# 《人工智能基础》

# 实验报告

项目名称： 十二生肖分类识别

项目地址链接：

[https://aistudio.baidu.com/aistudio/projectdetail/4214486?contributionType=1](https://aistudio.baidu.com/aistudio/projectdetail/4214486?contributionType=1（项目需要设置为公开）)

***注：请使用最新版本；***

***最新版本中包含VGG16和RepVGG各自以Adam为优化器，以交叉熵为损失函数训练得到的模型，均为各自效果最好的模型，实验数据中还用到了其他历史版本中的模型，部分实验模型因空间限制并未得到充分保存，仅留存实验数据于报告中。***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 成员 | 组长 | 组员1 | 组员2 | 组员3 | 组员4 |
| 学号 | 20301034 | 20301033 | 20301036 | 20301037 | 20301038 |
| 姓名 | 陈增耀 | 陈杰 | 韩熔 | 贺思超 | 何毅 |

***\*注:组员名字按开头字的拼英字母顺序排列***

目录

[一. 实验数据集 3](#_Toc8266)

[二. 网络结构 4](#_Toc29709)

[三. 损失函数及学习率 7](#_Toc16911)

[四. 优化器 8](#_Toc11932)

[五. 实验结果 9](#_Toc3565)

[1.ResNet模型 9](#_Toc18174)

[2.LeNet模型 11](#_Toc1258)

[3.GoogleNet模型 13](#_Toc1340)

[4.三种模型比较 14](#_Toc13752)

[六、总结与收获 14](#_Toc12505)

1. **实验数据集**
   * + 1. **数据预处理**

【说明对数据进行了哪些预处理工作，比如：裁剪、改变通道、改变格式、人脸检测等。说明处理的目的以及处理的结果（没有可以不写）。如果自己加入了新的数据，需要介绍新加入的数据（给出数量、尺寸、特点、如何采集的），以及加入新数据的目的。】

VGG16网络的输入为224\*224的RGB图像，所以将所有图像格式化为RGB格式，将原图随机裁切为224\*224大小，为了增加训练集样本多样性，训练集中的数据将有50%的概率水平翻转。最后将图像归一化，能够提供更快的收敛和更稳固的训练。

* + - 1. **数据集说明**

【介绍最后所用数据集的规模、训练集中的样本数、测试集中的样本数、输入图像的尺寸。】

训练集数目: 7096, 验证集数目: 639, 测试集数目: 601，输入图像尺寸224\*224

* + - 1. **数据加载**

【描述如何将训练数据和测试数据加载进入PaddlePaddle框架中的。说明加载的具体细节，比如：使用了PaddlePaddle里面哪种数据加载器，参数是如何设置的，这样设置的理由是什么。】

继承了paddle.io.Dataset创建自定义数据集，将创建的数据集进行实例化，接着使用了paddle.io.DataLoader将数据加载进paddlepaddle框架中的。训练集、验证集和测试集分别加载对应的数据集实例。

batch\_size开始时设置为128，主要对应着V100的16GB显存，尽量将显存利用率最大化以及最大化矩阵乘法并行化效率，较大的batch\_size收敛更快，需要训练的次数少，准确率上升稳定，但训练后期由于batch\_size较大导致其下降方向已经基本不再变化，跑完一次epoch所需的迭代次数变小，对参数的修正也会显得更加缓慢。

Shuffle设置为True，Shuffle可以防止训练过程中的模型抖动，有利于模型的健壮性，Shuffle可以防止过拟合，并且使得模型学到更加正确的特征。

Dropout设置为True，在向前传播的时候，让某个神经元的激活值以一定概率停止工作，Dropout能够有效缓解模型的过拟合问题，从而使得训练更深更宽的网络成为可能。

