Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/21/20231126

An overview of research results and applications in the field of semi-supervised semantic segmentation

Dingrui Shi

School of Data Sciences, Zhejiang University of Finance & Economics, Hangzhou, Zhejiang, China, 310018

19157706302@163.com

Abstract. Traditional segmentation approaches, supervised deep learning methods, and semi-supervised deep learning methods have all found widespread use as the field of semi-supervised semantic segmentation has advanced. These methods have developed and progressed over time, opening up novel avenues of research in the field of image segmentation and giving potent resources for tackling difficult practical issues. These developments have deepened our understanding of image segmentation and provided flexible and efficient solutions to challenges in practical applications, ranging from classical traditional approaches to supervised methods based on deep learning, and beyond to semi-supervised methods that leverage both labeled and unlabeled data. Focusing on their specialized applications in medical and remote sensing image processing, this paper presents a complete overview of the development status of these methods. This study's image segmentation solutions can help tackle actual-world issues where annotated data is rare or expensive to some extent.

Keywords: semi-supervised semantic segmentation, Generative Adversarial Networks, the pseudo-labeling method, consistency regularization.

1. Introduction

There have been many breakthroughs in the study of semi-supervised semantic segmentation. These include both supervised and unsupervised deep learning approaches. Features like pixel value similarity and edge detection are heavily relied upon by conventional segmentation techniques. Semantic segmentation uses techniques such as region expanding, edge detection, and graph cuts. To get the best results in semantic segmentation, supervised deep learning techniques use annotated data to train deep neural network models. U-Net, FCNs (Fully Convolution Networks), DeepLab, and other designs like them are all good examples. To improve model generalization and accuracy, semi-supervised learning approaches combine supervised and unsupervised learning, training on a small annotated dataset while also incorporating unlabeled data. These methods are useful for semantic segmentation tasks in real-world contexts where there is a lack of abundant labeled data. Current popular research topics in the field of semi-supervised learning include adversarial training, pseudo-labeling, and consistency regularization. These methods are crucial for training on sparse amounts of annotated and unlabeled data.

With the method of a literature review and analysis, this paper provides a brief overview of where technology stands today in the fields of adversarial learning, pseudo-labeling, and consistency regularization, and it also details the specific uses of semi-supervised segmentation in the processing of images from medical imaging and remote sensing. Accurate picture segmentation results are essential in many fields, but getting annotated data is typically difficult or expensive. Therefore, semi-supervised learning method research and implementation is crucial for solving these real-world issues.

The purpose of this work is to provide scholars with a thorough grasp of the many approaches used in semi-supervised semantic segmentation, as well as insights and suggestions for further refinement and application. In addition, the work provided in this paper offers practical answers to issues with picture segmentation in sectors like medical and remote sensing image analysis, which will undoubtedly have an effect on the growth and development of those disciplines.

2. Generative adversarial networks

With impressive results across a range of applications, including picture generation, object detection, and semantic segmentation, Generative Adversarial Networks (GANs) have emerged as a popular framework. The two networks that make up a GAN framework are the generator and the discriminator. The purpose of the generator is to acquire knowledge of the distribution of the target data so that random noise can be used to create synthetic images. The discriminator's job is to tell the difference between genuine images (from the true distribution) and forgeries produced by the generator [1].

Generative adversarial networks (GANs) can be used to create adversarial examples, which are slightly altered versions of input data that have visual similarities to the original but trick the model's output. These adversarial examples can be used to trick black-box models into giving misleading findings, which can then be exploited in attacks. By studying the strengths and weaknesses of their targets, GANs can provide more convincing adversarial instances. GANs can also be used to improve security defensive systems by, for instance, spotting adversarial cases and increasing model resilience. It is also important to strengthen adversarial example defenses and develop better defense mechanisms to ensure the security of target systems while employing GANs to improve black-box attack capabilities.

Black-box attacks seldom succeed because of the large gap between the attacker's and the target's models. To improve attack efficiency, numerous models might be attacked at once to produce more generalizable adversarial examples. By applying frequency domain transformations to the input image, we can improve the model and create more generalizable adversarial examples that can be used to attack standard training and defense models, overcoming a limitation of existing spatial transformations that prevents them from being generalized to a wide range of target models. Diverse spectral saliency maps are generated by frequency-domain transformations, which, when paired with existing attack methods, can enhance black-box attack capabilities [2].

