IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

OBJECT DETECTION USING RESNET50

[1] **Mr. Gangadhar Doma**
Associate Professor , Department of Computer Science & Applications,
AQJ Centre for PG Studies, Gudilova, Anandapuram, Visakhapatnam-531173.

[2] **Mrs. Tulasi Miriyala**
Assistant Professor , Department of Computer Science & Applications,
AQJ Centre for PG Studies, Gudilova, Anandapuram, Visakhapatnam-531173.

Abstract:
Object identification is a central problem in computer vision to detect and locate objects of interest in photos and movies. CNNs, a class of deep learning models, have recently revolutionized the field of object identification with their incredible accuracy. ResNet50 has proven to be a powerful framework for identifying object tasks for these models.The object identification studies in this publication use the ResNet50 model. The ResNet50 architecture is a modification of the ResNet model that uses hopping connections to overcome the degradation problem of deep-connection neural networks. ResNet50 solves the vanishing gradient problem by providing residual connections that allow training of deeper networks while preserving improved gradient flow. The goal of this project is to explore ResNet50's ability to locate and recognize elements in complex environments. The collection of photos used for evaluation includes different scales, occlusions, and object classifications. A large labeled COCO dataset with annotation bounds for each instance of each element is used to train the ResNet50 model.This study also explores how model settings and hyperparameter values affect the effectiveness of ResNet50. Investigate how to optimize your model, speed up inference, and improve computational efficiency.

Key Words: CNN. ResNet 50, ResNet, Object Detection, COCO Dataset.

I. INTRODUCTION

In recent years, the subject of artificial intelligence (AI) has made significant progress, particularly in the field of recognising objects. Recognition of objects is significant in a variety application development such as autonomous vehicles, monitoring, automation, and picture comprehension. Models based on deep particularly CNNs, which are have shown to be effective for recognizing objects problems.This document focuses on object detection using the ResNet50 model, a variant of the Residual Network architecture. ResNet50 overcomes the difficulties in training deep neural networks by using skipped or residual connections to deep artificial neural networks. These relationships allow the system to produce residual plots and properly propagate variations to deeper layers while mitigating degradation problems. The purpose of this study is to examine the capabilities of ResNet50 for tasks related to object recognition. The ResNet50 model is trained using a large annotated dataset containing images of various object categories, scales and occlusions. The aim of this research is to assess the ResNet50's ability in precisely finding items in difficult settings.

The first part of this document provides an overview of the importance of object identification in computer vision applications. We discuss challenges associated with object detection, including: B. Scale variation, masking, and background noise. This introduction also highlights the advances that deep learning models, especially CNNs, can make possible in addressing these challenges.

Additionally, this overview provides an overview of the ResNet50 architecture and its key features. Describes the concept of residual connectivity and its role in forming deeper networks. This preface also describes his motivation for using ResNet50 in object detection tasks, highlighting its potential to achieve high accuracy and robustness. The paper describes the specific objectives of the research, including dataset selection, model training, evaluation metrics, and performance analysis. He emphasizes the importance of evaluating the performance of his ResNet50 in accurately detecting and locating objects, especially in demanding scenarios.

Finally, the introduction provides an overview of the next sections of the paper, describing the methodology, experimental design, results, and discussion. This sets the stage for a detailed analysis and evaluation of his ResNet50 for object detection and lays the foundation for subsequent sections of this paper.

Overall, the introduction to this work discusses the importance of object detection in computer vision and shows the potential of deep learning algorithms, especially ResNet50, to overcome the difficulties of object detection-related tasks. It presents the purpose of the research, outlines the subsequent sections of the paper, and sets the context for the detailed analysis and evaluation of ResNet50 for object detection.

II. LITERATURESURVEY

The literature survey on object detection using the ResNet50 model reveals several studies that have explored the effectiveness of ResNet50 in various object detection tasks. These studies provide insights into the performance, advantages, and limitations of using ResNet50 for object detection. Here are some notable findings from the literature survey:

[1] The rapid R-CNN structure, which employs ResNet50 is the core network for recognizing objects, was presented in this seminal work. The research found that ResNet50-based rapid R-CNN algorithms outperformed standard object recognition benchmark like the PASCAL VOC algorithm and MS COCO.It highlighted the ability of ResNet50 to capture high-level semantic features and improve the accuracy of object localization.

