

[首页](#)[资讯](#)[深度资源](#)[产业视频](#)[GMIS峰会](#)[AI 商用搜索](#)[个人中心](#)[我的主页](#)[注销](#)

吴恩达deeplearning.ai第四课学习心得：卷积神经网络与计算机视觉

By 路雪 2017年11月22日 12:09

不久前，Coursera 上放出了吴恩达 deeplearning.ai 的第四门课程《卷积神经网络》。本文是加拿大国家银行首席分析师 Ryan Shrott 在完成该课程后所写的学习心得，有助于大家直观地了解、学习计算机视觉。



我最近在 Coursera 上完成了吴恩达教授的计算机视觉课程。吴恩达非常准确地解释了很多优化计算机视觉任务需要了解的复杂概念。我最喜欢的部分是神经风格迁移部分（第 11 课），你可以将莫内的绘画风格和你喜欢的任何图像结合起来创作自己的艺术作品。示例如下：



本文，我将讨论在该课程中学到的 11 堂重要的课。注意：这是 deeplearning.ai 发布的深度学习专门化的第四课。如果你想学习前三课，推荐以下文章：

- [入门 | 吴恩达 Deeplearning.ai 课程学习心得分享](#)
- [吴恩达 Deeplearning.ai 课程学习全体验：深度学习必备课程（已获证书）](#)

第 1 课：为什么计算机视觉发展迅速

大数据和算法的发展使智能系统的测试误差收敛至贝叶斯最优误差。这使得 AI 在很多领域中超越人类水平，包括自然感知任务。TensorFlow 的开源软件允许你使用迁移学习实现目标检测系统，可快速检测任意目标。有了迁移学习，你只需 100-500 个样本就可以使系统运行良好。手动标注 100 个样本工作量不是很大，因此你可以快速获得一个最小可行产品。

第 2 课：卷积的工作原理

吴恩达解释了如何实现卷积算子，并展示了它如何对图像进行边缘检测。他还介绍了其他滤波器，如 Sobel 滤波器，它赋予边缘的中心像素更多权重。吴恩达还介绍道：滤波器的权重不应该手动设计而应使用爬山算法（如梯度下降）学得。

第 3 课：为什么使用卷积网络？

关于为什么卷积网络在图像识别任务中表现如此好这个问题，吴恩达给出了多个哲学原因。他介绍了两个具体原因。一，参数共享（parameter sharing）：在图像某一部分有效的特征检测器对另一部分可能也有效。例如，边缘检测器可能对图像的很多部分都有用。参数共享需要较少的参数数量和较鲁棒的平移不变性。

