

## 146 | 计算机视觉领域的深度学习模型（一）：AlexNet

2018-09-07 洪亮劼

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 06:45 大小 3.10M



我们继续来讨论基于深度学习的计算机视觉技术。从今天开始，我们进入一个新的模块，我会结合几篇经典的论文，给你介绍几个专门为计算机视觉而提出来的深度学习模型。这些模型都在最近几年的深度学习发展中，起到了至关重要的作用。

我们这个系列要分享的第一篇论文题目是《基于深度卷积神经网络的图像网络分类》(ImageNet Classification with Deep Convolutional Neural Network) [1]。因为这篇文章的第一作者名字叫 Alex，所以文章提出的模型也经常被称为**AlexNet**。

那接下来我们就先介绍一下这篇论文的作者群。

第一作者就是亚力克斯·克里切夫斯基 (Alex Krizhevsky)。发表这篇论文的时候他在多伦多大学计算机系攻读博士学位；之后的 2013~2017 年间在谷歌任职，继续从事深度学习

的研究。

第二作者叫伊利亚·苏兹克维（Ilya Sutskever）。发表这篇论文的时候，苏兹克维也在多伦多大学计算机系攻读博士学位；之后到斯坦福大学跟随吴恩达做博士后研究。2013~2015 年间，他在谷歌担任研究科学家一职。2016 年之后，他参与共同创立了 OpenAI 并且担任研究总监这一职位。苏兹克维在深度学习方面已经发表了很多篇论文，目前论文的引用数已经超过 7 万次。

最后一位作者是杰弗里·辛顿（Geoffrey Hinton）。对于辛顿，我们就比较熟悉了，他是多伦多大学计算机系的教授，是机器学习，特别是深度学习的学术权威。可以说，几十年来，辛顿都在以神经网络为代表的深度学习领域深耕，即便是在其他学术思潮涌动的时候，他都能够坚持在深度学习这一领域继续钻研，这种精神让我们钦佩。

## 论文的主要贡献

如何来描述这篇论文的主要贡献呢？简而言之，这篇论文**开启了深度学习在计算机视觉领域广泛应用的大门**。通过这篇论文，我们看到了深度学习模型在重要的计算机视觉任务上取得了非常显著的效果。

具体来说，在 ImageNet 2012 年的比赛中，文章提到的模型比第二名方法的准确度要高出十多个百分点。能够达到这个效果，得益于在模型训练时的一系列重要技巧。这篇论文训练了到当时为止最大的卷积神经网络，而这些技巧使得训练大规模实用级别的神经网络成为可能。

## 论文的核心方法

要了解 AlexNet 的一些核心方法，我们就需要简单提一下 ImageNet 竞赛的数据集。这个数据集在当时有大约 120 万张训练图片，5 万张验证图片和 15 万张测试图片。这些图片属于 1000 个类别。这个数据集在当时来说应该算是无可争议的大型数据集。为了能够方便地处理这些图片，作者们把所有图片的分辨率都重新调整到了“256\*256”。AlexNet 直接在这些图片的 RGB 像素点上建模。

整个模型的架构是怎样的呢？AlexNet 一共包含 8 层网络结构，5 层全联通层（也就是前馈神经网络）。这 8 层网络架构总体来说是逐渐变小的一个趋势，也就是说每一层提取的信息越来越呈现高度的概括性。

那么在整个架构中，这篇文章提出的模型有哪些独到之处呢？

第一，**AlexNet 采用了“线性整流函数”（ReLU）来作为激活函数**。虽然这个选择在今天看来可以说是非常平常，甚至已经成为了神经网络建模的默认选项。但这个选择在当时还是很大胆的一种创新。这个创新带来了训练时间的大幅度减少，同时还能保持，甚至提升了模型性能。

第二，**整个模型的训练大量采用了 GPU，并且使用了多个 GPU 来进行计算**。这一点就在速度上和模型的大小上彻底解放了模型的选择。以前仅仅利用单个 GPU 的方式，没办法把所有的训练数据都放入一个 GPU 上。

第三，作者们介绍了一种叫作“**局部响应归一化**”（Local Response Normalization）的方法来对每层之间的单元进行归一。

如何进行最有效的归一，以及这些归一化有什么作用，这些问题一直都是深度学习研究面临的重要课题。从实际的使用上来看，这种局部响应归一化的方法在几年之后让位给了其他更为主流的归一方法。但是从这一个技术要点来看，**我们要想把深度学习模型真正应用到实际场景任务中，归一化是一个必不可少的组件**。

第四，作者们在 AlexNet 里面使用了所谓的“**重叠池化**”（Overlapping Pooling）这种方法。在普通的卷积神经网络中，“池化”的作用是从周围的单元中总结出必要的信息。一般来说，池化的过程中，并不重复覆盖相同的单元。也就是说，池化不会重叠。而在这篇论文中，作者们发现重叠池化能够降低错误率，虽然非常微量但是很重要。这个组件在之后的发展中并不多见。

除了在网络架构上的一些创新之外，AlexNet 的训练过程中最需要注意的是防止“**过拟合**”（Overfitting）。在很长的一段时间里，我们没有办法把深度神经网络模型应用在实际场景中，一个很重要的原因就是过拟合。可以说，如何防止神经网络模型过拟合，这个问题让研究人员伤透了脑筋。

所谓过拟合就是说模型在训练集上工作得很好，但是无法“**泛化**”（Generalization）到测试集，也就是没有出现过的数据上。无法泛化其实也就证明训练的模型对未知数据的预测能力很差。

这篇论文中主要提到了两种防止过拟合的方法。

第一种思路叫“**数据增强**”（Data Augmentation）。简单来说，这里的思路其实就是增加“虚拟数据”来增加数据的多样性，从而能够让模型更加健壮。那虚拟数据是怎么来的？虚拟数据其实来源于真实数据的变形。

第二种思路就是 **Dropout**。这种方法在当时看显得很随便，就是在训练的时候随机把一些单元置零。作者们发现在这样随机置零的过程后，模型会变得更加稳定。值得一提的是，Dropout 已经成为了这几年深度学习的一个标配。

## 小结

今天我为你讲了第一篇基于深度学习的经典论文，讨论了 AlexNet 这个模型，这个模型开启了深度学习全面进军计算机视觉领域的时代。

一起来回顾下要点：第一，AlexNet 模型在 ImageNet 2012 竞赛中胜出，让我们看到了深度学习模型在计算机视觉领域中所取得的显著效果；第二，我们讨论了 AlexNet 模型的四大创新之处，以及论文提出的两种防止过拟合的方法。

最后，给你留一个思考题，站在现在的立场，AlexNet 在当时的成功是否给你一些启发呢？

欢迎你给我留言，和我一起讨论。

## 参考文献

1. **ImageNet Classification with Deep Convolutional Neural Networks**, Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, NIPS, 2012.

# 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 145 | 基于深度学习的计算机视觉技术（三）：深度学习模型的优化

下一篇 147 | 计算机视觉领域的深度学习模型（二）：VGG & GoogleNet

## 精选留言

写留言

由作者筛选后的优质留言将会公开显示，欢迎踊跃留言。