

## 【总结】图像语义分割之FCN和CRF



ycszen · 1 年前

知



首发于  
智能单元

写文章

登录

(呕血制作啊！)前几天刚好做了个图像语义分割的汇报，把最近看的论文和一些想法讲了一下。所以今天就把它总结成文章啦，方便大家一起讨论讨论。本文只是展示了一些比较经典和自己觉得比较不错的结构，毕竟这方面还是有挺多的结构方法了。

## 介绍

**图像语义分割**，简单而言就是给定一张图片，对图片上的每一个像素点分类

从图像上来看，就是我们需要将实际的场景图分割成下面的分割图：



知

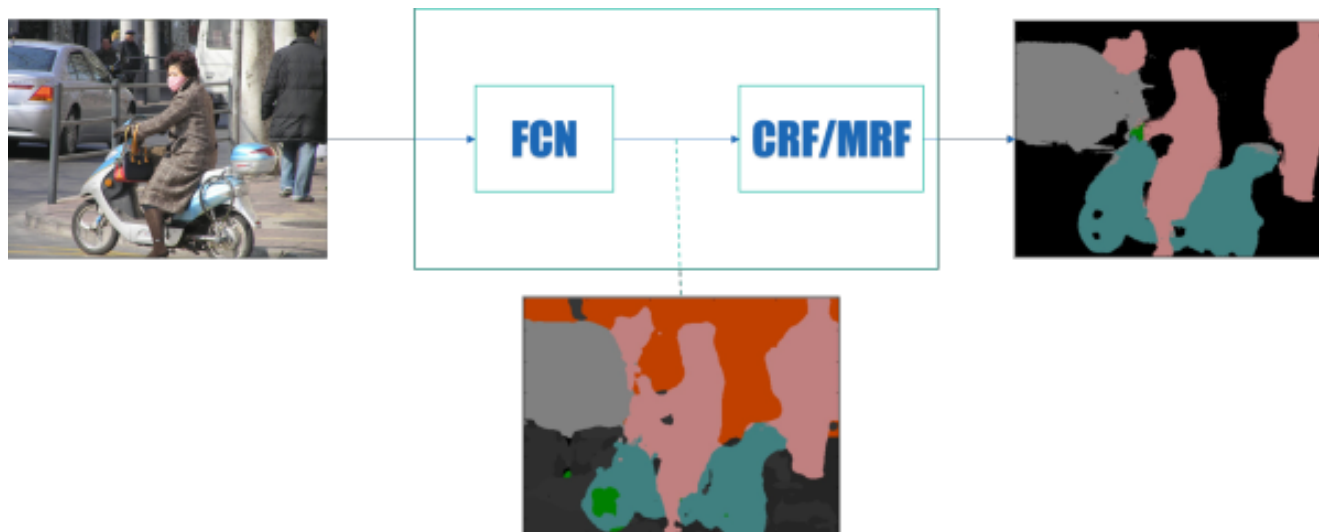


首发于  
智能单元

写文章

登录

不同颜色代表不同类别。经过我阅读“大量”论文（羞涩）和查看[PASCAL VOC Challenge performance evaluation server](#)，我发现图像语义分割从深度学习引入这个任务（FCN）到现在而言，一个通用的框架已经大概确定了。即：



- FCN-全卷积网络
- CRF-条件随机场
- MRF-马尔科夫随机场

前端使用FCN进行特征粗提取，后端使用CRF/MRF优化前端的输出，最后得到分割图。

接下来，我会从前端和后端两部分进行总结。

## 为什么需要FCN？

我们分类使用的网络通常会在最后连接几层全连接层，它会将原来二维的矩阵（图片）压扁成一维的，从而丢失了空间信息，最后训练输出一个标量，这就是我们的分类标签。

而图像语义分割的输出需要是个分割图，且不论尺寸大小，但是至少是二维的。所以，我们需要丢弃全连接层，换上全卷积层，而这就是全卷积网络了。具体定义请参看论文：[Fully Convolutional Networks for Semantic Segmentation](#)

