

深度学习的图像应用

Deep Learning for Image Application

Garry

大数据部门

2017 年 8 月 1 日

概要

① 从零实现一个神经网络

- 生成数据集
- 训练神经网络
- 预测

② 图像分类

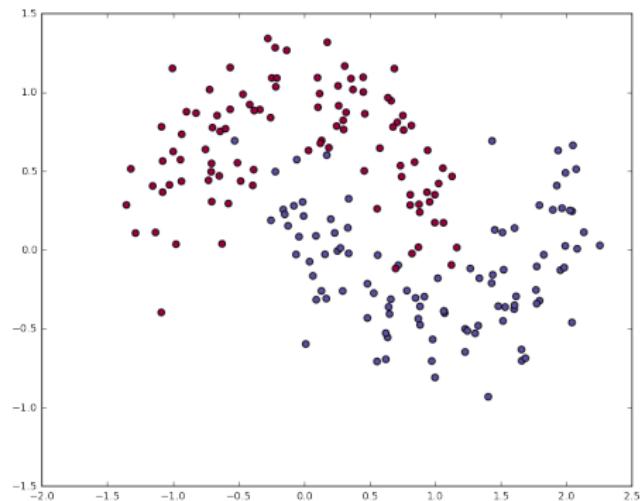
- AlexNet
- Demo

③ 图像检索

- 企业应用实例
- Demo

从零实现一个神经网络

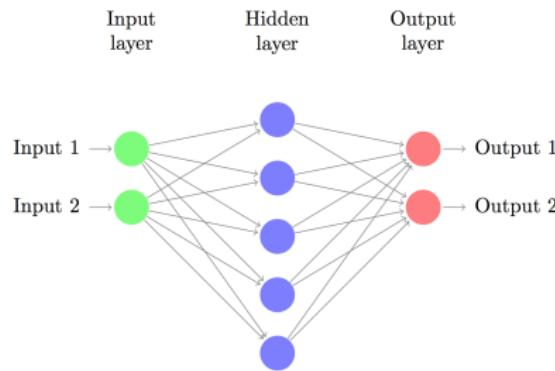
数据集



图：生成数据集

```
# Generate a dataset and plot it
np.random.seed(0)
X, y = sklearn.datasets.make_moons(200, noise=0.20)
plt.scatter(X[:,0], X[:,1], s=40, c=y, cmap=plt.cm.Spectral)
```

网络结构



图：网络结构

建立一个三层的神经网络（一个输入层，一个隐藏层，一个输出层）。输入层的维度由我们的数据决定。同样的，输出层的维度由类别的个数决定。

我们的网络如何预测?

$$\begin{aligned} z_1 &= xW_1 + b_1 \\ a_1 &= \tanh(z_1) \\ z_2 &= a_1 W_2 + b_2 \\ a_2 &= \hat{y} = \text{softmax}(z_2) \end{aligned} \tag{1}$$

- z_i 表示第 i 层的输入¹
- a_i 表示第 i 层经过激活函数运算后的输出
- W_1, b_1, W_2, b_2 是我们网络的参数, 也就是我们要从训练网络中学习的

如果我们的隐藏层有 500 个节点, 则 $W_1 \in \mathbb{R}^{2 \times 500}$, $b_1 \in \mathbb{R}^{500}$, $W_2 \in \mathbb{R}^{500 \times 2}$, $b_2 \in \mathbb{R}^2$.

¹ i 从 0 开始计数

学习参数

如果有 N 个训练样本, C 个类别, 则我们预测的 \hat{y} 关于实际类别 y 的损失如下:

$$L(y, \hat{y}) = -\frac{1}{N} \sum_{n \in N} \sum_{i \in C} y_{n,i} \log \hat{y}_{n,i} \quad (2)$$

我们的目标是找到是损失函数最小的参数.

梯度下降

使用梯度下降 (gradient descent) 来寻找最小值.

