

# 本科生毕业设计（论文）参考文献译文本

译文出处：Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]. Advances in neural information processing systems. 2012, 25(2): 1097-1105.

院 系 软件学院

专业班级 软件工程1205

姓 名 陈吕劼

学 号 U201217478

指导教师 管乐

2016年1月

**译文要求**

1. 译文内容须与课题（或专业内容）联系，并需在封面注明详细出处。
2. 出处格式为

图书：作者.书名.版本（第×版）.译者.出版地：出版者，出版年.起页～止页  
期刊：作者.文章名称.期刊名称，年号，卷号（期号）：起页～止页

1. 译文不少于5000汉字（或2万印刷符）。
2. 翻译内容用五号宋体字编辑，采用A4号纸双面打印，封面与封底采用浅蓝色封面纸（卡纸）打印。要求内容明确，语句通顺。
3. 译文及其相应参考文献一起装订，顺序依次为封面、译文、文献。
4. 翻译应在第七学期完成。

**译文评阅**

|  |
| --- |
| **导师评语**  应根据学校“译文要求”，对学生译文翻译的准确性、翻译数量以及译文的文字表述情况等做具体的评价后，再评分。 |
| 评分：\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_（百分制） 指导教师（签名）：\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  2016年1月22日 |

**基于深层卷积神经网络的ImageNet图像分类**

**摘要**

我们使用了大型深层卷积神经网络将2010年ImageNet LSVRC比赛中的120万张高分辨率图像分类成1000个不同的类别。在测试数据集上，我们分别在首个匹配和前五匹配准则下达到了37.5%和17.0%的错误率，而这个结果相比于以往最好的结果有了相当大的提升。我们的神经网络包含有6000万个参数和650000个神经单元，由5个卷积层和3个全连接层以及一个1000维的softmax分类层的组成，部分卷积层后还跟有最大池化层。为了加快学习速度，我们使用了非饱和神经元和利用GPU实现的非常快速的卷积操作。为减少全连接层中的过拟合问题，我们采用了最新开发被证明非常有效的名为“dropout”的正则化方法。我们也在2012年ILSVRC比赛中使用了这个模型的变种并获得优胜，相比于第二名26.2%的错误率，我们前五匹配准则下的错误率为15.3%。

1. **简介**

当前物体识别的实现都充分利用了机器学习方法，为了提升它们的性能，我们可以收集大量的数据，选择更强大的学习模型以及使用更好的技术来预防过拟合。直到最近，被标记的图像数据集依旧较小，大约成千上万张图片左右(例如., NORB [16], Caltech-101/256 [8, 9], and CIFAR-10/100 [12])。在这样大小的数据集下，简单识别任务可以很好得被处理，尤其是在通过了类别不变转换的增强后，比如，当前对于MNIST数字识别任务最好处理的错误率（小于0.3%）已经达到人类的水平[4]。但是，在现实情况中的物体则有相当大的变化，所以学习并识别它们需要更庞大的训练集。而且实际上，数据集太小的缺点已经被广泛认可(例如,Pintoetal.[21])，但是直到最近收集数以百万计被标记的图像才成为可能。最新更大型的数据集包括LabelMe[23]，拥有成千上万的完全划分图像，以及拥有超过22000个种类1500万张高像素图像的ImageNet[6]。

为了在上百万张图片中学习上千种类别的物体，我们需要一个拥有强大学习能力的模型。然而，物体识别任务的巨大复杂性使得即使在像ImageNet这样庞大的数据集下问题都难以解决，所以我们的模型需要一些先进的知识来弥补我们所缺失的数据。卷积神经网络构成了这样一类模型[16, 11, 13, 18, 15, 22, 26]，它的能力随网络的深度和广度变化而改变。同时，它能对图像的性质做出强大和大致正确的假设。因此，相比较于同样规模的标准前馈神经网络，卷积神经网络拥有更少的来连接和参数，所以更易训练，同时它理论上最佳性能也只有略微变差。

