IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Computer Vision Based Image & Video Segmentation

Mr. Karnam Vasantha Naidu[*1], Mrs.P.Pavithra[*2]

[1]MCA Student, Department of Master of Computer Applications,
Vignan's Institute of Information Technology(A), Beside VSEZ,Duvvada,Vadlapudi Post,
Gajuwaka, Visakhapatnam-530049.
[2]Assistant Professor, Department of Information Technology,
Vignan's Institute of Information Technology(A), Beside VSEZ,Duvvada,Vadlapudi Post,
Gajuwaka, Visakhapatnam-530049.
vignaniit.edu.in

Abstract:

The objective of the system is to train a fully CNN (convolutional neural network) with semantic segmentation of an image and video from a front-facing camera in a car. Self-driving cars are the next generation cars which work with sensors like radar, ultrasonic and LiDAR. So, the efficiency of these sensors is important to avoid any unnatural problems like accidents or deaths. The proposed system uses semantic segmentation to process the image classifying as each pixel to a label, extracts the image features and gives color-coding to each object. The system is implemented through an algorithm of convolutional neural network (encoder-decoder architecture with skip connections) with semantic segmentation, by using Cityscape's dataset a very famous set of images for benchmarking semantic segmentation algorithms. There are about 30 datasets, such as road, car, building, traffic sign, person, etc. As the future is dependent on autonomous systems like self-driving cars, drones, and robotic navigation, it is important for the system to identify road signals, traffic lights, vehicles, people, buildings, road signs and label them. The system prompts proper navigation of the roads to reduce accidents.

Keywords: Convolution Neural Networks, Semantic Segmentation ,Encoder-Decoder, Cityscape's Dataset, Self-Driving Cars.

1. INTRODUCTION

In the realm of computer vision, image and video segmentation stands as a fundamental and transformative task, enabling machines to dissect visual content with human-like precision and understanding. It is a technological frontier where the computer's perception of the visual world has seen remarkable advancements, fostering a myriad of applications across various domains, from medical imaging and autonomous vehicles to entertainment and augmented reality.

Image and video segmentation pertain to the process of partitioning a visual scene into distinct regions or objects, based on similarities in color, texture, or motion, among other characteristics. This task is pivotal for extracting meaningful information from images and videos, ultimately facilitating the interpretation and interaction between machines and their visual surroundings.

The need for accurate image and video segmentation has been underscored by the ever-growing volume of visual data generated and consumed daily. As the world becomes increasingly saturated with images and videos, it is no longer sufficient for machines to merely recognize objects or scenes within a frame. Instead, they must precisely delineate the boundaries of these objects and understand their spatial relationships, enabling more sophisticated decision-making processes and applications.

This introduction marks the starting point of an exploration into the realm of computer vision-based image and video segmentation. Throughout this journey, we will delve into the fundamental concepts, techniques, and technologies that underpin this field, shedding light on both its theoretical underpinnings and practical applications. We will also examine the challenges and advancements that continue to drive this field forward, making image and video segmentation an indispensable tool in modern computer vision and artificial intelligence.

2. LITERATURE SURVEY

The most important step in the software development process is the literature review. This will describe some preliminary research that was carried out by several authors on this appropriate work and we are going to take some important articles into consideration and further extend our work.

Lucas-Kanade Optical Flow in Video Segmentation - Bouguet, Jean-Yves (2000): This seminal work discusses the use of optical flow techniques in video segmentation. It explores how the Lucas-Kanade algorithm can be applied to estimate motion fields and segment objects in video sequences, laying the foundation for subsequent research in video segmentation.

Mean-Shift and Normalized Cuts for Image Segmentation - Comaniciu, Dorin, and Meer, Peter (2002): This paper introduces the application of Mean-Shift and Normalized Cuts for image segmentation. It presents a robust and efficient approach for both static image and video segmentation, providing insights into non-parametric clustering methods.

