

## 基于深度学习的图像语义分割研究综述

### 摘 要

本文第一部分简要介绍对于本门课程学习过程的心得体会与建议，心得体会按照章节依次列出，结合知识点谈了自己的困惑、解决过程以及学习心得，建议主要包括对于智慧树问答环节的改进以及教学内容可以删去已学过的、增添深度学习等热门内容。第二部分对基于深度神经网络的图像语义分割方法、研究现状和前景展望进行了简要的综述，提出了自己的一些看法和感想。根据样本标注类型和网络训练方式的不同，将现有的方法分为全监督学习图像语义分割方法和弱监督学习图像语义分割方法，对每类方法中代表性算法的效果以及优缺点进行了分析介绍。接着，针对图像语义分割领域从三个方面提出了自己的看法与感想。最后，预测并分析了图像语义分割技术面临的挑战以及未来的发展趋势。

**关键词：**心得与建议，深度学习，图像语义分割，全监督学习，弱监督学习

装

订

线

# Review of Image Semantic Segmentation Based on Deep Learning

## ABSTRACT

The first part of this article briefly introduces the experience and suggestions for the learning process of this course. The experience is listed in order according to the chapters. Combined with the knowledge points, I talk about my confusion, solving process and learning experience. The suggestions mainly include the improvement of the wisdom tree's Q&A session and for the teaching content, the learned part can be deleted and deep learning part can be added. The second part briefly summarizes the image semantic segmentation methods, research status and prospects based on deep neural networks, and puts forward some of my own views and feelings. According to the different types of sample annotations and network training methods, the existing methods are divided into full supervision learning-based methods and weak supervision learning-based methods. The paper systematically details the following aspects of the presented algorithms in semantic segmentation – theories, advantages and disadvantages, contributions. Moreover, I put forward my own views and impressions on the field of image semantic segmentation from three aspects. Finally, the summaries and future trends are given.

**Key words:** experience and suggestions, deep learning, image semantic segmentation, full supervision learning , week supervision learning

## 目 录

1	心得体会与建议.....	1
1.1	心得体会.....	1
1.2	建议.....	2
1.2.1	智慧树问答环节的改进.....	2
1.2.2	教学内容的改进.....	2
2	基于深度学习的图像语义分割研究综述.....	3
2.1	引言.....	3
2.2	全监督学习图像语义分割方法.....	3
2.2.1	基于 FCN 的方法.....	3
2.2.2	基于编解码器的方法.....	4
2.2.3	基于注意力机制的方法.....	5
2.2.4	基于 GAN 的方法.....	6
2.3	弱监督学习图像语义分割方法.....	6
2.3.1	基于边框级标注方法.....	6
2.3.2	基于涂鸦级标注方法.....	7
2.3.3	基于图像级标注方法.....	7
2.4	我的看法和感想.....	7
2.5	总结与展望.....	8
	参考文献.....	9

## 1 心得体会与建议

受疫情影响，这学期我们主要通过智慧树网课平台来进行学习，线上学习和线下学习确实有很大不同，比如智慧树上特有的学习习惯评分标准，即每天学习时长超过 25 分钟做一次规律学习记录，帮助我改掉了以前学习时间不均衡的坏习惯，一路走来，我有很多心得体会，也有一些建议。下面，我会根据章节简要说一下我的心得和建议。

### 1.1 心得体会

第一章主要是概论的介绍，人工智能主要有三大学派，包括符号主义、联结主义和行为主义。之前我一直以为人工智能就是深度学习，经过学习才发现原来它只属于其中的联结主义学派，我认为每一个学派都有自己的优势。例如，联结主义做感知非常有效，如做视觉语音识别和分类等效果显著，但做推理效果则不尽人意，符号主义则特别适合做推理。未来我们想要达到强人工智能，需要感知、认知、推理等功能，更需要三大学派的互相融合。

第二章主要讲解智能体的相关概念，智能体是通过传感器感知所处环境并通过执行器对该环境产生作用的事物。人们通常希望设计一种理性的智能体，我觉得只是单单理性还是不够的，还需要一定的不确定性，这是它成为强人工智能的必备条件，如果智能体仅仅是根据当前环境以及之前的知识储备机械式的思考，然后机械式的做出行动，这将是枯燥的，世界有趣的地方在于变化，设计智能体的困难和乐趣也在于应对这种变化，因此我认为一个好的智能体应该是具有不确定性的，换句话说，是不能让人一眼看穿的智能体。