1. **网络结构**

**【详细介绍所尝试的网络结构，给出网络结构图，说明：**

1）每个卷积层所采用的滤波器个数、卷积尺寸、滤波器深度、激活函数

2）池化层的个数、位置、池化窗口的尺寸及深度

3）全连接层的个数、每个全连接层的滤波器个数、尺寸、深度

4）输出层的神经元个数、激活函数】

我们组尝试了两种网络模型：VGG16和RepVGG。下面分别介绍这两种模型的结构与特点。

**1. VGG16**

（1）VGG16模型结构

VGG16模型在表格中有详细介绍，结构示意图**如图1所示**。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 卷积层13个 | | | | | | |
| 位置 | 1 | 2 | 4 | 5 | 7 | 8 |
| 输入数据尺寸 | 224\*224\*3 | 224\*224\*64 | 112\*112\*64 | 112\*112\*128 | 56\*56\*128 | 56\*56\*256 |
| 滤波器个数 | 64 | 64 | 128 | 128 | 256 | 256 |
| 卷积尺寸 | 3\*3 | 3\*3 | 3\*3 | 3\*3 | 3\*3 | 3\*3 |
| 滤波器深度 | 3 | 64 | 64 | 128 | 128 | 256 |
| 步长 | 1 | 1 | 1 | 1 | 1 | 1 |
| Padding填充 | same | same | same | same | same | same |
| 激活函数 | Relu | Relu | Relu | Relu | Relu | Relu |
|  |  |  |  |  |  |  |
| 9 | 11 | 12 | 13 | 15 | 16 | 17 |
| 56\*56\*256 | 28\*28\*256 | 28\*28\*512 | 28\*28\*512 | 14\*14\*512 | 14\*14\*512 | 14\*14\*512 |
| 256 | 512 | 512 | 512 | 512 | 512 | 512 |
| 3\*3 | 3\*3 | 3\*3 | 3\*3 | 3\*3 | 3\*3 | 3\*3 |
| 256 | 256 | 512 | 512 | 512 | 512 | 512 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| same | same | same | same | same | same | same |
| Relu | Relu | Relu | Relu | Relu | Relu | Relu |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| 池化层5个 | | | | | |  |
| 位置 | 3 | 6 | 10 | 14 | 18 |  |
| 输入数据尺寸 | 224\*224\*64 | 112\*112\*128 | 56\*56\*256 | 28\*28\*512 | 14\*14\*512 |  |
| 滤波器尺寸 | 2\*2 | 2\*2 | 2\*2 | 2\*2 | 2\*2 |  |
| 池化方法 | max pooling | max pooling | max pooling | max pooling | max pooling |  |
|  |  |  |  |  |  |  |
| 全连接层3个 | | | | |  |  |
| 位置 | Flatten | 19 | 20 | 21 |  |  |
| 输入数据尺寸 | 7\*7\*512 | 25088 | 1\*1\*4096 | 1\*1\*4096 |  |  |
| 滤波器尺寸 | 将数据拉平成向量 | 1\*1 | 1\*1 | 1\*1 |  |  |
| 滤波器个数 |  | 4096 | 4096 | 12 |  |  |
| 滤波器深度 |  | 1 | 4096 | 4096 |  |  |
|  |  |  |  |  |  |  |
| 输出层 | |  |  |  |  |  |
| 神经元个数 | 12 |  |  |  |  |  |
| 激活函数 | softmax |  |  |  |  |  |

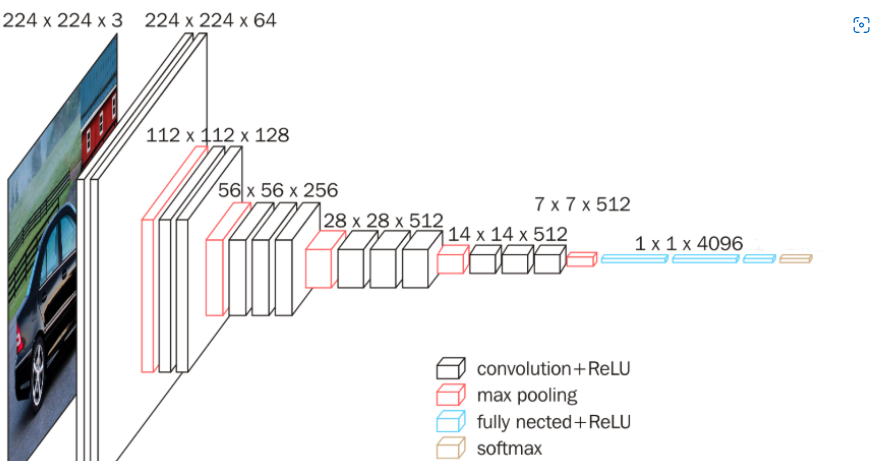
****

图1 VGG16模型结构

（2）VGG16模型的特点

VGG16具有巨大的参数数目，具有很高的拟合能力。VGG网络结构简洁，都是由小卷积核、小池化核、ReLU组合而成图2 残差模块的结构

VGG优点：小卷积核，小池化核，层数更深，卷积核堆叠的感受野，全连接转卷积；缺点：训练时间长，存储容量大。

**2. RepVGG**

（1）RepVGG模型结构

RepVGG模型包括44个卷积层，卷积尺寸3\*3，滤波器个数各有48、96、192个。激活层22个Relu，1个自适应池化层，22个RepVGG Block，最后有1个全连接层。**如图2所示。**