损失函数关于参数的梯度 (导数): $\frac{\partial L}{\partial W_2}, \frac{\partial L}{\partial b_2}, \frac{\partial L}{\partial W_1}, \frac{\partial L}{\partial b_1}$.

$$\delta_3 = \hat{y} - y$$

$$\delta_2 = (1 - \tanh^2 z_1) * \delta_3 W_2^T$$

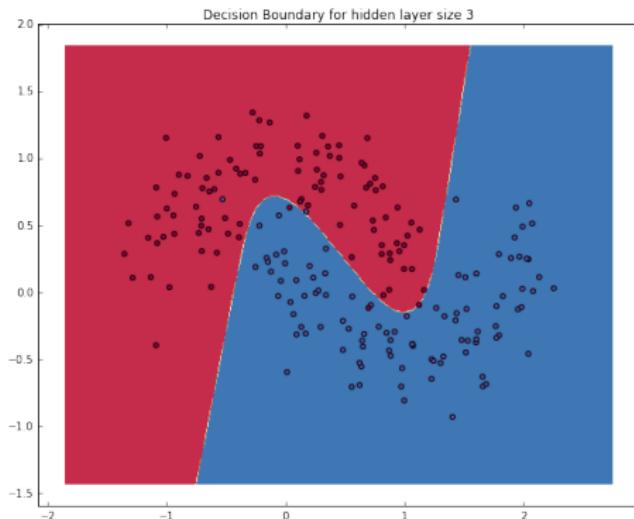
$$\frac{\partial L}{\partial W_2} = a_1^T \delta_3$$

$$\frac{\partial L}{\partial b_2} = \delta_3 \quad (3)$$

$$\frac{\partial L}{\partial W_1} = x^T \delta_2$$

$$\frac{\partial L}{\partial b_1} = \delta_2$$

预测结果



图：预测结果

图像分类

ImageNet

ImageNet 数据集: ImageNet 是一个按照 WordNet 层次组织的图像数据库（目前只有名词）。层次上的每个节点都有几百或几千个图像表示。目前，平均每个节点有超过 5 千张图像。²

²<http://www.image-net.org/>

ILSVRC2012

ILSVRC2012: 全称 Large Scale Visual Recognition Challenge 2012. 这个比赛的目的是利用人工标记的 ImageNet 的子集进行训练来判断照片中的内容, 从而实现检索和自动标记. 整体的目标是识别图像上出现的主要物体.³

³<http://image-net.org/challenges/LSVRC/2012/index>

ILSVRC2012 图像分类冠军: ImageNet Classification with Deep
Convolutional Neural Networks(Krizhevsky, Sutskever, and Hinton 2012)

AlexNet 结构

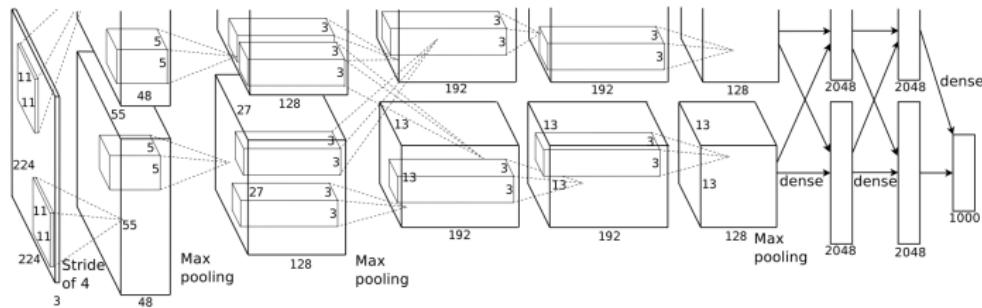


图: AlexNet 的 CNN 架构

数据集 **Caltech256**: 包含从 Google 图像搜索和 PicSearch.com 上获得的 30,607 张物体的图像. 这些图像通过人工判别被分配在 257 个类别中.

