青衣素

图像分割综述阅读——Understanding Deep Learning Techniques for Image Segmentation

> 最近开始研究图像分割的相关技术，打算从综述文章入手，先了解整个领域的研究情况，再具体到各个算法的实现与原理上面去。文章主要以翻译以及个人对文章的理解为主，翻译成中文，便于后续查找相关的知识点。这篇综述总结了到2019年为止常见的图像分割的方法，可以说是一个相关资料的大汇总吧。

Abstract

机器学习社区已经被大量基于深度学习的方法所淹没。卷积神经网络、递归神经网络、对抗神经网络、自编码等多种深部神经网络正有效地解决无约束环境下目标的检测、定位、识别和分割等具有挑战性的计算机视觉任务。在对目标检测或识别领域进行了大量分析研究的同时，在图像分割技术方面出现了许多新的深度学习技术。本文从分析的角度探讨了图像分割的各种深度学习技术。这项工作的主要目标是提供一个直观的理解，对图像分割领域作出重大贡献的主要技术。本文从传统的图像分割方法出发，阐述了深度学习对图像分割领域的影响。此后，大多数主要的分割算法都被逻辑地归类为专门针对其独特贡献的段落。有了大量直观的解释，读者就可以更好地看到这些过程的内部动态。

1 Introduction

图像分割可以定义为一种特定的图像处理技术，用于将图像分割成两个或多个有意义的区域。图像分割也可以看作是定义图像中不同语义实体之间边界的过程。从更技术的角度来看，图像分割是将标签分配给图像中的每个像素的过程，使得具有相同标签的像素相对于某种视觉或语义特性被连接（图1）。

图像分割包含了计算机视觉中的一类精细相关的问题。最经典的版本是语义分割[66]。在语义分割中，每个像素被分类为一组预定的类别中的一个，以使得属于同一类别的像素属于图像中的唯一语义实体。还值得注意的是，所讨论的语义不仅取决于数据，还取决于需要解决的问题。例如，对于行人检测系统，人应该属于同一部分，但是对于动作识别系统，可能有必要将不同的身体部位分为不同的类别。其他形式的图像分割可以集中在场景中最重要的对象上。由此产生了一类特殊的问题，即显着性检测[19]。该域的其他变体可能是前景背景分离问题。在许多系统中，例如图像检索或视觉问题解答，通常需要计算对象的数量。实例特定的细分解决了该问题。实例特定的分割通常与对象检测系统结合使用，以检测和分割场景中同一对象的多个实例[43]。时间空间中的分割也是一个具有挑战性的领域，并且具有各种应用。在对象跟踪方案中，像素级别分类不仅在空间域中执行，而且跨时间执行。流量分析或监视中的其他应用程序需要执行运动分割以分析运动对象的路径。在具有较低语义级别的分割领域中，过度分割也是一种常见的方法，其中将图像划分为非常小的区域以确保边界依从性，但以创建大量虚假边缘为代价。过度分割算法通常将其与区域合并技术结合使用以执行图像分割。甚至简单的颜色或纹理分割也可以在各种情况下使用。分割算法之间的另一个重要区别是需要用户交互。尽管希望拥有全自动系统，但用户的一点点互动可以提高质量在很大程度上进行细分。这尤其适用于我们处理复杂的场景，或者我们没有足够的数据来训练系统。

分割算法在现实世界中有多种应用。**在医学图像处理[123]中，我们还需要定位各种异常，例如动脉瘤[48]，肿瘤[145]，癌性元素（例如黑素瘤检测[189]）或手术过程中的特定器官[206]**(*-思考：医学图像的应用我觉得可能会是图像分割比较好的发展领域-*)。分割的另一个重要领域是监视。许多问题，如行人检测[113]、交通监视[60]，都需要对特定对象（如人或车）进行分割。其他领域包括卫星图像[11，17]、防御制导系统[119]、法医学，如人脸[5]、虹膜[51]和指纹[144]识别。一般的传统方法如直方图阈值化[195]、杂交[193，87]特征空间聚类[40]、基于区域的方法[59]、边缘检测方法[184]、模糊方法[39]、基于熵的方法[47]、神经网络（Hopfield神经网络[35]、自组织映射[27]、基于物理的方法[158]等。在这方面被广泛使用。然而，这种基于特征的方法有一个共同的瓶颈，那就是它们依赖于领域专家提取的特征的质量。一般来说，在图像分割中，人们往往会忽略潜在的或抽象的特征。另一方面，一般来说，深度学习解决了自动特征学习的问题。在这方面，计算机视觉中最常见的技术之一很快就被引入了卷积神经网络[110]，它通过反向传播学习一组级联卷积核[182]。从那时起，它已经显著地改进了诸如分层训练[13]、校正线性激活[153]、批处理规范化[84]、辅助分类器[52]、阿托罗斯卷积[211]、跳

公告

昵称： 青衣素
园龄： 1年6个月
粉丝： 0
关注： 0
+加关注

< 2021年10月 >

日	一	二	三	四	五	六
26	27	28	29	30	1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31	1	2	3	4	5	6

搜索

找找看
谷歌搜索

常用链接

我的随笔
我的评论
我的参与
最新评论
我的标签

最新随笔

1.210场周赛
2.第209场周赛
3.834. 树中距离之和
4.LCP 19.秋叶收藏集
5.113.路径总和II
6.KDD Debiasing top 方案
7.968.监控二叉树
8.第207场周赛
9.Tweedie损失函数
10.深度学习入门比赛——街景字符识别（五）

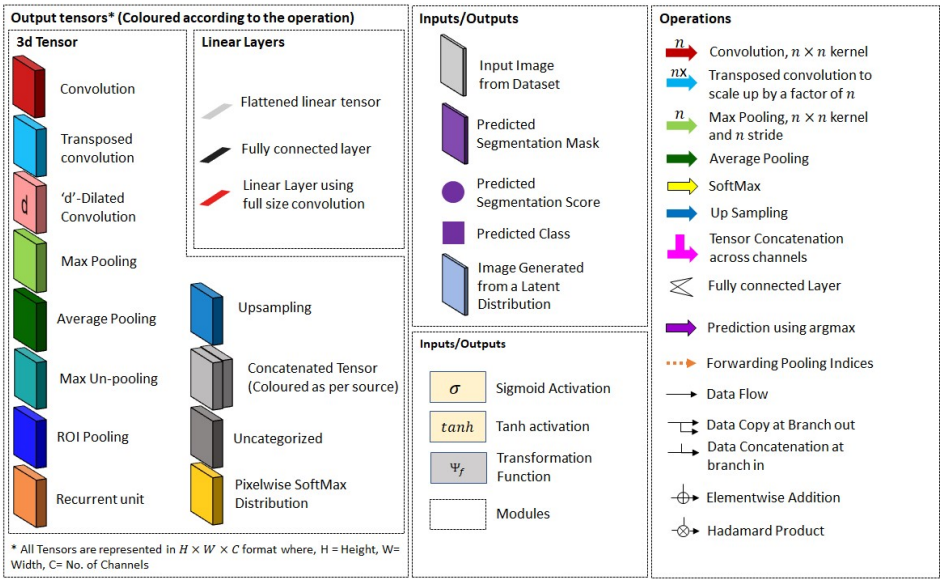
我的标签

leetcode(5)
街景字符识别(4)
深度学习比赛(4)
kaggle(2)
leetcode210场周赛(1)
力扣209周赛(1)
树形dp(1)
hard(1)
中等(1)
动态规划(1)
更多

过连接[78]、更好的优化技术[97]等特征。伴随着这些，出现了大量新的图像分割技术。各种这样的技术从流行的网络如AlexNet[104]、卷积自编码器[141]、递归神经网络[143]、残差网络[78]等中得到了启发。

2 Motivation

关于与图像分割相关的传统技术，已经有了许多评论和调查[61160]。他们中的一些人专门研究应用领域[107，123，185]，而另一些人则专注于特定类型的算法[20，19，59]。随着深度学习技术的到来，出现了许多新的图像分割算法。早期的研究[219]显示了基于深度学习的方法的潜力。最近的一些研究[68]涵盖了许多方法，并根据报告的性能对它们进行了比较。加西亚等人的工作。[66]列出了各种基于深度学习的分割技术。他们列出了各种最先进的网络在几个现代挑战中的表现。这些资源对于理解这个领域中的当前状态非常有用。虽然知道可用的方法对于开发产品是非常有用的，但是作为一名研究人员，要为这个领域做出贡献，就需要了解方法的基本机制，使他们有信心。在目前的工作中，我们的主要动机是回答为什么这些方法是按照它们的方式设计的问题。了解现代技术的原理将使它更容易应对新挑战并开发更好的算法。
我们的方法仔细分析了每种方法，以了解它们为什么在所做的事情上成功，以及为什么在某些问题上失败。意识到这种方法的利弊，可以开始新的设计，从正反两方面获益。我们推荐Alberto Garcia Garcia[66]的作品，以获得使用深度学习的一些最佳图像分割技术的概述，同时我们的重点是了解这些技术为什么、何时以及如何应对各种挑战。



2.1 Contribution

这篇论文的设计使新的研究人员获得了最大的利益。在深度学习时代之前，最初讨论了一些传统的技术来支持框架。逐渐地，人们讨论了影响深度学习开始的各种因素，以便读者对机器学习目前的发展方向有一个很好的了解。在随后的章节中，主要的深度学习算法以一种通用的方式进行了简要描述，以便在读者的头脑中建立一个更清晰的过程概念。随后讨论的图像分割算法被归为过去几年在这个领域中的主要算法家族。所有主要方法背后的概念都是用一种非常简单的语言和最少的复杂数学来解释的。几乎所有对应于主要网络的图都是使用如图2所示的公共表示格式绘制的。已经讨论过的各种方法都有不同的体系结构表示。统一的表示方案使用户可以了解网络之间的基本相似之处和不同之处。 最后，讨论了主要的应用领域，以帮助新研究人员追求自己选择的领域。（***注解：文章的叙述流程是由传统的图像分割算法——>目前流行的深度学习图像分割算法，对每个经典的网络进行简要的概述，大体了解每种算法的思路即可，详细的可以单独看各个算法的论文***）

3 Impact of Deep Learning on Image Segmentation

卷积神经网络或深度自编码等深度学习算法的发展不仅影响了目标分类等典型任务，而且在目标检测、定位、跟踪或图像分割等其他相关任务中也很有效。

3.1 Effectiveness of convolutions for segmentation

作为一种操作，卷积可以简单地定义为在将较小的核卷积到较大的图像上时，在核权重和输入值之间执行乘积和的函数。对于具有k个通道的典型图像，我们可以沿x和y方向卷积具有k个通道的较小尺寸的核，以获得二维矩阵格式的输出。有人观察到，在训练典型的CNN之后，卷积核倾向于生成关于对象的某些特征的激活图[214]。鉴于激活的性质，它可以看作是对象特定特征的分割遮罩。因此，生成特定于需求的分段的关键已经嵌入到这个输出激活矩阵中。大多数图像分割算法都是利用CNNs的这一特性，根据需要生成分割模板来解决这一问题。如下面在图3中所示，先前的层捕捉诸如轮廓或对象的小部分的局部特征。在后面的图层中，会激活更多的全局特征，例如场、人或天空。从这个图中还可以注意到，与后面的层相比，前面的层显示出更尖锐的激活。（***注解：个人理解，卷积操作是结合上下文，在局部区域提取像素图像的特征（相同的或者不同的），随着网络层数的增***

加，局部区域感受野不断增加，会提取全局特征）

Figure 3: Input image and sample activation maps from a typical CNN. (Top row) Input image and two activation maps from earlier layers showing part objects like t-shirts and features like contours. (Bottom row) shows activation maps from later layers with more meaningful activations like fields, people and sky respectively

3.2 Impact of larger and more complex datasets

深度学习给图像分割领域带来的第二个冲击是大量的数据集、挑战和竞争。这些因素鼓励世界各地的研究人员提出各种最先进的技术来实现跨领域的分割。表1列出了许多这样的数据集。

