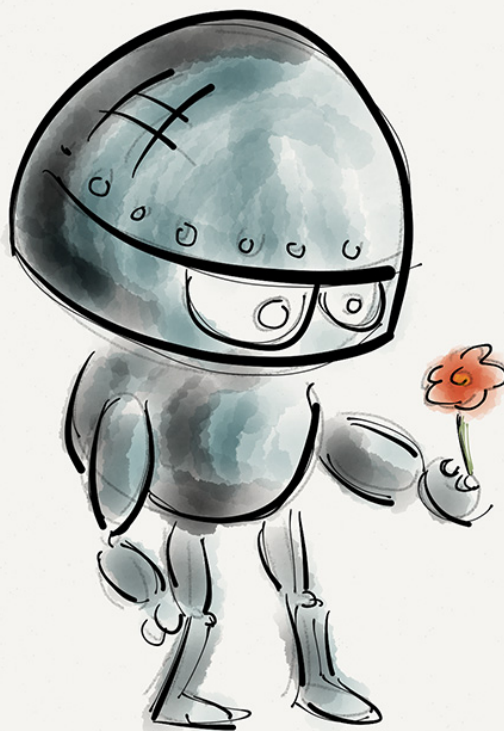


## 049 | 精读2017年ICCV最佳学生论文

2018-01-24 洪亮劫

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 07:20 大小 3.37M



周一我们认真剖析了 ICCV 2017 年的最佳研究论文 “Mask R-CNN”。今天我们来分享 ICCV 2017 的最佳学生论文《焦点损失用于密集物体检测》（[Focal Loss for Dense Object Detection](#)）。

可以说，这篇文章是我们周一分享的最佳论文的孪生兄弟。首先，这篇论文的作者群也基本是 Facebook 人工智能研究院的班底。其次，这篇文章解决的问题也很类似，也是物体识别和语义分割，只是不解决数据点分割的问题。

### 作者群信息介绍

除第一作者外，这篇论文的作者都来自 Facebook 的人工智能研究院。

第一作者林仓义 (Tsung-Yi Lin) , 目前在谷歌大脑 (Google Brain) 团队工作, 发表论文的时候在 Facebook 人工智能研究院实习。林仓义在台湾国立大学获得本科学位, 在加州大学圣地亚哥分校获得硕士学位, 2017 年刚从康奈尔大学博士毕业。博士期间, 他师从计算机视觉专家塞尔盖·比隆基 (Serge Belongie) , 发表了多篇高质量的计算机视觉论文。

第二作者皮里亚·高耶 (Priya Goyal) 是 Facebook 人工智能研究院的一名研究工程师。在加入 Facebook 之前, 皮里亚从印度理工大学获得了学士和硕士学位。

第三作者罗斯·吉尔什克 (Ross Girshick) , 第四作者何恺明, 还有最后一个作者皮奥特·多拉 (Piotr Dollár) , 这三位作者也是周一的最佳研究论文的作者, 我们已经介绍过了, 你可以回去再了解一下。

## 论文的主要贡献

我们首先来看一下这篇文章的主要贡献。

刚才我们已经简单地谈到了, **这篇文章要解决的问题, 就是对输入图像进行物体识别和语义分割这两个任务**。对于这个问题有两种主要的思路, 这两个思路都在不断地发展。

第一种思路, 那就是直接从输入图像入手, 希望能够从输入图像中提取相应的特征, 从而能够直接从这些特征中判断当前的图像区域是否属于某个物体, 然后也能够一次性地找到矩形框的位置用于定位这个物体。

这种思路虽然直观, 但有一个致命的问题, 那就是对于一个输入图像来说, 大量的区域其实并不包含目标物体, 因此也就可以被认为是学习过程中的“负例” (Negative Instance) 。如何有效地学习这么一个“不均衡” (Imbalanced) 的数据集是这种思路需要考虑的问题。

因为这个因素, 研究者们就开始思考另外一种思路, 那就是先学习一个神经网络用于找到一些候选区域, 然后在第二个阶段根据候选区域再去最终确定物体的类别和矩形框的位置。

在最近几年的实际评测中, 基于**两个阶段** (Two-stage) 的模型, 包括我们在上一篇分享中提到的 Faster R-CNN 以及其他变种一般都有比较好的表现。而基于**一个阶段** (One-stage) 的模型, 在这篇文章发布之前还不能达到两个阶段模型的水平。

本篇文章提出了一个新的目标函数，叫作“**焦点损失**”（Focal Loss），用于取代传统的“**交叉熵**”（Cross Entropy）的目标函数。这个新目标函数的主要目的就是让一个阶段模型能够在正负例比例非常不协调的情况下，依然能够训练出较好的模型，从而使得一个阶段模型在效果上能够和两个阶段模型媲美。同时，文章还提出了一种比较简单易用的深度网络结构，可以简单地训练出整个模型。

