IJMRSET
e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 6, June 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

6381 907 438 6381 907 438 ijmrset@gmail.com www.ijmrset.com

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 6, June 2024 |

| DOI:10.15680/IJMRSET.2024.0706139 |

A Review on Semantic Segmentation

M Sai Akash, S Praveena

I M. Tech, DECE, Department of ECE, MGIT, Hyderabad, India

Associate Professor, Department of ECE, MGIT, Hyderabad, India

ABSTRACT: Semantic segmentation, a key task in computer vision, involves predicting pixel-level labels to identify specific regions or objects within an image. It serves as a cornerstone in numerous applications, including scene understanding, human-machine interaction, computational photography, image search engines, and autonomous driving, where it aids in understanding object relationships. This paper provides a detailed survey of semantic segmentation techniques. We begin with an introduction to the task and its diverse applications. Subsequently, we categorize methods based on the number of input modalities, offering a comprehensive review of various approaches. We analyze the motivations behind these methods, evaluate their performance on benchmark datasets, and discuss their contributions and significance to the field. Lastly, we present a comparative overview of the datasets commonly employed in these studies.

KEYWORDS: Semantic Segmentation, Deep Learning, Neural Networks, Multi-Modal Approaches**I**

I. INTRODUCTION

Semantic Segmentation is a crucial component of image object detection, focusing on understanding an image at the pixel level. It involves assigning a class label to each pixel in an image, thereby predicting not only the object category but also delineating its boundaries. For instance, in autonomous vehicles operating in urban environments, semantic segmentation provides essential information about the road scene, such as the appearance (e.g., buildings, roads), shape (e.g., cars, pedestrians), and spatial relationships (context) between objects. This enables vehicles to make safe and deliberate navigation decisions. Similarly, in unstructured off-road settings, accurate identification of semantic classes like trails, grass, or rocks is vital for ensuring safety and preventing collisions.

Semantic segmentation's importance extends across computer vision tasks, as it forms the basis for critical decision-making in applications like autonomous driving. Historically, many techniques for semantic segmentation relied on hand-engineered features and treated pixel classification independently. Among these, methods like Random Forests and Boosting were particularly effective, predicting the class probabilities of central pixels. However, the advent of deep learning and the success of Convolutional Neural Networks (CNNs) in tasks like ImageNet classification significantly advanced the field.

The first notable attempt to adapt CNNs for variable-size input was based on LeNet, which recognized digit strings. Since then, deep learning has revolutionized semantic segmentation, yielding significant accuracy improvements. Consequently, most recent research has focused on deep learning approaches, leaving traditional techniques behind. This paper surveys deep learning-based semantic segmentation methods, categorizing them into two classes based on input modalities: unimodal (e.g., image only) and multimodal (e.g., image combined with CRF, RNN, or 3D point clouds).

Semantic segmentation is particularly suited for applications requiring precise pixel-wise classification. Its architecture facilitates the identification of objects, their dividing boundaries, and their spatial relationships within a scene. For example, an image may depict elements like the sky, land, and vegetation, with semantic segmentation capturing their appearances and interactions.

At its core, semantic segmentation involves assigning a state from a label set LLL to each pixel. Each label represents a distinct object or class, such as cars, roads, or backgrounds, with an additional state for non-relevant classes. Semantic segmentation has diverse applications, including scene understanding, human-machine interaction, computational photography, image search engines, and autonomous driving, where it helps infer relationships among objects. Its growing significance underscores its transformative impact on computer vision and machine learning.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 6, June 2024 |

| DOI:10.15680/IJMRSET.2024.0706139 |

II. RELATED WORK

Before delving into specific methods, it is essential to outline some foundational concepts that underpin semantic segmentation learning. First, the fundamental elements for deep learning using neural networks include the primary architectures, methodologies, and design considerations. Second, advanced techniques such as transfer learning play a crucial role during training, enhancing performance by leveraging pre-trained models. Lastly, when working with limited datasets, data preprocessing and augmentation techniques are indispensable for improving results.

Based on these foundational concepts, the methods for semantic segmentation can be divided into two primary categories: unimodal image-based approaches and multimodal approaches. Each category contains several methods, which we will explore in detail, focusing on their contributions to the advancement of semantic segmentation.

1) Unimodal Image-Based Approaches

Unimodal image-based approaches, particularly those using RGB images, are a major area of research in semantic segmentation due to their importance in visual recognition. Before the rise of deep learning, traditional methods relied on graphical models, such as Markov or Conditional Random Fields (MRFs or CRFs) [1]–[4]. These methods successfully segmented images into superpixels, extracting hand-crafted features from individual and neighboring segments, and employed graphical models to ensure label consistency across neighboring regions.

The advent of Convolutional Neural Networks (CNNs) revolutionized semantic segmentation by enabling automatic learning of robust features directly from datasets. A significant challenge was integrating CNNs into semantic segmentation tasks, which was ultimately addressed by the development of Fully Convolutional Networks (FCNs). This breakthrough led to numerous advancements, primarily distinguished by variations in network architecture.

1) Fully Convolutional Networks (FCNs):
Shelhamer et al. [5] introduced the first deep learning-based approach for semantic segmentation using fully convolutional networks (FCNs). This method enabled end-to-end training and pixel-to-pixel predictions, significantly enhancing performance compared to previous approaches.

The main contribution of FCNs was their ability to adapt convolutional neural networks for dense prediction tasks, generating output predictions of corresponding size for arbitrary input dimensions. The network was designed to ensure efficient and accurate inference.

