Datawhale零基础入门CV赛事
街景字符编码识别
分类模型介绍

分享人：张强

目录

contents

赛事资源

Datawhale

天池新人赛由天池与Datawhale联合发起，并提供学习内容和组织学习：

- Datawhale是一个专注于数据科学与AI领域的开源组织；
- CV直播PPT 可关注Datawhale，回复关键词 **CV直播** 下载；
- 同时可以加入Datawhale数据竞赛交流群，一起组队参赛，交流学习；

个人介绍

Datawhale

张强

- ✓ 计算机研究生，研究方向三维深度学习；
- ✓ 数据科学爱好者，CV学习者；
- ✓ Datawhale成员，开源贡献者；
- ✓ https://github.com/QiangZiBro

Datawhale CV小组开源项目：动手学CV-Pytorch版
https://github.com/datawhalechina/dive-into-cv-pytorch

本节讲解的各种模型比较：
https://github.com/QiangZiBro/cnn_models_comparation.pytorch

Datawhale 组队学习：
图像处理（上）：https://github.com/datawhalechina/team-learning
图像处理（下）：敬请关注

Part 1 分类模型基础概念

早期图像分类模型

机器学习：数据驱动的方法

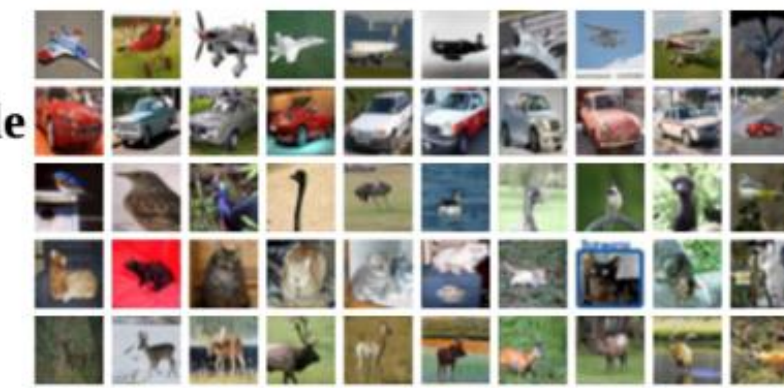
1.创建数据集，做标记
2.使用ML训练分类器
3.使用分类器测试新图片

```
def train(images, labels):
  # Machine learning!
  return model
```

```
def predict(model, test_images):
  # Use model to predict labels
  return test_labels
```

Example training set

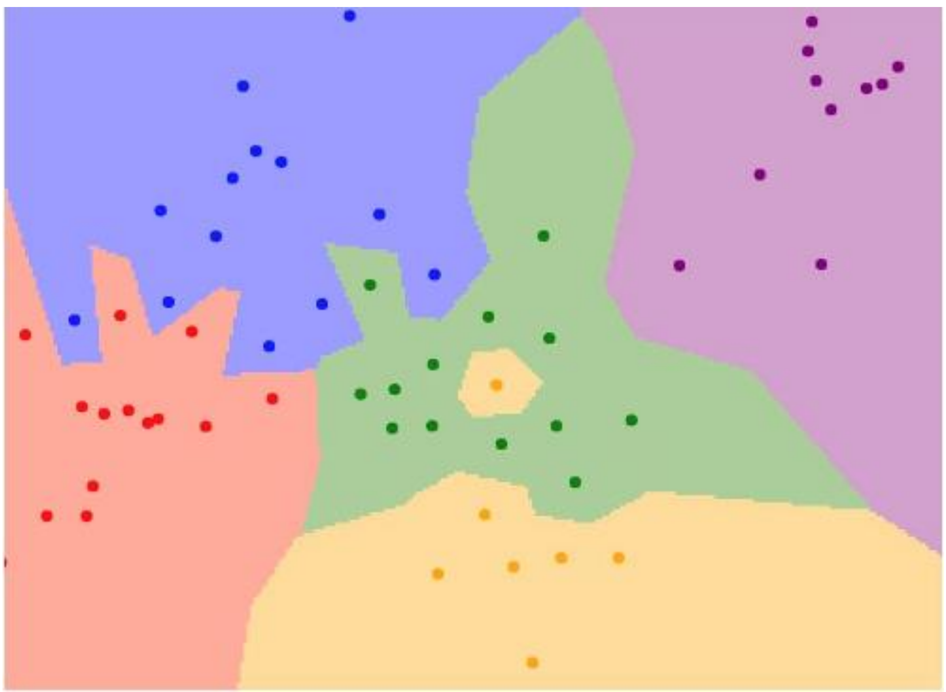
airplane
automobile
bird
cat
deer



http://cs231n.stanford.edu/slides/2019/cs231n_2019_lecture02.pdf

早期图像分类模型

近邻（Nearest Neighbor）



训练：存储所有图片

测试：找与测试图片**距离**最小的训练图片

早期图像分类模型

K近邻（k-nearest neighbor）

训练：存储所有图片

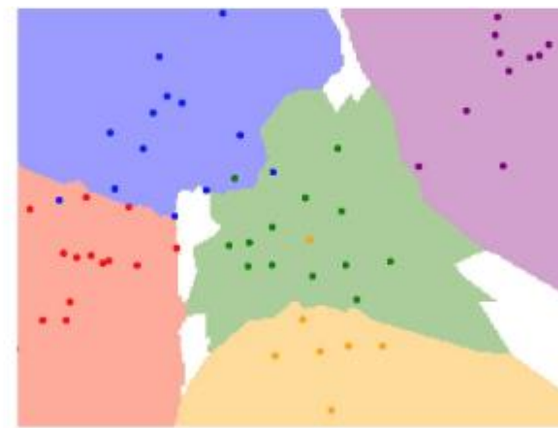
测试：找k个与测试图片**距离**最小的训练图片

超参数&难点

1. k?
2. 度量距离?

K = 1

K = 3



K = 5

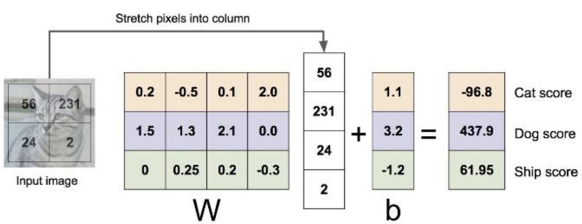
http://cs231n.stanford.edu/slides/2019/cs231n_2019_lecture02.pdf

