# **基于深度学习的狗种类识别**

**马群飞**

*华北电力大学，北京，中国，102200*

*学号：1172227207*

*导师：张莹*

*邮箱：1091714856@qq.com*

摘要

近些年来，深度学习以其效率高，可塑性好的特点被用到很多领域，其中在图像识别领域使用卷积神经网络技术能起到了非常好的效果。本文主要阐述了一个使用MXNET框架，通过迁移预训练过的残差神经网络来提取特征，并用这些特征训练一个三层神经网络来实现狗种类预测任务，从而达到根据狗图片识别其对应种类的效果，该方法的数据集来源于kaggle比赛上的狗种类数据集，要分的种类为120种，最后的识别准确结果以提交到kaggle上的csv文件所达到的loss分值为基准，实验表明本文构造额网络模型在验证集上达到了85.3%的准确率和0.48的损失值，提交kaggle发现网络模型在测试集上能达到0.489的损失值。

**关键字：**卷积神经网络 残差神经网络 kaggle 迁移

一、引言

据统计，狗的种类已有180种左右，虽然各种狗各有其特点，但人类很难针对这么多种狗加以分别，因此使用图像识别技术来帮助人类识别狗所对应的种类有很好的现实价值。

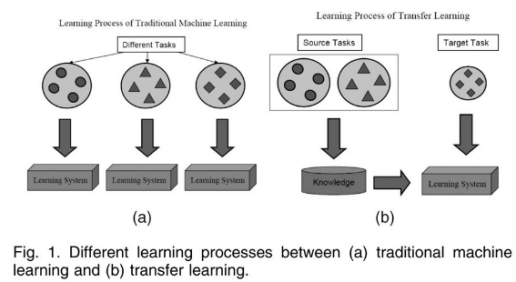
最近十年，深度学习在图像识别领域大放异彩，可以通过观察大型比赛ImageNet中发现，伴随着网络模型的不断改进，准确率在逐步提升，如2012年冠军用AlexNet模型错误率是16.4%，2014年冠军用VGGNET模型错误率是6.7%，带2015年的Resnet模型能达到3.5%，可以看出深度学习在图像识别领域所能起到的巨大效果，因此，对于狗种类识别问题我们可以尝试使用比较先进的残差网络模型来获取高的分类准确率。

1. 相关概念

### 2.1 迁移学习

迁移学习是利用之前学到的知识来帮助完成新环境下不同但相关的学习任务。迁移学习可以在当前任务缺乏高质量训练数据时，将之前任务中学到的知识应用到当前任务中。

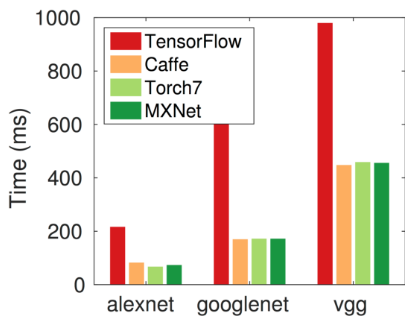
传统分类算法中，为了保障训练好的分类模型有较好的准确性和高可靠性，需要满足两个基本假设，一方面用于学习的训练样本和新的测试样本要满足独立同分布的条件，另一方面必须要有足够可利用的训练样本才能学习到一个好的分类模型，但是在实际应用中，这两点往往无法满足，比如股票数据就有很强的时效性，而且很所问题很难找到足够量的时效数据。因为人们能容易获取大量无标签的数据，而要获取有标签的数据则比较困难，尤其是一些高危险的任务，需要用虚拟环境去模拟。迁移学习可以解决那种少量有标签数据和大量无标签数据的问题，通过使用别的已经训练好的网络模型应用到当前少标签量的任务上，两者的比较如下图2.1，图中右侧表明迁移模型使用之前的knowledge知识库来帮助当前任务的学习。