## 前端结构

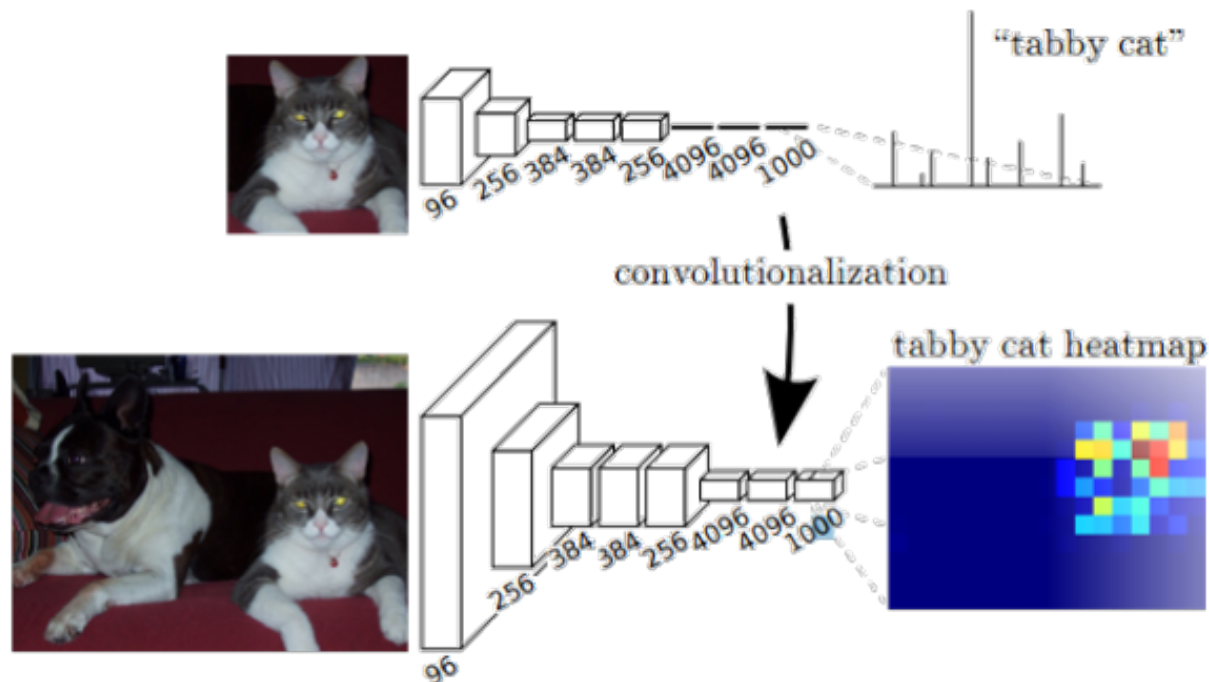
### FCN

此处的FCN特指[Fully Convolutional Networks for Semantic Segmentation](#)论文中提出的结构，而非广义的全卷积网络。

作者的FCN主要使用了三种技术：

- 卷积化（Convolutional）
- 上采样（Upsample）
- 跳跃结构（Skip Layer）

### 卷积化

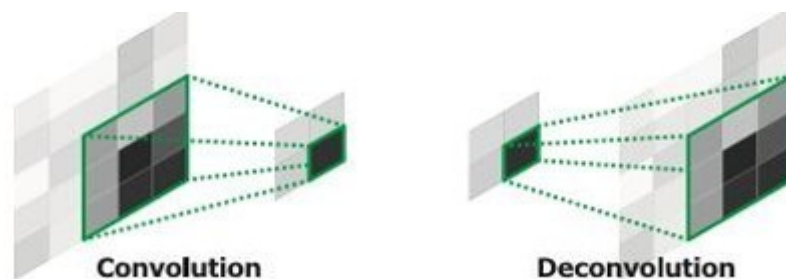


## 上采样

此处的上采样即是反卷积（Deconvolution）。当然关于这个名字不同框架不同，Caffe和Kera里叫Deconvolution，而tensorflow里叫conv\_transpose。CS231n这门课中说，叫conv\_transpose更为合适。

众所周知，普通的池化（为什么这儿是普通的池化请看后文）会缩小图片的尺寸，比如VGG16 五次池化后图片被缩小了32倍。为了得到和原图等大的分割图，我们需要上采样/反卷积。

反卷积和卷积类似，都是相乘相加的运算。只不过后者是多对一，前者是一对多。而反卷积的前向和后向传播，只用颠倒卷积的前后向传播即可。所以无论优化还是后向传播算法都是没有问题。图解如下：



但是，虽然文中说是可学习的反卷积，但是作者实际代码并没有让它学习，可能正是因为这个一对多的逻辑关系。代码如下：

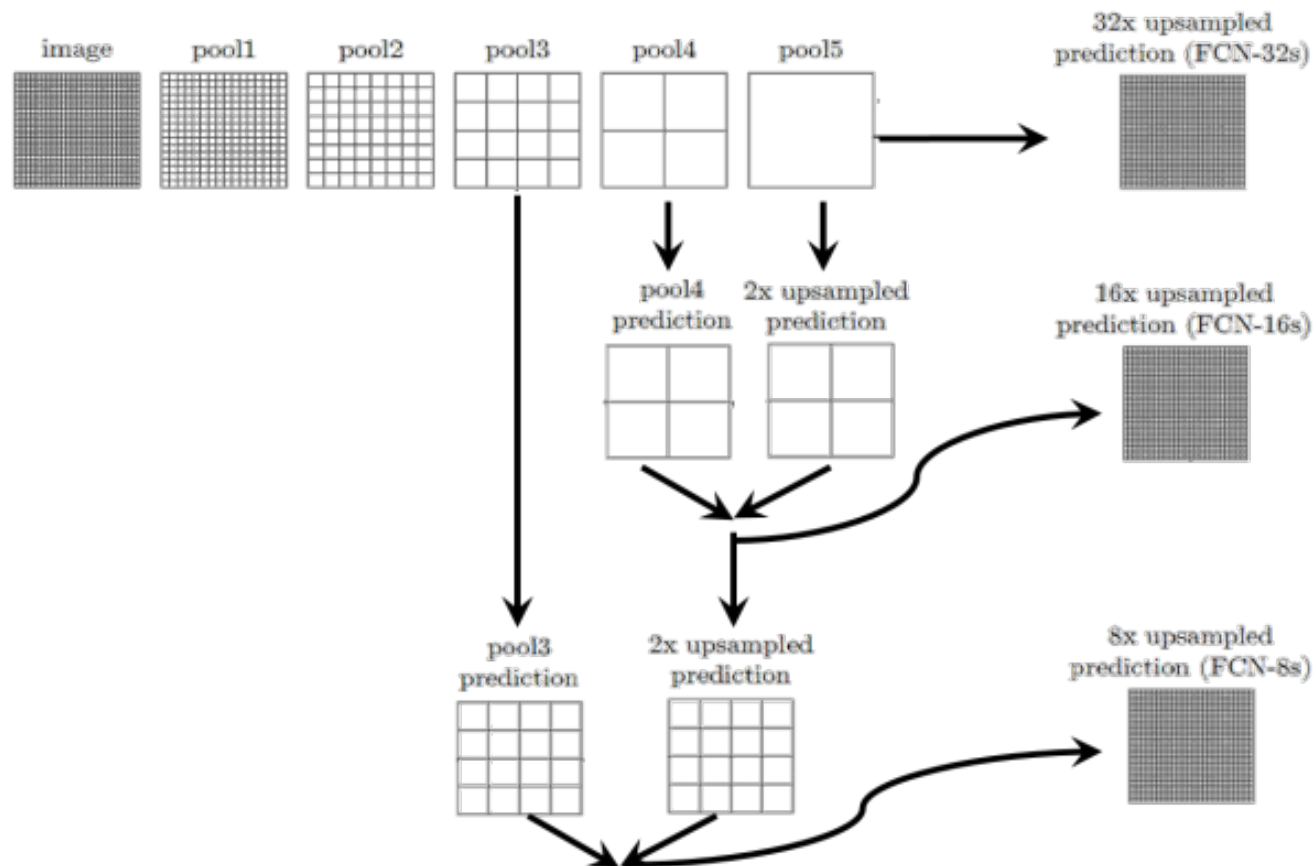
```
layer {
  name: "upscore"
  type: "Deconvolution"
  bottom: "score_fr"
  top: "upscore"
  param {
    lr_mult: 0
  }
  convolution_param {
    num_output: 21
    bias_term: false
    kernel_size: 64
  }
}
```

```
}  
}
```

可以看到lr\_mult被设置为了0.