早期图像分类模型

参数方法：线性分类

Algebraic Viewpoint

$f(x,W) = Wx$

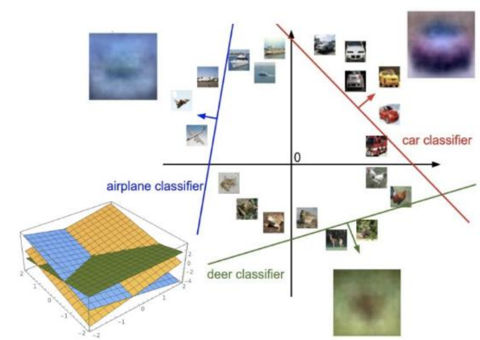


Visual Viewpoint

One template per class

Geometric Viewpoint

Hyperplanes cutting up space



http://cs231n.stanford.edu/slides/2019/cs231n_2019_lecture02.pdf

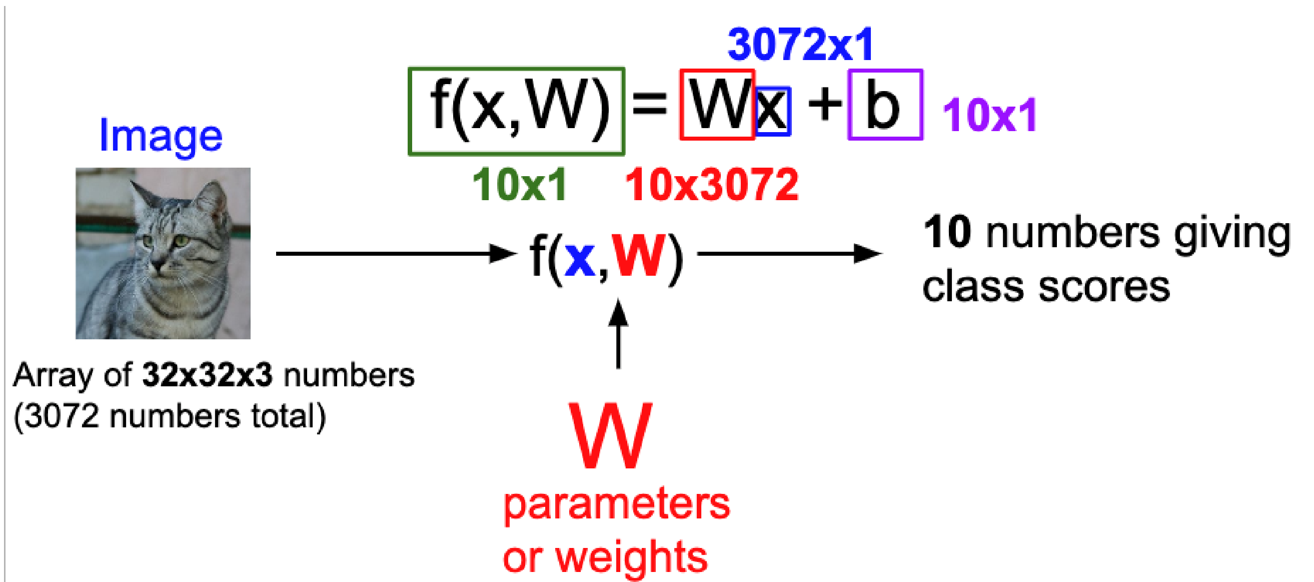
Datawhale

早期图像分类模型

参数方法：线性分类

估计W和b

回顾：线性代数 Ax = b



http://cs231n.stanford.edu/slides/2019/cs231n_2019_lecture02.pdf

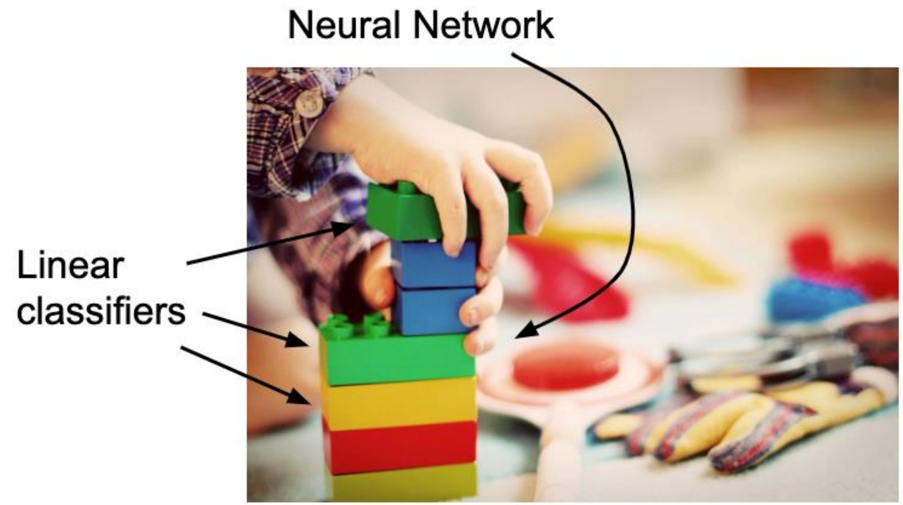
早期图像分类模型

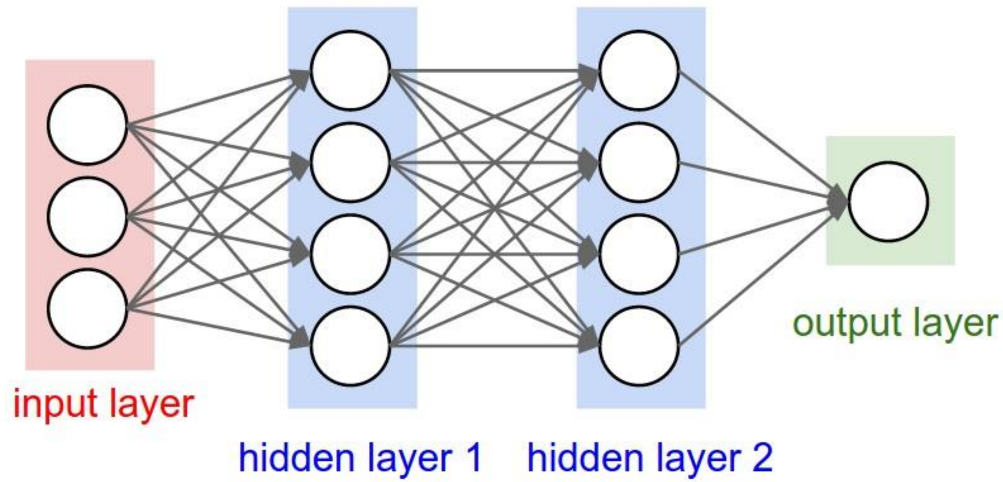
参数方法：多层感知机 （MLP）

输入32x32x3的图片

$$3072 \times 4096 + 4096 \times 4096 + 4096 \times 10 = 29401088 \approx \text{三千万}$$