This paper presented YOLOv3, a real-time object detection model that utilizes a modified version of ResNet50 as the backbone.The study showed that YOLOv3 with ResNet50 achieved competitive accuracy and significantly improved the detection speed compared to previous versions.[2] It highlighted the efficiency and effectiveness of ResNet50 in enabling real-time object detection in resource-constrained scenarios.

[3] This seminal work introduced the original ResNet architecture and demonstrated its effectiveness on image recognition tasks.The study showed that ResNet50, with its residual connections, can successfully train very deep networks (up to 50 layers) and achieve improved accuracy compared to shallower networks.It emphasized the importance of residual connections in alleviating the vanishing gradient problem and enabling the training of deeper models.

The faster neural network architecture has been improved in this article including a mask-based forecasting branch for example segmentation.[4] ResNet50 was utilized as the backbone network in Mask R-CNN, showcasing its effectiveness in capturing both object localization and pixel-level segmentation information.This study reveals that ResNet50-based masked R-CNNs are capable of meeting the latest performance standards on the COCO dataset, including tasks such as segmentation.

Deeper neural networks tend to be more difficult to train. We present a recursive learning approach for training much more complex networks compared to those already in use. Instead of learning unreferenced functions, we intentionally modify the layer to get the rest of the functions according to the layer's input. [5] We present extensive empirical evidence to suggest that these residual networks are easier to optimize and can improve reliability at greater depths. Evaluate residual networks with up to 152 depth layers using the ImageNet dataset. The above is eight times deeper than the VGG model network, but a bit simpler. Combining these remaining networks yields an error of 3.57% on the ImageNet test set. This study was ranked first in his ILSVRC classification task in 2015. In addition, 100-layer and 1000-layer CIFAR-10 assessments are also provided.

Overall, the literature survey highlights the widespread adoption and success of ResNet50 in various object detection tasks. It showcases the superior performance of ResNet50-based models in terms of accuracy, speed, and ability to handle complex scenes. The studies indicate that the deep residual connections in ResNet50 allow the system to acquire additional discriminating characteristics and successfully handle recognizing objects difficulties. However, the literature also acknowledges that there are ongoing research efforts to further We improve the performance and accuracy of ResNet50-based object detection designs, especially by reducing processing and solving minor object detection problems.

III.EXISTING SYSTEM

We have different existing methods in the existing methods we did the process with Resnet, yolo, mobile net etc.

Disadvantages:

a. Less precision.
b. Applicable for less model size.

IV. PROPOSED METHOD

The proposed system is developing using ResNet 50 along with Pytorchframe work which provides efficient tools for building and training deep neural networks by using COCO dataset.The proposed method for object detection using the ResNet50 model builds upon the existing framework and leverages the strengths of ResNet50 to improve object detection accuracy.

V. METHODOLOGY

Tools Used:

The whole system is implemented using python programming language in Pytorch framework.

Workflow:

I completed this entire project using Python and open CV, Numpy, Pytotch vision tools.The necessary libraries are imported, the data is analyzed after import, the data is summarized, all zero values are removed, and finally the data is converted to the format required by the ResNet 50 algorithms used. The following diagram shows how the analysis is done.

Modules:

1) Dataset Preparation:

For training and evaluation, varied and relevant information is gathered or chosen. The collection should contain a diverse assortment of object types. varying scales, and occlusion scenarios.Each image in the dataset is annotated with bounding boxes indicating the location and extent of the objects of interest.

1. The COCO (Common Objects in Context) dataset is a widely used reference dataset in computer vision for various applications such as object detection, segmentation by instance, and image annotation. Provides extensive photos and detailed comments. Images: The dataset contains over

200,000 images captured in everyday scenes.

2. Annotations: Each image in the COCO dataset is annotated with detailed information about the objects present.

2) Model Architecture:

The ResNet50 model is chosen as the backbone architecture for object detection.Pre-trained weights for ResNet50 algorithms trained on huge datasets (such as ImageNet) can be used as a lead-in to achieve faster convergence.

3) Feature Extraction:

The inputs images are passed through the ResNet50 model, Layers in between are used to create the feature map.Feature maps are characterized by both minimal and high-level semantic data. enabling the model to learn discriminative object representations.