代码 多伦多大学讲师 Michael Guerzhoy 在网站上提供了 AlexNet 的 TensorFlow 实现及权重 (weights).⁴

⁴http://www.cs.toronto.edu/~guerzhoy/tf_alexnet/

Caltech256 图片示例-背包

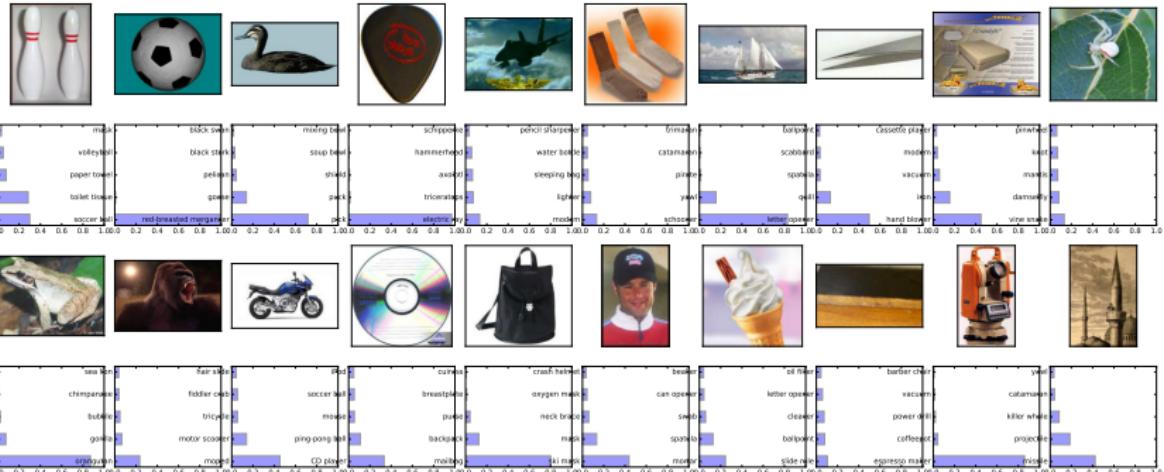


图：背包

Caltech256 图片示例-咖啡杯



图: 咖啡杯



图：分类结果

现场演示 (classification.py).....

技术前沿 (State of the Art)

模型	Top 5 准确率
ResNet 152	92.92%
ResNet 101	92.63%
ResNet 50	92.02%
VGG 16	89.88%
GoogLeNet	89.06%
Network in Network	81.21%
CaffeNet	79.93%
AlexNet	79.84%

表: 模型的准确率验证⁵

⁵<https://github.com/ethereon/caffe-tensorflow>

图像检索

Pinterest

Pinterest 是一个网络与手机的应用程序,可以让用户利用其平台作为个人创意及项目工作所需的视觉探索工具,同时也有人把它视为一个图片分享类的社交网站,用户可以按主题分类添加和管理自己的图片收藏,并与好友分享。⁶

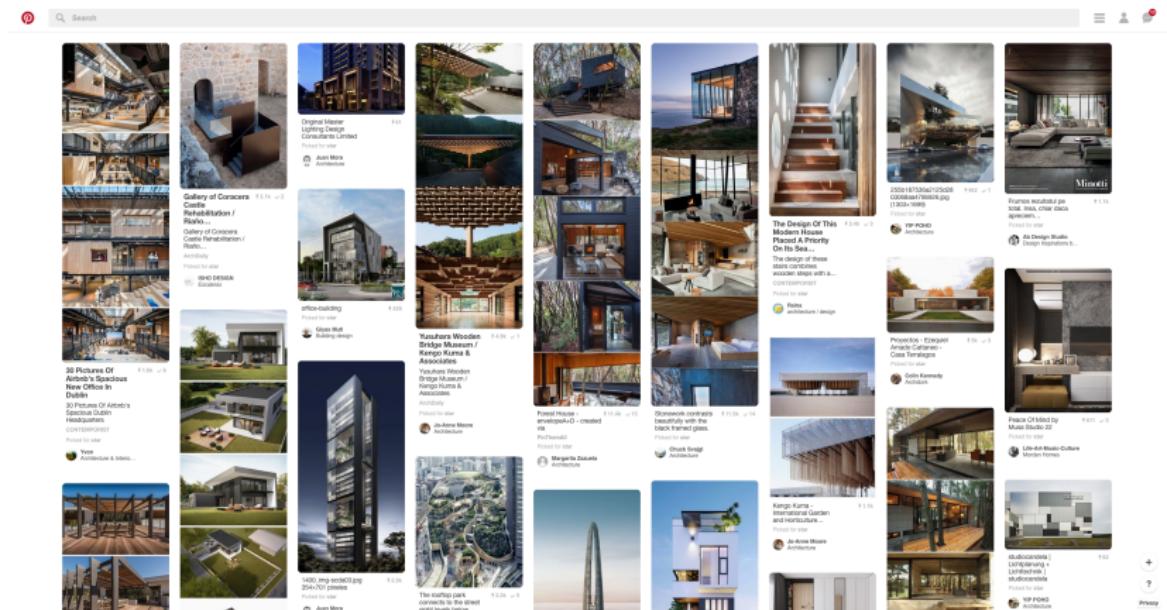


图: Pinterest 网站

⁶<https://zh.wikipedia.org/wiki/Pinterest>

论文

Pinterest 在 2014 年收购了初创公司 VisualGraph, 在它的平台上引入了视觉搜索. 在 2015 年,Pinterest 发表了一篇论文 Visual Search at Pinterest(Jing et al. 2015), 公开了它的系统架构:⁷