第三章主要讲解用于驱动理性智能体的搜索算法，我们小组实践课设计的智能军棋也是使用的  $\alpha$ - $\beta$  剪枝搜索算法，所以我对于这一章思考较多。开始时，我不理解无信息搜索中深度优先搜索的非完备性和非最优性，认为深度优先和宽度优先得到的最优解是一样的，因此它们都具有完备性和最优性。后来发现它不能保证在有解的情况下一定找到解，可能会出现陷入无穷分支导致死循环或者解不在最深结点的情况，至于我之前解的问题都是一些特殊问题，没有遇到非完备性和非最优性的情况，这也启示我学习的时候要注意掌握一般性原理，而非目前积累的那些微不足道的经验。

刚开始我们小组使用的是蒙特卡洛搜索的优化版，UCT（信心上限树）搜索，起因是清华的同学做过 UCT 搜索的智能四子棋，据说可以轻松打败  $\alpha$ - $\beta$  剪枝搜索。于是我们信心满满的使用 UCT 搜索做了一下智能体，结果发现军棋几乎是随机乱走的，而且会轻易被搜索层数很少的  $\alpha$ - $\beta$  剪枝搜索打败。后来查阅资料<sup>[1]</sup>发现，UCT 搜索算法在分支数较少，层数较少的搜索空间有很好的效果，而在军棋、围棋等分支数较多，层数较多的搜索空间很难收敛到最优解，需要大量的模拟操作，这也是 UCT 搜索在四子棋表现优秀而在军棋表现糟糕的原因。由此也启示我选择算法时要注意具体问题具体分析，不可人云亦云。

第四章和第五章主要讲解逻辑智能体以及必备的数学知识，逻辑智能体在知识库中应用推理来得到新的信息并作出决策，如果想要设计这样的智能体的话需要对推理有一定的认识。其中，我感触最深的是归结推理和反向链接，我认为它们都利用了一种正难则反的思想，归结推理是判

断结论能否由条件推出的一种方法，即先把结论取非，然后与条件合取，判断真假，相较于枚举推理，大大降低了时间复杂度；反向链接是由目标驱动，即由结论推测条件，相较于数据驱动的前向链接，也同样大大降低了复杂度。

第六章和第七章主要讲解智能体为应对不确定或部分可观察的环境所需要的概率推理等数学知识。初学时我对于贝叶斯网络中的条件独立关系不是很理解，教学视频里提及的也很少，后来我查阅资料，终于理解了给定父结点，一个结点条件独立于它的其他祖父结点和给定一个结点的父结点、子结点以及子结点的父结点，这个结点独立于网络中的所有其他结点<sup>[2]</sup>的含义。贝叶斯网络设计的非常巧妙，极大简化了完全联合分布所需的概率数目，这也启示我设计一个巧妙的数据结构对于简化算法有着极大的帮助。

第八章主要介绍了学习概论、决策树以及简单的人工神经网络。因为之前已经接触过一些有关图像分类和分割方面的神经网络知识，所以理解起来比较轻松。我最感兴趣的是一些工具函数的定义，比如决策树中信息熵的定义，仅仅利用了  $\log$  函数，就形象而贴切的描述出数据的混乱程度，简洁而有效，还有神经网络中的激活函数  $\text{relu}$  和  $\text{sigmoid}$ ，前者是仅仅将负值置为 0，后者仅仅用了  $e^x$  函数，就都增加了神经网络的非线性，提高了网络的泛化能力。

## 1.2 建议

总体上来说，讲授这门课的老师都很负责，讲解也很细致，思路都很清晰明了，但是还是有一些希望老师们可以采纳的建议。

### 1.2.1 智慧树问答环节的改进

智慧树里面提问的环节希望老师可以改进。这个提问功能应该是类似于一个论坛，谁有什么疑问都可以提出来，然后会的同学可以进行解答，以达到补充课堂教学的目的。但是到了后面就变味了，由于提问是要计分的，所以到了后期，同学们就开始疯狂的刷问题，有很多是重复之前别人发过的，也有很多是毫无意义的。

