



## FCN(1)——从分类问题出发



冯超

《强化学习精要》《深度学习轻松学》作者, 咨询请值乎

63 人赞了该文章

本文收录在[无痛的机器学习第一季](#)。

前面我们介绍了许多利用CNN进行分类任务的模型, 今天来看看用CNN做分割的模型。我们的主要内容来自这篇文章——《Fully Convolutional Networks for Semantic Segmentation》。

### 图像分割

首先我们来简单看看图像分割的任务是什么样的。所谓的图像分割就是将一幅图像中各个实体的边界确定下来, 这样我们可以通过寻找这个实体的边界确定实体的位置。在自然场景中我们可以根据画面中的实体进行划分, 在文档图片中, 我们也可以找出每一个文字的边界。这些都可以算是完成分割任务。

当然, 上面的描述还不太具有可操作性, 我们可以把问题做进一步地转化。我们可以把问题转换成一个分类问题, 我们判定每一个像素点的类别, 然后将相邻且类别相同的像素点聚集起来, 不就可以找到实体的边界了么? 这样也就是从分类问题出发解决分割问题的方式。

但是这样解决分割问题的方法还是存在一些问题: 之前我们对于一张图像, 最终只输出一个结果或者几个结果, 结果的数量并不多; 而现在我们输出的数量和像素的数量相同, 那么我们还能够获得和分类问题近似的精确率么? 这里面存在着一些需要解决的问题, 比如:

在分类问题中, 为了解决图像的transform invariance, 我们会加入pooling层减小维度, 这个被减小的维度如何能够再次放大呢?

### FCN的解决方案

FCN, 也就是Fully Convolutional Network, 是一个不包含全连接层的网络。这里面所谓的不包含全连接层, 实际上并不是标榜自己没有全连接层, 而是为了保证计算过程中每一层数字的相对位置。在我们通常的印象中, 全连接层需要把本来立体的图像拍平, 这样原本存在的空间特性将被抹掉。为了确保我们识别出来的类别能和原来每一个像素点的位置对上, 我们不能粗暴地把中间数据拍平, 这也是网络中不使用全连接层的原因。

我们可以利用常规的分类CNN网络得到一个接近最终结果的中间层, 这样层的数据往往已经可以代表了一些具有特定含义的特征, 而不再像原始的像素亮度那样含义晦涩。因此我们可以从这样的信息出发得到一些分类信息, 然后把这些分类信息重新映射到原始图片大小的区域上。

由于中间层的维度比原始图像小, 那么恢复到原始大小必然意味着一些插值的工作。一旦使用了插值的算法(比方说bilinear), 那么恢复的图像的精度一定会出现问题。所以大神们也曾经担心过这个问题。于是他们也想出过一些其他的方法。这些方法的特点就一条——

维度不缩小!

实际上想做到这一步从算法上来说也并不困难, 但是这样的方法会增加不少计算量, 而且也存在自己的问题。最终大神们选择了反卷积的方式完成了信息由小变大的工作。

反卷积可以让维度由小变大, 而且我们还可以通过学习其中的参数让这个变化的过程变得不那么简单粗暴。但是小维度毕竟是小维度, 谁也没法回避这个问题。就算采用反卷积的方式把维度扩大, 精度损失的问题依旧无法避免。所以我们还需要其他的办法。

这个办法就是融合。我们不仅仅使用较深层的特征信息, 还使用一些较浅的特征信息。我们知道较浅层的特征容易保留一些细节信息, 比方说边缘信息, 较深的特征容易保留一些类别信息, 那么如果我们把这两部分信息融合起来, 我们既可以保证找到的类别信息是准确的, 同时我们也可以保证一些边缘的位置信息能够找准。这样我们也就解决了这个问题。

文章中还提到了一个有意思的问题, 那就是**同样的问题——分类问题都是找图, 而现在的分割问题变成了一张图**

赞同 63

13 条评论

分享

收藏

...