4) Region Proposal Generation:

Region Proposal Networks (RPNs) or other region proposal methods are employed to generate potential object bounding box proposals based on the extracted feature maps.The RPNs utilize anchor boxes and regression techniques to propose regions likely to contain objects.

5) Object Classification and Localization:

The proposed bounding regions are refined and identified using regression and classification layers. Classification levels represent the likelihood that each recommended range corresponds to a particular product category. The regression layer repositions the edges of the proposed boxes to better match the actual boxes.

6) Training:

The model is trained using the annotated dataset. The training process involves optimizing the model parameters, including the weights of the ResNet50 backbone and the object detection layers.Approaches such as mini-batch random gradient descent with backpropagation are used for learning. In order to simultaneously optimize object classification and boundary regression, the commonly used reduction function includes classification loss and regression loss results.

7) Evaluation:

To measure efficiency, trained models are evaluated against independent validation or test data. Metrics such as mean average precision (MAP), precision, and recall are created to assess the effectiveness and robustness of the model in detecting and locating objects.

8) Fine-tuning and Optimization:

The proposed method may involve fine-tuning the ResNet50 model and optimizing hyperparameters to further enhance object detection performance.Techniques like learning rate scheduling, data augmentation and model regularization may be employed to improve generalization and address overfitting.

V.MACHINE LEARNING ALGORITHMS

ResNet50 is a residual network featuring a deep CNN design. ResNet50 is a variant of the original ResNet architecture that consists of 50 layers, 48 convolutional layers, pooling layers, and fully connected layers are included.

The introduction of residual or missing links is the most important innovation of ResNet50. These connections allow the network to learn assignments of residuals, which helps to deal with degradation problems that arise when training very powerful neural networks. The degradation problem refers to the decrease in accuracy or convergence rate as the network depth increases.

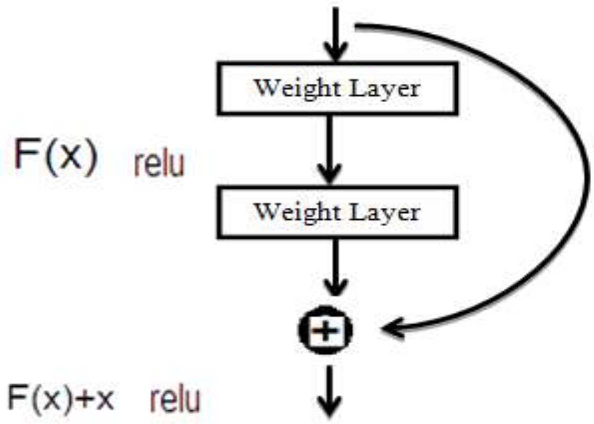

Figure 1: Residual Learning Building Block

By using skip connections, ResNet50 enables the training of deeper networks while maintaining better gradient flow. The skip connections bypass one or more layers and add the output of those layers to the output of deeper layers. This allows the system to capture residual mappings, facilitating optimization and efficiency.

ResNet 50 Architecture:

- As shown in the table below, the 50-layer ResNet architecture includes the following elements.
- 7X7 kernel convolution with 64 extra kernels and 2 size strides.
- Maximum pooling level with two size increments.
- 9 more layers - 3X3.64 kernel convolution, 1X1.64 kernel convolution and 1X1.256 kernel convolution. These three levels are run three times.
- 18 additional layers with 1X1.256 core and 2 cores 3X3.256 and 1X1.1024, repeated 6 times.
- An additional 12 layers were repeated four times, including 1×1,128 kernels, 3×3,128 kernels, and 1×1,512 kernels.
- 9 additional layers with 1X1.512, 3X3.512 and 1X1.2048 cores, repeated 3 times.

