

lhanchao的博客

个人资料



lhanchao

原创

粉丝

喜欢

评论

68

226

13

168



等级：

博客5

访问量：28万+

积分：3142

排名：1万+

介绍

关注计算机视觉和深度学习小白一枚，评论和私信可能回不及时，可以邮件联系：lhanchao@mail.ustc.edu.cn。博客可能也写的有问题，希望大家一起交流

文章搜索

文章分类

Python相关

(6)

Tensorflow学习笔记

(2)

Ubuntu相关

(7)

OpenCV

(10)

深度学习

(15)

三维重建

(7)

杂记

(4)

C++

(8)

特征点匹配

(7)

LintCode

(12)

文章存档

2017年11月

(2)

2017年7月

(1)

2017年6月

(2)

2017年5月

(2)

2017年4月

(2)

展开

阅读排行

Tensorflow学习笔记 (8) —...

(21847)

特征点匹配——SIFT算法详解

(14576)

特征点匹配——ORB算法介绍

(13898)

Opencv感兴趣区域复制，画...

(12300)

神经网络中前向传播和反向传...

(11863)

使用visualSFM和meshlab进...

(11672)

[深度学习]Ubuntu16.04 + G...

(10034)

特征点匹配——使用基础矩阵...

(8865)

Ubuntu 16.04无法安装第三...

(8704)

三维重建 (四) PMVS算法 th...

(8402)

最新评论

三维重建 (二) Sift特征提取与匹配qq_41532245

: 楼主，你好，我想问一下这种代码直接就可以提取出特征点了，但是sift算法每一步怎么体现出来的呢，还...

libSVM + VS2013 + ...大萌籽

: 跑起来了，学习了！赞！

使用visualSFM和meshl...MoniWong

: 您好，您在文中所重建的模型拍了多少图片啊？不知道为什么我生成稠密点云后点云依旧很稀疏。

Opencv中的Mat类使用方法总结guoji_

: get

[深度学习] Batch Normalization算法介绍

标签：

深度学习

Batch-Norm

BN

2017年04月21日 11:34:31

6740人阅读

评论(1)

收藏

举报

分类：

深度学习 (14)

版权声明：本文为博主原创文章，未经博主允许不得转载。 <https://blog.csdn.net/lhanchao/article/details/70308092>

目录(?)

很早就打算写这篇博客了，最近遇到的问题比较多，所以拖了又拖，今天问题似乎解决了，等着程序运行的时候再来回顾一下Batch Normalization算法。

Batch Normalization是2015年Google研究员在论文《Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift》一文中提出的，同时也将BN应用到了2014年的GoogLeNet上，也就是Inception-v2。

BN算法在很大程度上加速了训练过程，放宽了网络初始化的条件，论文中还提出有了BN可以在不使用Dropout，同时也可以在一定程度上提升网络的识别效果，在之后的ResNet等等新网络中有广泛的应用。

下面我们来详细的看一下BN算法。

1. 要解决的问题

自从2012年以来，CNN网络模型取得了非常大的进步，而这些进步的推动条件往往就是模型深度的增加。从AlexNet的几层，到VGG和GoogleNet的十几层，甚至到ResNet的上百层，网络模型不断加深，取得的效果也越来越好，然而网络越深往往就越难以训练。我们知道，CNN网络在训练的过程中，前一层的参数变化影响着后面层的变化（因为前面层的输出是后面的输入），而且这种影响会随着网络深度的增加而不断放大。在CNN训练时，绝大多数都采用mini-batch使用随机梯度下降算法进行训练，那么随着输入数据的不断变化，以及网络中参数不断调整，网络的各层输入数据的分布则会不断变化，那么各层在训练的过程中就需要不断的改变以适应这种新的数据分布，从而造成网络训练困难，难以拟合的问题。

（可以这样想，比如网络中每一层都是一个人，今天前面一层的人说要你往左走3,明天有让你往左走2,第三天又让你往右走5,结果三天下来，你还在原地，这样就让你的进度变慢了。PS：我是这样理解的，如有问题，请指出）BN算法解决的就是这样的问题，他通过对每一层的输入进行归一化，保证每层的输入数据分布是稳定的，从而达到加速训练的目的。