通过BP进行训练



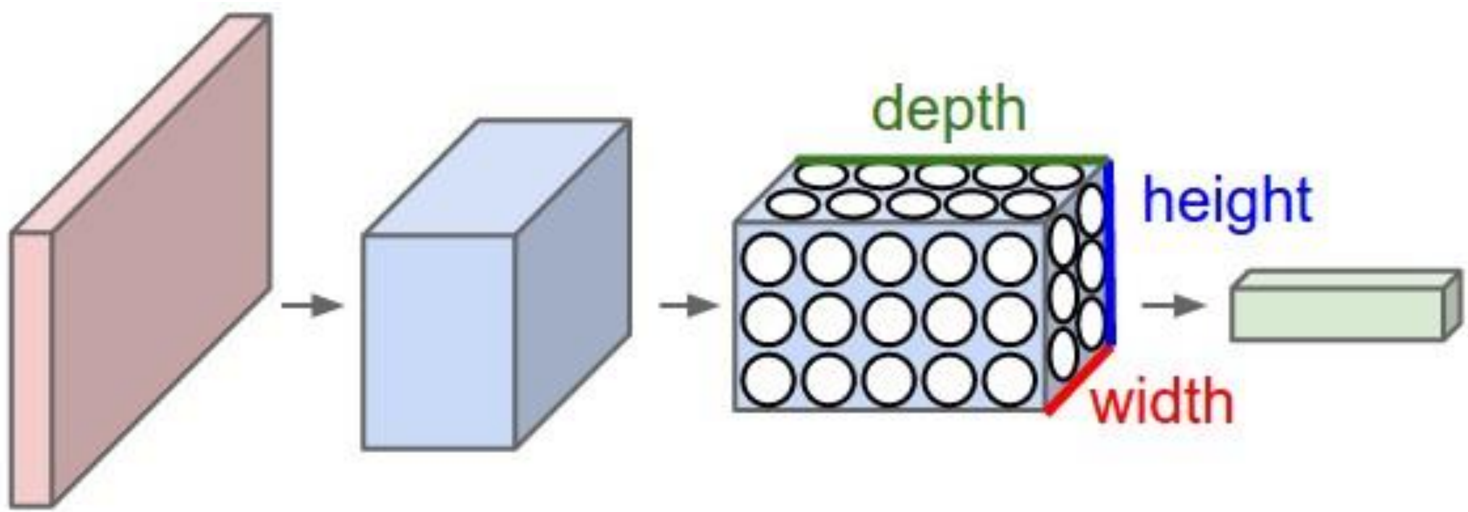


http://cs231n.stanford.edu/slides/2019/cs231n_2019_lecture02.pdf

早期图像分类模型

总结：对图像分类
- kNN等传统机器学习方法耗时大
- 线性分类模型，MLP参数大多，不适用图像数据

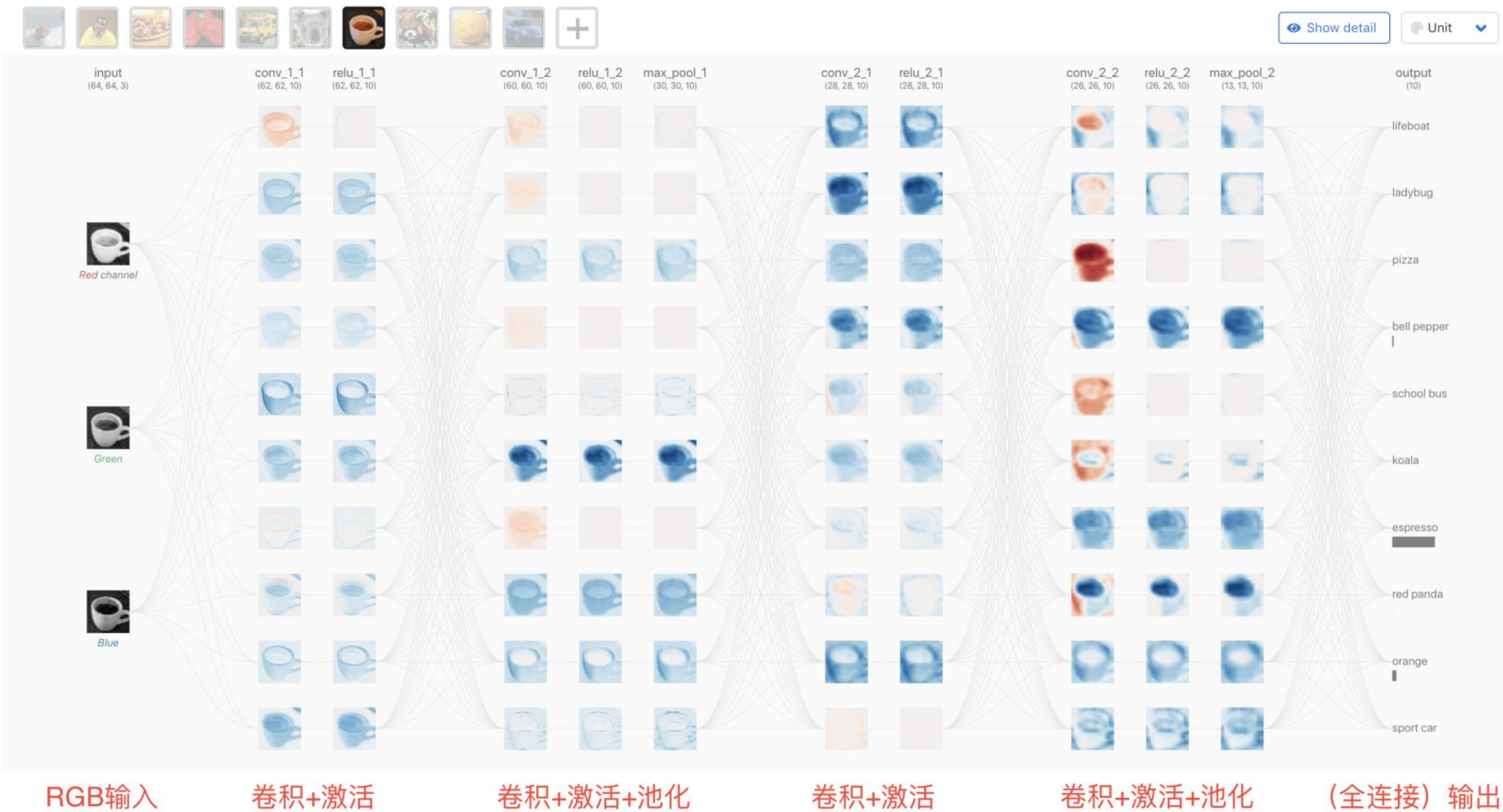
CNN：卷积神经网络
大大减少参数数量，提高识别准确度



http://cs231n.stanford.edu/slides/2019/cs231n_2019_lecture02.pdf

CNN

卷积神经网络例子

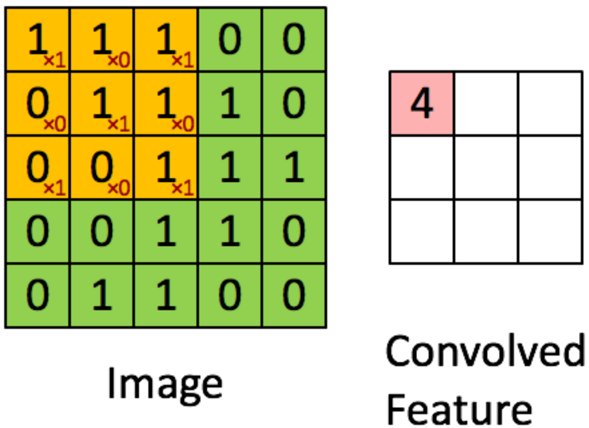


https://poloclub.github.io/cnn-explainer/

Datawhale

CNN基础

卷积操作



- 感受野 (receptive field)
- 卷积核 (filter)
- 卷积 (convolution)

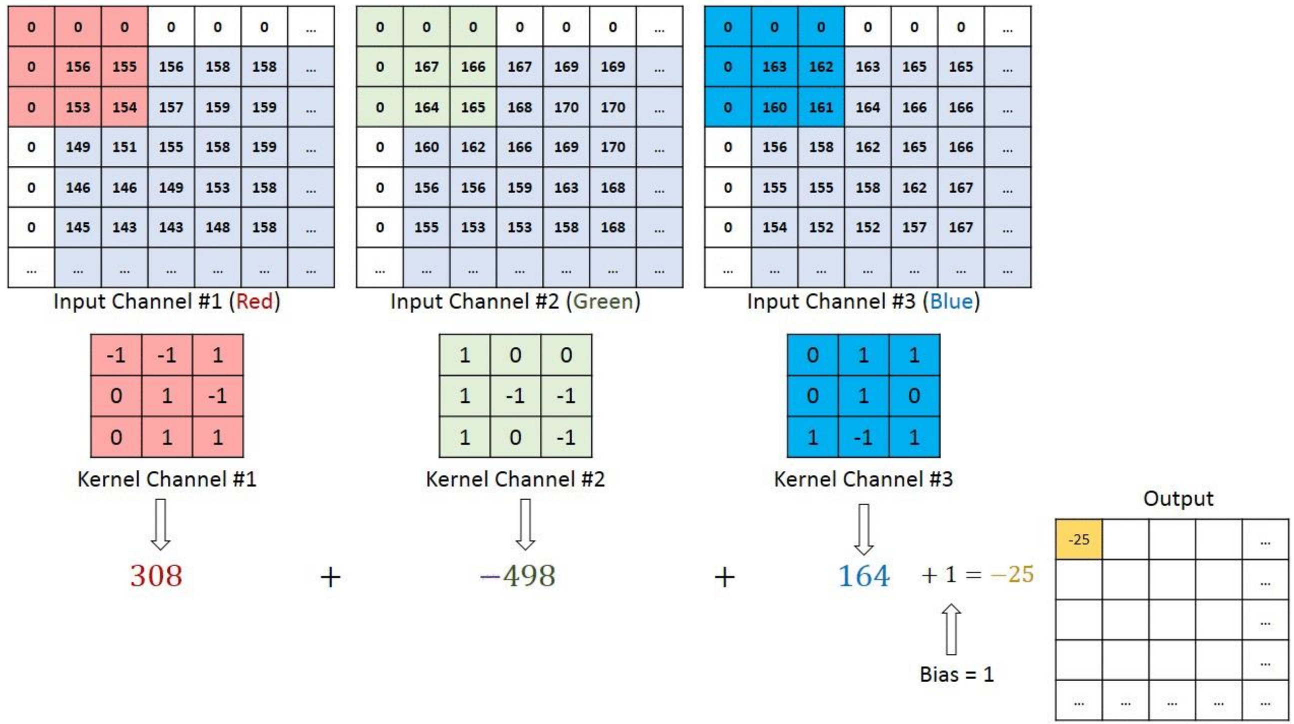
得到结果:

- 单张图片
- 单个卷积核
- 输出单通道图片

理解CNN

多输入通道卷积

多个输入通道的图片
对应多个卷积核
输出单通道图片



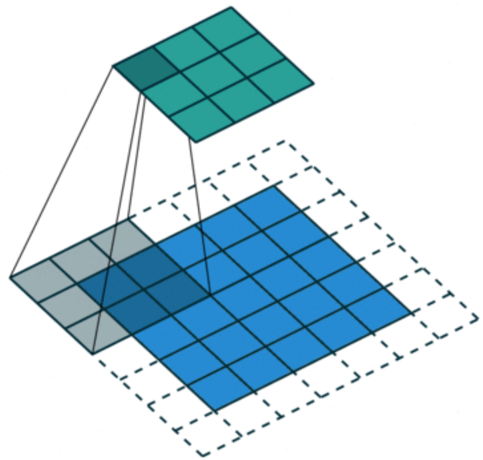
多输出通道卷积

多个通道的图片
对应多组卷积核
输出多通道图片

Datawhale

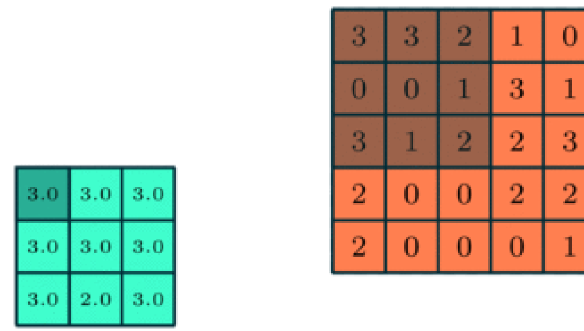
理解CNN

步长与填充
(stride & padding)



stride=2
padding=1

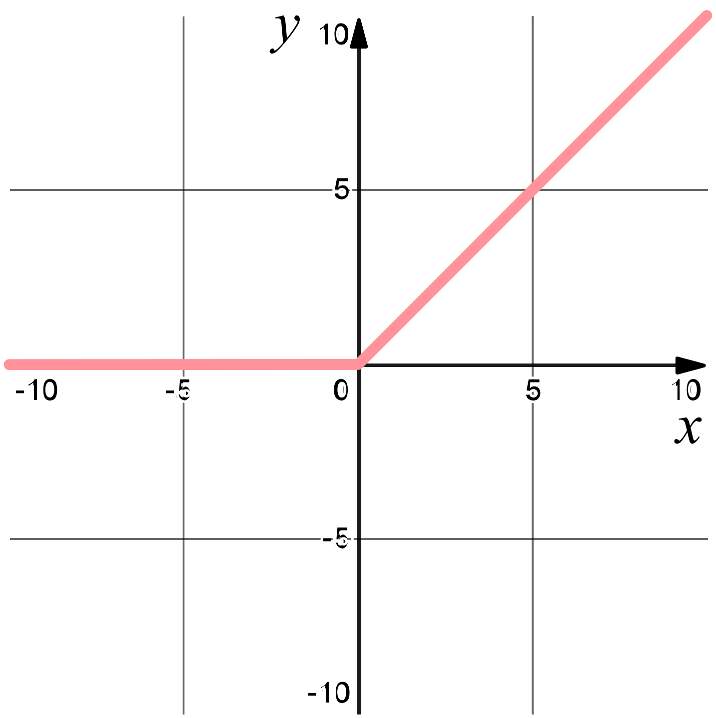
池化层
(pooling layer)



