

# 机器学习纳米学位（进阶）开题报告

---

龚宇海

2018年4月24日

## 项目背景

猫狗大战来自于Kaggle上举办的一次有趣的竞赛，目标是训练一个模型，然后通过上传一张猫或狗的照片，来分辨它是猫，还是狗，涉及到机器学习领域中的深度学习和图像分析方向。

项目需要使用到卷积神经网络（Convolutional Neural Network, CNN），是神经网络的一种，它在图像识别和分类领域已被证明非常有效。在卷积神经网络中，卷积的主要目的是从输入图像中提取特征。通过使用输入数据中的小方块来学习图像特征，卷积保留了像素间的空间关系。

## 问题描述

从Kaggle上获得的数据来自于真实世界中不同地区不同人所拍摄的猫或狗的图像。图像的分辨率质量参差不齐，猫和狗的形态各异，品种多样，背景环境也非常复杂，包括可能含有的异常图片，比如图像中根本就没有猫或狗，而是人或建筑物。这些都增加了分类的难度。

## 输入数据

项目的数据集由Kaggle竞赛项目中提供，训练文件夹包含12,500张标记为猫的图像和12,500张标记为狗的图像。测试文件夹包含12,500个图像，根据数字ID命名，未做标记。对于测试集中的每个图像，模型需要预测图像是狗的概率（1=狗，0=猫），很明显，这是一个二分类问题。

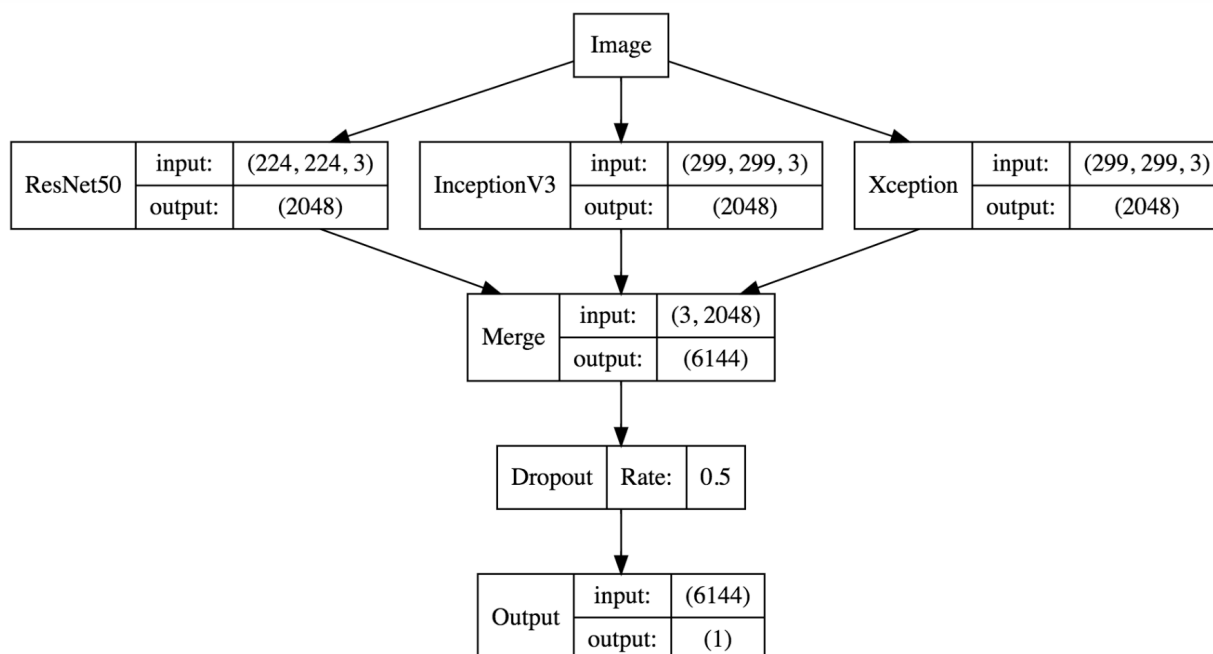
## 解决办法

项目使用开源机器学习库Keras作为主要开发框架，以最流行的TensorFlow作为Keras的后端程序，实现卷积神经网络。在构建卷积神经网络时，启用GPU支持可以大大的提升计算效率，所以还需安装CUDA。使用AWS p2.xlarge（或p3.2xlarge）上的针对深度学习的服务器实例，可以降低成本，节约时间，提高开发效率。

