Object Segmentation Using Machine Learning and Computer Vision Technique

Vaishali Aggarwal, Muzahid Islam, Muzammil Azeem and Chetan Mehta

May 12, 2023

Object Segmentation using Machine Learning and Computer Vision Technique

Vaishali Aggarwal, Muzahid Islam, Muzammil Azeem, Chetan Mehta

Department of Computer Science and Engineering, Chandigarh University, India

***Abstract*— In computer vision, the process of separating objects from their backgrounds in still or moving images is known as object segmentation. Due to the varying item looks, illumination, and occlusions, this work is difficult. Machine learning approaches have made considerable strides in object segmentation in recent years, delivering cutting-edge results on a number of benchmark datasets. This article presents an overview of current developments in computer vision and machine learning methods for object segmentation. In order to segment objects, standard computer vision techniques, including thresholding, region-based segmentation, and edge detection, are first reviewed. We discuss a range of popular deep learning techniques, such as convolutional network, Mask R-CNN, and U-Net, exploring their benefits and drawbacks. lastly, we analyze the current progress in object segmentation, from semi-supervised and unsupervised segmentation approaches to video object segmentation and domain adaptation. Additionally, we consider some of the unresolved challenges in object segmentation, including the issues of dealing with small objects, occlusions, and adapting to new domains.**

***Index Terms* — convolutional neural network (CNN), computer vision (CV), Deep learning (DL), Earth observation (EOb)**

***Keywords:* - artificial intelligence (AI), machine learning, deep learning, neural networks, convolutional neural networks, CNN, image segmentation, object detection, Earth observation,**

INTRODUCTION

Identification and separation of distinct objects from an image or video stream is the process of "object segmentation." Numerous computer vision applications, including driverless cars, security systems, and medical imaging, need the completion of this job. Machine learning techniques have enabled considerable improvements in object segmentation in recent years.

Deep learning algorithms in particular have demonstrated astounding performance in object segmentation tasks. These models are able to reliably recognize things in a variety of settings and circumstances because they learn from a significant quantity of annotated data. In addition, segmentation accuracy may be increased by combining machine learning models with computer vision techniques, including edge detection, region expanding, and watershed algorithms.

The most recent approaches to object segmentation utilising machine learning and computer vision will be examined in this research article. The many techniques and algorithms utilised in object segmentation, such as fully convolutional network, Mask R-CNN, and U-Net, will be briefly discussed. We will also go over the difficulties associated with object segmentation, including occlusion, object deformation, and complex backgrounds, as well as provide solutions to these problems.

We will investigate the techniques and evaluate the proficiency of multiple object segmentation models utilizing regular assessment metrics such as Intersection over Union (IoU), precision, and recall. we will contrast the performance of these models on benchmark datasets, pinpointing their advantages and drawbacks.

LITERATURE REVIEW

The technique of locating and separating various items within an image or a video sequence is known as "object segmentation." In many computer vision applications, including object recognition, tracking, and scene comprehension, this is an essential step. Deep learning approaches in particular have demonstrated astounding performance in object segmentation tasks in recent years. We will explore some current research papers on object segmentation using machine learning and computer vision techniques in this survey of the literature.

"Fully Convolutional Network for Semantic Segmentation" by Jonathan Long, Trevor Darrell, and Evan ShelHamer (2015)

For semantic segmentation, Fully Convolutional Network (FCN) are a well-liked deep learning architecture. It converts image classification networks into completely convolutional networks that can produce pixel-wise segmentation maps using a fully convolutional method. The authors demonstrated that FCN outperforms earlier approaches based on hand-operated functions and graphical models in order to attain state-of-the-art performance on the challenging PASCAL VOC 2012 dataset.

"Mask R-CNN" by Ross Girshick, Piotr Dollár, Georgia Gkioxari, and Kaiming He (2017)

A segmentation branch is added to the Faster R-CNN object identification framework using the deep learning architecture known as Mask R-CNN. A brand-new ROIAlign layer is included, enabling precise pixel-level alignment between maps and the input image. The authors demonstrated that Mask R-CNN attained cutting-edge performance on the COCO dataset, a significant object detection and segmentation benchmark.

"U-Net: Convolutional neural Networks for Biomedical Image Segmentation" by Philipp Fischer, Olaf Ronneberger, and Thomas Brox (2015)

The U-Net deep learning architecture was created for the segmentation of biomedical pictures, specifically for the segmentation of individual cells or nuclei in microscope images. An expanding path reconstructs the segmentation map from the feature representation in the architecture, while a contracting path converts the fed image into a low-dimensional feature representation. On a number of biomedical picture segmentation datasets, the authors showed higher performance.

