**(一)基于深卷积神经网络的 ImageNet 分类**

**ImageNet Classification with Deep Convolutional**

**Neural Networks**

**--Alex Krizhevsky,Ilya Sutskever,Geoffrey E. Hinton、NIPS2012（神经信息处理系统进展大会）**

**一、科学问题**

**1.1 本文所涉及科学问题**

目标识别，把图像分类。

训练了大型深度卷积神经网络，将ImageNet LSVRC-2010比赛中的120万个高分辨率图像分类为1000个不同的类别。

**1.2 同行专家如何解决**

目前的对象识别方法主要是使用机器学习方法。

**1.3 本文所解决的问题**

降低目标识别，图像分类的错误率

为了使训练更快，本文使用了非饱和神经元和卷积运算的非常有效的GPU实现。为了减少全连接层中的过度拟合，本文采用了最近开发的称为“压差”的正则化方法，该方法被证明是非常有效的。

**1.4 本文解决方案效果**

在测试数据上，本文实现了前三和前五的错误率分别为37.5％和17.0％，这比之前的最新技术水平要好得多。

在ILSVRC-2012竞赛中获得了从以前的26.2％降低到15.3％的前5个测试错误率。

**二、研究内容**

**2.1 理论与方法介绍**

本文需要一个具有大量学习能力的模型。卷积神经网络（CNN）构成了一类这样的模型[16,11,13,18,15,22,26]。它们的容量可以通过改变它们的深度和宽度来控制，并且它们还对图像的性质（即统计的平稳性和像素依赖性的局部性）做出强有力且大部分正确的假设。因此，与具有类似大小的层的标准前馈神经网络相比，CNN具有更少的连接和参数，因此它们更容易训练，而它们的理论上最佳性能可能仅略微更差。

**2.2 验证分析与实验效果**

本文的网络实现了前三名和前五名测试集错误率分别为37.5％和17.0％5。在ILSVRC2010竞赛期间取得的最佳表现为47.1％和28.2％，此后最佳公布的结果分别为45.7％和25.7％

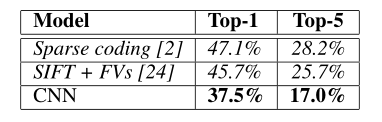


表1：ILSVRC2010测试集的结果比较。斜体是其他人获得的最佳结果。

本文还参加了ILSVRC-2012竞赛中的模型，并在表2中报告了本文的结果。对表2的5个CNN在整个2011年秋季发布预训练的两个CNN的预测进行平均后，误差率为15.3％而第二好的比赛项目实现了26.2％的错误率。

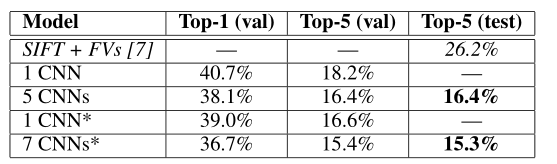


表2：ILSVRC-2012验证和测试集的错误率比较。斜体是其他人获得的最佳结果。带有星号\*的型号经过“预训练”，可对整个ImageNet 2011秋季版本进行分类。

**三、论文存在问题及后续研究重点**

**3.1 论文存在问题**

如果移除单个卷积层，本文的网络性能会下降。例如，删除任何中间层会导致网络的前1个性能损失约2％。所以深度对于实现本文的结果非常重要。

**3.2 后续研究重点**

本文希望在未来的视频序列上使用非常大且深度的卷积网络，其中时间结构这些信息在静态图像中缺失或不那么明显时提供了非常有用的信息。

**四、该问题相关研究成果**

**4.1 相关论文一**

**（1）题目**：Large scale visual recognition challenge 2010.

大规模视觉识别挑战2010。

**（2）作者介绍**：Olga Russakovsky authorJia DengHao SuJonathan KrauseSanjeev SatheeshSean MaZhiheng HuangAndrej KarpathyAditya KhoslaMichael BernsteinAlexander C. BergLi Fei-Fei

**（3）摘要**: ImageNet大规模视觉识别挑战是对象类别分类和检测数百个对象类别和数百万图像的基准。从2010年到现在每年都在进行挑战，吸引了来自50多个机构的参与。本文描述了这个基准数据集的创建以及结果可能导致的对象识别的进展。本文讨论了收集大规模地面实况注释的挑战，突出了分类对象识别的关键突破，详细分析了大规模图像分类和目标检测领域的现状，并比较了状态。 - 人类精确度的计算机视觉准确性。本文总结了挑战5年中的经验教训，并提出了未来的方向和改进。

