“猫狗大战”项目开题报告

2018 年 4 月 25 日

1. 领域背景

“猫狗大战”项目的领域背景是深度学习中的计算机视觉。根据维基百科的定义，“计算机视觉”是一个交叉学科领域，该领域研究计算机如何对数字图像或者视频产生高层次的理解。它使得计算机模仿人类视觉系统并自动处理相关任务成为可能，例如对象识别，动作追踪，实景再现和图像复原等等。

计算机视觉的发展最早要追溯到 20 世纪 60 年代，那个时候世界上一些研究人工智能的大学就已经开展了相关领域的研究，目的是为了赋予智能机器人“看得见”的能力。后来在 20 世纪 70 年代进行的相关研究为这一领域奠定了基础，这些成果至今仍被沿用（包括边缘抽取，线段标注，多面体建模，动作估计和通过多个小结构的连接构成的对象表示等等）。20 世纪 80 年代则开展了严格的数学分析和量化研究。到 20 世纪 90 年代末，计算机视觉与计算机图形学的交叉研究开始增多。最近几年计算机视觉研究的主流是与机器学习技术（特别是深度学习）和各种复杂优化方法相结合的，基于特征的方法。

本项目要解决的问题就是计算机视觉领域中的“图像识别”问题，具体来说就是训练计算机识别图片中是猫还是狗。这一问题已经有

了很多成熟的解决方案。近几年的研究成果提出了很多以卷积神经网络为核心的图像识别解决方案，例如：

1. VGGNet
2. ResNet
3. Inception v3
4. Xception

我本人之所以选择这个项目，是因为自己对深度学习中计算机视觉这个细分领域很感兴趣，希望自己能够掌握这个领域的理论基础，培养工程实践能力，将来把图像识别技术应用于自动识别害虫，为农作物病虫害防治提供基于深度学习的解决方案。

2. 问题描述

“猫狗大战”要解决的问题是训练计算机识别数字图片中是猫还是狗。这对于人类来说并不是什么大问题，而对于计算机来说有不小的难度，因为它所看到的图片是由数字构成的点阵。目前该问题已经有了很多的解决方案，其中基于卷积神经网络的深度学习解决方案效果最好。该问题属于二分类问题，可以被量化，因为我们最终的目标是通过训练得到一个函数 y=f(X)，函数的输入是数字图像，输出是二分类的概率估计 p 和 q，既图像有多大的概率是猫（p）或狗（q），并满足 p + q = 1。可以通过对预测得到的分类结果和实际的分类结果之间的差异来评估性能，因此是可测量的。最后，该问题显然是可复现的，定义良好的问题，所以它一定能够被很好地

用各种方法解决。

3. 数据集

本项目的数据集来自 Kaggle，可以直接从网站上项目主页的“Data”标签下载，包括两个压缩文件，一个叫 train（训练集）一个叫 test（测试集）。解压之后的训练集包括 25000 张图片，都为 JPEG 格式，是按照“猫/狗.编号”来命名的，比如“cat.0.jpg”。图片包含猫和狗的各种姿态的照片，排在前面的 12500 张图片全部是猫，剩下的全部是狗。通过对图片的大致浏览，发现某些图片还包含有人类，如图 1 所示。有的图片则同时出现了狗和猫，但被标记为猫，如图 2 所示。甚至还发现了异常值，如图 3 所示，一张人类的照片被标记为了猫，类似这样的异常值肯定会对分类准确性造成一定影响。