stride=1

理解CNN

激活函数——Relu



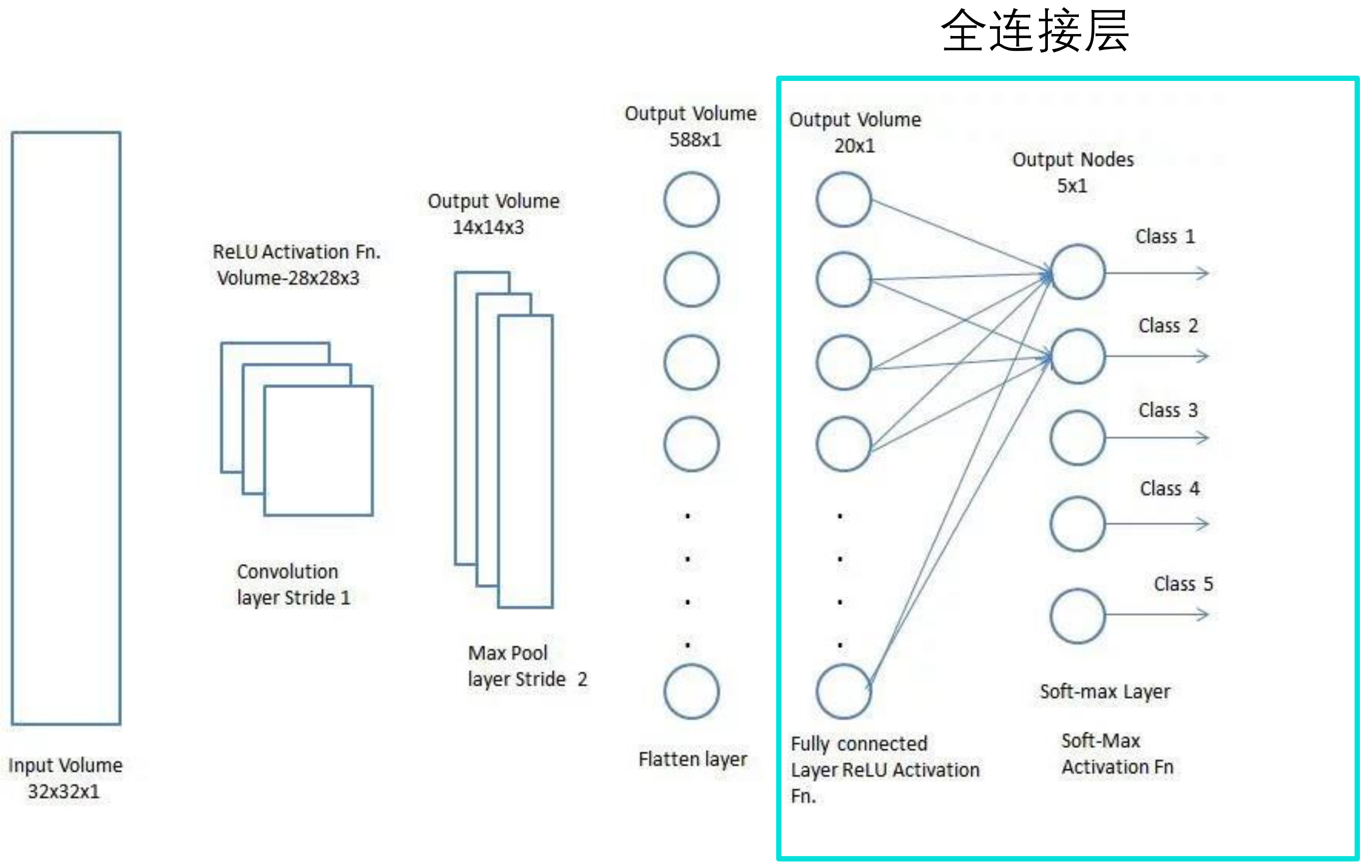
$$W_1(W_2X) = ?$$

$$W_1\sigma(W_2X) = ?$$

目的：引入非线性因素，
让网络的可表达性更强

理解CNN

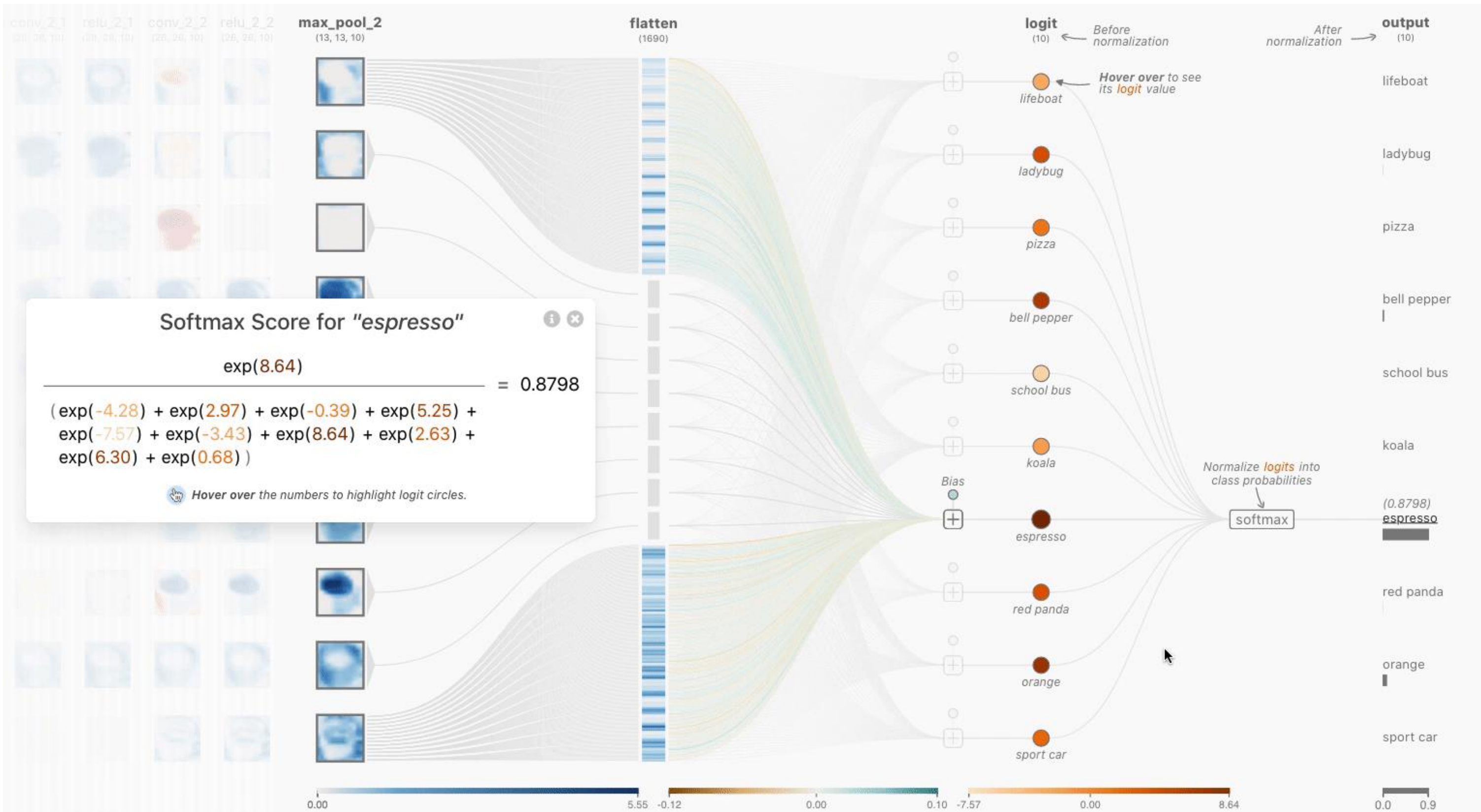
Datawhale



理解CNN

Softmax层

任意数字→0~1间的概率



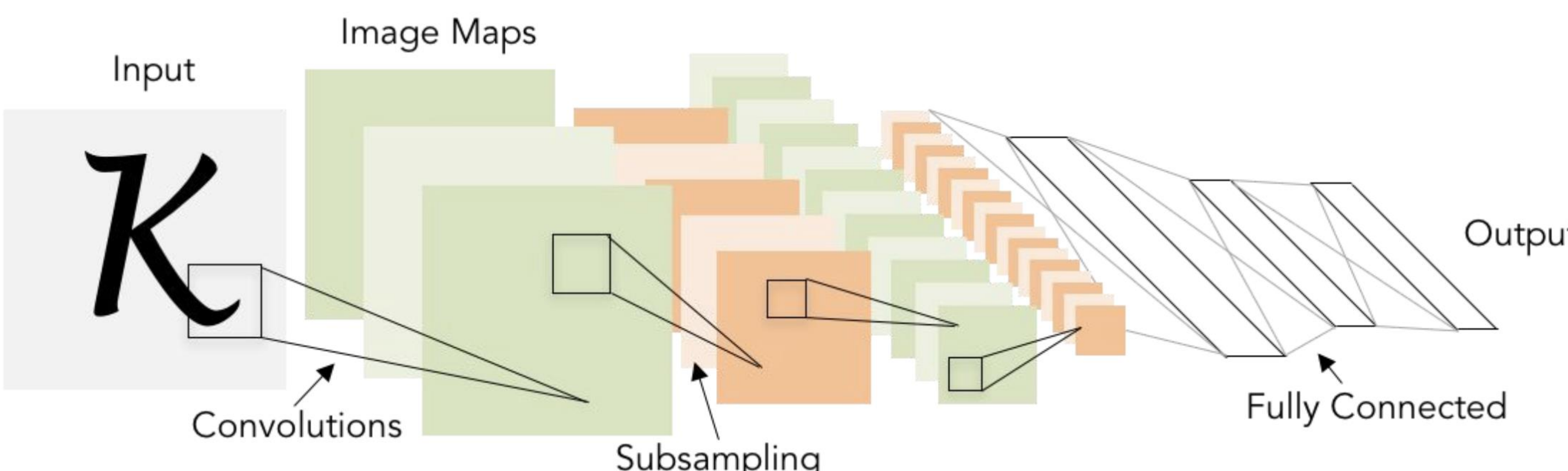
Part 2 案例学习

案例学习

Datawhale

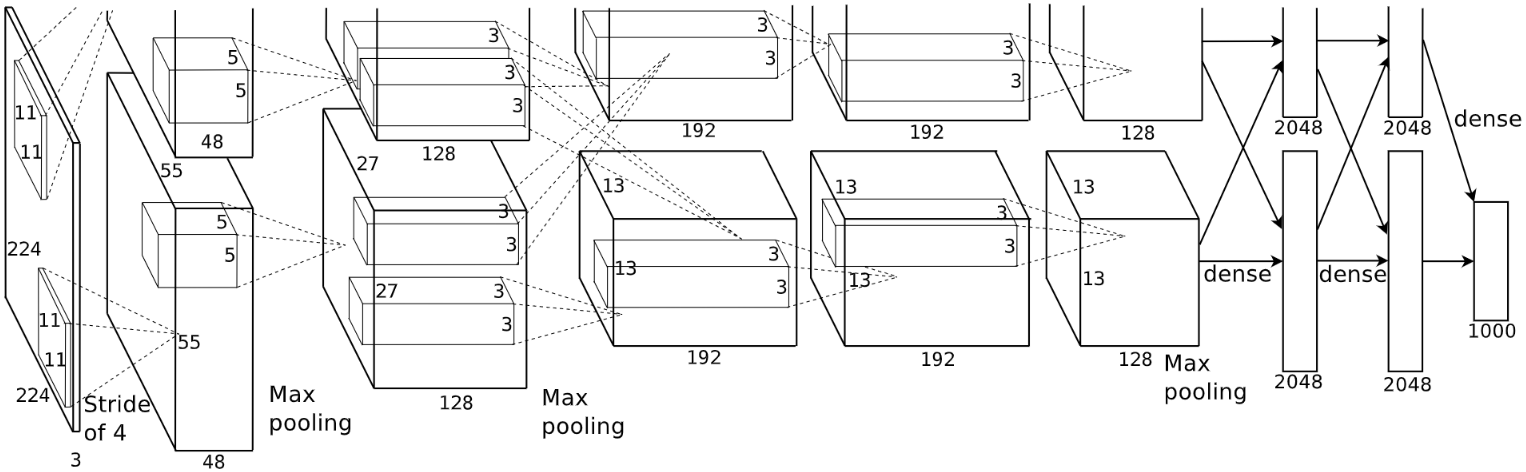
LeNet（LeCun et al., 1998 ）

- 卷积核5x5，步长1
- 下采样（池化）2x2，步长为2



案例学习

AlexNet（Krizhevsky et al. 2012）



输入：227x227x3 RGB图像

CONV1：96 组 11x11的卷积核，步长为4

练习：第一层输出的特征图大小是多少？

宽和高：（227-11）/ 4+1=55

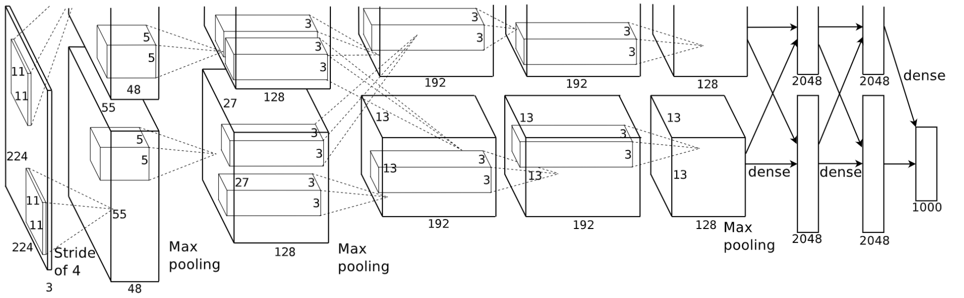
尺度 55x55x96

Datawhale

案例学习

AlexNet（Krizhevsky et al. 2012）

[227x227x3] INPUT
[55x55x96] CONV1: 96 11x11 filters at stride 4, pad 0
[27x27x96] MAX POOL1: 3x3 filters at stride 2
[27x27x96] NORM1: Normalization layer
[27x27x256] CONV2: 256 5x5 filters at stride 1, pad 2
[13x13x256] MAX POOL2: 3x3 filters at stride 2
[13x13x256] NORM2: Normalization layer
[13x13x384] CONV3: 384 3x3 filters at stride 1, pad 1