**4.2 相关论文二**

**（1）题目**：Multi-column deep neural networks for image classification

用于图像分类的多列深度神经网络

**（2）作者介绍**：Dan Ciregan ; Ueli Meier ; Jürgen Schmidhuber

**（3）摘要**：传统的计算机视觉和机器学习方法无法将人类的表现与诸如识别手写数字或交通标志等任务相匹配。本文的生物学合理，广泛和深入的人工神经网络架构可以。卷积赢家 - 通吃神经元的小（通常是最小的）感受域产生大的网络深度，导致与视网膜和视觉皮层之间的哺乳动物中发现的稀疏连接的神经层大致一样多。只有获胜的神经元才会被训几个深层神经专栏成为以不同方式预处理的输入专家;他们的预测是平均的。显卡允许快速培训。在极具竞争力的MNIST手写基准测试中，本文的方法是第一个实现近人类表现的方法。在交通标志识别基准测试中，它比人类高出两倍。本文还在众多常见的图像分类基准上改进了最新技术

**4.3 相关论文三**

**（1）题目**：Convolutional networks and applications in vision

卷积网络和视觉应用

**（2）作者介绍**：Yann LeCun ; Koray Kavukcuoglu ; Clement Farabet

**（3）摘要**：智能任务，如视觉感知，听觉感知和语言理解，需要构建世界（或“特征”）的良好内部表征？在保留相关信息的同时，必须对输入的无关变化保持不变。机器学习的一个主要问题是如何自动学习这些好的功能。卷积网络（ConvNets）是一种具有生物启发性的可训练架构，可以学习不变的特征。 ConvNets中的每个阶段都由滤波器组，一些非线性和特征池层组成。通过多个阶段，ConvNet可以学习多级功能层次结构。虽然ConvNets已成功部署在从OCR到视频监控的许多商业应用中，但它们需要大量标记的训练样本。本文描述了新的无监督学习算法，以及允许使用非常少的标记样本训练ConvNets的新的非线性阶段。描述了用于越野移动机器人的视觉对象识别和视觉导航的应用

**(二)** **可视化和理解卷积网络**

**Visualizing and Understanding Convolutional NetworksMatthew D. Zeiler、Rob Fergus，Procedia Computer Science**

**一、科学问题**

**1.1 本文所涉及科学问题**

（填写所读论文涉及的科学问题，所谓科学问题一般为本论文的大方向）

大型卷积网络模型在ImageNet基准测试中表现出了令人印象深刻的分类性能（Krizhevsky等，2012）。然而，没有明确的理解他们为什么表现如此之好，或者如何改进它们

**1.2 同行专家如何解决**

（个人感觉主要为国内外当前研究现状，一般在论文的引言中都有介绍。）

Krizhevsky等人在ImageNet 2012分类基准测试中表现出创纪录的表现，主要方法有：（i）有更大规模的培训集，有数百万个带标签的例子; （ii）强大的GPU实现，使非常大的模型的训练变得实用，以及（iii）更好的模型正则化策略，例如Dropout（Hinton等，2012）。

**1.3 本文所解决的问题**

（解决什么问题，效果如何，一般摘要及总结中都会有）

本文介绍了一种新颖的可视化技术，可以深入了解中间特征层的功能和分类器的操作。这些可视化用于诊断角色，使本文能够找到优于Krizhevsky等人的模型架构。在ImageNet分类基准上。本文还对不同模型层进行磨损消耗研究，以发现不同模型层的性能贡献。

**1.4 本文解决方案效果**

（使用本文的方法可以达到怎样的效果，一般在总结中会提到）

本文以多种方式探索了大型卷积神经网络模型，这些模型经过了图像分类训练。首先，本文提出了一种可视化模型中活动的新方法。这表明这些特征远非随机，无法解释的模式。相反，当本文提升层次时，它们显示出许多直观上令人满意的属性，例如组合性，增加不变性和类别歧视。本文还展示了如何使用这些可视化来调试模型的问题以获得更好的结果，例如改进Krizhevsky等人。 （Krizhevsky等，2012）令人印象深刻的ImageNet 2012结果。然后，本文通过一系列遮挡实验证明，该模型虽然经过分类训练，但对图像中的局部结构非常敏感，并且不仅仅使用广泛的场景环境。对该模型的消融研究表明，对网络而言，最小深度，而不是任何单个部分，对模型的性能至关重要。

最后，本文展示了ImageNet训练模型如何能够很好地推广到其他数据集。对于Caltech-101和Caltech-256，数据集足够相似，本文可以击败报告的最佳结果，在后一种情况下可以获得相当大的优势。