2. BN算法流程

BN算法的理论推导我这里就不多说了（原因是理解的不透彻= =），这里只讲一下算法的流程。首先，BN算法在每一次iteration中的每一层输入都进行了归一化，将输入数据的分布归一化为均值为0,方差为1的分布，如下式：

$$\hat{x}^{(k)} = \frac{x^k - E[x^k]}{\sqrt{Var[x^k]}}$$

其中， x^k 表示输入数据的第 k 维， $E[x^k]$ 表示该维的平均值， $\sqrt{Var[x^k] + \epsilon}$ 表示标准差。

但是这种做法有一个致命的缺点，尽管这样把每层的数据分布都固定了，但是这种分布不一定是前面一层的学习到的数据分布，这样强行归一化就会破坏掉刚刚学习到的特征，BN算法的第二步就解决了这个缺点。

BN算法在第二步中设置了两个可学习的变量 γ 和 β ，然后用这两个可学习的变量去还原上一层应该学到的数据分布

$$y^{(k)} = \gamma^k \hat{x}^{(k)} + \beta^{(k)}$$

添加这种操作的目的是还原出上一层需要学习的数据分布，这样**BN就把原来不固定的数据分布全部转换为固定的数据分布，而这种数据分布恰恰就是要学习到的分布**，从而加速了网络的训练。BN算法的流程如下图所示：

特征点匹配——FREAK算法介绍
qq_27705465 : [reply]lhanchao[/reply] opencv中，freak类没有特征提取的方式，会...
[深度学习] Batch Norm...
陈立里 : 楼主，请问如果Convolution 层后面跟着BN，那么Convolution 层的bias_...
[深度学习] Ubuntu16.04...
qq_40300210 : [reply]wspba[/reply] 你懂个积极
[Python] caffe.io...
lhanchao : [reply]baiyu33[/reply] 谢谢提醒
[Python] caffe.io...
baiyu33 : 拜托写清楚，不仅仅是[0~255]和[0,1]的区别，rgb和bgr也是反的。
使用visualSFM和meshl...
梦之七月未央 : 您好，没找到网格化那个选项，是版本不对吗

联系我们



请扫描二维码联系客服
✉webmaster@csdn.net
☎400-660-0108
💬QQ客服 🗣客服论坛

关于 招聘 广告服务 百度
©1999-2018 CSDN版权所有
京ICP证09002463号

经营性网站备案信息
网络110报警服务
中国互联网举报中心
北京互联网违法和不良信息举报中心

Input: Values of x over a mini-batch: $\mathcal{B} = \{x_{1...m}\};$
Parameters to be learned: γ, β
Output: $\{y_i = \text{BN}_{\gamma, \beta}(x_i)\}$

$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m} \sum_{i=1}^m x_i$$
$$\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_{\mathcal{B}})^2$$
$$\hat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}}$$
$$y_i \leftarrow \gamma \hat{x}_i + \beta \equiv \text{BN}_{\gamma, \beta}(x_i)$$

// mini-batch mean

// mini-batch variance

// normalize

// scale and shift

👍

1

🔖

💬

Algorithm 1: Batch Normalizing Transform, applied to activation x over a mini-batch. <http://blog.csdn.net/lhanchao>

2.1 BN算法在训练和测试时的应用

BN算法在训练时的操作就如我们上面所说，首先提取每次迭代时的每个mini-batch的平均值和方差进行归一化，再通过两个可学习的变量恢复要学习的特征。
但是在实际应用时就没有mini-batch了，那么BN算法怎样进行归一化呢？实际上在测试的过程中，BN算法的参数就已经固定好了，首先进行归一化时的平均值和方差分别为：

$$E[x] = E_B[\mu_B]$$
$$Var[x] = \frac{m}{m-1} E_B[\sigma_B^2]$$

即平均值为所有mini-batch的平均值的平均值，而方差为每个batch的方差的无偏估计(无偏估计介绍见[什么是无偏估计？——知乎](#))

最终BN算法的训练和测试的流程如下图所示：

Input: Network N with trainable parameters $\Theta;$
subset of activations $\{x^{(k)}\}_{k=1}^K$
Output: Batch-normalized network for inference, $N_{\text{BN}}^{\text{inf}}$
1: $N_{\text{BN}}^{\text{tr}} \leftarrow N$ // Training BN network
2: **for** $k = 1 \dots K$ **do**
3: Add transformation $y^{(k)} = \text{BN}_{\gamma^{(k)}, \beta^{(k)}}(x^{(k)})$ to $N_{\text{BN}}^{\text{tr}}$ (Alg. 1)
4: Modify each layer in $N_{\text{BN}}^{\text{tr}}$ with input $x^{(k)}$ to take $y^{(k)}$ instead
5: **end for**
6: Train $N_{\text{BN}}^{\text{tr}}$ to optimize the parameters $\Theta \cup \{\gamma^{(k)}, \beta^{(k)}\}_{k=1}^K$
7: $N_{\text{BN}}^{\text{inf}} \leftarrow N_{\text{BN}}^{\text{tr}}$ // Inference BN network with frozen parameters