[13x13x384] CONV4: 384 3x3 filters at stride 1, pad 1
[13x13x256] CONV5: 256 3x3 filters at stride 1, pad 1
[6x6x256] MAX POOL3: 3x3 filters at stride 2
[4096] FC6: 4096 neurons
[4096] FC7: 4096 neurons
[1000] FC8: 1000 neurons (class scores)

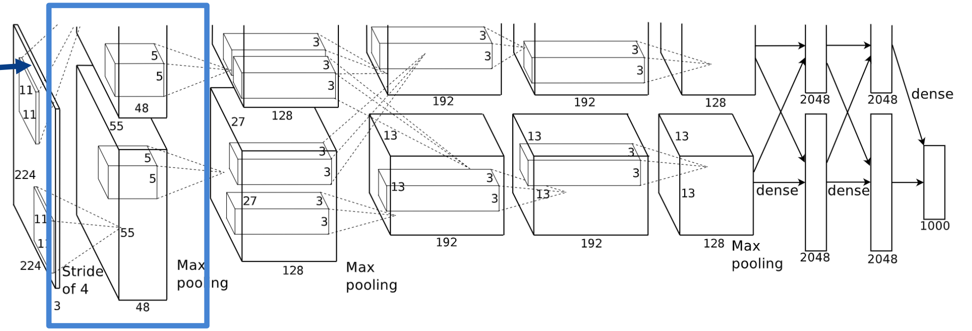


细节
- 首先使用得ReLU
- 使用LRN层（现在用的不多）
- 数据增强
- 丢弃法，概率0.5
- 批量大小128
- SGD增量0.9
- 学习率1e-2

案例学习

AlexNet（Krizhevsky et al. 2012）

[227x227x3] INPUT
[55x55x96] CONV1: 96 11x11 filters at stride 4, pad 0
[27x27x96] MAX POOL1: 3x3 filters at stride 2
[27x27x96] NORM1: Normalization layer
[27x27x256] CONV2: 256 5x5 filters at stride 1, pad 2
[13x13x256] MAX POOL2: 3x3 filters at stride 2
[13x13x256] NORM2: Normalization layer
[13x13x384] CONV3: 384 3x3 filters at stride 1, pad 1
[13x13x384] CONV4: 384 3x3 filters at stride 1, pad 1
[13x13x256] CONV5: 256 3x3 filters at stride 1, pad 1
[6x6x256] MAX POOL3: 3x3 filters at stride 2
[4096] FC6: 4096 neurons
[4096] FC7: 4096 neurons
[1000] FC8: 1000 neurons (class scores)