### 1.2.2 教学内容的改进

有很多重复的知识点，比如逻辑推导已经在离散数学课程中学习过，概率推导也在概率论课程中学习过，所以就没必要在人工智能这门课从头再学，感觉可以删掉这些，然后加入一些目前大家都很想了解的深度学习有关的内容，这样的话应该会更受同学们的欢迎。

## 2 基于深度学习的图像语义分割研究综述

### 2.1 引言

近年来，人机交互、自动驾驶、医学影像、虚拟现实等人工智能领域发展迅速，语义分割作为这些领域的研究重点，具有极高的研究价值和广泛的应用背景。图像语义分割<sup>[3]</sup>是将图片作为输入，通过一系列技术操作将其分割解析为与语义类别相关的不同图像区域，包括背景（草地、天空等）和离散对象（人、汽车等）。由于分割对象常常受到遮挡、割裂，加之背景光线等环境因素的影响，图像语义分割有很大的难度。

传统的图像分割方法<sup>[4]</sup>根据图像的颜色、空间结构和纹理信息等特征进行处理分析，有基于灰度阈值的图像分割技术、基于边缘特征的分割技术、基于局部区域的分割技术和基于聚类分析、模糊集理论、图论等特定理论的图像分割技术。但这些传统方法大都需要针对不同问题调节参数，鲁棒性较差，精度也较低，难以达到实际应用的需要。

近来，深度学习技术广泛应用于计算机视觉等领域，为图像语义分割技术带来了全新的机遇。基于深度学习的图像分割技术的思路<sup>[5]</sup>是：直接向深层神经网络输入大量原始图像数据，利用训练好的深度学习网络，对图像数据进行复杂处理，得到高层次的抽象特征，最后直接输出带有像素类别标签的与输入图像具有相同分辨率的分割图像。基于深度学习方法的图像分割技术很好地解决了传统算法的问题，成为近年来图像分割领域的研究热点。

本文对前人提出的基于深度学习的语义分割方法进行简要的总结和分析，根据<sup>[6]</sup>标注类型和学习方式的不同，将之分为全监督学习图像语义分割方法和弱监督学习图像语义分割方法。前者使用经过专家精确加工的像素级别标注作为训练样本，后者则使用弱标注数据作为样本进行训练。

### 2.2 全监督学习图像语义分割方法

#### 2.2.1 基于 FCN 的方法

这一类的代表网络有 FCN、Deeplab、Deeplab-V2、Deeplab-V3 等。

Long 等人<sup>[7]</sup>于 2014 年设计了一种兼容任意尺寸图像、以全监督学习方式进行图像语义分割的全卷积神经网络（FCN）。FCN 将 CNN 模型中的全连接层替换为全卷积层，使用跨层（skip

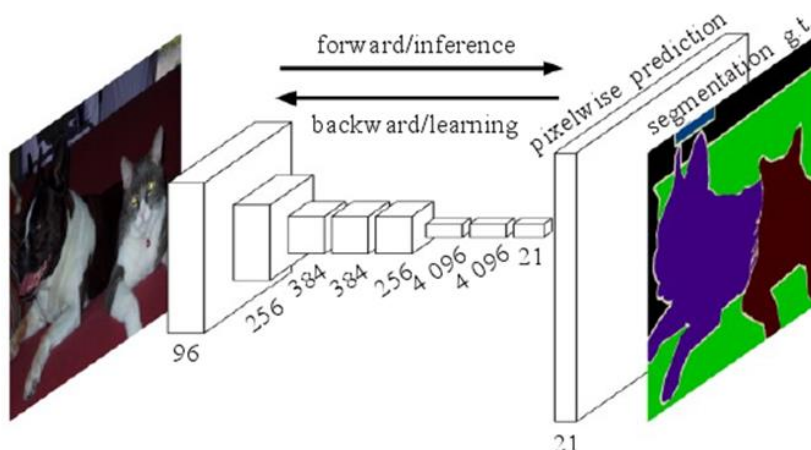


图 2.1 FCN 架构

layer) 方法组合中间卷积层产生的特征图, 再通过双线性插值 (BI) 算法进行上采样, 成功地将图像分类网络转变为图像分割网络, 将粗糙的分割结果转换为精细的分割结果。

