

Machine Learning Engineer Nanodegree

Capstone Proposal

俞伟山

August 30th, 2018

Proposal

猫狗大战 ([Dogs vs. Cats Redux: Kernels Edition](#)) 。

项目背景

此项目最早源于 `kaggle` 2013年的[Dogs vs. Cats](#)比赛。那时，网站为了防止恶意攻击，一般会提供一些验证问题，用来区别人和机器，即 `CAPTCHA` (Completely Automated Public Turing test to tell Computers and Humans Apart)。这些问题，要设计得容易让人解决，而让计算机不好解决。在那时，`Asirra` (Animal Species Image Recognition for Restricting Access)，就是人容易解决，而计算机不好解决的问题。

猫狗大战属于图像识别的问题，那时已经有了一些机器学习算法应用于图像识别，文献[1]使用机器学习算法，可以在狗狗的图像识别中，达到 `80%` 的分类准确率。

随着这几年机器学习的发展，特别是深度学习和图像分析的发展，各种深度学习框架、`ConvNet` 模型相继出现。`kaggle` 于2017年，再次举办了猫狗大战的比赛，排在第一名的，`LogLoss` 得分达到了 `0.03302`。

图像识别问题，属于 `计算机视觉` 领域。`计算机视觉` 是一个跨学科的领域，它处理计算机如何高度理解数字图像或视频的问题。它包含如何自动从图像和视频中抽取、分析和理解一些有用的信息。

图像识别是 `计算机视觉` 的典型问题。目前，最好的解决图像识别问题的算法是基于 `ConvNet` 的算法。2012年，`Alex Krizhevsky` 使用 `AlexNet` 在 `ImageNet` 举办的 `ILSVRC-2012` 中达到了 `15.3%` 的top-5错误率[2]，领先第二名的 `26.2%`，被认为是深度学习的革命。

本人对 `计算机视觉` 比较感兴趣，而用深度 `ConvNet` 来处理图像识别问题又是目前比较常见的操作，所以，我选择这个毕业项目。

问题描述

使用深度学习方法识别一张图片是猫还是狗。

- 输入：一张彩色图片
- 输出：狗的概率

数据集和输入

数据集使用 [Dogs vs. Cats Redux: Kernels Edition](#)提供的数据集，

可以用[kaggle api](#)工具，下载整个数据集：

```
kaggle competitions download -c dogs-vs-cats-redux-kernels-edition
```

数据集包含训练集和测试集。

训练集包含25,000张猫和狗的图片，猫和狗的图片各占一半，图片以 `label.id.jpg` 形式命名。

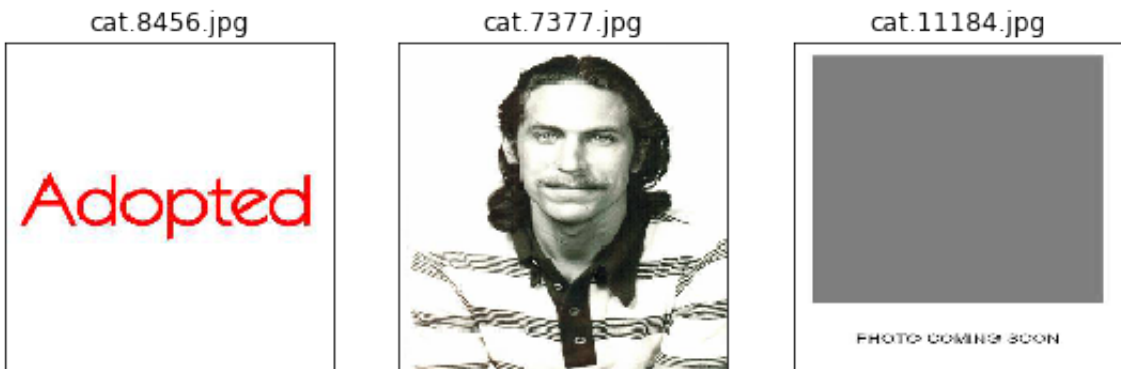
label	id	filename
cat	0	cat.0.jpg
cat	1	cat.1.jpg
cat
cat	12499	cat.12499.jpg
dog	0	dog.0.jpg
dog	1	dog.1.jpg
dog
dog	12499	dog.12499.jpg