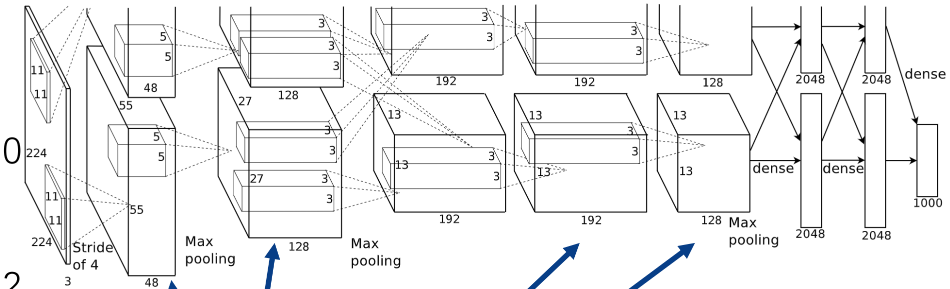


历史因素:
双GPU训练， GTX 580 GPU 3GB,
特征图在通道上等分到了两个GPU

Datawhale

案例学习

AlexNet（Krizhevsky et al. 2012）



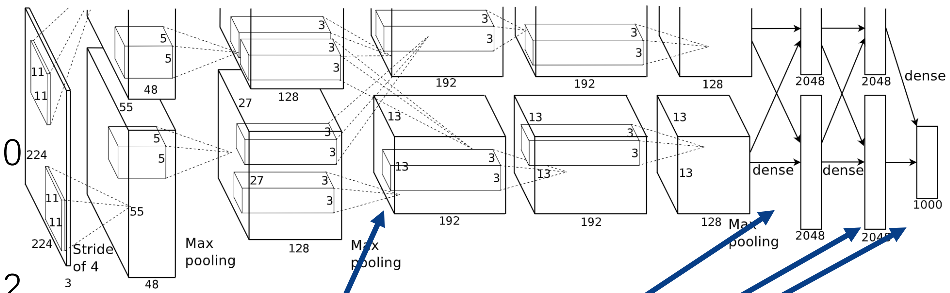
[227x227x3] INPUT
[55x55x96] CONV1: 96 11x11 filters at stride 4, pad 0
[27x27x96] MAX POOL1: 3x3 filters at stride 2
[27x27x96] NORM1: Normalization layer
[27x27x256] CONV2: 256 5x5 filters at stride 1, pad 2
[13x13x256] MAX POOL2: 3x3 filters at stride 2
[13x13x256] NORM2: Normalization layer
[13x13x384] CONV3: 384 3x3 filters at stride 1, pad 1
[13x13x384] CONV4: 384 3x3 filters at stride 1, pad 1
[13x13x256] CONV5: 256 3x3 filters at stride 1, pad 1
[6x6x256] MAX POOL3: 3x3 filters at stride 2
[4096] FC6: 4096 neurons
[4096] FC7: 4096 neurons
[1000] FC8: 1000 neurons (class scores)

conv1，conv2，conv4，conv5
和特征图的连接只在单GPU上

Datawhale

案例学习

AlexNet（Krizhevsky et al. 2012）



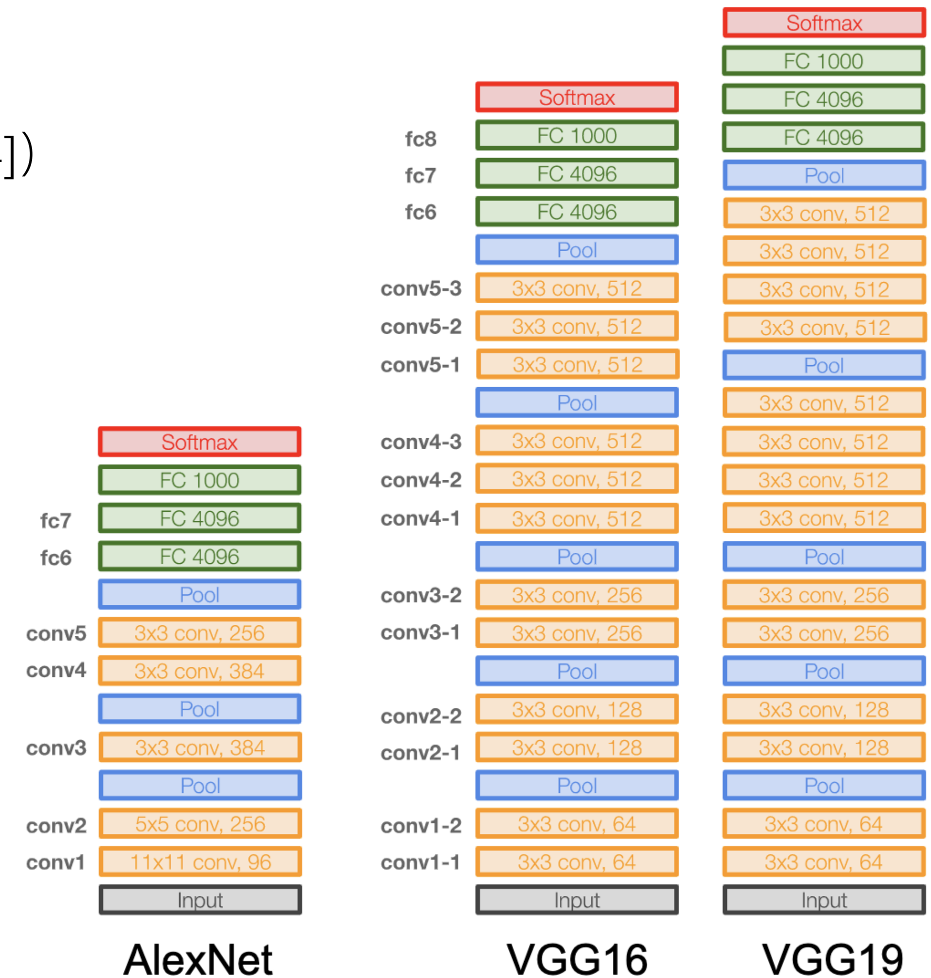
[227x227x3] INPUT
[55x55x96] CONV1: 96 11x11 filters at stride 4, pad 0
[27x27x96] MAX POOL1: 3x3 filters at stride 2
[27x27x96] NORM1: Normalization layer
[27x27x256] CONV2: 256 5x5 filters at stride 1, pad 2
[13x13x256] MAX POOL2: 3x3 filters at stride 2
[13x13x256] NORM2: Normalization layer
[13x13x384] CONV3: 384 3x3 filters at stride 1, pad 1
[13x13x384] CONV4: 384 3x3 filters at stride 1, pad 1
[13x13x256] CONV5: 256 3x3 filters at stride 1, pad 1
[6x6x256] MAX POOL3: 3x3 filters at stride 2
[4096] FC6: 4096 neurons
[4096] FC7: 4096 neurons
[1000] FC8: 1000 neurons (class scores)

conv3，FC6，FC7，FC8
在GPU间进行交流

案例学习

VGG（[Simonyan and Zisserman, 2014]）

- ILSRC14分类任务第2名，定位任务第一名
- 训练步骤和AlexNet类似
- 没有LRN
- VGG16和VGG19，后者性能稍好
- 使用了3个3x3卷积核来代替7x7卷积核，使用了2个3x3卷积核来代替5*5卷积核