针对 FCN 经过池化操作, 特征图的分辨率不断降低, 部分像素的空间位置信息丢失和未能有效考虑图像上下文信息, 缺乏空间一致性导致分割结果不够精细的问题, Chen 等人<sup>[8]</sup>提出了 Deeplab 网络, 该网络结合深度卷积神经网络 (DCNN) 和全连接条件随机场 (FCCRF) 优化粗分割图像, 提高深度网络定位精确性。

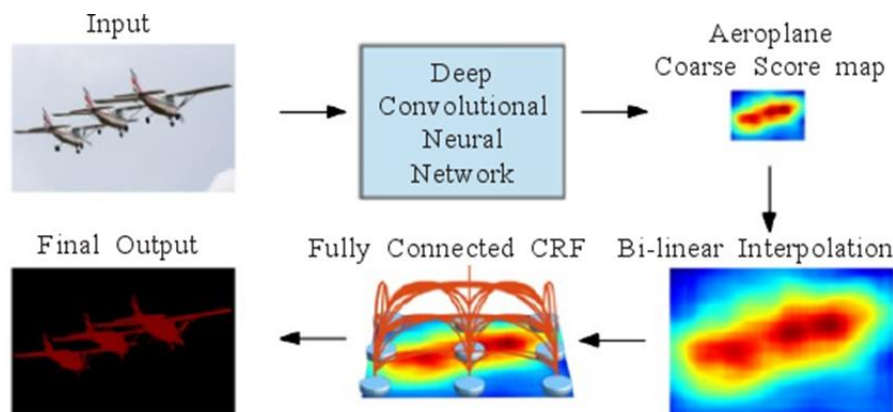


图 2.2 Deeplab 架构

Chen 等人<sup>[9]</sup>在此基础上提出了 Deeplab-V2 网络, 该网络采用空洞卷积在不增加参数量的情况下有效扩大滤波感受野来混合更丰富的上下文信息, 采用空洞空间金字塔 (ASSP) 池化方法捕获多尺度的对象和图像上下文特征。

随后, Chen 等人<sup>[10]</sup>又提出了 Deeplab-V3 网络, 该网络放弃了条件随机场 (CRF) 操作, 改进了 ASSP 模块, 加入了批规范化 (BN) 操作, 为了防止空洞卷积感受野的扩张率过大导致的“权值退化”现象, 增加了全局平均池化结构, 获得了比 Deeplab-V2 网络更加良好的效果。

## 2.2.2 基于编解码器的方法

这一类的代表网络有 U-net、SegNet、Bayesian SegNet、DeconvNet 等。

Ronneberger 等人<sup>[11]</sup>提出了一种编码-解码结构的网络模型 U-net, 该网络由收缩路径和扩张路径组成, 收缩路径通过下采样捕捉图片中的上下文信息, 逐层提取影像特征, 是一个编码器, 扩张路径通过上采样还原影像的位置信息, 逐步恢复物体细节和图像分辨率。该网络解决了图像语义分割大量样本和计算资源的需求, 适用于较少样本的影响分割。

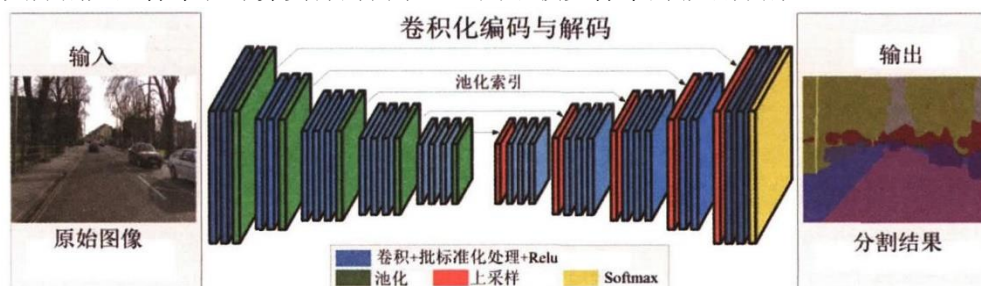


图 2.3 SegNet 架构

Ronneberger 等人<sup>[12]</sup>又针对自动驾驶汽车和智能机器人问题提出了 SegNet 网络, 该网络基



于先验概率计算每一个像素点的分类，编码器由全卷积网络构成，解码器由反卷积网络构成，利用转置卷积和上池化操作进行上采样处理。

针对先验概率无法给出分类结果置信度的问题，文献<sup>[13]</sup>基于 SegNet 网络提出了 Bayesian SegNet 网络。该网络在每个卷积层后增加了一个 DropOut 层，有效防止了过拟合的发生，同时，还引入贝叶斯网络和高斯过程，基于后验概率计算像素类别，使得网络能更合理地模拟事件概率。

