13][14]. Labeling cells, cell clusters, and microbial byproducts in biofilms take up to two days per image on average [15][16][17]. Therefore, it is highly beneficial to develop high-performance deep segmentation networks that can train with scantly annotated training data.

2. Semantic Segmentation with Scantly Annotated Data

Semantic segmentation [4] is one of the challenging image analyses tasks that has been studied earlier using image processing algorithms and more recently using deep learning networks; see [6][10][11][18] for detailed surveys. Several image processing algorithms based on methods including clustering, texture and color filtering, normalized cuts, superpixels, graph and edge-based region merging, have been developed to perform segmentation by grouping similar pixels and partitioning a given image into visually distinguishable regions [6]. More recent supervised segmentation approaches based on [4] use fully connected networks (FCNs) to output spatial maps instead of classification scores by replacing the fully connected layers with convolutional layers. These spatial maps are then up-sampled using deconvolutions to generate pixel-level label outputs. Other decoder variants to transform a classification network to a segmentation network include the SegNet [19] and the U-Net [20].

Currently, deep learning-based approaches are perhaps the de facto choice for semantic segmentation. Recently, Sehar and Naseem [11], reviewed most of the popular learning algorithms (~120) for semantic segmentation tasks, and concluded the overwhelming success of deep learning compared to the classical learning algorithms. However, as pointed out by the authors, the need for large volumes of training data is a well-known problem in developing segmentation models using deep networks. Two main directions that were explored earlier for addressing this problem are the use of limited dense annotations (scant annotations) and the use of noisy image-level annotations (weakly supervised annotations). Active learning and semi-supervised learning are two popular methods in developing segmentation models using scant annotations and are described below.

2.1. Active Learning for Segmentation

In the iterative active learning approach, a limited number of unlabeled images are selected in each iteration for annotation by experts. The annotated images are merged with training data and used to develop the next segmentation model, and the process continues until the model performance plateaus on a given validation set. Active learning approaches can be broadly categorized based on the criteria used to select images for annotation and the unit (images, patches, and pixels) of annotation. For instance, in [21], FCNs are used to identify uncertain images as candidates, and similar candidates are pruned leaving the rest for annotation. In [22], the drop-out method from [23] is used to identify candidates and then discriminatory features of the latent space of the segmentation network are used to obtain a diverse sample. In [24], active learning is modeled as an optimization problem maximizing Fisher information (a sample has higher Fisher information if it generates larger gradients with respect to the model parameters) over samples. In [25], sample selection is modeled as a Boolean knapsack problem, where the objective is to select a sample that maximizes uncertainty while keeping annotation costs below a threshold. The approach in [21] uses 50% of the training data from the MICCAI Gland challenge (85 training, 80 test) and lymph node (37 training, 37 test) datasets; [22] uses 27% of the training data from MR images dataset (25 training, 11 test); [24] uses around 1% of the training data from an MR dataset with 51 images; and [25] uses 50% of the training data from 1,247 CT scans (934 training, 313 test) and 20% annotation cost. Each of these works produces a model with the same performance as those obtained by using the entire training data.

The unit of annotation for most active learning approaches used for segmentation is the whole image. Though the approach in [25] chooses samples with least annotation cost, it requires experts to annotate the whole image. An exception to these are [24][26][27], where 2D patches are used as the unit of annotation. While active learning using pixel-level annotations (as used by SSPA approach) is rare, some recent works show how pixel-level annotations can be cost effective and produce high-performing segmentation models [28]. Pixel-level annotations require experts to be directed to the target pixels along with the surrounding context, and such support is provided by software prototypes, including those such as the PIXELPICK described in [28]. There are several domain-specific auto-annotators exist for medical images and authors have also developed a domain-specific auto-annotator for biofilms that will be released soon to that community.

2.2. Semi-Supervised Segmentation with Pseudo-Labels

Semi-supervised segmentation approaches usually augment manually labeled training data by generating pseudo-labels for the unlabeled data and using these to generate segmentation models. As an exception, the approach in [29] uses K-means along with graph cuts to generate pseudo-labels and use these to train a segmentation model, which is then used to produce refined pseudo-labels, and the process is repeated until the model performance converges. Such approaches do not use any labeled data for training. A more typical approach in [30] first generates a segmentation model by training on a set of scant expert annotations, and the model is then used to assign pseudo-labels to unlabeled training data. The final model is obtained by training it on the expert-labeled data along with pseudo-labeled data until the performance converges. For a more comprehensive discussion on semi-supervised approaches, please see [10][18].