案例学习

Datawhale

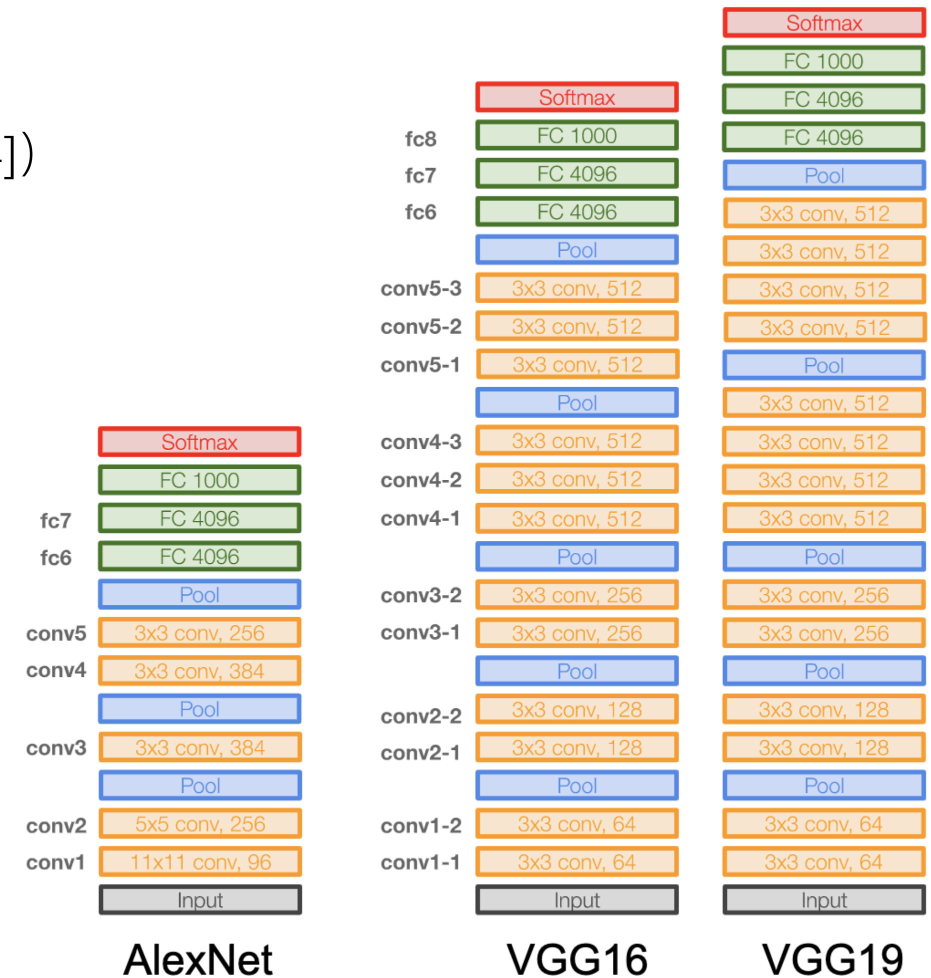
VGG（[Simonyan and Zisserman, 2014]）

Q1：为什么用小卷积核？

- 堆叠的小卷积核和大卷积核的接收野相同
- 参数更少$3C^2\ vs.\ 7^2C^2$

Q2：3个3x3的卷积核接收野是多少？

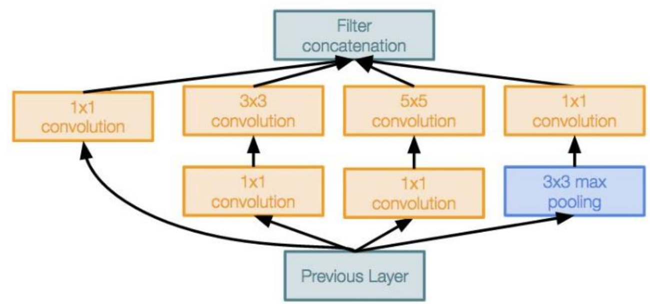
- 7x7



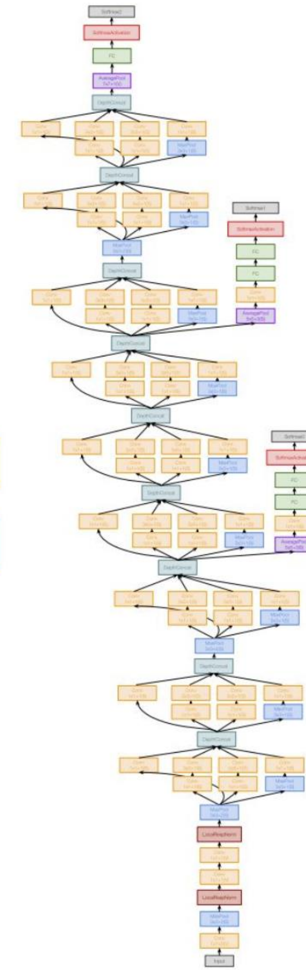
案例学习

GoogLeNet（ Szegedy et al., 2014）

- 22 层
- 高效地Incept模块
- 没有全连接层
- 500万参数，仅为AlexNet的1/12
- ILSVRC14分类挑战赛冠军



Inception module

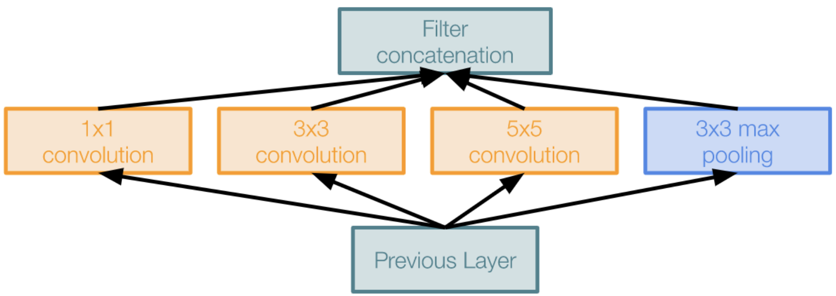


Datawhale

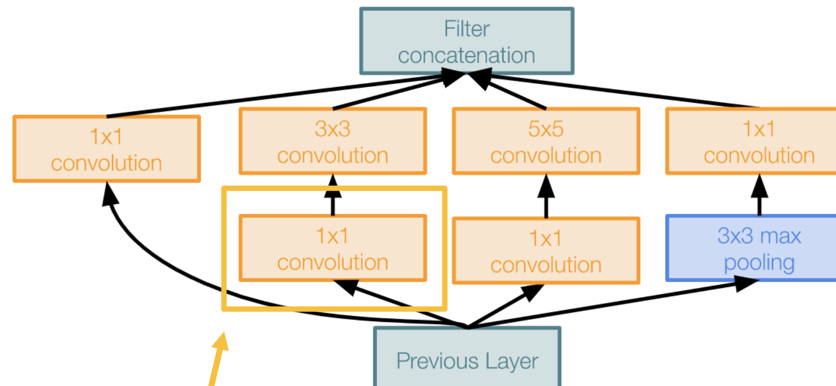
案例学习

GoogLeNet（Szegedy et al., 2014）

两种Inception模块



Naive Inception module



Inception module with dimension reduction

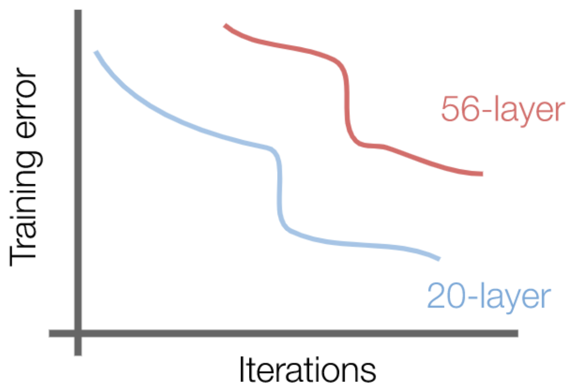
- 并行的4路操作：3个卷积，1个池化
- 输出在通道上连接
- 计算消耗大

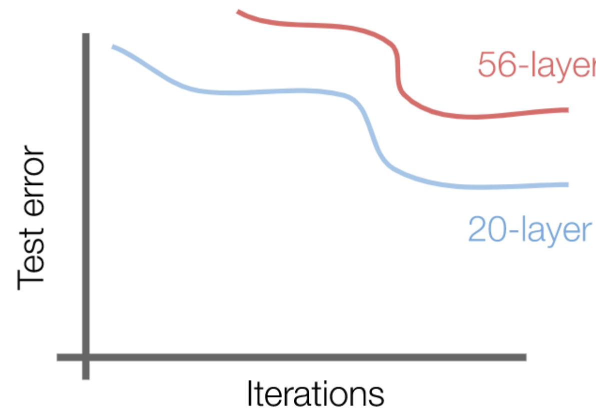
降维：减少通道数，即减少了参数

案例学习

ResNet([He et al., 2015])