Fully Convolutional Design: The authors redefined the architecture of CNNs by replacing fully connected layers with fully convolutional layers, allowing the network to process inputs of any size and produce output maps of corresponding size. This transformation interpreted fully connected layers as 1×11 times 11×1 convolutional layers, extracting channel-wise information from input images. **Skip Connections**: To enhance semantic accuracy, skip connections were used to combine information from both coarse, deep layers and fine, shallow layers. The deep layers captured high-level appearance information, while the shallow layers refined this data for more precise predictions. This approach significantly improved reconstruction accuracy and prevented chaotic outputs. **End-to-End Training**: The network was trained using established architectures like AlexNet, VGG, and GoogLeNet, with VGG serving as the primary model backbone. The use of pre-trained networks enabled the authors to achieve state-of-the-art results.

The FCN approach resulted in efficient, pixel-wise prediction and enabled dense segmentation of images with superior accuracy. The innovative use of skip connections and the redefinition of fully connected layers marked a significant leap in the field of semantic segmentation, setting the stage for future developments. This transformation and its impact are illustrated in Figure 1 of the original work.

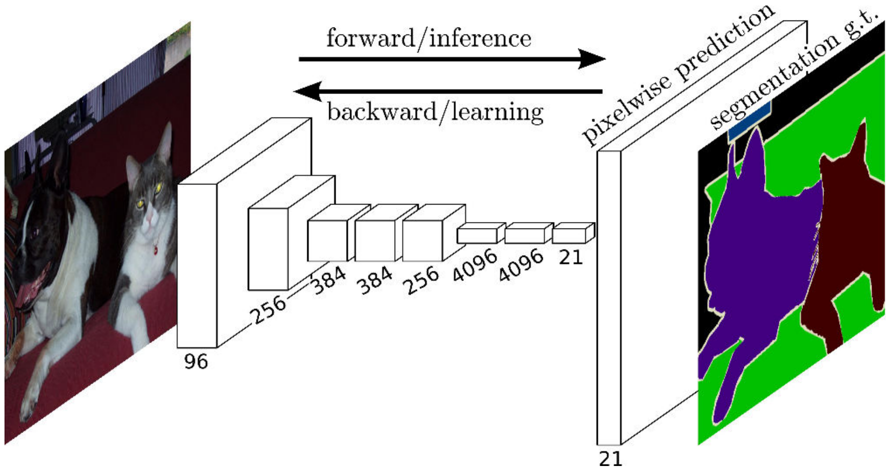

FIGURE 1:THE ARCHITECTURE OF FULLY CONVOLUTIONAL NETWORKS AND HOW THE NEW NETWORK GET THE FINAL PIXEL PREDICTION FOR SEMANTIC SEGMENTATION.

2) Urban Scene Segmentation with Laser-Constrained CRFs:
In 2017, Alvis et al. [6] introduced a novel approach for urban scene segmentation that combines appearance features extracted from images with the global constraints of superpixels derived from 3D point clouds, leveraging Conditional Random Fields (CRFs). This method was particularly valuable for robotics, where understanding the surrounding environment is crucial for autonomous navigation. Robots typically employ a variety of sensors, including color cameras, inertial measurement units, and 3D laser scanners, to gather multimodal data. The more diverse the sensor modalities, the better the performance of the segmentation task. However, integrating these different modalities into a cohesive solution remains a significant challenge.

For autonomous robots, semantic segmentation plays a key role in enabling them to comprehend their environment and make informed decisions. The data used for this task is typically in the form of either images or 3D point cloud data. As seen in previous methods, researchers have relied on CRFs and the Maximum A Posteriori (MAP) solution for inference-based segmentation.

In this paper, the authors proposed a new CRF inference method designed for scene segmentation that incorporates global constraints, ensuring that nodes within a set are assigned the same class label. To optimize computational efficiency and reduce processing time, the model uses a relaxed quadratic CRF approach, with its MAP solution derived through the widely-used stochastic gradient descent method. The proposed method was evaluated using both image and 3D point cloud data collected in urban environments, where image data provides the necessary appearance features for CRF, while 3D point clouds offer global spatial constraints across sets of nodes. Comparative analysis with belief propagation, traditional quadratic programming relaxation, and higher-order potential CRFs demonstrated the advantages of this new method.

Conclusion: Semantic segmentation plays a crucial role in the field of computer vision, and through our research, we aim to provide a foundational understanding of neural network-based methods. By comparing different approaches, we can gain insights into the evolution of semantic segmentation and trace the changes in its underlying structure. This understanding is valuable for new researchers, as it may inform and inspire future research efforts. Our goal is to conduct a more detailed comparison in the future, given the significant impact semantic segmentation has on both computer vision and machine learning.

REFERENCES

[1] C. Farabet, C. Couprie, L. Najman, and Y. LeCun. Learning Hierarchical Features for Scene Labeling. In IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), vol.35, no. 8, pp. 1915-1929, 2013.
[2] L. Ladicky, P. Sturgess, K. Alahari, C. Russell, and P. H. S. Torr. What, Where and How Many? Combining Object Detectors and CRFs. In Proc. European Conf. on Computer Vision (ECCV), 2010.
[3] B. Micusik, J. Kosecka, and G. Singh. Semantic Parsing of Street Scenes from Video. In Intl J. Rob. Res. (IJRR), vol. 31, no. 4, pp. 484-497, 2012.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 6, June 2024 |

| DOI:10.15680/IJMRSET.2024.0706139 |

[4] J. Xiao and L. Quan. Multiple View Semantic Segmentation for Street View Images. In Proc. IEEE Intl Conf. on Computer Vision (ICCV), 2009.
[5] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 39(4):640-651.
[6] Alvis C D, Ott L, Ramos F. Urban scene segmentation with laser-constrained CRFs[C]// Ieee/rsj International Conference on Intelligent Robots and Systems. IEEE, 2016:2147-2154