**图像分割技术**

目录

[1 前言 2](#_Toc118466948)

[1.1 概述 2](#_Toc118466949)

[1.2 形式化 2](#_Toc118466950)

[2 基于视觉的图像分割 4](#_Toc118466951)

[3 基于学习的图像分割 5](#_Toc118466952)

[3.1 导论 5](#_Toc118466953)

[3.2 基于聚类的图像分割 5](#_Toc118466954)

[3.3 基于朴素贝叶斯的图像分割 5](#_Toc118466955)

[3.4 基于马尔科夫网的图像分割 6](#_Toc118466956)

[4 基于网络的图像分割 8](#_Toc118466957)

[引言 8](#_Toc118466958)

[图像分割的分类 8](#_Toc118466959)

[图像分割的发展脉络 8](#_Toc118466960)

[深度学习方法中主要的模型分类 8](#_Toc118466961)

[前置知识（Preliminary）——常用的基础网络模型 9](#_Toc118466962)

[Convolutional Neural Networks (CNNs) 9](#_Toc118466963)

[Recurrent Neural Networks (RNNs) and the LSTM 9](#_Toc118466964)

[Encoder-Decoder and Auto-Encoder Models 9](#_Toc118466965)

[Generative Adversarial Networks (GANs) 9](#_Toc118466966)

[经典模型 9](#_Toc118466967)

[Fully Convolutional Networks (FCNs) 9](#_Toc118466968)

[U-Net 10](#_Toc118466969)

[Feature Pyramid Network (FPN) 10](#_Toc118466970)

[Mask R-CNN (Regional CNN) 10](#_Toc118466971)

[未来展望 11](#_Toc118466972)

[参考文献 12](#_Toc118466973)

# 1 前言

## 1.1 概述

在图像处理与计算机视觉领域，**图像分割(image segmentation)**是在像素级别将一个完整图像划分为若干具有特定语义**区域(region)**或**对象(object)**的过程。每个分割区域是一系列拥有相似特征——例如颜色、强度、纹理等的像素集合，因此图像分割也可视为以图像属性为特征空间，为全体像素赋予标签的分类问题。

图像分割是高级图像处理的基础技术，它将原始冗余而繁杂的图像，转化为一种更具意义且简单紧凑的组织形式。在智能安防、卫星遥感、医学影像处理、生物特征识别等领域[1]，图像分割通过提供精简且可靠的图像特征信息，有效地提高后续从而利于后续图像分析、理解等技术的计算效率，具有重要意义。如图1.1.1所示，在自动驾驶领域，通过图像分割可以快速识别出车道线、指示牌、交通信号灯等重要交通信息。

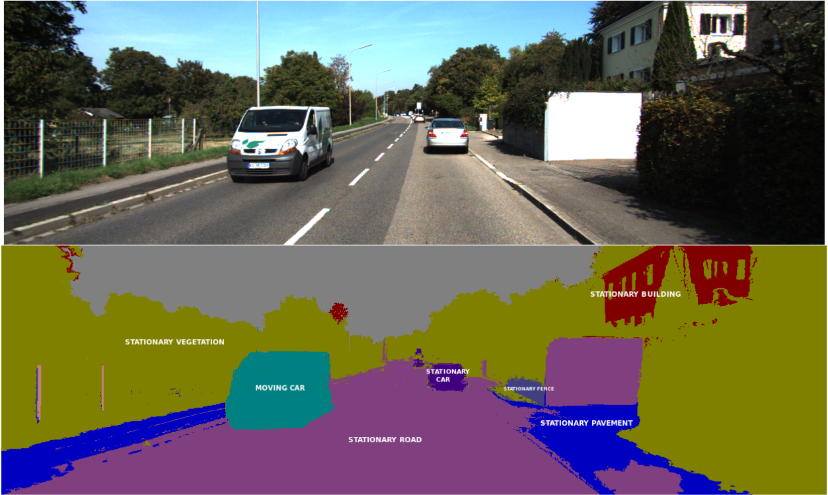


图1.1.1 图像分割在自动驾驶领域的应用

## 1.2 形式化

将一幅的图像视为全体像素的集合



其中表示位于坐标处的像素，则图像分割的目标就是将其划分为若干非空子集，满足：

1. **完备性**。满足，说明图像中所有像素均被分割；
2. **独立性**。满足对有，说明子集彼此独立，图像中不存在被重复分割的像素；
3. **一致性**。满足对有，其中表示对区域像素某个属性进行逻辑判断，说明分割得到的子集内部存在相似特性；
4. **互斥性**。满足对有，说明不同子集间的像素不存在相似特征；
5. **连通性**。对有是连通的，即区域内任意像素都存在到其他像素的通路。

