激光与光电子学进展
Laser & Optoelectronics Progress
ISSN 1006-4125,CN 31-1690/TN

《激光与光电子学进展》网络首发论文

题目：道路场景语义分割综述
作者：王龙飞，严春满
收稿日期：2020-07-09
网络首发日期：2020-10-15
引用格式：王龙飞，严春满．道路场景语义分割综述[J/OL]．激光与光电子学进展.
https://kns.cnki.net/kcms/detail/31.1690.TN.20201015.0915.002.html

网络首发时间：2020-10-15 13:45:53
网络首发地址：https://kns.cnki.net/kcms/detail/31.1690.TN.20201015.0915.002.html

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，为自动驾驶汽车的安全行驶提供了保障。本文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，随着深度学习的出现和不断发展,本文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，其中重点介绍基于深度学习的语义分割技术，并将其按照强监督、弱监督、无监督三种不同的网络训练方式进行阐述；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：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

A Survey of Semantic Segmentation of Road Scenes

Wang Longfei，Yan Chunman*,

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

Abstract Image semantic segmentation is an important research field of computer vision and 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. This paper starts with the definition of semantic segmentation of road scenes, and discusses the current challenges in this field; Secondly, with the emergence and continuous development of deep learning, this paper divides semantic segmentation technology into traditional segmentation technology, traditional segmentation technology combined with deep learning and segmentation technology based on deep learning. This paper focuses on the semantic segmentation technology based on deep learning, and explain it according to three different network training methods: strong supervision, weak supervision, and unsupervised. Then the data sets and performance evaluation indicators related to the semantic segmentation of road scenes are summarized, and based on this comparison, the segmentation results of common image semantic segmentation methods are analyzed. Finally, the challenges faced by the road scene semantic segmentation technology and the future development direction are prospected.

Key words： 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。针对道路场景进行语义分割是对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，实现道路场景图像在像素级别上的分类。在自动驾驶的技术组成单元中,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,使其做出正确的判断,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，是一项十分具有研究意义的技术，成为了当前研究的热点。

机遇与挑战并存,在自动驾驶技术发展势头十分迅猛之时,其核心道路场景的语义分割也面临着许多挑战。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精确性会受到不同行驶区域

基金项目：国家自然科学基金（61861041）
*E-mail: yancha02@163.com

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的不同也会使得目标物体与图片上差异较大,进而影响分割的效果。这些因素都极大的提升了图像语义分割的难度,进而影响到无人驾驶的实现。综上所述,在无人驾驶的研究中，道路场景的语义分割是一项十分关键并且充满挑战性的技术。

语义分割技术作为目前计算机视觉研究的热点方向，已有一些文献[3-13]对其成果进行了综述，但是，针对于无人驾驶领域的道路场景语义分割方法并没有全面的综述性的文献，因此本文进行了相关工作，第 1 节对自动驾驶技术和其核心道路场景语义分割技术进行了概述；第 2 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归纳与总结，并对不同时期的语义分割方法进行了综述；第 3 节重点对当前的研究热点：基于深度学习的语义分割方法进行了分析，通过监督信息的不同，将这个时期的语义分割方法分为强监督、弱监督与无监督 3 种类别并对每一类别进行进一步的分类与分析；第 4 节针对道路场景，对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进行了总结，并对本文提到所有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对比和分析；第五节，对本文的工作进行总结，对本文研究方向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，如何在两者中取得平衡是道路场景语义分割的关键，目前基于强监督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 图像语义分割发展历史

从演变过程来讲，图像的语义分割技术的发展过程如图 1 所示。（基于深度学习的语义分割阶段的详细分类见图 11 和图 17）

（1）传统的语义分割阶段。由于所处时期计算处理能力有限，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。以此方式得到的分割结果相对粗陋，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。从一定程度上讲，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，而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。该种方法传承使用了传统的分割方法，因此，该种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，所以当前情况下，基于深度学习技术的语义分割方法得到了普及和推广，并且取得了比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。

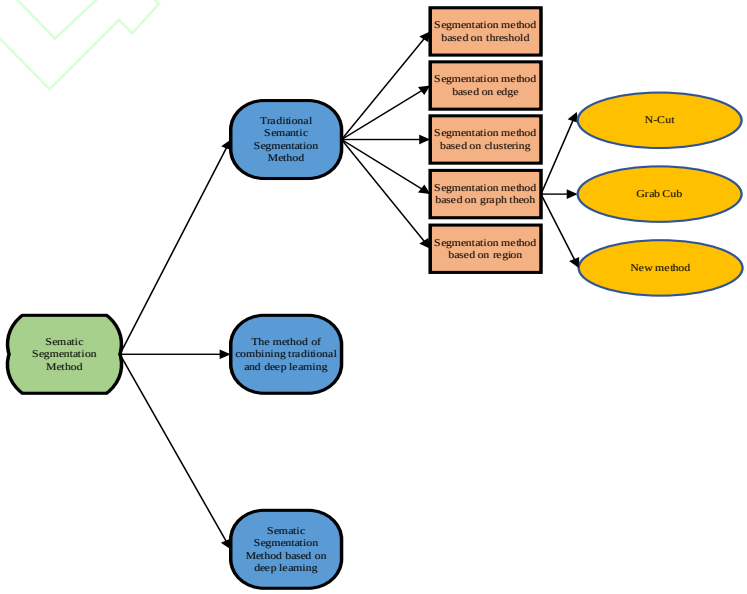

图 1 图像语义分割发展历史

Fig. 1 Development history of image semantic segmentation

2.1 传统的图像语义分割算法

在深度学习还没有广泛应用与计算机视觉之前，传统图像语义分割方法为了将目标与背景分离，利用多种特征，例如：颜色、灰度、纹理、几何形状等把图像划分成多个独立的区域。这一阶段的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]，将在下文中进行说明。

2.1.1 Normalized cut 图像分割算法

Jianbo Shi 等学者在本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，将其定义为“图”并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：以图为单位，然后计算权重图(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）作为其中的一个重要的方法，我们可以通过图 2 权重图来理解。我们把图 2 看成一个整体 G，我们的任务是把它分割成两个部分。从图中不难看出，中间横着的两条权重为 0.1 的边就是最小化切割。通过这个思路进行最小化分割就可以很完美的把这个整体 G 分割成两部分，但是最小化切割也存在边缘角元素缺失等缺陷，如图 3 所示，这使得最终的结果存在偏差。