Noh 等人<sup>[14]</sup>在 FCN 的基础上提出一个完全对称的 DeconvNet 网络，该网络利用反卷积替换 BI 算法，建立了一种完全对应机制，使用反卷积网络提取精细边界，使用 FCN 提取总体形状，更好地反映了物体的细节，提高了分割效果。

### 2.2.3 基于注意力机制的方法

这一类的代表网络有 PSANet、CCNet、BiSeNet、HMANet、DANet 等。

视觉注意力机制是人类视觉所特有的大脑信号处理机制。而深度学习网络的注意力机制借鉴了人类的视觉注意力机制，基本思想是在运算过程中忽略无关信息、关注重点信息<sup>[3]</sup>。

Zhao 等人<sup>[15]</sup>将注意力机制引入到语义分割任务中，提出了 PSANet 网络，通过预测注意力图来学习聚合每个位置的上下文信息。但这些注意力图都比较大，会占据大量 GPU 内存并且会提高计算复杂度。

CCNet<sup>[16]</sup>模块可以插入任意完全卷积的神经网络，BiSeNet<sup>[17]</sup>网络无需任何上采样操作即可整合全局语境信息，提高了分割速度，HMANet<sup>[18]</sup>是一个用于航空图像语义分割的混合多注意力网络。

自注意力机制也在图像语义分割任务上取得显著效果。DANet<sup>[19]</sup>是一种新型的场景语义分割网络，利用“自注意力机制”捕获丰富的语义信息。该网络在带有空洞卷积的 ResNet 网络架构的尾部添加两个并行的注意力模块：位置注意力模块和通道注意力模块。为了更好地利用两

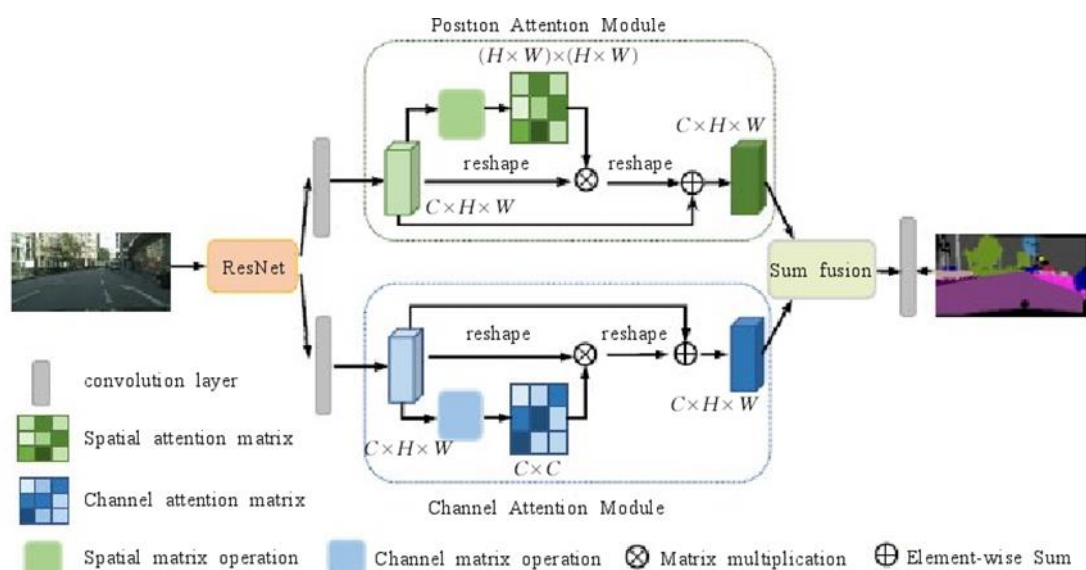


图 2.4 DANet 架构

个注意力模块的全局语义信息，网络还将模块的输出经过一个卷积层后进行主元素相加实现特



征融合，最后通过一个卷积层得到最终预测结果。

## 2.2.4 基于 GAN 的方法

使用条件随机场（CRF）进行结构化预测存在计算量大、模型复杂等问题，在图像语义分割过程中，使用 GAN 代替 CRF 捕获图像上下文信息可以很好地避免上述问题，还可以在不增加模型的训练时间和复杂度的情况下增加长距离空间标签的连续性。