# 2 基于视觉的图像分割

阈值(根据灰度直方图分割)、边缘检测、形态学算子

# 3 基于学习的图像分割

## 3.1 导论

## 3.2 基于聚类的图像分割

## 3.3 基于朴素贝叶斯的图像分割

## 3.4 基于马尔科夫网的图像分割

在图像分割中，通常默认图像中某像素点只受相邻像素的影响，较远处的像素对该像素没有作用，或者说其作用已被包含在相邻像素内，例如当前像素语义是天空，那么近邻像素也很可能表示天空。形式化地，像素的邻域定义为



其中表示两个像素间的欧式距离，表示的是邻域的阶次，阶次越高像素包含的邻点越多，且满足当阶次时，；此外，邻域满足互易性，即有。常用的邻域如图3.2.1所示。



(a) 一阶邻域 (b) 二阶邻域 (c) 五阶邻域

图3.2.1 常见的邻域系统

这种邻域特性类似于马尔科夫链的无后效性，即任意一个状态包含了所有历史状态的信息，每个状态的推演只与上一个状态有关。由于图像是二维数据，因此用经典的无向图模型——马尔科夫随机场代替一维的马尔科夫链进行建模，如图3.2.1所示。马尔科夫随机场中的全局马尔科夫性、局部马尔科夫性和成对马尔科夫性，恰好表征了像素只受邻域影响的假设偏好。

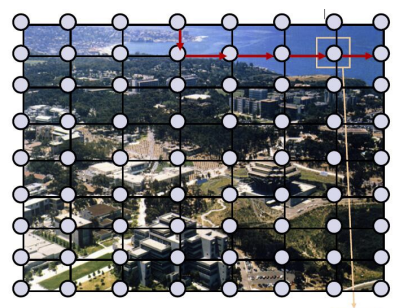


图3.2.1 马尔科夫随机场

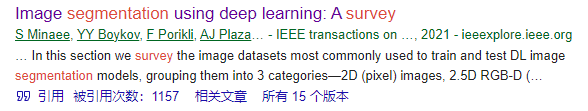
马尔科夫随机场是一个偶对，其中

* 网络结构是无向图，是图形中所有节点——随机变量的集合；是所有连边——变量间相关依赖的集合；
* 概率分布是在网络结构上的因子分解，由Hammersley-Clifford定理可得



其中是图中全体极大团集合，是极大团对应的随机变量集合，因子在马尔科夫随机场中也称为团位势。

# 4 基于网络的图像分割



## 引言

### 图像分割的分类

语义分割、实例分割和全景分割。

Image segmentation can be formulated as the problem of classifying pixels with semantic labels (semantic segmentation), or partitioning of individual objects (instance segmentation), or both (panoptic segmentation).

### 图像分割的发展脉络

Numerous image segmentation algorithms have been developed in the literature, from the earliest methods, such as thresholding [4], histogram-based bundling, regiongrowing [5], k-means clustering [6], watershed methods [7], to more advanced algorithms such as active contours [8], graph cuts [9], conditional and Markov random fields [10], and sparsity-based [11], [12] methods. In recent years, however, deep learning (DL) models have yielded a new generation of image segmentation models with remarkable performance improvements, often achieving the highest accuracy rates on popular benchmarks (e.g., Fig. 1). This has caused a paradigm shift in the field.

### 深度学习方法中主要的模型分类

1) Fully convolutional networks

2) Convolutional models with graphical models

3) Encoder-decoder based models

4) Multiscale and pyramid network based models

5) R-CNN based models (for instance segmentation)

6) Dilated convolutional models and DeepLab family

7) Recurrent neural network based models

8) Attention-based models

9) Generative models and adversarial training

10) Convolutional models with active contour models

11) Other models

## 前置知识（Preliminary）——常用的基础网络模型

（感觉这部分没有必要讲，随便找个硕士论文都有🤣，文档里我尝试用一段话介绍）

### Convolutional Neural Networks (CNNs)

由卷积层、非线性层（激活函数）和池化层搭建而成，卷积层相当于加权求和、激活函数就是非线性化（否则无论多少层神经网络都跟一层的一样，因为都是线性的）、池化层负责降维减少模型参数。每层共享权重，越高层感受野越大，越能进行高级的特征表示。