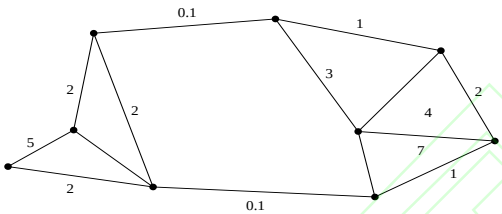

图 2 权重图

Fig. 2 Weight map

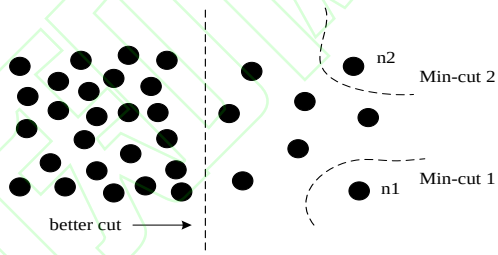

图 3 最小化切割效果图

Fig. 3 The effect picture of minimizing cutting

2.1.2 Grab cut 图像分割算法

本世纪初由欧美研究机构提出的 GraphCut[16-17]算法同样以图片分割为理论基础；该机构在研究过程中利用到混合高斯模型以及吉尔斯能量方程， 基于 RGB 实现建模效果并形成算法的全局基础，在求得方程最优解过程中采用迭代方式，最终获取高斯模 型的最优参数解。该算法的提出显著拓宽图像分割领域并实现对彩色图像的分割。

Grab cut 虽然在分割性能上有所提升但是它在使用上的便利性和广泛性较差，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，用户的初始化不佳也会影响其分割效果。刘磊等[18]将高阶势能项引入 Grab Cut，使得其可以更好的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

传统方法比较流行的有以下三种检测法。最初，2011 年由 Arbeláez 等人[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 方法对任一像素作边缘的实际概率进行合理测算，而后将该测算结果形成不同的闭合区域，随后利用 UCM 法使不同的闭合区间进行转化，形成层次分明的树状结构。随着研究不断深入，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，该检测法主要由不同的决策树进行组合形成分类器。2017 年，包括 Pont-Tuest[21]在内的研究小组综合了以上两种检测方法，提出了新的检测方法—即 MCG 算法。该方法在检测分割过程中，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，而后使用随机法形成的分类器做进一步的分割处理。可以说，该方法实现了传统模式、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表现在重点使用表层特征和外部结构特点完成图像分割，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，推动了语义分割技术的变革。业内专家学者在进行语义分割研究过程中，逐步开发了深入学习的思维方式，引入了深度学习算法模型。按照首先采用传统方法进行初步分割，得出目标区域，而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，最终实现分割目标区域、完成自动标注的目的。

从源头上讲，Farabet 首次于 2013 年创新使用深度学习技术，优化调整语义分割方法。新形成的分割方法运用卷积神经网原理，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。随后，通过超像素以及无参数多级解析等多个处理过程形成最终结果。与此同时期，包含 Couprie 在内的另一个科研团队，在分割室内场景过程中，综合运用了图像和深度图技术。该团队使用的算法在基本流程上相对简单，首先对图像的滤波特征、卷积特征等进行合理提取，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，可以使用该分类器做进一步的分类。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，使用超像素分割法对弱边界的图像区域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更能获取更多，更高级的语义信息来表达图像中的信息。在深度学习引入语义分割领域以来，作为衡量语义分割效果的重要指标，如何提高分割精度一直是研究的热门方向。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进步，是该方向具有里程碑意义的成果。因此，许多基于 FCN 的语义分割方法相继出现，本章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，主要优缺点如表 1 所示。

表 1 强监督、弱监督及无监督语义分割方法的优缺点对比

Table 1 Comparison of advantages and disadvantages of strong supervised, weakly supervised and unsupervised semantic segmentation methods

Type	Advantages	Disadvantages
Strong supervision	Based on densely annotated data sets, with high segmentation accuracy	Excessive reliance on the data set marked by dense set, inability to migrate, and poor segmentation accuracy for unknown scenes
Weak supervision	Only need image-level annotated data set to complete training	Need to train a large number of data sets, which takes a long time and the accuracy is lower than that of strong supervision
Unsupervised	It does not rely on the manual intensive annotation data set, and has strong adaptability to unknown environment	It is difficult to adapt and the segmentation accuracy is not high at present

3.1 强监督语义分割方法

对样本进行人工标注可以体现出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，增强分割精度。可以说，强监督学习模型是当前应用程度最广的分割模型，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，全卷积神经网络（FCN）模型的提出无疑是具有里程碑意义的，其重要的意义在于为后期的模型算法研究指出了全新的方向。在 FCN 网络模型中，其研究重点在于以

像素为单位，深度探究语义分割的基本原理和有效过程。其网络结构示意图如图 4 所示。

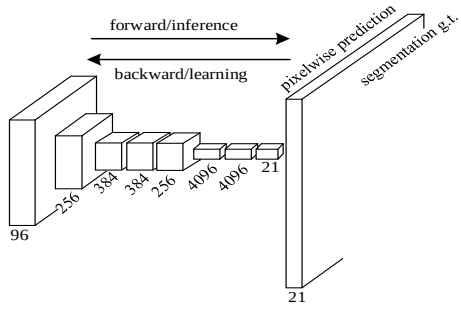

图 4 全卷积网络结构示意图

Fig.4 Structural diagram of fully convolutional network

在全卷积网络（FCN）结构中以一定数量其尺寸固定的卷积层，起到常规卷积网络中全连接层的作用；在全卷积网络（FCN）中包括的卷积层与采样层则分别涉及到上下、正反等多种类型，且上述层次在空间任意平移时保持结构不变。恢复图片原始分辨率大小是全卷积网络的常见应用场景，处理过程中利用到反卷积形式。全卷积网络（FCN）中通过多个固定尺寸的卷积层承担传统结构中全连接层的任务，这种结构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，与神经网络在图片中的自由滑动密切相关。然而全卷积网络（FCN）仍然保留使用了卷积神经网络（CNN）中的池化层，池化层使得卷积神经网络增加了感受野并且进行了融合特征，但是连续的下采样，会导致细节丢失，极大地影响了分割的结果，例如，处理较大物体的时候，容易将标签分配错误，处理较小的物体的时候，容易忽略其存在。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,根据这些方法的改进特点不同,我们将其划分为 6 类: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、基于循环神经网络的分割方法和基于生成对抗网络的分割方法。如图 5 所示。