**图2.1 迁移学习对比**

### 2.2 MXNET框架

MXnet是一个开源的深度学习计算平台，它是DMLC分布式机器学习通用工具包的重要部分。MXnet的优点是，轻量化、可移植性高、也可轻松扩展到多个GPU和多台机器，并且高效利用显存，同时速度比Caffe快，占用内存也比Caffe小，IO要求也比Caffe低。而且MXnet能够同时支持符号式编程和声明式编程，在命令式编程上MXNet提供张量运算，而声明式编程中MXNet支持符号表达式，用户可以自由的混合它们来快速实现自己的想法。例如我们可以用声明式编程来描述神经网络，并利用系统提供的自动求导来训练模型。另一方面，模型的迭代训练和更新模型法则中可能涉及大量的控制逻辑，因此我们可以用命令式编程来实现。同时我们用它来进行方便的调式和与主语言交互数据，因此Mxnet有很大的使用方面的便利性，并且MXnet有很快的网络训练速度，如下图2.2使用MXnet与其他框架在训练模型时间上进行比较，可以看出MXNet有很好的计算速度，而且MXnet 最近被亚马逊AWS 选为官方深度学习平台，可见MXnet有很大的影响力。

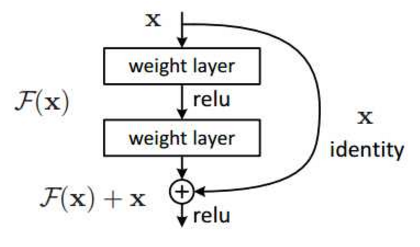


**图2.2 MXnet效果**

### 2.3 残差神经网络

残差神经网络是2015年由微软研究院四位华人提出的，在Imgnet上夺冠的神经网络，它是有152层深的神经网络，在2015年的比赛中取到了3.5%错误率，ResNet的结构可以极快的加速超深网络的训练，其模型主要是在传统网络中通过加入跳跃连接，构建如下图2.3的残差学习模块，通过这种参数控制跳跃的方式来克服网络越深训练难度越大的问题。

残差神经网络核心的思想是针对随着网络深度增加出现梯度消散的问题，会设计公式为的残差单元结构，这样有一部分x是不经过矩阵乘法和非线性变化的，可以通过控制保留系数C来控制有多少x需要进行非线性变化，哪些是直接通过的，这样就可以通过计算残差让网络学习出合适的参数值，当层数较深达到饱和时，就会令保留系数变大，从而基本相当于y=x的全等映射层，从而能解决梯度消散的问题。



**图2.3 残差学习单元**

三、算法说明

### 3.1观察数据集

为了能够便于更好的设计算法，我们先对给定的数据集使用图表进行展示和观察，以及查看训练集合测试集的数据量。可以通过给定的csv文件找到各种类图片的数量，可以观察到各品种给定的图片数量太少了，所以需要我们扩充数据集或者使用迁移的方法来制定后面的模型设计策略，而且通过随机展示图片的尺寸可观察到图片的尺寸是不一致的，所以需要对图片的尺寸进行规范化处理。

### 3.2数据集预处理

由于给定的数据集中测试图片和训练图片都是分别放在一个大文件中的，没有按种类分离，因此为了便于以后的操作，便于对数据集的利用，我们这里可以使用建立软链接的形式，把训练数据集根据原数据文件和csv文件将训练数据集分类进行重新组合在一起，这样的放置形式相当于创建一个快捷方式，对软链接文件的编辑不会影响到原文件，但是能够能通过软链接文件拿到和查看原文件。

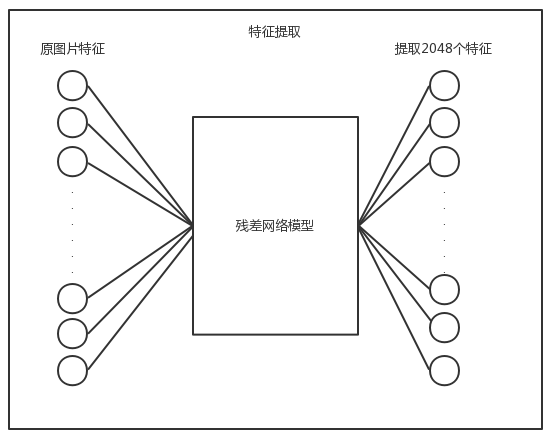
### 3.3特征提取

这部分是因为数据量太少，无法使用如此小的数据量训练一个较好的神经网络，因此我们可以使用迁移学习的策略，使用MXNET框架中自带的残差网络来对所有样本数据进行样本提取，这样之后进行网络训练时候就会使用这些特征数据进行网络训练。