Table 1: A list of various datasets in the image segmentation domain

Category	Dataset
Natural Scenes	Berkeley Segmentation Dataset [140]
	PASCAL VOC [54]
	Stanford Background Dataset [72]
	Microsoft COCO [122]
	MIT Scene parsing data(ADE20K) [222] [223]
	Semantic Boundaries Dataset [75]
	Microsoft Research Cambridge Object Recognition Image Database (MSRC) [188]
Video Segmentation Dataset	Densely Annotated Video Segmentation(DAVIS) [168]
	Video Segmentation Benchmark(VSB100) [64]
	YouTube-Video object Segmentation [209]
Autonomous Driving	Cambridge-driving Labeled Video Database (CamVid) [23]
	Cityscapes: Semantic Urban Scene Understanding [41]
	Mapillary Vistas Dataset [155]
	SYNTHIA: Synthetic collection of Imagery and Annotations [178]
	KITTI Vision Benchmark Suite [67]
	Berkeley Deep Drive [212]
	India Driving Dataset(IDD) [202]
Aerial Imaging	Inria Aerial Image Labeling Dataset [134]
	Aerial Image Segmentation Dataset [213]
	ISPRS Dataset collection [57]
	Google Open Street Map [8]
	DeepGlobe [49]
Medical Imaging	DRIVE:Digital Retinal Images for Vessel Extraction [191]
	Sunnybrook Cardiac Data [171]
	Multiple Sclerosis Database [129] [25]
	IMT: Intima Media Thickness Segmentation Dataset [148]
	SCR: Segmentation in Chest Radiographs [201]
	BRATS: Brain Tumor Segmentation [146]
	LITS: Liver Tumour Segmentation [74]
	BACH: Breast Cancer Histology [6]
	IDRiD: Indian Diabetic Retinopathy Image Dataset [169]
	ISLES: Ischemic Stroke Lesion Segmentation [135]
Saliency Detection	MSRA Salient Object Database [37]
	ECSSD: Extended Complex Scene Saliency Dataset [187]
	PASCAL-S DATASET [117]
	THUR15K: Group Saliency in Image [36]
	JuddDB: MIT saliency benchmark [18]
	DUT-OMRON Image Dataset [210]
Scene Text Segmentation	KAIST Scene Text Database [112]
	COCO-Text [203]
	SVT: Street View Text Dataset [205]

4 Image Segmentation using Deep Learning

如前所述，卷积在生成语义激活图方面非常有效，该语义激活图具有固有地构成各种语义段的组件。已经实现了各种方法来利用这些内部激活来分割图像。表2总结了主要的基于深度学习的分割算法，并简要描述了它们的主要贡献。

Table 2: A summary of major deep learning based segmentation algorithms.
Abbreviations: S: Supervised, W: Weakly supervised, U: Unsupervised, I: Interactive, P: Partially Supervised, SO: Single objective optimization, MO: Multi
objective optimization, AD: Adversarial Learning, SM: Semantic Segmentation,
CL: Class specific Segmentation, IN: Instance Segmentation, RNN: Recurrent
Modules, E-D: Encoder Decoder Architecture

Method	Year	Supervision					Learning			Type			Modules		Description
		S	W	U	I	P	SO	MO	AD	SM	CL	IN	RNN	E-D	
Global Average Pooling	2013		✓				✓				✓				Object specific soft segmentation
DenseCRF	2014					✓	✓			✓					Using CRF to boost segmentation
FCN	2015	✓					✓			✓					Fully convolutional layers
DeepMask	2015	✓						✓			✓				Simultaneous learning for segmentation and classification
U-Net	2015	✓					✓			✓				✓	Encoder-Decoder with multiscale feature concatenation
SegNet	2015	✓					✓			✓				✓	Encoder-Decoder with forwarding pooling indices
CRF as RNN	2015	✓						✓		✓			✓		Simulating CRFs as trainable RNN modules
Deep Parsing Network	2015	✓						✓							Using unshared kernels to incorporate higher order dependency
BoxSup	2015		✓							✓					Using bounding box for weak supervision
SharpMask	2016	✓						✓			✓			✓	Refined Deepmask with multi layer feature fusion
Attention to Scale	2016	✓					✓			✓					Fusing features from multi scale inputs
Semantic Segmentation	2016	✓							✓	✓					Adversarial training for image segmentation
Conv LSTM and Spatial Inhibition	2016	✓						✓				✓	✓		Using spatial inhibition for instance segmentation
JULE	2016			✓				✓		✓			✓		Joint unsupervised learning for segmentation
ENet	2016	✓					✓			✓					Compact network for realtime segmentation
Instance aware segmentation	2016	✓						✓				✓			Multi task approach for instance segmentation
Mask RCNN	2017	✓						✓		✓					Using region propsal network for segmentation
Large Kernel Matters	2017	✓					✓			✓				✓	Using larger kernels for learning complex features
RefineNet	2017	✓					✓			✓				✓	Multi path refinement module for fine segmentation
PSPNet	2017	✓					✓			✓					Multi scale pooling for scale agnostic segmentation
Tiramisu	2017	✓					✓			✓				✓	DenseNet 121 feature extractor
Image to Image Translation	2017	✓							✓	✓				✓	Conditional GAN for translation image to segment maps
Instance Segmentation with attention	2017	✓						✓				✓	✓		Attention modules for image segmentatoin
W-Net	2017			✓				✓		✓				✓	Unsupervised segmentation using normalized cut loss
PolygonRNN	2017				✓		✓			✓			✓		Generating contours by RNN
Deep Layer Cascade	2017	✓					✓			✓					Multi level approach to handle pixels of different complexity
Spatial Propagation Network	2017	✓					✓			✓					Refinement using linear label propagation
DeepLab	2018	✓					✓			✓					Atrous convolution, Spatial pooling pyramid, DenseCRF
SegCaps	2018	✓						✓							Capsule Networks for Segmentation
Adversarial Collaboration	2018			✓				✓							Adversarial collaboration between multiple networks
Superpixel Supervision	2018			✓				✓							Using superpixel refinement as supervisory signals
Deep Extreme Cut	2018				✓		✓			✓					Using extreme points for interactive segmentation
Two Stream Fusion	2019				✓		✓			✓					Using image stream and interaction stream simultaneously
SegFast	2019	✓					✓			✓				✓	Using depth-wise separable convolution in SqueezeNet encoder

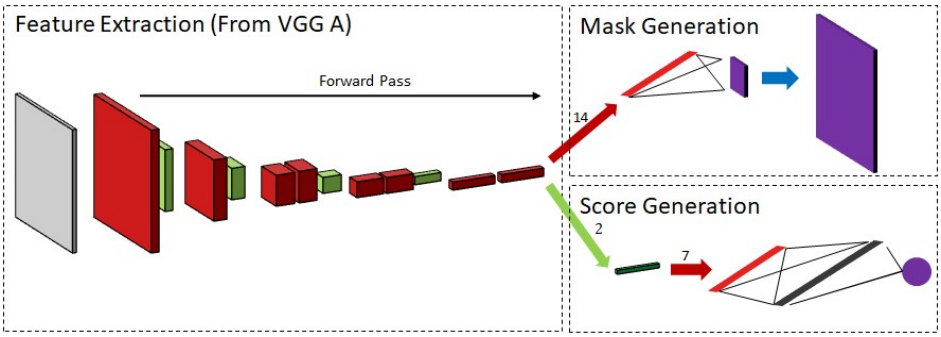
4.1 Convolutional Neural Networks

卷积神经网络是计算机视觉中最常用的方法之一，为了更好地完成分割任务，它采用了许多简单的改进。

4.1.1 Fully convolutional layers

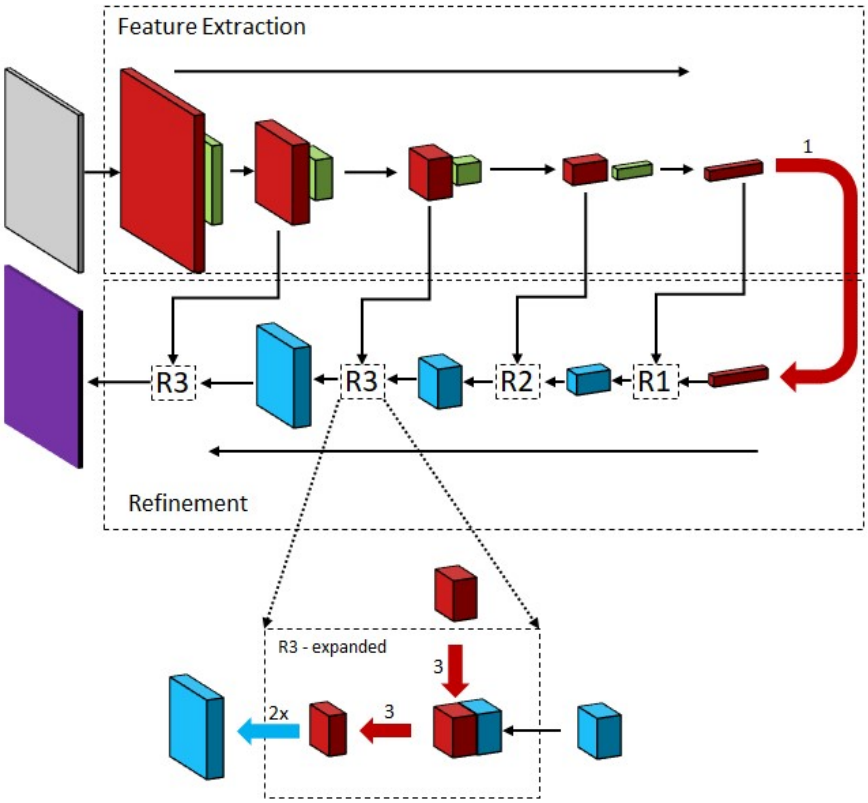
分类任务通常需要以类数上的概率分布形式的线性输出。为了将二维激活图的体积转换成线性层，它们通常被展平。扁平的形状允许完全连接的网络执行以获得概率分布。然而，这种重构会丢失图像中像素之间的空间关系。在完全卷积神经网络（FCN）[130]中，最后卷积块的输出直接用于像素级分类。FCNs首先在PASCAL VOC 2011分割数据集[54]上实现，像素精度为90.3%，平均IOU为62.7%。避免完全连接的线性层的另一种方法是使用全尺寸平均池将一组二维激活图转换为一组标量值。由于这些集合标量被连接到输出层，因此对应于每个类的权重可用于对前几层中对应的激活映射执行加权求和。这个过程称为全局平均池（GAP）[121]可以直接用于各种训练网络，如残差网络，以找到可用于像素级分割的对象特定激活区。这种算法的主要问题是由于中间子采样操作而导致的锐度损失。子采样是卷积神经网络中增加核感觉面积的常用操作。它的意思是，当激活映射在随后的层中减小时，卷积在这些层上的内核实际上对应于原始图像中更大的区域。但是，它在处理过程中减小了图像大小，当采样到原始大小时，图像大小会失去锐度。已经实施了许多方法来处理这个问题。对于完全卷积模型，可以使用前几层的跳过连接来获得更清晰的激活版本，从中可以划出更细的片段（参见图4）。另一项工作显示了如何使用高维核来捕获FCN模型的全局信息，从而创建了更好的分割遮罩[165]。分割算法也可以看作是边界检测技术。从这个角度来看，卷积特征也非常有用[139]。虽然早期的层可以提供精细的细节，但后期的层更关注较粗的边界。

深度掩模和锐器掩模 DeepMask[166]是Facebook人工智能研究（F AIR）一个与图像分割相关的项目的名字。它展示了与FCN模型相同的思想流派，只是模型能够执行多任务（参见图5）。