## 论文的核心方法

在这一节，我们来讲一讲“焦点损失”的含义。因为这是一个新的目标函数，建议你还是阅读原文来理解这个目标函数的数学性质。这里，我们针对这个新的目标函数进行一个高度概括性的解释。

我们从普通的二分分类问题中常用的交叉熵，我们简称为 CE 目标函数说起。首先，我们认为模型预测类别是正例的概率是  $P$ 。CE 目标函数基本上可以认为是这个概率的对数的负数，也就是在机器学习中经常使用的“**负对数似然**”（Negative Log Likelihood）。模型的目的是最小化“负对数似然”，从而学习模型参数。

作者们观测到这么一个现象，那就是 CE 目标函数在  $P$  是一个比较大的数值时，比如大于 0.5 的时候，依然会有一个“损失”（Loss）。什么意思呢？就是说，某一个数值点，我们现在已经知道它可能是正例的可能性大于 0.5 了，也就是我们其实已经大体知道这个结果了，但是目标函数依然认为学习算法需要去对这个数据点进行作用，从而减少这个“损失”。

这其实也就是整个问题的核心，那就是传统的 CE 目标函数，并没有指导机器学习算法用在“应该使劲”的地方，而是分散到了一些原本已经不需要再去关注的数据点上。当然，这也就造成了学习困难的局面。

这篇文章提出的“焦点损失”对 CE 进行了一个看上去很小的改动，那就是在 CE 目标函数的“负对数似然”之前乘以一个“相反概率”的**系数**，并且这个系数有一个指数参数去调节这个系数的作用。如果你对该内容感兴趣，建议你参考原论文查看细节。如果对细节不感兴趣，那重点理解这个目标函数的作用就可以了。

**“焦点损失”有两个性质。**第一，当一个数据点被分错类的时候，并且这个数据点的真实概率很小，那么，损失依然和 CE 类似。当一个数据点的真实概率趋近 1，也就是原本算法就可以比较自信的时候，损失会相对于 CE 变小。第二，刚才所说的系数起到了一个调节作用，决定究竟需要对哪些“容易分类的数据点”降低损失到什么程度。

文章在新的“焦点损失”的基础上提出了一个新的网络结构叫 RetinaNet，使用一个阶段的思路来解决物体检测和语义分割的任务。这里我简要概括一下 RetinaNet 的一些特点。

第一，RetinaNet 使用了 ResNet 来从原始的输入图像中抽取基本的图像特性。

第二，文章采用了一种叫 FPN（Feature Pyramid Net）的网络架构来对图像的不同分辨率或者不同大小的情况进行特性抽取。

第三，和 Faster R-CNN 相似的，RetinaNet 也是用了 Anchor 的思想，也就是说从小的一个移动窗口中去寻找一个比较大的矩形框的可能性。

最后，RetinaNet 把从 FPN 抽取出来的特性用于两个平行的网络结构，一个用于物体分类，一个用于矩形框的定位。这一点很类似两个阶段模型的做法。

## 方法的实验效果

作者们使用 RetinaNet 在目前流行的图像物体检测任务数据集 COCO 上做了检测。首先，RetinaNet 的“平均精度”（Average Precision）要好于之前的所有一个阶段模型，初步验证了提出的目标函数和网络架构的优越性。并且，在实验中，作者们分别使用了不同的“焦点损失”指数参数来展示这个参数对于结果的重要性。同时，作者们还展示了，RetinaNet 能够比 Faster R-CNN 这种经典的两阶段模型，以及一些变种在实验结果上至少持平甚至要更好。

## 小结

今天我为你讲了 2017 年 ICCV 的最佳学生论文，这篇文章介绍了目前在图像物体识别中的最新目标函数“焦点损失”的大概内容。

一起来回顾下要点：第一，我们简要介绍了这篇文章的作者群信息。第二，我们分析了这篇文章要解决的问题和主要贡献。第三，我们详细介绍了文章提出方法的核心内容。

最后，给你留一个思考题，除了这篇文章介绍的更改目标函数的方法，针对不平衡的数据集，你觉得还有哪些通常使用的方法？

欢迎你给我留言，和我一起讨论。

---

# AI 技术内参

你的360度人工智能信息助理

洪亮劼

Etsy 数据科学主管  
前雅虎研究院资深科学家



新版升级：点击「👤 请朋友读」，10位好友免费读，邀请订阅更有**现金**奖励。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 048 | 精读2017年ICCV最佳研究论文

下一篇 050 | 如何将“深度强化学习”应用到视觉问答系统？

## 精选留言 (1)

写留言



林彦

2018-01-24



传统的机器学习针对数据集不平衡可以对比例小的过采样，比例大的欠采样，还有对目标函数根据数据比例进行缩放。