为了提高网络的适应能力，我们会在往预训练模型喂养前先对图片进行预处理，包括对特征值的归一化，对图片设置成统一的规格，特征矩阵的偏转等操作。

对样本特征进行处理之后，我们就可以将这里特征放入到预训练过的残差网络来提取特征，如图3.1，提取的过程是将数据按批次在残差网络中从头到尾跑一遍，把最后的网络输出结果作为提取出的特征结果，这里会分别对训练数据和测试数据进行特征的提取。

进行特征提取后，为了便于以后的观察和对网络的训练，我们会将所有提取出来的特征保存到hdf5文件，来便于提取和使用。



**图3.1 特征提取过程**

### 3.4数据训练和测试

将样本特征进行提取后，我们就可以使用这些特征数据来训练一个适合狗分类任务的神经网络。

这里我们会设计一个三层的神经网络，最后输出层的网络深度是120层，因为我们要根据输出的数值来对样本进行分类，对网络的每层的深度可以根据其效果进行调整，优化器我们会使用SGD，激活函数选择通用的relu函数，最后的输出层会使用softmax+交叉熵函数形式来获得网络的损失值输出。

在进行网络的训练前，我们会将训练集进行分离，分出来20%的数据量作为验证集，剩余的80%的数据集作为喂养数据集，这样可以便于使用验证集检验网络的预测效果，并可通过绘制损失值曲线来观察网络是否过拟合，使用数据集训练一定迭代次数后，我们可以得到一个较好的网络。

实际训练时会将数据集按批次放入网络进行训练，训练好网络后，我们可以将测试集输入神经网络来获取每个测试集所对应的softmax输出值，并根据要提交的样例csv格式将所有测试结果放入进行信息保存和在kaggle官网中提交。

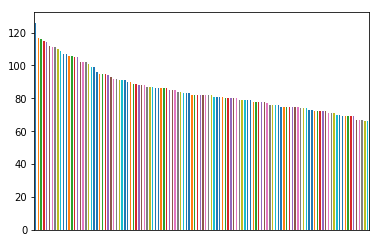
**表1：提交结果样式**

|  |  |  |  |
| --- | --- | --- | --- |
| Id | 种类输出值 | 种类输出值 | ......种类输出值 |
| 测试图id | Softmax值 | Softmax值 | .............. |
| 测试图id | Softmax值 | Softmax值 | .............. |
| ....... | Softmax值 | Softmax值 | .............. |

四、实验结果及分析

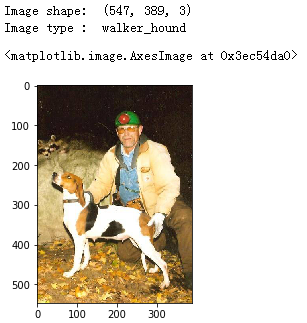
### 4.1 观察数据集

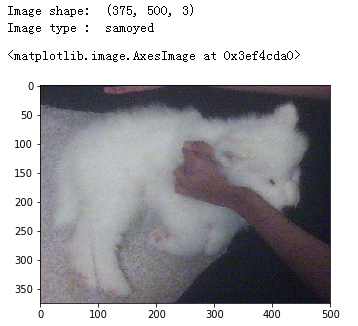
首先我们可以下载kaggle竞赛提供的数据集，因为已提供了有标签的训练数据，因此可根据数据集使用图像来展示狗种类数量分布图，分布图如图4.1所示，共有120种狗图片，大部分的种类图片数量范围为70-120，训练集的图片数总量为10222张图，测试集图片总量为10357张图，随机的图片展示尺寸有的是（240,288,3），有的是（375,500,3）可知给定的数据集图片是尺寸不一的。