它有两个主要分支，来自一个共享的要素表示。其中一个为中心对象创建像素级分类或概率掩码，第二个分支生

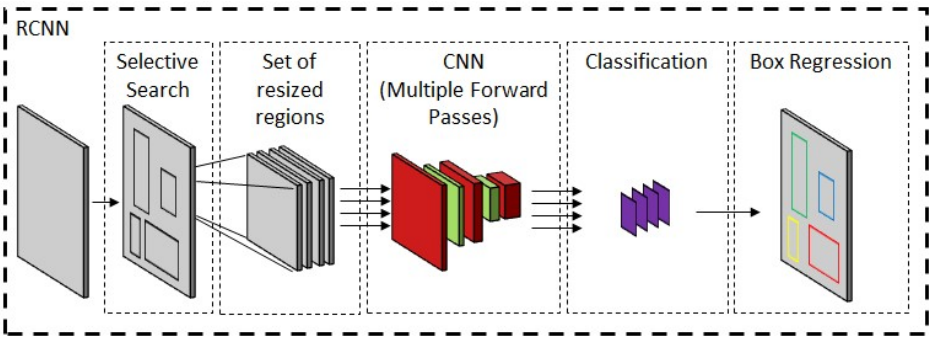
成与对象识别精度相对应的分数。该网络加上16步的滑动窗口，在图像的不同位置创建对象片段，而分数有助于识别哪些片段是好的。该网络在SharpMask [167]中得到了进一步升级，在每个步骤中，通过使用卷积细化以自上而下的方式组合来自每一层的概率蒙版以生成高分辨率蒙版（请参见图6）。 Sharpmask的平均召回率为39.3，胜过Deepmask，后者在MS COCO Segmentation Dataset上的得分为36.6。

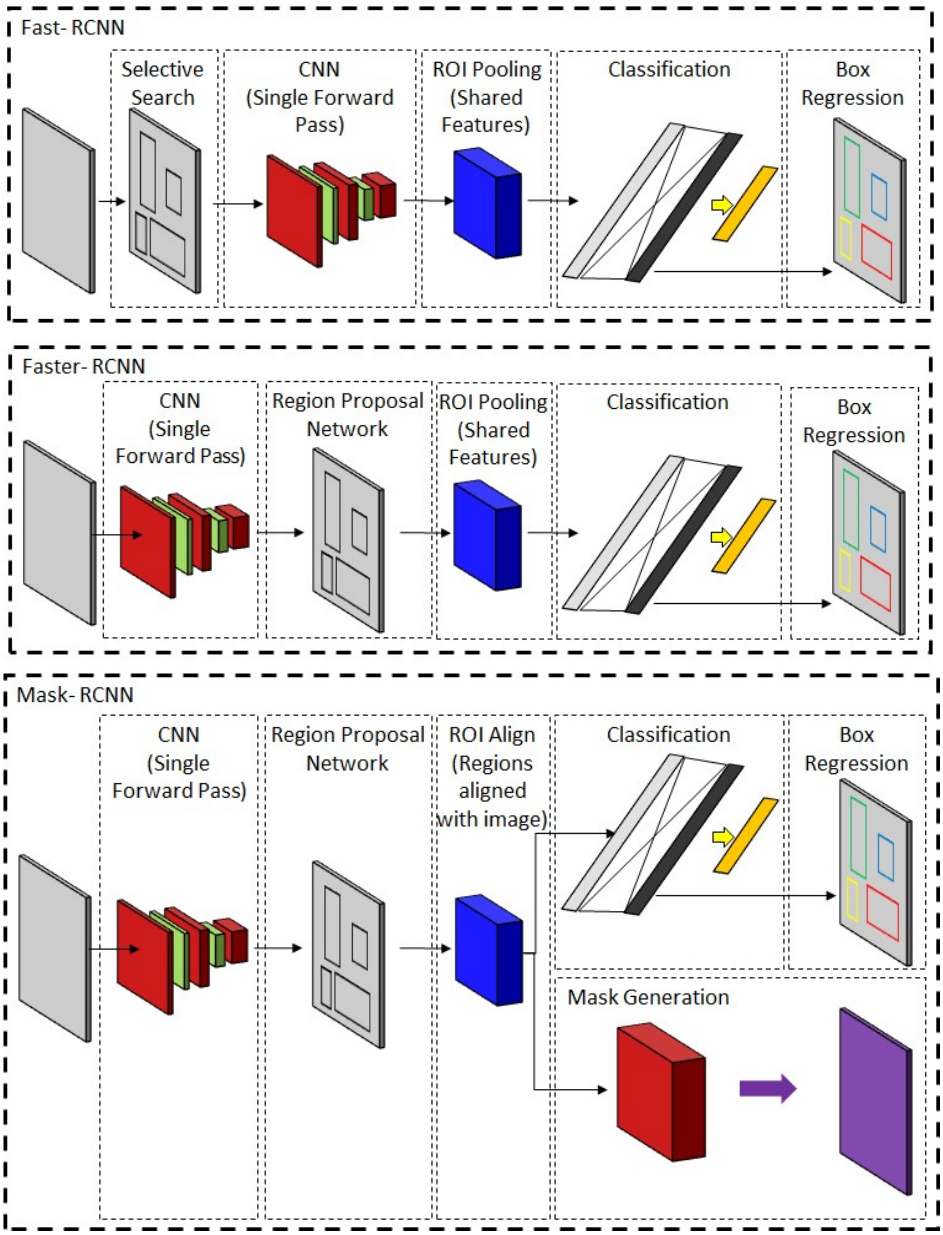


4.1.2 Region proposal networks

另一个类似的图像分割一起飞速发展的是目标定位。这样的任务涉及到在图像中定位特定对象。此类问题的预期输出通常是与查询对象对应的一组边界框。虽然严格地说，这些算法中的一些并没有解决图像分割问题，但是它们的方法与这个领域是相关的。

RCNN (Region-based Convolutional Neural Networks) CNNs的引入在计算机视觉领域提出了许多新的问题。其中一个主要问题是像AlexNet这样的网络是否可以被扩展以检测多个对象的存在。基于区域的CNN[70]或更常见的R-CNN使用选择性搜索技术来提出可能的目标区域，并在裁剪窗口上执行分类，以验证基于输出概率分布的合理定位。（***注解：RCNN是目标检测经典算法，个人觉得目标检测和图像分割具有很强的关联性，在做分割的时候，或许可以利用目标检测的相关手段***）**选择性搜索技术[198，200]分析各种方面，例如纹理，颜色或强度，以将像素聚类为对象。**与这些段相对应的边界框通过分类网络来列出一些最敏感的框。最后，用简单的线性回归网络可以得到更紧密的坐标。这项技术的主要缺点是它的计算成本。网络需要为每个包围盒命题计算一个前向通过。跨所有框共享计算的问题是，框的大小不同，因此无法实现大小一致的特征。在改进的Fast R-CNN[69]中，提出了ROI（感兴趣区域）池，其中感兴趣区域被动态地池化以获得固定大小的特征输出。此后，网络主要被候选区域建议的选择性搜索技术所限制。在快速RCNN[175]中，不依赖于外部特征，而是使用中间激活映射来提出边界框，从而加快特征提取过程。边界框代表对象的位置，但它们不提供像素级分段。更快的R-CNN网络被扩展为Mask R-CNN[76]，其并行分支执行像素级对象特定的二进制分类以提供准确的片段。使用掩模RCNN，COCO[122]测试图像的平均精度为35.7。RCNN算法家族如图7所示。区域建议网络经常与其他网络相结合[11844]以给出实例级分段。RCNN在HyperNet[99]的名字下进一步改进，使用了特征抽取器的多层特征。区域建议网络也已经实现，例如特定的分割。如前所述，像RCNN这样的方法的目标检测能力通常与分割模型相结合，以便为同一对象的不同实例生成不同的遮罩[43]。

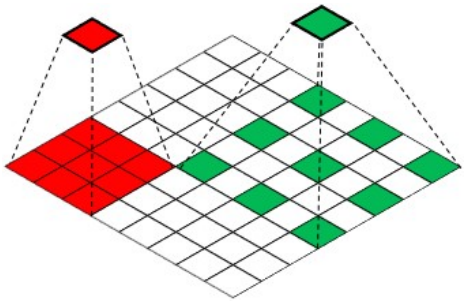




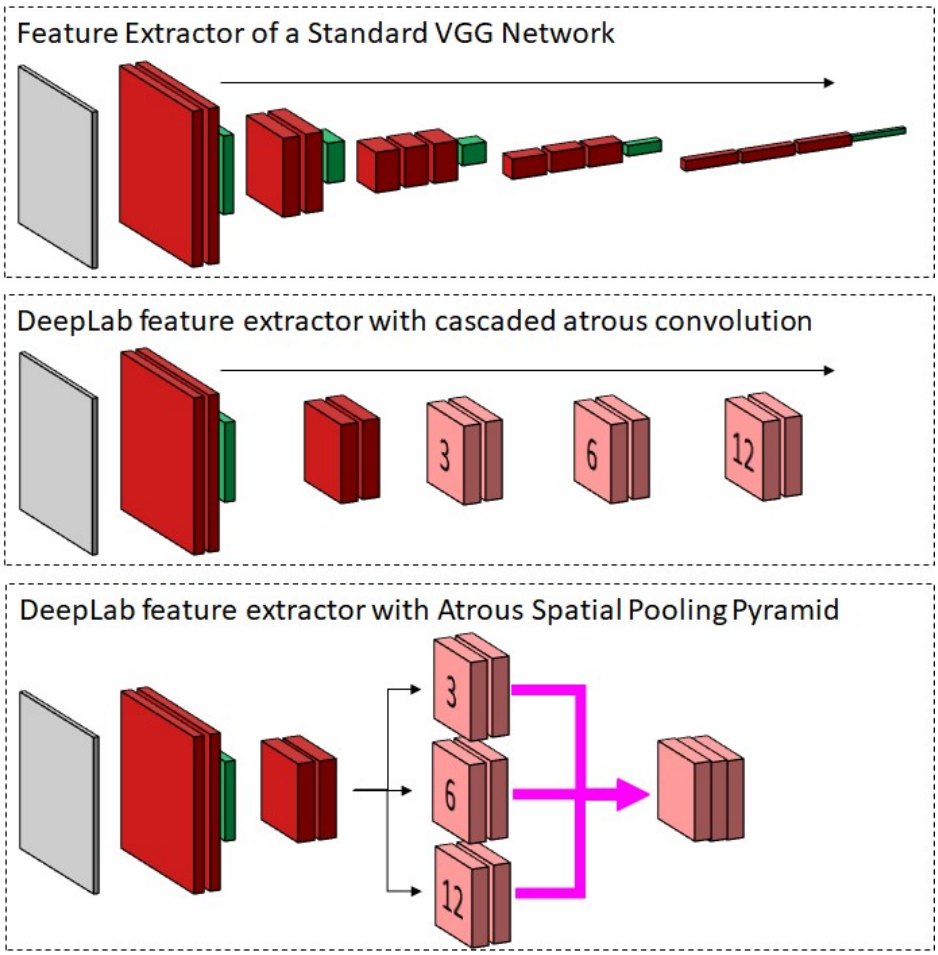
4.1.3 DeepLab

虽然像素级分割是有效的，但两个互补的问题仍然影响着性能。首先，较小的内核大小无法捕获上下文信息。在分类问题中，这是通过使用池层来处理的，池层相对于原始图像增加了内核的感觉区域。但在分割中，会降低分割输出的清晰度。由于可训练参数的数量显著增加，较大内核的替代使用往往较慢。为了解决这个问题，DeepLab[30，32]系列算法演示了各种方法的使用，如atrus卷积[211]、空间池金字塔[77]和完全连接的条件随机场[100]，以高效地执行图像分割。DeepLab算法能够在PASCAL VOC 2012数据集上获得79.7的平均值[54]。

原子/膨胀卷积 任何一层卷积核的大小决定了网络的感觉反应区域。当较小的内核提取局部信息时，较大的内核试图关注更多的上下文信息。然而，较大的内核通常带有更多的参数。例如，要有6×6的感觉区域，必须有36个神经元。为了减少CNN中参数的数量，通过像池技术这样的技术，在更高层增加了感觉区域。池层减少了图像的大小。当一个图像由一个2×2的核（步长为2）合并时，图像的大小减少了25%。在原始图像中，3×3的核对应着6×6的更大的感觉区域。然而，与以前不同的是，卷积核只需要18个神经元（每层9个）。在分割的情况下，池会产生新的问题。图像尺寸的减小会导致生成的图像段的锐度损失，因为缩小后的地图会缩放到图像尺寸。为了同时处理这两个问题，扩张或萎缩的卷积起着关键的作用。萎缩/扩张卷积在不增加参数数目的情况下增加视野。如图8所示，膨胀系数为1的3×3核可以作用于图像中5×5的区域。核的每一行和每一列都有三个神经元，它们与图像中的强度值相乘，强度值由1的膨胀因子分离。通过这种方式，核可以跨越更大的区域，同时保持较低的神经元数量，并保持图像的清晰度。除了DeepLab算法，atrus卷积[34]也被用于基于自动编码器的体系结构。