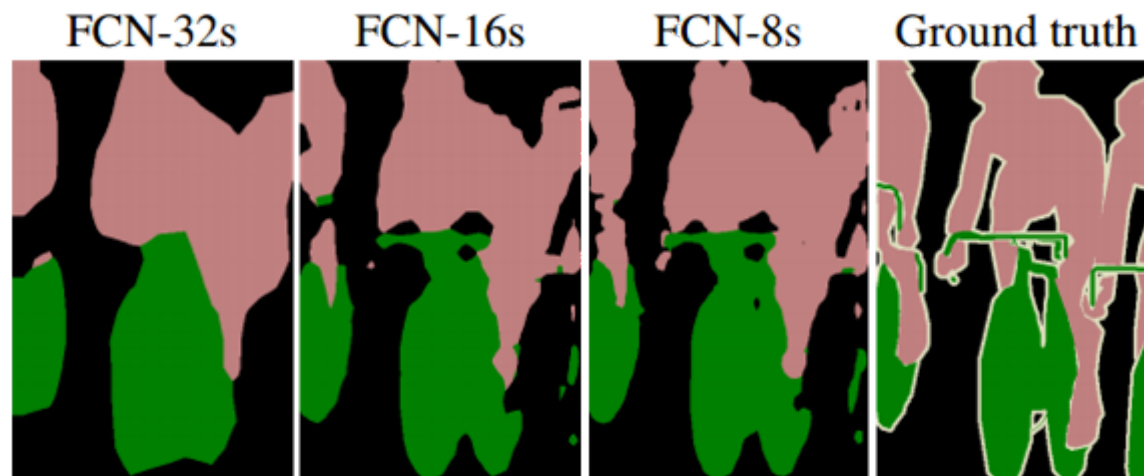
## 跳跃结构

（这个奇怪的名字是我翻译的，好像一般叫忽略连接结构）这个结构的作用就在于优化结果，因为如果将全卷积之后的结果直接上采样得到的结果是很粗糙的，所以作者将不同池化层的结果进行上采样之后来优化输出。具体结构如下：



而不同上采样结构得到的结果对比如下：





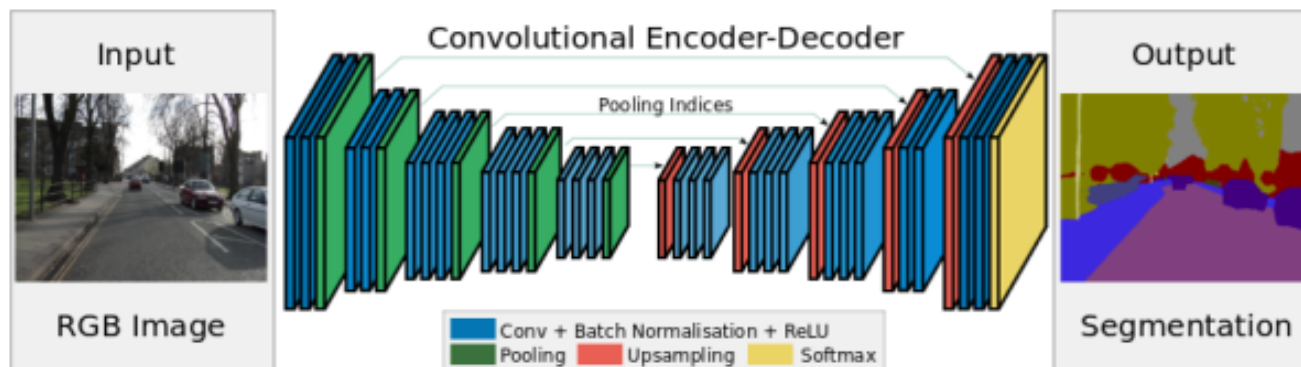
当然，你也可以将pool1，pool2的输出再上采样输出。不过，作者说了这样得到的结果提升并不大。

