

# AlexNet

[论文链接](#)

## Introduction

介绍图像识别任务的背景以及面临的挑战。

强调了卷积神经网络的重要性。

## Dataset

介绍了ImageNet的基本信息。本文使用了ImageNet的子集ILSVRC-2010作为训练集。

训练集不做任何预处理，只做一次裁剪，什么特征也不需要提取(作者一直重复，应该很骄傲)

ImageNet的分辨率不固定，所以将短边缩减到256像素

## Architecture

- 第一层卷积：输入为 $3 \times 224 \times 224$ 的RGB图片，卷积核为 $11 \times 11$ ，stride为4。输出尺寸为 $54 \times 54 \times 96$ ，输出到两个GPU上。
- 第二层卷积：输入为 $54 \times 54$ ，卷积核为 $5 \times 5$ ，步幅为1，填充为2，输出为256，池化核为 $2 \times 2$ ，步幅为2，池化后为 $27 \times 27$ 。GPU独自进行卷积，无通讯。
- 第三层卷积：输入为 $27 \times 27$ ，卷积核为 $3 \times 3$ ，步幅为1，填充为1，输出为384。

GPU还是每个GPU上有自己的卷积核，但是每个卷积核会同时将第二个卷积层中的卷积结果作为输入，通讯一次。

- 第四层卷积：输入为 $27 \times 27$ ，卷积核为 $3 \times 3$ ，步幅为1，填充为1，输出为384。
- 第五层卷积：输入为 $27 \times 27$ ，卷积核为 $3 \times 3$ ，步幅填充为1，输出为256，池化核为 $2 \times 2$ ，步幅为2，池化后为 $13 \times 13$ 。
- 展平层：展平成一维向量 $13 \times 13 \times 256 = 43648$ ，传入全连接层。
- 第一层全连接层：神经元数量为4096，使用Relu激活函数。
- 第二层全连接层：神经元数量为4096，使用Relu激活函数。
- 第三层全连接层：神经元数量为1000，使用Softmax激活函数。

高宽慢慢变小、深度慢慢增加，随着深度的增加，慢慢地将空间信息压缩，直到最后每一个像素能够代表前面一大块的像素。

通道数慢慢增加，可以认为每个通道数是去看一种特定的模式，这时候人已经看不懂了，只有神经网络能看懂。

例如3个通道为RGB，192个通道可以简单地认为，能够识别图片中的192种不同的模式

使用Relu激活函数，相比tanh和sigmoid更快（作者很急）

机器识别新图片时候会分成4096份进行对比。

AlexNet模型并行（model parallel）当时是为了节约时间，其实是不必要的，所以在之后七八年内没几人使用。

但现在自然语言处理火了模型过大又开始用了。

过几年老黄的计算卡突飞猛进到一张卡Z100等于很多张H100的时候可能又不用了。

## Reducing Overfitting

有6000万参数，所以必须好好处理过拟合。

### Data Augmentation

- 最简单最粗暴的方法：在256×256的图片上随机扣224×224扩大数据集。
- PCA：降低维度，保留关键信息，即使图片修改颜色神经网络也能看出来。

### Dropout

通过以0.5的概率将每个隐藏神经元的输出设置为零，在每次输入时随机采样一个不同的神经网络架构，但所有这些架构共享权重。

Dropout减少了神经元之间的复杂共适应性，使得神经元被迫学习更加稳健的特征。它有效地防止了过拟合。但是耗费了时间。

## Details of learning

SGD、batchsize、momentum、weight decay、learning rate、训练时间、计算卡。

流水账，也没讲为什么要使用这些参数。

## Results

模型在图像识别竞赛多么多么优秀.....

### Qualitative Evaluations

两张卡其中一张卡与颜色无关，另一张卡颜色相关。这种情况每次都出现，且与权重无关。（诡异炼丹）

上图训练成功。

## Discussion

神经网络深度很重要

使用未监督学习可能会更好

未来要应用到视频（Yolo? ? ?）