尽管卷积神经网络有着吸引人的特质以及相对高效的网络结构，但是它应用在大型高分辨率图像上依旧十分耗费资源。幸运的是，如今的GPU已经配备了对二维卷积的高度优化实现以致足以支持大型卷积神经网络的训练，而且例如ImageNet的现代数据集也包含足够多的带标签的图像来避免训练模型中的严重过拟合。

本文具体的贡献如下：我们利用2010和2012年ILSVRC比赛中的ImageNet子训练集，训练了一个至今为止最大的卷积神经网络，并且获得了到目前为止在数据集上发表过的最好的成绩。我们写了一个对于二维卷积和其他所有训练卷积神经网络操作的高度优化的GPU实现，而这些都是开源可用的。在第3节，我们会详细介绍网络包含的一系列新鲜且特别的提高性能和减少训练时间的特性。即使拥有120万张图片的训练集，网络规模带来的过拟合依然成为一个重要的问题，所以在第四节，我们会介绍一些有效的技术来预防过拟合。我们最终的网络包含5个卷积层和3个全连接层，而且它的深度似乎非常重要。我们发现无论去除任何一个卷积层（每一层都包含不超过1%的模型参数），都会导致性能较差。

最后，网络的规模主要受限于目前GPU的可用内存和我们愿意容忍的训练时间量。在使用两个GTX 580 3GB GPU的情况下，我们的网络训练了五到六天的时间。我们所有的实验表明更快的GPU和更大的数据集就能简单得使我们的结果变好。

1. **数据集**

ImageNet是一个拥有大约22000个种类超过1500万标记好高质量图片的数据集。这些图片从网络上收集并通过亚马逊的Mechanical Turk众包工具被人标记。从2010年开始，作为帕斯卡可视对象挑战的一部分，名为ImageNet大型视觉识别挑战（ILSVRC）的竞赛每年举办一次。这项比赛使用ImageNet的一个子数据集，其中包含1000个种类各大约1000张图片总共大约120万张训练图片和50000张验证图像以及150000张测试图片。

2010年的ILSVRC比赛是唯一一届测试数据集标记可用的，所以我们在此基础上做了大多数的实验。由于我们同样参加了2012年的ILSVRC比赛，我们在第6节也呈现了这一年数据集的结果，而这一年的测试数据的标记是未知的。在ImageNet上，习惯呈现两个错误率：首次匹配错误率和前五匹配错误率。前五匹配错误率是指正确的标记不在模型给出的前五种可能的标记之中的概率。

虽然ImageNet的图像分辨率很高，但我们的系统要求一个固定的输入维数。因此，我们对图像进行采样得到固定大小256 x 256的图像。对于给定的矩形图像，我们首先将较短的边缩小到256像素长，然后裁剪正中间的256 x 256大小的一块作为结果。除了减去每个像素在训练过程中的平均活度，我们并未使用任何其他方法来预处理图像，所以都是直接使用每个像素的RGB值来训练网络。

1. **体系结构**

图二显示了我们网络的结构。它包含八个学习层——五个卷积层和三个全连接层。接下来，我将介绍一些关于网络结构的新颖或特别的特性。3.1到3.4节将根据重要性从高到低罗列这些特性。

3.1 ReLU非线性

根据输入x，建立一个神经元的激活函数f的一般方式为或者。考虑到使用梯度下降方法的训练时间，这些饱和非线性方法远慢于非饱和非线性方法。根据Nair和Hinton的研究成果[20]，我们侵向于使用整合线性单元（ReLU）非线性激活函数的神经单元。在训练深层卷积神经网络时使用ReLU要比使用tanh快上好几倍。从图1中显示了在CIFAR-10数据集上使用一个四层的卷积神经网络训练时达到25%错误率时的周期次数。这幅图显示出我们如果使用传统的饱和神经元模型是无法训练如此庞大的神经网络的。

