激光与光电子学进展 先进成像

道路场景语义分割综述

王龙飞，严春满*

西北师范大学物理与电子工程学院，甘肃 兰州 730030

摘要 图像语义分割是计算机视觉的重要研究领域，是场景理解的关键技术之一。在无人驾驶领域，通过对道路场景进行高质量的语义分割，可为自动驾驶汽车的安全行驶提供保障。首先从道路场景语义分割的定义出发，探讨了目前该领域面临的挑战；其次，将语义分割技术划分为传统的分割技术，传统与深度学习相结合的分割技术和基于深度学习的分割技术，重点介绍了基于深度学习的语义分割技术，并按照强监督、弱监督、无监督三种不同的网络训练方式对其进行了阐述；然后总结与道路场景语义分割相关的数据集以及性能评价指标，并在此基础上进行对比，分析常见的图像语义分割方法的分割结果；最后，对道路场景语义分割技术面临的挑战以及未来的发展方向进行了展望。

关键词 机器视觉；计算机视觉；语义分割；卷积神经网络；自动驾驶；数据集

中图分类号 TP391 **文献标志码** A **doi**：10.3788/LOP202158.1200002

Review on Semantic Segmentation of Road Scenes

Wang Longfei, Yan Chunman*

School of Physics and Electronic Engineering, Northwest Normal University, Lanzhou, Gansu 730030, *China*

Abstract Image semantic segmentation is an important research field of computer vision and also one of the key technologies for scene understanding. In the field of unmanned driving, high-quality semantic segmentation of road scenes provides a guarantee for the safe driving of autonomous vehicles. First, this paper starts with the definition of semantic segmentation of road scenes and discusses the current challenges in this field. Second, this paper divides the semantic segmentation technology into a traditional segmentation technology, a traditional segmentation technology combined with deep learning and a segmentation technology based on deep learning, focuses on the semantic segmentation technology based on deep learning, and elaborates it according to three different network training methods of strong supervision, weak supervision and unsupervison. Then, the datasets and performance evaluation indicators related to the semantic segmentation of road scenes are summarized and compared, and the segmentation results using the common image semantic segmentation methods are analyzed. Finally, the challenges faced by the road scene semantic segmentation technologies and the future development direction are prospected.

Key words machine vision; computer vision; semantic segmentation; convolutional neural network; automatic driving; data set

OCIS codes 150.1135;100.5010;100.4996

1 引 言

图像语义分割技术在实际中的应用广泛，典型应用场景有自动驾驶和医学图像识别等，其中针对道路场景的语义分割[1-2]正是自动驾驶的核心技术之一。针对道路场景的语义分割是将采集到的道路场景图像中的每个像素都划分到对应的类别，以实现道路场景图像在像素级别上的分类。在自动驾驶

收稿日期：2020-07-09；**修回日期**：2020-08-17；**录用日期**：2020-09-27

基金项目：国家自然科学基金(61861041)

E-mail：*yancha02@163.com

的技术组成单元中，环境信息的处理是一个关键部分，这就需要高水平的道路场景语义分割等相关技术为智能车辆提供重要的路况信息，保证自动驾驶汽车能够安全行驶。因此，在自动驾驶领域中，道路场景的语义分割技术发挥着十分重要的作用，是当前研究的热点。

在自动驾驶中，准确性和实时性是十分重要的指标。但是实际语义分割中的精确性会受到不同行驶区域的影响，首先要克服不同目标对象的相异性和相似目标对象的相似性，其次还要注意分割对象所处场景的复杂性，最后一些外界因素如光照、拍摄条件、拍摄设备和拍摄距离也会使得目标物体与图片差异较大，进而影响分割的效果。这些因素都极大提升了图像语义分割的难度，进而影响无人驾驶的实现。综上所述，在无人驾驶领域中，道路场景的语义分割是一项关键并且充满挑战的技术。

语义分割技术是目前计算机视觉研究的热点研究方向，已有一些文献[3-13]对其成果进行了综述。但是，关于无人驾驶领域的道路场景语义分割方法，还没有全面的综述性文献。因此，本文进行了相关工作，第一节对自动驾驶技术及其核心道路场景语义分割技术进行了概述；第二节以深度学习的出现作为划分点，对图像语义分割技术的发展历史进行了归纳与总结，并对不同时期的语义分割方法进行了综述；第三节重点对基于深度学习的语义分割方法进行了分析，通过不同的监督信息，将语义分割方法分为强监督、弱监督与无监督三种类别，并对每一类别进行进一步的分类与分析；第四节针对道路场景，总结了当前适用于道路场景语义分割的数据集以及相应的评估指标，对本文提到的语义分割方法的技术特性与工作性能进行了测评与分类总结，并针对道路场景语义分割的特点与需求，对这些分割方法进行了对比和分析；第五节，对本文的工作进行了总结，对道路场景语义分割技术的发展趋势进行了总结与展望。通过分析可以看出，针对道路场景的语义分割对分割的精准性和实时性有着较高的要求，如何权衡两者是道路场景语义分割的关键，目前基于强监督的语义分割方法依旧是道路场景语义分割的主流方法，基于弱监督和无监督的方法是该领域未来热门的研究方向。

2　图像语义分割的发展历史

从演变过程来看，图像的语义分割技术的发展过程如图 1 所示。

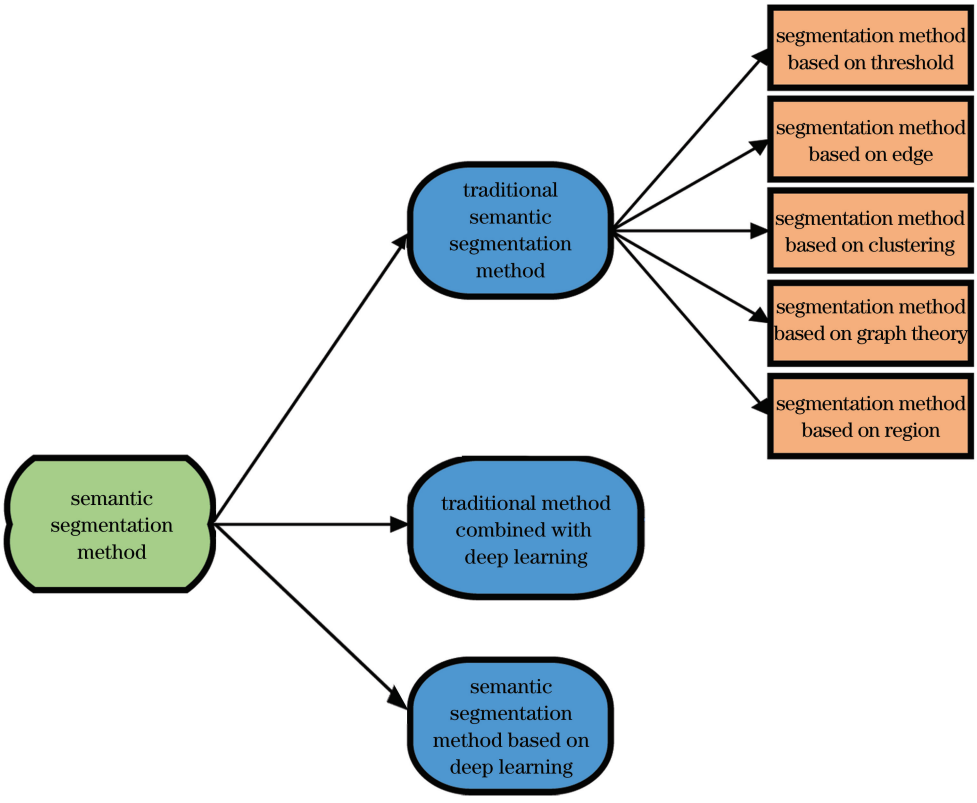

图 1　图像语义分割的发展历史

Fig. 1 Development history of image semantic segmentation

1)传统的语义分割阶段。由于当时的计算处理能力有限,该时期的语义分割算法主要依靠图像纹理、颜色以及其他一些简易的表层特征进行图像分割。分割结果相对粗陋,精度较低,且无相关标注。

2)传统方法与深度学习相结合的语义分割阶段。该方法类似于目标检测法,在分割过程中首先使用传统方法对图像进行初步处理,形成图像级的效果,然后使用卷积神经网络(Convolutional Neural Networks,CNN)中的特征分类器进行语义分割,最终形成图像分割效果。该方法传承了传统的分割方法,存在一定的不足和缺陷,其准确性也相对较低。

3)基于深度学习的语义分割阶段。在鲁棒性特征的自主学习和分类的过程中,深度学习技术表现出了不可比拟的优势和特点,有相对强大的能力。目前,基于深度学习技术的语义分割方法得到了普及和推广,并且取得了较前两类方法更好的效果。该类方法也是本文重点讨论的内容。

2.1　传统的图像语义分割算法

在深度学习广泛应用于计算机视觉之前,为了将目标与背景分离,传统图像语义分割方法利用多种特征,如颜色、灰度、纹理、几何形状等,把图像划分成多个独立的区域。传统图像语义分割方法包括基于阈值的分割、基于边缘的分割、基于聚类的分割、基于图论的分割以及基于区域的分割。其中最常用的是基于图论的分割,而“Normalized cut”和“Grab cut”算法是基于图论分割法的最常用的技术[14],将在下文进行说明。

2.1.1　Normalized cut 图像分割算法

研究者在 21 世纪初提出了一种全新的图像分割方法,该分割方法以图片为单位,并将其作为分割图像的依据,因此这种语义分割方法被定义为 Normalized cut 算法[15]。其实现图像分割的思路是:以图片为单位,计算权重图(weighted graph),然后将其分割成一些具有相同特征的区域。其中最小分割算法(Min-cut algorithm)作为一个重要的方法,可以很完美地把待分割整体分割成两部分,但是最小化分割也存在边缘角元素缺失等缺陷,这使得最终的结果存在偏差。

2.1.2　Grab cut 图像分割算法

Grab cut[16-17]算法同样以图片分割为理论基础,利用混合高斯模型以及吉尔斯能量方程,基于颜色空间实现建模效果,采用迭代方式求得方程的最优解,最终获取高斯模型的最优参数解。该算法的提出显著拓宽了图像分割领域并实现了彩色图像的分割。

Grab cut 虽然在分割性能上有所提升,但是便利性较差,很多系统无法使用该技术,而且还需要考虑操作者的稳定性。刘磊等[18]将高阶势能项引入 Grab cut,使得其可以更好地描述像素的细节和关联信息,从而提高了模型的分割精度。

2.1.3　最新的传统语义分割算法

2011 年,Arbeláez 等[19]综合运用 GPB(Globalized Probability of Boundary)和 UCM(Ultrametric Contour Map)两种方法进行检测,并提出了一种全新的检测算法即轮廓检测法。该算法首先利用 GPB 方法对任一像素边缘的实际概率进行合理测算,然后针对这些测算结果形成的不同闭合区域,利用 UCM 法进行转化,形成层次分明的树状结构。随着研究的不断深入,2016 年,Zhang 等[20]提出了随机决策森林分割法,与轮廓法不同的是,该检测法主要是利用不同的决策树进行组合以形成分类器。2017 年,Pont-Tuest 等[21]综合了以上两种检测方法,提出了新的检测方法即 MCG 算法。该方法首先使用 GPB-UCM 法对图像轮廓进行分割处理,得到不同的块状结构,然后使用随机法形成的分类器进行进一步的分割处理。该方法实现了传统模式和传统方法的优化升级。