这是第一种结构，也是深度学习应用于图像语义分割的开山之作，所以得了CVPR2015的最佳论文。但是，还是有一些处理比较粗糙的地方，具体和后面对比就知道了。

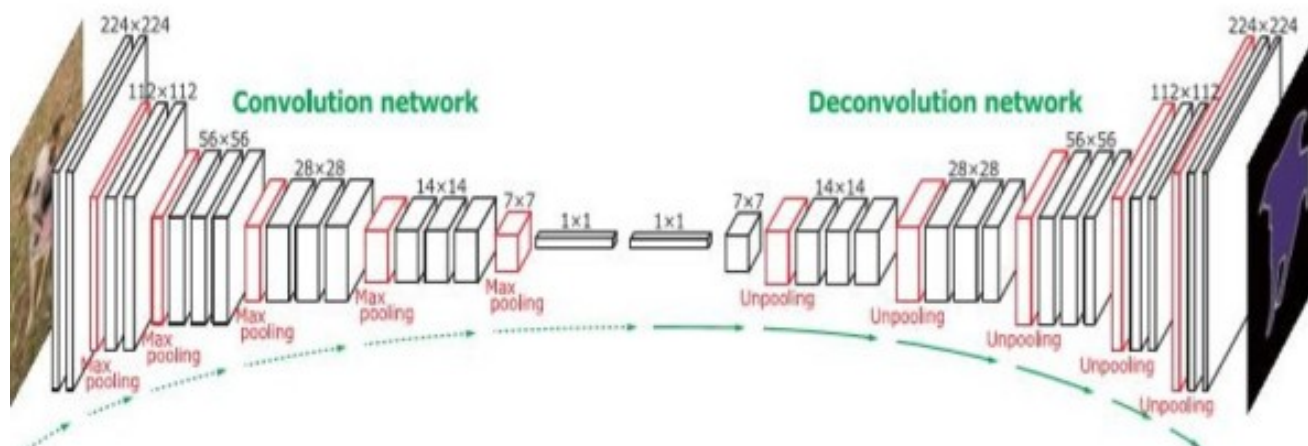
## SegNet/DeconvNet

这样的结构总结在这儿，只是我觉得结构上比较优雅，它得到的结果不一定比上一种好。

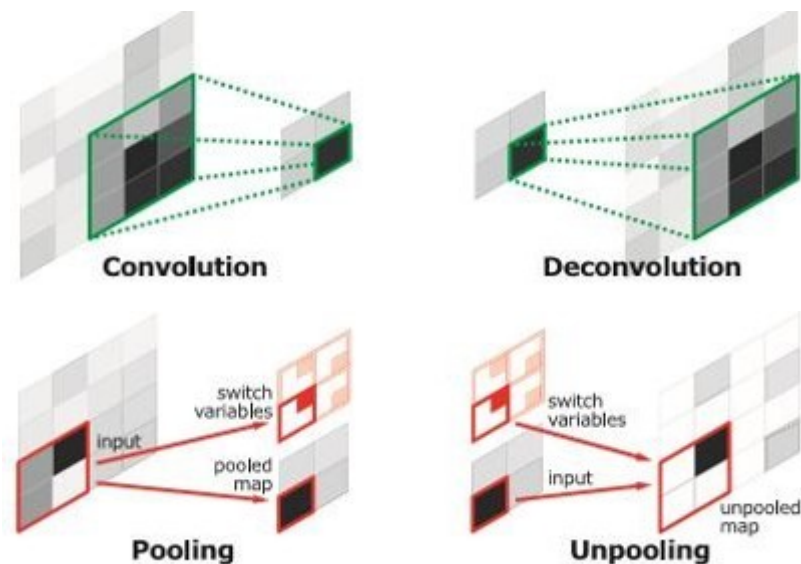
### SegNet



## DeconvNet



这样的对称结构有种自编码器的感觉在里面，先编码再解码。这样的结构主要使用了反卷积和上池化。即：



反卷积如上。而上池化的实现主要在于池化时记住输出值的位置，在上池化时再将这个值填回原来的位置，其他位置填0即OK。

## DeepLab

接下来介绍一个很成熟优雅的结构，以至于现在的很多改进是基于这个网络结构的进行的。

首先这里我们将指出一个第一个结构FCN的粗糙之处：为了保证之后输出的尺寸不至于太小，FCN的作者在第一层直接对原图加了100的padding，可想而知，这会引入噪声。

而怎样才能保证输出的尺寸不会太小而又不会产生加100 padding这样的做法呢？可能有人会说减少池化层不就行了，这样理论上是可以的，但是这样直接就改变了原先可用的结构了，而