对数据进行预处理，并从数据集中提取特征。区分训练集和验证集，然后进行模型训练，根据结果进行再优化，调整参数，直到取得满意的结果。根据项目要求，必须达到Kaggle Public Leaderboard前10%，即小于0.04141。

## 基准模型

在使用Keras深度学习库的官网上看到，已经有人给我们提供了一些可使用的模型，并且标注了模型的准确率，利用现有的模型以及预训练出的权重，就能很好的表示猫和狗的特征。综合考虑，基准模型选择的是InceptionV3、ResNet50和Xception这三个模型组合。结构图如下：



## 评估指标

训练一个可以有效运行的神经网络，需要使用损失函数。对于二分类问题，常使用交叉熵作为损失函数对模型的参数求梯度进行更新：

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

其中n是样本数量，y是标签，y=1表示狗，y=0表示猫， $\hat{y}$ 表示样本为狗的概率。

损失函数的值越小，说明模型对于数据集的分类效果越好。

## 设计大纲

1、从模型结构示意图中可以看出，预先需要准备的是输入的图片数据。对初始数据进行预处理是接触项目必须要做的事情，发现数据的特点，才能找到分析数据的切入点。发现图片名称是有序排列的，这对于读取文件带来了便利。根据培神项目指引，可以用符号链接的方式，对数据文件重新归类。

2、由于是使用AWS的云服务器，它按时间计费的方式，如果长期在云端调试代码，执行测试的话，预计经济成本会大幅提高，所以将特征向量导出后训练的方式，不失为一种节约成本的好办法。所以下一步就是将特征向量导出为文件，然后下载到本地进行编码开发。在提取特征时，可以参考Keras官方提供的[应用代码实例](#)。

3、将下载到本地的特征向量导入后，开始构建模型。通常情况下使用训练集的20%作为验证集，开始训练数据。然后使用CAM可视化模型，观察卷积神经网络学习的过程，得知学习到的哪些特征点是做出最后判断的依据。

4、最后对模型进行Fine-tune，获得满意的结果后，在测试集上进行测试。将每一步的分析过程整理成文档，撰写毕业报告，补充必要的参考引用等。

参考引用：

- [1] Y. Bengio, P. Simard, and P. Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2):157–166, 1994.
- [2] C. M. Bishop. *Neural networks for pattern recognition*. Oxford university press, 1995.
- [3] W. L. Briggs, S. F. McCormick, et al. *A Multigrid Tutorial*. Siam, 2000.
- [4] K. Chatfield, V. Lempitsky, A. Vedaldi, and A. Zisserman. The devil is in the details: an evaluation of recent feature encoding methods. In *BMVC*, 2011.
- [5] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The Pascal Visual Object Classes (VOC) Challenge. *IJCV*, pages 303–338, 2010.
- [6] S. Gidaris and N. Komodakis. Object detection via a multi-region & semantic segmentation-aware cnn model. In *ICCV*, 2015.
- [7] R. Girshick. Fast R-CNN. In *ICCV*, 2015.
- [8] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *CVPR*, 2014.
- [9] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In *AISTATS*, 2010.
- [10] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio. Maxout networks. *arXiv:1302.4389*, 2013.
- [11] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu. Deeply-supervised nets. *arXiv preprint arXiv:1409.5185*, 2014.
- [12] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.
- [13] Y. Movshovitz-Attias, Q. Yu, M. C. Stumpe, V. Shet, S. Arnoud, and L. Yatziv. Ontological supervision for finegrained classification of street view storefronts. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1693–1702, 2015.
- [14] R. Pascanu, T. Mikolov, and Y. Bengio. On the difficulty of training recurrent neural networks. *arXiv preprint arXiv:1211.5063*, 2012.
- [15] D. C. Psichogios and L. H. Ungar. Svd-net: an algorithm that automatically selects network structure. *IEEE transactions on neural networks/a publication of the IEEE Neural Networks Council*, 5(3):513–515, 1993.
- [16] L. Sifre and S. Mallat. Rotation, scaling and deformation invariant scattering for texture discrimination. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, Portland, OR, USA, June 23-28, 2013, pages 1233–1240, 2013.
- [17] N. Silberman and S. Guadarrama. Tf-slim, 2016.
- [18] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [19] C. Szegedy, S. Ioffe, and V. Vanhoucke. Inception-v4, inception-resnet and the impact of residual connections on learning. *arXiv preprint arXiv:1602.07261*, 2016.

[20] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1–9, 2015.