空间金字塔池 R-CNN引入了空间金字塔池[77]，其中ROI池显示了使用多尺度区域进行对象定位的好处。然而，在DeepLab中，为了改变视野或感觉区域，萎缩性卷曲优先于合并层。为了模拟ROI池的效果，将多个具有不同膨胀度的萎缩卷积分支组合在一起，利用多尺度特性进行图像分割。



全连通条件随机场 条件随机场是一种无向判别概率图模型，常用于各种序列学习问题。与离散分类器不同，在对样本进行分类时，它会考虑其他相邻样本的标签。图像分割可以看作是一个像素分类序列。像素的标签不仅依赖于其自身的强度值，还依赖于相邻像素的值。这种概率图形模型的使用通常用于图像分割领域，因此它值得专门一节（第4.1.4节）。

4.1.4 Using inter pixel correlation to improve CNN based segmentation(利用像素间相关性改进CNN分割)

使用概率图形模型，如马尔可夫随机场（MRF）或条件随机场（CRF）进行图像分割，即使不包括基于CNN的特征抽取器，也有其自身的发展。CRF或MRF的主要特征是具有一元和成对分量的能量函数。

$$E(x) = \underbrace{\sum_i \theta_i(x_i)}_{\text{unary potential}} + \underbrace{\sum_{ij} \theta_{ij}(x_i, x_j)}_{\text{pairwise potential}}$$

非深度学习方法侧重于建立有效的成对势函数，如利用长程依赖关系、设计高阶势函数和探索语义标签的上下文；基于深度学习的方法侧重于产生强的一元势函数，并使用简单的成对成分来提高性能。crf通常以两种方式与基于深度学习的方法相结合。一个作为单独的后处理模块，另一个作为端到端网络（如深度解析网络[128]或空间传播网络[126]）中的可训练模块。

利用CRFs改进全卷积网络 最早启动这种边界细化范例的实现之一是[101]的工作，它引入了用于图像分割的完全卷积网络，很有可能为图像中的对象绘制粗段。然而，获得更清晰的片段仍然是个问题。在[29]的工作中，输出

像素级别的预测被用作完全连接的CRF的一元电势。 对于图像中的每对像素i和j，成对电位定义为：

$$\theta_{ij}(x_i, x_j) = \mu(x_i, x_j)\left[w_1 \exp\left(-\frac{||p_i - p_j||^2}{2\sigma_\alpha^2} - \frac{||I_i - I_j||^2}{2\sigma_\beta^2}\right) + w_2 \exp\left(-\frac{||p_i - p_j||^2}{2\sigma_\gamma^2}\right)\right]$$

这里，u（Xi，XJ）=1，否则0和W1，W2是给内核的权重。表达式使用两个高斯核。第一种是双边核，它依赖于RGB通道中的像素位置（pi，pj）及其对应的强度。第二个核仅依赖于像素位置。σα、σβ和σγ控制高斯核的尺度。这种成对势能函数设计背后的直觉是，确保RGB通道中强度相似的邻近像素被分类在同一类下。这个模型后来也被包括在一个叫做DeepLab的流行网络中（参见第4.1.3节）。在不同版本的DeepLab算法中，CRF的使用能够显著提高Pascal 2012数据集上的平均IOU（在某些情况下高达4%）。

CRF as RNN 尽管CRF对于任何基于深度学习的语义图像分割体系结构来说都是一个有用的后处理模块[101]，但其主要缺点之一是不能用作端到端体系结构的一部分。在标准CRF模型中，成对势可以用加权高斯和来表示。然而，由于精确极小化是困难的，CRF分布的平均场近似被认为是用一个更简单的形式来表示分布，这个简单的形式只是独立边缘分布的乘积。这种自然形式的平均场近似不适合反向传播。在[221]的工作中，这个步骤被一组卷积运算所代替，该卷积运算在循环管道上迭代，直到达到收敛。正如他们在工作中所报告的，与BoxSup的71.0和DeepLab的72.7相比，使用提议的方法获得了74.7的mIOU。操作顺序最容易解释如下。

1.初始化：一元位势上的SoftMax操作可以给我们提供要处理的初始分布。

2.消息传递：使用两个高斯核、一个空间核和一个双边核进行卷积。与CRF的实际实现类似，为了有效地计算全连通CRF，在构建超自面体网格时也会发生splatting和slicing

3.加权滤波器输出：用1×1核与所需信道数卷积，滤波器输出可以加权和求和。通过反向传播可以很容易地学习权重。

4.相容性变换：考虑一个相容性函数来跟踪不同标签之间的不确定性，一个简单的1×1卷积（输入和输出通道数相同）就足以模拟这种情况。与指定相同惩罚的potts模型不同，这里可以学习兼容性函数，因此是一个更好的选择。

5.增加一元电位：这可以通过从一元电位的相容性变换中减去惩罚来实现

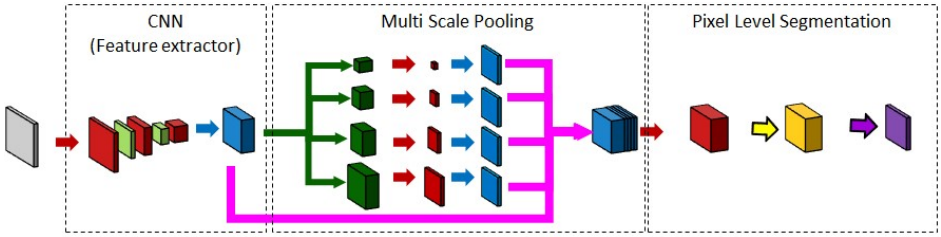
6.规范化：输出可以用另一个简单的softmax函数规范化。

合并高阶依赖项 另一个端到端网络受CRFs的启发，将高阶关系合并到一个深层网络中。在深度解析网络中，使用一系列特殊的卷积和池操作来增强来自标准VGGlike特征提取程序（但池操作较少）的像素级预测。首先，通过在特征映射的不同位置上实现大的非共享卷积核的局部卷积，获得与平移相关的、对长距离相关性建模的特征。与标准crf类似，空间卷积基于局部标签上下文惩罚概率映射。最后，使用块最小池（block min pooling）在深度上执行像素最小池（pixel wise min pooling），以接受惩罚最小的预测。类似地，在文献[126]的工作中，我们提出了一个行/列传播模型来计算图像上的全局成对关系。从稀疏变换矩阵中提取一个稠密的相似矩阵，根据像素的相似性对粗略预测的标签进行重新分类。

4.1.5 Multi-scale networks(多尺度网络)

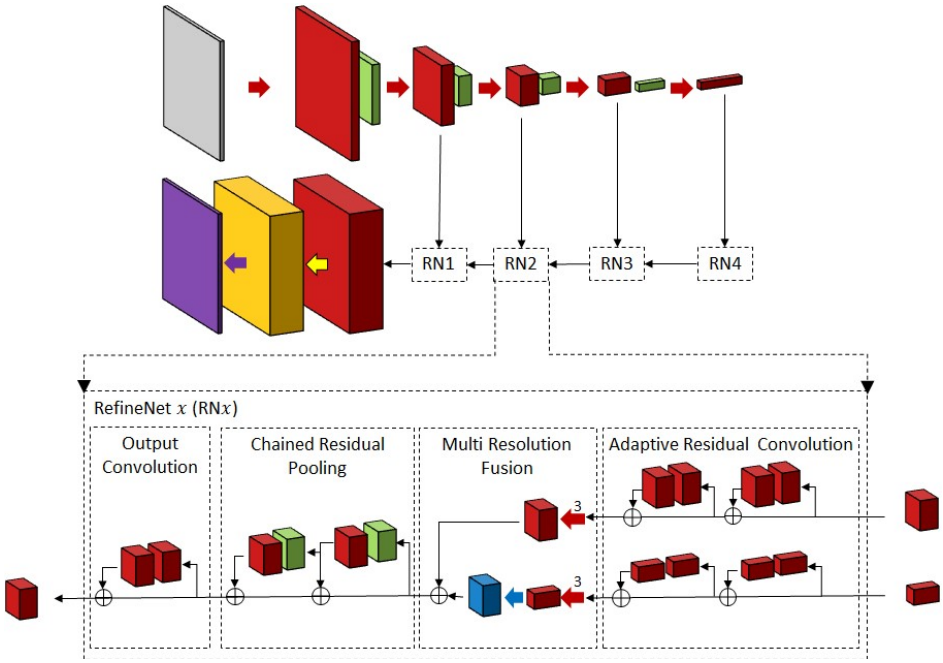
自然场景图像分割的一个主要问题是，感兴趣的对象的大小是非常不可预测的，因为在现实世界中，对象可能有不同的大小，对象可能看起来更大或更小，这取决于对象和相机的位置。CNN的性质决定了精细的小尺度特征在早期层中被捕获，而当一个特征在网络的深度上移动时，这些特征对于更大的对象变得更加具体。例如，场景中的一辆微型车由于像合用或下采样这样的操作，被捕获到更高层的可能性要小得多（***注解：个人理解，小尺度的特征在浅层网络中被捕获，随着网络层数的增加，感受野不断增大，尺度越大的特征将会被捕获***）。从不同尺度的特征图中提取信息以创建不确定图像中对象大小的分段通常是有益的。多尺度自动编码器模型[33]考虑激活不同分辨率以提供图像分割输出。

PSPNet 金字塔场景解析网络[220]建立在FCN之上基于像素级别的分类网络。来自ResNet-101网络的特征图通过多尺度池化层转换为不同分辨率的激活，然后再对其进行升采样并与原始特征图连接以进行分割（请参见图10）。利用辅助分类器进一步优化了ResNet等深层网络的学习过程。不同类型的池模块关注激活映射的不同区域。不同大小的核池（如1×1、2×2、3×3、6×6）会查看激活图的不同区域，以创建空间池金字塔。在ImageNet场景解析挑战中，PSPNet的平均IoU为57.21，FCN为44.80，SegNet为40.79。



RefineNet 使用CNN最后一层的特征可以为对象段生成软边界。这一问题在具有空洞卷积的DeepLab算法中得以避免。RefineNet[120]采用了另一种方法，即细化中间激活图并将其分层连接，以组合多尺度激活，同时防止锐度损失。该网络由ResNet的每个块的单独的RefineNet模块组成。每个RefineNet模块由三个主要模块组成，即剩余卷积单元（RCU）、多分辨率融合（MRF）和链式剩余池（CRP）（参见图11）。RCU块由一个自适应卷积集组成，该卷积集微调用于分割问题的ResNet权重的预训练权重。MRF层使用卷积和上采样层融合不同分辨率的激活，以创建更高分辨率的地图。最后，在CRP层中，多个大小的内核被用于激活以从大图像区域捕获背景上

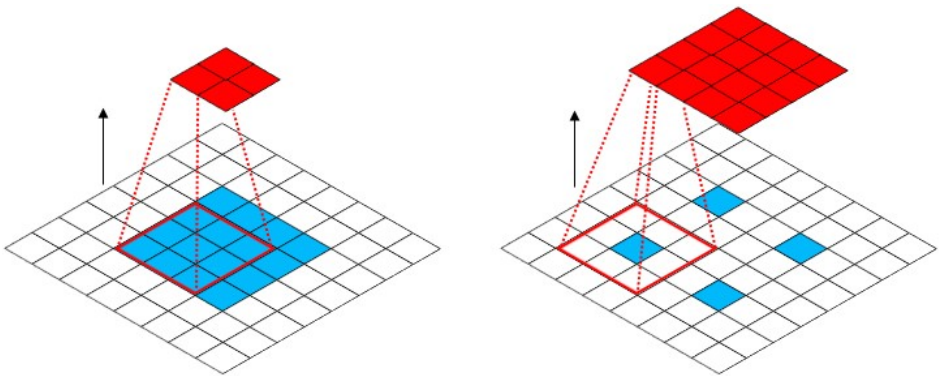
下文。RefineNet在Person-Part数据集上进行了测试，其IOU为68.6，而DeepLab-v2的IOU为64.9，两者都使用ResNet-101作为特征提取程序。