"Deep Lab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs" by Liang-Chieh Chen, Alan L. Yuille, Kevin Murphy, George Papandreou, and Iasonas Kokkinos (2017)

Dilated convolutions and fully linked conditional random fields (CRFs) are combined in the Deep Lab deep learning architecture for semantic segmentation to increase segmentation accuracy. The network can now capture multi-scale context without adding further parameters thanks to the authors' innovative atrous convolutions variation of dilated convolutions. On a number of benchmark datasets, including PASCAL VOC 2012 and Cityscapes, the architecture delivered cutting-edge performance.

"Attention U-Net: Learning Where to Look for the Pancreas" by Steven McDonagh, Nils Y. Hammerla, Bernhard Kainz, Matthew Lee, Jo Schlemper, Loic Le Folgoc, Kazunari Misawa, Mattias Heinrich, Matthew Lee, (2018)

Attention U-Net is a variant of U-Net that includes an attentional mechanism to figure out where to focus on the input image for better segmentation performance. The authors applied the architecture to the task of pancreatic segmentation in CT scans and demonstrated improved performance compared to the baseline U-Net architecture.

"Adversarial Learning for Semi-supervised Semantic Segmentation" by Yunchao Wei, Xiaodan Liang, Yu-Wing Tai, Yau Pun Chen, and Chi-Keung Tang (2018)

A semi-supervised adversarial learning approach for semantic segmentation is suggested in this paper. The authors present a generator network that creates high-quality segmentation maps from unannotated images as well as a discriminator network that can tell the difference between segmentation maps that are actually formed and those that aren't. The authors showed that the suggested method outperformed the competition on a number of benchmark datasets, including PASCAL VOC 2012 and Cityscapes.

"Generative Adversarial Networks for Medical Image Segmentation: A Review" by Yuhao Cheng, Zongwei Zhou, and Junjie Cao (2020)

As shown in this review study, significant progress has been made in the use of using generative adversarial networks (GANs) to segment medical images. The authors talk about a number of GAN-based techniques for segmenting medical images, such as conditional GANs, adversarial autoencoders, and cycle-consistent GANs. They also compare the various methodologies and indicate potential future research possibilities.

Various object segmentation models will be examined in this study, and their efficacy will be evaluated using accepted metrics including Intersection over Union (IoU), precision, and recall. The models' performance on tested datasets will then be compared, and their advantages and disadvantages will be noted.

Fig. 1. Image to Model to Generate PredictWe will investigate the techniques and evaluate the proficiency of multiple object segmentation models utilizing regular assessment measure such as Intersection over Union (IoU), precision, and recall. we will contrast the performance of these models on benchmark datasets, pinpointing their advantages and drawbacks.

Fig 2. Masking the Results to our Image

METHODOLOGY

Research Methodology

Here is a Research Methodology that can be followed for object segmentation using machine learning and computer vision techniques:

Problem Definition: Define the problem statement and the scope of the research. Determine the type of object segmentation task, such as instance segmentation or semantic segmentation, and identify the specific object classes to be segmented.

Data Collection: collect a sizable dataset of images and the segmentation masks that go with them. Use publicly available datasets or collect data using custom data acquisition methods. Make sure your data set is diverse and representative of real world scenarios.

Pre-processing: Preprocess data by resizing images, normalizing pixel values, and applying transformations such as rotation, scaling, and reflection to enhance data.

Model Selection: Choose a suitable machine learning model for the task, similarly convolutional neural network (CNNs), and select an appropriate architecture based on the specific requirements of the task.

Training: Train the selected model using the pre-processed data. To avoid overfitting, experiment with various hyper parameters like learning rate and batch size and employ strategies like regularization.

Evaluation: Evaluate the trained model using standard evaluation metrics, similarly intersection-over-union (IoU) and mean average precision (mAP). Use validation data to tune the model and ensure that it generalizes well to new data.

Analysis: Analyze the results and identify the strengths and weaknesses of the model, Analyse errors to learn about different kinds of errors. your model has encountered and identify areas for improvement.

Optimization: Optimize the model by fine-tuning the hyper parameter and experimenting with different architectures and optimization algorithms.

Deployment: Deploy the trained model to perform object segmentation on new images. Integrate the model into a larger system, such as a robot or an autonomous vehicle, if applicable.

Machine Learning Techniques for Object Segmentation:

Object segmentation is the process of identifying and separating objects from the background in an image or video. Machine learning techniques can be used to perform this task automatically.

