

# 基于深度卷积神经网络的小样本分割算法综述

CV开发者都爱看的 极市平台 2022-02-05 22:00

↑ 点击[蓝字](#) 关注极市平台



来源 | AI算法修炼营

编辑 | 极市平台

## 极市导读

本文综合介绍了基于深度卷积神经网络的小样本分割相关工作，并整理了方法分类、数据集、评测指标、实验结果等内容。>>加入极市CV技术交流群，走在计算机视觉的最前沿

### 介绍

深度卷积神经网络在图像分类、目标检测、语义分割等许多视觉理解任务上都取得了重大突破。一个关键的原因是**大规模数据集的可用性**，比如 ImageNet，这些数据集支持对深度模型的培训。然而，数据标记是昂贵的，特别是对于密集的预测任务，如语义分割和实例分割。

此外，在对模型进行训练之后，很难将模型应用于新类的预测。与机器学习算法不同的是，人类只看到几个例子就能很容易地从图像中分割出一个新概念。

人类和机器学习算法之间的差距激发了对**小样本学习**的研究，其目的是**学习一个模型，可以很好地推广到具有稀缺标记的训练数据的新类别**。

小样本分割的终极目的是**利用支持集中的 K 个训练图像对来“学习”一个模型**，使得该模型能对训练图像中出现的类别的新样本实现分割。

## 相关工作

### 2.1 元学习

元学习解决的是学习如何学习的问题。元学习的思想是学习「学习（训练）」过程。主要有基于记忆 Memory 的方法、基于预测梯度的方法、利用 Attention 注意力机制的方法、借鉴 LSTM 的方法、面向 RL 的 Meta Learning 方法、利用 WaveNet 的方法、预测 Loss 的方法等等。

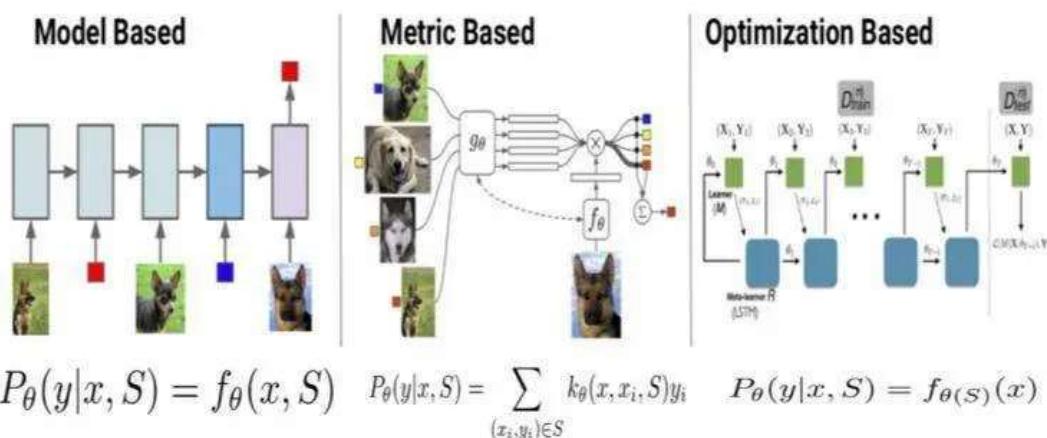
### 2.2 小样本学习

小样本学习是元学习在监督学习领域的应用，Few-shot Learning

模型大致可分为三类：Model Based，Metric Based 和 Optimization Based。

其中 Model Based 方法旨在通过模型结构的设计快速在少量样本上更新参数，直接建立输入  $x$  和预测值  $P$  的映射函数；Metric Based 方法通过度量 batch 集中的样本和 support 集中样本的距离，借助最近邻的思想完成分类。

Optimization Based 方法认为普通的梯度下降方法难以在 few-shot 场景下拟合，因此通过调整优化方法来完成小样本分类的任务。



### 2.3 语义分割

语义分割就是按照“语义”给图像上目标类别中的每一点打一个标签，使得不同种类的东西在图像上被区分开来。可以理解成像素级别的分类任务。

输入: ( $H \times W \times 3$ ) 正常的图片；输出: ( $H \times W \times \text{class}$ ) 可以看为图片上每个点的 one-hot 表示，每一个 channel 对应一个 class，对每一个 pixel 位置，都有 class 数目个 channel，每个 channel 的值对应那个像素属于该 class 的预测概率。