4.2 Convolutional autoencoders（卷积自动编码器）

最后一小节讨论用于执行像素级分类以处理图像分割问题的判别模型。另一种思路是从自动编码器中获得灵感。传统上，自动编码器用于从输入样本中提取特征，同时试图保留大部分原始信息。自动编码器基本上由将输入表示形式从原始输入编码为可能的较低维中间表示形式的编码器和尝试从中间表示形式重构原始输入的解码器组成。损失是根据原始输入图像和重建输出图像之间的差异来计算的。解码器部分的生成性经常被修改并用于图像分割。与传统的自动编码器不同，在分割过程中，损失是根据重建的像素级类分布和期望的像素级类分布之间的差异来计算的。与RCNN或DeepLab算法的分类方法相比，这种分割方法更像是一种生成过程。这种方法的问题在于防止在编码过程中过度抽象图像。这种方法的主要好处是能够产生更清晰的边界，而复杂程度要低得多。与分类方法不同，解码器的生成特性可以学习基于提取的特征创建精细的边界。影响这些算法的主要问题是抽象级别。人们已经看到，如果不进行适当的修改，缩小特征地图的大小会在重建过程中造成不一致。在卷积神经网络的范例中，编码基本上是一系列卷积和汇集层或跨步卷积。然而，重建可能会很棘手。从低维特征解码的常用技术是转置卷积或非冷却层。与常规卷积特征提取方法相比，使用基于自动编码器的方法的主要优点之一是可以自由选择输入大小。通过巧妙地使用下采样和上采样操作，可以输出与输入图像具有相同分辨率的像素级概率。这一优点使得具有多尺度特征转发的编码器-解码器体系结构在输入大小不是预先确定的并且需要与输入大小相同的输出的网络中变得无处不在。

转置卷积 转置卷积（也称为分数步卷积）被引入来逆转传统卷积操作的影响[156，53]。它通常被称为反卷积。然而，在信号处理中定义的反卷积与转置卷积在基本公式上是不同的，尽管它们有效地解决了相同的问题。在卷积运算中，输入的大小根据内核的填充量和跨距而变化。如图12所示，跨步2将创建激活次数为跨步1的一半。为了使转置卷积起作用，填充和步幅的控制方式应使大小变化相反。这是通过扩大输入空间来实现的。注意，与空洞卷积不同，空洞卷积的内核被放大，这里的输入空间被放大。（***注解：转置卷积作用就是实现上采用，上采用常采用的方法之一***）

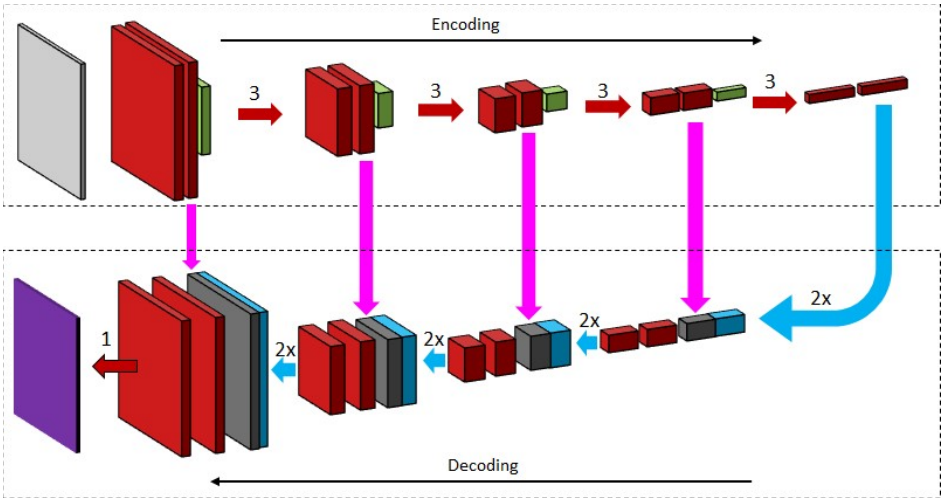


上池化 另一种减小激活大小的方法是通过池层。跨距为2的2×2池层将图像的高度和宽度减少了2倍。在这样的池层中，像素的2×2邻域被压缩为单个像素。不同类型的池以不同的方式执行压缩。最大池考虑4个像素中的最大激活值，而平均池取相同的平均值。对应的上池化将单个像素解压缩到2×2像素的邻域，以使图像的高度和宽度加倍。（***注解：简单来说，就是池化的反向操作***）

4.2.1 Skip Connections

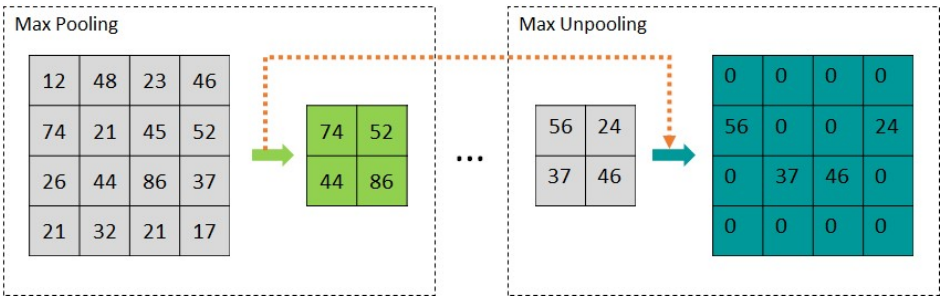
卷积神经网络中经常使用线性跳跃连接来改善大量层的梯度流[78]。随着网络深度的增加，激活图倾向于关注越来越多的抽象概念。Skip连接已经被证明是非常有效的，它可以将不同层次的抽象组合在一起，生成清晰的分割图。

U-NET 2015年提出的U-Net体系结构，被证明对诸如神经元结构分割、放射照相和细胞追踪挑战等各种问题相当有效[177]。该网络的特点是具有一系列卷积和最大池层的编码器。解码层包含镜像卷积序列和转置卷积序列。如前所述，它表现为一个传统的自动编码器。以前有人提到过抽象层次对图像分割质量的影响。为了考虑不同层次的抽象，U-Net实现跳过连接，以将未压缩的激活从编码块复制到解码块之间的镜像对应块，如图13所示。U-Net的特征抽取器也可以升级，以提供更好的分割图。绰号为“The 100 layers Tiramisu”[88]的网络应用了U-Net的概念，使用了基于密集网络的特征抽取器。其他现代变化包括胶囊网络的使用[183]以及局部约束路由[108]。U-Net被选为ISBI小区跟踪挑战赛的获胜者。在PhC-U373数据集中，平均IoU为0.9203，而第二好的为0.83。在DIC-HeLa数据集中，平均IoU为0.7756，明显好于次优方法（仅为0.46）。



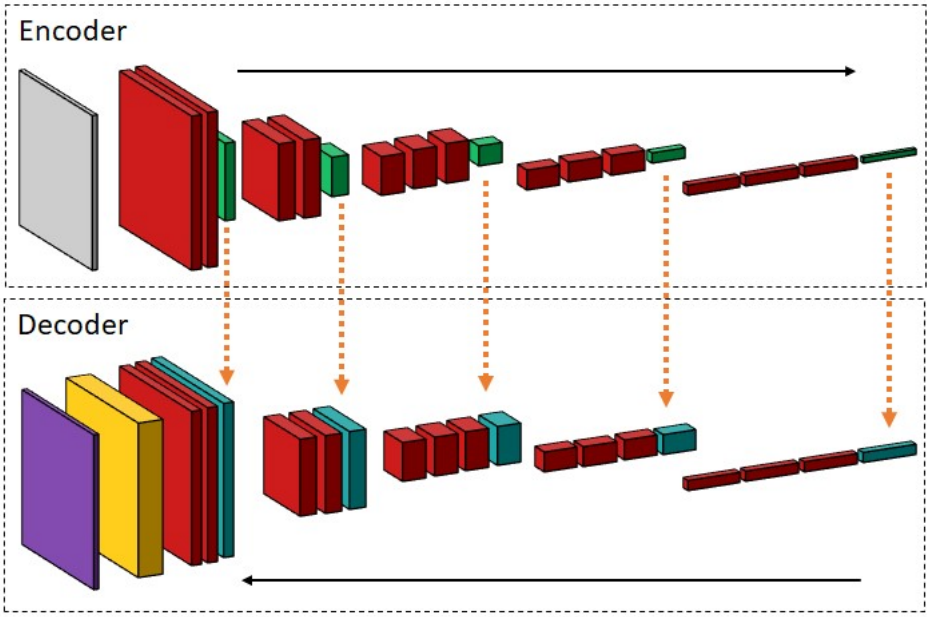
4.2.2 Forwarding pooling indices（前向池化索引）

由于各种原因，最大池一直是减少激活映射大小的最常用技术。激活表示图像区域对特定内核的响应。在最大池中，通过仅考虑在该区域内获得的最大响应，将像素区域压缩为单个值。如果典型的自动编码器在编码阶段将2×2像素的邻域压缩为单个像素，则解码器必须将像素解压缩为2×2的相似尺寸。通过转发池索引，网络在执行最大池时基本上记住4个像素中最大值的位置。与最大值对应的索引被转发到解码器（参见图14），以便在取消池操作的同时，可以将来自单个像素的值复制到下一层中2×2区域中的相应位置[215]。其余三个位置的值在随后的卷积层中计算。如果在不知道池索引的情况下将值复制到随机位置，则会出现分类不一致，尤其是在边界区域。



SegNet SegNet算法[9]于2015年推出，以在复杂的室内和室外图像上与FCN网络竞争。该体系结构由5个编码块和5个解码块组成。编码块遵循VGG-16网络特征抽取器的结构。每个块是一个多重卷积、批处理规范化和ReLU层的序列。每个编码块以存储索引的最大池层结束。每个解码块以使用保存的池索引的未冷却层开始（参见图15）。来自编码器中第i块的最大池层的索引被转发到解码器中第（L-i+1）块的最大未冷却层，其中L是每个编码器和解码器中块的总数。SegNet架构的mIoU为60.10，而DeepLab LargeFOV为53.88，FCN为49.83，

Deconvnet为59.77。

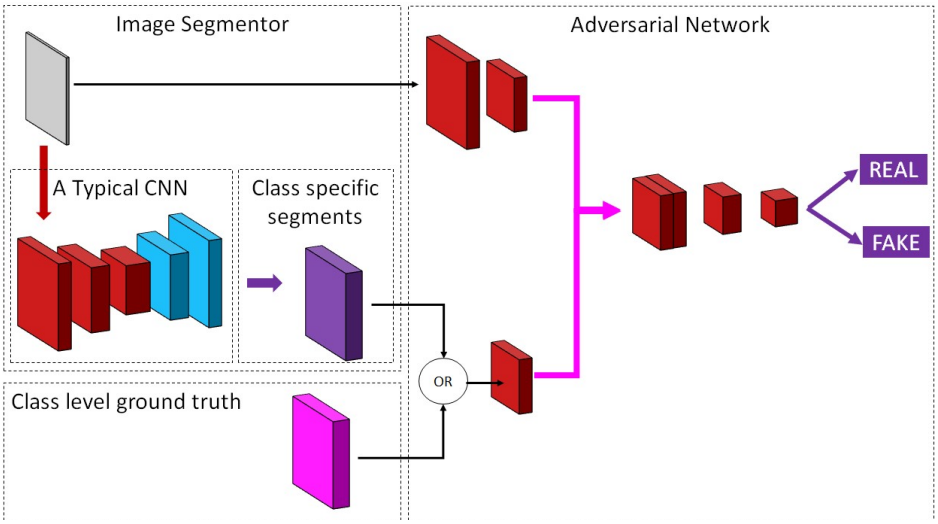


4.3 Adversarial Models（对抗模式）

到目前为止，我们已经看到了像FCN，DeepMask，DeepLab这样的纯粹区分模型，它主要为类的每个像素生成一个概率分布。此外，自动编码器将分割视为生成过程，但最后一层通常连接到像素级软最大分类器。对抗学习框架从不同的角度来处理优化问题。生成性对抗网络（GANs）作为一种具有显著性能的生成性网络，得到了广泛的应用。对抗学习框架主要由两个网络组成：生成网络和鉴别网络。生成器G尝试使用名为pz（z）的噪声输入先验分布从训练数据集中生成图像。网络G（z；θG）表示由具有权重θG的神经网络表示的可微函数。鉴别器网络尝试正确猜测输入数据是来自训练数据分布（pdata（x））还是由生成器G生成。鉴别器的目标是更好地捕捉假图像，而生成器试图更好地愚弄鉴别器，从而在生成更好输出的过程中。整个优化过程可以写成一个最小-最大问题，如下所示：