One popular approach of object segmentation is using deep learning model, similarly convolutional neural network (CNNs). convolutional neural network (CNNs) can be trained on large datasets of labelled images to learn how to recognize and segment objects. The most popular kind of CNN employed for this purpose is a fully convolutional network, which substitutes convolutional layers for the fully linked layers of a conventional CNN to enable pixel-level segmentation.

The training process involves feeding the CNN with input images and their corresponding segmentation masks, which show where the objects in the picture are located. The CNN learns to map input images to their corresponding segmentation masks, enabling it to segment new images it has not seen before.

Another approach is to use unsupervised learning techniques, such as clustering or thresholding, to segment objects based on their visual characteristics. These techniques can be useful when there are not enough labelled data available or when the objects in the image have very distinct visual features.

Overall, machine learning is a powerful tool for object segmentation, and its applications are widespread, including in fields such as medical imaging, robotics, and autonomous vehicles.

Convolutional Neural Network (CNNs) in Object Segmentation

Convolutional Neural Network (CNNs) are commonly used for object segmentation tasks. The following is a high-level overview of how CNNs can be used for object segmentation:

Pre-processing: The input image is pre-processed by resizing it to a standard size and normalizing the pixel values.

Convolutional Layers: An image that has been previously processed is fed into a CNN that has numerous convolutional layers. A collection of filters are applied by each convolution layer to the input image in order to detect various aspects of the image, such as edges, corners, and texture. Each convolution layer's output is passed on to the following layer.

Down sampling Layers: To decrease the spatial dimension of the feature map while keeping crucial information, the output of the convolutional layer is passed through numerous down sampling layers, also known as pooling layers.

Up sampling Layers: Once the feature maps have been reduced in size, the network then applies up sampling layers, also known as deconvolution layers, to increase the dimensions of the feature maps. The Aim is to create a segmentation mask that is identical to the size of the source picture.

SoftMax Layer: SoftMax, the network's final layer, produces a probability map of the input image. The likelihood of each pixel being the same for each class in the dataset is represented by that pixel on the probability map.

Loss Function: Using a loss function, the projected probability map is contrasted with the actual segmentation mask. The objective is to reduce the discrepancy between expected and actual segmentation masks.

Optimization: In order to optimize the network parameters and reduce the loss function, the network is trained using backpropagation and gradient descent.

Prediction: Once the neural network is trained, it can be used to anticipate segmentation masks for new images. The network receives the input image and outputs the segmentation mask as predicted.

Fig. 3. filtering and sorting with anchors visualizes each stage of the main stage, including presentations and thought networks. High- and low-quality anchors, as well as the improvement of the anchor field

Object Segmentation Datasets:

There are several publicly available datasets that can be used for object segmentation tasks:

PASCAL_(VOC): The PASCAL Visual Object Classes (VOC) dataset consists images of objects belonging to 20 different classes, including people, animals, vehicles, and household objects. Each image has pixel-level segmentation masks, object bounding boxes, and object class labels.

COCO: Images of objects from 80 different classifications, including common things, common people, and animals, can be found in the Common Objects in Context (COCO) dataset. Each image is notated with object bounding boxes and pixel-level segmentation masks.

Cityscapes: Images of urban street scenes with pixel-level annotations for object classes like vehicles, people, and buildings may be found in the Cityscapes dataset.

ADE20K: The ADE20K dataset contains images of indoor and outdoor scenes, with pixel-level annotations for 150 object categories such as furniture, plants, and vehicles.

ImageNet: The ImageNet dataset contains over a million images of objects belonging to 1,000 different classes. While it does not contain pixel-level segmentation annotations, it can be used for training object recognition models, which can then be adapted for object segmentation tasks.

Analysis

Here are some key analysis aspects that can be performed on object segmentation using machine learning and computer vision techniques:

Performance Evaluation: Use the proper measures, such as Intersection over Union (IoU), Mean Average Precision (mAP), and F1 Score, to assess the model's performance. The effectiveness of the model's ability to segment the objects and distinguish between various item classes will be assessed using these metrics.

Error Analysis: Analyse errors to learn what kinds of mistakes the model makes. This can help identify areas where the model needs improvement and guide the direction of future research. For example, if the model frequently misclassifies certain object classes or struggles to segment objects in certain scenarios, these areas can be targeted for improvement.

Complexity Analysis: To Analyze the model's complexity in terms of the quantity of parameters, the duration of inference, and the amount of memory used. This can help determine the practical feasibility of deploying the model in real-world scenarios.

Generalization Analysis: Examine the model's propensity for generalization to fresh, untested data. This can be accomplished by testing the model with data that isn't from the training set. A model that generalizes well will be able to perform well on the test set, while a model that overfits to the training set will perform poorly on the test set.