A number of adversarial training approaches have been proposed to improve model robustness, although these approaches generally compromise model accuracy. Nonetheless, users may have varying expectations for the level of resilience and accuracy of the model, suggesting that a new model will need to be retrained for each user. An Adversarial Training with Dynamic Balancing (ATDB) approach, which can dynamically change the trade-off between standard accuracy and robust accuracy during testing, is needed to solve this problem. In contrast to traditional model-conditional training frameworks, the inputs to One-Shot Adversarial Training (OAT) are control hyperparameters. The trained model can readily adapt to varying requirements and maintain a high degree of accuracy during testing. By using dual batch normalization, researchers can train inside the same model without degrading performance by combining standard feature statistics with adversarial feature statistics in a way that is statistically independent. In addition, trade-offs between accuracy, resilience, and runtime efficiency are possible by extending OAT to a One-Time Adversarial Training with Scalability (OATS) framework. This method drastically lessens resource demands while permitting alterations based on user requirements, boosting the model's usability and adaptability [3].

3. The pseudo-labeling method

Pseudo-labeling is a method of supervised learning that uses both labeled and unlabeled data to boost model accuracy. In this technique, the highest anticipated probability serves as a pseudo-label for data

DOI: 10.54254/2755-2721/21/20231126

that has not been labeled. Using pseudo-labels helps define classes more precisely and reduce the size of the learnt classes. But inaccurate pseudo-labels can add unnecessary noise and throw off the model's inference. Overfitting problems can arise when pseudo-labels are overconfident since they add nothing new to the model.

The reliability of pseudo-labels is usually quite high. This means that the performance of the model will suffer greatly if training is carried out with low-quality pseudo-labels. Consequently, the uncertainty of output values can be estimated as a substitute measure of confidence by taking into account methods for estimating the uncertainty of deep neural networks. The performance of the model can be improved by selecting credible samples for pseudo-labeling using the probabilities from the softmax layer alongside uncertainty estimations.

In order to avoid wasting time and effort, the pseudo-labeling method only uses high-confidence predictions as pseudo-labels. Everyone agrees that every pixel matters, no matter how unclear its label. A new approach is proposed in this article that makes use of entropy as a measure of the accuracy of prediction findings, with low entropy indicating accurate predictions and high entropy indicating inaccurate ones. All unlabeled data, including those with inaccurate predictions, can have an effect on the training process since they can be used as partial negative samples for certain classes. To boost model efficiency, this technique makes complete use of uncertainty data [4].

4. Consistency regularization

Overfitting can be prevented in supervised learning by including additional loss terms. One form of regularization used in semi-supervised learning is called consistency regularization. In this approach, variations of the same image are handled as independent inputs, and consistency constraints are applied to the predictions made from each. Data augmentation methods like random cropping and flipping can be used to meet this restriction. The method relies on the smoothness assumption and the cluster assumption, which state that similar data points provide similar results while data points with distinct labels are separated in low-density regions.

To ensure that the classifier's pairwise distance distribution matches the distribution in the latent space, a novel consistency regularization method, called hypersphere space consistency regularization, can be proposed for semi-supervised and weakly supervised learning by drawing on insights from some geometric theories. In machine learning, training stability is enhanced by restricting the output space to the unit hypersphere, where features well-clustered on the hypersphere are linearly separable from the rest of the feature space [5].

To introduce invariance or robustness into the model, self-supervised learning promotes features collected from two distinct augmented images to pull against each other. Consistency regularization is intelligently guided by the confidence between model predictions from two substantially enhanced images, allowing the semi-supervised learning framework to also learn self-supervised operations. Specifically, it entails tripling the number of forks from two to three, with one fork representing a weak perturbation and the other two representing significant perturbations. Moreover, the confidence of pseudo-labels derived from each substantially augmented image can be measured, and the consistency loss can be weighted according to this confidence [6].