**二、研究内容**

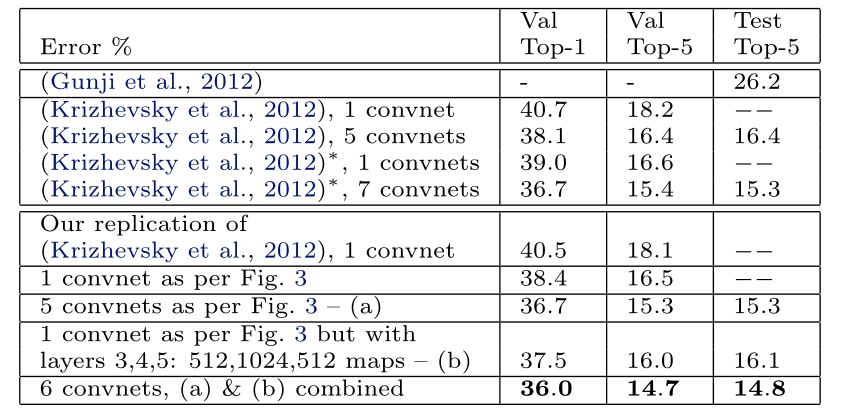
**2.1 理论与方法介绍**

（论文主要研究内容的提出，主要技术路线、理论与方法介绍）

在本文中，本文介绍了一种可视化技术，该技术揭示了激发模型中任何层的单个特征映射的输入激励。它还允许本文在训练期间观察特征的演变并诊断模型的潜在问题。本文提出的可视化技术使用多层反卷积网络（deconvnet），如（Zeiler等人，2011）所提出的，将特征激活投射回输入像素空间。本文还通过遮挡输入图像的部分来执行分类器输出的灵敏度分析，从而揭示场景的哪些部分对于分类是重要的。

**2.2 验证分析与实验效果**

（论文中的实验分析和实验效果）



IMAGENET 2012分类错误率。表示在Imagenet 2011和2012培训集上培训的模型。

**三、论文存在问题及后续研究重点**

**3.1 论文存在问题**

本文的预测模型不太适用于PASCAL数据，可能遭受数据集偏差（Torralba＆Efros，2011），尽管它仍然在最佳报告结果的3.2％之内，尽管没有调整任务。

**3.2 后续研究重点**

允许每个图像有多个对象的不同损失函数，使性能提高。使网络也能够解决对象检测问题。

**四、该问题相关研究成果**

**4.1 相关论文一**

**（1）题目**：Greedy Layer-Wise Training of Deep Networks深度网络的贪婪层智能培训

**（2）作者介绍**：Yoshua Bengio, Pascal Lamblin, Dan Popovici, Hugo Larochelle

**（3）摘要**: 电路的复杂性理论强烈表明，在表示某些功能所需的计算元素方面，深层架构比浅层架构更有效（有时是指数级）。深层多层神经网络具有许多级别的非线性，允许它们紧凑地表示高度非线性和高度变化的函数。然而，直到最近还不清楚如何训练这样的深度网络，因为从随机初始化开始的基于梯度的优化似乎经常陷入不良解决方案中。 Hinton等人。最近为深度信念网络（DBN）引入了一种贪婪的分层无监督学习算法，这是一种具有多层隐藏因果变量的生成模型。在上述优化问题的背景下，本文凭经验研究该算法并探索变量以更好地理解其成功并将其扩展到输入是连续的或者输入分布的结构未充分揭示要预测的变量的情况。在监督的任务中。本文的实验也证实了这样的假设：贪婪的无层监督训练策略主要通过初始化接近良好局部最小值的区域中的权重来促进优化，从而产生内部分布式表示，这是输入的高级抽象，带来更好概括

**4.2 相关论文二**

**（1）题目**：Multi-column deep neural networks for image classificationeralization.用于图像分类的多列深度神经网络。

**（2）作者介绍**：Ciresan, D; Meier, U; Schmidhuber, J

**（3）摘要**：传统的计算机视觉和机器学习方法无法将人类的表现与诸如识别手写数字或交通标志等任务相匹配。本文的生物学合理，广泛和深入的人工神经网络架构可以。卷积赢家 - 通吃神经元的小（通常是最小的）感受域产生大的网络深度，导致与视网膜和视觉皮层之间的哺乳动物中发现的稀疏连接的神经层大致一样多。只有获胜的神经元才会被训几个深层神经专栏成为以不同方式预处理的输入专家;他们的预测是平均的。显卡允许快速培训。在极具竞争力的MNIST手写基准测试中，本文的方法是第一个实现近人类表现的方法。在交通标志识别基准测试中，它比人类高出两倍。本文还在众多常见的图像分类基准上改进了最新技术。

**4.3 相关论文三**

**（1）题目**：Efficient learning of sparse, distributed, convolutional feature representations for object recognition.

有效学习用于对象识别的稀疏，分布式，卷积特征表示。

**（2）作者介绍**：Kihyuk Sohn ; Dae Yon Jung ; Honglak Lee ; Alfred O. Hero。

**（3）摘要**：信息图像表示对于实现对象识别任务中的最新性能非常重要。在用于开发图像表示的特征学习算法中，受限制的玻尔兹曼机器（RBM）具有良好的表达能力并构建有效的表示。但是，培训成果管理制的困难一直是其广泛使用的障碍。为了解决这个难题，本文展示了混合模型和RBM之间的联系，并为利用这些连接的RBM提供了一种有效的训练方法。据本文所知，这是第一项工作，表明RBM可以通过几乎没有超参数调整来训练，以提供与混合模型类似或明显更好的分类性能（例如，高斯混合模型）。除了这种有效的训练之外，本文还评估了卷积训练的重要性，与非卷积训练相比，卷积训练可以捕获更大的空间背景，冗余更少。总的来说，本文的方法使用单一类型的功能在Caltech 101/256数据集上实现了最先进的性能。