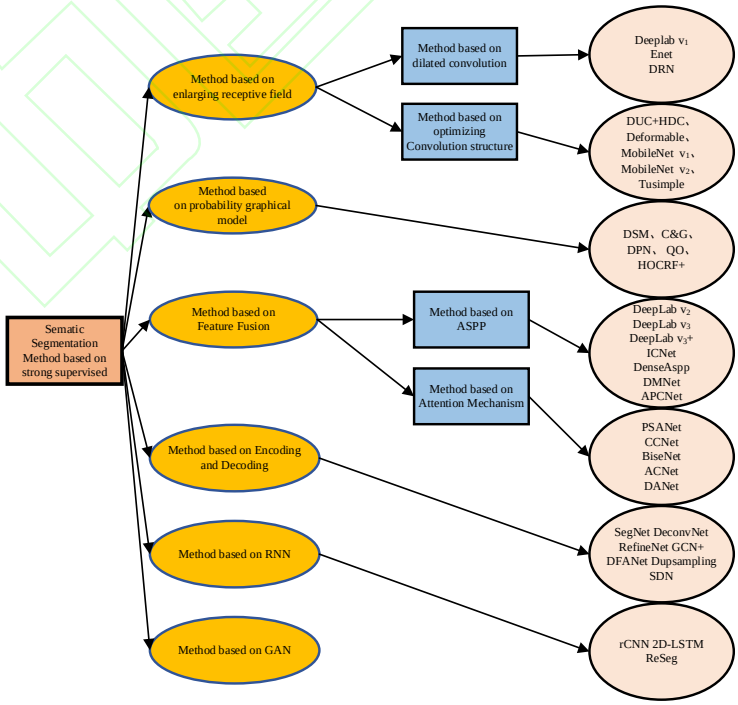

图 5 基于强监督的语义分割方法

Fig.5 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]是由 Fisher 等人提出的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。该方法由学者 chen 于近年来提出，对卷积神经网 络分辨率的影响着重体现在计算特征响应领域。扩展卷积模式的提出与空洞卷积相呼应，对其作用的描述 以多种卷积的感受野为依据；如图 6 所示，选择 3×3 卷积，对比扩张系数为 1、2、4 下相同卷积的感受野， 不难看出与扩张系数呈正相关关系，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 作用是十分显著的。且扩张卷积的使用能够扩充卷积的堆叠效应，同样提升感受野的大小。空洞卷积的使 用则专注于分辨率和计算响应能力的提升，降低计算中对参数和计算过程的依赖度，仅需要输入较少的参 数或因子实现对卷积核感受野的扩大效应，同时有助于前后内容的获取。

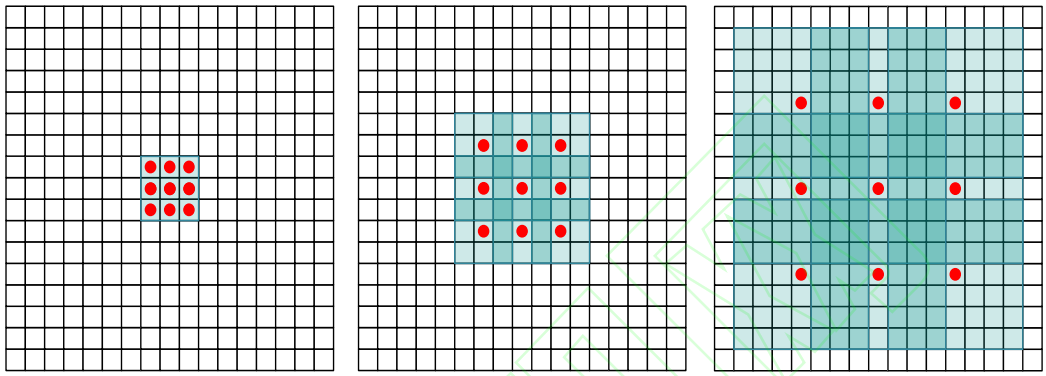

图 6 扩张卷积示意图 （a）普通卷积 （b）扩张率为 2 的扩张卷积 （c）扩张率为 4 的扩张卷积

Fig.6 Schematic diagram of expansion convolution (a) ordinary convolution (b) expansion convolution with an expansion rate of 2 (c) expansion convolution with an expansion rate of 4

DeepLab v1[26]网络模型由 Chen 等人于 2014 年提出，DeepLab v1 创新性地将空洞卷积应用到 VGG16 网络，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，恢复感受野至原图像大小，提升了模型分割的准确率。2016 年，Adam Paszke[27]提出了一种实时分割的模型—ENet。该种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，以此调整感受野的实际区域大小，有效破解了特征分辨率持续下降等不良问题。该算法模型中，运用参数较少，运行速度较快，一定程度上推动了实时分割技术的发展。2017 年， Fisher Yu等人[28] 提出了 DRN 网络模型。研究表明，该模型立足 ResNet 网络基础，运用空洞卷积对普通卷积进行替换操作，以此维持原图像的实际分辨率和原网络的有效感受野区域。该模型中，由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，以此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，需要借助移除残差和最大池化层等方法进行操作处理。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，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不断的下降，会造成部分空间信息遗失。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的获取图像的深度特征，能扩大感受野范围，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，以至于出现数据遗失、消息丢失等不良问题。在文献[29]中，研究人员运用了混合扩张卷积(hybrid dilated convolution, HDC)对扩张卷积进行了替代操作，同时运用了稠密上采样的算法取代了 BI 算法。由于 HDC 卷积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。从一定程度上讲，以上方法可以有效增强感受野，但是由于卷积核的形状相对固定，模拟几何变换的处理能力相对较弱，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。文献[30]中，研究人员在进行卷积处理过程中，运用了有一定偏移量的采样操作，引入了可学习的一个偏移量，最终调整卷积核的形状，使其具有可变性，以此得出了可变形卷积(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，不断提高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，会通过较少的计算量，来降低性能消耗。实践活动中，在移动设备应用端使用的分割模型，通常情况下包括逐点卷积、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都运用不同的卷积核。客观的说，深度卷积的实际分割效果并不良好，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]中提出了在深度卷积开始之前，通过升维来不断提高卷积维度，使其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，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能得以平衡。

Lin 等在研究中提出，综合运用 CRF、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。该种方法能获取相对丰富的数据信息，能提高运行效率，然而在结构预测过程中，仅能将图像输入到一元项或成对项，在中高项中却难以实现结构预测，由此形成的分割效果实际精度相对不高。对此，Arnab等人[32]提出可以将两种不同形式的高阶势能项(higher order potential, HOP)内嵌到 CNN 中，从而开展深度的训练，不断提升分割质量。此外，为了优化分割模型、提高分割质效，Vemulapalli等人[33]提出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 所涉及的方法都比较注重对扩张率不同的空洞卷积进行串行操作，以此不断增强感受野，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，占用大量的运行空间，消耗大量的内存。3.1.2 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更好地利用全局有效信息，提高分割进度。基于特征融合的方法，通过融合不同层次、不同区域特征来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。从而有效避免基于扩大感受野的分割方法和基于概率特征图的分割方法所导致的问题。

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。于 2016 年提出的 DeepLabv2[37]在 DeepLabv1 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。通过采样能够利用不同比例实现对上下文的捕捉，该过程中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。DeepLabv3[38-39]改进了 ASPP 结构，引入 Resnet block 模块并以提取显著性特征为目标，提出空洞卷积模式并实现模块与空间要素的池化效应。结合上述两种方法，包含 Yang 在内的研究团队进行深入研究，创新连接方法，提出了 DenseASPP[40]网络。在街景分类过程中，该网络运用了全新的方式对扩张卷积进行连接操作，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。不过通过 He 等科研团队研究发现，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。为此，该团队提出动态多尺度网络(Dynamic Multi-scale Network, DMNet) [41]，并提出可借助该网络实现对动态卷积语义的感知和估计。为了网络提高聚合上下文信息的能力，Zhao 等人提出金字塔场景解析网络(pyramid scene parsing network, PSPNet) [42]，随后 Zhao 等人又从压缩 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) [43]。He 等提出的自适应金字塔上下文网络(APCNet) [44]，综合运用多个自适应模块，对多层级的上下文表示进行合理构建。Wu等人[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。该种方法能有效获取具有高分辨率映射特征的样本，且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。在实践过程中，