基于 GAN 的这类方法一般使用 FCN、SegNet 等分割网络作为生成器网络，原始输入图像经过生成器网络处理后得到预测的分割图像，将预测的分割图像作为人造样本、真实的标签图像作为真实样本输入判别器网络，判别器网络学习真实样本和人造样本的区别，并基于博弈思想进行对抗训练，内部的反馈机制会对生成器网络与判别器网络进行调节，经过数次迭代训练，生成器网络的分割准确率和判别器网络的鉴别能力都会不断提高<sup>[6]</sup>。

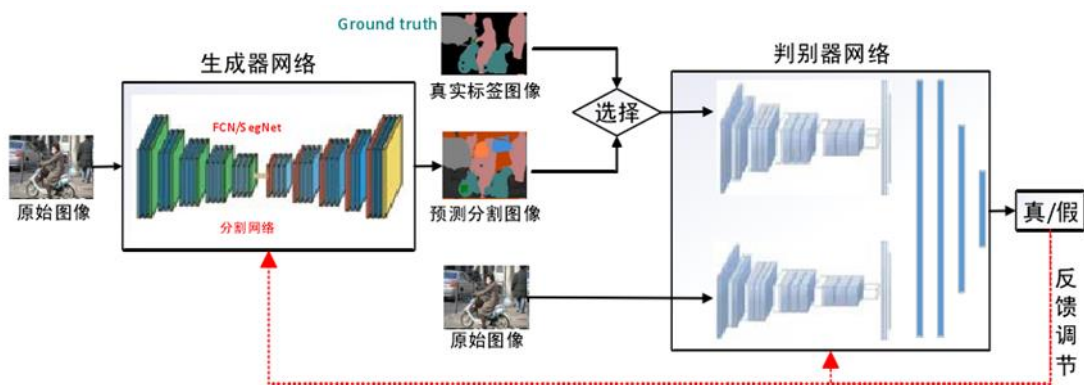


图 2.5 基于 GAN 的方法架构

2016 年，Luc 等人<sup>[20]</sup>首次将 GAN 引入图像分割领域，提出一种图像分割的新方法，使用判别器识别真实标签与分割图像，缩小标签和分割图像之间的不一致性，提高分割网络的分割准确率。Hoffman 等人<sup>[21]</sup>将 GAN 与领域适应性思想结合，将源域与目标域共享标记空间，并通过最优化目标损失函数来减少全局偏移和特定偏移的影响，突出用于图像语义分割的领域适应性框架。

## 2.3 弱监督学习图像语义分割方法

### 2.3.1 基于边框级标注方法

这一类的代表方法有 BoxSup 网络和 DeepCut 方法。

边框级标注泛指一些人工标记的边界框，基于边框级标注的方法使用边框级标注图像作为训练样本，该方法时间成本较低，可以大大节省人力物力，并且依然包含了丰富的监督信息和物体位置。

Dai 等人<sup>[22]</sup>以 FCN 为基础网络，用边框级标注的图像作为训练样本，通过循环迭代的方式不断提高分割准确率，提出了 BoxSup 网络。首先，利用 MCG 算法得到初步的目标候选区域；其次，将其作为监督信息输入到 FCN 中进行优化和更新；待 FCN 输出精度更高的候选区域后，再将这个候选区域重新输入到 FCN 中进行训练，如此循环迭代，直至最后准确率收敛。