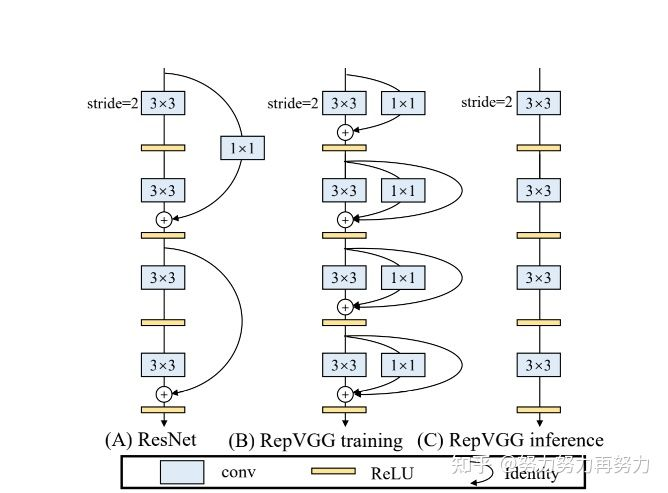


图2 RepVGG模型结构

（2）RepVGG模型的特点

RepVGG将训练推理网络结构进行独立设计，在训练时使用高精度的多分支网络学习权值，在推理时使用低延迟的单分支网络，然后通过结构重参数化将多分支网络的权值转移到单分支网络。RepVGG性能达到了SOTA，思路简单新颖。

RepVGG优点：简单即快速、内存使用经济、灵活，训练时的多分支结构；缺点：模型较大，结构没有VGG简单。

**三、损失函数及学习率**

1. **损失函数**

我们组尝试了2种损失函数，分别是：交叉熵损失函数、负对数似然损失函数 。

（1）交叉熵损失：

图形用户界面, 文本, 应用程序

描述已自动生成

交叉熵能够衡量同一个随机变量中的两个不同概率分布的差异程度，交叉熵作为损失函数在进行梯度下降计算的时候可以避免。出现梯度弥散，导致学习速率下降。

（2）负对数似然损失：

图形用户界面, 文本, 应用程序

描述已自动生成

Nll是交叉熵的底层实现，softamax+log+nllLoss=crossEntropyLoss，探究其与交叉熵损失的差距。

在反向传播过程中，更新每个权值，将其输出差值与输入激活相乘，以便得到该权值梯度。从权值中减去梯度的比值（百分比）。该比值称为学习率。

2. **学习率**

我们组尝试了6种学习率，分别是：0.000005、0.00001、0.00005、0.0001、0.0005、0.001

**四、****优化器**

【说明尝试了几种优化器，分别是什么优化器，各有什么原理、特点。介绍最终所采用的优化器及选择它的原因】

我们组尝试了两种优化器：Adam和Momentum优化器，分别介绍如下。

* + - 1. **Adam优化器**

原理：

文本

描述已自动生成

特点：

Adma吸收了Adagrad（自适应学习率的梯度下降算法）和动量梯度下降算法的优点，既能适应稀疏梯度（即自然语言和计算机视觉问题），又能缓解梯度震荡的问题

* + - 1. **Momentum优化器**

原理：

文本

描述已自动生成

特点：

Momentum梯度下降算法在与原有梯度下降算法的基础上，引入了动量的概念，使网络参数更新时的方向会受到前一次梯度方向的影响，换句话说，每次梯度更新都会带有前几次梯度方向的惯性，使梯度的变化更加平滑；Momentum梯度下降算法能够在一定程度上减小权重优化过程中的震荡问题。

* + - 1. **优化器的选择**

我们最终选择了更好的Adam优化器，最常用的Adam优化器，Adam优化器相较于Momentum优化器还有着收敛速度快、调参容易等优点，Momentum优化器增加了稳定性：不只是依赖于当前时刻的梯度，还考虑了以往时刻的梯度。在梯度方向有所改变的维度上更新速度变慢，可加快收敛并减小震荡；收敛速度更快：惯性加持，在梯度方向不变的维度上速度更快；还有一定摆脱局部最优的能力。