为了获取长距离上下文信息，专家学者提出了扩张卷积等创新方法。然而，该类方法在获取信息过程中，并不能形成密集的信息。因此，在进行语义分割过程中，Zhao等人[46]合理引入了注意力机制，提出了 PSANet 网络模型，预先绘制注意力图[47]来聚合不同位置的信息。实践发现，该方法需要借助巨大的注意力图，来实现对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 3 种模块进行详细介绍：CCNet[48]算法模块能够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，就能实现对全局信息的整合操作，有效降低运行成本，提高计算速度；ACNet[50]模块综合运用力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。为此，多位学者纷纷将该机制进行合理运用，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]等创新方法纷纷涌现出来。在进行语义分割过程中，合理引入注意机制，借助该模块对有关信息进行学习，通过对注意力机制的调整和优化，形成全新的十字交叉模块和自注意模块，以此对全局信息进行获取，对各层级信息进行感受，使得捕获信息、获得内部特征变得更为容易。

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由一系列卷积－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，以此形式可有效解决分辨率不断下降的问题，可以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 7 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，又在 SegNet 网络基础上提出了 Bayesian SegNet 网络，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等人[56]基于 FCN 提出了一个完全对称的 DeconvNet 网络，该网络利用 FCN 与反卷积网络进行互补，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。

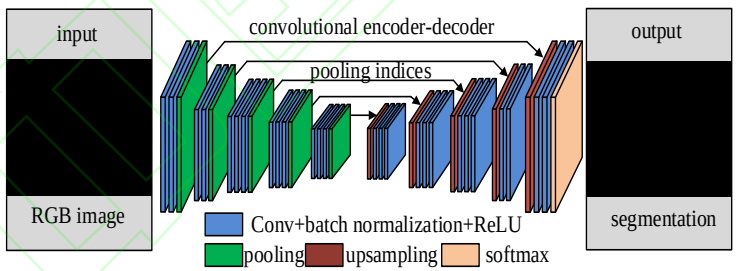

图 7 SegNet 网络结构示意图

Fig.7 Structural diagram of SegNet network

Olaf Ronneberger 等学者提出架构相对匀称的 U-net 网络，该网络的编码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，而现如今许多场合需要对 3D 场景进行分割处理。于是，Milletari 等人提出 V-Net 网络，该模型是一种将 3D 体积、全卷积与神经网络相结合的三维对称语义分割模型[57]，该模型解决了训练标注数据集不足的问题，与其它分割模型相比具有计算优势。语义分割过程中丢失数据信息的问题不间断发生，为此，Lin 等人于 2017 年提出了 RefineNet[58]网络。该网络采用的是链式残差连接的模式，对分割过程中缺失的信息能够进行追回和有效融合，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，该模型都能产生相对优越的分割效果。多年以来，专家学者们在针对如何利用编码解码器结构来改良分割方法、提高分割效率等方面进行了深入的研究和探索，并提出了一系列卓有成效的研究成果。影响较大的如：

AdamPaszke 等人运用提前采样的方法，持续不断减少对解码器的运用，以此形成简化版的 ENet 结构，实现减低网络冗余、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，多位学者从不同角度、不同方面入手做出了多项改进：（1）不断提高语义分割的实际速度，例如运用 ENet、LEDNet 等模型，有力推动了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，另有一种应用较广、效果良好的运算模型——即循环神经网络(Recurrent Neural Network, RNN) [60]模型。该种模型的主要优势特征在于，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。以此思想为先导，Visin等人[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，在处理图像过程中，灵活性、多样性相对较差。为此，Liang等人[63]提出了 Graph-LSTM 网络，该网络设置任意超像素为参考的有效节点，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 FCCRF 之间缺乏有效的交互联系,Zheng等人[64]提出了 CRFasRNN 网络模型，将 CRF 的有关学习、推理过程融入 RNN 的运算中。

RNN 能够保留有关信息，能够实现对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，并融入到神经网络结构中，以此形式对卷积层空间特征进行有效提取，也能实现对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(generative adversarial network, GAN)[65]在一定程度上也能实现对 CRF 的替代，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、图像之间的不一致性，Luc等人[66]于 2016 年首次引入了 GAN 技术，在进行语义分割过程中，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 网络不能很好的破解像素类别不一致、不均衡的现实问题，Xue等[67]研究创造了一种基于多尺度、多层次函数的对抗网络模型，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 模型还具有能够识别数据真假，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，要实现对小样本特征的持续有效学习，关键问题就是如果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，尤其针对大数据图像，该方法的解释性、可延伸性存在不足，仍有较大的提升空间。

3.2 弱监督语义分割方法

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训练有一定的局限性，因此弱监督语义分割开始涌现。本文根据不同类型的监督信息，将弱监督学习图像语义分割方法分为 6 类：基于边界框级标注方法、基于涂鸦级标注方法、基于点级标注方法、基于图像级标注方法、基于混合标注方法以及基于附加数据源方法。

3.2.1 基于边界框级标注方法

基于边界框标注方法将包括整个物体的矩形区域作为训练样本，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等人[68]借助 FCN 网络，合理运用候选区域的研究思维，提了出 BoxSup 网络模型。该模型中，预先设定以边界框标注的图像作为训练样本，选用 MCG 算法进行计算，可以形成原始候选区域，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；而后，进一步预测候选区域的有效范围，并对该区域重复进行优化升级，反复操作直至结果收敛到合理范围内即可。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，普遍使用的是简单迭代方式进行模型训练，由此形成的最终结果往往与实际标签之间有较大的差异。以 Song 为代表的研究团队，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 model, BCM) [70]模型对不相干区域进行删除操作，由此形成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，可以最大限度的减低错误标注形成的不良影响。

3.2.2 基于涂鸦级标注方法

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实际任务量。

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，选定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，而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，主要是针对上个步骤已形成的图像进行模型训练，最终形成合理的分割结果。

3.2.3 基于点级标注方法

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，同样预算前提下，点级监督的监督效果更佳，最终结果更为优越。

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

3.2.4 基于图像级标注方法

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。Pinheiro 等在研究过程中，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。Papandreou 等人使用了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。Durand 等人对相关特征图进行分解操作，形成初步的多通道特征。不同通道有各自不同的局部特征，经过池化操作后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。在实践操作过程中，可以借助对目标区域的扩展，对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。以 Kolesnikov[74]为代表的研究团队提了 SEC(Seed, Expand, and Constrain)算法。受到该算法的影响和启示,Huang 等人[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]研究人员将 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]研究人员有效运用了 AffinityNet 网络，并形成准确的分割标签，以此来弥补相关信息的缺失。在研究过程中，Zhou等人[78]对监督信息进行了图像级的标注，并运用响应峰值持续提高实例分割效率。对比来看，该分割方法运算过程简单，样本获取成本较低，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等人[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，以此提出了一种 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工作，得出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 基于混合标注方法

综上所述，以上方法在减低成本、减少时间方面都有显著优势，能大幅度减少训练数据的实际需求。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。