Comparative Analysis: Compare the performance of the developed model to existing state-of-the-art models in the literature. This can help determine how well the model performs relative to other models and identify areas for improvement.

Visualization Analysis: Visualize the segmentation results to gain insights into how the model is segmenting the objects. This can help identify areas where the model may be struggling or making errors and guide the direction of future research.

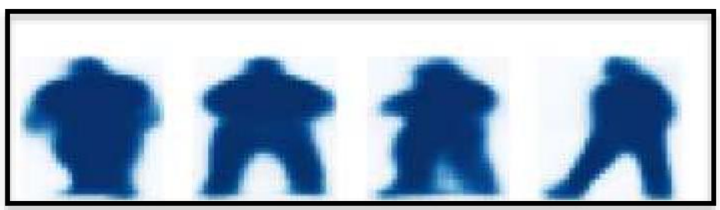


Fig. 4. the appropriately positioned on the image by means of mask generation.

OUTCOMES

The result of object segmentation using machine learning and computer vision technique can be depending on the specific problem, dataset, and model used.

Fig. 5. In the second step, Bounding Box Refinement is an example of final detecting boxes that is applied to them (dotted lines).

However, here are some typical results that can be expected from a well-designed and trained object segmentation model:

High Segmentation Accuracy: A well-trained object segmentation model should be able to accurately segment the objects in the input image. Metrics related Intersection over Union (IoU), Mean Average Precision (mAP), and F1 Score can be used to quantify this. High accuracy is crucial for applications such as medical imaging, autonomous driving, and robotics.

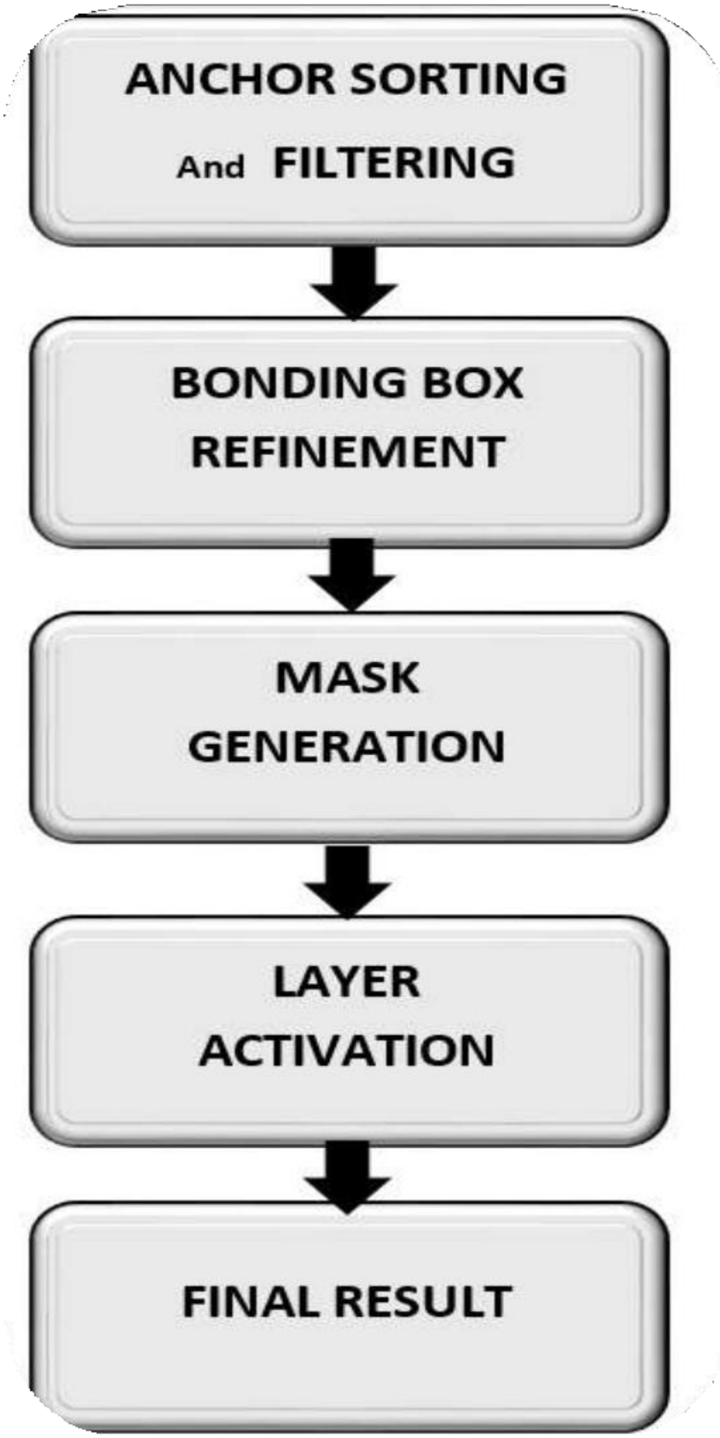
Fast Inference Time: The inference time of the model, i.e., the time it takes to segment an input image, should be fast enough to be used in real-time applications. This can be achieved through model optimization techniques such as model pruning, quantization, and compression.