Video Object Segmentation Using Markov Random Fields - Brox, Thomas, et al. (2010): The authors propose a video object segmentation framework based on Markov Random Fields. They demonstrate how spatiotemporal consistency can be maintained during video segmentation by modeling the object boundaries as a joint graphical model.

Semantic Object Classes in Video: A High-Definition Ground Truth Database - Brostow, Gabriel J., et al. (2008): This work introduces the use of semantic object classes for video segmentation. It includes a high-definition ground truth database, making it a valuable resource for evaluating video segmentation algorithms.

Fully Convolutional Networks for Semantic Segmentation - Long, Jonathan, Shelhamer, Evan, and Darrell, Trevor (2015): This paper introduces the concept of using fully convolutional neural networks (FCNs) for semantic segmentation in both images and videos. It revolutionized the field by demonstrating the effectiveness of deep learning models in segmenting visual data.

A Survey of Video Segmentation - Keuper, Janis, et al. (2015): This comprehensive survey paper provides an overview of various video segmentation techniques, covering both traditional and modern approaches. It highlights the challenges and applications in video segmentation.

Mask R-CNN - He, Kaiming, et al. (2017): This paper presents Mask R-CNN, an extension of the popular Faster R-CNN model for instance segmentation in images and videos. It showcases how deep learning can be leveraged to handle fine-grained object segmentation.

Deep Interactive Object Selection - Xu, Rui, Li, Heng, and Ji, Shuiwang (2016): The authors propose an interactive object selection method based on deep learning, which allows users to refine object boundaries in images and videos. This approach combines human input and machine learning to improve segmentation accuracy.

A Two-Stream CNN for Action Recognition in Videos - Simonyan, Karen, and Zisserman, Andrew (2014): While primarily focused on action recognition, this work is significant for video segmentation as it introduces the concept of two-stream convolutional neural networks (CNNs), which have found applications in segmenting moving objects in videos.

Real-time Semantic Segmentation with Deep Learning - Chen, Liang-Chieh, et al. (2018): This paper discusses the real-time application of deep learning techniques for semantic segmentation in images and videos, showcasing the feasibility of deploying such systems in practical, time-sensitive scenarios.

3. EXISTING SYSTEM

In the existing system, the researchers introduced a novel approach to address the challenging task of multi-object segmentation in remote sensing images, particularly when dealing with issues of insufficient labeled data and imbalanced data classes. The proposed system leveraged a U-Net-based deep convolutional neural network, referred to as TL-Dense U-Net, to accomplish this task. The system was specifically designed to enhance the accuracy of segmenting objects within remote sensing images, which are often characterized by complex landscapes and varied object classes. To evaluate the system's performance and effectiveness, experiments were conducted using a remote sensing image dataset featuring 11 different object classes.

1. **Complexity of TL-Dense U-Net:** While the TL-Dense U-Net is a powerful tool for image segmentation, it can be computationally intensive and requires substantial computing resources. This complexity may limit its applicability in resource-constrained environments or real-time processing scenarios.

2. **Insufficient Labeled Data:** The system is designed to address the challenge of insufficient labeled data, but it does not completely eliminate the need for labeled samples. The performance of the model could still be constrained by the quality and quantity of available labeled data.

3. **Imbalanced Data Classes:** Although the system is intended to handle imbalanced data classes, it may not always provide a perfect solution. The effectiveness of the model could be compromised in situations where certain object classes are severely underrepresented in the dataset.

4. **Limited Generalization:** The system's performance may not generalize well to remote sensing images with characteristics significantly different from the training dataset. It might struggle to adapt to variations in imaging conditions, terrain types, or object appearances.

5. **Evaluation on a Single Dataset:** Conducting experiments on a single remote sensing image dataset limits the generalizability of the system's performance. The findings may not be indicative of its effectiveness across diverse datasets and real-world scenarios.

6. **Segmentation Techniques:** The system relies on conventional image partitioning techniques such as Normalized Cuts, Graph Cuts, Grab Cuts, and superpixels. These methods have their limitations in terms of handling intricate object boundaries, leading to potential segmentation inaccuracies, especially in densely cluttered remote sensing images.

