

“猫狗大战”开题报告

项目背景：

近年来，深度学习技术的提出与发展，让人工智能领域再度成为了人们的焦点，各个方向百花齐放，其中“计算机视觉”也取得了不错的进展。如何利用好深度学习技术，自动学习和识别图像的特征，让机器具有人类的“视觉”也成了亿万学者追求的目标。

问题描述：

“猫狗大战”项目要求根据给定的数据集（包含很多张不重复的猫和狗的图片）和对应的标签（猫或狗）来训练一个端到端的模型，使得模型能够自动识别出未知图片（测试集）是猫还是狗。此模型可以是深度学习中的常用的神经网络及其变更，而识别效果好坏可以利用测试集上分类准确率来衡量。

数据或输入：

由于“猫狗大战”项目拟设计成端到端的网络，因此整个系统的输入应当是图片，而输出为类别标签，无需手工干扰或手工提取特征。也正如此，模型能个自我学习到完整图片的显著特征和隐藏特征，这些特征对于猫狗识别具有很好的可区分性。而图片对应的标签则用于纠正模型的学习，告诉模型学的对不对，以此训练模型。

解决方法描述：

要得到一个具有猫狗分类能力的系统，我们要搭建一个深度学习模型，模型分输入部分，结构部分及输出部分。其中输入部分应当是整个数据集中的训练集部分，输入部分能够将图片转换为像素值传入到模型结构部分；而模型结构部分能够将像素映射到某些高维空间，从中找出数据的分布与规律，最终映射成特征向量，供输出部分做分类；模型的输出部分将上一部分的隐藏层特征进行激活（如二分类任务的 **sigmoid** 等），得到识别到猫或狗的概率，概率较大的即为最终结果。

评估标准：

由于“猫狗大战”项目是一个二分类任务，即“是猫”或“是狗”的任务。因此在训练阶段，可以使用二分类的损失函数（如 **sigmoid** 交叉熵）作为训练误差；对于测试阶段，我们可以使用测试集的分类准确率来衡量模型的好坏，另外，若样本分布不均，可以使用 **F1 score** 来衡量整体性能。

基准模型：

计算机视觉成为研究焦点不得不归功于卷积神经网络（CNN^[1]）的提出与发展。得益于 CNN 对图像强大的学习能力，本项目拟采用 CNN 作为基准模型，且早期的 CNN（如 AlexNet^[2]）也经常被作为该领域一系列任务的 Baseline 模型，此外，多数研究成果均为 CNN 的改进或结合。

项目设计：

完整的猫狗分类系统需要分为三个过程：

- 1.数据预处理：此部分主要做归一化处理，让 0-255 的像素值映射到 0-1 之间便于模型学习；
- 2.模型框架：搭建一个深度学习模型，输入预处理后的图像，传送到隐藏层，得到输出结果，根据预测结果与真实标签构造误差函数，反向调整整个模型的参数直到损失收敛。
- 3.测试模型：利用全新的未知的图片（测试集）来测试模型的好坏。

参考文献：

- [1]Yann LeCun, Gradient-Based Learning Applied to Document Recognition, 1998
- [2]Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105.