只是简单的，上显然不能因以前的结构参数进行，所以，这里使用了一个

知



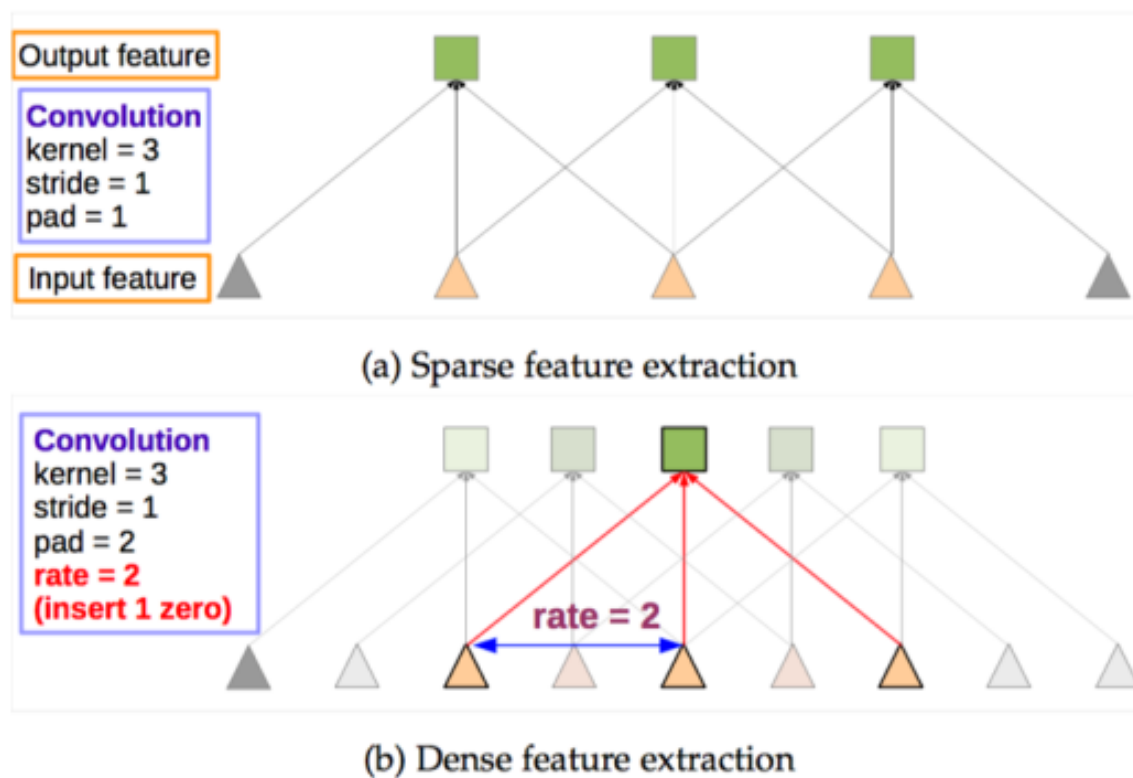
首发于  
智能单元

写文章

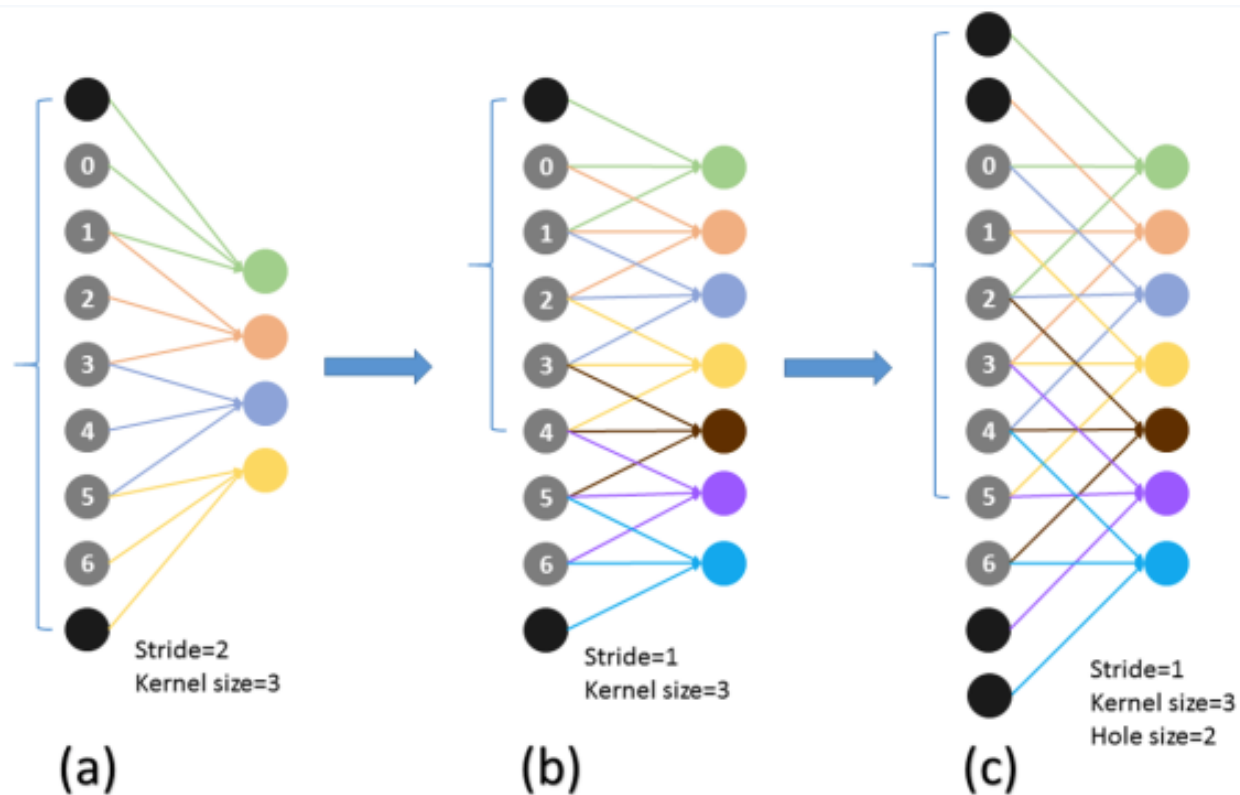
登录

非常优雅的做法：将pooling的stride改为1，再加上 1 padding。这样池化后的图片尺寸并未减小，并且依然保留了池化整合特征的特性。

但是，事情还没完。因为池化层变了，后面的卷积的感受野也对应的改变了，这样也不能进行 fine-tune 了。所以，Deeplab提出了一种新的卷积，带孔的卷积：Atrous Convolution.即：

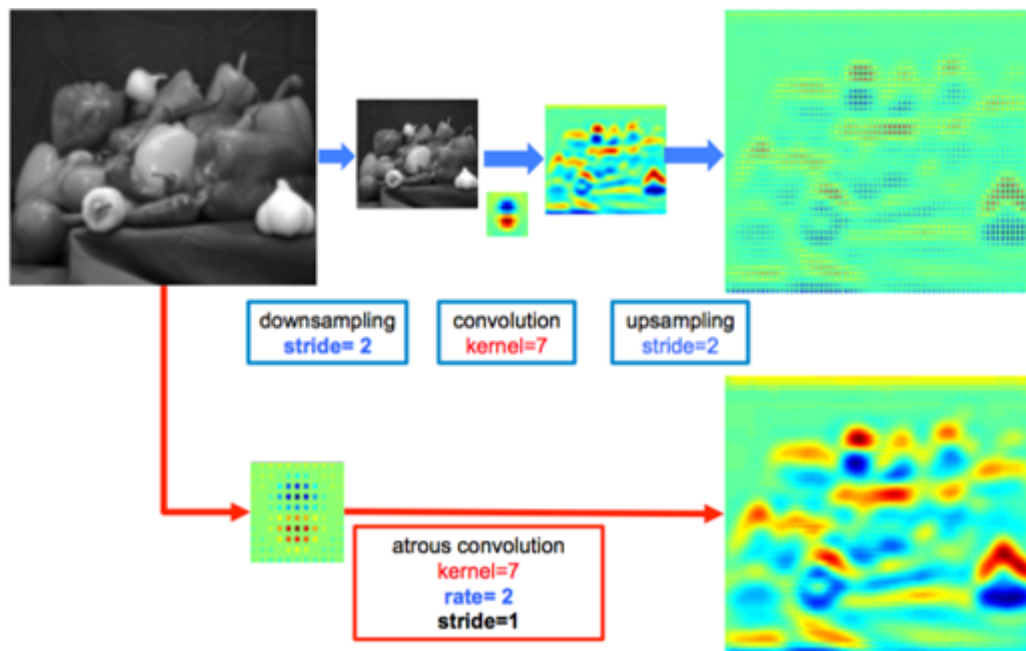


而具体的感受野变化如下：



a为普通的池化的结果，b为“优雅”池化的结果。我们设想在a上进行卷积核尺寸为3的普通卷积，则对应的感受野大小为7。而在b上进行同样的操作，对应的感受野变为了5。感受野减小了。但是如果使用hole为1的Atrous Convolution则感受野依然为7。