半监督学习的分割方法，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。Papandreou 等研究人员提出的随机梯度下降(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等研究人员[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的降低人工交互的操作方式。所以，当前的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，而且视频的传播在当前情况下更为普遍。Hong 等研究人员在进行搜索时，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屏资料。同时，合理运用分类器对相关视频区间进行优化处理，获取更优的检索结果。此外，该研究小组引入了注意力机制用于新型编解码结构中，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 无监督语义分割方法

大量研究数据说明：如果某个神经网络有大量的数据训练基础，则该网络往往拥有相对良好的运行属性。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。对此，有效的破解方式是采取相对密集的手动标注方法，并反复进行网络训练。另外的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值。

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。Tzeng[82]于 2015 年提出了 DC 方法，能够有效运用二元域分类器实现对标签的均匀布局。此后，经过深入研究，Tzeng等[83]又进一步提出了新的判别域适用（Adversarial Discriminative Domain Adaptation, ADDA）法，借助对抗训练模式来不断优化相关模型。为了解决跨域分割问题 Hoffman等[84]提出 FCNWild 方法。包括 Zhang[85]在内的研究小组提出了 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，对运用合成图像提升语义分割质量性能进行了深入的探索研究，形成了初步的研究成果。

图 8 将弱监督和无监督语义分割方法的分类进行了汇总。

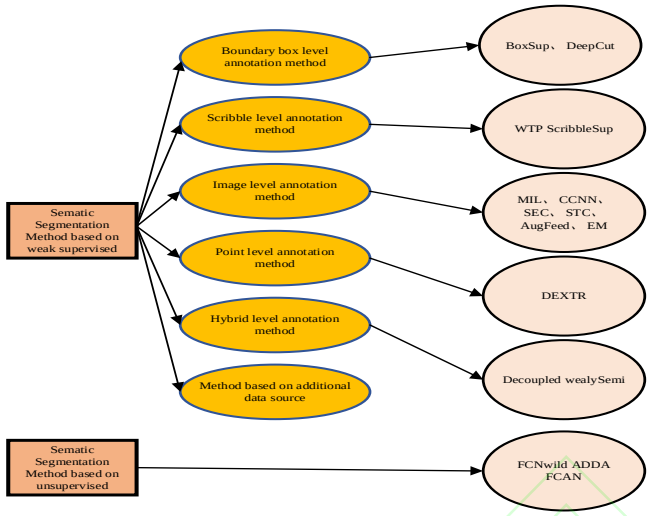

图 8 基于弱监督和无监督的语义分割方法

Fig.8 Semantic segmentation method based on weak supervision and unsupervised

4 城市道路场景数据集以及性能评价指标

本节中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

大量的研究人员专注于城市道路场景数据集，在实际场景下尝试使用多个传感器捕获多维度信息，并提供了大量的精细标注，构成大型城市道路数据集，这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对比如表 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

Data set name	Proposal time	Classes	Total amount of data	Areas	Environment
CamVid[86]	2009	32	700	Europe	Day
KITTI[87]	2013	10	/	Germany America	Day
Oxford Robotcar[88]	2014	/	2000，0000	Oxford, UK	All weather conditions
Cityscapes[89]	2016	34	20000	Germany Switzerland France	Climate (spring/summer/autumn)
SYNTHIA[90]	2016	11	13407	Rendering city	Rendering environment
Comma.ai	2016	/	/	America	/
Mapillary Vistas[91]	2017	66	25000	America Europe Africa Asia Oceania	Sunny, Rain, Snow, Fog, Dusk, Day, Night
Apollo Scape [92]	2018	28	143906	China	Day, Snow, Rain, Fog
BDD100K[93]	2018	10	10000	Multiple cities around the world	Various scenes
Udacity's Driving[94]	2018	/	/	/	/

NuScenes	2019	23	140，0000	Boston Singapore	Day
D²-City	2019	12	/	China	Complex weather
Waymo[95]	2019	/	3000 driving clips	America	Day, Night, Dawn, Dusk, Rainy Sunny days

/表示该方法未提供相应的结果。

表 3 常见的交通标志数据集

Table 3 Common traffic sign datasets

Data set name	Summary
KUL Belgium Traffic Sign[96]	A dataset of traffic signs in Belgium
German Traffic Sign[97]	German traffic Annotated dataset
STSD[98]	More than 20,000 images with 20% labels, containing 3488 traffic signs.
LISA[99]	7855 annotations on more than 6610 frames.
Tsinghua-Tencent 100K[100]	Data set for cooperation between Tencent and Tsinghua，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、使其发挥重大的作用，务必需要对分割框架的结构属性和实际效能进行合理的测评。在进行评估时，需要选用科学严谨、多样化的维度、公平合理的指标来进行比较。下文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要求，因此需要用运行时间去衡量分割方法的实时性的优劣。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

当前已经具备多种指标用来评价像素语义分割性能的优劣，通过像素准确率（PA）参数体现像素与类对应关系的准确度，对其求平均值则获取平均准确率（mPA）；同时包括交并比类指标，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(PA)的计算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比例，进而作为体现分割评价指标中正确率的依据。相关公式如（1）所示

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

公式中 p_{ii} p_{ij} 代表不同数量含义，其中前者对应准确分类的像素，后者则对应分类错误的像素，通过 j 和 i 体现不同的归属类。

像素与划分类的对应关系不一定是准确的，并通过平均准确率(MPA)指标反馈准确比例，定义如（2）所示

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

平均交并比(MIoU)是模型对每一类预测的结果和真实值的交集与并集的比值，求和再平均的结果，定

义如式(3)所示

$$mIoU = \frac{1}{n+1}\sum_{i=0}^{n}\frac{p_{ij}}{\sum_{j=0}^{n}p_{ij} + \sum_{j=0}^{n}p_{ij} - p_{ij}} \tag{3}$$

频率加权交并比(FWIoU)是对 mloU 改进后的新的评价指标，旨在对每个像素的类别按照其出现的频率进行加权，定义如式(4)所示。

$$FMIoU = \frac{1}{\sum_{i=0}^{n}\sum_{j=0}^{n}P_{ij}}\sum_{i=0}^{n}\frac{\sum_{i=0}^{n}P_{ij}P_{ii}}{\sum_{j=0}^{n}P_{ij} + \sum_{j=0}^{n}P_{ij} - P_{ii}} \tag{4}$$

其中，mIoU 指标的代表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

本文研究的核心领域自动驾驶领域更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考察了它们的实时性。

4.3.1 传统语义分割方法实验对比

在传统图像语义分割方法中，N-Cut，Grab Cut 等经典算得到了广泛的应用，但这些算法的效率较低。在此基础上 GPB-UCM，Random Decision Forest，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的质量以及算法时间复杂度上都有更好的表现，但是由于传统分割方法在完成语义分割任务上存在的限制，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

Table 4 Analysis and summary of traditional image semantic segmentation methods

Methods	Year	Contribution	PGM	data set	mIoU(%)
Normalized cut	2000	The graph is divided into k subgraphs and the cut of K subgraphs is minimized	/	/	/
Grab Cut	2004	Use image texture and boundary information, relying on a small amount of manual intervention to obtain better foreground and background segmentation	/	/	/
GPB-UCM	2011	Calculate the probability of each pixel as an edge, detect the target contour, generate a contour map, and complete the segmentation. Complex steps and high complexity	/	BSDS	/
Ramdom Decision Forest	2016	Combine multiple decision trees into a classifier	/	/	/
MCG	2017	On the basis of GPS-UCM, using the generated multiple contour segmentation blocks, combined with the random forest classifier to get prediction object	/	BSDS	/