网络越深，误差越大。这不是过拟合的原因，而是一个优化的问题，网络越深，越难优化

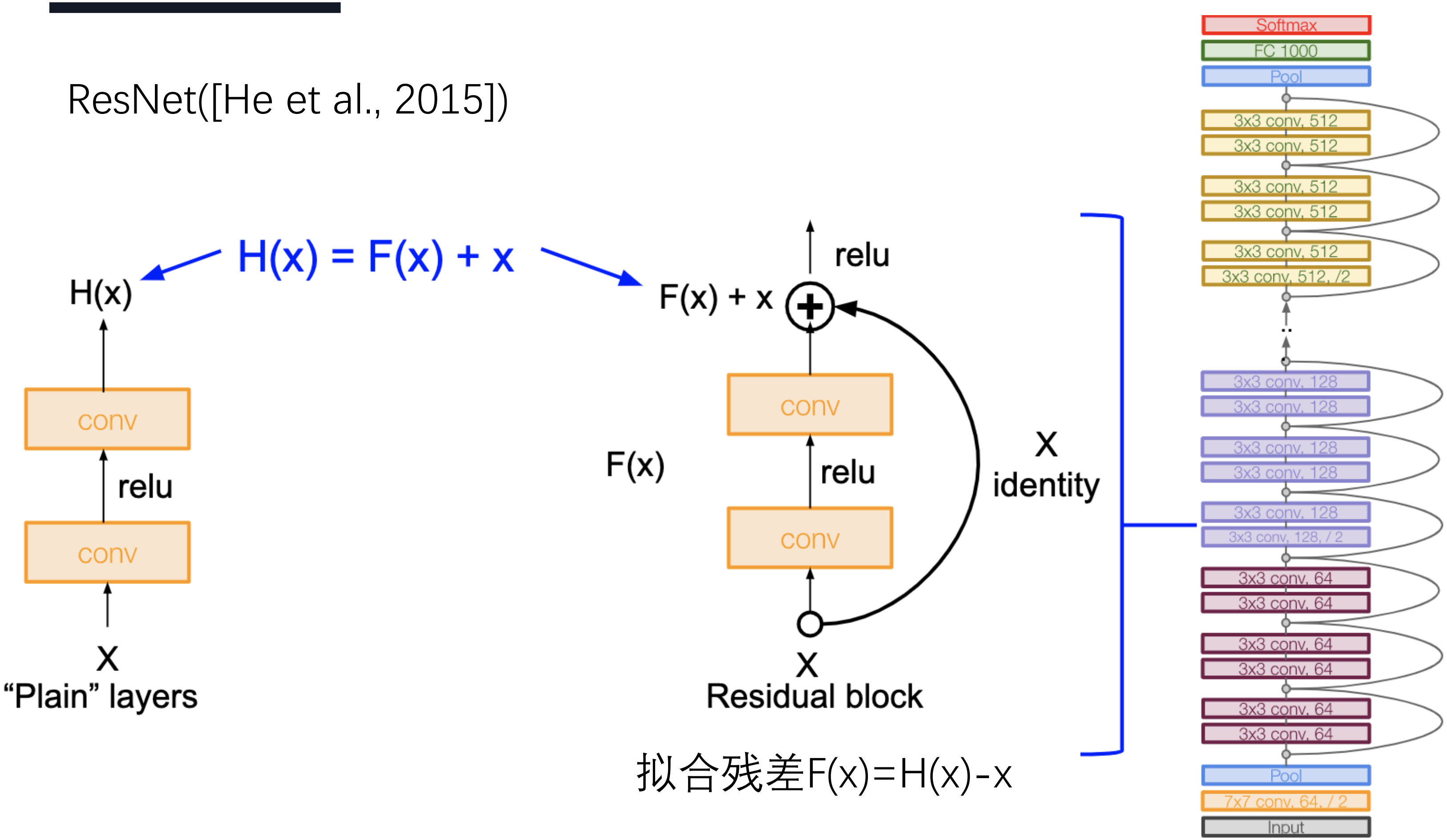




解决：希望更深的模型至少要比更浅的模型要好，将前面学习的内容copy到后面来

案例学习

ResNet([He et al., 2015])



拟合残差F(x)=H(x)-x

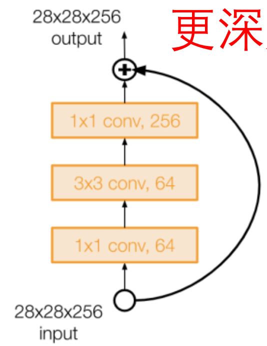
案例学习

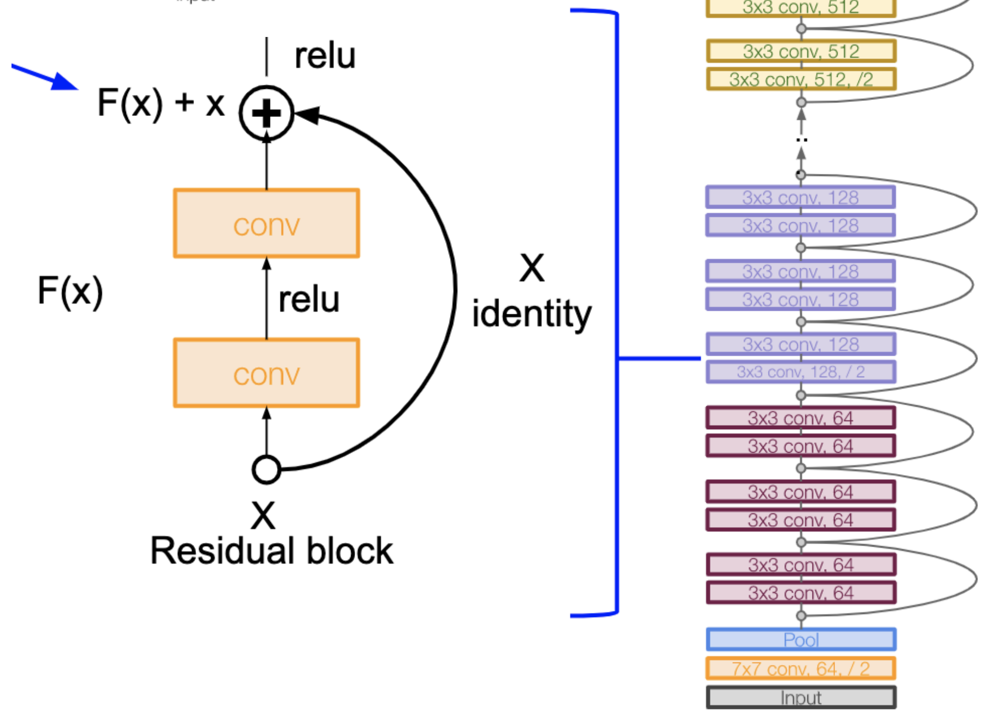
ResNet([He et al., 2015])

模型细节

- 残差块为基本单元
- 每个残差块有两个3x3卷积
- 周期性地2倍增卷积核组数，和使用下采样2倍减图像尺寸
- 网络初始有单独的卷积层
- 只有一个全连接层

更深层使用Bottleneck





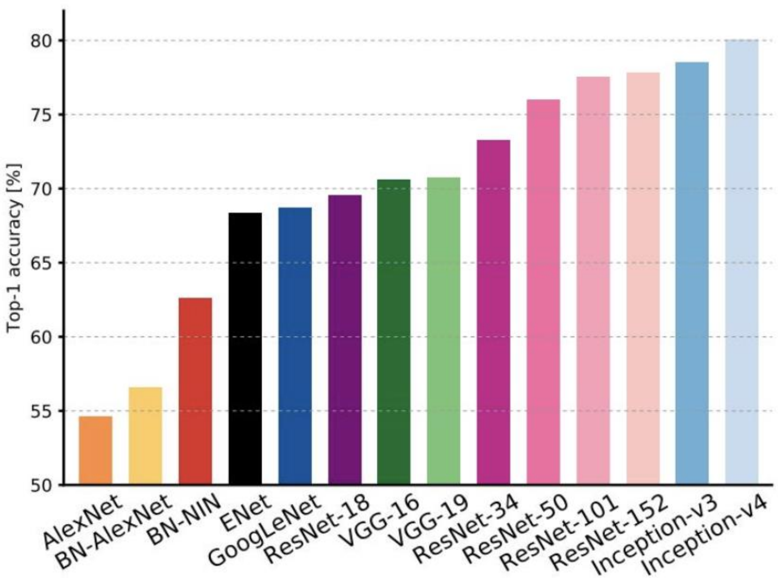
Datawhale

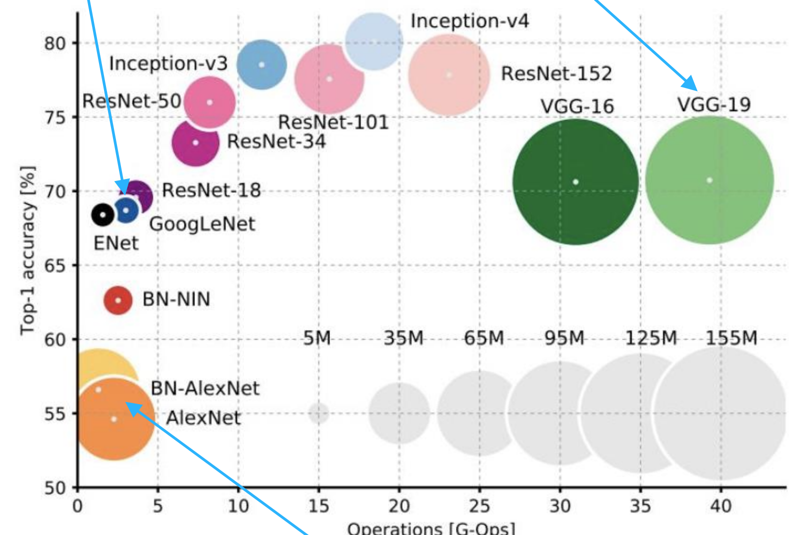
Datawhale

网络比较

GoogLeNet：最高效

VGG：最高内存，最多计算





An Analysis of Deep Neural Network Models for Practical Applications, 2017.

AlexNet：精度低，
内存大，计算少

总结

如何提升分类模型性能？

- 改善ResNet：SENet等
- dropout
- batch normalization
- 1x1卷积
- more…

Datawhale CV小组开源项目：动手学CV-Pytorch版
https://github.com/datawhalechina/dive-into-cv-pytorch

本节讲解的各种模型比较：
https://github.com/QiangZiBro/cnn_models_comparation.pytorch

参考

1.CS231N http://cs231n.stanford.edu/
2.CS230 https://cs230.stanford.edu/
3.CNN explainer https://poloclub.github.io/cnn-explainer/
4.LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.
5. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., & Anguelov, D. & Rabinovich, A.(2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).
6. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
7. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Datawhale

Part 3 Q&A

问答

- 你对比赛有什么问题?
- 你对学习有什么问题?
- 你对PPT内容有什么问题?

Datawhale

一个专注于AI领域的开源组织