2.2　传统方法与深度学习相结合的图像语义分割方法

研究发现,传统方法的突出特点是重点使用表层特征和外部结构特点来完成图像分割,然后进行人工标注[22]。现代科技的进步推动了深度学习技术的持续发展和语义分割技术的变革。研究者将深度学习算法模型引入语义分割研究中,首先采用传统方法进行初步分割,得出目标区域,然后使用卷积神经网络对目标的特征进行深入学习,并形成科学合理的分类器,最终实现目标区域的分割并完成自动标注。

该分割方法基于卷积神经网络原理,对卷积网络进行训练,借助分割树、超像素等技术获取原始的轮廓分割区域,实时监督卷积网络,并进行深度学习,通过超像素分割以及无参数多级解析等多个处理过程获得最终结果。同时期,研究者在分割室内场景过程中,综合运用了图像和深度图技术,基本流程相对简单:首先对图像的滤波特征、卷积特征等进行合理提取,通过融合不同尺度、不同结构、不同层级的特征图,构建科学的分类器。RGB 图像经过超

像素分割[23]后，可以利用该分类器进行进一步的分类。不可忽视的是，超像素分割法存在诸多不稳定因素，容易产生不合理、错误的分类结果。此外，超像素分割法对弱边界图像区域的处理存在一定的难度和局限性。

3　基于深度学习的语义分割方法

近年来，随着深度学习的快速发展，语义分割研究领域也取得了突破性进展。与传统的语义分割方法相比，基于深度学习的语义分割方法能获取更多、更高级的语义信息来表达图像中的信息。自语义分割领域引入深度学习以来，作为衡量语义分割效果的重要指标，分割精度一直是研究的热点。全卷积神经网络(FCN)[24]模型初步实现了像素级的语义分割，使得该领域的分割精度有了跨越式提高。许多基于 FCN 的语义分割方法相继出现，本节将详细介绍基于深度学习的语义分割方法，根据网络训练方式的不同，将其划分为基于强监督的语义分割方法、基于弱监督的语义分割方法以及基于无监督的语义分割方法，它们主要的优缺点如表 1 所示。

表 1　基于强监督、弱监督及无监督的语义分割方法的优缺点对比

Table 1 Advantage and disadvantage comparison among strongly supervised, weakly supervised and unsupervised semantic segmentation methods

Type	Advantage	Disadvantage
Strongly supervised	High segmentation accuracy based on densely annotated datasets	Being excessively dependent on dataset marked by dense set, inability to migrate, and poor segmentation accuracy for unknown scenes
Weakly supervised	Only image-level annotated dataset required to complete training	Large number of datasets needed, long time, and lower accuracy than that of strong supervision
Unsupervised	Being independent on manual intensive annotation dataset and strongly adaptable to unknown environment	Being difficult to adapt and no high segmentation accuracy at present

3.1　基于强监督的语义分割方法

样本的人工标注可以体现大量有用的局部数据和细节特征，能在一定程度上大幅提升训练效果，提高分割精度。强监督学习模型是当前应用最广的分割模型，也是效果最佳、影响范围最大的算法模型。在语义分割的理论研究历程中，FCN 模型的提出无疑是具有里程碑意义的，其为后期的模型算法研究指出了全新的方向。FCN 网络的结构示意图如图 2 所示。

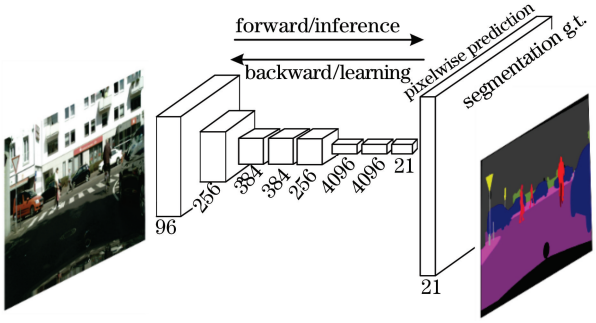

图 2　全卷积网络的结构示意图[24]

Fig. 2 Structural diagram of fully convolutional network[24]

FCN 中的卷积层与采样层分别涉及到上下、正反等多种类型，且这些结构类型在空间中任意平移时保持不变。恢复图片的原始分辨率大小是全卷积网络的常见应用场景，处理过程中常利用反卷积形式。在 FCN 结构中，一定数量的尺寸固定的卷积层起到常规卷积网络中全连接层的作用，这种结构可以让神经网络在图片中自由密切滑动以提升卷积神经网络的滑动灵活性，最终生成的预测图中包含稠密的输出图像。然而，FCN 仍然保留了卷积神经网络(CNN)中的池化层，池化层增加了卷积神经网络的感受野，但是连续的下采样会导致细节丢失，极大地影响分割的结果。同时，较高的采样率会导致特征图大小和空间信息的损失。针对上述问题，在 FCN 基础上，研究者又提出了一系列新方法，我们将其划分为 6 类，即基于扩大感受野的分割方法、基于概率图模型的分割方法、基于特征融合的分割方法、基于编码器-解码器的分割方法、基于循环神经网络(RNN)的分割方法和基于生成对抗网络(GAN)的分割方法，如图 3 所示，其中 ASPP 为空洞空间卷积池化金字塔。

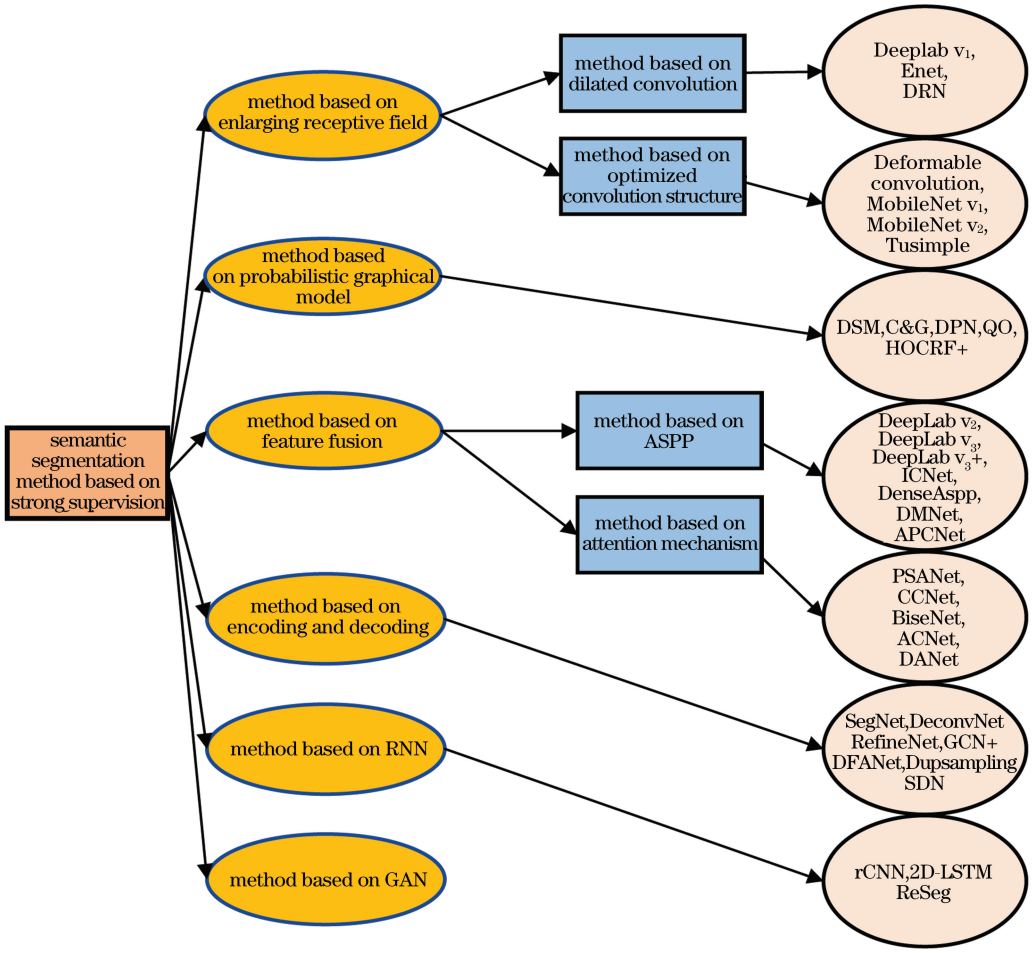

图 3　基于强监督的语义分割方法

Fig. 3 Semantic segmentation method based on strong supervision

3.1.1　基于扩大感受野的方法

空洞卷积[25]可用于密集预测的卷积层，又名扩张卷积。空洞卷积以保证图像分辨率属性为基础，在不减小覆盖范围的同时提升感受野。该方法对卷积神经网络分辨率的影响着重体现在特征响应计算领域。扩展卷积模式与空洞卷积相呼应。如图 4 所示，选择 3×3 卷积，对比扩张系数为 1、2、4 时的感受野，不难看出，感受野与扩张系数呈正相关，扩张系数为 4 时感受野大小为扩张系数为 1 时的 5 倍，因此扩展卷积的作用是十分显著的。扩展卷积能够扩充卷积的堆叠效应，提升感受野的大小。空洞卷积则专注于分辨率和计算响应能力的提升，降低计算过程对参数的依赖度，仅需要输入较少的参数或因子，就能实现卷积核感受野的扩大，同时有助于前后内容的获取。

DeepLab v1 网络模型由 Chen 等[26]提出，DeepLab v1 创新性地将空洞卷积应用到 VGG16 网络中，通过将 VGG16 的全连接层转换为卷积层，并将 VGG 模型第四个和第五个池化层之后的所有卷积层分别调整为不同扩张率的空洞卷积，感受野被恢复至原图像大小，提升了模型分割的准确率。2016 年，Paszke 等[27]提出了一种实时分割的模型即 ENet。该分割模型主要运用了 bottleneck 模块思维方式，对多个空洞卷积进行串行操作，以调整感受野的实际区域大小，有效解决了特征分辨率持续下降等问题。该算法模型运用的参数较少，运行速度较快，在一定程度上推动了实时分割技术的发展。2017 年，Yu 等[28]提出了 DRN 网络模型。研究表明，该模型以 ResNet 网络为基础，运用空洞卷积对普通卷积进行替换操作，以维持原图像的实际分辨率和原网络的有效感受野区域。该模型利用两个不同扩张率的空洞卷积，对 ResNet 的末尾卷积层进行替换操作，以不断增强空间有效信息。为了避免空洞卷积的循环利用引发的棋盘效应，需要移除残