/表示该方法未提供相应的结果。

4.3.2 基于强监督语义分割方法实验对比

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

Classification		Model name	Year	Key technology	PGM	Dataset	mloU（%）
		FCN	2014	Upsampling，Skip Layer	/	PASCALVOC 2012/Cityscapes	62.2/65.3
Method based on enlarging receptive field	Method based on dilated convolution	DeepLab v1	2014	Upsampling, structure prediction	CRF	PASCAL VOC 2012/Cityscapes	71.6/63.1
		ENet	2016	Decomposition filter, dilated convolution	/	Cityscapes/CamVid	58.3/51.3
		DRN	2017	Dilated convolution	/	/	/
	Method based on optimizing convolution structure	DUC+HDC	2017	DUC+HDC	/	PASCAL VOC 2012/Cityscapes	83.1/80.1
		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; 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 Improved ASPP	CRF	PASCAL VOC 2012	86.9
		DeepLab v3+	2018	ASPP module with separable convolution，Skip join fusion of different level features	/	PASCAL VOC 2012/Cityscapes	89.0/82.1
		ICNet	2017	Cascade model, feature fusion	/	Cityscapes/CamVid	70.6/67.1
		DenseASPP	2018	ASPP, Densely connected networks to improve receptive field	/	Cityscapes	80.6
		DMNet	2019	Dynamic convolution module Context-Aware Correlation Filter	/	PASCAL VOC 2012	84.4
		APCNet	2019	GLA，ACM	/	PASCAL VOC 2012	84.2
		PSANet	2018	Attention mechanism	/	PASCAL VOC 2012/Cityscapes	85.7/80.1
		CCNet	2018	Dilated convolution，Feature weighted fusion	/	Cityscapes	81.4

Method based on Attention Mechanism	BiseNet	2018	Spatial path，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，Feature weighted fusion	/	PASCAL VOC 2012/Cityscapes	82.6/81.5
Method based On Encoding and Decoding	SegNet	2015	Deconvolution, upsampling, dropout layer	/	CamVid	55.6
	DeconvNet	2015	Deconvolution, Unpooling	/	PASCAL VOC 2012	69.6
	RefineNet	2017	Bilinear Interpolation Skip join and Residual join	/	Cityscapes	73.6
	GCN+	2017	Large kernel convolution, 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	Capture 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	Extend the function of ReNet	/	CamVid	/
Method based on GAN	/	2016	GAN Adversarial Training	/	PASCAL VOC 2012	54.3
	/	2016	GAN Domain Adaptation	/	Cityscapes	67.8

/表示该方法未提供相应的结果。

从上表可以看出，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的测试。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%，分割精度基本满足对街道场景图像语义分割的精度要求，然而这些算法实时性上有所欠缺。ENet、ESPNet、ICNet 和 BiSeNet 这四种算法虽然分割准确率不如上述算法，但由于尺寸小，计算成本轻等特点，这些算法具有实时性强的优势。

针对算法参数数量和运行速率，本文从强监督学习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

Table 6 Speed analysis of algorithm

Model	Parameters	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 这 3 种分割算法速度较快，实时性强，适用于实时图像语义分割。其中，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，以在分类能力和感受野之间取得平衡，是目前在分割效率和准确性之间达到均衡最突出的算法之一。而 FCN、和基于 FCN 的 DeepLab v1、DeepLab v2 运行时间较长,无法满足实时图像分割的需求。而在 DeepLab 系列中，DeepLab v3+分割效果最好，主要是其吸取 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，从而实现图像语义分割精度和速度的均衡。其他很多算法的分割速度都比 FCN 要低,也同样无法满足实时图像分割的需求,不适用于动态场景分割的任务。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 基于弱监督语义分割方法实验对比

表 7 列举了在最具代表性的数据集上基于深度学习的弱监督学习图像语义分割方法实验结果对比。主要比较的因素有监督信息、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

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

Supervision information	Model name	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, 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

/表示该方法未提供相应的结果。

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。而边界框标签的形式虽然比较复杂，但是能够提供目标位范围的监督信息，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，虽然目前基于弱监督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，但是由于包含的有用信息过少，在分割效果和分割性能上与强监督语义分割算法差距较大，很多算法还不太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方向。

4.3.4 基于无监督语义分割方法实验对比

表 8 列举了在最具代表性的数据集上基于深度学习的弱监督学习图像语义分割方法实验结果对比。主要比较的因素有关键技术、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

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

Model name	Year	Key technology	Dataset	mIoU (%)
FCNWild	2016	Domain adaptive fully 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

从概念上讲，无监督分割主要是使用虚拟场景，对现实场景进行数据标注，完成语义分割，以此形式降低标注成本，简化分割过程。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，当前情况下无监督分割法的有效精度并不高，如何进一步提升分割精度、提高分割质量也将会逐渐成为未来课题研究的重点和热点。

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。减小标注成本的弱监督和无监督方法目前分割效果不明显，分割边界粗糙且不连续，如何提高其分割精度将成为今后研究的热门方向。

5 结束语

近年来，随着自动驾驶等应用不断发展，对模型尺寸、计算成本、分割精度等方面提出了更高的要求。本文首先介绍了基于道路场景语义分割的发展现状与挑战，说明研究语义分割技术是一项意义重大，十分重要的任务。以深度学习为基本特征，划分语义分割技术为传统模式，传统与深度学习相结合的模式以及基于深度学习的模式，重点致力于深度学习这一正在崛起的研究领域，将其进一步细分为强监督学习图像语义分割方法，弱监督学习图像语义分割方法和无监督学习图像语义分割方法。对每类方法的代表性算法结合本文研究的道路场景进行了研究、分析和对比,并概括总结了每类方法的技术特点和优缺点，分析哪一种算法更适合于道路场景的分割，使其更好地应用到无人驾驶领域。总体来看，利用深度学习来对道路场景进行语义分割技术在不断进步，但是也有一些需要改进和继续研究的方向。

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，例如在行驶的过程中周围天气的变化，交通指示灯的变化，以及对来往车辆和行人的判断，这就要求对输入的分割对象要有很精确的分割。

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的网络模型的重点指标，但是随着无人驾驶技术的不断成熟，分割效率对其产生的影响越来越大，这就需要在维持高精确率的基础上尽量缩短响应时间。

3）弱监督或无监督语义分割技术。弱监督和无监督的语义分割技术目前的分割效果还不明显，如何利用尽量少的标注信息来提高网络模型的精度会是未来发展的趋势。

4）三维数据的应用[103-104]。三维数据对真实场景至关重要，现如今更多的语义分割方法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一个研究热点。

参考文献

[1] Zhou Jimiao, Li Bijun, Chen Shizeng. A real-time road scene segmentation method based on multi-layer 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] Liuyuan Deng, Ming Yang et al, 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 Xuan, Wang Liang, Ding Qi. 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(02):440-468.