1. **实验结果**

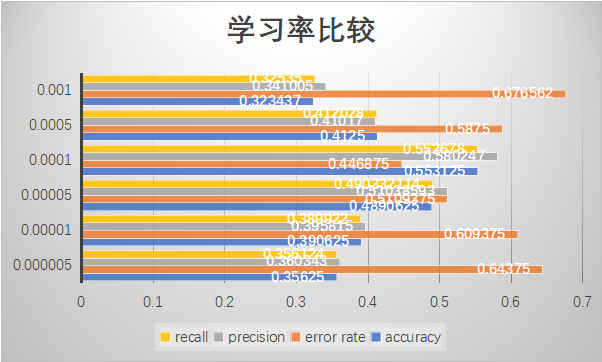
【用表格或图的形式，对比各种尝试（使用不同的网络模型、损失函数、优化器）的实验结果（recall，precision，accuracy，error-rate）；分析、解释实验过程中出现的一些现象。】

**1. 不同学习率实验结果比较**

使用VGG16网络，Adam优化器，交叉熵损失函数，batch\_size=128，epochs=20

表1 VGG16模型取不同学习率的实验结果

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 学习率 | 0.000005 | 0.00001 | 0.00005 | 0.0001 | 0.0005 | 0.001 |
| accuracy | 0.35625 | 0.390625 | 0.489063 | 0.553125 | 0.4125 | 0.323437 |
| error rate | 0.64375 | 0.609375 | 0.510938 | 0.446875 | 0.5875 | 0.676562 |
| precision | 0.36034357 | 0.395815 | 0.510386 | 0.580247 | 0.41017 | 0.341005 |
| recall | 0.35612422 | 0.389922 | 0.490332 | 0.552678 | 0.412028 | 0.32535 |



学习率为0.0001时，accuracy、precision、recall显著高于其他五组数据，表明此时整体的预测准确程度较高、对正样本结果中的预测准确程度较高、实际为正的样本中被预测为正样本的概率较高； error rate显著低于其他五组数据，表明此时整体的预测错误率较低。总体看来，学习率为0.0001时模型的预测能力最强。

学习率设置过小，收敛速度会非常慢，会增大找到最优值的时间，而且有可能进入局部极值点就收敛，找不到真正的最优解。所以随学习率的增大，预测能力会逐渐增强。

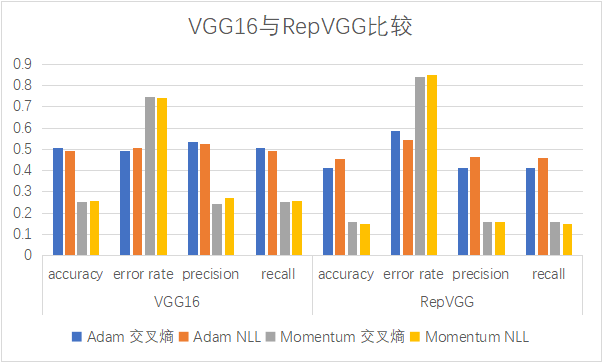
如果学习率太高，网络在训练过程中可能会跳过最小值点。更糟糕的是，高学习率可能会导致loss不断变大，这样就脱离了模型的学习目标。所以学习率超过0.0001时，错误率反而增大，预测准确率降低。

**2. 不同神经网络实验结果比较**

Epochs=20,学习率=0.0001

表2 不同神经网络实验结果比较

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 优化器 | | Adam | | Momentum | |
| 损失函数 | | 交叉熵 | NLL | 交叉熵 | NLL |
| VGG16 | accuracy | 0.50625 | 0.49375 | 0.253125 | 0.25625 |
| error rate | 0.49375 | 0.50625 | 0.746875 | 0.74375 |
| precision | 0.532701 | 0.525507 | 0.242026 | 0.271008 |
| recall | 0.505401 | 0.492008 | 0.252789 | 0.256476 |
| RepVGG | accuracy | 0.496527 | 0.45625 | 0.157812 | 0.148437 |
| error rate | 0.503472 | 0.54375 | 0.842187 | 0.851562 |
| precision | 0.5030656 | 0.465634 | 0.159138 | 0.155777 |
| recall | 0.4985417 | 0.458458 | 0.157989 | 0.149538 |