不会出现过拟合的问题？最终的实验结果告诉了我们，过拟合的问题似乎并不存在，但是当初大家还是担心了一下，也曾经想过采用采样的方式解决这个问题。对于最终输出的每一个像素的类别信息，我们并不把所有的像素点的结果计算到loss中进行反向传播，而是只取其中一部分的像素点。这个想法是有点道理的，因为每一个紧密相邻的像素点之前的特征差距可能并不大，如果每一个像素点都计算在内，那么就相当于我们对某一组特征增加了很高的权重。但好在我们对所有像素点都增加权重的话，这个影响还是会抵消的。

最后提一下分割问题和分类问题在evaluation方面的不同。对于经典的分类问题，我们常用的loss是cross entropy loss，精确率判断则是最终的类别判断的正确率。而在分割问题中，我们除了判断每一个像素点的“分类”loss和精确率，我们还可以计算IoU。它的全称是Intersection of Union。我们有模型预测的边界和Ground Truth的边界。我们计算两个边界相交和它们相并的比例，也可以判断最终的分割效果。如果以IOU作为评价标准，那么只要主体部分能够分割正确，那么我们就可以拿到比较高的分数。所以添加过多浅层的信息可能不会对最终结果造成很大的准确率提升。

未来

从上面的内容中，我们基本上完成了对FCN的基本介绍。FCN是一个将High-level问题的模型框架应用到Low-level问题的成功案例。但是，从已经走过的历史长河来看，这个方法开辟一片崭新的领域，但是并没有走在这个领域的前沿。所以相关的应用在此我们也不赘述了，留给大家自己开发了。

曾经在图像分割问题中，我们经常使用概率图模型的方式进行建模求解，正如一些童鞋所知。后来的大神们尝试了深度CNN和CRF模型结合的手段进行尝试，并获得了非常不错的成绩。在后面的篇幅里我们就来看看两种方法是如何结合在一起的。

当然，我们花了很长的时间介绍了CNN模型的方方面面，所以一路看下来的各位对CNN已经比较熟悉，所谓没吃过猪肉也见过猪跑了。但是CRF的内容在此之前我们并未涉及到，因此想搞清楚CRF对于初学者还是有点压力，下面我们会尽可能地用最通俗的语言介绍下CRF。

广告时间

更多精彩尽在《深度学习轻松学：核心算法与视觉实践》！

编辑于 2017-11-22

机器学习    深度学习 (Deep Learning)

文章被以下专栏收录



无痛的机器学习

专栏主营业务：让更多人能看的懂的机器学习科普+进阶文章。欢迎各位大神投稿或协...

进入专栏

推荐阅读



一些分类方法  
(Classification) 的总结 I

话说

发表于话说



图像分类 (零)

射命丸咲

发表于Pytho...



Jan Stypka

文本多标签分类工具—  
Magpie

杜润之

▲ 赞同 63 ▼    13 条评论    分享    ★ 收藏    ...

13 条评论

切换为时间排序

写下你的评论...

 SCP-1731 年前

resnet and fcn ?

赞

 冯超 (作者) 回复 SCP-1731 年前

only fcn

赞 查看对话

 孟昱先1 年前

谢谢分享~一个小问题，IoU里的o似乎应该是over不是of

1

 冯超 (作者) 回复 孟昱先1 年前

不好意思，我去查一下确定下吧。不过两个单词背后表达的含义是一样的

赞 查看对话

 Pygmalion Chen1 年前

前辈你好，请问全连接层把图像拍平时什么意思啊，不太理解？

赞

 冯超 (作者) 回复 Pygmalion Chen1 年前

我的表述不够专业，拍平相当于改变维度，从过去单张图的c\*h\*w变成了单一维度

赞 查看对话

 Pygmalion Chen 回复 冯超 (作者)1 年前

明白了，谢谢啦！

赞 查看对话

 知乎用户1 年前

“这个被减小的纬度如何能够再次放大呢”-----“维度”

赞

 冯超 (作者) 回复 知乎用户1 年前

谢谢，我会尽快修改

1 查看对话

 郭帆1 年前

前辈您好，请问当反向传播时，梯度是如何穿过crop层和denconv层的？

赞

 知乎用户 回复 冯超 (作者)1 年前


A over B是B分之A或者A比B的意思，IoU里就是intersection比上union的比值，如果是of的就解释不通了吧

赞 查看对话

 知乎用户6 个月前

请问fcn得到的feature map怎么用于分类问题啊，谢谢

赞

 孙小也

赞同 6313 条评论分享收藏...

您好，您提到的，计算loss时，如果把所有的pixel都计算在内，相邻的类别相同可能会被赋了更大的权值，但是这个影响可以被抵消，能说下是怎么抵消的吗？

👍 赞