**FCN** 是语义分割的开山之作，主要特色有两点：全连接层换成卷积层，不同尺度的信息融合 FCN-8S,16s,32s。

**U-net** 用于解决小样本的简单问题分割，比如医疗影片的分割。它遵循的基本原理与 FCN 一样：

1. Encoder-Decoder 结构：前半部分为多层卷积池化，不断扩大感受野，用于提取特征。后半部分上采样回复图片尺寸。
2. 更丰富的信息融合：如灰色剪头，更多的前后层之间的信息融合。这里是把前面层的输出和后面层 concat（串联）到一起，区别于 FCN 的逐元素加和。

不同 Feature map 串联到一起后，后面接卷积层，可以让卷积核在 channel 上自己做出选择。注意的是，在串联之前，需要把前层的 feature map crop 到和后层一样的大小。

**SegNet** 和 U-net 在结构上其实大同小异，都是编码-解码结果。区别在于，SegNet 没有直接融合不同尺度的层的信息，为了解决为止信息丢失的问题，SegNet 使用了带有坐标 (index) 的池化。

在 Max pooling 时，选择最大像素的同时，记录下该像素在 Feature map 的位置（左图）。在反池化的时候，根据记录的坐标，把最大值复原到原来对应的位置，其他的位置补零（右图）。后面的卷积可以把 0 的元素给填上。这样一来，就解决了由于多次池化造成的位置信息的丢失。

**Deeplab V1** 不同于之前的思路，他的特色有两点：

1. 由于 Pooling-Upsample 会丢失位置信息而且多层上下采样开销较大，把控制感受野大小的方法化成：带孔卷积 (Atrous conv)。

2. 加入 CRF (条件随机场) , 利用像素之间的关连信息: 相邻的像素, 或者颜色相近的像素有更大的可能属于同一个 class。

**PSPnet:** 前面的不同 level 的信息融合都是融合浅层和后层的 Feature Map, 因为后层的感受野大, 语义特征强, 浅层的感受野小, 局部特征明显且位置信息丰富。

PSPnet 则使用了空间金字塔池化, 得到一组感受野大小不同的 feature map, 将这些感受野不同的 map concat 到一起, 完成多层次的语义特征融合。

**Deeplab V2** 在 v1 的基础上做出了改进, 引入了 ASPP (Atrous Spatial Pyramid Pooling) 的结构, 如上图所示。我们注意到, Deeplab v1 使用带孔卷积扩大感受野之后, 没有融合不同层之间的信息。

ASPP 层就是为了融合不同级别的语义信息: 选择不同扩张率的带孔卷积去处理 Feature Map, 由于感受野不同, 得到的信息的 Level 也就不同, ASPP 层把这些不同层级的 feature map concat 到一起, 进行信息融合。

**Deeplab v3** 在原有基础上的改动是: 1. 改进了 ASPP 模块。2. 引入 Resnet Block。3. 丢弃 CRF。

新的 ASPP 模块: 1. 加入了 Batch Norm。2. 加入特征的全局平均池化 (在扩张率很大的情况下, 有效权重会变小) 。全局平均池化的加入是对全局特征的强调、加强。

在旧的 ASPP 模块中: 我们以为在扩张率足够大的时候, 感受野足够大, 所以获得的特征倾向于全局特征。但实际上, 扩张率过大的情况下, Atrous conv 出现了“权值退化”的问题, 感受野过大, 都已近扩展到了图像外面, 大多数的权重都和图像外围的 zero padding 进行了点乘, 这样并没有获取图像中的信息。有效的权值个数很少, 往往就是 1。于是我们加了全局平均池化, 强行利用全局信息。

**Deeplab v3+** 可以看成是把 Deeplab v3 作为编码器 (上半部分) 。后面再进行解码, 并且在解码的过程中在此运用了不同层级特征的融合。此外, 在 encoder 部分加入了 Xception 的结构减少了参数量, 提高运行速递。

