

# 基于卷积神经网络实现猫狗识别

## 项目背景

图像识别，是利用计算机对图像进行处理、分析和理解，用来识别各种不同模式的目标和对象的技术。数字图像处理和识别的研究开始于 1965 年。数字图像与模拟图像相比具有存储、传输方便可压缩、传输过程中不易失真、处理方便等优势，这些为图像识别技术的发展提供了强大的动力。图像识别的发展经历了三个阶段：文字识别、数字图像处理与识别、物体识别。

卷积神经网络<sup>[1]</sup> (Convolutional Neural Network, CNN) 是一种前馈神经网络，它是一种专门用来处理具有类似网络结构的数据的神经网络。该神经网络使用卷积这种数学运算来替代一般神经网络中用到的矩阵乘法运算。卷积网络在诸多应用领域都表现优异。

猫狗识别是众多图像识别中的一个具体类别，从浅层的角度来看，猫狗识别的并不能带来实际的意义，但是从技术的深层原因来探讨，可以发现，处理此类图像识别的技术是通用的，例如可以应用在肿瘤图像阳阴性的判别、车牌号识别、人脸识别等很多方面，而高效率的图像识别可以极大的提高这些领域的精确度，造福社会，意义重大。

近年来卷积神经网络的发展，给大型图像处理带来了坚实的理论基础<sup>[2]</sup>，该技术对于大型图像的处理有着非常出色的表现。卷积神经网络可以避免对图像的复杂的前期预处理，可以直接输入原始图像，所以在图像处理方面得到了广泛的应用。另一方面，多家科技公司提出了自家在神经网络方面的基础框架，如 Tensorflow、caffe、mxnet 等，使得普通工程师就能简单地应用神经网络来解决现实中复杂的难题，并且，越来越多类似 Imagenet<sup>[3]</sup> 这种视觉识别方面的比赛，将图像识别的理论和实践不断的推向新高度，带来更多的可行性方案和理论基础。

## 问题描述

该项目需要根据数据集中的数据，构建卷积神经网络模型，输入一张彩色图片，识别出该图片中是猫还是狗，用数据集中的图片输入模型，使模型不断训练，然后调节超参数，不断优化模型，直至在测试中模型的正确率达到一定程度。解决这个问题方法有多种，1、可以使用主成分分析 (PCA) 对图片进行降维，然后使用支持向量机 (SVM) 进行模型的训练；2、使用普通的神经网络对图片进行处理、分析和识别；3、使用卷积神经网络，利用其特有的卷积层和池化层来对图片进行处理、分析和识别。

## 数据或输入

数据集的来源是 Kaggle 网站中的挑战比赛 “Dogs vs. Cats Redux: Kernels Edition” 的数据，数据集将通过一个写好的图片输入函数来输入，该函数使用 python 语言编写，只需要在模型中调用该函数便可以实现数据集的输入。该数据集中拥有 25000 张猫和狗的图片，50%为猫，50%为狗，其中的每一张图片都拥

有一个标签来指明图中动物的种类。另外还有包括 12500 张图片的测试集，测试集中的图片没有标签，需要训练好的模型根据图片预测其种类。这是一个庞大的数据集，其数据量足够用来对模型进行训练。再将图片输入模型之前，将会先对数据集进行预处理，其中包括对图像的缩放，数据集的分割，将其分为多个 batch<sup>[4]</sup>，对图片的标签进行独热编码。之后，将图像输入到卷积神经网络中进行训练，并将判定结果和标签进行对比，判断其是否分类正确。该数据集非常大，想要一次完全训练需要消耗大量的时间和计算机性能，因此需要将庞大的数据集进行分割，并在前期调参过程中使用单个 batch 来训练，当模型的准确率较高时，再使用多个 batch 进行训练，这样既节省时间，也避免消耗大量性能。训练集图片的标签是 cat 或是 dog，这非常不利于最终结果的比对分析，因此将标签进行独热编码是提高效率非常有效的方法。