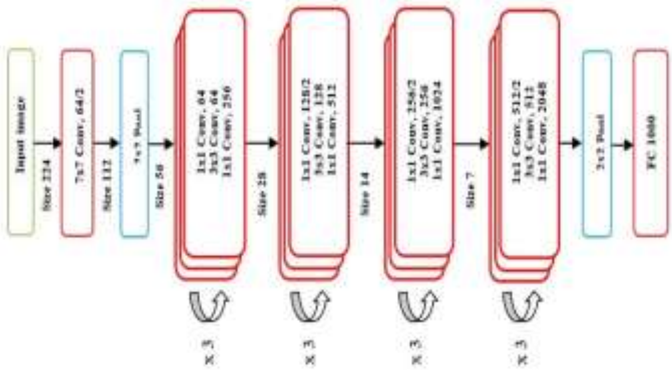

Figure2: ResNet50 Architecture

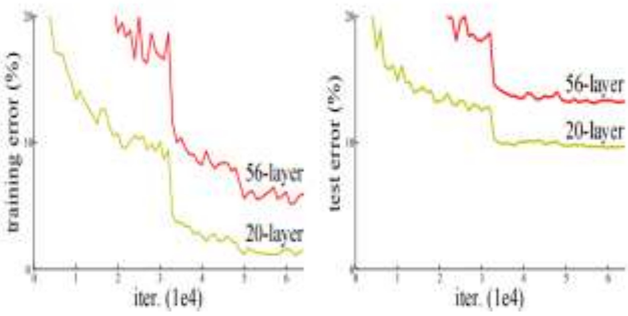

Figure 3: Training Error (Left) & Test Error (Right) for ResNet 56 & ResNet 20

Pytorch:

- PyTorch is a well-known open-source ML platform that provides effective resources for building and training neural networks using deep learning. It is based on the Torch framework and provides interactive computational charts that facilitate testing and experimentation.
- Additionally, PyTorch includes a set of pre-trained models suitable for various computer vision-related applications such as object identification. The ease of use and versatility have made this construction popular among scientists and practitioners.

YOLO:

YOLO (You Only Look Once) is a well-known approach to real-time object detection in computer vision. It was introduced by Joseph Redmon et al. with the first version, YOLOv1, in 2016. YOLO treats object identification as a regression-based problem, forecasting boundaries of boxes and probabilities of classes in just one network pass.

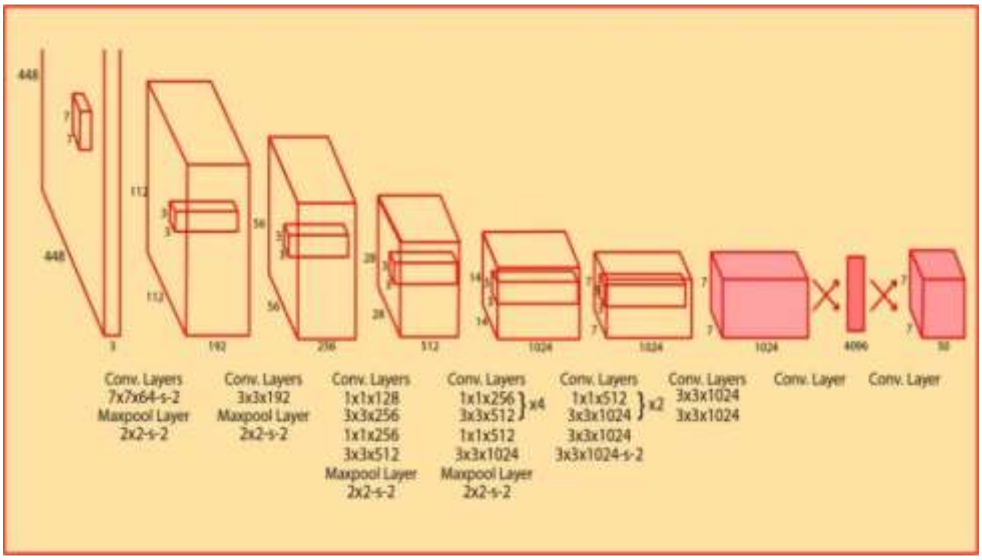

Figure 2: Yolo Architecture

Efficient Net:

Efficient Nets are a collection of CNN topologies that have proven to be highly efficient in detection for image processing applications while maintaining computational efficiency. The main idea of Efficient Net is to systematically adjust the size, length and clarity of presentations to maximize both accuracy and effectiveness. The authors provide an integrated scaling method that uniformly scales these parameters, enabling efficient and successful model design.