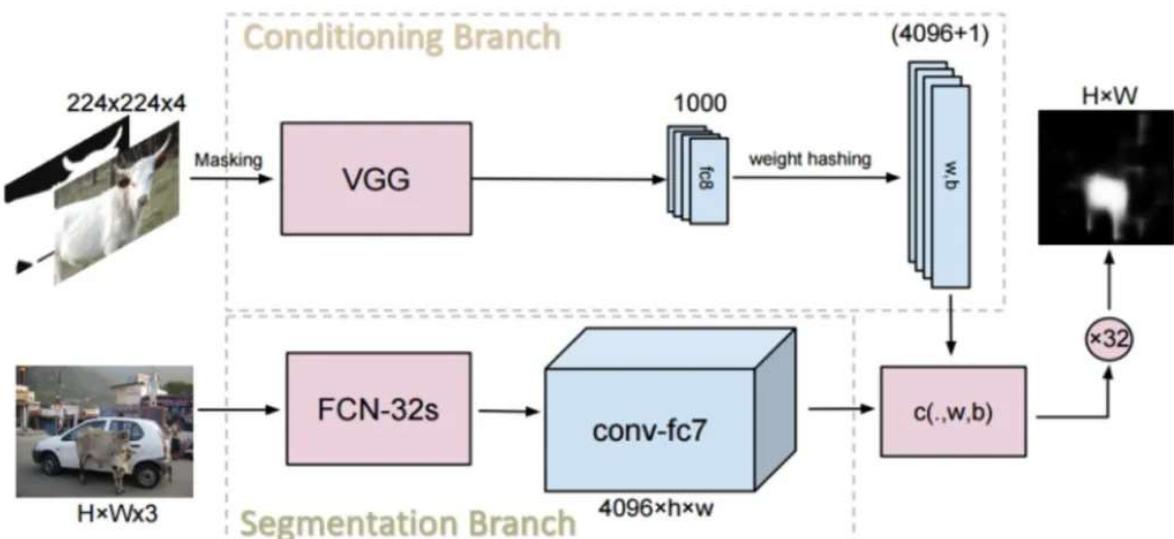
## 方法总结分类

### 3.1 OSLSM

motivation 就是学习一个 one-shot 的分割模型，首次提出双分支的网络用于 few-shot segmentation，条件分支用 VGG 提取特征，生成权重 ( $w, b$ )，分割分支用 FCN-32s 结构对 query image 进行特征提取，将其与条件分支所得参数进行点乘再通过  $\sigma$  函数得到分割结果。

得到一个分割图，上采样到图像大小，利用某个阈值产生分割的二值图。

输出的时候，为了是参数量与分割分支的特征图的通道数相对应，采用 weight hashing 的策略，将输出的 1000 维向量映射为 4097 维 ( $w:4096, b:1$ )，这种映射机制是建模成固定权重参数的全连接层来实现的。



### 3.2 co-FCN

文章的主要比较惊奇的一点是仅对原始图像做一些稀疏的标注（目标位置上点几个关键点，背景位置上点几个关键点）就却能够实现对目标的像素级的分割。

网络结构和 BMVC 那篇 paper 设置类似，也是采用双分支结构，将标注信息与原始图像 concat 后输入 conditioning branch 得到输入图像的 embedding。

利用 segmentation branch 对 query image 进行特征提取，并将结果与 conditioning branch 得到的 embedding 进行 concat，再进行像素级分割。

### **3.3 AMP-2**

**motivation:** 如何得到一个更好的原型。

**方法:** 在 task 流中，不断地更新每个类别地原型。

### **3.4 SG-One**

**训练任务的流程:**

网络包含一个主干网络 Stem (代表的是 VGG-16 的前 3 个 Block) 和两个分支：

Guidance Branch 和 Segmentation Branch，需要注意的是两个分支共用了三个卷积块，个人的理解是，如果两个分支完全独立，那么 Guidance Branch 产生的指导就会一成不变；

这就失去了意义，作者在论文最后的消融实验中做了相关的分析与实验；

与 Segmentation Branch 产生交互后就可以在优化分割损失的同时对引导的特征图进行优化，使之匹配相应的真实 mask。网络在训练的时候加载了在 ILSVRC 数据集上预训练的权重。

**测试任务的流程：**

### **主要的创新点：**

- a. 提出使用 masked average pooling 来提取 support set 的中目标的表征向量；
- b. 采用余弦相似度来度量 query set 的表征向量与 support set 的表征向量之间距离，用于指导 query set 的分割；

### **为什么 masked average pooling 会有用？**

解释如下：全卷积网络（FCN）能够保留输入图像的中每个像素相对位置；所以通过将二值 mask 与提取到的特征图相乘就可以完全保留目标的特征信息，排除掉背景等无关类别的特征。

