ISSN: 2664-9640

# Development and Enlightenment of Global Image Segmentation Technology

Xiaoyan Huang

College of Electronic and Information, Southwest Mingzu University, Chengdu 610041, China.

## Abstract

**In recent years, deep learning has achieved great success in natural language processing, autonomous driving, computer vision, multimedia and other fields. Image vision accounts for 80% of the input of all human sensory perception information, and is the most common and important perceptual link. If artificial intelligence is a technological revolution, the center of revolution must occur in the field of computer image vision, not in other fields.This paper will focus on the basic AI image classification technology and the most complex AI image segmentation technology.**

## Keywords

**Deep Learning, Computer Image Vision, AI Image Classification, AI Image Segmentation.**

## 1. Introduction

With the development of computer technology and the improvement of software and hardware strength, artificial intelligence technology is in full swing, and the purpose of artificial intelligence is to let computers learn to see, hear and read with powerful computing power, and then understand various forms of information in the real world. In other words, the computer's understanding of images, speech, and text basically constitutes our current artificial intelligence. Deep learning is a major breakthrough in the field of artificial intelligence in the past decade, and the existing deep learning model belongs to the computer neural network, which is influenced by the nerves of the human brain.

Metaprocessing information and cognitive things are inspired to arise and are used to solve various machine learning problems. In recent years, deep learning has achieved great success in natural language processing, autonomous driving, computer vision, multimedia and other fields. Image Vision accounts for 80% of the input of all human sensory perception information, and is the most common and important perceptual link. Therefore, if artificial intelligence is a technological revolution, the center of the revolution must occur in the field of computer image vision, and not in other fields. As one of the most important information transmission carriers in human life, images are the most intuitive and clear way to display various scenes, objects and events in real life. Therefore, image understanding technology is widely applied in various fields of human production and life. The vigorous development of science and technology and the increasing demand of human beings are the strongest driving forces for development in this era, which has pushed various industries into the wave of artificial intelligence. In addition to promoting the development of the machine vision field, the development of artificial intelligent image technology can also be widely used in other information technology fields such as big data and the Internet of Things, creating a variety of integrated applications. Therefore, artificial intelligence image technology has gradually become a major mainstream branch in the field of science and industry. When the human brain needs to understand the image information, it will first identify various objects, especially key objects, in specific scenes according to past experience and knowledge, and on this basis, combine the relationship between the objects to understand the semantics of the entire image and then make a judgment. Artificial intelligence

ISSN: 2664-9640

is designed to imitate the memory, recognition and learning capabilities of the human brain with the powerful computing power of computers. But it's not easy to accurately model the human brain. In recent years, with the joint efforts of domestic and foreign scientific researchers, artificial intelligence image understanding technology has continued to develop and progress, and the image understanding effect has been continuously improved, gradually meeting people's needs for image understanding in various scenarios.

## 2. The Current Status of the Development of Artificial Intelligence Image Technology

The current artificial intelligence image technology is mainly based on the principle of neural network deep learning, relying on meticulous algorithms and powerful computing hardware to achieve image understanding. Deep learning uses multi-layer linear and nonlinear transformations to make the network have strong learning ability and expression ability, especially in extracting implicit abstract features and environmental information. In an image, various implicit factors are often associated in a complex manner and relatively abstract, while deep learning can extract the complex features implied by the image through multi-layer transformation and processing of the image, so that image understanding becomes simple. Therefore, the application of deep learning to image tasks can not only save manpower and material resources, but also improve the efficiency and accuracy of task execution.

### 2.1. International Progress

In 1986, the Journal of Nature published a well-known backpropagation algorithm for training neural networks proposed by Rumelhart, Hinton, and Williams, which is still widely used today. Later, with the rapid development of computer hardware, the emergence of big data, and the advancement of model design and training methods, neural networks have been booming. In 2012, Hinton's research team used deep learning to build the AlexNet network structure, the most influential breakthrough in deep learning in the field of computer vision. Hinton currently holds a key role in Google Brain, and under his leadership, Google's image recognition and Android audio recognition performance has been greatly improved. "Deep learning" has changed from a marginal topic to a core technology relied on by Internet giants such as Google, and since then deep learning has flourished in the field of computer vision.In a variety of applications, deeper and more complex neural network structures are constantly being proposed. At the same time, the ImageNet dataset has become a common dataset in the field of images, and the milestone full-convolutional neural networks (FCN) have been developed. FCN is the first network to implement end-to-end deep neural network training on variable graphs to realize image semantic segmentation. A semantic segmentation model based on a combination of convolutional networks and fully connected CRFs enables more accurate positioning of object boundaries.