- Apache Hadoop
- Caffe 卷积神经网络框架 (AlexNet,VGG,Hamming Distance)
- Cascading(批量处理)
- PinLater(消息)
- Apache HBase(存储)

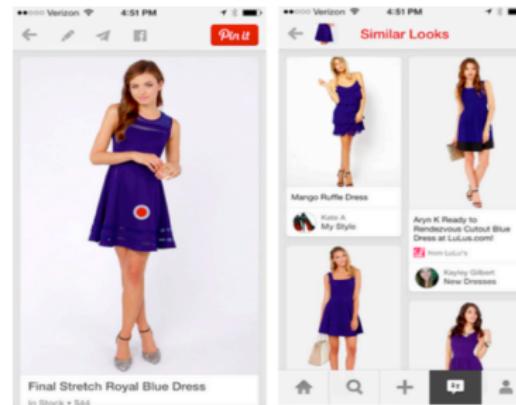


图: 用户一点击红点, 系统会显示与检索物体相似的产品

⁷https://en.wikipedia.org/wiki/Reverse_image_search

实验步骤

- Caltech256 数据集作为待检索的图片库
- 抽取 AlexNet 的 fc6 输出作为的图像的特征 (4096 维)
- 选择 cosine 作为图像特征之间距离 (效果比欧式距离好)
- 检索时, 先抽取输入图像的特征, 计算与图片库中每个图像的距离, 然后排序

计算向量 u 与 v 之间的相关距离:

$$1 - \frac{u \cdot v}{\|u\|_2 \|v\|_2} \quad (4)$$

实例



图：图像检索实例

图像变换检索实验



图：图像变换检索（文字水印，旋转（nearest），裁剪，镜像）

速度（一次实验，非统计平均）query cost: 7.590151, dataset: 30185, query: 5

现场演示 (visual_search.py).....

淘宝图片搜索

淘宝网 Taobao.com

更多市场

搜索

上传图片就能搜同款啦!

分类导航：上衣、裙装、下装、箱包、鞋、配饰、零食、美妆、**瓶饮**、家具、玩具

筛选条件：女 帧选主体

外观相似宝贝



¥46.00 包邮

0人付款

雄泰真空304不锈钢保温杯男女水杯户外便携茶杯办公室带漏情侣杯

天成五金机电

河北 石家庄



¥48.00

0人付款

功夫茶具套装耐热耐高温玻璃茶具
花茶大容量飘逸杯茶吹制玻璃杯

bg4uby

江苏 扬州



¥69.00 包邮

2人付款

爱屋格林手绘陶瓷盖+印花修身马克杯
(礼盒装),时尚创意带盖杯子

买买买moneymoneymoney



100%正品 全国包邮

¥27.50 包邮

0人付款

时尚运动水杯 D-1002 隔水茶壶 泡茶
杯简约防漏密封杯子 带茶漏

给力每天10

浙江 宁波



淘宝图像搜索存在的问题

搜索出来的商品类别与上传图片并不太一致，比如上传一张马克杯图像，返回结果排在前面的是保温杯、茶具等。

可能的方案：先通过图像分类算法，识别物体类别，且在计算图片库特征时也按类别将图像分开，在搜索时只搜索相关分类的图像，这样还能减少计算量，加快检索速度。

可以改进的地方

- ① 实验过程中对图像的预处理比较“粗暴”
- ② 特征相似度衡量方法的选择比较简单
- ③ 检索时查询全库，未做优化方面的探究

谢谢!

参考文献

Jing, Yushi et al. (2015). "Visual search at pinterest". In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, pp. 1889–1898.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton (2012). "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*, pp. 1097–1105.