8: **for** $k = 1 \dots K$ **do**
9: // For clarity, $x \equiv x^{(k)}, \gamma \equiv \gamma^{(k)}, \mu_{\mathcal{B}} \equiv \mu_{\mathcal{B}}^{(k)}$, etc.
10: Process multiple training mini-batches \mathcal{B} , each of size m , and average over them:
$$E[x] \leftarrow E_{\mathcal{B}}[\mu_{\mathcal{B}}]$$
$$\text{Var}[x] \leftarrow \frac{m}{m-1} E_{\mathcal{B}}[\sigma_{\mathcal{B}}^2]$$

11: In $N_{\text{BN}}^{\text{inf}}$, replace the transform $y = \text{BN}_{\gamma, \beta}(x)$ with $y = \frac{\gamma}{\sqrt{\text{Var}[x] + \epsilon}} \cdot x + \left(\beta - \frac{\gamma E[x]}{\sqrt{\text{Var}[x] + \epsilon}}\right)$
12: **end for**

👍

1

🔖

💬

Algorithm 2: Training a Batch-Normalized Network <http://blog.csdn.net/lhanchao>

2.2 BN算法在CNN中的应用

BN算法在CNN中往往放在每个卷积层之后，ReLU操作之前。在CNN中操作时，BN算法把每个特征图看作一个神经元，计算该特征图对应数据的均值和方差进行归一化，并且每个特征图对应两个学习变量 γ 和 β 。详细介绍见<http://blog.csdn.net/hjimce/article/details/50866313>中关于BN在CNN中的使用方法介绍。
最后介绍一下BN在Caffe中的使用

```
1 layer {  
2   bottom: "Convolution1"  
3   top: "Convolution1"
```

```
4   name: "bn1"
5   type: "BatchNorm"
6   batch_norm_param {
7     use_global_stats: true
8   }
9   include {
10    phase: TEST
11  }
12 }
13 layer {
14   bottom: "Convolution1"
15   top: "Convolution1"
16   name: "bn1"
17   type: "BatchNorm"
18   include {
19     phase: TRAIN
20   }
21 }
22 layer {
23   bottom: "Convolution1"
24   top: "Convolution1"
25   name: "scale1"
26   type: "Scale"
27   scale_param {
28     bias_term: true
29   }
30 }
```


1





Caffe中并没有实现BN层分为两个部分BatchNorm和Scale层，其中BatchNorm又根据use_global_stats的值分为两种， use_global_stats为true时该层的参数保持不变， 这种应该应用在test的情况下， 而train的过程中要保证use_global_stats一定不为true，当然，默认情况下use_global_stats为false。

关于Batch Normalization的更多介绍可以看原文也可以参考以下链接：
[深度学习中 Batch Normalization为什么效果好？——知乎](#)
[深度学习（二十九）Batch Normalization 学习笔记](#)
[解读Batch Normalization](#)

- 上一篇

[深度学习]Deep Residual Learning for Image Recognition(ResNet,残差网络)阅读笔记
- 下一篇

[深度学习] RCNNs系列（1） Ubuntu下Faster RCNN配置及训练和测试自己的数据方法

您还没有登录,请[\[登录\]](#)或[\[注册\]](#)

查看评论



陈立里

1楼 2018-03-08 09:55发表

楼主，请问如果Convolution 层后面跟着BN，那么Convolution 层的bias_term: 在训练和测试的时候各为什么？
还有一些网络文件里面，有
param {
lr_mult: 1
decay_mult: 1
}
这种参数，而且有时候在一层里不止加一个， 楼主你知道这是什么意思吗？

深度学习中的Batch Normalization

 whitesilence 2017年07月21日 17:06  9477

在看 ladder network(<https://arxiv.org/pdf/1507.02672v2.pdf>) 时初次遇到batch normalization （BN）. 文中说BN能加速收敛等好...