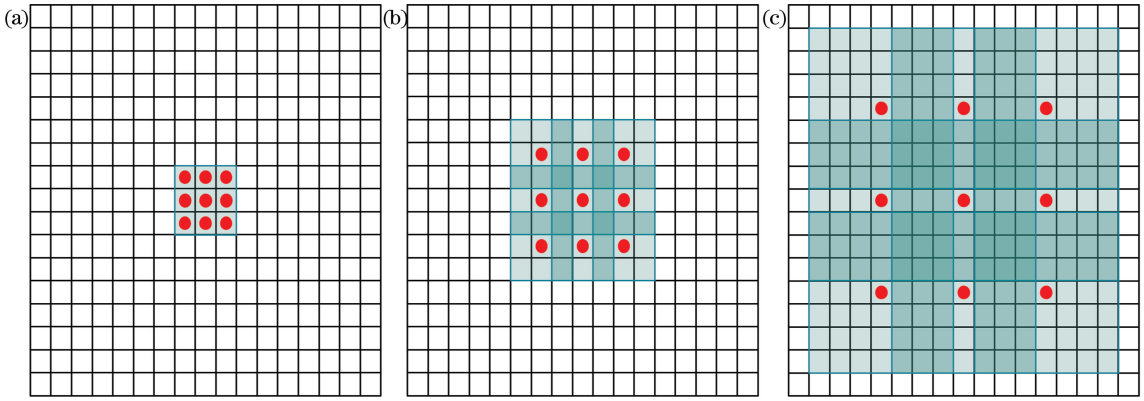

图 4　扩张卷积示意图[25]。(a)普通卷积；(b)扩张率为 2 的扩张卷积；(c)扩张率为 4 的扩张卷积

Fig. 4 Schematic of expansion convolution[25]. (a) Ordinary convolution; (b) expansion convolution with expansion rate of 2; (c) expansion convolution with expansion rate of 4

差和最大池化层等，最后通过全卷积等方法实现像素的输出操作。

使用 CNN 方法对图像进行语义分割时，其中的池化操作过程将会不断增大感受野的有效范围，并融合背景信息。但不可忽视的是，该过程同样会使图像分辨率持续下降，造成部分空间信息的遗失。针对该类问题，比较合理的解决思路是优化卷积结构，并使用优化后的卷积结构进行卷积和池化等操作。

运用扩张卷积的方式可以快速有效地获取图像的深度特征，扩大感受野范围，并能保留特定像素的位置信息。然而，在进行卷积操作处理时，容易形成一定的空间漏洞，出现数据遗失、消息丢失等问题。在文献[29]中，研究者运用混合扩张卷积(hybrid dilated convolution，HDC)对扩张卷积进行了替代操作，同时运用稠密上采样的算法取代了 BI 算法。HDC 方式包含了一系列的扩张卷积模块，能够进一步扩大感受野，同时维持局部信息有关特征。以上方法可以有效扩大感受野，但是由于卷积核的形状相对固定，这些方法模拟几何变换的能力相对较弱，适应图形变化的能力较差，提取不规则形状物体特征的能力也较差。在文献[30]中，研究者在进行卷积处理时，通过运用有一定偏移量的采样操作，引入了一个可学习的偏移量，最终调整了卷积核的形状，使其具有可变性，并提出了可变形卷积(deformable convolution)的基本概念。该种卷积模式能有效扩大感受野，增大图像区域，提高语义分割对图形变换的自适应能力，并提高了分割的精度和准确度。在进行深度可分离卷积(depthwise separable convolution)的过程中，较少的计算量降低了性能消耗。应用于移动设备的分割模型通常有逐点卷积和深度卷积两种模式。其中，逐点卷积主要运用 1×1 卷积，深度卷积则在各个通道运用不同的卷积核。深度卷积的实际分割效果并不好，往往只能对低维度空间的基本特征进行提取。为了解决该问题，文献[31]在深度卷积开始之前，不断提高卷积维度，使深度卷积能够在高维度空间运行。

3.1.2　基于概率图模型的分割方法

概率图模型(Probabilistic Graphical Model，PGM)用于 CNN 的后期处理，以结构化预测的方式有效地优化物体边界，捕获图像上下文信息，使得局部特征与全局特征的利用率得以平衡。

综合运用条件随机场(Conditional Random Fields，CRFs)、CNN 两种模型，对信息传递过程中的相关信息进行合理预测，能够有效降低冗余计算量，进而实现运算效率的提升。该方法能获取相对丰富的数据信息，提高运行效率。然而，在结构预测过程中，该方法仅能将图像输入到一元项或成对项中，在中高项中难以实现结构预测，分割精度相对不高。因此，Arnab 等[32]将两种不同形式的高阶势能项(higher order potential，HOP)内嵌到 CNN 中，以训练深度，分割质量得到了提高。此外，为了优化分割模型并提高分割质效，Vemulapalli 等[33]使用高斯条件随机场(Gaussian conditional random field，GCRF)优化分割结果。部分学者对 FCN 和 CRF 两种模型进行了融合操作，提出了两种不同的分割模型，即 SegModel[34] 网络模型和 DFCN-DCRF[35] 网络模型。

3.1.3　基于特征融合的分割方法

3.1.1 节中的方法均利用扩张率不同的空洞卷积进行串行操作，以此不断增大感受野，并对语义特征进行深入提取。然而，循环反复利用空洞卷积势必会产生棋盘效应，也会使部分特征遗失，并占用大量的运行空间，消耗大量的内存。3.1.2 节中的基于概率模型图的方法也存在计算量过大、训练时间长、消耗大量内存等方面的问题。特征融合是指将提取出的特征图进行相加或拼接融合。在特征提取阶段，通过融合多尺度的特征信息，丰富特征图的语义信息。在特征的利用阶段，通过融合不同层级的特征，利用全局有效信息，提高分割精度。基于特征融合的方法通过融合不同层次、不同区域的特征，捕获图像中隐含的上下文信息，能有效提高分割速率和分割效能，也能大幅度降低运行消耗。

Lin 等[36] 提出了特征金字塔网络（Feature Pyramid Networks）。在结构设置过程中，该网络通过调整高层特征、低层特征的连接形式，丰富各尺度下特征的语义信息。DeepLab v2[37] 在 DeepLab v1 的基础上引入了带孔卷积和金字塔池化（ASPP），并将 VGG-16 网络换成 ResNet 网络。通过采样，利用不同比例实现了上下文的捕捉，该过程以输入多种采样率的空洞卷积为基础。分类效果的提升则以卷积图像特征的挖掘以及内容图像特征的提取为基础，且上述处理以不影响特征图的分辨率为前提。DeepLab v3[38-39] 改进了 ASPP 结构，通过引入 Resnet block 模块并以提取显著性特征为目标，提出了空洞卷积模式，实现了模块与空间要素的池化效应。结合上述两种方法，Yang 等[40] 进行了深入研究，提出了 DenseASPP 网络。在街景分类过程中，该网络运用全新的方式对扩张卷积进行了连接操作，以此获取密集程度更高的采样点和有效范围更广的接收野。He 等[41] 研究发现，尽管 ASPP 能够对图形尺度变化进行一定的处理，但是难以在尺度、扩张率的变化中实现新的平衡。因此，该团队提出了动态多尺度网络（Dynamic Multi-scale Network，DMNet），并借助该网络实现了动态卷积语义的感知和估计。为了提高网络聚合上下文信息的能力，Zhao 等[42] 提出了金字塔场景解析网络（pyramid scene parsing network，PSPNet），随后 Zhao 等[43] 又从压缩 PSPNet 的角度出发，提出了具有实时分割特点的图像级联网络（image cascade network，ICNet）。He 等[44] 提出了自适应金字塔上下文网络（APCNet），综合运用多个自适应模块，对多层级的上下文表示进行了合理构建。Wu 等[45] 针对扩张卷积的替代网络，于 2019 年提出了全新的联合金字塔取样模型。该方法能有效获取具有高分辨率映射特征的样本，可以大幅度降低精度损失和内存消耗。

传统的固定卷积结构主要借助 FCN 的分割框架获取信息，但只能获取短距离信息。为了获取长距离上下文信息，研究者提出了扩张卷积等方法。然而，该类方法在获取信息的过程中，并不能形成密集的信息。因此，在进行语义分割的过程中，Zhao 等[46] 合理引入了注意力机制，提出了 PSANet 网络模型，通过预先绘制注意力图[47] 来聚合不同位置的信息。该方法借助巨大的注意力图，实现各像素之间关系的计算，在运行过程中计算相对复杂，内存使用率相对较高。为了切实提高分割质效，研究人员提出了一系列创新网络模块。下面对其中比较流行的三种模块进行详细介绍：CCNet[48] 算法模块通过插入完全卷积的任意神经网络，能够进行高端分割；BiSeNet[49] 模块不需要进行任何采样处理，就能实现全局信息的整合操作，有效降低了运行成本，提高了计算速度；ACNet[50] 模块综合运用自注意力辅助算法模式和平行分支架构，对深度图像特征进行了平衡操作。近年来，自注意力机制在语义分割实践中取得了越来越显著的成效。研究者将该机制纳入语义分割的基本过程中。为了有效降低时空复杂度，双重注意网络（Dual Attention Networks，DANet）[51] 等创新方法被提出。在进行语义分割的过程中，合理引入注意力机制，对有关信息进行学习，通过调整和优化注意力机制，形成全新的十字交叉模块和自注意力模块，以此获取全局信息，对各层级信息进行感受，使得信息和内部特征的捕获变得更为容易。

3.1.4　基于编码-解码器的方法

该方法的基本思路是编码器通过一系列卷积-池化操作，提取图像的主要特征信息，再通过解码器的上采样-转置卷积结构，逐步恢复图像的空间维度。依托编码器-解码器的基本方法，可以对低分辨率的图形进行特征处理和上采样操作，可有效解决分辨率下降的问题，并高度还原像素的时空信息和图形的维度数据。

SegNet[52] 和 U-net[53] 是两个典型的用于图像语义分割的编码-解码器结构，SegNet 网络结构图

如图 5 所示。SegNet 采用 VGG-16 网络，利用该网络输出稠密的特征图，通过对稀疏图像的卷积计算实现对稠密图的恢复[54]。随后，研究者又在 SegNet 网络的基础上提出了 Bayesian SegNet 网络，通过引入贝叶斯网络和高斯过程，解决了先验概率无法给出分类结果置信度[55]的问题，提升了网络的学习能力。Noh 等[56]基于 FCN 提出了一个完全对称的 DeconvNet 网络，该网络基于 FCN 与反卷积网络的互补，使用 FCN 提取总体形状，利用反卷积网络提取精细边界，既能应对不同尺度大小的物体，又能更好地识别物体的细节，提高了分割的效率。

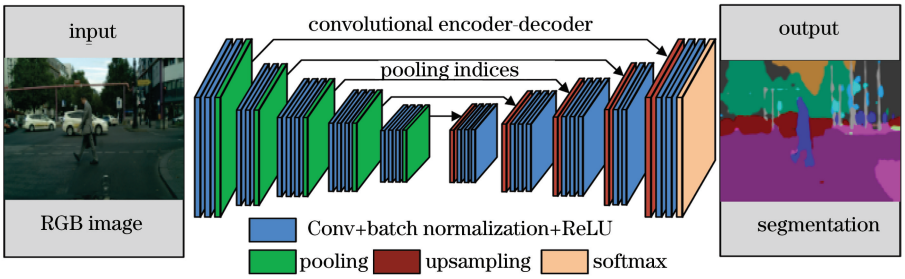

图 5　SegNet 网络的结构示意图[52]