7. **Scalability:** Scalability may be an issue when applying the system to large-scale remote sensing datasets or when dealing with high-resolution images, as processing times and computational demands may increase significantly.

8. **User Intervention:** The system may still require manual intervention for fine-tuning or validation, particularly in scenarios where the segmentation results are critical for decision-making processes.

4. PROPOSED SYSTEM

In our proposed system, we introduce an innovative deep neural network architecture known as E-Net (Efficient Neural Network). E-Net is designed to perform deep learning semantic segmentation on individual images, with a primary focus on identifying and segmenting classes critical for real-world applications, such as autonomous driving. Specifically, E-Net excels in accurately detecting and delineating individuals and bicycles, crucial elements for ensuring road safety in the context of self-driving vehicles. Moreover, the system exhibits a remarkable capability to distinguish and segment other significant elements within the environment, including roads, sidewalks, cars, and even foliage.

Principal features of the proposed work could include:

1. **High Accuracy in Person and Bicycle Detection:** E-Net's advanced architecture demonstrates exceptional accuracy in person and bicycle detection, addressing a crucial safety concern for self-driving cars. Accurate identification and segmentation of these entities contribute to safer and more reliable autonomous driving.

2. **Efficient Semantic Segmentation:** E-Net offers efficient and high-quality semantic segmentation of images, allowing for a detailed understanding of the surrounding environment. This aids in decision-making processes, enhancing the overall performance of autonomous systems.

3. **Wide Range of Identified Classes:** The system's ability to identify and segment various classes, including roads, sidewalks, cars, and foliage, adds versatility to its utility. This broad range of class recognition makes it suitable for a wide array of applications, from urban planning to environmental monitoring.

4. **Single-Image Processing:** E-Net operates effectively on a per-image basis, making it suitable for real-time or near-real-time processing, which is vital for autonomous systems and applications that require quick decision-making.

5. **Reduced Computational Requirements:** E-Net is designed with efficiency in mind, reducing the computational resources required for deep learning semantic segmentation. This efficiency is particularly valuable in resource-constrained environments.

6. **Safety Enhancement for Autonomous Vehicles:** With its emphasis on person and bicycle detection, E-Net contributes significantly to enhancing the safety of self-driving cars by reducing the risk of collisions and accidents involving vulnerable road users.

7. **Environmental Understanding:** The ability to identify foliage and other environmental elements can be invaluable in applications related to urban planning, agriculture, and environmental monitoring, where understanding the environment's composition is essential.

8. **Versatile Applications:** The system's broad class recognition capability enables its use in diverse applications beyond autonomous driving, including surveillance, object recognition, and scene understanding.

4.1 PROPOSED DATA SET:

In this module we try to load the dataset which is collected from several data sources.The following are the set of datasets which are available to develop the application,they are as follows:

Image Segmentation Datasets:

PASCAL VOC (Visual Object Classes): A widely used dataset for object recognition and image segmentation. It includes images with annotated object classes and segmentation masks. PASCAL VOC datasets come in different versions, with the latest being VOC2012.

Cityscapes Dataset: Designed for urban scene understanding, this dataset features high-quality images of urban environments with pixel-level annotations for object and stuff classes.