Batch Normalization 的原理解读

 ZhikangFu 2016年11月29日 14:13  3796

1:motivation 作者认为：网络训练过程中参数不断改变导致后续每一层输入的分布也发生变化，而学习的过程又要使每一层适应输入的分布，因此我们不得不降低 学习率、小心地初始化。作者将分布发生变化称...

Batch Normalization导读

 malefactor 2016年05月24日 19:08  44241

Batch Normalization作为最近一年来深度学习的重要成果，已经广泛被证明其有效性和重要性。本文对原始论文进行导读，帮助读者更好地理解BatchNorm。 ...

深度学习（二十九）Batch Normalization 学习笔记

近年来深度学习捷报连连，声名鹊起，随机梯度下架成了训练深度网络的主流方法。尽管随机梯度下降法，将对于训练深度网络，简单高效，但是它有个毛病，就是需要我们人为的去选择参数，比如学习率、参数初始化等，这些...

 hjimce 2016年03月12日 17:00  90902

Batch Normalization 学习笔记

 leayc 2017年08月28日 11:31  984

本文是对批标准化方法的一则学习笔记

论文笔记-Batch Normalization

 u012816943 2016年06月17日 11:30  7024

Batch Normalization：减弱 internal covariate shift，使训练加快，并且可以不再用dropout和LRN。...



1

关于Batch Normalization的另一种理解

 Alchipmunk 2017年01月11日 16:00  1513

Batch Norm可谓深度学习中非常重要的技术，不仅可以使训练更深的网络变容易，加速收敛，还有一定正则化的效果，可以防止模型过拟合。在很多基于CNN的分类任务中，被大量使用。但我最近在图像超分辨...



解读Batch Normalization

 elaine_bao 2016年03月14日 22:22  37272

本文转载自：http://blog.csdn.net/shuzfan/article/details/50723877 本次所讲的内容为Batch Normalization，简称BN，来源于《...

Batch Normalization理解

 life_is_amazing 2016年07月06日 15:36  3020

原文地址：http://blog.csdn.net/malefactor/article/details/51476961 作者：张俊林 Batch Normalization作为...

解读Batch Normalization

 shuzfan 2016年02月23日 16:03  14675

目录 目录 1-Motivation 2-Normalization via Mini-Batch Statistics 测试 BN before or after Activation 3-Expe...


我读Batch Normalization

 xuanwu_yan 2016年09月18日 19:48  1344

论文地址：Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift 论...

Batch Normalization简介



2017年11月24日 17:20 3.36MB 

batch-normalization 技术分析

 guoyuhaoaaa 2017年02月07日 10:39  581

这两天重新回顾了一下batch-normalization技术，主要参考了论文《Batch Normalization: Accelerating Deep Network Training by R...

总结2: Batch Normalization反向传播公式推导及其向量化

1. 背景介绍上周学习了吴恩达的Deep Learning专项课程的第二门课，其中讲到了Batch Normalization（BN）。但在课程视频中，吴恩达并没有详细地给出Batch Normali...

 pan5431333 2017年09月21日 15:47  1195

Batch Normalization 简单理解

 meanme 2015年09月23日 14:47  18827

1：背景由于在训练神经网络的过程中，每一层的 params是不断更新的，由于params的更新会导致下一层输入的分布情况发生改变，所以这就要求我们进行权重初始化，减小学习率。这个现象就叫做intern...

《Batch Normalization Accelerating Deep Network Training by Reducing Inte...

《Batch Normalization Accelerating Deep Network Training by Reducing Internal Covariate Shift》阅读笔记...

 happynear 2015年03月13日 12:42  43021

Batch Normalization详解

 qq_34484472 2017年09月14日 16:38  936

Batch Normalization详细介绍

【机器学习】covariate shift现象的解释

 mao_xiao_feng 2017年01月10日 22:10  7503

一、什么是covariate shift？在论文中经常碰到covariate shift这个词，网上相关的中文解释比较少 你可能会在介绍深度学习Batch Normalization方法的论文到中看...

Tensorflow中dropout和batch normalization

 Tcorpion 2018年01月04日 19:35  324

直接上代码探讨：def batch_norm_layer(x, train_phase, scope_bn):##x is input-tensor, train_phase is tf.Varia...

深入浅出——深度学习中的Batch Normalization使用

《Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shi

u010402786

2016年04月24日 15:33

8383



1