2.3. The SSPA Approach

The segmentation with scant pixel annotations (SSPA) approach seamlessly integrates active learning and semi-supervised learning approaches with pseudo-labels to produce high-performing segmentation models with cost-effective expert annotations. Similar to the semi-supervised approach in [29], the SSPA does not require any expert annotation to produce the base model. It uses an image processing algorithm based on the watershed transform [31] to generate pseudo-labels. The base model generated using these pseudo-labels is then successively refined using active learning. However, unlike the prior active learning approaches used for segmentation, it employs image entropy instead of image similarity to select top-k high entropy or low entropy images for expert annotation. Further, unlike most of the earlier active learning approaches for segmentation (with the exception of [28]), the unit of annotation is a pixel, targeting uncertain pixels only while other pixels are labeled based on the behavior learned by the models.

In the *SSPA* approach, expert annotations are obtained on demand only for the training samples identified in each active learning step. Further, the unit of annotation is a pixel, and the process is terminated when the model performance plateaus or no further refinements are possibly similar to [29]. The *SSPA* approach outperforms state-of-the-art results in multiple datasets including those used in [32].

The SSPA uses the watershed algorithm to generate pseudo-segmentation masks. This algorithm [31][33][34][35] treats an image as a topographic surface with its pixel intensities capturing the height of the surface at each point in the image. The image is partitioned into basins and watershed lines by flooding the surface from minima. The watershed lines are drawn to prevent the merging of water from different sources. The variant of watershed algorithm used, the marker-controlled watershed algorithm

(MC-WS) [36], automatically determines the regional minima and achieves better performance than the regular one. MC-WS uses morphological operations [37] and distance transforms [38] of binarized images to identify object markers that are used as regional minima.

In Petit et al. [39], the authors proposed a ConvNets-based strategy to perform segmentation on medical images. They attempted to reduce the annotation effort by using a partial set of noisy labels such as scribbles, bounding boxes, etc. Their approach extracts and eliminates ambiguous pixel labels to avoid the error propagation due to these incorrect and noisy labels. Their architecture consists of two stages. In the first stage, ambiguity maps are produced by using K FCNs that perform binary classification for each of the K classes. Each classifier is given the input of pixels only true positive and true negative to the given class and the rest are ignored. In the second stage, the model trained at the first stage is used to predict labels for missing classes, using a curriculum strategy [40]. The authors stated that only 30% of training data surpassed the baseline trained with complete ground-truth annotations. Even though this approach allows recovering the scores obtained without incorrect/incomplete labels, it relies on the use of a perfectly labeled sub-dataset (100% clean labels). This approach was further extended to an approach called INERRANT [41] to achieve better confidence estimation for the initial pseudo-label generation, by assigning a dedicated confidence network to maximize the number of correct labels collected during the pseudo-labeling stage.

Pan et al. [42] proposed a label-efficient hybrid supervised framework for medical image segmentation, where the annotation effort is reduced by mixing a large quantity of weakly annotated labels with a handful of strongly annotated data. Mainly two techniques, namely dynamic instance indicator (DII) and dynamic co-regularization (DCR), are used to extract the semantic clues while reducing the error propagation due to strongly annotated labels. Specifically, DII adjusts the weights for weakly annotated instances based on the gradient directions available in strongly annotated instances, and DCR handles the collaborative training and consistency regularization. The authors stated that the proposed framework shows competitive performance only with 10% of strongly annotated labels, compared to the 100% strongly supervised baseline model.

Zhou et al. [43] recently proposed a watershed transform-based iterative weakly supervised approach for segmentation. This approach first generates weak segmentation annotations through image-level class activation maps, which are then refined by watershed segmentation. Using these weak annotations, a fully supervised model is trained iteratively. However, this approach carries many downsides, such as no control over initial segmentation error propagation in the iterative training, requires many manual parameterization during weak annotation generation, and lack of grasping fuzzy, low-contrast and complex boundaries of the objects [44][45]. Segmentation error propagation through iterations can adversely impact model performance, especially in areas requiring sophisticated domain expertise. In such cases, it may be best to seek expert help in generating segmentation ground truth to manage boundary complexities of the objects and mitigate the error propagation of weakly supervision.

References

1. Farabet, C.; Couprie, C.; Najman, L.; Lecun, Y. Learning hierarchical features for scene labeling. IEEE Trans. Pattern Anal. Mach. Intell. 2013, 35, 1915–1929.

2. Hariharan, B.; Arbeláez, P.; Girshick, R.; Malik, J. Simultaneous detection and segmentation. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 297–312.

3. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans. Pattern Anal. Mach. Intell. 2017, 40, 834–848.

4. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.

