

# Network In Network(NIN):2014

National University of Singapore, Singapore

## 摘要:

我们提出了一个新的网络架构叫做NIN (Network in Network) 来增强模型在感受野内的图像局部的分辨能力。传统的卷积层一般是通过让输入经过一个线性滤波器之后再用非线性激活函数进行处理,但是在我们的方法中构建了一个结构更复杂的微型网络来对感受野内更复杂的数据进行处理。我们将这个微型网络设置为多层感知器。特征图是通过类似CNN的方式在输入上滑动得到的。深度NIN可以通过堆叠上述的模块得到。通过微型网络进行增强的局部建模,同时在特征图进入分类头的时候使用了全局平均池化层,相较于传统的全连接层更具可解释性,同时也能防止过拟合现象的发生。我们通过使用NIN在CIFIA-10和CIFIA-100数据集上取得了SOTA,和在SVHN和MNIST数据集上相对可以的表现

## 论文出发点或背景

卷积核的底层核心还是线性激活的方式,我们认为这种方式对抽象数据的处理水平很低,通过将GLM (generalized linear model) 替换为近似能力更高的模型可以增强网络对于局部的抽象能力。我们所说的抽象是指该特征对于相同的概念的变体是相同的。当样本是线性可分离时, GLM对数据的抽象表现较好。因此传统的CNN假设the latent concept是线性可分的,然而一些数据是高度非线性的,因而CNN经常通过后接非线性函数来捕获非线性特征。在NIN中我们通过代替GLM为以可以通过反向传播算法训练的符合通用近似定理的多层感知器为基础的微型网络,增强局部抽象能力。

## 论文创新思路

当样本是线性可分离时, GLM对数据的抽象表现较好。因此传统的CNN假设the latent concept是线性可分的,然而一些数据是高度非线性的,因而CNN经常通过后接非线性函数来捕获非线性特征,有时也会通过使用多尺寸的卷积核进行一定的补偿。在NIN中我们通过代替GLM为以可以通过反向传播算法训练的符合通用近似定理的多层感知器为基础的微型网络,增强局部抽象能力。

利用全局池化层替换网络的全连接层:

可解释性: 在特征映射和类别之间架起了桥梁

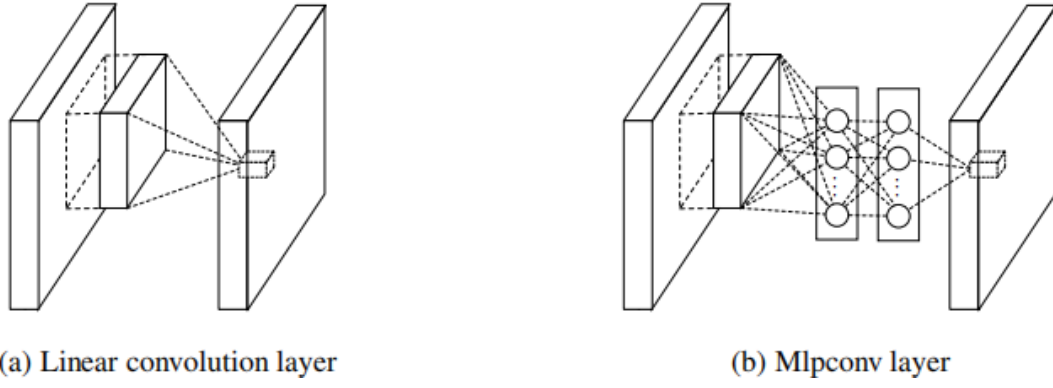
防止过拟合: 对比全连接层, 全局池化平均本身是一个结构正则化器, 可以防止对整体结构的过拟合

## 论文方法的大概介绍

the micro network is integrated into CNN structure in pursuit of better abstractions for all levels of features.

选择多层感知器的原因: 1.多层感知器本身就是一个深度模型, 符合特征重用的思想2.和卷积神经网络的结构相兼容, 使用反向传播算法进行参数更新

在NIN中我们通过代替GLM为以可以通过反向传播算法训练的符合通用近似定理的多层感知器为基础的微型网络, 增强局部抽象能力。



$$\begin{aligned}
 f_{i,j,k_1}^1 &= \max(w_{k_1}^1{}^T x_{i,j} + b_{k_1}, 0). \\
 &\vdots \\
 f_{i,j,k_n}^n &= \max(w_{k_n}^n{}^T f_{i,j}^{n-1} + b_{k_n}, 0).
 \end{aligned}$$