5. Semi-supervised segmentation in medical and remote sensing images

5.1. Semi-supervised segmentation in the medical field

In medical applications, accurate semantic segmentation results are crucial because they give the anatomical data doctors need to diagnose and cure patients. In the field of biomedical picture segmentation, where diverse organs and patterns need to be separated, supervised deep learning algorithms have recently surpassed the state-of-the-art thanks to the advent of convolutional neural networks. In order to train, these strategies require a sizable amount of labeled data with predetermined pixel or voxel categories. However, it is difficult to get densely labeled medical images and takes a great deal of effort and domain-specific expertise. Several solutions have been presented to this issue, and

DOI: 10.54254/2755-2721/21/20231126

data augmentation is one of them. Through the use of generative adversarial networks (GANs) or simple linear combinations, data augmentation can increase the size of a labeled dataset. Data augmentation approaches can improve segmentation performance, but they have their limits due to the artificial nature of the generated images and labels. Semi-supervised learning-based methods are an improved alternative because they may make use of both labeled and unlabeled data. This strategy has promise for medical picture segmentation challenges since it improves model performance and generalizability by combining small amounts of labeled data with a large amount of unlabeled data. Semi-supervised approaches, which combine supervised and unsupervised learning, are a powerful tool for medical picture segmentation [7]. The core components of a deep segmentation network in the semi-supervised setting are the encoder E and the decoder D. In order to extract features from input images, the encoder E uses a series of downsampling blocks, whereas the decoder D uses a series of upsampling blocks to restore features to pixel-level predictions. During the pre-training phase, we exclusively use unlabeled data to train the encoder E, which uses self-supervised global contrastive learning to extract features from images. To further train the encoder E and decoder D to acquire pixel-level feature representations, supervised local contrastive learning is done on partially labeled data. This semi-supervised framework uses both unlabeled and labeled data to its full potential, learning global features through self-supervised pre-training and acquiring pixel-level features through supervised training to boost the deep segmentation network's performance and representation ability [8].

5.2. Application of semi-supervised segmentation in remote sensing

Numerous disciplines, including climate science, environmental monitoring, military surveillance, and land resources, rely heavily on remote sensing technology as a crucial data collecting and processing method. Classifying images via remote sensing has its own unique quirks. Spectral feature redundancy, same-spectral-different-object, and different-spectral-same-object are just a few examples of the difficulties associated with hyperspectral remote sensing photos. The categorization outcomes for high-resolution remote sensing images are strongly influenced by the consistency and efficiency of their spatial characteristics. Due to their great spatial resolution and intricate details, high-resolution remote sensing photos cannot be semantically segmented using existing techniques for natural images. One of the most active areas of study in remote sensing image processing is semantic segmentation of high-resolution remote sensing images.

Several obstacles exist for semantic segmentation of high-resolution remote sensing images. First, high-resolution remote sensing images exhibit considerable intra-class differences and minor inter-class differences due to factors like sensor angle and ambient changes, which makes successful feature extraction more challenging. Second, training on high-resolution remote sensing images might be problematic since their huge dimensions necessitate splitting them up into tiny sub-images, which can break up the overall image. Finally, high-resolution remote sensing image semantic segmentation using supervised algorithms is labor-intensive and time-consuming since it requires a large number of labels at the pixel level.

Recent studies have demonstrated that some of these problems can be solved by including generative adversarial networks (GANs) into natural picture semantic segmentation. With GANs, semantic annotations aren't necessary because the models can learn image structure, keep spatial label continuity in segmentation findings, and train with unlabeled data. Incorporating GANs into a semi-supervised semantic segmentation model for high-resolution remote sensing images allows for accurate semantic segmentation with less labeled samples needed. An end-to-end fully convolutional network can be developed as the discriminator to efficiently handle the peculiarities of high-resolution remote sensing images. By properly restricting the gradient updates of the segmentation network, this network is able to learn the local structural information of the pictures while preserving label continuity in the overall image segmentation results. In addition, attention-selection-based adversarial loss functions can be implemented. This model uses convolutional neural networks to extract characteristics from high-resolution remote sensing photos, which is an improvement over typical non-deep learning segmentation methods. In addition, the spatial continuity of segmentation results for large-sized remote sensing

DOI: 10.54254/2755-2721/21/20231126

images is effectively preserved using this method, which is an improvement over existing deep learning-based remote sensing image semantic segmentation models. To further enhance the precision of segmentation outcomes, attention-based loss functions are built. Semantic segmentation problems in high-resolution remote sensing photos can be solved with the help of the proposed model design and optimization methodologies, leading to more reliable and accurate findings. The remote sensing image processing community stands to benefit greatly from this method's comprehensive applicability, since it provides solid backing for relevant studies and applications [9].