Fig. 5 Structural diagram of SegNet network[52]

研究者提出了架构相对匀称的 U-Net 网络，该网络的编码和解码结构作用不同且相互配合，起到完善细节、恢复效果的作用。尽管 U-Net 语义分割模型实现了很好的分割效果，但只能处理 2D 图像。研究者提出了 V-Net 网络[57]，该模型是一种将 3D 体积、全卷积与神经网络结合的三维对称语义分割模型，解决了训练标注数据集不足的问题，与其他分割模型相比，具有计算优势。为了解决语义分割过程中数据信息的丢失问题，Lin 等[58]于 2017 年提出了 RefineNet 网络。该网络采用的是链式残差连接模式，能够追回并有效融合分割过程中缺失的信息，进而形成相对清晰的预测图像。综合来看，该方法能有效融合高层特征和低层特征，并运用了恒等映射、残差连接等先进的思维方式，能得到良好的训练效果。在不同的场景环境中，该模型都能获得相对优越的分割效果。研究者们基于编码和解码器结构，深入研究了分割方法，提出了一系列卓有成效的研究成果。如运用提前采样的方法，减少解码器的应用，以此形成简化版的 ENet 结构，实现降低网络冗余、减少参数数量等基本目标。

为了合理改良编码-解码模型结构，研究者进行了多项改进：1）不断提高语义分割的实际速度，例如运用 ENet、LEDNet 等模型，有力推动实时分割目标的实现；2）对多个分辨率特征进行有效融合，例如 DUpsamling[59]模块，可用于学习采样；3）不断扩展感受野的有效范围，提高分割精度，例如 GCN 模块；4）对多尺度、多层级的信息进行捕获，确保有效恢复目标信息，例如 SDN 模块。

3.1.5　基于循环神经网络的方法

深度学习中另一种应用较广、效果良好的运算模型是循环神经网络（Recurrent Neural Network，RNN）[60]模型。该模型的主要优势是除了学习当前信息之外，还可综合运用序列信息，构筑全局建模算法，提高图形信息的综合利用率。基于此思想，Visin 等[61]综合运用 CNN 方式获取的局部信息、RNN 方式获取的全局特征，同时借鉴图像分割模型，提出了 ReSeg 网络。受到图像分割网络 ReNet 的影响，Li 等[62]提出了 LSTM-CF（Long Short-Term Memorized Context Fusion）网络模型，该模型能有效利用深度图像和光度两个基本特征。然而，由于该模型仅使用了 LSTM 方式，图像处理的灵活性和多样性相对较差。因此，Liang 等[63]提出了 Graph-LSTM 网络，该网络将任意超像素设置为参考的有效节点，并为图像构建了自适应图像。此后，该团队调整了该网络的运行模式，从编码分层的角度进行了结构优化。考虑到 FCN 与全连接的 CRF 模型（FC-CRF）之间缺乏有效的交互联系，Zheng 等[64]提出了 CRFasRNN 网络模型，将 CRF 的有关学习、推理过程融入到 RNN 的运算中。

RNN 能够保留有关信息，实现历史数据和历史记忆的递归处理，能够对图像内的序列信息进行提取操作，同时也能对图像语义关系进行合理建模，获取有关数据信息。与此同时，该网络模式能与卷积层深入结合，通过融入到神经网络结构中，能对卷积层空间特征进行有效提取，也能实现

像素特征的深度提取。

3.1.6　基于生成对抗网络的方法

与金字塔网络结构类似，生成对抗网络[65]在一定程度上也能替代 CRF，完成图像信息特征的获取，可以在不额外增加训练时间和训练难度的情况下，实现空间的连续性扩展，确保空间特征的一致性。

为了持续减小标签和图像之间的不一致性，Luc 等[66]于 2016 年首次引入了 GAN 技术，在进行语义分割的过程中，运用判别器对标签和分割领域进行识别操作。在医学领域，由于 U-Net 网络不能很好地解决像素类别不一致、不均衡的现实问题，Xue 等[67]提出了一种基于多尺度、多层次函数的对抗网络模型，在图形分割过程中，运用判别器对分割对象的局部属性、全局结构特点进行深入学习，以此获取不同像素间的有效空间关系。除此之外，GAN 模型还具有识别数据真假并持续产生新数据的能力。由于特征学习具备一定的关联性，因此，为了实现对小样本特征的持续有效学习，要将对抗学习合理运用到弱监督学习或半监督学习中。GAN 模型在运用过程中存在一定的不稳定性，尤其是针对大数据图像，该方法的解释性和可延伸性存在不足。

3.2　基于弱监督的语义分割方法

基于强监督的语义分割方法需要大量像素级标注训练样本，由于获取像素级的语义标注样本需要消耗大量的时间和精力，并且通过像素级标注样本进行训练有一定的局限性，因此基于弱监督的语义分割开始涌现。本文根据不同类型的监督信息，将基于弱监督的图像语义分割方法分为 6 类：基于边界框级标注的方法、基于涂鸦级标注的方法、基于点级标注的方法、基于图像级标注的方法、基于混合标注的方法以及基于附加数据源的方法。

3.2.1　基于边界框级标注的方法

基于边界框标注的方法是将包括整个物体的矩形区域作为训练样本，并提供标注信息。虽然该标注方法是众多标注方法中较为复杂的一种，但是其包含了更多的语义信息，成本较低，分割性能较好。

Dai 等[68]借助 FCN 网络，通过合理运用候选区域，提了出 BoxSup 网络模型。该模型以边界框标注的图像作为训练样本，选用 MCG 算法进行计算，形成原始候选区域，随后将该内容以“监督信息”的形式录入 FCN 网络，实现进一步的优化升级；然后，进一步预测候选区域的有效范围，并对该区域进行反复优化升级，直至结果收敛到合理范围内。面对分类问题，DeepCut[69]主要通过反复更迭操作来进行图像分割，以此不断提高分割准确率和图像精度。

在传统的弱监督学习过程中，普遍使用简单迭代方式进行模型训练，最终结果往往与实际标签有较大的差异。Song 等[70]在进行图像分割方法研究时，合理运用边界框驱动分类区域掩蔽（box-driven class-wise masking，BCM）模型，对不相干区域进行删除操作，由此获得像素级的分割区域和填充率，随后运用填充率引导的自适应损失（filling rate guided adaptive loss，FR-loss）算法模型，对提案中已完成标注的错误像素进行纠正和删除。该模型算法主要依托边界框监督算法对图像数据进行标注和分割，可以最大限度地降低错误标注形成的不良影响。

3.2.2　基于涂鸦级标注的方法

该方法在标注过程中合理设定训练样本，采取的分割方法也相对简单。样本以涂鸦级图像为主，获取难度相对较低，有效减少了人工标注的任务量。

文献[71]提出了利用随机涂鸦的点作为监督信息的标注方法。该方法在实际操作过程中，使用像素点对图像进行标注，设定涂鸦点为监督信息，有效结合了监督信息、CNN 网络模型的函数优势，得到了良好的分割结果。文献[72]提出了 ScribbleSup 模型算法。该模型以一些包含涂鸦线条或涂鸦点的图像为样本，并以涂鸦方式展开标注。该算法大致可以分为两个阶段：第一阶段为自动标记阶段，主要依据涂鸦线条形成不同形态的像素块，然后以该像素块为基本节点进行自动建模，最终对所有图像进行标注处理；第二阶段为图像训练阶段，主要是针对第一阶段已形成的图像进行模型训练，最终获得合理的分割结果。

3.2.3　基于点级标注的方法

从本质上讲，实例点的标注方法是一种弱标注的方法，主要通过提供位置信息、标识中心位置等方式来实现。与其他算法相比，在同样预算前提下，点级监督的监督效果更佳，最终效果更为优越。

为了获取良好的分割效果，研究者有效融合了点级监督、损失函数的优势特征，进一步强化了语义分割的效果监督。考虑到分割对象包含四个极端点，Maninis 等[73]以此为基础提出了可以实现半自动分割的 CNN 架构即 DeepExtreme Cut（DEXTR）。

3.2.4　基于图像级标注的方法

对比来看，图像级标注有着多重优势和特点，标注过程相对简单，不需要使用像素标注，样本获取相对容易，整体工作量相对较小。因此，该方法也逐渐在弱监督学习过程中成为主流方法。研究者在研究过程中，通过合理引入多实例的学习模型，对图形标签、像素间的关联结构进行科学搭建，并运用超像素等算法对各类标签进行平滑操作。研究者使用期望值最大化的方式，对像素级标签进行合理预测和评估，并以此标签作为训练样本，对数据模型进行更新，最大限度完成期望值的最大化处理。研究者对相关特征图进行了分解操作，形成了初步的多通道特征。不同通道有不同的局部特征，经过池化操作形成多通道的基本特征图，随后对该图进行特征标签信息学习。

与像素级标注相比，图像级标注的方法显得有些简单粗陋，很难取得良好的、符合预期的分割效果。在实际操作过程中，可以借助目标区域的扩展、监督信息的挖掘等多种方式，实现图像级标注质量的有效提升。Kolesnikov 等[74]提了 SEC（Seed，Expand，and Constrain）算法。受到该算法的影响和启示，Huang 等[75]使用 SRG 区域增长方法，对种子区域进行监督，获取相关信息，最终形成科学合理的像素标签。受到空洞卷积的影响和启发，Wang 等[76]将 MDC（multi-dilated convolutional）算法模型运用到图像去噪领域。在外部数据缺失或监督信息遗失不全的情况下，Ahn 等[77]通过有效运用 AffinityNet 网络，获得了准确的分割标签，以此来弥补相关信息的缺失。在研究过程中，Zhou 等[78]对监督信息进行了图像级的标注，并利用响应峰值持续提高实例分割效率。对比来看，该分割方法的运算过程简单，样本获取成本较低，通过分类标注即可达到分割的目的，逐步提高了语义分割质量和逐点定位实际效果。

此外，Wei 等[79]以显著性为基本特征，对额外知识信息进行了有效的提取，并提出了一种 SCT 模型算法。该方法对具有显著性特征的区域进行了由下向上的检测，得到了区域图与标签信息之间的关系，随后逐步推断出图像的分割掩码，并以此为监督信息展开学习训练。

3.2.5　基于混合标注的方法

综上所述，以上方法在降低成本、减少时间方面都有显著优势，能大幅度减少数据训练的实际需求。然而，不可忽视的是，弱标注方法存在一定的局限性，单独一种标注数据并不能取得良好的分割效果。在进行标注的过程中，如果能融合其他类型数据，实现优势互补，就能提高分割效果。

基于半监督学习的分割方法在通常情况下使用两种标注图像，其中像素级的相对较少，弱标记的相对较多。研究者提出了随机梯度下降（stochastic gradient descent，SGD）[80]算法模型，通过对两种图像的组合操作，取得了单一类型图像不可比拟的优越性能。Hong 等[81]提出了 DecoupledNet 半监督的分割框架模型。该模型对分割、分类项目进行区别操作，其中分类网络在模型学习过程中主要运用了图像级数据，随后使用训练实例对分割网络进行优化和升级。由于不存在重复循环的操作，因此该方法有相对良好的扩展性。