### **3.5 PANet**

#### **创新点：**

利用了 prototypes 上的度量学习，无参数。

提出 prototypes 对齐正则化，充分利用 support 的知识。

对于带有弱注释的少样本直接使用。

用同一个 backbone 来提取 support 和 query 的深度特征，然后使用 masked average pooling 从 support 的特征将不同的前景物体和背景嵌入不同的 prototypes 中，每个 prototype 表示对应的类别，这样 query 图像的每个像素通过参考离它的嵌入表达最近的特定类的 prototype 来标记，得到 query 的预测 mask 后。

训练的时候，得到 mask 后，再将刚才提取的 query feature 和 mask 作为新的“support set”，将之前的 support set 作为新的“query set”，再用“support set”对“query set”做一波预测，然后再算一个 loss

prototype 紧凑且鲁棒的对每个语义类别进行表达；mask 标记那块就是无参数量学习，通过和嵌入空间的逐像素匹配来执行分割

执行一个 prototype 对齐正则化，用 query 和他的 mask 建立新的 support，然后用这个来预测原始的 support set 的分割，实验证明能鼓励 query 的 prototype 对齐他们的

support 的 prototype，只有训练的时候这么做（反向再推一次，看看是否真的相似）

### 3.6 CANet

#### **主要贡献：**

开发了一种新颖的双分支密集比较模块，该模块有效地利用来自CNN的多级特征表示来进行密集的特征比较。

提出迭代优化模块，以迭代方式改进预测结果。迭代细化的能力可以推广到具有少量镜头学习的看不见的类，以生成细粒度图。

采用注意机制有效地融合来自 k-shot 设置中的多个支持示例的信息，其优于单次结果的不可学习的融合方法。

证明给定的支持集具有弱注释，即边界框，我们的模型仍然可以获得与昂贵的像素级注释支持集的结果相当的性能，这进一步减少了新类别对于少数镜头分割的标记工作量。

### 3.7 PGNet

在 CANet 基础上加了一个图注意力机制

### **3.8 FWB**

**两个主要创新点：**

## Feature Weighting

支持集前后景差异标准化向量：

最大化特征差异：

最终得到的关联向量：

标准化处理后的余弦相似性：

Feature Boosting

## 数据集介绍

**PASCAL-5i**

**MS COCO-20i**

## FSS-100

### 评测指标介绍

**Mean-IoU:**

IoU 就是每一个类别的交集与并集之比，而 mIoU 则是所有类别的平均 IoU。

**FB-IoU:**

前景和背景一起的准确率

结果陈列

**如果觉得有用，就请分享到朋友圈吧！**



**极市平台**

专注计算机视觉前沿资讯和技术干货，官网：[www.cvmart.net](http://www.cvmart.net)

617篇原创内容

---

公众号

△点击卡片关注极市平台，获取最新CV干货

公众号后台回复“**transformer**”获取最新Transformer综述论文下载～

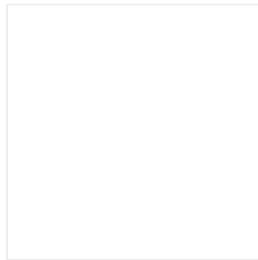
**极市干货**

**课程/比赛：**珠港澳人工智能算法大赛 | 保姆级零基础人工智能教程

**算法trick：**目标检测比赛中的tricks集锦 | 从39个kaggle竞赛中总结出来的图像分割的Tips和Tricks

**技术综述：**一文弄懂各种loss function | 工业图像异常检测最新研究总结（2019-2020）

## # CV技术社群邀请函 #



△长按添加极市小助手

添加极市小助手微信 (**ID : cvmart4**)

备注：姓名-学校/公司-研究方向-城市（如：小极-北大-目标检测-深圳）

即可申请加入极市**目标检测/图像分割/工业检测/人脸/医学影像/3D/SLAM/自动驾驶/超分辨率/姿态估计/ReID/GAN/图像增强/OCR/视频理解**等技术交流群

每月大咖直播分享、真实项目需求对接、求职内推、算法竞赛、干货资讯汇总、与 **10000+** 来自港科大、北大、清华、中科院、CMU、腾讯、百度等名校名企视觉开发者互动交流~

觉得有用麻烦给个在看啦~

阅读原文 阅读 1951

分享 收藏

赞 6 在看 5

分享此内容的人还喜欢

真·机械朋克！物理学家用扬声器搭建神经网络，登上今日Nature  
量子位 赞 142

写下你的留言