## 解决方法

解决猫狗图像识别的较优方法是使用卷积神经网络对图像进行分析、检测、识别。通过对输入数据的预处理后，输入构建的卷积网络，该网络包括输入层、卷积层、最大池化层、扁平化层、全连接层以及输出层。通过大量分批次的输入数据的训练、调参、优化，最终达到训练效果。

## 评估标准

对于猫狗识别这个问题，和数据集角度分析，这是一个二分类问题，并且检测结果的错误率不会带来重大的事故，仅仅只是识别错误的影响，因此正确率就可以很好的衡量这个模型的性能，而无需用到准确率、召回率和 F 分数这些评估标准；评估模型一个简单、常用的指标是正确率，正确率是正确预测的数量除以数据的总数量而得到的，是评价一个分类模型好坏的重要指标；但是，正确率无法反映出模型误判的具体程度，显然一个误判程度更低的模型要由于误判程度较高的模型，因此选用 Logloss 来作为评估标准会更合适，它可以避免正确率的局限性。Logloss 的计算公式如下：

$$\text{Logloss} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{i,j} \log(p_{i,j})$$

其中 N 为样本数，M 为类别数，该项目属于二分类问题，所以公式中的 M 值为 2，根据这个公式就可以计算出 Logloss 的值作为评估标准。

## 基准模型

输入数据的类型(图像)，是一种已知的拓扑结构，在这种前提下，使用卷积神经网络是最好的选择。具有衰减学习率以及动量<sup>[5]</sup>的 SGD 是优化算法的一个合理的选择。因为数据集的数量没有达到千万级别，所以应该在一开始就包含一些温和的正则化。提前终止是常用的一种方案，另一个非常合理的选择是 Adam 算法<sup>[6]</sup>，Dropout<sup>[7]</sup>也是一个容易实现的，兼容性强并且出色的正则化项，可以有效地预防卷积网络过拟合。参数共享是一个可以提高效率的有效方法，该方法可以显著的减小模型所占用的内存。此外，该模型应该具备基本的输入层、卷积层、

最大池化层、扁平化层、全连接层和输出层。其中每一层之间的激活函数都应当选用修正线性单元(ReLU)。作为模型的基准阈值,应该在 Kaggle 排行榜的前 10%,也就是在 Public Leaderboard 上的 logloss 要低于 0.06127。

## 项目设计

项目基于 Google 公司的机器学习开源框架 Tensorflow<sup>[8]</sup>搭建卷积神经网络,实施过程分为 5 块: 1、对数据进行预处理; 2、构建模型,并输入数据进行模型的训练; 3、对模型的输出进行准确率的计算,参照可以调节的超参数,将他们进行可视化,对可视化的图表进行参数研究,分析出参数与正确率的内在联系,并就其中联系对超参数<sup>[9]</sup>进行调节; 4、不断优化模型,使其正确率达到预想水平; 5、在训练好的模型上利用测试集进行数据的测试。

在模型训练过程中需要反复的进行增量式的改动,如调节超参数或者改进算法。算法的改进,可以从多方面考虑,比如正则化算法的选择,有批标准化、Dropout、提前终止等。数据的分析主要体现在超参数调节和 Logloss 之间的内在联系,比如可以使用散布矩阵(scatter matrix)来可视化超参数和 Logloss 之间的相关性。

## 参考文献

- [1] R. Girshick et al. (Microsoft). Region-based convolutional networks for accurate object detection and segmentation , 2016
- [2] C. Szegedy et al. (Google). Going deeper with convolutions , 2015
- [3] O. Russakovsky et al. Imagenet large scale visual recognition challenge, 2015
- [4] S. Lofte and C. Szegedy(Google). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, 2015
- [5] I. Sutskever et al. (Hinton). On the importance of initialization and momentum in deep learning, 2013
- [6] D. Kingma and J. Ba. Adam: A method for stochastic optimization, 2014
- [7] N. Srivastava et al. (Hinton). Dropout: A simple way to prevent neural networks from overfitting, 2014
- [8] M. Abadi et al. (Google). TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems, 2016
- [9] J. Bergstra and Y. Bengio. Random search for hyper-parameter optimization, 2012