过去几年，深度学习（Deep learning）在解决诸如视觉识别(visual recognition)、语音识别(speech recognition)和自然语言处理(natural language processing)等很多问题方面都表现出非常好的性能。在不同类型的深度神经网络当中，卷积神经网络是得到最深入研究的。早期由于缺乏训练数据和计算能力，要在不产生过拟合(overfitting)的情况下训练高性能卷积神经网络是很困难的。标记数据和近来GPU的发展，使得卷积神经网络研究涌现并取得一流结果。本文中，我们将纵览卷积神经网络近来发展，同时介绍卷积神经网络在视觉识别方面的一些应用。

CNN快速发展，得益于LeNet-5、Alexnet、ZFNet、VGGNet、GoogleNet、ResNet等不同结构的设计出现。

1.CNN的基本结构

卷积神经网络的结构有很多种，但是其基本架构是相似的，拿LeNet-5为例来介绍，如下图，它包含三个主要的层——卷积层（ convolutional layer）、池化层（ pooling layer）、全连接层（ fully-connected layer）

   图中的卷积网络工作流程如下，输入层由32×32个感知节点组成，接收原始图像。然后，计算流程在卷积和子抽样之间交替进行，如下所述：

    第一隐藏层进行卷积，它由6个特征映射组成，每个特征映射由28×28个神经元组成，每个神经元指定一个 5×5 的接受域；

    第二隐藏层实现子抽样和局部平均，它同样由 6 个特征映射组成，但其每个特征映射由14×14 个神经元组成。每个神经元具有一个 2×2 的接受域，一个可训练系数，一个可训练偏置和一个 sigmoid 激活函数。可训练系数和偏置控制神经元的操作点。

    第三隐藏层进行第二次卷积，它由 16个特征映射组 成，每个特征映射由 10×10 个神经元组成。该隐藏层中的每个神经元可能具有和下一个隐藏层几个特征映射相连的突触连接，它以与第一个卷积 层相似的方式操作。

    第四个隐藏层进行第二次子抽样和局部平均计算。它由 16 个特征映射组成，但每个特征映射由 5×5 个神经元组成，它以 与第一次抽样相似的方式操作。

    第五个隐藏层实现卷积的最后阶段，它由 120 个神经元组成，每个神经元指定一个 5×5 的接受域。

    最后是个全连接层，得到输出向量。

    相继的计算层在卷积和抽样之间的连续交替，我们得到一个“双尖塔”的效果，也就是在每个卷积或抽样层，随着空 间分辨率下降，与相应的前一层相比特征映射的数量增加。卷积之后进行子抽样的思想是受到动物视觉系统中的“简单的”细胞后面跟着“复杂的”细胞的想法的启发而产生的。

其中，卷积层，用来学习输入数据的特征表征。卷积层由很多的卷积核（convolutional kernel）组成，卷积核用来计算不同的feature map；

激励函数（activation function）给CNN卷积神经网络引入了非线性，常用的有sigmoid 、tanh、 ReLU函数；

池化层降低卷积层输出的特征向量，同时改善结果（使结构不容易出现过拟合），典型应用有average pooling 和 max pooling；

全连接层将卷积层和Pooling 层堆叠起来以后，就能够形成一层或多层全连接层，这样就能够实现高阶的推力能力

2.CNN卷积神经网络的改进策略

  自从2012年AlexNet 取得成功以后，科研工作者提出了很多改进CNN的方法，基本都是从以下六个方面入手：convolutional layer、pooling layer、activation function、loss function、regularization 、optimization

2.1convolutional layer

1). Network in network

It replaces the linear filter of the convolutional layer by a micro network

2). Inception module : 是继承了NIN的扩展. 使用不同size的filter来捕获不同size的visual patterns.

2.2 pooling layer

池化层是CNN的重要组成部分，它通过减少卷积层之间的连接数量来降低计算的复杂度。

1).Lp pooling：Lp 池化是建立在复杂细胞运行机制的基础上，受生物启发而来

2). Mixed pooling：combination of max pooling and average pooling

3).Stochastic pooling：Stochastic pooling 是受 droptout 启发而来的方法

4).Spectral pooling

5). Spatial pyramid pooling

空间金字塔池化可以把任何尺度的图像的卷积特征转化成相同维度，这不仅可以让CNN处理任意尺度的图像，还能避免 cropping和warping操作，导致一些信息的丢失，具有非常重要的意义。

一般的CNN都需要输入图像的大小是固定的，这是因为全连接层的输入需要固定输入维度，但在卷积操作是没有对图像尺度有限制，所有作者提出了空间金字塔池化，先让图像进行卷积操作，然后转化成维度的特征输入到全连接层，这个可以把CNN扩展到任意大小的图像。

2.3activation function

一个合适的激励函数可以有效地提高CNN的运算性能。常用的非线性激活函数有sigmoid、tanh、relu等等，前两者sigmoid/tanh比较常见于全链接层，后者relu常见于卷积层。

sigmoid函数图像曾在神经网络和深度学习（一）中介绍过，这里再介绍ReLU、LReLU、PReLU、RReLU、ELU几种激励函数的特性曲线

2.4 Loss function

1). Softmax loss :

2). Hinge loss : to train large margin classifiers such as SVM

3). Contrastive loss : to train Siamese network

2.5 Regularization 正则化

通过正则化可以有效的减小CNN的过拟合问题。这里介绍两种正则化技术——Dropout 、 DropConnect

Dropout是指在模型训练时随机让网络某些隐含层节点的权重不工作，不工作的那些节点可以暂时认为不是网络结构的一部分，但是它的权重得保留下来(只是暂时不更新而已) ； 而DropConnect 是DropOut的进一步发展，不再和DropOut一样随意设置输出神经元的值为0，而是随机设置权重矩阵W 为0。

2.6 Optimization

这里介绍几种优化CNN的关键技术：

1). Weights initialization

2). Stochastic gradient descent

3). Batch Normalization

4). Shortcut connections

3. 加速CNN计算速度

3.1 FFT : 使用快速傅里叶变换，可以重复利用一些单元，比如输出梯度的傅里叶变换

3.2 Matrix Factorization : 矩阵因子分解，可以减小计算量，来加速CNN 的训练

3.3 Vector quantization : 矢量量化（VQ）是用来压缩密集的连接层，使得CNN模型变得更小

4. CNN 的主要应用

使用CNN,可以使得以下几个应用达到最佳的（state-of-the-art）性能:

1). Image Classification

2). Object Tracking

3). Pose Estimation

4). Text Detection

5). Visual Saliency detection

6). Action Recognition

7). Scene Labeling

总结： 通过本篇论文可以对CNN卷积神经网络的基本框架有了一定的了解，知道了提高CNN性能的方法，从哪些方面来入手来完善CNN网络结构，还有CNN主要的应用，如果想要做这个几个方面的研究，那么CNN是必须要深入了解的。接下来的工作就是将文章中提到的几种CNN结构（LeNet-5、Alexnet、ZFNet、VGGNet、GoogleNet、ResNet）熟悉，并且尽快应用起来，在此基础上从第二部分的六个方面提出自己的优化方法。

---------------------

作者：Michael\_liangliang

来源：CSDN

原文：https://blog.csdn.net/laingliang/article/details/53073591

版权声明：本文为博主原创文章，转载请附上博文链接！