DeepCut 方法<sup>[23]</sup>是在 GrabCut 算法的基础上加入 CRF 和 CNN 形成的，该方法将分类问题看

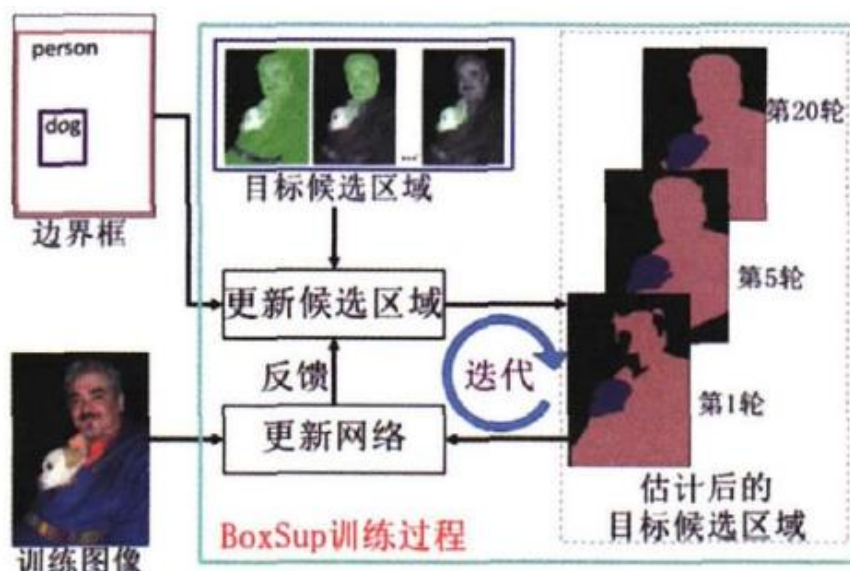


图 2.6 BoxSup 架构

作在稠密连接的条件随机场下的能量最小化问题，通过在 CNN 中进行迭代训练，逐步提高图像的分割精度。

### 2.3.2 基于涂鸦级标注方法

涂鸦级标注是通过图像进行涂写得到任意形式的线段，是最方便用户的标注方式。基于涂鸦级标注的方法使用涂鸦级标注的图像进行训练，分割过程较为简单，分割样本也极易获取。

Lin 等人<sup>[24]</sup>提出基于用户交互的 ScribbleSup 方法，该方法利用图模型训练卷积网络，用来对简笔标注的图像进行语义分割，基于图模型将简笔标注的信息结合空间约束、外观和语义内容，传播到未经过标记的像素上并学习网络参数。分为自动标记阶段和图像训练阶段，自动标记阶段首先根据涂鸦线条对图像生成像素块，然后每个像素块作为图模型中的一个结点，再用 GraphCut 算法建模自动完成对所有训练图像的标注；图像训练阶段是将自动标记阶段完成的标注图像送到 FCN 中去训练，得到语义分割的结果。

### 2.3.3 基于图像级标注方法

图像级标注使用类别标签作为训练标注，无需进行像素标注，制作更简单、工作量更小，是如今弱监督图像分割领域研究的热点。但图像级标注仅仅提供了物体种类信息，缺少形状、位置信息，因此分割难度也很大。

Pinheiro 等人<sup>[25]</sup>采用多示例学习模型构建图像标签与像素之间的关联性，使用超像素和 MCG 算法等平滑每个像素级的类标签。该方法首先使用 ImageNet 图像级标签对模型进行训练，利用 CNN 生成特征平面，然后将这些特征平面通过聚合层对模型进行约束，该模型也得到了很好的分割结果。

## 2.4 我的看法和感想

想要了解进而研究某一领域，需要做好三件事情，首先，要搞清楚这一领域到底是做什么的，然后，需要了解目前存在的尚未开始解决或者是尚未解决彻底的难题，最后，需要根据前人的研

究提出一些自己认为能够解决这些难题的看法。

放到基于深度学习的图像语义分割领域来说，第一个问题关于它是做什么的，前面已经探讨过，所以不再赘述；第二个问题是图像语义分割有什么困难，为什么就解决不好。其实经过前面的探讨，发现困难还是很多的，比如分割边缘不准，其实这一点放在人类身上也是存在的，海天一色是难以分辨的，只不过机器更容易被边缘迷惑，常用于自动汽车驾驶的 SegNet 网络在分割路面情况的图片时，树木和车对应的边缘分割区域边缘一直在震荡、变形，这恰恰说明边缘是极难分割的；又比如标注的费时费力，甚至标注本身就是有噪点不精确的，如果仅凭人工的话，标注进展是极其缓慢的，精确标注一张图至少也要五六分钟，训练一个成熟的网络，最起码也要上万张的训练图像，这是极为耗费人力的，因此目前的公开的数据集都很珍贵；又比如对被遮挡区域进行建模的问题，因为我们是人类，会在脑海中想象物体的全貌，所以对被遮挡物体的分辨是容易的，但是我们如何把这种能力赋予机器呢？我们如何让网络也具有这种想象力呢？这绝不是仅仅加几个通道就能解决的问题。