第二个原因是连接的稀疏性：每个输出层只是少量输入的函数（尤其是，滤波器的尺寸）。这大幅降低了网络的参数数量，加快了训练速度。

第 3 课：为什么使用 Padding？

Padding 通常用于保证卷积过程中输入和输出张量的维度是一样的。它还可以使图像边缘附近的帧对输出的贡献和图像中心附近的帧一样。

第 4 课：为什么使用最大池化？

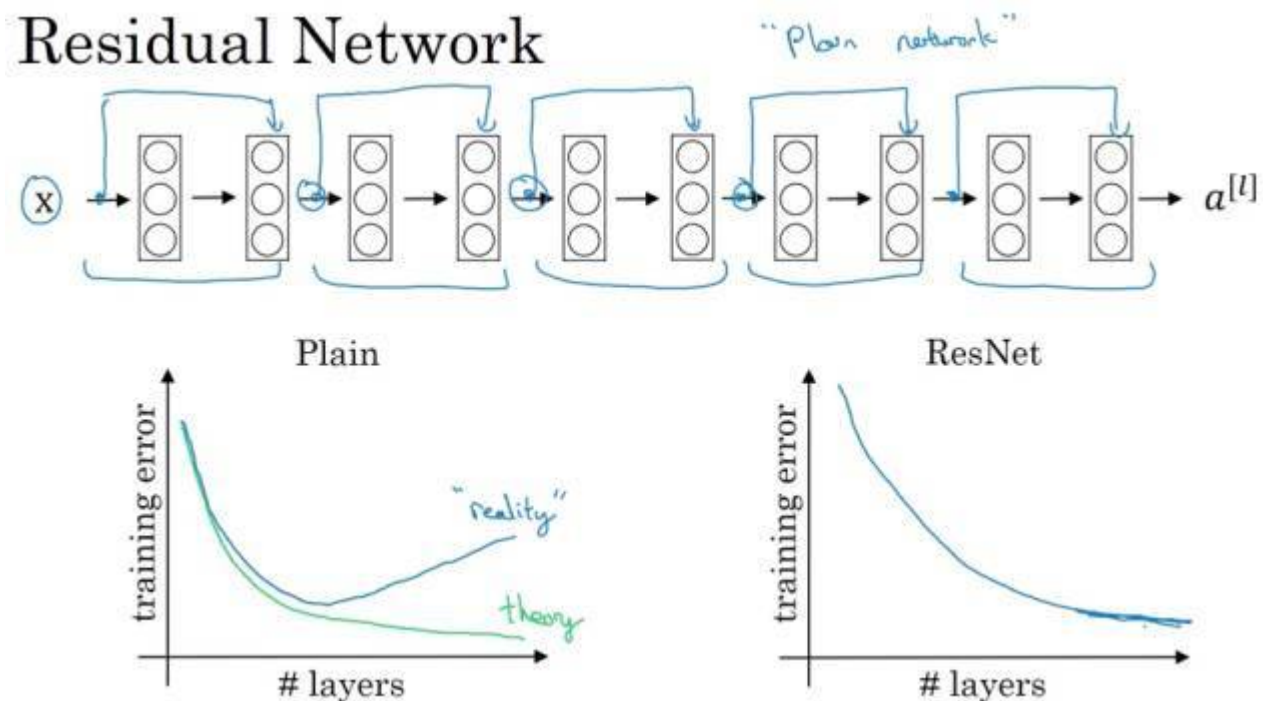
实证研究表明最大池化在卷积神经网络中非常有效。通过对图像进行下采样，我们可以减少参数数量，使特征在缩放或方向变化时保持不变。

第 5 课：经典的网络架构

吴恩达展示了 3 种经典的网络架构：LeNet-5、AlexNet 和 VGG-16。他的主要思想是：高效网络中层的通道规模不断扩大，宽度和高度不断下降。

第 6 课：为什么 ResNets 有效？

对于普通的网络来说，由于梯度下降或爆炸，训练误差并不是随着层数增加而单调递减的。这些网络具备前馈跳过连接（feed forward skipped connection），在性能不出现下降的情况下训练大型网络。