测试集包含12,500张待分类的图片，图片根据序号命名。

观察图片，可以发现猫或狗的场景有很多。比如，拿猫的图片来讲，有些整屏都是猫，有些干扰信息比较多（比如和人的合影），有些图里有很多猫，有些还是卡通猫。



我们也会观察到一些异常值。如下图所示：



这几张图和猫或狗都没有关系，但训练数据把它们放到猫的分类下。

图片的像素大小和长宽比例都不固定，如下图所示：



对于图片大小不一致的问题，我们可以用 `keras` 的 `ImageDataGenerator` 或 `cv2` 的 `resize` 方法，把图片数据处理成 `ConvNet` 需要的输入shape。

我们还需要把训练数据分成训练集和验证集，而测试集只在最后测试时使用，避免测试数据渗透到训练过程中。通过 `sklearn` 的 `train_test_split` 方法，把训练集进一步分成0.8的训练集和0.2的验证集。

```
from sklearn.model_selection import train_test_split
X_train, X_valid, y_train, y_valid = train_test_split(X, y, test_size=0.2)
```

解决办法

解决这个图像识别的问题，仍使用 `ConvNet` 模型。我们将基于 `ImageNet` 上比较成熟的 `ConvNet` 模型，来解决猫狗大战问题。

基准模型

我们将使用 `VGGNet` [3] 作为基准模型。`VGGNet` 是牛津大学 `Visual Geometry Group` 和 `Google` 的 `DeepMind` 公司的研究员一起研发的深度卷积神经网络。`VGGNet`探索了卷积神经网络的深度与其性能之间的关系，通过反复堆叠3x3的小型卷积核和2x2的最大池化层，`VGGNet`成功地构筑了16~19层深的卷积神经网络。`VGGNet`相比之前state-of-the-art的网络结构，错误率大幅下降，并取得了ILSVRC 2014比赛分类项目的第2名和定位项目的第1名。

根据 `kaggle` 上目前的[Leaderboard](#)排行榜，要进入前10%，则 `LogLoss` 至少要少于 0.06127。我们把基准阈值设置为 0.06127。

评估指标

项目的评估指标参见 [Dogs vs. Cats Redux: Kernels Edition Evaluation](#)。它根据 `LogLoss` 来评估，`LogLoss` 越低越好。

`LogLoss` 定义如下：

$$\text{LogLoss} = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

其中：

- `n` 是测试集的大小。

- \hat{y}_i 是预测的图片是狗的概率。
- y_i 是真实值，是狗时为1，是猫时为0。
- $\log()$ 是自然对数。

在 `keras` 中，`loss函数` 使用 `binary_crossentropy` 即是 `LogLoss`。

设计大纲

`Sinno Jialin Pan` 的迁移学习调查，提出一种 `Feature-representation-transfer` 的迁移学习方法[4]。而文献[5] 采用迁移学习中特征提取的思想，使用 `ImageNet` 上预训练的模型(比如 VGG16)的特征，再加自己的全连接层，取得在小数据集上训练出自己非常强大的图片分类模型的效果。

借鉴迁移学习的思路，设计步骤如下：

1. 用 `opencv` 读取图片数据。
2. 用 `keras` 提供的 `ImageDataGenerator` 对图片数据大小进行处理并流化产生batch。
3. 用 `ImageNet` 上的预训练模型Model对图片数据进行预测。
4. 提取Model中最后一个卷积层的 `feature map`，并保存。
5. 综合多个Model的 `feature map` 作为输入层，添加我们的 `Dense全连接层`，并训练这个最终模型。

我们将首先尝试 `ImageNet` 上 `top-5 accuracy` 较高的 `xception` [6]、`Inception v3` [7]和 `DenseNet` [8]这几个网络的预训练结果。

参考文献

- [1] Philippe Golle. Machine Learning Attacks Against the Asirra CAPTCHA. 2008.
- [2] Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. 2012.
- [3] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015.
- [4] Sinno Jialin Pan and Qiang Yang Fellow. A Survey on Transfer Learning. IEEE, 2009.
- [5] François Chollet. Building powerful image classification models using very little data. <https://blog.keras.io/building-powerful-image-classification-models-using-very-little-data.html>. Published: 2018-01-29.
- [6] François Chollet. Xception: Deep Learning with Depthwise Separable Convolutions. 2017.
- [7] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Zbigniew Wojna. Rethinking the Inception Architecture for Computer Vision. 2015.
- [8] Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger. Densely Connected Convolutional Networks. 2016.