3.2.6　基于附加数据源的方法

上文中所阐述的涂鸦标注、点级标注内容，一般情况下直接获取的难度较大，需要借助人工交互方式才能取得。与像素标注内容相比，该类标注信息的获取难度相对较小，然而实施弱监督学习的主要目的就是为了尽可能地减少人工交互。所以，研究人员通常会引入部分附加数据，并使用强度较大的监督信息，以避免使用人工标注。

与单张图像相比，视频信息的获取难度相对较小，而且目前视频的传播更为普遍。研究者在进行搜索时，将类标签作为关键词，以 web 库作为搜索源，运用全自动的检索方式获取有关视频资料。同时，合理运用分类器对相关视频区间进行优化处理，可获取更优的检索结果。此外，研究者在新型编解码结构中引入了注意力机制，能对无相关性的知识进行迁移操作，使其运转到弱监督的分割操作过程中。

3.3　基于无监督的语义分割方法

大量研究数据表明：如果某个神经网络有大量的训练数据，则该网络往往具有相对良好的运行属性。在实践过程中，如果设定一定规模的数据集，经过良好训练的网络通常不会有良好的表现。有效的解决方式是采取相对密集的手动标注方法，反复进行网络训练。另一解决办法是综合运用自动语义标注的电脑进行数据合成，进而反复开展数据训练。在此过程中，循环往复的数据合成训练会在一定程度上降低数据的使用性能，减小运行效果。综合来说，最佳的处理办法是引入无监督适

用方法，构建合理的标记区域，持续降低标注数据的误差。

在通常情况下，无监督方式自适应训练的基本过程是借助 Domain Shift 最小化来构建合理的跨域。研究者提出了 DC[82] 方法，有效运用二元域分类器实现标签的均匀布局。此后，经过深入研究，Tzeng 等[83] 进一步提出了新的对抗判别域适用（Adversarial Discriminative Domain Adaptation，ADDA）法，借助对抗训练模式，不断优化相关模型。为了解决跨域分割问题，Hoffman 等[84] 提出 FCNWild 方法。Zhang 等[85] 提出了 FCAN 自适应网络模型，该模型有效结合了图像域和特征域双重自适应网络，利用用合成图像，提升了语义分割质量。

图 6 对基于弱监督和无监督的语义分割方法的分类进行了汇总。

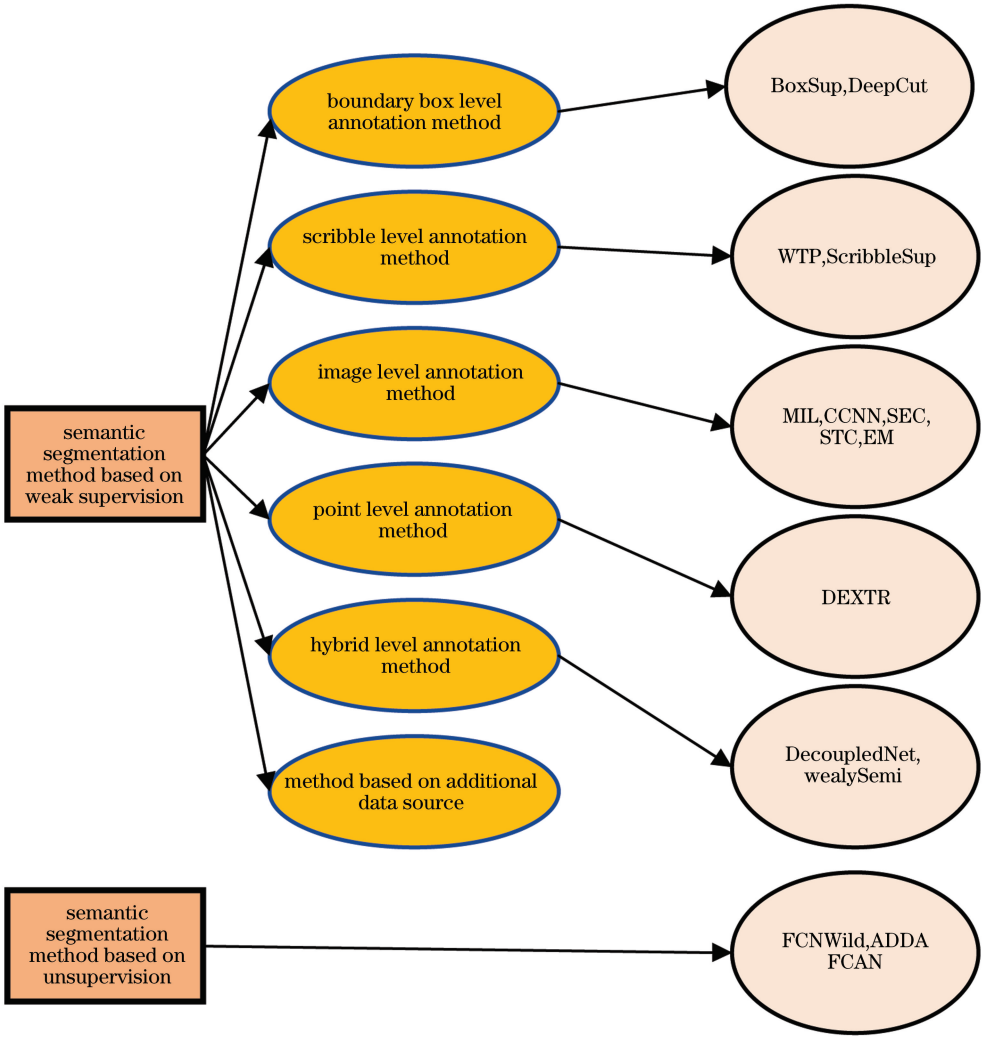

图 6　基于弱监督和无监督的语义分割方法

Fig. 6 Semantic segmentation method based on weak supervision and unsupervision

4　城市道路场景数据集以及性能评价指标

本节以道路场景为核心背景，阐述道路场景语义分割中常用的数据集和评价道路场景语义分割效果常用的性能指标，然后将不同的语义分割方法在不同数据集下进行性能对比，分析和总结适用于道路场景语义分割的方法。

4.1　城市道路场景数据集

大量研究人员专注于研究城市道路场景数据集，在实际场景下使用多个传感器捕获多维度信息，通过大量精细标注，构建了大型城市道路数据集，极大地促进了复杂城市街景下视觉理解的发展。常用的自动驾驶数据集如表 2 所示，常见的交通标志数据集如表 3 所示。

表 2　常用的自动驾驶数据集

Table 2 Common automatic driving datasets

Dataset	Year	Number of categories	Total amount of data	Area	Environment
CamVid[86]	2009	32	700	Europe	Day
KITTI[87]	2013	10		Germany and America	Day
Oxford Robotcar[88]	2014		2×10^7	Oxford	All weather conditions
Cityscapes[89]	2016	34	20000	Germany , Switzerland and France	Spring, summer , and autumn
SYNTHIA[90]	2016	11	13407		Various scenes
Comma. ai	2016			America	
Mapillary Vistas[91]	2017	66	25000	America ,Europe ,Africa, Asia ,and Oceania	Complex weather
Apollo Scape [92]	2018	28	143906	China	Complex weather
BDD100K[93]	2018	10	10000	Multiple cities around the world	Various scenes
Udacity's Driving[94]	2018	3, 8	9420, 15000		
NuScenes	2019	23	14×10^5	Boston and Singapore	Day
D^2-City	2019	12		China	Complex weather
Waymo[95]	2019		3000	America	Complex weather

表 3　常见的交通标志数据集

Table 3 Common traffic sign datasets

Dataset	Summary
KUL Belgium Traffic Sign[96]	Dataset of traffic signs in Belgium
German Traffic Sign[97]	German traffic annotated dataset
STSD[98]	More than 20,000 images containing 3488 traffic signs
LISA[99]	7855 annotations with more than 6610 frames
Tsinghua-Tencent 100K[100]	Dataset with 100000 pictures, including 30000 traffic sign examples

4.2　性能评价指标

在分割实践中，为了取得良好的效果，需要对分割框架的结构属性和实际效能进行合理的测评。下面从执行时间和准确度两个方面对分割网络的性能指标进行阐述。

4.2.1　运行时间

运行时间或处理速度是一个非常有价值的度量标准，在许多应用领域，实时性是一个十分重要的性能，因此需要用运行时间衡量分割方法的实时性。但是，由于硬件后端实现水平的不同，运行时间很难进行比较。因此应在相同的条件下，通过运行时间的比较来评判分割方法的分割效率。对于无人驾驶等对实时性要求较高的领域，运行时间是非常重要的评价标准。

4.2.2　准确度

当前已有多种指标可评价像素语义分割性能的优劣，像素准确率(PA)能体现像素与类对应关系的准确度，对其求平均值则可获取平均准确率(mPA)，还有交并比类指标，例如平均交并比(mIoU)、频率加权交并比(FWIoU)等。常使用mIoU来衡量语义分割模型的性能。

像素准确率的计算是以预测类别准确的像

素数为对象，通过比值计算获取准确率，计算公式为

$$P=\frac{\sum_{i=0}^{n}p_{ii}}{\sum_{i=0}^{n}\sum_{j=0}^{n}p_{ij}},\tag{1}$$

式中：P 为像素准确率；p_{ii} 为准确分类的像素数量；p_{ij} 为分类错误的像素数量；i 和 j 为归属类编号；n 为类别总数。

像素与划分类的对应关系不一定是准确的，因此通过平均准确率(mPA)指标表征准确性，计算公式为

$$M=\frac{1}{n+1}\sum_{i=0}^{n}\frac{p_{ii}}{\sum_{j=0}^{n}p_{ij}},\tag{2}$$

式中：M 为平均准确率。

平均交并比的计算公式为

$$m=\frac{1}{n+1}\sum_{i=0}^{n}\frac{p_{ii}}{\sum_{j=0}^{n}p_{ij}+\sum_{j=0}^{n}p_{ji}-p_{ii}},\tag{3}$$

式中：m 为平均交并比。

频率加权交并比是 mloU 改进后的新评价指标，计算公式为

$$F=\frac{1}{\sum_{i=0}^{n}\sum_{j=0}^{n}p_{ij}}\sum_{i=0}^{n}\frac{\sum_{j=0}^{n}p_{ij}p_{ii}}{\sum_{j=0}^{n}p_{ij}+\sum_{j=0}^{n}p_{ji}-p_{ii}},\tag{4}$$

式中：F 为频率加权交并比。mIoU 指标的表征性和简单性非常突出，是目前图像语义分割领域使用频率最高和最常见的准确度评价指标。

4.3　算法性能对比

本文研究的自动驾驶领域需要实时高效的分割网络，除了对比不同分割网络的分割准确率外，还针对适用于道路场景语义分割的网络，从参数数量和运行速率两方面研究了它们的实时性。

4.3.1　传统语义分割方法的实验对比

在传统图像语义分割方法中，N-Cut 和 Grab cut 等经典算法得到了广泛的应用，但这些算法的效率较低。在此基础上，GPB-UCM，Random Decision Forest 和 MCG 等改进算法吸取了经典算法的优点，在生成的图像分割块质量以及算法时间复杂度上都有更好的表现，但是，传统分割方法在分类数量、分割精度等方面无法满足道路场景语义分割的要求。传统图像语义分割方法的分析归纳如表 4[9]所示