5. Dai, J.; He, K.; Sun, J. Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1635–1643.

6. Zhu, H.; Meng, F.; Cai, J.; Lu, S. Beyond pixels: A comprehensive survey from bottom-up to semantic image segmentation and cosegmentation. J. Vis. Commun. Image Represent. 2016, 34, 12–27.

7. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Martinez-Gonzalez, P.; Garcia-Rodriguez, J. A Survey on Deep Learning Techniques for Image and Video Semantic Segmentation. Appl. Soft Comput. 2018, 70, 1568–4946.

8. Zhao, B.; Feng, J.; Wu, X.; Yan, S. A survey on deep learning-based fine-grained object classification and semantic segmentation. Int. J. Autom. Comput. 2017, 14, 119–135.

9. Thoma, M. A survey of semantic segmentation. arXiv 2016, arXiv:1602.06541.

10. Lateef, F.; Ruichek, Y. Survey on semantic segmentation using deep learning techniques. Neurocomputing 2019, 338, 321–348.

11. Sehar, U.; Naseem, M.L. How deep learning is empowering semantic segmentation. Multimed. Tools Appl. 2022, 1573–7721.

12. Chakravarthy, A.D.; Bonthu, S.; Chen, Z.; Zhu, Q. Predictive Models with Resampling: A Comparative Study of Machine Learning Algorithms and their Performances on Handling Imbalanced Datasets. In Proceedings of the 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA), Boca Raton, FL, USA, 16–19 December 2019; pp. 1492–1495.

13. Abeyrathna, D.; Subramaniam, M.; Chundi, P.; Hasanreisoglu, M.; Halim, M.S.; Ozdal, P.C.; Nguyen, Q. Directed Fine Tuning Using Feature Clustering for Instance Segmentation of Toxoplasmosis Fundus Images. In Proceedings of the 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE), Cincinnati, OH, USA, 26–28 October 2020; pp. 767–772.

14. Halim, S.M.; (Byers Eye Institute at Stanford University, Palo Alto, CA, USA). Personal communication, 2020.

15. Abeyrathna, D.; Life, T.; Rauniyar, S.; Ragi, S.; Sani, R.; Chundi, P. Segmentation of Bacterial Cells in Biofilms Using an Overlapped Ellipse Fitting Technique. In Proceedings of the 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Houston, TX, USA, 9–12 December 2021; pp. 3548–3554.

16. Bommanapally, V.; Ashaduzzman, M.; Malshe, M.; Chundi, P.; Subramaniam, M. Self-supervised Learning Approach to Detect Corrosion Products in Biofilm images. In Proceedings of the 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Houston, TX, USA, 9–12 December 2021; pp. 3555–3561.

17. Kalimuthu, J.; (Civil and Environmental Engineering Department, South Dakota School of Mines Technology, Rapid City, SD, USA). Personal Communication, 2022.

18. Tajbakhsh, N.; Jeyaseelan, L.; Li, Q.; Chiang, J.N.; Wu, Z.; Ding, X. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. Med. Image Anal. 2020, 63, 101693.

19. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 2017, 39, 2481–2495.

20. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.

21. Yang, L.; Zhang, Y.; Chen, J.; Zhang, S.; Chen, D.Z. Suggestive Annotation: A Deep Active Learning Framework for Biomedical Image Segmentation. In Proceedings of the MICCAI, Quebec City, QC, Canada, 11–13 September 2017.

22. Ozdemir, F.; Peng, Z.; Tanner, C.; Fuernstahl, P.; Goksel, O. Active learning for segmentation by optimizing content information for maximal entropy. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support; Springer: Berlin/Heidelberg, Germany, 2018; pp. 183–191.

23. Gal, Y.; Ghahramani, Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In Proceedings of the International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; pp. 1050–1059.

24. Sourati, J.; Gholipour, A.; Dy, J.G.; Kurugol, S.; Warfield, S.K. Active deep learning with fisher information for patch-wise semantic segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support; Springer: Berlin/Heidelberg, Germany, 2018; pp. 83–91.

25. Kuo, W.; Häne, C.; Yuh, E.; Mukherjee, P.; Malik, J. Cost-sensitive active learning for intracranial hemorrhage detection. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 16–20 September 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 715–723.

26. Zheng, H.; Yang, L.; Chen, J.; Han, J.; Zhang, Y.; Liang, P.; Zhao, Z.; Wang, C.; Chen, D.Z. Biomedical image segmentation via representative annotation. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 5901–5908.

27. Sourati, J.; Gholipour, A.; Dy, J.G.; Tomas-Fernandez, X.; Kurugol, S.; Warfield, S.K. Intelligent labeling based on fisher information for medical image segmentation using deep learning. IEEE Trans. Med. Imaging 2019, 38, 2642–2653.