图 1 人类抱着猫咪的图片

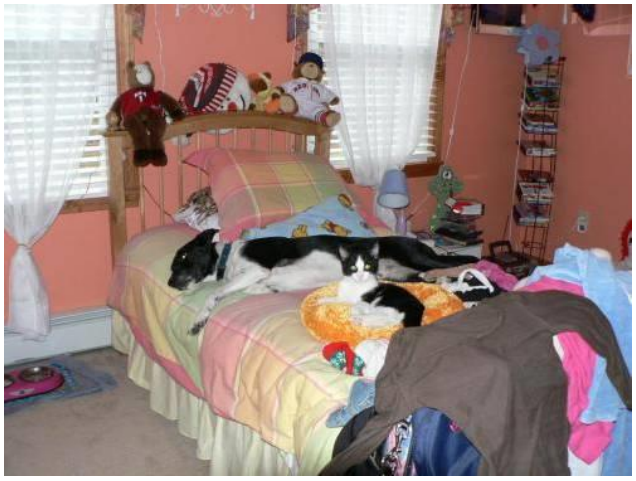

图 2 同时出现猫和狗的图片

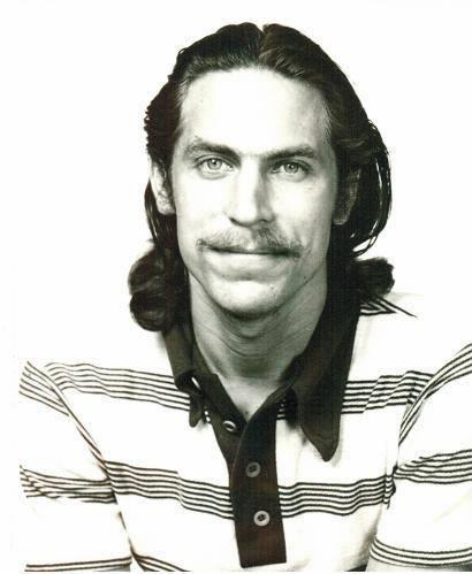

图 3 一张人类的图片被标记为猫

另外，图片的尺寸不一致，在输入到神经网络之前，应该调整图片的大小。具体地，需要按照神经网络输入层的要求对图片进行 resize 操作。训练集数据需要进一步分为训练集和验证集并打乱顺序，这可以提高模型在测试集上的泛化能力。测试集则包含 12500 张图片，以数字编号，猫狗的出现顺序已随机打乱。

4. 解决方案描述

拟采用“卷积神经网络”作为本项目的解决方案，这种特殊类型的

深度神经网络在计算机视觉领域有着广泛的应用并表现良好。具体地，将采取迁移学习的策略，使用在 ImageNet 上预训练过的四种卷积网络 VGGNet，ResNet，Inception v3 和 Xception 导出特征向量，然后再在特征向量的基础上构建模型进行分类。这种方法有效的原因是卷积网络是一种分层架构，前几层能识别图像中的一些简单的图案，例如边缘等等，这些往往是每个图像识别问题所共有的特征，因此可以复用并节约时间，只需要训练和调整最后几层即可。

5. 基准测试模型

按照毕业项目要求，我最后所得到的模型要能进入 Kaggle 排行榜前 10%，即在 Public Leaderboard 上 LogLoss 值低于 0.06127。

6. 评价指标

本项目要得到的最终结果是一个二分类问题，所以这里将用准确率结合二元交叉熵损失函数作为算法性能好坏的评估指标，将根据训练集和测试集的损失函数表现来评估算法性能。如果验证集损失还在下降，那么需要增加模型复杂度或者多训练几代；如果验证集损失上升，则出现过拟合，需要正则化或 Dropout 防止过拟合；如果验证集的损失出现震荡，则需要减小学习率；如果验证集的损失趋于稳定，则可以减少训练代数。

$$C = \frac{-1}{n}\sum_{x}[ylnp + (1-y)ln(1-p)]$$

- C：损失函数（Cost Function）；
- n：数据集数量；
- y：分类为狗（y=1）或猫（y=0）；
- p：模型预测分类为狗的概率。

7. 项目设计

7.1 数据清洗和可视化

从上面的分析可知，训练集中的某些图片是有问题的，明明是人却标记成了猫，这样的图片肯定会对模型性能造成影响。首先要找出所有这样的问题图片，这里将使用 OpenCV 中的 Haar feature-based cascade classifiers。OpenCV 提供了很多预训练的人脸检测模型，他们以 XML 文件形式保存，我将下载其中的 haarcascade_frontalface_alt.xml，并使用它写一个人脸识别器。然后检测每一张训练集中的图片，找出有人脸的图片，人工检查这些图片，将既没有猫又没有狗的图片删除掉。最后对清洗后的数据进行简单的可视化分析，统计一下猫的图片有多少张，狗的图片有多少张。

7.2 数据预处理

查阅 ImageDataGenerator 文档可知，它需要将不同种类的图片分到单独的子文件夹中。所以这里需要对训练集的数据进行预处理，

首先创建 data 文件夹，然后在 data 文件夹下分别创建 train 和 validation 子文件夹，再在这两个文件夹下都分别创建 dogs 和 cats 两个子文件夹。将图片的 95%作为训练集，5%作为测试集分配到各个文件夹中。为了节约空间可以采用符号链接的方式。

7.3 导出特征向量

使用 ImageNet 预训练的 VGGNet，ResNet，Inception v3 和 Xception，去掉各自的顶层（全连接层）然后进行训练，分别导出各自的特征向量并保存到磁盘上。

7.4 构建融合模型

采用迁移学习的思路，载入保存的特征向量做模型融合，添加全连接层然后构建模型，并可视化模型的基本架构。

7.5 模型训练和优化

最后，训练模型，画出损失函数图像并评估模型的性能。若达不到基准测试的要求则需要对模型进行调参。若出现过拟合，则可以调参最后的全连接层，增加正则化或使用 Dropout，或进行数据增强，或调参更多的卷积层，直到达到基准测试要求。

8. 参考文献

[1] Isma Hadji and Richard P. Wildes. What Do We Understand About Convolutional Networks? arXiv:1803.08834v1 [cs.CV] 23 Mar 2018

[2] Karen Simonyan & Andrew Zisserman. VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION. arXiv:1409.1556v6 [cs.CV] 10 Apr 2015

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition. arXiv:1512.03385v1 [cs.CV] 10 Dec 2015

[4] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Zbigniew Wojna. Rethinking the Inception Architecture for Computer Vision. arXiv:1512.00567v3 [cs.CV] 11 Dec 2015

[5] Francois Chollet. Xception: Deep Learning with Depthwise Separable Convolutions. arXiv:1610.02357v3 [cs.CV] 4 Apr 2017