第 7 课：使用迁移学习

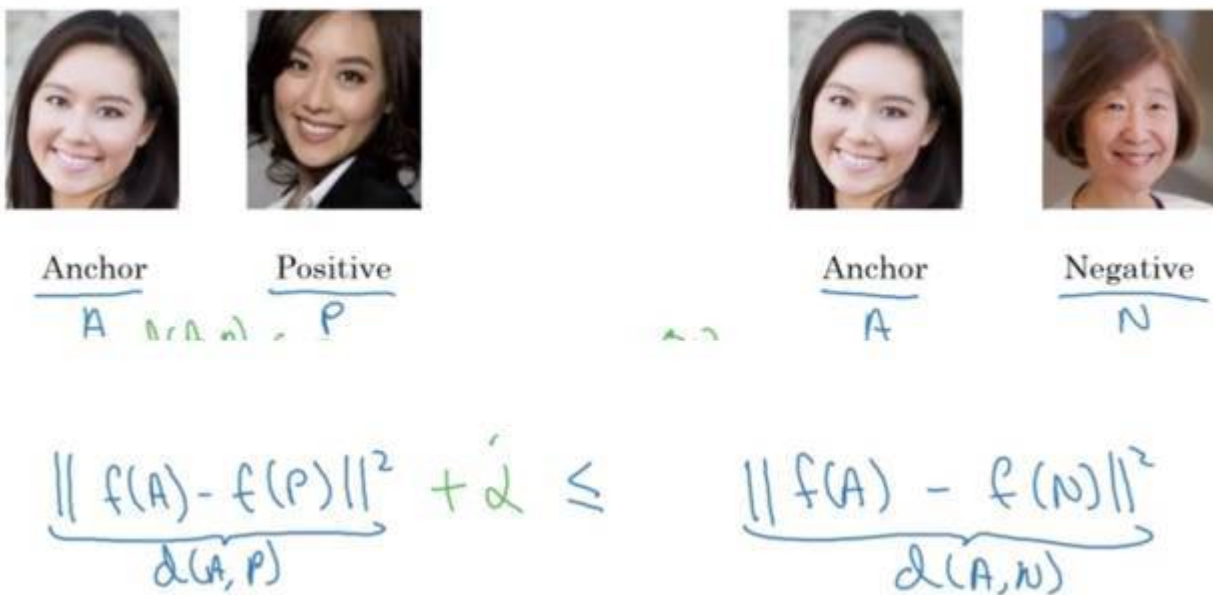
使用 GPU 从头开始训练大型神经网络，如 Iception 可能需要数周时间。因此，我们需要下载预训练网络中的权重，仅仅重训练最后的 softmax 层（或最后几层），以减少训练时间。原因在于相对靠前的层倾向于和图像中更大的概念相关——如边缘和曲线。

第 8 课：如何在计算机视觉竞赛中获胜

吴恩达认为我们应该独立训练一些网络，平均它们的输出结果以获取更好的性能。数据增强技术——如随机裁剪图片、水平翻转和垂直轴对称调换也可以提升模型性能。最后，你需要从开源的实现和预训练模型上开始，针对目标应用逐渐调整参数。

第 9 课：如何实现目标检测

吴恩达首先解释了如何进行图像关键点检测。基本上，这些关键点距离训练输出的范例很远。通过一些巧妙的卷积操作，你会得到一些输出量，其中显示目标在特定区域的概率以及目标的位置。他还解释了如何使用交并比（Intersection-over-Union, IoU）公式评估目标检测算法的有效性。最后，吴恩达将所有这些内容整合在一起，介绍了著名的 YOLO 算法。



第 10 课：如何实现人脸识别

人脸识别是 one-shot 学习问题，通常，你只有一张图像来识别这个人。解决的办法是学习一个相似性函数，给出两个图像之间的差异程度。所以，如果图像是同一个人，函数的输出值较小，而不同的人则相反。

吴恩达给出的第一个解决方案叫作 siamese 网络。该方法将两个人的图像输入到同一个网络中，然后比较它们的输出。如果输出类似，则是同一个人。该神经网络的训练目标是：如果输入图像是同一个人