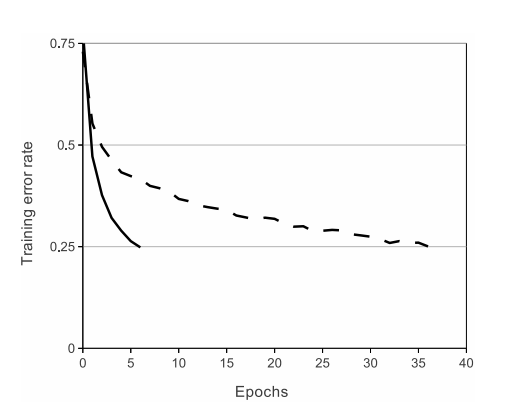


图1: 在一个四层卷积层的神经网络上对CIFAR-10使用ReLUs（实线）达到25%训练错误率是在同等网络上使用tanh神经元（虚线）速度的6倍。每一个神经网络的学习速率都是相互独立且尽可能得快并且没有使用任何的正规化方法。虽然对于不同的神经网络这种效果的大小各异，但是使用ReLU方法训练都要比同等环境下的饱和神经元方法快上几倍。

我们并不是最先想到要替换卷积神经网络中传统神经元的人。例如，Jarrett等人[11]宣称非线性方法在Caltech-101数据集上配合他们对应的正则化方法以及本地平均池化方法效果很好。然而，在这个数据上最主要的考虑是要避免过拟合，所以他们所努力探索的和我们之前提到使用ReLU来提升适应数据集的能力有所不同。更快的学习速度对于在大型数据集上使用大型模型训练的性能有很影响。

3.2 多GPU训练

一个GTX 580 GPU只拥有3GB的内存，这会限制可训练的模型的最大规模。事实证明120万张图片是足以训练神经网络的，但是这对于单一GPU来说规模太大，因此我们将网络拆分到两个GPU上。如今的GPU因为可以相互直接读取和写入对方的内存而不经过主机的机器内存，所以非常适合于跨GPU的并行运算。并行策略是将内核（或神经元）平分到每个GPU上，还有一个小策略是GPU仅在某些特定层通信。这意味着，例如第三层内核接受所有来自第二层的内核为输入，然而第四层内核只接受在同一个GPU上的第三层的内核为输入。选择连接模式是一个交叉验证的问题，但是这个使得我们可以精准地调整所有的通信直到所有结果的满足要求。

1. **降低过拟合**

××××××××××××××××××××××××，其×××××可表示如下：

 (2-1）

 (2-2）

×××××××××××××××××××××××××××× (如表2-1所示)

表2-1□××××××××××

|  |  |  |  |
| --- | --- | --- | --- |
| ××××× | ××× | ××× | ××× |
| ××××× | ×× | ×× | ×× |
| ××××× | ×× | ×× | ×× |
| ××××× | ×× | ×× | ×× |
| ××××× | ×× | ×× | ×× |

（表标题：位于表格上方，宋体5号，字母、阿拉伯数字为Time New Roman 5号，表内容：宋体5号，字母、阿拉伯数字为Time New Roman 5号）

（“□”表示空格）

××××××××××××××××××××××××××× (如图2-1所示)



图2-1□××××××××××

（图标题：位于图下方，宋体5号，字母、阿拉伯数字为Time New Roman 5号）

.......

.......

.......

.......

**参考文献**（宋体5号加粗）

[1]□王静康,张凤宝,夏淑倩等.论化工本科专业国际认证与国内认证的“实质性”.高等工程教育研究,2014,5:1-4

[2]□Stone J A, Howard L P. A simple technique for observing periodic nonlinearities in Michelson interferometers. Precision Engineering,1998,22(4):220-232

[3]□朱印红,袁衍明.Dreamweaver完美网页设计——技术入门篇.(第一版).北京:中国电力出版社,2006:19～20

[4]□Lewis S L. Physics and chemistry of the solar system.北京:北京大学出版社,2014.1～2

[5]□陈剑.上博简《民之父母》“而得既塞於四海矣”句解释[EB/OL］.简帛研究网站，http://www.bamboosilk.org/Wssf/2003/chenjian03.htm．2003-01-18

( 宋体5号)

（可加页，A4纸双面打印）

**参考文献原文**