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z)))] \quad (3)$$

本文还从对抗学习的角度探讨了分词问题。分割网络被视为生成每个类的分割掩模的生成器，而鉴别器网络试图预测一组掩模是来自地面真值还是来自生成器的输出[133]。该过程的示意图如图20所示。此外，条件GANs被用于执行图像到图像的翻译[86]。该框架可用于图像的语义边界与输出分割图不一定重合的图像分割问题，例如，在创建建筑物的框架示意图的情况下。



4.4 Sequential Models（顺序模型）

到目前为止，几乎所有讨论的技术都涉及到语义图像分割。另一类分割问题，即实例级分割需要稍有不同的方法。与语义图像分割不同，这里将同一对象的所有实例分割为不同的类。这种类型的分割问题大多是作为一种学习来处理的，即给出一系列的对象段作为输出。因此，序贯模型在此类问题中发挥了作用。一些常用的主要体系结构是卷积LSTMs、递归网络、基于注意的模型等。

4.4.1 Recurrent Models（回归模型）

传统的LSTM网络采用全连接权值对序列输入下的长、短期记忆进行建模。但它们无法捕捉图像的空间信息。此外，图像的全连接权值在很大程度上增加了计算量。在卷积LSTM[176]中，这些权重被卷积层所代替（参见图）。卷积LSTMs已经在一些工作中用于实例级分割。通常它们被用作对象分割网络的后缀。像LSTM这样的递归模型的目的是在序列输出的不同时间戳中选择对象的每个实例。该方法已在FCN和U-NET等对象分割框架下实现[28]。（***注解：即在完成语义分割的基础上加上卷积LSTM模块作为后续实例分割任务的解决方案***）

4.4.2 Attention Models（注意力模型）

卷积LSTMs可以在不同的时间戳上选择不同的对象实例，而注意力模型则可以更好地控制单个实例的本地化过程。控制注意力的一个简单方法是空间抑制[176]。设计空间抑制网络以学习一个偏差参数，该参数将从将来的激活中切断先前检测到的片段。随着专用注意模块和外部存储器的引入，注意模型得到了进一步的发展。在文献[174]中，将实例分割网络划分为4个模块。首先，外部内存提供前面所有步骤中的对象边界细节。其次，盒子网络尝试预测对象的下一个实例的位置，并为第三模块（即分割模块）输出图像的子区域。分割模块类似于前面讨论的卷积自动编码器模型。第四个模块根据预测的片段是否符合对象的适当实例来评分。当分数低于用户定义的阈值时，网络终止。

4.5 Weakly Supervised or Unsupervised Models（弱监督和无监督模型）

神经网络通常采用反向传播等算法进行训练，其中参数w根据其局部偏导数（相对于使用损失函数f获得的误差值E）进行更新。

$$w = w + \Delta w = w - \eta \frac{\delta E}{\delta w}$$

损失函数通常用目标值和预测值之间的距离来表示。但在许多情况下，图像分割需要使用不带真实性注释的数据。这导致了无监督图像分割技术的发展。实现这一目标的直接方法之一是使用在具有类似样本和基本事实的其他较大数据集上预先训练的网络，并在特征地图上使用聚类算法（如K-means）。然而，这种半监督技术对于具有唯一样本空间分布的数据样本是低效的。另一个缺点是网络被训练成在与测试数据仍然不同的输入分布上执行。这不允许网络充分发挥其潜力。在完全无监督的分割算法中，关键问题是开发一种能够测量分割或像素簇质量的损失函数。由于这些限制，当涉及到弱监督或无监督的方法时，文献数量相对较少。

4.5.1 Weakly supervised algorithms（弱监督算法）

即使在缺乏适当的像素级注释的情况下，分割算法也可以利用较粗的注释（例如边界框，甚至图像级标签[161116]）来执行像素级分割。

利用边界框 从数据标注的角度来看，与像素级分割相比，定义边界框的代价要低得多。具有边界框的数据集的可用性也比具有像素级分段的数据集大得多。边界框可用作生成像素级别分割图的弱监督。在文献[42]中，BoxSup使用区域建议方法（如选择性搜索）生成分割建议。在此基础上，采用多尺度组合分组法对候选模板进行组合，目标是选择具有最高IOU的最优组合。该分割图用于调整传统的图像分割网络，如FCN。在帕斯卡VOC 2012测试集中，BoxSup能够达到75.1百万单位，而FCN为62.2，或DeepLab CRF为66.4。

4.5.2 Unsupervised Segmentation（无监督算法）

与有监督或弱监督分割不同，无监督图像分割算法的成功主要取决于学习机制。下面介绍一些常见的方法。

学习多重目标 在无监督学习中，一种最普遍的方法是考虑多个目标，这些目标的设计是为了在没有基本事实的情况下能够很好地进行分割。

一个常见的变体称为JULE或联合无监督学习的深度表示已被用于一些应用中，其中有一个与地面真实性缺乏样本。JULE的基础在于训练序列模型和深层特征提取模型。该方法主要是作为一种图像聚类算法引入的。然而，它已经扩展到其他应用，如图像分割。能够进行这些分割的关键目标是开发适当的目标函数。在JULE中，目标函数考虑了簇内样本之间的亲和力，以及簇与其邻域之间的负亲和力。聚集聚类是在递归网络的时间戳上进行的。在[149]的著作中，JULE尝试的是图像补丁，而不是整个图像样本。通过一个类似的目标函数，它能够将补丁划分为预先定义的类。在这种情况下，JULE被用来为下一次迭代提供聚集聚类信息作为监控信号。

另一种学习多个目标的方法是在具有独立工作的神经网络之间进行对抗性协作[172]，如单眼深度预测、估计摄像机运动、检测光流和将视频分割到静态场景和运动区域。通过竞争性协作，每个网络竞相解释属于静态类或移动类的相同像素，然后这些像素与调节器共享其学习到的概念以执行运动分割。

使用优化模块进行自我监控 使用其他无监督过分割技术可用于监督深层特征提取器[92]。通过强制多个约束，如特征之间的相似性、空间连续性、轴内规范化。所有这些目标都是通过反向传播优化的。通过使用SLIC [1]之类的标准算法从图像中提取超像素来实现空间连续性，并且超像素中的所有像素都必须具有相同的标签。利用两个分割图之间的差异作为监督信号来更新权值。

其他相关的无监督语义信息提取技术： 没有注释的学习总是很有挑战性的，因此有很多文献提出了许多有趣的解决方案。使用CNNs来解决从图像中衍生出来的拼图问题[157]可以用来学习对象各个部分之间的语义连接。提出的上下文无关网络以一组图像块为输入，试图建立它们之间正确的空间关系。在这个过程中，它同时学习特定于对象部分的特征及其语义关系。这些基于补丁的自我监控技术可以使用上下文信息进一步改进[152]。使用上下文编码器【164】还可以导出图像的各个部分之间的空间和语义关系。上下文编码器基本上是，cnn训练来生成图像的任意区域，该区域受其周围信息的约束。另一个提取语义信息的例子可以在图像着色过程中找到[218]。彩色化过程需要像素级的理解对象对应的语义边界。其他的自我监控技术可以利用视频中的运动线索来分割对象的各个

部分，以便更好地理解语义[216]。该方法可以学习语义分割、人的句法分析、实例分割等任务的结构特征和连贯特征。

4.5.3 W-Net

W-Net[207]的灵感来源于之前讨论过的U-Net。W-Net体系结构由两个级联的U-Net组成。第一U-Net充当编码器，将图像转换为其分段版本，而第二U-Net尝试从第一U-Net的输出（即分段图像）重建原始图像。两个损失函数同时最小化。其中之一是由第二U网给出的输入图像和重建图像之间的均方误差。第二个损失函数来自标准化割[186]。硬标准化切割公式为，

$$Ncut_K(V) = \sum_{k=1}^{K} \frac{\sum_{u \in A_k, v \in V - A_k} w(u,v)}{\sum_{u \in A_k, t \in V} w(u,t)}$$

其中，k段中的Ak是像素集，V是所有像素的集，w测量两个像素之间的权重。
但是，此功能不可区分，因此无法进行反向传播。因此，提出了一种函数的软版本。

$$J_{soft-Ncut}(V,K) = K - \sum_{k=1}^{K} \frac{\sum_{u \in V} p(u = A_k) \sum_{v \in V} w(u,v) p(v = A_k)}{\sum_{u \in V} p(u = A_k) \sum_{t \in V} w(u,t)}$$

其中，p（u=Ak）表示节点u属于Ak类的概率。利用全连通条件随机场进一步细化了输出分割图。剩下的不重要的片段使用层次聚类进一步合并。

4.6 Interactive Segmentation（交互式分割）

图像分割是计算机视觉领域中最困难的挑战之一。在许多图像过于复杂、噪声大或光照条件差的情况下，用户的一点交互和引导可以显著提高分割算法的性能。即使是在深度学习范式之外，交互式分割也在蓬勃发展。然而，卷积神经网络具有强大的特征提取能力，可以减少交互量，达到一定程度。

4.6.1 Two stream fusion（双流融合）

交互式分割最直接的实现之一是具有两个并行分支，一个来自表示交互式流的图像，另一个来自图像，并将它们融合以执行分割[83]。交互输入以表示正类和负类的不同颜色点的形式出现。通过一些后处理，其中交互作用图的强度是根据与点之间的欧几里德距离计算的，我们得到两组图（每类一个），它们看起来像以点为中心的模糊voronoi单元。将映射按元素相乘以获得交互流的图像。这两个分支由卷积序列和池层组成。在每个分支的末端获得特征的Hadamard积被发送到融合网络，在融合网络中生成低分辨率分割图。另一种方法是跟随FCN的足迹，融合多尺度特征，得到更清晰的分辨率图像。

4.6.2 Deep Extreme Cut（深度极大值切割）

与双流方法相反，deep extreme cut[138]使用一条管道从RGB图像创建分割图。此方法要求用户提供4个点，表示对象边界中的四个极端区域（最左边、最右边、最上面、最下面）。通过从这些点创建热图，一个4通道输入被输入到DenseNet101网络。网络的最终特征映射被传递到金字塔场景解析模块中，用于分析全局上下文以执行最终的分割。该方法在PASCAL测试集上可以得到80.3的mIOU。

4.6.3 Polygon-RNN（多边形RNN）

Polygon RNN[26]采用与其他方法不同的方法。从典型VGG网络的不同层次提取多尺度特征，并将这些特征串接起来，为递归网络创建特征块。反过来，RNN应该提供一个点序列作为表示对象轮廓的输出。该系统主要设计为交互式图像标注工具。用户可以用两种不同的方式进行交互。首先，用户必须为感兴趣的对象提供一个紧边界框。其次，在建立多边形之后，允许用户编辑多边形中的任何点。但是，这种编辑不用于对系统的任何进一步培训，因此为改进系统提供了一条小途径。

4.7 Building more efficient networks（建立更有效的网络）

虽然许多复杂的网络具有许多奇特的模块，可以提供非常好的语义分割质量，但将这样的算法嵌入到现实系统中则是另一回事。许多其他因素，如硬件成本、实时响应等，都带来了新的挑战。效率也是创建消费者级系统的关键。

4.7.1 ENet

ENet[163]提出了一些有趣的设计选择，以创建一个具有少量参数（37万）的非常浅的分割网络。它不是像SegNet或U-Net那样的对称编码器-解码器架构，而是具有更深的编码器和更浅的解码器。不是在池后增加通道大小，而是并行池操作与stride 2的卷积一起执行，以减少总体特征。为了提高学习能力，与ReLU相比，PReLU被使用，使得传递函数保持动态，以便它可以模拟ReLU的作业以及所需的身份函数。这通常是ResNet中的一个重要因素，但是由于网络很浅，使用PReLU是一个更明智的选择。除此之外，使用因子化过滤器还允许使用较少数量的参数。