表 4　传统图像语义分割方法的分析归纳[9]

Table 4 Analysis and summary of traditional image semantic segmentation methods[9]

Method	Year	Contribution
Normalized cut	2000	Dividing graph into k subgraphs and then minimizing them
Grab cut	2004	Using image texture and boundary information dependent on small amount of manual intervention to obtain better foreground and background segmentation
GPB-UCM	2011	Using probability of each pixel as an edge, detecting target contour, generating contour map, and completing segmentation with complex steps and high complexity
Random Decision Forest	2016	Combining multiple decision trees into classifier
MCG	2017	On basis of GPS-UCM, using generated multiple contour segmentation blocks when combined with random forest classifier to get prediction object

4.3.2　基于强监督的语义分割方法的实验对比

基于强监督的图像语义分割方法的分析归纳如表 5 所示。

表 5　基于强监督的图像语义分割方法的分析归纳

Table 5 Analysis and summary of image semantic segmentation method based on strong supervision

Method		Model	Year	Key technology	PGM	Dataset	mIoU / %
Method based on enlarging receptive field	Method based on dilated convolution	DeepLab v1	2014	Upsampling and structure prediction	CRF	PASCAL VOC 2012, Cityscapes	71.6, 63.1
		ENet	2016	Decomposition filter and dilated convolution		Cityscapes, CamVid	58.3, 51.3
		DRN	2017	Dilated convolution			
	Method based on optimizing convolution structure	Deformable	2017	Deformable convolution		PASCAL VOC 2012	75.3
		MobileNet V1	2017	Depth separable convolution		COCO	70.6
		MobileNet V2	2018	Improved depth separable convolution		COCO	71.7
		TuSimple	2018	Upsampling convolution and mixed dilated convolution		PASCAL VOC 2012	83.1
Method based on probability graphical model		DSM	2016	Modeling CRF through CNN	CRF	PASCAL VOC 2012	78.0
		C&G	2016	Embedding CRF into CNN	CRF	PASCAL VOC 2012	78.1
		DPN	2015	Integrating CNN with MRF	MRF	PASCAL VOC 2012	77.5
		QO	2016	Quadratic optimization	G-CRF	PASCAL VOC 2012	80.2
		HOCRF+	2016	Embedding CRF into CNN	HOCRF	PASCAL VOC 2012	77.9
Method based on feature fusion	Method based on ASPP	DeepLab v3	2017	Improved dilated convolution and improved ASPP	CRF	PASCAL VOC 2012	86.9
		DeepLab v3+	2018	ASPP module with separable convolution and skip join fusion of different level features		PASCAL VOC 2012, Cityscapes	89.0, 82.1
		ICNet	2017	Cascaded model and feature fusion		Cityscapes, CamVid	70.6, 67.1
		DenseASPP	2018	ASPP and densely connected networks to improve receptive field		Cityscapes	80.6
		DMNet	2019	Dynamic convolution module and context-aware correlation filter		PASCAL VOC 2012	84.4
		APCNet	2019	GLA and ACM		PASCAL VOC 2012	84.2
	Method based on attention mechanism	PSANet	2018	Attention mechanism		PASCAL VOC 2012, Cityscapes	85.7, 80.1
		CCNet	2018	Dilated convolution and feature weighted fusion		Cityscapes	81.4
		BiseNet	2018	Spatial path and context path		Cityscapes, CamVid	78.9, 68.7
		ACNet	2019	Three parallel branch architecture and attention assistant module integrating attention mechanism		NYUDv2	48.3
		DANet	2019	Dilated convolution, deconvolution and feature weighted fusion		PASCAL VOC 2012, Cityscapes	82.6, 81.5

续表

Method	Model	Year	Key technology	PGM	Dataset	mIoU / %
Method based on encoding and decoding	SegNet	2015	Deconvolution, upsampling and dropout layer		CamVid	55.6
	DeconvNet	2015	Deconvolution and unpooling		PASCAL VOC 2012	69.6
	RefineNet	2017	Bilinear interpolation skip join and residual join		Cityscapes	73.6
	GCN+	2017	Large kernel convolution and global convolution network		PASCAL VOC 2012, Cityscapes	82.2, 76.9
	DFANet	2019	Deep feature polymerization network		Cityscapes, CamVid	70.3, 64.7
	DUpsampling	2019	Fusion of different resolution features		PASCAL VOC 2012	88.1
	SDN	2019	Capturing multi-scale context information to ensure fine recovery of target location information		PASCAL VOC 2012, CamVid	86.6, 71.8
Method based on RNN	rCNN	2014	Multi size input window		SIFT Flow	
	2D-LSTM	2015	Four different directions of RNN		SIFT Flow	
	ReSeg	2016	Extending of ReNet function		CamVid	
Method based on GAN		2016	GAN adversarial training		PASCAL VOC 2012	54.3
		2016	GAN domain adaptation		Cityscapes	67.8

从表 5 可以看出，PASCAL VOC 2012 数据集更多地应用于静态图像的测试；Cityscapes 和 CamVid 数据集更多地应用于动态场景和实时性较高的场景。针对道路场景语义分割，基于 Cityscapes 数据集，DeepLab v3＋、DenseASPP、DUC＋HDC、PSPNet、PSANet、CCNet 和 DANet 等算法的 mIoU 值均超过了 80%，分割精度基本满足街道场景图像语义分割的精度要求，但实时性有所欠缺。ENet、ESPNet、ICNet 和 BiSeNet 算法虽然分割准确率不如上述算法，但由于尺寸小、计算成本小等特点，这些算法具有实时性强的优势。

针对算法的参数数量和运行速率，本文从基于强监督的图像语义分割方法中选择了代表性较强、实时性高的几种算法，在 Cityscapes 测试数据集上进行了分析对比，其速度分析对比如表 6[6] 所示。

表 6　算法速度分析[5]

Table 6 Speed analysis of algorithms[5]

Model	Parameter	Time /ms	mIoU /%
FCN-8		500	63.1
DeepLab	250.8	4000	63.1
SegNet	29.5	89.2	57
CRF-RNN		700	74.7
ENet	0.4	135.4	57
DeepLab v2	44	4000	70.4
PSPNet	250.8	1288	81.2
DUC ＋ HDC		900	80.1
DenseASPP	28.6	500	80.6
ESPNet	0.4		60.3
BiSeNet1	5.8	13	68.4
BiSeNet2	49	21	74.7
DeepLab v3＋	200＋	600	82.1
ICNet	26.5	33	69.5
DAFNet	7.8	10	71.3

从表 6 中可以看出，在分割速度上，各类算法还是有较大的差异，其中 BiSeNet、ICNet 和 DFANet 算法的速度较快，实时性强，适用于实时图像语义分割。BiSeNet 提出了用于高分辨率图像的浅层网络和快速下采样的深度网络，在分类能力和感受野之间取得了平衡。而 FCN 和基于 FCN 的 DeepLab v1、DeepLab v2 的运行时间较长，无法满足实时图像分割的需求。在 DeepLab 系列中，DeepLab v3＋的分割效果最好，主要是其吸取了 DeepLab 系列方法的优点，并结合深度可分离卷积，使模型得到简化，提高了分割效率，从而实现了图像语义分割精度和速度的均衡。其他算法的分割速度都比 FCN 低，也无法满足实时图像分割的需求，不适用于动态场景分割。因此，在无人驾驶领域，平衡分割精度与分割速度依然是最重要的任务。

4.3.3　基于弱监督的语义分割方法实验对比

在最具代表性的数据集上，对基于弱监督的图像语义分割方法进行了对比，如表 7[5] 所示，主要比较的因素为监督信息、关键技术、是否使用 PGM 方法、实验数据集和评价指标。

表 7　基于弱监督的图像语义分割方法的分析归纳[5]

Table 7 Analysis and summary of image semantic segmentation method based on weak supervision[5]

Supervision information	Model	Year	Key technology	PGM	Dataset	mIoU /%
Frame level	BoxSup	2015	MCG		PASCAL VOC 2012/ PASCAL-CONTEXT	75.2/ 40.5
	DeepCut	2016	CRF	CRF		
Scribble level	WTP	2016	Objectness		PASCAL VOC 2012	49.1
	ScribbleSup	2015	Hyperpixel	CRF	PASCAL VOC 2012	71.3
Image level	MIL	2015	MCG		ImageNet	42.0
	CCNN	2015	Class Size		PASCAL VOC 2012	42.4
	SEC	2016	Saliency detection algorithm	CRF	PASCAL VOC 2012	50.7
	STC	2015	Saliency detection algorithm	CRF	PASCAL VOC 2012	49.8
	AugFeed	2016	MCG	CRF	PASCAL VOC 2012	54.34
	EM	2017	Saliency detection algorithm	CRF	PASCAL VOC 2012	58.71
Image level and pixel level	Decoupled	2015		CRF	PASCAL VOC 2012	66.6
Image level, frame level and pixel level	WeaklySemi	2015		CRF	PASCAL VOC 2012	73.9

从表 7 中可以看出，在基于弱监督的语义分割方法中，虽然图像级标签比较容易获得，但是它包含的有用信息过少，不足以获得准确的分割结果。边界框标签的形式虽然比较复杂，但是能够提供目标位置范围内的监督信息，所以相比于其他基于弱监督的语义分割方法，具有较好的分割结果。总体来说，虽然基于弱监督的图像分割技术大大减少了数据集的标注要求，降低了研究成本，但是包含的有用信息过少，在分割效果和分割性能上与基于强监督的语义分割算法差距较大，不能满足无人驾驶领域的分割要求，不过这将会是未来该领域研究的热点。

4.3.4　基于无监督的语义分割方法实验对比

在最具代表性的数据集上，对基于无监督的图像语义分割方法进行了对比，如表 8[9] 所示，主要比较的因素为关键技术、是否使用 PGM 方法、实验数据集和评价指标。

表 8　基于无监督的图像语义分割方法的分析归纳[9]

Table 8 Analysis and summary of image semantic segmentation method based on unsupervision[9]

Model	Year	Key technology	Dataset	mIoU /%
FCNWild	2016	Domain adaptive full convolution adversarial training	Cityscapes	27.1
ADDA	2017	Adversarial training	NYU Depth v2	
FCAN	2018	Image domain adaptive network and feature adaptive network	Cityscapes	47.75

基于无监督的图像语义分割方法主要是使用虚拟场景，对现实场景进行数据标注，进而完成语义分

割。该方法降低了标注成本，简化了分割过程，但是需要对虚拟场景、现实场景之间的差别有客观的认识和理解，纹理、光照等方面的差异往往能够降低现实场景中的图像分割精度和准确度，产生一定的分割偏差。大量研究数据表明，目前基于无监督的图像语义分割方法的有效精度并不高，进一步提升其分割精度、提高分割质量将会是未来研究的重点和热点。

通过上述分析可以看出，在自动驾驶领域，基于强监督的语义分割方法依旧是目前主流的道路场景分割方法，在考虑分割精度的同时也要考虑分割效率。基于弱监督和无监督的图像语义分割方法减小了标注成本，但目前分割效果不明显，分割边界粗糙且不连续，提高其分割精度是今后研究的热点。