**图4.1 狗种类的图片分布**

随机挑选的狗图片尺寸如下面两张图。

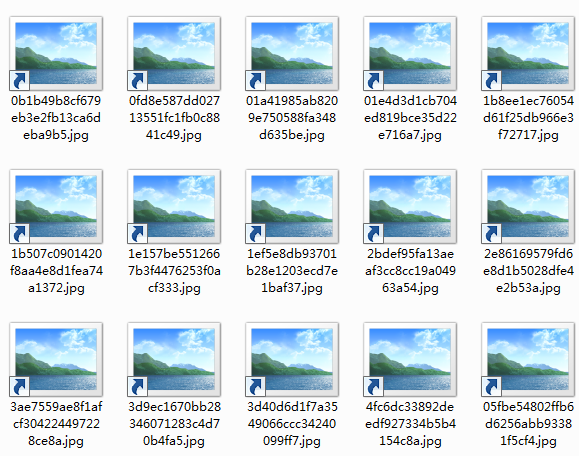


**图4.2 随机输出的图片尺寸展示**  

**图4.3 随机输出的图片尺寸展示**

### 4.2 建立软连接

为了便于以后的数据使用，这里会针对训练集和测试集建立数据软链接，进行数据的自定义重新组合，如这里会将所有训练数据根据图片归属的类别重新分布图片，如图4.4，将同种类的图片软链接放到一个文件夹中保存,这样后面进行使用时候就可以根据狗软链接文件所在的文件夹知道其对应的类别。



**图4.4 软链接展示**

### 4.3 特征提取

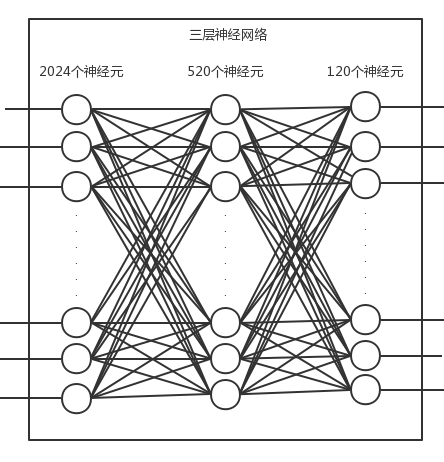
进行特征提取时，我们会先对样本进行预设置，对样本进行的处理有将特征值归一化，将图片尺寸调整到224\*224尺寸大小，进行样本调整后会将所有数据放入到预训练的模型中提取样本特征，按批次将数据放入到残差网络中进行模型的提取，我们指定的每批次样本数量为32个，针对10222个训练样本需要320次迭代，针对10357个测试样本需要324次迭代，进行特征提取的过程如图所示，每1W的样本使用网络提取特征使用了将近半小时，如图4.5为提取训练集特征进度过程。



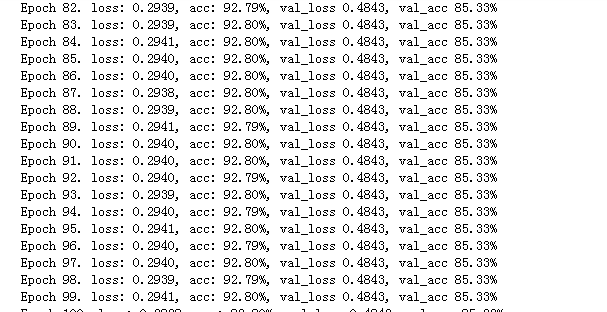
**图4.5 提取训练集特征**

### 4.4 训练模型

这里会搭建一个3层神经网络，输入层是2048个神经元，中间层有520个神经元，最后的输出有120个神经元，激活函数使用relu函数，最后损失函数计算用交叉熵损失函数，网络结构如图4.6。