[5] Jing Zhuangwei, Guan Haiyan, Peng Daifeng, et al. Review of image semantic segmentation based on deep neural network [J]. Computer Engineering,2020:1-30.
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,2020:1-30.

[6] Wang Yu, Zhang Huanjun, Huang Haixin. Review of image semantic segmentation algorithm based on deep learning [J]. Electronic technology application, 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(06):23-27+36.

[7] Zhang Xiangfu, Liu Jian, Shi Zhangsong, et al. Review of semantic segmentation based on deep learning [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 Huilan, Zhang Yun. Review 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 Yanran, Chen Qingliang, Wu Junjun. Review of image semantic segmentation methods for complex environment [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 Huiyu, Wu Junjun. Review 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[J]. arXiv preprint arXiv:2001.05566, 2020.

[12] Zhang Jiaying, Zhao Xiaoli, Chen Zheng. Review of point cloud semantic segmentation based on deep learning [J]. Laser & Optoelectronics Progress, 2020,57 (4): 28-46.
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):28-46.

[13] Tian Qichuan, Meng Ying. Image semantic segmentation based on convolutional neural network [J]. Journal of Chinese Mini-Micro 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 Yupeng, Zhao Weidong, Wang Zhicheng, et al. Research on image-based image segmentation[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] Qiuhua Zheng,Wenqing Li, et al. An Interactive Image Segmentation Algorithm Based on Graph Cut[J]. Procedia Engineering,2012,29.
[17] Han Xu. Research on Grabcut based Automatic Image segmentation algorithm [D]. Beijing: Beijing Printing Institute, 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 Lei, Shi Zhiguo, Su Haoru, et al. Image segmentation based on high order Markov random fields [J]. 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] Pablo Arbeláez, Maire M , Fowlkes C , et al. Contour Detection and Hierarchical Image Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2011, 33(5):898-916.
[20] Zhang C , Xue Z , Zhu X , et al. Boosted random contextual semantic space based representation for visual recognition[J]. Information ences, 2016, 369:160-170.
[21] Pont-Tuset Jordi, Arbelaez Pablo, 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 Abdel Hamid, Helaly Hany Ahmed. Comparison Between Digital and Manual Marking for Toric Intraocular Lenses: A Randomized Trial.[J]. Medicine,2015,94(38).
[23] Wang Chunyao, Chen Junzhou, Li Wei. A review of research on super pixel segmentation algorithm [J]. Computer application research, 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] Evan Shelhamer, Jonathan Long, Trevor Darrell. Fully Convolutional Networks for Semantic Segmentation[M]. IEEE Computer Society, 2017，39（4）：640-651.
[25] Yu Fisher，Koltun Vladlen．Multi-scale context aggregation by dilated convolutions[J]. arXiv preprint arXiv: 1511．07122，2015.
[26] Chen L C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs[J].International Conference on Learning Representations, 2014(4):357-361.
[27] Paszke Adam，Chaurasia Abhishek，et al. ENet: a deep neural network architecture for real-time semantic segmentation［J］.arXiv preprint arXiv: 1606．02147.
[28] Yu Fisher，Koltun Vladlen，Funkhouser Thomas．Dilated residual networks［C］//IEEE Conference on Computer Vision and Pattern Recognition(CVPR) ，2017: 636-644.
[29] Fang, Y, Li, Y, et al, Face completion with Hybrid Dilated Convolution. Signal Processing-Image Communication, 80,2020:115664.
[30] Dai J, Qi H, Xiong Y, Li Y, Zhang G, Hu H, Wei Y. Deformable convolutional networks. In: Proc. of the IEEE Int’l Conf. on Computer Vision. 2017. 764−773.
[31] Ghiasi G, Fowlkes CC. Laplacian reconstruction and refinement for semantic segmentation. ArXiv preprint arXiv:1605.02264, 2016.
[32] Arnab A, Jayasumana S, Zheng S, et al. Higher order conditional random fields in deep neural networks[C]//European Conference on Computer Vision. Springer, Cham, 2016: 524-540.
[33] Vemulapalli R, Tuzel O, Liu M Y, et al. Gaussian conditional random field network for semantic

segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 3224-3233.

[34] Shen F, Gan R, Yan S, et al. Semantic segmentation via structured patch prediction, context CRF and guidance CRF[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition （CVPR）.2017: 5178-5186.

[35] Jiang J, Zhang Z, Huang Y, 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). IEEE, 2017: 525-530.

[36] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.

[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 & Machine Intelligence, 2016, 40(4):834-848.

[38] Wang P, Chen P, Yuan Y. Understanding convolution for semantic segmentation[C]//2018 IEE winter conference on applications of computer vision (WACV). IEEE, 2018: 1451-1460.

[39] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2881-2890.

[40] Yang M, Yu K, Zhang C, et al. DenseASPP for semantic segmentation in street scenes[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018: 3684-3692.

[41] He J, Deng Z, Qiao Y. Dynamic Multi-scale Filters for Semantic Segmentation[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision.2019: 3561-3571.

[42] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2881-2890.

[43] Zhao H, Qi X, Shen X, et al. ICNet for real-time semantic segmentation on high-resolution images[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 405-420.

[44] He J, Deng Z, Zhou L, et al. Adaptive pyramid context network for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition（CVPR）. 2019: 7519-7528.

[45] Wu H, Zhang J, Huang K, et al. FastFCN: Rethinking dilated convolution in the backbone for semantic segmentation[J]. arXiv preprint arXiv:1903.11816, 2019.

[46] Zhao H, Zhang Y, Liu S, et al. PSANet: Point-wise spatial attention network for scene parsing[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018:267-283.

[47] Yuan Jiajie, Zhang Ling, Chen Yunhua. Deep neural network image recognition based on attention convolution module [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] Shouting Feng,Zhongshuo Zhuo,Daru Pan,Qi Tian. CCNet: A cross-connected convolutional network for segmenting retinal vessels using multi-scale features[J]. Neurocomputing, 2020,392.

[49] Yu C, Wang J, Peng C, et al. Bisenet: Bilateral segmentation network for real-time semantic segmentation[C]//Proceedings of the European Conference on Computer Vision (ECCV).2018: 325-341.

[50] Luo C , Xin W , Li X , et al. ACNet: Attention-based Convolution Network with Additional Discriminative Features for DCM Classification (S)[C]// The 31st International Conference on Software Engineering and Knowledge Engineering. 2019,155.

[51] Xue H , Liu C , Wan F , et al. DANet: Divergent Activation for Weakly Supervised Object Localization[C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2019:6589-6598.

[52] A. Krizhevsky, I. Sutskever, G.E. Hinton. ImageNet classification with deep convolutional neural networks[C]. Advances in neural information processing systems, 2012, 1097-1105.

[53] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C] //International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.