5　结束语

近年来，随着自动驾驶等应用的不断发展，人们对模型尺寸、计算成本、分割精度等方面提出了更高的要求。介绍了道路场景语义分割的发展现状与挑战。将语义分割技术划分为传统模式、传统与深度学习相结合的模式以及基于深度学习的模式，重点介绍了基于深度学习的模式，将其进一步细分为基于强监督的图像语义分割方法、基于弱监督的图像语义分割方法和基于无监督的图像语义分割方法。针对道路场景，对每类方法的代表性算法进行了分析和对比，概括总结了每类方法的技术特点和优缺点。总体来看，基于深度学习对道路场景进行语义分割的技术还在不断发展，但是也有一些需要改进的地方。

1）语义分割算法的精度有待进一步提高。无人驾驶的核心在于对周围环境的精细化感知和判断，例如行驶过程中周围天气的变化、交通指示灯的变化以及来往车辆和行人，这就要求对输入的分割对象进行精确的分割。

2）实时语义分割技术[101-102]。现阶段精确率依然是评价语义分割网络模型的重点指标，但是随着无人驾驶技术的不断成熟，分割效率的影响越来越大，这就需要在维持高精确率的基础上尽量缩短响应时间。

3）基于弱监督或无监督的语义分割技术。目前基于弱监督和无监督的语义分割技术的分割效果还不理想，利用尽量少的标注信息来提高网络模型的精度是未来发展的趋势。

4）三维数据的应用[103-104]。三维数据对真实场景至关重要，目前大多语义分割方法的分割对象都是二维场景，因此三维数据的应用将会是未来的研究热点。

参　考　文　献

[1] Zhou J M, Li B J, Chen S Z. A real time semantic segmentation method based on multi-level feature fusion[J]. Bulletin of Surveying and Mapping, 2020(1): 10-15.
周继苗, 李必军, 陈世增. 一种多层特征融合的道路场景实时分割方法[J]. 测绘通报, 2020(1): 10-15.

[2] Deng L Y, Yang M, Liang Z D, et al. Fusing geometrical and visual information via superpoints for the semantic segmentation of 3D road scenes[J]. Tsinghua Science and Technology, 2020, 25(4): 498-507.

[3] Liu S T, Yin F L. The basic principle and its new advances of image segmentation methods based on graph cuts[J]. Acta Automatica Sinica, 2012, 38(6): 911-922.
刘松涛, 殷福亮. 基于图割的图像分割方法及其新进展[J]. 自动化学报, 2012, 38(6): 911-922.

[4] Tian X, Wang L, Ding Q. Review of image semantic segmentation based on deep learning[J]. Journal of Software, 2019, 30(2): 440-468.
田萱, 王亮, 丁琪. 基于深度学习的图像语义分割方法综述[J]. 软件学报, 2019, 30(2): 440-468.

[5] Jing Z W, Guan H Y, Peng D F, et al. Survey of research in image semantic segmentation based on deep neural network[J]. Computer Engineering, 2020, 46(10): 1-17.
景庄伟, 管海燕, 彭代峰, 等. 基于深度神经网络的图像语义分割研究综述[J]. 计算机工程, 2020, 46(10): 1-17.

[6] Wang Y, Zhang H J, Huang H X. A survey of image semantic segmentation algorithms based on deep learning[J]. Application of Electronic Technique, 2019, 45(6): 23-27,36.
王宇, 张焕君, 黄海新. 基于深度学习的图像语义分割算法综述[J]. 电子技术应用, 2019, 45(6): 23-27,36.

[7] Zhang X F, Liu J, Shi Z S, et al. Review of deep learning-based semantic segmentation[J]. Laser & Optoelectronics Progress, 2019, 56(15): 150003.
张祥甫, 刘健, 石章松, 等. 基于深度学习的语义分割问题研究综述[J]. 激光与光电子学进展, 2019, 56(15): 150003.

[8] Luo H L, Zhang Y. A survey of image semantic segmentation based on deep network[J]. Acta Electronica Sinica, 2019, 47(10): 2211-2220.
罗会兰, 张云. 基于深度网络的图像语义分割综述[J]. 电子学报, 2019, 47(10): 2211-2220.

[9] Wang Y R, Chen Q L, Wu J J. Research on image semantic segmentation for complex environments[J]. Computer Science, 2019, 46(9): 36-46.
王嫣然, 陈清亮, 吴俊君. 面向复杂环境的图像语义分割方法综述[J]. 计算机科学, 2019, 46(9): 36-46.

[10] Kuang H Y, Wu J J. Survey of image semantic segmentation based on deep learning[J]. Computer Engineering and Applications, 2019, 55(19): 12-21,42.
邝辉宇, 吴俊君. 基于深度学习的图像语义分割技术研究综述[J]. 计算机工程与应用, 2019, 55(19): 12-21,42.

[11] Minaee S, Boykov Y, Porikli F, et al. Image segmentation using deep learning: a survey[EB/OL]. (2020-01-15)[2020-06-15]. https://export.arxiv.org/pdf/2001.05566.

[12] Zhang J Y, Zhao X L, Chen Z. Review of semantic segmentation of point cloud based on deep learning[J]. Laser & Optoelectronics Progress, 2020, 57(4): 040002.
张佳颖, 赵晓丽, 陈正. 基于深度学习的点云语义分割综述[J]. 激光与光电子学进展, 2020, 57(4): 040002.

[13] Tian Q C, Meng Y. Image semantic segmentation based on convolutional neural network[J]. Journal of Chinese Computer Systems, 2020, 41(6): 1302-1313.
田启川, 孟颖. 卷积神经网络图像语义分割技术[J]. 小型微型计算机系统, 2020, 41(6): 1302-1313.

[14] Khan M W. A survey: image segmentation techniques[J]. International Journal of Future Computer and Communication, 2014,3(2): 89-93.

[15] Yang Y P, Zhao W D, Wang Z C, et al. Research on graph-based Normalized Cut image segmentation method[J]. Computer and Modernization, 2010(1): 113-116.
杨宇鹏, 赵卫东, 王志成, 等. 基于图论的Normalized Cut图像分割方法研究[J]. 计算机与现代化, 2010(1): 113-116.

[16] Zheng Q H, Li W Q, Hu W H, et al. An interactive image segmentation algorithm based on graph cut[J]. Procedia Engineering, 2012, 29: 1420-1424.

[17] Han X. Research on aotomatic image segmentation algorithm based on Grab Cut[D]. Beijing: Beijing Institute of Graphic Communication, 2018: 8-9.
韩旭. 基于Grab Cut的图像自动分割算法研究[D]. 北京:北京印刷学院,2018:8-9.

[18] Liu L, Shi Z G, Su H R, et al. Image segmentation based on higher order Markov random field[J]. Journal of Computer Research and Development, 2013, 50(9): 1933-1942.
刘磊, 石志国, 宿浩茹, 等. 基于高阶马尔可夫随机场的图像分割[J]. 计算机研究与发展, 2013, 50(9): 1933-1942.

[19] Arbeláez P, Maire M, Fowlkes C, et al. Contour detection and hierarchical image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(5): 898-916.

[20] Zhang C J, Xue Z, Zhu X B, et al. Boosted random contextual semantic space based representation for visual recognition[J]. Information Sciences, 2016, 369: 160-170.

[21] Pont-Tuset J, Arbeláez P, Barron J T, et al. Multiscale combinatorial grouping for image segmentation and object proposal generation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(1): 128-140.

[22] Elhofi A H, Helaly H A. Comparison between digital and manual marking for toric intraocular lenses: a randomized trial[J]. Medicine, 2015, 94(38): e1618.

[23] Wang C Y, Chen J Z, Li W. Review on superpixel segmentation algorithms[J]. Application Research of Computers, 2014, 31(1): 6-12.
王春瑶, 陈俊周, 李炜. 超像素分割算法研究综述[J]. 计算机应用研究, 2014, 31(1): 6-12.

[24] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.

[25] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[EB/OL]. (2016-04-30)[2020-06-15]. https://arxiv.org/abs/1511.07122.

[26] Chen L C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs[EB/OL]. (2016-06-07)[2020-06-10]. https://arxiv.org/abs/1412.7062v2.

[27] Paszke A, Chaurasia A, Kim S, et al. ENet: a deep

neural network architecture for real-time semantic segmentation [EB/OL]. (2016-06-07) [2020-06-15]. https://arxiv.org/abs/1606.02147.

[28] Yu F, Koltun V, Funkhouser T. Dilated residual networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 636-644.

[29] Fang Y C, Li Y F, Tu X K, et al. Face completion with hybrid dilated convolution[J]. Signal Processing: Image Communication, 2020, 80: 115664.

[30] Dai J F, Qi H Z, Xiong Y W, et al. Deformable convolutional networks[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 764-773.

[31] Ghiasi G, Fowlkes C C. Laplacian pyramid reconstruction and refinement for semantic segmentation[M]// Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9907: 519-534.

[32] Arnab A, Jayasumana S, Zheng S, et al. Higher order conditional random fields in deep neural networks[M]// Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9906: 524-540.

[33] Vemulapalli R, Tuzel O, Liu M Y, et al. Gaussian conditional random field network for semantic segmentation [C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 3224-3233.

[34] Shen F L, Gan R, Yan S C, et al. Semantic segmentation via structured patch prediction, context CRF and guidance CRF [C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5178-5186.

[35] Jiang J D, Zhang Z J, Huang Y Q, et al. Incorporating depth into both CNN and CRF for indoor semantic segmentation[C]//2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), November 24-26, 2017, Beijing, China. New York: IEEE Press, 2017: 525-530.

[36] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.

[37] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.

[38] Wang P Q, Chen P F, Yuan Y, et al. Understanding convolution for semantic segmentation[C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV), March 12-15, 2018, Lake Tahoe, NV, USA. New York: IEEE Press, 2018: 1451-1460.

[39] Chen L C, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [M] // Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 833-851.

[40] Yang M K, Yu K, Zhang C, et al. DenseASPP for semantic segmentation in street scenes [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 3684-3692.

[41] He J J, Deng Z Y, Qiao Y. Dynamic multi-scale filters for semantic segmentation[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 3561-3571.

[42] Zhao H S, Shi J P, Qi X J, et al. Pyramid scene parsing network [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6230-6239.

[43] Zhao H S, Qi X J, Shen X Y, et al. ICNet for real-time semantic segmentation on high-resolution images[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11207: 418-434.

[44] He J J, Deng Z Y, Zhou L, et al. Adaptive pyramid context network for semantic segmentation [C] //

2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 7511-7520.

[45] Wu H K, Zhang J G, Huang K Q, et al. FastFCN: rethinking dilated convolution in the backbone for semantic segmentation [EB/OL]. (2019-03-28) [2020-06-15]. https://arxiv.org/abs/1903.11816.

[46] Zhao H S, Zhang Y, Liu S, et al. PSANet: point-wise spatial attention network for scene parsing [M]// Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11213: 270-286.

[47] Yuan J J, Zhang L, Chen Y H. Deep neural network based on attention convolution module for image recognition [J]. Computer Engineering and Applications, 2019, 55(8): 9-16.
袁嘉杰, 张灵, 陈云华. 基于注意力卷积模块的深度神经网络图像识别[J]. 计算机工程与应用, 2019, 55(8): 9-16.