4.7.2 Deep Layer Cascade（深层级联）

深层级联[116]解决了几个挑战，并作出了两个重大贡献。首先，分析了不同类别像素级分割的难度。在级联网络中，较容易的片段在早期被发现，而后一层则集中在需要更精细片段的区域。其次，提出的层级联可以与Inception-ResNet-V2（IRNet）等常用网络一起使用，在一定程度上提高了速度甚至性能。IRNet的基本原理是创建一个多级管道，在每个阶段中，一定数量的像素将被分类到一个段中。在早期阶段，最简单的像素将被分

类，而不确定性较大的较难像素将进入后期阶段。在随后的阶段中，卷积将仅发生在那些在前一阶段无法分类的像素上，同时将更难的像素转发到下一阶段。通常，提出的模型分为三个阶段，每个阶段向网络添加更多的卷积模块。在分层级联的情况下，VOC12测试集的mIOU为82.7，DeepLabV2和Deep解析网络是最接近的竞争对手，mIOU分别为79.7和77.5。在速度方面，每秒处理23.6帧，而SegNet为14.6 fps，DeepLab-V2为7.1 fps。

4.7.3 SegFast

另一个名为SegFast[159]的最新实现能够构建一个只有60万个参数的网络，这使得一个网络可以在大约0.38秒的时间内完成前向传递，而无需GPU。该方法将深度可分离卷积的概念与SqueezeNet的火模相结合。SqueezeNet引入了fire模块的概念，以减少卷积权重的数量。通过深度可分卷积，参数的数量进一步下降。他们还提出了使用可分离的差分转置卷积进行解码。即使有如此多的功能缩减方法，性能也相当于其他流行的网络，如SegNet。

4.7.4 Segmentation using superpixels（使用超混合进行分割）

过分割算法[1]在基于局部信息将图像分割成小块上得到了很好的发展。使用补丁分类算法，这些超混合可以转换为语义段。然而，由于过分割过程不考虑邻域关系，因此有必要将其纳入到斑块分类算法中。与像素级分类相比，执行面片分类要快得多，因为超像素的数量远小于图像中像素的数量。Farabet等人首次将CNNs用于超混合水平分类。[55]。然而，仅仅考虑没有上下文的超混合会导致错误的分类。在文献[46]的工作中，在补丁分类过程中，通过考虑不同级别的邻域超像素来捕获多个级别的上下文。通过使用加权平均、最大投票或基于不确定性的方法（如dempster-shafer理论）融合不同层次上下文的块级概率，提出了一种非常有效的分割算法。达斯等人的作品。与Farabet等人的74.56%相比，能够获得77.14%[46]的像素级精度。[55]。超混合可以用来建立有效的语义分割模型，反之亦然。相反，语义分割的基本事实可以用来训练网络执行过度分割[197]。利用卷积特征进行过分割，同时特别注意语义边界，可以计算出像素的相似度。

5 Applications

图像分割是计算机视觉领域中最常解决的问题之一。它经常被其他相关任务如目标检测、目标识别、场景分析、图像描述生成等所增强。因此，这一分支的研究发现在各种现实生活场景中有着广泛的应用。

5.1 Content-based image retrieval (CBIR)（基于内容的图像检索）

随着internet上结构化和非结构化数据的不断增加，开发高效的信息检索系统显得尤为重要。因此，CBIR系统一直是一个利润丰厚的研究领域。许多其他相关问题，如可视化问答、基于交互式查询的图像处理、描述生成等。图像分割在许多情况下是有用的，因为它们代表了不同对象之间的空间关系[12，127]。实例级分段对于处理数值查询至关重要[217]。无监督方法[90]对于处理大量的非注释数据特别有用，这在这个工作领域非常常见。

5.2 Medical imaging（医学成像）

图像分割的另一个主要应用领域是医疗保健领域。许多诊断程序都涉及到处理对应于不同类型成像源和身体不同部位的图像。一些最常见的任务类型是分割有机元素，如血管[58]、组织[91]、神经[132]等。其他类型的问题包括异常的定位，如肿瘤[224145]，动脉瘤[48131]等等。显微图像[85]还需要各种各样的分割，如细胞或细胞核检测、细胞计数、细胞结构分析等。这个领域面临的主要挑战是缺乏大量的数据来应对具有挑战性的疾病，由于所涉及的成像设备类型不同，图像质量也不同。医疗程序不仅涉及到人类，也涉及到其他动植物。

5.3 Object Detection（目标检测）

随着深度学习算法的成功，与自动目标检测相关的研究领域也出现了激增。机器人机动性[114]、自主驾驶[196]、智能运动检测[192]、跟踪系统[204]等应用。在智能机器人自主决策的帮助下，深海[190，106]或太空[181]等极为偏远的地区可以得到有效的探索。在国防等部门，无人驾驶飞行器或无人机[154]被用来探测偏远地区的异常或威胁[119]。分割算法在卫星图像的各种地理统计分析中有着重要的应用[109]。在图像或视频后期制作等领域，对诸如图像抠像[115、115]，合成[24]和rotoscoping [2]之类的各种任务执行分割通常是必不可少的。

5.4 Forensics

虹膜[125，65]、指纹[94]、手指静脉[170]、牙科记录[93]等生物特征验证系统涉及各种信息区域的分割，以便进行有效的分析。

5.5 Surveillance（监督）

监视系统[147，95，89]与各种问题有关，如遮挡，照明或天气状况。此外，监视系统还可以对来自高光谱源的图像进行分析[4]。监视系统还可以扩展到各种应用，如目标跟踪[82]、搜索[3]、异常检测[173]、威胁检测[137]、交通控制[208]等。图像分割对于将感兴趣的对象与自然场景中的杂波分离起着至关重要的作用。

6 Discussion and Future Scope（讨论和未来展望）

本文讨论了各种方法，强调了它们的主要贡献、优缺点。有这么多不同的选择，仍然很难为一个问题选择正确的方法。选择正确算法的最佳方法是首先分析影响选择的变量。

影响基于深度学习方法性能的一个重要方面是数据集和注释的可用性。在这方面，表1提供了属于不同领域的数据集的简明清单。在处理其他小规模数据集时，通常在类似域的较大数据集上预先训练网络。有时可能有大量的样本可用，但像素级分割标签可能不可用，因为创建它们是一个累赘的问题。即使在这些情况下，对网络的其他相关问题（如分类或定位）进行预训练也有助于学习一组更好的权重。

在这方面，必须作出的一个相关决定是在有监督、无监督或弱监督算法中进行选择。在当前的情况下，存在着大

量的有监督方法，但是无监督和弱监督算法仍然远远没有达到饱和水平。这是图像分割领域的一个合理的关注点，因为数据收集可以通过许多自动化的过程来进行，但是对它们进行完美的注释需要手动操作。这是研究人员可以在构建端到端可扩展系统方面做出贡献的最突出的领域之一，该系统可以对数据分布进行建模，确定最佳的类数并在完全不受监督的域中创建准确的像素级分割图 。弱监督算法也是一个要求很高的领域。收集与分类或本地化等问题对应的注释要容易得多。利用这些标注来指导图像分割也是一个很有前途的领域。
建立用于图像分割的深度学习模型的下一个重要方面是选择合适的方法。经过预训练的分类器可以用于各种完全卷积的方法。在大多数情况下，通过将网络中不同深度的信息进行融合，可以实现多尺度的特征融合。预先训练的分类器，如VGGNet、ResNet或DenseNet，也经常用于编码器体系结构的编码器部分。这里，还可以将信息从各个编码器层传递到解码器的相应大小相似的层以获得多尺度信息。编码器-解码器结构的另一个主要优点是，如果仔细设计下采样和上采样操作，则可以生成与输入大小相同的输出。与FCN或DeepMask等简单卷积方法相比，这是一个主要的优点。这消除了对输入大小的依赖，因此使系统更具可伸缩性。这两种方法在语义分割问题中最为常见。但是，如果需要更精细级别的实例特定段，则通常需要与与对象检测对应的其他方法相结合。利用边界盒信息是解决这些问题的一种方法，而其他方法使用基于注意力的模型或递归模型为对象的每个实例提供作为片段序列的输出。
在测量系统性能时，可以考虑两个方面。一个是速度，另一个是准确度。条件随机场是最常用的后处理模块之一，用于优化其他网络的输出。CRFs可以模拟为RNN，创建端到端可训练模块，以提供非常精确的分割图。其他细化策略包括使用过分割算法（如超级像素）或使用人类交互来指导分割算法。在速度增益方面，可以使用深度可分卷积、核分解、减少空间卷积等策略来实现网络的高度压缩。这些策略可以在很大程度上减少参数的数量，而不会使性能降低太多。最近，生成性对抗网络的受欢迎程度急剧上升。然而，它们在分割领域的应用还相当少，只有少数几种方法可以解决这个问题。考虑到他们所取得的成功，它肯定有潜力大幅度改善现有的系统。
图像分割的未来在很大程度上取决于可用数据的质量和数量。虽然互联网上有大量的非结构化数据，但缺乏准确的注释是一个值得关注的问题。尤其是像素级的注释，如果没有人工干预，很难获得。最理想的方案是利用数据分发本身来分析和提取表示概念而不是内容的有意义的片段。这是一项非常艰巨的任务，尤其是当我们处理大量非结构化数据时。关键是将数据分布的表示映射到问题陈述的意图，这样派生的段在某种程度上是有意义的，并且有助于系统的总体目的。

7 Conclusion

图像分割出现了一种新的基于深度学习的算法。从基于深度学习的算法的发展开始，我们已经彻底解释了与基于深度学习的图像分割相关的各种最新算法的优缺点。简单的解释使读者能够掌握最基本的概念，这些概念有助于基于深度学习的图像分割算法的成功。图中遵循的统一表示方案可以突出各种算法的异同。在今后的理论调查工作中，可以结合实证分析的方法进行探讨。

Supplementary Information（补充资料）

深度学习前的图像分割

阈值：最直接的图像分割可以是基于相对于强度值的阈值为每个图像分配类标签。最早的算法之一，通常被称为Otsu的[199]方法，选择一个关于最大方差的阈值点。许多现代方法被应用于模糊逻辑[39]或非线性阈值[193]。早期的方法主要集中在二值化阈值上，而在随后的几年中也出现了多类分割。
聚类方法： 像K-means[194]这样的聚类方法可以将图像聚类成多个类。这是一个非常简单的过程，对于感兴趣的对象相对于背景的对比度很高的图像，可以产生很好的效果。其他的聚类方法结合了模糊逻辑[16150]甚至多目标优化[10]。
基于直方图的方法： 基于直方图的方法[195193]在语义分割方面提供了一个更加全局的视角。通过分析直方图的波峰和波谷，可以将图像适当地分割成若干个最佳的分割段。不像k-means这样的聚类算法，不需要事先知道聚类的数量。
边缘检测 研究图像分割问题的另一个角度是考虑对象之间的语义边界[184]。语义图像分割与边缘检测算法有着密切的关系，其原因有很多，如图像中的单个对象往往被强度梯度发生急剧变化的边缘分割。利用边缘检测和语义分割概念的一种常见方法是基于超像素的处理[45，38]。
区域成长的方法： 虽然基于强度的方法在图像聚类中非常有效，但它们没有考虑局部性因素。区域生长方法依赖于假设公共区域内的相邻像素具有一些共同的性质。这类方法通常从种子点开始，在语义边界内缓慢增长[59]区域是通过基于区域内方差或能量合并相邻较小区域而增长的。许多常见的算法，如Mumford Shah[151]或Snakes算法[96]这些方法的其他变体依赖于lambda连通性并基于像素强度增长
基于图的方法： 通过将像素或像素组视为节点，从而将图像转换为加权无向图，可以使用图分割算法来考虑局部性的上下文。图切割算法可有效地用于获取段。概率图模型，如马尔可夫随机场（MRF）[80]可用于创建基于先验概率分布的像素标记方案。MRF试图最大化基于一组特征对像素或区域进行正确分类的概率，概率图形模型，如MRF或其他类似的基于图形的方法[56]也可以被视为能量最小化问题[47]。在这方面，模拟退火[124]也可以是一个恰当的例子。这些方法可以根据图的能量来选择划分图。
流域转换 分水岭算法[73]假设图像的梯度为地形表面。类似于这样一个表面中的水流线，具有最高梯度的像素充当分割的轮廓
基于特征的技术 颜色、纹理、形状、梯度等各种特征可用于训练机器学习算法，如神经网络[35、27、105]或支持向量机，以执行像素级分类。在深度学习开始之前，全连接神经网络被有效地用于语义分割。然而，完全连接的网络会对较大的图像产生巨大的内存挑战，因为每个层都伴随着O（n2）的可训练权重，其中n是每个输入图像的最终激活图的高度和权重。