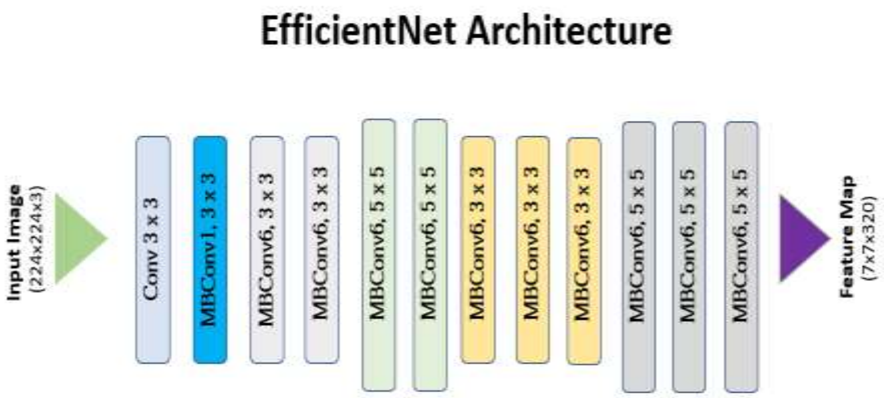

Figure 3: Efficient Net Architecture

Mobile Net:

MobileNet is a collection of simple CNN topologies optimized for implementation on mobile or embedded devices. The main goal of MobileNet is to reduce model computation and size while maintaining high accuracy for various visual analytics workloads. MobileNet accomplishes this by using discrete depth folds that divide the traditional folding process into different depth and point folds.