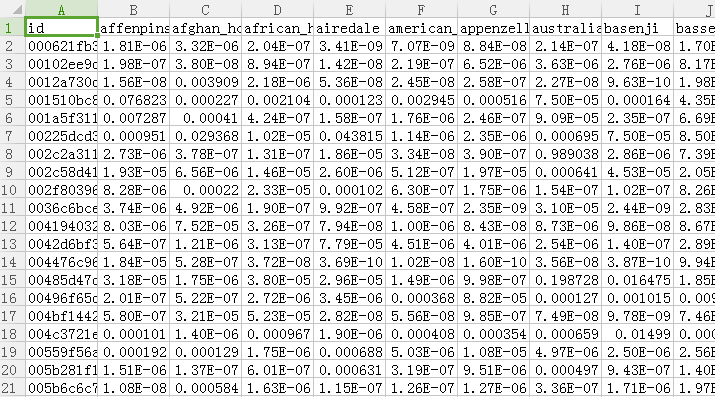
**图4.6 三层神经网络结构图**

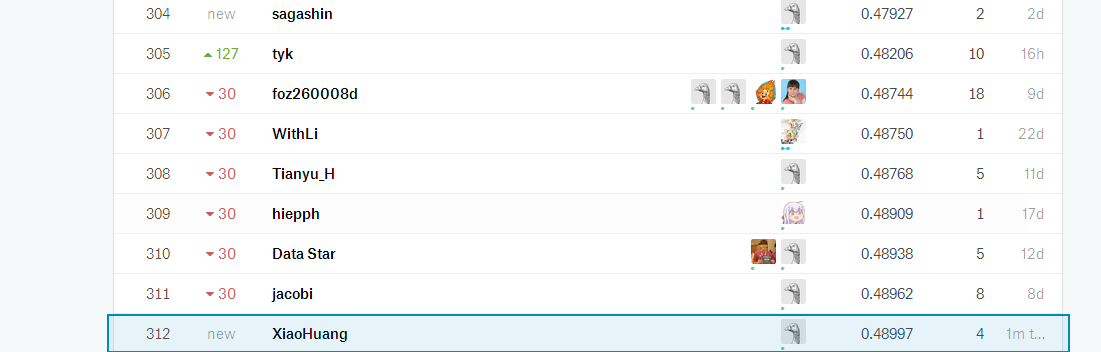
进行网络训练时会先将训练集分80%作为喂养数据集，20%作为验证集，初始学习率是0.001,使用SGD优化器对网络进行训练，指定每次训练时的批次数据大小为64个，执行100次数据迭代，并每个20次迭代将学习率缩小一半，这样更有利于找到局部最小值，并在每批次迭代中输出批次内喂养数据和验证数据样本对应的平均准确率和损失函数值，通过实验可获得如图4.7对应的网络训练输出结果，验证集准确率能起到一定的代表作用，可知网络能接近达到85.3%的预测准确率。

**图4.7网络训练输出**

### 4.5训练集预测

为了能进一步的验证训练出的网络模型的效果，我们可以使用测试集投入网络来获取输出，将使用测试样本获得的softmax输出拼接保存到csv文件中，如图4.8防止保存，提交到官网可发现，官网上对应损失值得评分为0.489，排名为312名，排在前45%，如图4.9所示，说明效果还是不错的。

**图4.8 csv提交样式**

 **图4.9 kaggle排名展示图**

五、结论

本文使用Kaggle比赛中的数据集，实现了一个基于迁移学习的神经网络模型，通过迁移过来的预训练残差网络模型来提取特征，并构造神经网络结构后用监督学习方法进行模型的训练，使得最后的模型适合于分类任务，后期通过对网络深度的调整，参数及学习率动态变化等方法使得模型的准确率逐步提升，最后通过实验验证可知，网络模型在验证集上能得到85.4%的准确率和0.48的损失函数值，在往kaggle提交的测试集上达到了0.489的损失函数值，从而可知本文实现的方法有85%的高概率能对狗图像正确分类。