所以，Atrous Convolution能够保证这样的池化后的感受野不变，从而可以fine tune，同时也能保证输出的结果更加精细。即：



## 总结

这里介绍了三种结构：FCN, SegNet/DeconvNet, DeepLab。当然还有一些其他的结构方法，比如有用RNN来做的，还有更有实际意义的weakly-supervised方法等等。

## 后端

终于到后端了，后端这里会讲几个场，涉及到一些数学的东西。我的理解也不是特别深刻，所以欢迎吐槽。

### 全连接条件随机场(DenseCRF)



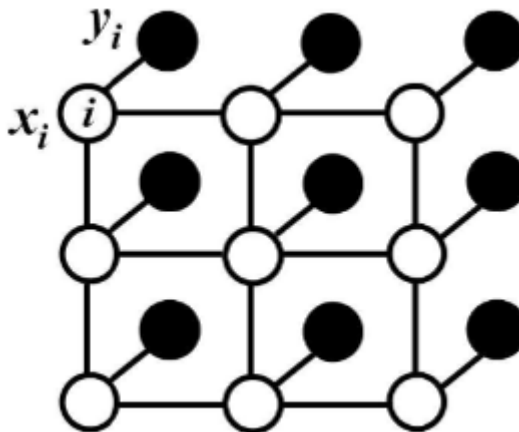
首发于  
智能单元

知

写文章

登录

对于每个像素  $i$  具有类别标签  $x_i$  还有对应的观测值  $y_i$ ，这样每个像素点作为节点，像素与像素间的关系作为边，即构成了一个条件随机场。而且我们通过观测变量  $y_i$  来推测像素  $i$  对应的类别标签  $x_i$ 。条件随机场如下：



条件随机场符合吉布斯分布：(此处的  $x$  即上面说的观测值)

$$P(\mathbf{X} = \mathbf{x}|\mathbf{I}) = \frac{1}{Z(\mathbf{I})} \exp(-E(\mathbf{x}|\mathbf{I}))$$

其中的  $E(\mathbf{x}|\mathbf{I})$  是能量函数，为了简便，以下省略全局观测  $\mathbf{I}$ ：

$$E(\mathbf{x}) = \sum_i \Psi_u(x_i) + \sum_{i < j} \Psi_p(x_i, x_j)$$

其中的一元势函数  $\sum_i \Psi_u(x_i)$  即来自于前端FCN的输出。而二元势函数如下：

二元势函数就是描述像素点与像素点之间的关系，鼓励相似像素分配相同的标签，而相差较大的像素分配不同标签，而这个“距离”的定义与颜色值和实际相对距离有关。所以这样CRF能够使图片尽量在边界处分割。

而全连接条件随机场的不同就在于，二元势函数描述的是每一个像素与其他所有像素的关系，所以叫“全连接”。

关于这一堆公式大家随意理解一下吧... ..而直接计算这些公式是比较麻烦的（我想也麻烦），所以一般会使用平均场近似方法进行计算。而平均场近似又是一堆公式，这里我就不给出了（我想大家也不太愿意看），愿意了解的同学直接看论文吧。

## CRFasRNN

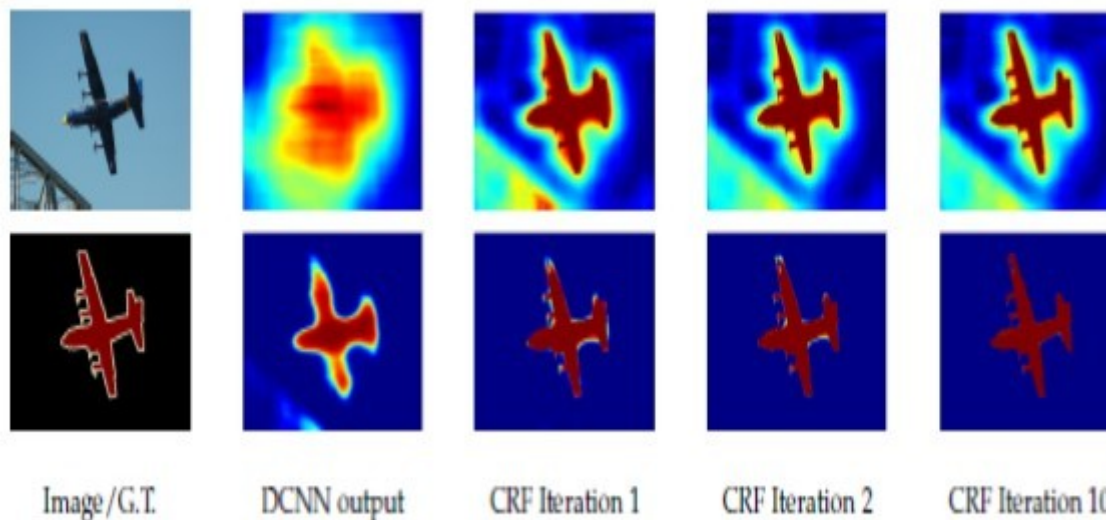
最开始使用DenseCRF是直接加在FCN的输出后面，可想这样是比较粗糙的。而且在深度学习中，我们都追求end-to-end的系统，所以CRFasRNN这篇文章将DenseCRF真正结合进了FCN中。