### 2.2. Domestic Progress

At the beginning of domestic progress, some traditional image segmentation methods were used in China, which were often used for the segmentation of medical images, satellite images, and other simple images. The traditional image segmentation method is mainly based on the spatial characteristics of the image, as well as the regional distribution of each object in the image to achieve image segmentation, such as the image texture, spatial structure and color and other characteristics, the common methods are threshold-based segmentation method, area-based segmentation method, edge-based segmentation method, and based on some specific theories of segmentation methods. These methods often rely on people's prior knowledge to select parameters and design models, and the segmentation effect of complex classes is poor, consuming a lot of manpower and material resources. Later, with the development and

ISSN: 2664-9640

application of deep learning technology, many image methods based on deep learning have emerged, and deep learning has greatly simplified the process of performing image understanding and can produce higher quality results. At present, the major Internet companies in China are also focusing on the development of artificial intelligence. In upstream, Cambrian and Horizon assisted in chip production; Sagitar innovation and Radium intelligent power sensor manufacturing; Baidu, Alibaba, Tencent and Huawei help cloud computing, provide powerful computing brains for human intelligence, share computing platforms and allocate resources rationally; Baidu is also powering artificial intelligence in system architecture. In the middle reaches, Megvii Technology, SenseTime, Tuyang Information, Geling Deep Pupil and Yitu Technology are committed to technology research and service. In downstream industry applications, the above companies have their own deployments in various key artificial intelligence image application fields such as intelligent driving, intelligent security, intelligent medical care, and smart home.

## 3. Key Technology in the Field of Artificial Intelligence Image Analysis

Image understanding tasks based on artificial intelligence image technology Mainly have three core tasks: image classification, object detection, and image segmentation. Image classification aims to classify the content in the image, and object detection aims to identify the location and category of the target of interest in the image; The image segmentation tasks mainly include image semantic segmentation, image instance segmentation, and newborn image panorama segmentation tasks. Image semantic segmentation aims to classify the entire image at the pixel level, that is, to classify each pixel in the image with a clear semantic category label. Instance segmentation and semantic segmentation have similarities, but it only detects the instance target in the image and segments it, at the same time, it is necessary to distinguish between different instances, it can be seen that the difficulty of image segmentation is higher than image classification, it can not only be applied to flat or cubic static images, but also applied to video processing. The further need for image understanding has given rise to a more comprehensive task - image panorama segmentation. Image panorama segmentation can be seen as a combination of semantic segmentation and instance segmentation of two tasks, for the category of uncountable objects, distinguish their pixel categories, for countable objects categories, not only to identify the semantic category to which the target belongs, but also to distinguish each instance. The task can provide more comprehensive scene information and can be widely used in the understanding of various natural scenes.

### 3.1. Artificial Intelligence Image Classification Technology

Image classification is an important topic in the field of artificial intelligence images. In traditional machine learning methods, the standard process for identifying and classifying various images is to first extract features, then screen features, and finally send feature vectors into a suitable classifier to achieve feature classification. The AlexNet network structure proposed by Hinton's research team consisting of 5 convolutional layers plus 2 layers of fully connected layers integrates the above three steps with the help of deep learning, and carries out information mining and feature extraction of different angles of the input image layer by layer. AlexNet won the 2012 ILSVRC Challenge with a low error rate of 15.4%, and convolutional neural networks (CNNs) have since become the dominant method in image classification. Some of the current better performing image classification networks such as GoogleNet and ResNet networks are deep convolutional neural networks. The convolutional Neural Network (CNN) was inspired by Hubel and Wiesel's study of the sensory field in the cat visual cortex, and its earliest application was the handwritten digital recognition problem, which has been excellent in the treatment of this problem, which is a feed-forward neural network. At present, convolutional neural networks are used in almost all visual tasks. ResNet is not only used for image classification, but also for typical backbones used in feature extraction.

## 3.2. Artificial Intelligence Image Segmentation Technology

Image segmentation technology based on artificial intelligence mainly includes semantic segmentation, instance segmentation and panoramic segmentation. It can be seen that image semantic segmentation is to classify images at the pixel level, and instance segmentation only focuses on the recognition and segmentation of several instance-level objects, while panoramic segmentation is to combine the above tasks to classify and identify objects in the whole image.