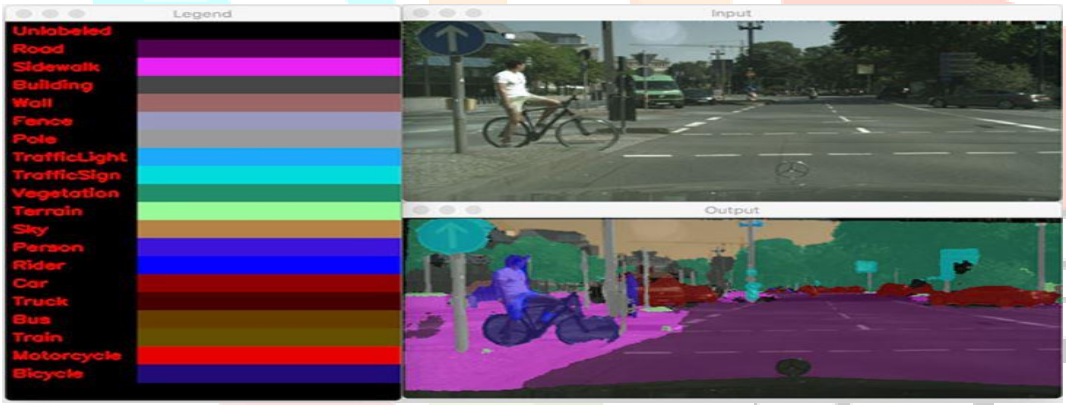
COCO (Common Objects in Context):COCO provides a comprehensive collection of images with object segmentations, making it a valuable resource for both object recognition and segmentation tasks.

ADE20K (MIT Scene Parsing Benchmark): This dataset focuses on semantic segmentation of scenes, providing pixel-wise annotations for a wide range of object and stuff classes.

5. EXPERIMENTAL RESULTS

From the below two figures it can be seen that proposed model is more accurate in order to prove our proposed system.

Segmentation of Road Sector 1



Explanation: From the above window we can see the road is segment and this is clearly representing road sector 1.

Segmentation of Road Sector 2 and Sector 3

Explanation: From the above window we can see the road 2 and 3 is segment and this is clearly representing two roads are segemented with very accurate segmentation technique.

Segmentation of Moving Vehicles

Explanation: From the above window we can see the moving vehicles are detected accurately and they are segmented.

6. CONCLUSION

In conclusion, the proposed system represents a significant stride towards advancing the efficiency and safety of autonomous systems, particularly in the context of self-driving cars. By employing a fully Convolutional Neural Network (CNN) equipped with semantic segmentation capabilities, the system effectively processes images and videos captured by front-facing car cameras, discerning the intricate details of the road environment. This approach is indispensable as it ensures the precise identification and labeling of objects and elements within the scene, encompassing everything from road signs and vehicles to pedestrians and buildings.

Declaration

1. All authors do not have any conflict of interest.
2. This article does not contain any studies with human participants or animals performed by any of the authors.

References

1) Long, J., Shelhamer, E., & Darrell, T. (2015). Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

2) He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (ICCV).

3) Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2015). The Pascal Visual Object Classes Challenge: A Retrospective. International Journal of Computer Vision, 111(1), 98-136.

4) Brox, T., Bruhn, A., Papenberg, N., & Weickert, J. (2004). High accuracy optical flow estimation based on a theory for warping. In European Conference on Computer Vision (ECCV).

5) Perazzi, F., Pont-Tuset, J., McWilliams, B., Van Gool, L., Gross, M., & Sorkine-Hornung, A. (2016). A Benchmark Dataset and Evaluation Methodology for Video Object Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

6) Khoreva, A., Benenson, R., Omran, M., Hein, M., & Schiele, B. (2017). Lucid Data Dreaming for Multiple Object Tracking. In Proceedings of the IEEE International Conference on Computer Vision (ICCV).

7) Pont-Tuset, J., Arbelaez, P., Barron, J., Marques, F., Malik, J., & Van Gool, L. (2017). Multiscale Combinatorial Grouping for Image Segmentation and Object Proposal Generation. IEEE Transactions on Pattern Analysis and Machine Intelligence.

8) Brostow, G. J., Shotton, J., Fauqueur, J., Cipolla, R. (2008). Segmentation and recognition using structure from motion point clouds. In European Conference on Computer Vision (ECCV).

9) Perazzi, F., Krähenbühl, P., Pritch, Y., Hornung, A. (2012). Saliency Filters: Contrast Based Filtering for Salient Region Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

10) Hou, X., Yu, L., Wang, S., Gao, Z., Chen, D. (2016). A Novel Saliency Detection Model in RGB-D Image. IEEE Transactions on Image Processing.