这篇文章也使用了平均场近似的方法，因为分解的每一步都是一些相乘相加的计算，和普通的加减（具体公式还是看论文吧），所以可以方便的把每一步描述成一层类似卷积的计算。这样即可结合进神经网络中，并且前后向传播也不存在问题。

当然，这里作者还将它进行了迭代，不同次数的迭代得到的结果优化程度也不同（一般取10以内的迭代次数），所以文章才说是as RNN。优化结果如下：





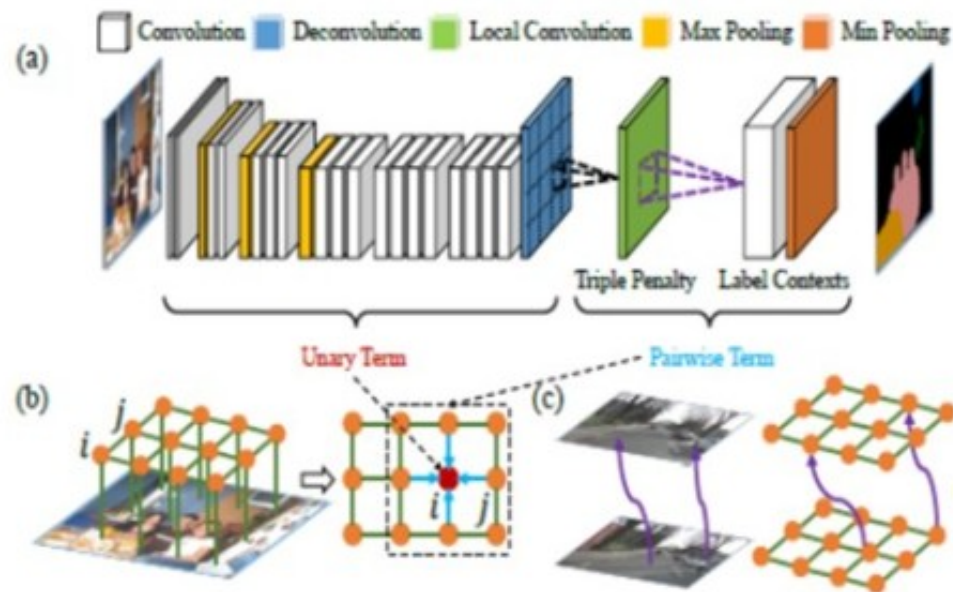


### 马尔科夫随机场(MRF)

在Deep Parsing Network中使用的是MRF，它的公式具体的定义和CRF类似，只不过作者对二元势函数进行了修改：

$$\Psi(y_i^u, y_i^v) = \sum_{k=1}^K \lambda_k u_k(i, u, j, v) \sum_{\forall z \in N_j} d(j, z) p_z^v$$

其中，作者加入的  $\lambda_k$  为label context，因为  $u_k$  只是定义了两个像素同时出现的频率，而  $\lambda_k$  可以对一些情况进行惩罚，比如，人可能在桌子旁边，但是在桌子下面的可能性就更小一些。所以这个量可以学习不同情况出现的概率。而原来的距离  $d(i, j)$  只定义了两个像素间的关系，作者在这儿加入了个triple penalty，即还引入了  $j$  附近的  $z$ ，这样描述三方关系便于得到更充足

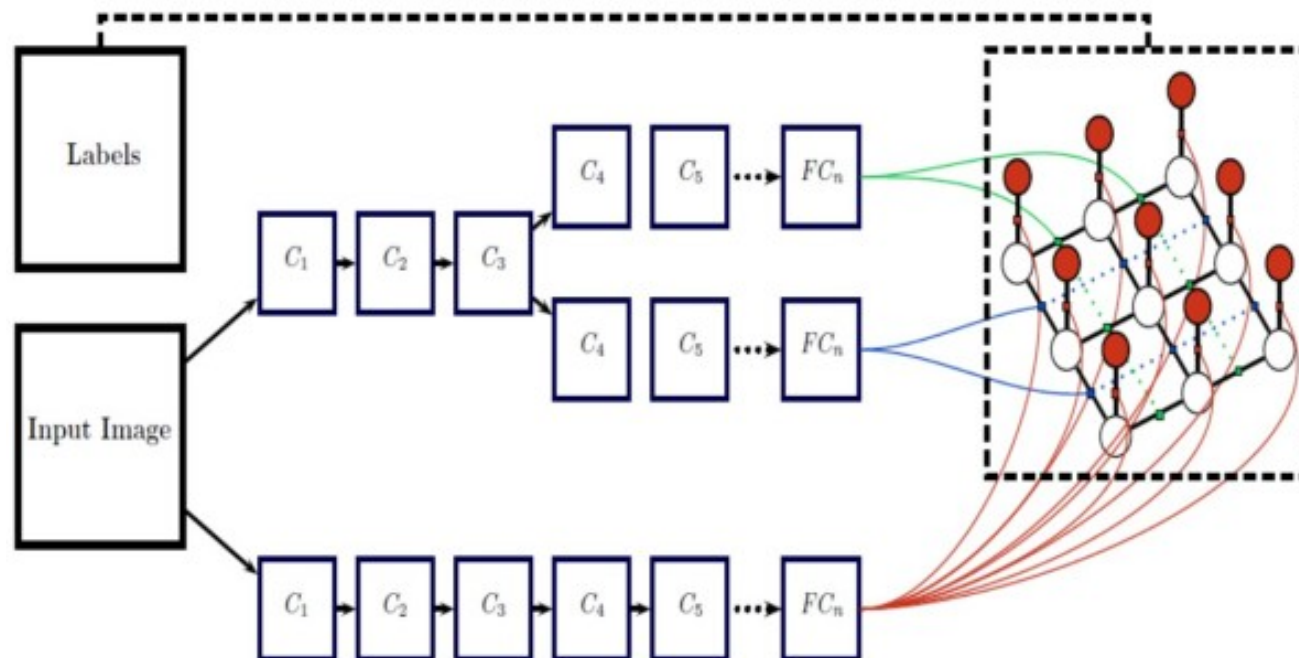


这个结构的**优点**在于：

- 将平均场构造成了CNN
- 联合训练并且可以one-pass inference，而不用迭代

### 高斯条件随机场(G-CRF)

这个结构使用CNN分别来学习一元势函数和二元势函数。这样的结构是我们更喜欢的：



而此中的能量函数又不同于之前：

$$E(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T (\mathbf{A} + \lambda \mathbf{I}) \mathbf{x} - \mathbf{B} \mathbf{x}$$