### 3.2.1. Artificial Intelligence Image Semantic Segmentation Technology

Image semantic segmentation aims to classify the entire image from the pixel level, that is, to distinguish a semantic category label for each pixel in the image. With the application of full-convolutional networks in deep learning to image semantic segmentation tasks, semantic segmentation tasks have developed as a typical dense prediction model. Image semantic segmentation technology plays an important role in many life scenarios. In medical image processing, image semantic segmentation can help doctors to identify diseased tissue, thus providing a more reliable medical solution; In automatic driving, the street traffic scene can be divided and understood, and the road, traffic sign, people and vehicles can be divided, so as to provide drivers with information on assisted driving, and effectively improve driving efficiency and safety; In face recognition, image semantic segmentation technology realizes face identity recognition by determining the contours of the face, facial features and hair, and can even conduct emotional analysis, or estimate gender.In smart agriculture, image semantic segmentation is possible The intelligent identification of crops and other weeds, so as to accurately control the weeding behavior of weeding machinery, can greatly reduce the agricultural labor workload, improve agricultural production efficiency and reduce production costs. In addition, image semantic segmentation technology can also be applied in clothing analysis, natural scene understanding, and military reconnaissance and other fields. Semantic segmentation is gradually entering people's production and life in various ways, changing people's lifestyles and living habits, and how to accurately and quickly perform image segmentation has become a major difficult point in the research process.

The population needs to go through three stages, namely feature extraction, feature fusion, and upsampling prediction. The feature extraction stage is mainly based on operations such as convolution to extract deep abstract semantic information in the image. The feature fusion stage is mainly to further improve the semantic segmentation performance of images, which is divided into two ways: early fusion and late fusion. The main purpose of the upsampling stage is to upsample the features to the size of the original plot, thereby outputting the final semantic segmentation prediction plot, and the commonly used upsampling methods are mainly bilinear difference method and deconvolution method.

### 3.2.2. Artificial Intelligence Image Instance Segmentation Technology

Mask R-CNN technology. Most instance segmentation methods divide the task into two steps, first making a region proposal, and then classifying and segmenting the proposed area, which makes them unable to obtain real-time speed even when the image size is reduced. The single-stage instance segmentation method generates a location-sensitive map that is assembled into the final mask by location-sensitive pooling or by combining semantic segmentation logic and direction prediction logic. These methods still require some time-consuming post-processing and cannot be done in real time. Yolact technology. Learning prototypes (also known as "glossaries" or "codebooks") have been widely explored in the field of computer vision. Classical representations include text and visual vocabulary, and prototypes for target detection have been designed in their work, using prototypes to represent features.

The Yolact network uses a prototype assembly mask to perform instance segmentation, and the prototype is specific to each image. At the same time, the prototype in Yolact has an interesting emergence behavior, that is, through different activations in the prototype, the network can locate the instance itself, that is, the prototype also has a positioning function, and even a large object can be directly positioned without relying on the box branch. Yolact splits the instance segmentation task into two parallel subtasks, and after extracting the features, it sends the features to the prediction branch and the prototype branch, predicts the branch output, detects the target category, position and corresponding mask coefficient, and integrates the output of the pre-test branch and the prototype branch through the mask integration operation to obtain the final instance segmentation prediction result.

3.2.3. Artificial Intelligence Image Panoramic Segmentation Technology

The common panoramic segmentation method of artificial intelligence image panoramic segmentation technology adopts a shared backbone network structure based on multi-task learning. The final step of panoramic segmentation is the back-end integration of semantic segmentation results and instance segmentation results, which have two main problems of partial instance mask overlap and partial pixel category conflicts, and most of the panoramic segmentation methods use heuristic rules to solve these problems.

Shared trunk method. Subtask information federation method. Since semantic segmentation and instance segmentation tasks have many similarities, some methods will combine the information of each subtask to enhance the consistency between subtasks, thereby alleviating pixel conflicts and other problems during back-end fusion, and ultimately improving the panoramic segmentation effect. Learnable de-overlapping methods. When performing back-end fusion, the instance segmentation result includes overlapping areas, and the panoramic partition does not allow the existence of overlapping areas, which is a major problem of the back-end fusion surface, in the existing panoramic segmentation study, the method to solve the masking problem includes using heuristic rules to determine the masking order of each object, and predicting the object to which each pixel belongs. The SOGNet approach addresses the key issue of back-end integration. In addition to this, OANet can efficiently and efficiently predict instances and semantic segmentation in a single network.