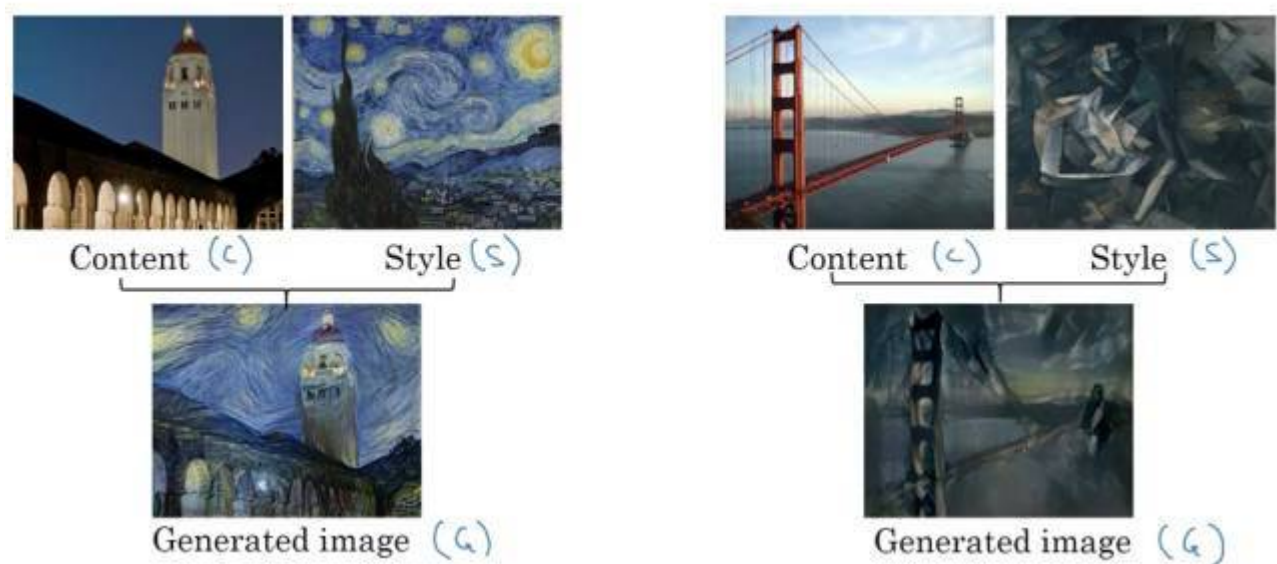
人的，则编码距离相对较小。

他给出的第二个解决方法是三重损失法（triplet loss method）。这个想法的核心在于图片的三个维度（Anchor（A）、Positive（P）和 Negative（N））在训练过后，A 和 P 的输出距离大大小于 A 和 N 的输出距离。

第 11 课：如何使用神经风格迁移创造艺术品

吴恩达在课程中解释了如何结合图像内容和风格生成全新图像。如下图所示：

Neural style transfer



神经风格迁移的关键是理解卷积神经网络每个网络层中学习的视觉表征。前面的层学习简单的特征，如边缘，后面的特征学习复杂的物体，如脸、脚、汽车。

为了构建神经风格迁移图像，你简单地定义一个代价函数，即内容图像和风格图像中相似点结合后的凸函数。具体来说，该代价函数可表示为：

$$J(G) = \alpha * J_{\text{content}}(C, G) + \beta * J_{\text{style}}(S, G)$$

其中，G 是生成的图像，C 是内容图像，S 是风格图像。该学习算法简单地使用梯度下降最小化与生成图像 G 有关的代价函数。

步骤如下：

1. 随机生成 G 。
2. 使用梯度下降最小化 $J(G)$ ，也就是 $G := G - dG(J(G))$ 。
3. 重复步骤 2。

结论

通过完成该课程，你会对大量计算机视觉文献有个直观的理解。课程中的作业也会让你动手实现这些方法。虽然完成这门课后，你不能直接成为计算机视觉领域的专家，但这门课程可能成为你在计算机视觉领域开启一种思路或者事业的一个踏板。我的GitHub 地址：<https://github.com/ryanshrott>。



原文链接：<https://towardsdatascience.com/computer-vision-by-andrew-ng-11-lessons-learned-7d05c18a6999>

声明：本文由机器之心编译出品，原文来自Medium，转载请查看要求，机器之心对于违规侵权者保有法律追诉权。



说点什么....

[提交评论](#)



[关于我们寻求报道商务合作服务条款](#)

©2017版权所有 机器之心（北京）科技有限公司

京 ICP 备 12027496

全球人工智能信息服务

[友情链接](#)

[Synced Global](#)[机器之心](#) [Medium](#) [博客PaperWeekly](#)[网易智能](#)[动脉网](#)[硬蛋网](#)



联系电话: +86 010-57150141

联系邮箱: contact@jiqizhixin.com