[48] Feng S T, Zhuo Z S, Pan D R, et al. CcNet: a cross-connected convolutional network for segmenting retinal vessels using multi-scale features [J]. Neurocomputing, 2020, 392: 268-276.

[49] Yu C Q, Wang J B, Peng C, et al. BiSeNet: bilateral segmentation network for real-time semantic segmentation[M]// Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11217: 334-349.

[50] Luo C, Xin W, Li X J, et al. ACNET: attention-based convolution network with additional discriminative features for DCM classification [EB/OL]. [2020-06-15]. http://ksiresearchorg.ipage.com/seke/seke19paper/seke19paper_155.pdf.

[51] Xue H L, Liu C, Wan F, et al. DANet: divergent activation for weakly supervised object localization [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea. New York: IEEE Press, 2019: 6588-6597.

[52] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[53] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[M]// Navab N, Hornegger J, Wells W, et al. Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.

[54] Wu Z S, Fu W P, Han G N. Road scene understanding based on deep convolutional neural network[J]. Computer Engineering and Applications, 2017, 53(22): 8-15.
吴宗胜, 傅卫平, 韩改宁. 基于深度卷积神经网络的道路场景理解[J]. 计算机工程与应用, 2017, 53(22): 8-15.

[55] Yan Y Y, Qu X X, Zhu Q Y, et al. Confidence measure method of classification results based on outlier detection[J]. Journal of Nanjing University (Natural Science), 2019, 55(1): 102-109.
严云洋, 瞿学新, 朱全银, 等. 基于离群点检测的分类结果置信度的度量方法[J]. 南京大学学报(自然科学), 2019, 55(1): 102-109.

[56] Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1520-1528.

[57] Li Q B, Su D. Multi-organ abdominal image segmentation based on V-Net[J]. Digital Technology & Application, 2019(1): 89,91.
李庆勃, 苏丹. 基于 V-Net 的腹部多器官图像分割[J]. 数字技术与应用, 2019(1): 89,91.

[58] Lin G S, Milan A, Shen C H, et al. RefineNet: multi-path refinement networks for high-resolution semantic segmentation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5168-5177.

[59] Tian Z, He T, Shen C H, et al. Decoders matter for semantic segmentation: data-dependent decoding enables flexible feature aggregation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 3121-3130.

[60] Yang L, Wu Y X, Wang J L, et al. Research on recurrent neural network[J]. Journal of Computer Applications, 2018, 38(S2): 1-6,26.
杨丽, 吴雨茜, 王俊丽, 等. 循环神经网络研究综述[J]. 计算机应用, 2018, 38(S2): 1-6,26.

[61] Visin F, Romero A, Cho K, et al. ReSeg: a

recurrent neural network-based model for semantic segmentation[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 26-July 1, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 426-433.

[62] Li Z, Gan Y K, Liang X D, et al. LSTM-CF: unifying context modeling and fusion with LSTMs for RGB-D scene labeling[M]// Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9906: 541-547.

[63] Liang X D, Shen X H, Feng J S, et al. Semantic object parsing with graph LSTM[M]//Leibe B, Matas J, Sebe N, et al. Computer vision - ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 125-143.

[64] Zheng S, Jayasumana S, Romera-Paredes B, et al. Conditional random fields as recurrent neural networks[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1529-1537.

[65] Wang K F, Gou C, Duan Y J, et al. Generative adversarial networks: the state of the art and beyond [J]. Acta Automatica Sinica, 2017, 43(3): 321-332.
王坤峰, 苟超, 段艳杰, 等. 生成式对抗网络 GAN 的研究进展与展望[J]. 自动化学报, 2017, 43(3): 321-332.

[66] Luc P, Couprie C, Chintala S, et al. Semantic segmentation using adversarial networks[EB/OL]. (2016-11-25)[2020-06-15]. https://arxiv.org/abs/1611.08408v1.

[67] Xue Y, Xu T, Zhang H, et al. SegAN: adversarial network with multi-scale L_1 loss for medical image segmentation[J]. Neuroinformatics, 2018, 16(3/4): 383-392.

[68] Dai J F, He K M, Sun J. BoxSup: exploiting bounding boxes to supervise convolutional networks for semantic segmentation[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1635-1643.

[69] Rajchl M, Lee M C H, Oktay O, et al. DeepCut: object segmentation from bounding box annotations using convolutional neural networks[J]. IEEE Transactions on Medical Imaging, 2017, 36(2): 674-683.

[70] Song C F, Huang Y, Ouyang W L, et al. Box-driven class-wise region masking and filling rate guided loss for weakly supervised semantic segmentation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 3131-3140.

[71] Bearman A, Russakovsky O, Ferrari V, et al. What's the point: semantic segmentation with point supervision[M]// Leibe B, Matas J, Sebe N, et al. Computer vision - ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9911: 549-565.

[72] Lin D, Dai J F, Jia J Y, et al. ScribbleSup: scribble-supervised convolutional networks for semantic segmentation[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 3159-3167.

[73] Maninis K K, Caelles S, Pont-Tuset J, et al. Deep extreme cut: from extreme points to object segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 616-625.

[74] Kolesnikov A, Lampert C H. Seed, expand and constrain: three principles for weakly-supervised image segmentation[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9908: 695-711.

[75] Huang Z L, Wang X G, Wang J S, et al. Weakly-supervised semantic segmentation network with deep seeded region growing[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7014-7023.

[76] Wang Y J, Wang G D, Chen C, et al. Multi-scale dilated convolution of convolutional neural network for image denoising[J]. Multimedia Tools and Applications, 2019, 78(14): 19945-19960.

[77] Ahn J, Kwak S. Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City,

UT, USA. New York: IEEE Press, 2018: 4981-4990.

[78] Zhou Y Z, Zhu Y, Ye Q X, et al. Weakly supervised instance segmentation using class peak response [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 3791-3800.

[79] Wei Y C, Liang X D, Chen Y P, et al. STC: a simple to complex framework for weakly-supervised semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(11): 2314-2320.

[80] Mukhopadhyay S. Stochastic gradient descent for linear systems with sequential matrix entry accumulation [J]. Signal Processing, 2020, 171: 107494.

[81] Hong S, Noh H, Han B. Decoupled deep neural network for semi-supervised semantic segmentation [EB/OL]. (2015-06-17) [2020-06-15]. https://arxiv.org/abs/1506.04924.

[82] Donahue J, Jia Y Q, Vinyals O, et al. DeCAF: a deep convolutional activation feature for generic visual recognition[EB/OL]. (2013-10-06)[2020-06-15]. https://arxiv.org/abs/1310.1531.

[83] Tzeng E, Hoffman J, Saenko K, et al. Adversarial discriminative domain adaptation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2962-2971.

[84] Hoffman J, Wang D Q, Yu F, et al. FCNs in the wild: pixel-level adversarial and constraint-based adaptation[EB/OL]. (2016-12-08)[2020-06-15]. https://arxiv.org/abs/1612.02649v1.

[85] Zhang Y H, Qiu Z F, Yao T, et al. Fully convolutional adaptation networks for semantic segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 6810-6818.

[86] Brostow G J, Fauqueur J, Cipolla R. Semantic object classes in video: a high-definition ground truth database[J]. Pattern Recognition Letters, 2009, 30(2): 88-97.

[87] Geiger A, Lenz P, Stiller C, et al. Vision meets robotics: the KITTI dataset[J]. The International Journal of Robotics Research, 2013, 32(11): 1231-1237.

[88] Maddern W, Pascoe G, Linegar C, et al. 1 year, 1000 km: the Oxford RobotCar dataset[J]. The International Journal of Robotics Research, 2017, 36(1): 3-15.

[89] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 3213-3223.

[90] Ros G, Sellart L, Materzynska J, et al. The SYNTHIA dataset: a large collection of synthetic images for semantic segmentation of urban scenes [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 3234-3243.

[91] Neuhold G, Ollmann T, Bulò S R, et al. The mapillary vistas dataset for semantic understanding of street scenes [C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 5000-5009.

[92] Huang X Y, Wang P, Cheng X J, et al. The ApolloScape open dataset for autonomous driving and its application [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(10): 2702-2719.

[93] Yu F, Chen H F, Wang X, et al. BDD100K: a diverse driving dataset for heterogeneous multitask learning [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 2633-2642.

[94] Buyval A, Gabdullin A, Mustafin R, et al. Realtime vehicle and pedestrian tracking for Didi udacity self-driving car challenge[C]//2018 IEEE International Conference on Robotics and Automation (ICRA), May 21-25, 2018, Brisbane, QLD, Australia. New York: IEEE Press, 2018: 2064-2069.

[95] Gu Z C, Li Z H, Di X, et al. An LSTM-based autonomous driving model using a Waymo open dataset[J]. Applied Sciences, 2020, 10(6): 2046.

[96] Gudigar A, Chokkadi S, Raghavendra U, et al. An efficient traffic sign recognition based on graph embedding features [J]. Neural Computing and

Applications, 2019, 31(2): 395-407.

[97] Houben S, Stallkamp J, Salmen J, et al. Detection of traffic signs in real-world images: the German traffic sign detection benchmark [C] // The 2013 International Joint Conference on Neural Networks (IJCNN), August 4-9, 2013, Dallas, TX, USA. New York: IEEE Press, 2013: 1-8.

[98] Zhu Y Y, Zhang C Q, Zhou D Y, et al. Traffic sign detection and recognition using fully convolutional network guided proposals [J]. Neurocomputing, 2016, 214: 758-766.

[99] Lee E, Kim D. Accurate traffic light detection using deep neural network with focal regression loss [J]. Image and Vision Computing, 2019, 87: 24-36.

[100] Song S J, Que Z Q, Hou J J, et al. An efficient convolutional neural network for small traffic sign detection [J]. Journal of Systems Architecture, 2019, 97: 269-277.

[101] Lu W C, Pang Y W, He Y Q, et al. Real-time and accurate semantic segmentation based on separable residual modules [J]. Laser & Optoelectronics Progress, 2019, 56(5): 051005.
路文超, 庞彦伟, 何宇清, 等. 基于可分离残差模块的精确实时语义分割[J]. 激光与光电子学进展, 2019, 56(5): 051005.

[102] Cai Y, Huang X G, Zhang Z A, et al. Real-time semantic segmentation algorithm based on feature fusion technology [J]. Laser & Optoelectronics Progress, 2020, 57(2):021011.
蔡雨, 黄学功, 张志安, 等. 基于特征融合的实时语义分割算法[J]. 激光与光电子学进展, 2020, 57(2):021011.

[103] Yang J, Dang J S. Recognition and segmentation of three-dimensional point cloud based on deep cascade convolutional neural network [J]. Optics and Precision Engineering, 2020, 28(5): 1187-1199.
杨军, 党吉圣. 采用深度级联卷积神经网络的三维点云识别与分割[J]. 光学 精密工程, 2020, 28(5): 1187-1199.

[104] Zhang A W, Liu L L, Zhang X Z. Multi-feature 3D road point cloud semantic segmentation method based on convolutional neural network [J]. Chinese Journal of Lasers, 2020, 47(4): 0410001.
张爱武, 刘路路, 张希珍. 道路三维点云多特征卷积神经网络语义分割方法[J]. 中国激光, 2020, 47(4): 0410001.