### Recurrent Neural Networks (RNNs) and the LSTM

每一时刻的输入是外界输入以及上一个隐藏状态，输出为一个外界输出以及下一个隐藏状态。但面对较长的序列信息时不能很好地利用相互的信息，同时还会受到梯度消失或爆炸的问题。LSTM改良了这些问题。

### Encoder-Decoder and Auto-Encoder Models

常说的U-Net就是一种E-D结构，encoder负责将输入压缩成不同层次的特征（就像前文所说的卷积），然后decoder负责通过不同层次的特征得到输出。这种结构非常适合sequence-to-sequence或者image-to-image的问题，例如自然语言处理（Natural Language Processing, NLP）或者各种CV任务（深度估计、图像分割、超像素等）。Auto-Encoder就是特殊情况。

### Generative Adversarial Networks (GANs)

GAN就是博弈的感觉，minmax问题。

## 经典模型

### Fully Convolutional Networks (FCNs)

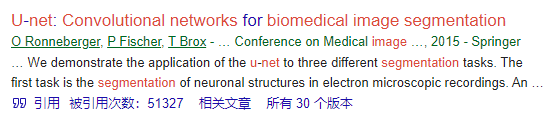


Long et al. [30] proposed Fully Convolutional Networks (FCNs), a milestone in DL-based semantic image segmentation models.

…, by removing all fully-connected layers such that the model outputs a spatial segmentation map instead of classification scores.

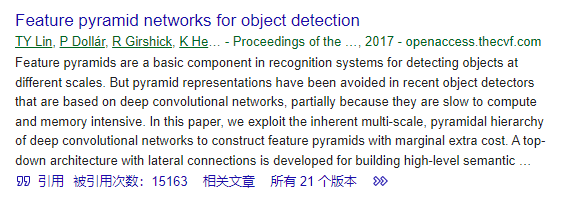
…, it is too computationally expensive for real-time inference, it does not account for global context information in an efficient manner, and it is not easily generalizable to 3D images.

### U-Net



Ronneberger et al. [47] proposed the U-Net (Fig. 14) for efficiently segmenting biological microscopy images. The U-Net architecture comprises two parts, a contracting path to capture context, and a symmetric expanding path that enables precise localization. The U-Net training strategy relies on the use of data augmentation to learn effectively from very few annotated images. It was trained on 30 transmitted light microscopy images, and it won the ISBI cell tracking challenge 2015 by a large margin.

### Feature Pyramid Network (FPN)



One of the most prominent models of this sort is the Feature Pyramid Network (FPN) proposed by Lin et al. [53], which was developed for object detection but was also applied to segmentation.

To merge low and high resolution features, the FPN is composed of a bottom-up pathway, a top-down pathway and lateral connections. The concatenated feature maps are then processed by a 3 × 3 convolution to produce the output of each stage. Finally, each stage of the top-down pathway generates a prediction to detect an object.

### Mask R-CNN (Regional CNN)



In particular, the Faster R-CNN [61] architecture (Fig. 17) uses a region proposal network (RPN) that proposes bounding box candidates. The RPN extracts a Region of Interest (RoI), and an RoIPool layer computes features from these proposals to infer the bounding box coordinates and class of the object. Some extensions of R-CNN have been used to address the instance segmentation problem; i.e., the task of simultaneously performing object detection and semantic segmentation.

He et al. [62] proposed Mask R-CNN (Fig. 18), which outperformed previous benchmarks on many COCO object instance segmentation challenges (Fig. 19), efficiently detecting objects in an image while simultaneously generating a high-quality segmentation mask for each instance. Essentially, it is a Faster R-CNN with 3 output branches—the first computes the bounding box coordinates, the second computes the associated classes, and the third computes the binary mask to segment the object. The Mask R-CNN loss function combines the losses of the bounding box coordinates, the predicted class, and the segmentation mask, and trains all of them jointly.

## 未来展望

1. 更加具有挑战性的数据集；
2. 结合DL与更早期的图像分割模型；
3. 可解释的深度神经网络模型；
4. 弱监督或无监督学习；
5. 实时计算的模型；
6. 节省内存的模型；
7. 更多的应用。

# 参考文献

1. 黄鹏,郑淇,梁超.图像分割方法综述[J].武汉大学学报(理学版),2020,66(06):519-531
2. Minaee S, Boykov Y Y, Porikli F, et al. Image segmentation using deep learning: A survey[J]. IEEE transactions on pattern analysis and machine intelligence, 2021.