# 4. Challenges and Problems Facing the Development of Artificial Intelligence Images in China

Under the rapid development of artificial intelligence deep learning, image technology has also made rapid progress, playing a huge role in automatic driving, security inspection monitoring, and medical image analysis. However, human beings' continuous pursuit of application experience is still driving the development of artificial intelligent image technology, which has broad prospects and many problems.

## 4.1. More Image Data Requirements

More image data demand in the "data, algorithm, computing power" of the three major human intelligence development of the driving force, the most important thing at the moment is the data, data sets in artificial intelligence has a pivotal position, the current data sets for image semantic segmentation are mainly Cityscapes, Microsoft COCO and PASCAL VOC series. Experiments and competitions on these datasets are refreshed every year, but although the models show high performance on the datasets, they may not be universal in real-life scenarios, in addition to the time and labor costs of expanding the datasets. To save costs, the researchers began exploring semantic segmentation of semi-supervised and weakly supervised, using unlabeled data to train models together with labeled data, improving semantic segmentation

ISSN: 2664-9640

performance or using annotations that are weaker than pixel-level annotations to oversee semantic segmentation networks.

### 4.2. Higher Image Dimension Requirements

In real-life applications, traditional 2D image semantic segmentation still has certain limitations, and it is difficult to directly know the spatial position of the object from the 2D image, as well as its layout in the overall space. At present, the three-dimensional point cloud can characterize the three-dimensional information of the image, in order to obtain the semantic information of the point cloud in the scene, the semantic segmentation technology based on the point cloud has been developed.However, 3D data often faces extremely expensive computational costs and the cost of data acquisition and annotation, and technicians find a balance between the cost of solving accurate 3D information and the limitations of traditional 2D segmentation, that is, using the top-down semantic map to perceive the orientation and layout of the surrounding environment objects, and this process of obtaining a top-down semantic map from the first-view 2D image is called cross-perspective semantic segmentation.

### 4.3. Faster Processing Speed Demand

In the application of automatic driving, semantic segmentation technology is mainly used to segment the image received by the vehicle, assist the vehicle for automatic navigation, obstacle recognition and traffic sign recognition, etc., so it has high requirements for the running speed, storage space, and accuracy of the semantic segmentation algorithm. In addition, other tasks, including face segmentation, also have high requirements for the real-time semantic segmentation, in order to achieve the deployment of real-time applications, high-precision speed is one of the directions of semantic segmentation technology.

## References

[1] H.T. Lu, M.K. Luo: New Progress in Computer Vision Research Based on Deep Learning, Data Acquisition and Processing, Vol. 37 (2022) No.02, p.247-278.(In Chinese)

[2] Y.N. Geng, S.S. Liu, T. Liu: Review of Pedestrian Detection Techniques Based on Computer Vision, Computer Applications, Vol. 41 (2021) No.01, p.43-50.(In Chinese)

[3] Y.P. He: Exploration on the Image Processing Technology of Computer Vision Algorithms, Electronic World, Vol. 16 (2021) No.04, p.96-97.(In Chinese)

[4] X.C. Wu: Research on Image Correction Technology under Computer Vision Algorithm, Journal of Heihe College, Vol. 22 (2020) No.11, p.184-185.(In Chinese)

[5] X.Q. Ma: Research on Image Recognition and Recovery of Computer Vision Technology, Electronic Technology and Software Engineering, Vol. 48 (2020) No.23, p.120-121.(In Chinese)

[6] Q. Cui: Research on Image Processing Technology Based on Computer Vision Algorithm, Computer Products and Circulation, Vol. 15 (2020) No.05, p.111.(In Chinese)

[7] Y. Li: Based on Computer Figure Image Design and Visual Communication Design, Satellite TV and Broadband Multimedia, Vol. 31 (2020) No.05, p.61-62.(In Chinese)

[8] J.H. Cheng, G.H. Zeng, J. Liu: Research on Complex Background Image Classification Methods Based on Deep Learning , Electronic Technology, Vol. 33 (2020) No.12, p.59-66.(In Chinese)

[9] G.F. Li: An Analysis of Image Processing Technology Based on Computer Vision Algorithm, Information Record Materials, Vol. 20 (2019) No.06, p.125-126.(In Chinese)

[10] S. Zhao: Analysis and Research of Content-based Image Division and Classification Methods, Technology Communication, Vol. 37 (2010) No.18, p.242-243.(In Chinese)