Robustness to Variations: The model should be robust to variations in the input data, such as changes in lighting, pose, and occlusion. A robust model can perform well in a wide range of scenarios and can be deployed in real-world applications.

Generalization to New Data: The model need to have good generalizability to fresh, untested data. This can be assessed by putting the model to the test on a dataset that isn't the training set. A model that generalizes well can be used in a wide range of applications and can adapt to new scenarios.

Fig.6 activations of layers It can be helpful to frequently examine the activations at various layers to search for warning indicators of danger (all zeros or random noise). putting all the elements together to create the finished product

Performance over State-of-the-Art: A well-designed and trained object segmentation model should outperform existing state-of-the-art models in terms of accuracy, speed, and robustness. it can validate the effectiveness of the proposed approach and contribute to the advancement of the field.



flow of data

CONCLUSION

One of the most important tasks in computer vision is object segmentation, which has several uses. The traditional object segmentation techniques require manual input and are time-consuming, but with the advancements in The process can be automated using machine learning and computer vision techniques. Machine learning techniques such as supervised learning and unsupervised learning have been used to automate object segmentation, and CNNs such as U-Net, Mask R-CNN, and FCNs have been widely used for object segmentation. Object segmentation datasets such as PASCAL VOC, COCO, and ImageNet are used to train and evaluate machine learning algorithms.

In this article, we looked at the newest progress in computer vision and machine learning techniques for object segmentation. We looked into the restrictions of traditional methods for computer vision and examined several renowned deep learning models for object segmentation, outlining their pros and cons. Moreover, we investigated current developments in object segmentation and pointed out outstanding issues in this domain. We are sure that this guide will be useful for specialists and scholars who are focused on object segmentation.

REFERENCES

1. Deep Learning of Representations: Looking Forward. In Statistical Language and Speech Processing; Dediu, A.H., Martin-Vide, C., Mitkov, R., Truthe, B., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 1–37.

2. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. Nature 2015, 521, 436–444. [CrossRef]

3.. ImageNet Classification with Deep Convolutional Neural Networks. In Advances in Neural Information Processing Systems; Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2012; Volume 25, pp. 1097–1105.

4. Deep learning for computer vision: A brief review. Comput. Intell. Neurosci. 2018, 2018, 7068349. [CrossRef]

5. Shrestha, A.; Mahmood, A. Review of Deep Learning Algorithms and Architectures. IEEE Access 2019, 7, 53040–53065. [CrossRef]

6. Zhang, L.; Zhang, L.; Du, B. Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. IEEE Geosci. Remote Sens. Mag. 2016, 4, 22–40. [CrossRef]

7. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. IEEE Geosci. Remote Sens. Mag. 2017, 5, 8–36. [CrossRef]

8. Ball, J.E.; Anderson, D.T.; Chan, C.S., Sr. Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community. J. Appl. Remote Sens. 2017, 11, 1–54. [CrossRef]

9. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat. Deep learning and process understanding for data-driven Earth system science. Nature 2019, 566, 195–204. [CrossRef]

12. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Available online: https://www.tensorflow.org/ (accessed on 1 April 2020).

14. Bengio, Y.; Courville, A.; Vincent, P. Representation learning: A review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. 2013, 35, 1798–1828. [CrossRef]

15. Dahl, G.E.; Ranzato, M.; Mohamed, A.R.; Hinton, G. Phone Recognition with the Mean-Covariance Restricted Boltzmann Machine. In Proceedings of the 23rd International Conference on Neural Information Processing Systems—Volume 1; Curran Associates Inc.: Red Hook, NY, USA, 2010; pp. 469–477.

16. Dahl, G.E.; Yu, D.; Deng, L.; Acero, A. Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition. Trans. Audio Speech and Lang. Proc. 2012, 20, 30–42. [CrossRef]

17. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.E.; Mohamed, A.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.N.; et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. IEEE Signal Process. Mag. 2012, 29, 82–97. [CrossRef]

18. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. Science 2006, 313, 504–507. [CrossRef] [PubMed]