6. Conclusion

To improve the model's generalization ability and accuracy, semi-supervised learning mixes supervised and unsupervised learning by training on a small labeled dataset and then using unlabeled data to supplement the training. This is especially important for semantic segmentation tasks where labeled data is rare in practical applications. Adversarial learning methods, pseudo-labeling approaches, and consistency regularization are some of the most popular areas of study right now. Generative adversarial networks (GANs) provide the basis for adversarial learning approaches, which have showed promise in applications like semantic segmentation. The generator network learns how the target data is distributed, and the discriminator network can tell the difference between authentic and fabricated images. The adversarial samples used in attacking and defending models can be generated with the help of GANs. Semi-supervised learning approaches like pseudo-labeling use the highest predicted probabilities from unlabeled data as "pseudo-labels" to boost model accuracy. To prevent the introduction of noisy samples, pseudo-labels can be generated from only the most trustworthy samples by taking uncertainty estimation into account. To prevent overfitting, semi-supervised learners might employ consistency regularization, a form of regularization. This approach utilizes data augmentation approaches to increase model performance while applying consistency requirements by taking into account alternate copies of the same image.

Because it is so challenging to collect labeled data, semi-supervised learning approaches have promise for use in the medical and remote sensing picture segmentation fields. These strategies efficiently combine small amounts of labeled data with massive amounts of unlabeled data to boost model performance and generalization. To solve segmentation problems, semi-supervised methods integrate supervised and unsupervised learning by building deep segmentation networks using encoders and decoders.

There are still a few limitations in our understanding of semi-supervised semantic segmentation algorithms, despite the fact that they have been thoroughly examined and studied in this paper. To begin, this research mostly relied on literature review and analysis techniques rather than actual surveys or experimental validations. This means that the efficacy and performance of semi-supervised semantic segmentation algorithms in real-world settings lack direct empirical research backing. Second, it's possible that some recent developments in the field were not accounted for in this study. New techniques and tools are being developed and introduced into the area on a regular basis. Therefore, it may be necessary to enhance the overall knowledge of the research field. Semantic segmentation is only one area where semi-supervised learning techniques like self-supervised learning, transfer learning, etc. could benefit from more investigation and study. These techniques have shown promise in other areas, and they may be able to help with semantic segmentation if applied properly.

References

[1] Pel'aez-Vegas, A., Mesejo, P., & Luengo, J. (2023). A Survey on Semi-Supervised Semantic Segmentation. ArXiv, abs/2302.09899.
[2] Long, Y., Zhang, Q., Zeng, B., Gao, L., Liu, X., Zhang, J., & Song, J. (2022). Frequency Domain Model Augmentation for Adversarial Attack. ArXiv, abs/2207.05382.
[3] Wang, H., Chen, T., Gui, S., Hu, T., Liu, J., & Wang, Z. (2020). Once-for-All Adversarial Training: In-Situ Tradeoff between Robustness and Accuracy for Free. ArXiv, abs/2010.11828.

[4] Wang, Y., Wang, H., Shen, Y., Fei, J., Li, W., Jin, G., Wu, L., Zhao, R., & Le, X. (2022). Semi-Supervised Semantic Segmentation Using Unreliable Pseudo-Labels. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4238-4247.

[5] Tan, C., Gao, Z., Wu, L., Li, S., & Li, S. (2022). Hyperspherical Consistency Regularization. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 7234-7245.

[6] Kim, J., Min, Y., Kim, D., Lee, G., Seo, J., Ryoo, K., & Kim, S. (2022). ConMatch: Semi-Supervised Learning with Confidence-Guided Consistency Regularization. ArXiv, abs/2208.08631.

[7] Hu, X., Zeng, D., Xu, X., & Shi, Y. (2021). Semi-supervised Contrastive Learning for Label-efficient Medical Image Segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention.

[8] Hu, X., Zeng, D., Xu, X., & Shi, Y. (2021). Semi-supervised Contrastive Learning for Label-efficient Medical Image Segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention.

[9] LIU Yu-Xi,ZHANG Bo,WANG Bin.Semi-supervised semantic segmentation based on Generative Adversarial Networks for remote sensing images[J].Journal of INFRARED AND MILLIMETER WAVES,2020,39(04):473-482.