[54] Wu Zongsheng, Fu Weiping, HAN Gaining. Understanding of Road Scenarios based on deep convolutional neural networks [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 Yunyang, Qu Xuexin, Zhu Quanyin, et al. Measurement method of confidence 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]//Proceedings of the IEEE international conference on computer vision. 2015: 1520-1528.

[57] Li Qingbo, Su Dan. Multiple organ image segmentation of abdomen based on V-Net[J]. Digital Technology and Application,2019，37（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,2019,37(1):89+91.

[58] Lin, G，Milan, A， et al．RefineNet: Multi-path Refinement Networks for High-Resolution Semantic Segmentation[C]/ / IEEE Conference on Computer Vision and Pattern Recognition(CPVR)，2017: 5168-5177.

[59] Tian, Zhi, et al. “Decoders Matter for Semantic Segmentation: Data-Dependent Decoding Enables Flexible Feature Aggregation.” 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019：3126–3135.

[60] Yang Li, Wu Yuxi, Wang Junli, Liu Yili. Review of cyclic neural networks [J]. Computer applications, 2018,38 (S2): 1-6 + 26.
杨丽,吴雨茜,王俊丽,刘义理.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, Ciccone M, Romero A, et al. Reseg: A recurrent neural network-based model for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops（CVPRW）. 2016: 426-433.

[62] Li Z, Gan Y, Liang X, et al. Lstm-cf: Unifying context modeling and fusion with ISTMs for RGB-d scene labeling[C]//European conference on computer vision. Springer, Cham, 2016: 541-557.

[63] Liang，Xiaodan，et al. Semantic object parsing with graph LSTM[C]//European Conference on Computer Vision,2016: 125-143.

[64] Zheng S, Jayasumana S, Romera-Paredes B, et al. Conditional random fields as recurrent neural networks[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1529-1537.

[65] Wang Kunfeng, Gou Chao, Duan Yanjie, et al. Research progress and Prospect of generative countermeasure network Gan [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[J]. arXiv preprint arXiv:1611.08408, 2016.

[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, He K, Sun J. BoxSup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. IEEE，International Conference on Computer Vision（ICCV），2015：1635-1643.

[69] Rajchl Martin, Lee Matthew C H, Oktay Ozan, et al. DeepCut: Object Segmentation From Bounding Box Annotations Using Convolutional Neural Networks[J]. IEEE transactions on medical imaging,2017,36(2)：10-11.

[70] Song C, Huang Y, Ouyang W, et al. Box-driven class-wise region masking and filling rate guided loss for weakly supervised semantic segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition（CVPR）, 2019: 3136-3145.

[71] Bearman A, Russakovsky O, Ferrari V, et al. What's the point: Semantic segmentation with point supervision. In: Proc. of the European Conf. on Computer Vision. Springer-Verlag, 2016. 549−565.

[72] Lin D, Dai J, Jia J, et al. ScribbleSup: Scribble-supervised convolutional networks for semantic segmentation. In 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016. 3159−3167.

[73] Maninis K K, Caelles S, Pont-Tuset J, et al. Deep extreme cut: From extreme points to object segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 616-625.

[74] Kolesnikov A, Lampert C H. Seed, expand and constrain: Three principles for weakly-supervised image segmentation [C]//European Conference on Computer Vision. Springer, Cham, 2016: 695-711.

[75] Huang Z, Wang X, Wang J, et al. Weakly-supervised semantic segmentation network with deep seeded region growing[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7014-7023.

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

[77] Ahn J, Kwak S. Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 4981-4990.

[78] Zhou Y, Zhu Y, Ye Q, et al. Weakly supervised instance segmentation using class peak response[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 3791-3800.

[79] Wei Y, Liang X, Chen Y, et al. Stc: A simple to complex framework for weakly-supervised semantic segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 39(11): 2314-2320.

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

[81] Hong S, Noh H, Han B. Decoupled deep neural network for semi-supervised semantic segmentation[C]//Advances in neural information processing systems. 2015: 1495-1503.

[82] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. In Proc.ICML, 2014. 1, 2, 5.

[83] Tzeng E, Hoffman J,Saenko K, et al. Adversarial discriminative domain adaptation[C]//IEEE Conference on Computer Vision and Pattern Recognition. Hawaii: IEEE, 2017:4.

[84] Hoffman J, Wang D, Yu F, et al. FCNs in the wild: Pixel level adversarial and constraint-based adaptation[J]. ArXiv preprint arXiv: 1612. 02649,2016.

[85] Zhang Y. Qiu Z, Yao T, et al, Fully Convolutional Adaptation Networks for Semantic Segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 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, Urtasun R. Vision meets robotics: The KITTI dataset. The Int'l Journal of Robotics Research, 2013,32(11):1231-1237.

[88] Will Maddern, Geoffrey Pascoe, Chris Linegar, et al. 1 year, 1000 km: The Oxford RobotCar dataset. 2017,

36(1):3-15.
[89] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 3213-3223.
[90] German Ros, Laura Sellart, Joanna Materzynska, et al. The synthia. dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In IEEE Conference on Computer Vision and Pattern Recognition, pages 3234-3243, 2016.
[91] Neuhold G , Ollmann T , Bulo S R , et al. The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes[C]// IEEE International Conference on Computer Vision. IEEE Computer Society, 2017.27(3):22-29.
[92] Wang P, Huang X, Cheng X, et al. The Apollo Scape open dataset for autonomous driving and its application[J]. IEEE transactions on pattern analysis and machine intelligence, 2019.
[93] Yu F , Chen H , Wang X , et al. BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning[J]. 2018：1805.04687
[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. IEEE, 2018.8460913.
[95] Gu Z , Li Z , Di X , et al. An LSTM-Based Autonomous Driving Model Using Waymo Open Dataset[J]. ArXiv preprint arXiv: 1912.04838,2020.
[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]// International Joint Conference on Neural Networks. IEEE, 2013:1-8.
[98] Zhu Y , Zhang C , Zhou D , et al. Traffic sign detection and recognition using fully convolutional network guided proposals[J]. Neurocomputing, 2016, 214(nov.19):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(JUL.):24-36.
[100] Shijin Song, Zhiqiang Que, et al. An efficient convolutional neural network for small traffic sign detection[J]. Journal of Systems Architecture,2019,97.
[101] Lu Wenchao, Pang Yanwei, He Yuqing, et al. Precise real-time semantic segmentation based on separable residual module [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 Yu, Huang Xuegong, Zhang Zhian, et al. Real time semantic segmentation algorithm based on feature fusion [J]. Progress in laser and optoelectronics, 2020,57 (2): 021001.
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(02):021001.
[103] Yang Jun, Dang Jisheng. 3D point cloud recognition and segmentation using deep cascade convolution neural network [J]. Optical 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 Aiwu, Liu Lulu, Zhang Xizhen. Multi feature convolution neural network semantic segmentation method for road 3D point cloud [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.

网络首发：

标题：道路场景语义分割综述
作者：王龙飞,严春满
收稿日期：2020-07-09
录用日期：2020-09-27
DOI：10.3788/lop58.120003

引用格式：
王龙飞,严春满. 道路场景语义分割综述[J].激光与光电子学进展,2021,58(12):120003.