从跨通道池化的角度来看，上式等价于一个正常的卷积级联的跨参数池化，也等价于一个 $1 \times 1$ 卷积核，这种交互方式加强了信息的复杂性和交互性

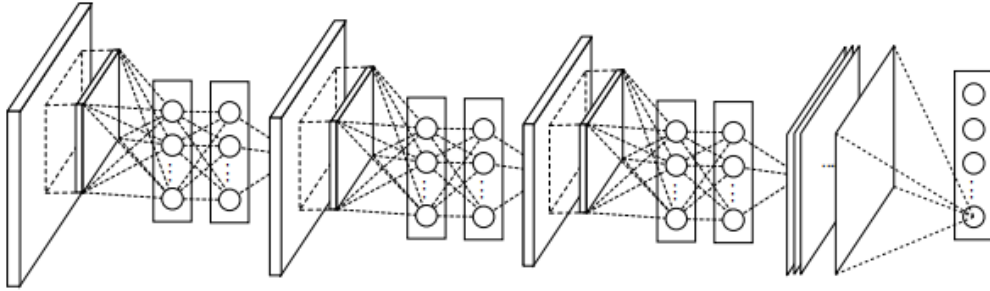


Figure 2: The overall structure of Network In Network. In this paper the NINs include the stacking of three mlpconv layers and one global average pooling layer.

$n$ 代表了多层感知器的层数。ReLU被用作每个全连接层的激活函数，利用全局池化层替换网络的全连接层，通过将特征映射和类被强制对应起来，更适合卷积结构，特征图也被赋予了类别置信度的含义，同时全局平均池化层没有参数进行优化，可以被看作是一个结构正则化器，能够避免过拟合。

## 实际效果

发现在mlpconv层之间引入Dropout可以将测试误差减少20%以上

Table 1: Test set error rates for CIFAR-10 of various methods.

Method	Test Error
Stochastic Pooling [11]	15.13%
CNN + Spearmint [14]	14.98%
Conv. maxout + Dropout [8]	11.68%
<b>NIN + Dropout</b>	<b>10.41%</b>
CNN + Spearmint + Data Augmentation [14]	9.50%
Conv. maxout + Dropout + Data Augmentation [8]	9.38%
DropConnect + 12 networks + Data Augmentation [15]	9.32%
<b>NIN + Dropout + Data Augmentation</b>	<b>8.81%</b>

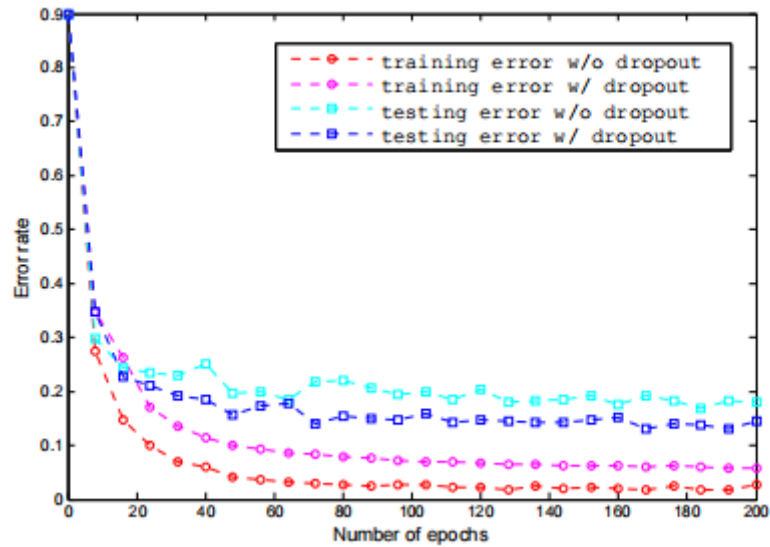


Table 2: Test set error rates for CIFAR-100 of various methods.

Method	Test Error
Learned Pooling [16]	43.71%
Stochastic Pooling [11]	42.51%
Conv. maxout + Dropout [8]	38.57%
Tree based priors [17]	36.85%
<b>NIN + Dropout</b>	<b>35.68 %</b>

Table 4: Test set error rates for MNIST of various methods.

Method	Test Error
2-Layer CNN + 2-Layer NN [11]	0.53%
Stochastic Pooling [11]	0.47%
NIN + Dropout	0.47%
<b>Conv. maxout + Dropout [8]</b>	<b>0.45 %</b>

Table 5: Global average pooling compared to fully connected layer.

Method	Testing Error
mlpconv + Fully Connected	11.59%
mlpconv + Fully Connected + Dropout	10.88%
mlpconv + Global Average Pooling	10.41%

## 个人对这篇论文的理解

论文主要是提出卷积神经网络对于非线性的数据抽象能力不够，除去非线性的激活函数，本身只有一个线性上的运算，为了加强这种抽象能力，作者提出将卷积得到的特征图紧接着放入一个多层感知器中，通过大量神经元的运算增加网络的复杂度和表达能力，全连接层的实现是通过 $1 \times 1$ 卷积实现的。全局平均池化层的使用减小了网络分类head的需要学习的参数，因而过拟合的风险也小了，同时实验也证明了全局平均池化层的有效性。总而言之，作者的思路是通过引入更加复杂的结构来增强网络的表达能力、特征提取能力，而不是简单地堆叠卷积神经网络的深度。