28. Shin, G.; Xie, W.; Albanie, S. All You Need Are a Few Pixels: Semantic Segmentation With PixelPick. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, Montreal, BC, Canada, 11–17 October 2021; pp. 1687–1697.

29. Zhang, L.; Gopalakrishnan, V.; Lu, L.; Summers, R.M.; Moss, J.; Yao, J. Self-learning to detect and segment cysts in lung ct images without manual annotation. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 1100–1103.

30. Bai, W.; Oktay, O.; Sinclair, M.; Suzuki, H.; Rajchl, M.; Tarroni, G.; Glocker, B.; King, A.; Matthews, P.M.; Rueckert, D. Semi-supervised learning for network-based cardiac MR image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Quebec City, QC, Canada, 11–13 September 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 253–260.

31. Vincent, L.; Soille, P. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. IEEE Comput. Archit. Lett. 1991, 13, 583–598.

32. Chakravarthy, A.D.; Chundi, P.; Subramaniam, M.; Ragi, S.; Gadhamshetty, V.R. A Thrifty Annotation Generation Approach for Semantic Segmentation of Biofilms. In Proceedings of the 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE), Cincinnati, OH, USA, 26–28 October 2020; pp. 602–607.

33. Grau, V.; Mewes, A.U.; Alcañiz, M.; Kikinis, R.; Warfield, S.K. Improved watershed transform for medical image segmentation using prior information. IEEE Trans. Med. Imaging 2004, 23, 447–458.

34. Grau, V.; Kikinis, R.; Alcañiz, M.; Warfield, S.K. Cortical gray matter segmentation using an improved watershed transform. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology—Proceedings, Cancun, Mexico, 17–21 September 2003; Volume 1, pp. 618–621.

35. Ng, H.P.; Ong, S.H.; Foong, K.W.C.; Goh, P.S.; Nowinski, W.L. Medical Image Segmentation Using K-Means Clustering and Improved Watershed Algorithm. In Proceedings of the 2006 IEEE Southwest Symposium on Image Analysis and Interpretation, Hangzhou, China, 2–4 November 2001; pp. 61–65.

36. Beucher, S.; Meyer, F. The Morphological Approach to Segmentation: The Watershed Transformation. Available online: https://www.researchgate.net/profile/Serge-Beucher/publication/230837870_The_Morphological_Approach_to_Segmentation_The_Watershed_Transformation/links/00b7d5319b Morphological-Approach-to-Segmentation-The-Watershed-Transformation.pdf (accessed on 23 May 2022).

37. Salembier, P. Morphological multiscale segmentation for image coding. Signal Process. 1994, 38, 359–386.

38. Malpica, N.; De Solórzano, C.O.; Vaquero, J.J.; Santos, A.; Vallcorba, I.; García-Sagredo, J.M.; Del Pozo, F. Applying watershed algorithms to the segmentation of clustered nuclei. Cytometry: J. Int. Soc. Anal. Cytol. 1997, 28, 289–297.

39. Petit, O.; Thome, N.; Charnoz, A.; Hostettler, A.; Soler, L. Handling missing annotations for semantic segmentation with deep convnets. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support; Springer: Berlin/Heidelberg, Germany, 2018; pp. 20–28.

40. Bengio, Y.; Louradour, J.; Collobert, R.; Weston, J. Curriculum learning. In Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, QC, Canada, 14–18 June 2009; pp. 41–48.

41. Petit, O.; Thome, N.; Soler, L. Iterative confidence relabeling with deep ConvNets for organ segmentation with partial labels. Comput. Med. Imaging Graph. 2021, 91, 101938.

42. Pan, J.; Bi, Q.; Yang, Y.; Zhu, P.; Bian, C. Label-efficient Hybrid-supervised Learning for Medical Image Segmentation. arXiv 2022, arXiv:2203.05956.

43. Zhou, H.; Song, K.; Zhang, X.; Gui, W.; Qian, Q. WAILS: Watershed Algorithm With Image-Level Supervision for Weakly Supervised Semantic Segmentation. IEEE Access 2019, 7, 42745–42756.

44. Zhou, S.; Nie, D.; Adeli, E.; Yin, J.; Lian, J.; Shen, D. High-resolution encoder–decoder networks for low-contrast medical image segmentation. IEEE Trans. Image Process. 2019, 29, 461–475.

45. Ning, Z.; Zhong, S.; Feng, Q.; Chen, W.; Zhang, Y. SMU-Net: Saliency-Guided Morphology-Aware U-Net for Breast Lesion Segmentation in Ultrasound Image. IEEE Trans. Med. Imaging 2022, 41, 476–490.