神经网络简史

利用McCulloch和Pitts提出的人工神经元的初始命题和感知器的后续模型，可以模拟具有可学习权重的线性模型。但线性系统的局限性很快就显现出来了，因为人们发现它们无法学习具有因变量的情况，比如在异或问题中[162]。自从引入Neocognitron（一种分层的自组织神经结构）以来，几乎十年后，多层模型才开始出现。然而，问题是扩展了基于随机梯度的多层学习的思想。这就是反向传播的概念出现的时候。通过反向传播，可见层的误差可以作为偏导数链传播到中间层的权重。引入非线性激活函数（如sigmoid单元）可以使这些中间梯度不间断地流动，并解决异或问题。虽然很明显，更深的网络提供了更好的学习能力，但它们也会导致梯度消失或爆炸[14]。这是一个很难处理的问题，特别是在序列网络中，直到长短期记忆细胞[81]被提出取代传统的递归神经网络[143]。卷积神经网络[110]作为端到端分类器被引入，成为现代计算机视觉领域最重要的贡献之一。在接下来的十年里，随着辛顿提出了受限的玻尔兹曼机器（boltzmann machines）[13]和深度信念网络（deep-believe networks）[79]，深度学习开始蓬勃发展。

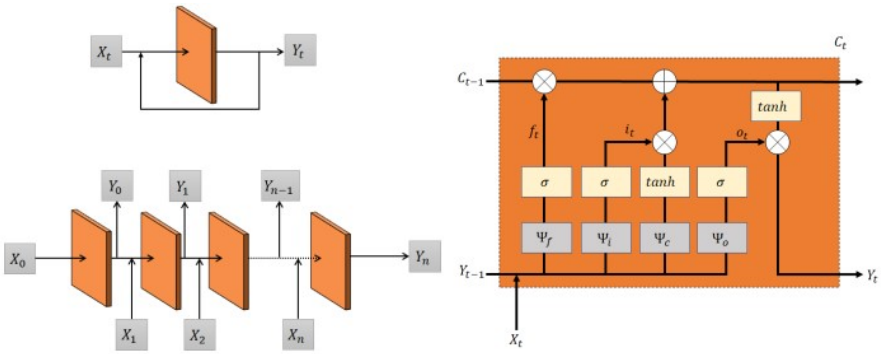
硬件方面的进展 触发深度学习开始的另一个重要因素是并行计算能力的可用性。早期对基于CPU的体系结构的依赖为大量浮点操作造成了瓶颈。集群是一种成本高昂的东西，因此研究在有大量资金的组织中相当本地化。但随着Nvidia开发的图形处理单元（gpu）[7]的出现，其用于访问并行计算核的底层cudapi神经学习系统得到了显著的提升。这些图形卡以更便宜的速度提供了成百上千的计算核心，这些核心被有效地构建来处理基于矩阵的操作，这是神经网络的理想选择。

较大的数据集 随着基于GPUs神经网络研究的有效性越来越广泛。但另一个重要的推动因素是新的和具有挑战性的数据集和挑战的涌入。从MNIST数据集[111]开始，数字识别成为新系统的一个常见挑战。上世纪90年代末，Yann LeCun、Yoshua Bengio和Geoffrey Hinton在加拿大高等研究院（Canadian institute of advanced research）推出了用于对象分类的自然对象CIFAR数据集[103]，从而使深入学习得以继续。随着名为Imagenet的大型视觉识别挑战赛的开展，这项挑战进一步扩展到了1000级Imagenet数据库[50]的新高度。到目前为止，这一直是作为任何新的对象分类算法基准的主要挑战之一。随着人们转向更复杂的数据集，图像分割也变得具有挑战性。PASCAL-VOC图像分割[54]、ILSVRC场景分析[15]、DAVIS Challenge[168]、Microsoft-COCO[122]等许多数据集和竞争对手的出现，也推动了图像分割的研究。

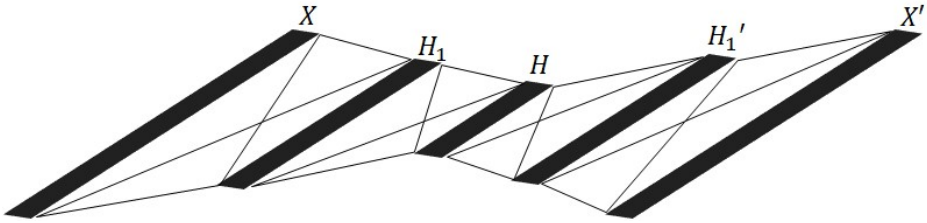
典型深度学习模式的大类

随着深度学习的出现，许多有趣的神经网络被提出来解决各种挑战

序列模型： 深网络的最早问题是递归神经网络[143]。递归网络的特点是反馈回路允许它们接受一系列输入，同时在每个时间步上携带学习到的信息。然而，在长的输入链上，可以看到信息随着时间的推移由于梯度消失而丢失。消失或爆炸梯度主要是由于偏导数的长乘法链可以将结果值推到几乎为零或很大的值，进而导致不重要或太大的权重更新[14]。解决这一问题的第一次尝试是由长-短记忆结构[81]提出的，在这种结构中，相关信息可以通过公路通道进行长距离传播，公路通道仅受加法或减法的影响，因此保留了梯度值。序列模型在计算机视觉中有着广泛的应用，如视频处理、实例分割等。图17示出了典型递归网络的一般和展开版本的示例。

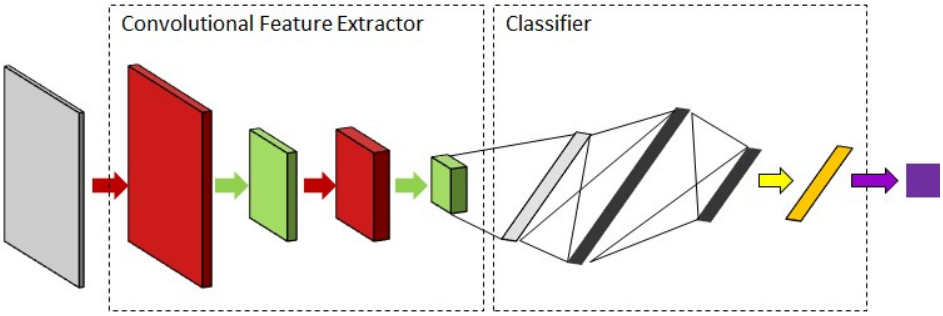


自编码器： 自多层感知器引入自联想网络[102]以来，自动编码器就一直存在。自动编码器的原理是将原始输入编码成一个潜在的表示。解码器网络试图从编码的表示重建输入。基于输入与重构输出之差的损失函数最小化保证了中间表示中的信息损失最小。这个隐藏的表示是实际输入的压缩形式。由于它保留了输入图像的大多数定义属性，因此通常用作进一步处理的特征。自动编码器由两个主要阶段组成，即编码和解码阶段。经过训练，编码器可以很容易地用作特征提取器。解码器部分可用于生成目的。许多作品将解码器的生成特性用于各种图像分割应用[98]。下图（18）显示了具有完全连接的线性层的自动编码器的表示。

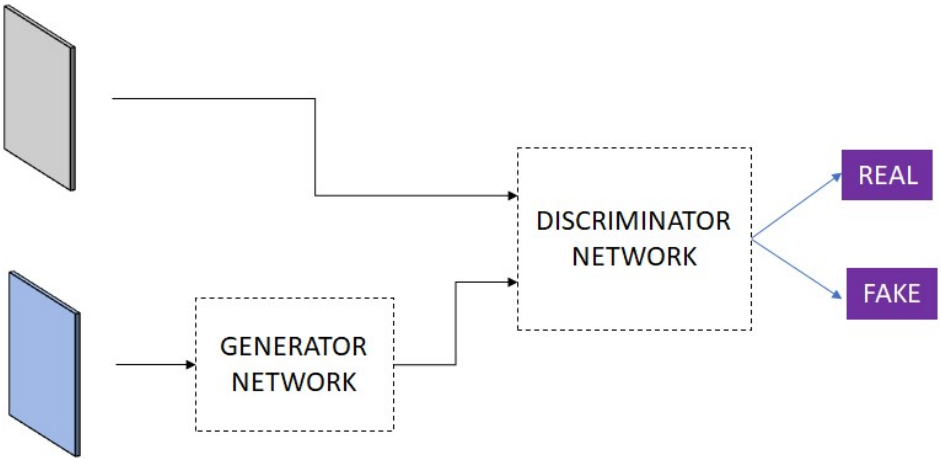


卷积神经网络： 卷积神经网络[110104]可能是计算机视觉深度学习的重要发明之一。卷积核常被用于复杂图像的特征提取。然而，设计内核并不是一件容易的事情，特别是对于像自然图像这样的复杂数据。利用卷积神经网络，基于交叉熵或均方误差等误差函数，可以通过反向传播随机初始化和迭代更新核函数。CNN中常见的其他操作有池、批处理规范化、激活、剩余连接等。池层增加卷积核的感受野。批处理规范化[84]是指涉及跨批处理的

激活规范化的泛化过程。激活函数是感知器学习的重要组成部分。自AlexNet[104]引入以来，整流线性单元（ReLU）[153]一直是选择的激活函数。ReLU（校正线性单元）提供0或1的梯度，因此，可以防止梯度消失或爆炸，还可以导致激活的稀疏性。最近，另一种有趣的梯度增强方法出现在应用程序剩余连接中。剩余连接[78]提供了梯度流动的替代路径，该路径没有抑制梯度的操作。残差连接在许多情况下也被用来提高分割图像的质量。图19示出了卷积特征提取器和全连接分类器。



生成模型 生成模型可能是计算机视觉中深度学习的最新吸引力之一。当序列模型如长-短期记忆或门控递归单元能够生成矢量元素序列时，由于空间的复杂性，在计算机视觉中要困难得多。最近，各种方法，如变分自动编码器[98]或对抗性学习[136，71]在生成复杂图像方面变得非常有效。生成性可以非常有效地用于生成分割遮罩等任务。图20示出了通过对抗性学习来学习的生成性网络的典型示例。



分类: 论文阅读

标签: 综述, 图像分割论文

青衣素
关注 - 0
粉丝 - 0
+加关注

0 0

« 上一篇： kaggle——predict futures sales
» 下一篇： 《推荐系统实践》笔记——第2章

posted @ 2020-04-25 14:56 青衣素 阅读(2194) 评论(0) 编辑 收藏 举报

编辑推荐：

- 肢体识别与应用
- .NET 排序 Array.Sort 实现分析
- .Net 微服务实战之可观测性
- 使用 three.js 实现炫酷的酸性风格 3D 页面
- 一个故事看懂 CPU 的 TLB

最新新闻：

- 没想到苹果才是最赚钱的游戏公司！比索尼微软任天堂加起来还多（2021-10-04 18:52）
- 网络巨头的AI大战，百度，阿里，腾讯你们累吗？（2021-10-04 18:51）
- 监视员工闹上法庭！谷歌开除员工引争议，说好的「不作恶」？（2021-10-04 18:44）
- 云基础设施支出减少 2.4%：公有云下降 6.1%，私有云增长 7.8%（2021-10-04 10:16）
- 贝佐斯遭蓝色起源员工倒戈：为追赶马斯克急功近利，安全措施不充分也敢载人上天（2021-10-04 10:09）

» 更多新闻...