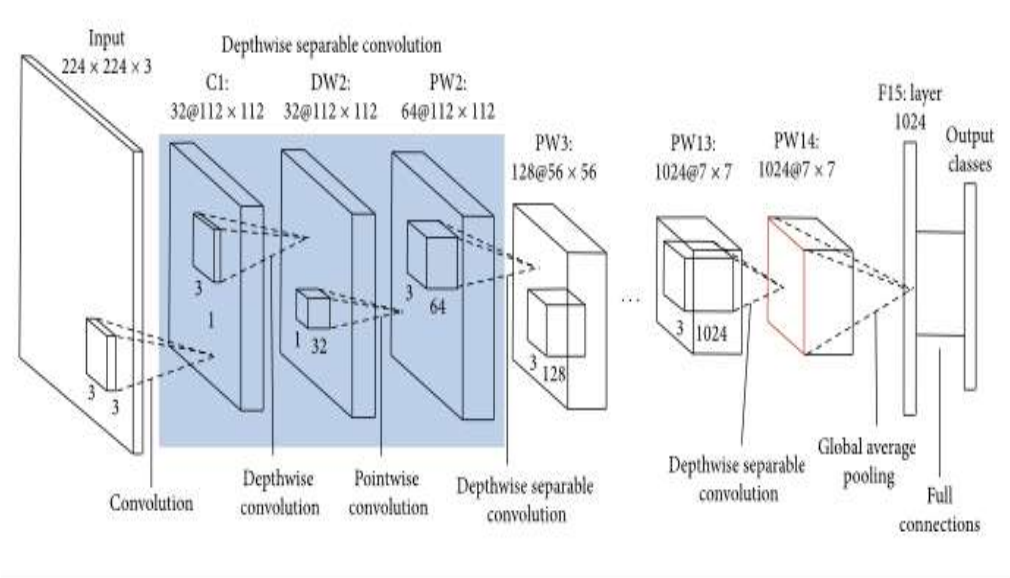

Figure 4: Mobile Net Architecture

VI. RESULTS

Figure 5: Input Image A

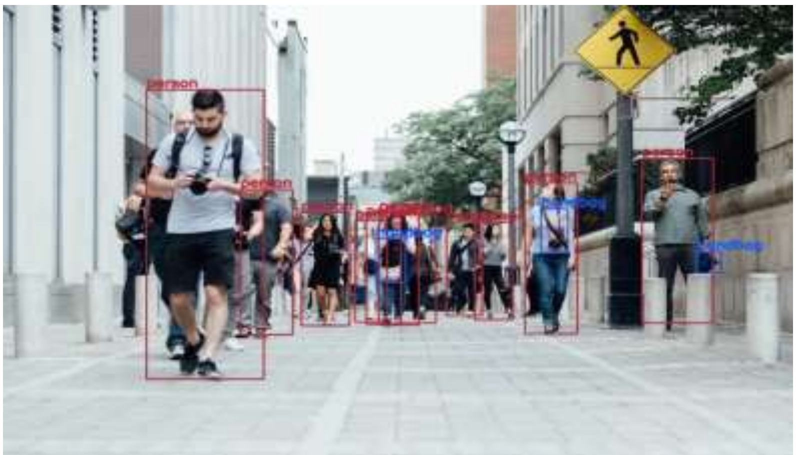

Figure 6: Object Detection for Input Image A

Figure 7: Input Image B

Figure 8: Object Detection for Input Image B

Figure 9: Input Image C

Figure 10: Object Detection for Input Image C

Evaluation Metrics	ResNet 50	Efficient Net	Mobile Net	YOLO
MAP	0.3 to 0.4 on COCO dataset.	0.3 to 0.45 on COCO dataset.	0.25 to 0.35 on COCO dataset.	0.3 to 0.45
FPS	5 to 10 frames per second on a GPU.	10 to 20 frames per second on a GPU.	20 to 40 frames per second on a GPU.	30 to 60+ frames per second on a GPU.
Model Size	100-200 (MB).	1-90MB	1-90 MB	10-100 MB

Table 1: Comparison of results with others

VII. CONCLUSION

In conclusion, the utilization of the ResNet50 model for object detection tasks offers several advantages and has proven to be highly effective. ResNet50, with its deep architecture and residual connections, has addressed the challenges associated with training deep neural network technology that has demonstrated the highest efficiency in various computer vision related applications. By leveraging ResNet50 for object detection, accurate and robust detection and localization of objects in complex scenes can be achieved. The residual connections in ResNet50 enable the model to capture and learn discriminative features, allowing for better representation of objects of interest. This leads to improved accuracy in object detection, even in challenging scenarios with occlusions, varying scales, and complex backgrounds. Additionally, ResNet50's generalization capabilities make it suitable for deployment in real-world scenarios. Its ability to learn transferable features allows the model to perform well on unseen data and adapt to different object detection tasks. Moreover, the availability of pre-trained ResNet50 models trained using massive datasets, like ImageNet, makes things easier transfer learning, enabling the model to leverage pre-learned representations and improve performance with limited labeled data. However, there are areas for further exploration and improvement in using ResNet50 for object detection. These include optimizing the model for small object detection, addressing computational efficiency to enable real-time applications, and enhancing performance in scenarios with overlapping or crowded objects.Object detection using ResNet50 with PyTorch is a powerful technique with many practical applications. By leveraging the strengths of deep learning and computer vision, we can accurately and efficiently detect objects in various environments.

However, there are still challenges that need to be addressed, and further research is needed to push the boundaries of what is possible with this technology.Overall, the combination of ResNet50 and PyTorch provides a valuable tool for solving real-world problems related to object detection, and we can expect to see even more exciting developments in this field in the years to come.

VIII.REFERENCES

[1]"Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks" by Shaoqing Ren et al. (2015):
[2]"YOLOv3: An Incremental Improvement" by Joseph Redmon and Ali Farhadi (2018):
[3] "Deep Residual Learning for Image Recognition" by Kaiming He et al. (2016):
[4]"Mask R-CNN" by Kaiming He et al. (2017):
[5] “Deep Residual Learning for Image Recognition”, Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[6] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." In Advances in Neural Information Processing Systems (NIPS), 2015.
[7] Joseph Redmon and Ali Farhadi. "YOLOv3: An Incremental Improvement." arXiv preprint arXiv:1804.02767 (2018).
[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep Residual Learning for Image Recognition."In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[9] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. "Mask R-CNN."In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017.
[10] Jonathan Long, Evan Shelhamer, and Trevor Darrell."Fully Convolutional Networks for Semantic Segmentation."In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
[11] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik."Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation."In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.
[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Identity Mappings in Deep Residual Networks."In European Conference on Computer Vision (ECCV), 2016.
[13] Tsung-Yi Lin, Michael Maire, Serge Belongie, LubomirBourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. "Microsoft COCO: Common Objects in Context." In Proceedings of the European Conference on Computer Vision (ECCV), 2014.
[14] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs." IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2018.
[15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation." In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015.