而当  $(\mathbf{A} + \lambda \mathbf{I})$  是对称正定时，求  $E(\mathbf{x})$  的最小值等于求解：

$$(\mathbf{A} + \lambda \mathbf{I}) \mathbf{x} = \mathbf{B}$$

而G-CRF的优点在于：

- 二次能量有明确全局

## 感悟

- FCN更像一种技巧。随着基本网络（如VGG，ResNet）性能的提升而不断进步。
- 深度学习+概率图模型（PGM）是一种趋势。其实DL说白了就是进行特征提取，而PGM能够从数学理论很好的解释事物本质间的联系。
- 概率图模型的网络化。因为PGM通常不太方便加入DL的模型中，将PGM网络化后能够是PGM参数自学习，同时构成end-to-end的系统。

完结撒花

## 引用

[1][Fully Convolutional Networks for Semantic Segmentation](#)

[2][Learning Deconvolution Network for Semantic Segmentation](#)

[3][Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials](#)

[4][Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs](#)

[5][Conditional Random Fields as Recurrent Neural Networks](#)

[6][DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs](#)

知



首发于  
智能单元

写文章

登录

[8][Fast, Exact and Multi-Scale Inference for Semantic Image Segmentation with Deep Gaussian CRFs](#)

[9][SegNet](#)

转载须全文转载且注明作者和原文链接，否则保留维权权利

「真诚赞赏，手留余香」

赞赏

2 人赞赏



深度学习（Deep Learning）

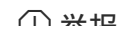
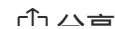
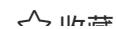
Caffe（深度学习框架）

semantic segmentation

知



首发于  
智能单元



写文章

登录



### 文章被以下专栏收录



**智能单元**  
聚焦通用人工智能

[进入专栏](#)

**Semantic Segmentation Paper Reading**  
紧随图像语义分割的进展

[进入专栏](#)

### 86 条评论

写下你的评论...



**御宅暴君**

文章的引用太乱，我重新整理成保持与文章小章节顺序一致，且所有超链接尽可能指向当前最新的 arxiv 且被格式化完整的 paper title. FCN:

[\[1605.06211\] Fully Convolutional Networks for Semantic Segmentation](#)

SegNet:

[A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation](#)

DeconvNet:

[\[1505.04366\] Learning Deconvolution Network for Semantic Segmentation](#)

DeepLab:



全连接条件随机场(DenseCRF):

[Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials](#)

CRFasRNN:

[\[1502.03240\] Conditional Random Fields as Recurrent Neural Networks](#)

马尔科夫随机场(MRF):

[\[1509.02634\] Semantic Image Segmentation via Deep Parsing Network](#)

高斯条件随机场(G-CRF):

[\[1603.08358\] Fast, Exact and Multi-Scale Inference for Semantic Image Segmentation](#)

[with Deep Gaussian CRFs](#)

9 个月前

12 赞

以上为精选评论



**Xenophon Tony**

好棒 ~ 谢谢作者 ~ ~

1 年前



**ycszen (作者)** 回复 **Xenophon Tony**

有用就好~

1 年前

查看对话



**ICOZ**

非常感谢，请问您写过关于cnn的文章吗 或者推荐的文章

知



首发于  
智能单元

写文章

登录



**ycszen (作者)** 回复 **ICOZ**

[查看对话](#)

就是普通的CNN吗？这个可以看看三巨头合出的nature的论文:Deep Learning

1 年前

1 赞



**柴云**

好棒~~

1 年前



**Jianping Shi**

总结的不错，有兴趣来商汤实习么？会有非常多实际中需要用segmentation解决的问题，也会有不少适合research的topic 有兴趣可以私聊☺

1 年前



**crackhopper**

BTW, 概率图应该是PGM。probablistic graph model，如果我没记错。

1 年前



**ycszen (作者)** 回复 **crackhopper**

[查看对话](#)

呃.....对。我的失误

1 年前





我找时间翻译到julia语言里.....  
1 年前

1

2

3

4

...

9

[下一页](#)

## 推荐阅读



### 财务尽调过程中如何“防雷”

作者：刘嘯-VC2017年8月12日星期六由于IPO审核进程加快，不少公司准备加快冲击IPO的进程，如... [查看全文](#) >

投行小兵 · 13 天前 · 编辑精选 · 发表于 小兵研究精华



### 被告存在“不适格”吗？


\*本文经授权发布，谢绝无授权转载\*在一些诉讼案件中，常常会听到被告答辩主张不是“适格被告... [查看全文](#) >

建纬（北京）律师事务所 · 6 天前 · 编辑精选

知



首发于  
智能单元

 写文章

[登录](#)



这两年，我关注了很多关于保险的微信公众号，个人感觉在众多保险理念和保险知识科普文章之外... [查看全文](#) >

sky · 12 天前 · 编辑精选



## 这锅红烧牛腩是我不外传的撩男大杀器

我的文章都 先在公众号发布的，拒绝任何没有告知过我的抄袭和转载。公共号：食色信也。很不... [查看全文](#) >

村姑信 · 5 个月前 · 编辑精选 · 发表于 食色信也