第三个问题是基于前人的经验提出自己独到的解决方案，由于这里仅仅是读了一些文献，做了一些简单的综述，因此提出解决方案是不可能了，但是可以简单地说一下见解。从之前提到过的各种深度学习网络来看，很多都是在原有经典网络的基础上修修补补得来的，并不是说这种情况不好，只是提一些建议，在修改经典网络的时候，我们需要关注的是大方向而不应该被细节局限了创造力，以 U-net 网络为例，它的大方向应该就是提出了一种较为对称的编码-解码结构，我们在使用它解决问题或者改进它的时候，不应该纠结于编码、解码部分的深度到底选几层，而是应该从提出这个经典网络的本质出发，发挥自己的创造力加以改进。当然，能够完完全全自己提出一种崭新的网络肯定是更好的。

## 2.5 总结与展望

近来，深度学习技术已经广泛应用到图像语义分割领域。本文主要对基于深度学习的图像语义分割的研究进行了大致的分类，根据标注类型和学习方式的不同，将之分为全监督学习图像语义分割方法和弱监督学习图像语义分割方法，又对其中每类方法的代表性算法进行了研究与分析。

经过这一系列总结，我认为未来该领域的研究方向可能有应用于场景解析的图像语义分割、实时图像语义分割、应用于三维数据的语义分割、应用于三维数据的语义分割、应用于视频数据的语义分割以及未来网络架构的轻量化。其中，随着 AR、VR 应用和自动驾驶、行人检测等领域的发展，三维点云数据的分割和未来网络架构的轻量化很有可能成为未来语义分割的研究热点。

## 参考文献

- [1] 季辉, 丁泽军. 双人博弈问题中的蒙特卡洛树搜索算法的改进 [J]. 计算机科学, 2018, 45(01):140-143.
- [2] Russell S J, Norvig P. Artificial Intelligence: A Modern Approach, Third Edition. 2013.
- [3] 景庄伟, 管海燕, 彭代峰, 于永涛. 基于深度神经网络的图像语义分割研究综述 [J/OL]. 计算机工程:1-30[2020-07-04]. <https://doi.org/10.19678/j.issn.1000-3428.0058018>.
- [4] 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(06):1302-1313.
- [5] 梁新宇, 罗晨, 权冀川, 肖铠鸿, 高伟嘉. 基于深度学习的图像语义分割技术研究进展 [J]. 计算机工程与应用, 2020, 56(02):18-28.
- [6] 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.
- [7] Chen L C, Papandreou G, Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. IEEE Trans. on Pattern Analysis & Machine Intelligence, 2014, 39(4):640-651. [doi: 10.1109/TPAMI.2016.2572683]
- [8] 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.
- [9] 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.
- [10] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017. World Health Organization. Factors regulating the immune response: report of WHO Scientific Group [R]. Geneva: W H O, 1970.
- [11] 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.
- [12] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation [J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(12): 2481-2495.

- 
- [13] Kendall A, Badrinarayanan V, Cipolla R. Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding[J]. Proceedings of the British Machine Vision Conference (BMVC), 2017, 57:1-12.
  - [14] 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.
  - [15] 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.
  - [16] Huang Z, Wang X, Huang L, et al. Ccnet: Criss-cross attention for semantic segmentation[C]//Proceedings of the IEEE International Conference on Computer Vision. 2019:603-612.
  - [17] 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.
  - [18] Niu R. HMANet: Hybrid Multiple Attention Network for Semantic Segmentation in Aerial Images[J]. arXiv preprint arXiv:2001.02870, 2020.
  - [19] Fu J, Liu J, Tian H, et al. Dual attention network for scene segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 3146-3154.
  - [20] Luc P, Couprie C, Chintala S, et al. Semantic segmentation using adversarial networks[J]. arXiv preprint arXiv:1611.08408, 2016.
  - [21] 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.
  - [22] Dai J, He K, Sun J. Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 1635-1643.
  - [23] 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, 2016, 36(2): 674-683.
  - [24] Lin D, Dai J, Jia J, et al. Scribblesup: Scribble-supervised convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 3159-3167.
  - [25] Pinheiro P O, Collobert R. From image-level to pixel-level labeling with convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1713-1721.
-