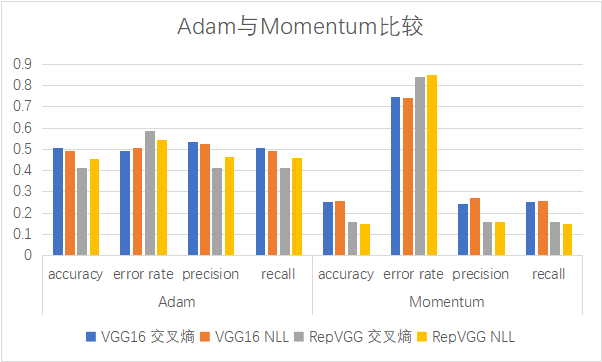
RepVGG相对VGG16来说在除在Adam优化器与NLL损失函数的情况下差距较小，其他情况下准确率都比VGG16小了10%，可能是VGG16在20个epoch内的提升更加明显，以及RepVGG在相同的batch\_size=128下无法发挥全部的实力，显存空余较多，对参数的修正更为缓慢。

1. **不同优化器实验结果比较**

Epochs=20,学习率=0.0001

表3 不同优化器实验结果比较

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 网络 | | VGG16 | | RepVGG | |
| 损失函数 | | 交叉熵 | NLL | 交叉熵 | NLL |
| Adam | accuracy | 0.50625 | 0.49375 | 0.4965277 | 0.45625 |
| error rate | 0.49375 | 0.50625 | 0.5034722 | 0.54375 |
| precision | 0.532701 | 0.525507 | 0.5030656 | 0.465634 |
| recall | 0.505401 | 0.492008 | 0.4985417 | 0.458458 |
| Momentum | accuracy | 0.253125 | 0.25625 | 0.157812 | 0.148437 |
| error rate | 0.746875 | 0.74375 | 0.842187 | 0.851562 |
| precision | 0.242026 | 0.271008 | 0.159138 | 0.155777 |
| recall | 0.252789 | 0.256476 | 0.157989 | 0.149538 |



在其他可控因素不变的情况下，

相比于Momentum，Adam在accuracy，precision，recall三个方面的数值明显更高，即使用Adam优化器的情况下，对于各方面的预测准确度和预测正样本概论较高；同时error rate相比于使用Momentum的情况，出现了大幅度降低。

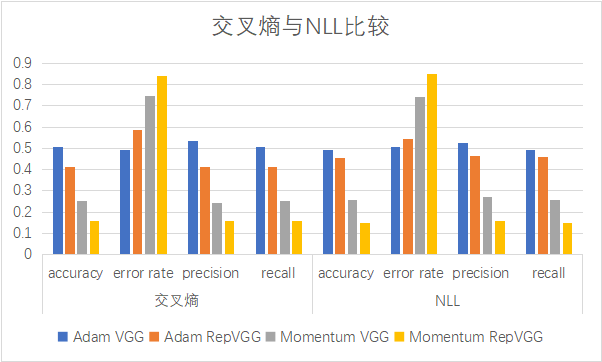
从理论上分析，Adam优化器不但吸收了Momentum的动量梯度下降的优点，能够缓解梯度震荡的问题，同时也能通过自适学习率应梯度下降算法，解决稀疏数据的问题。因此从理论上看，Adam优化器显然好于Momentum优化器，而实际实验数据也印证了这一点。因此采用Adam优化器

1. **不同损失函数实验结果比较**

Epochs=20,学习率=0.0001

表4 不同损失函数实验结果比较

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 优化器 | | Adam | | Momentum | |
| 网络 | | VGG | RepVGG | VGG | RepVGG |
| **交叉熵** | accuracy | 0.50625 | 0.466527 | 0.253125 | 0.157812 |
| error rate | 0.49375 | 0.503472 | 0.746875 | 0.842187 |
| precision | 0.532701 | 0.5030656 | 0.242026 | 0.159138 |
| recall | 0.505401 | 0.4985417 | 0.252789 | 0.157989 |
| **NLL** | accuracy | 0.49375 | 0.45625 | 0.25625 | 0.148437 |
| error rate | 0.50625 | 0.54375 | 0.74375 | 0.851562 |
| precision | 0.525507 | 0.465634 | 0.271008 | 0.155777 |
| recall | 0.492008 | 0.458458 | 0.256476 | 0.149538 |



交叉熵和NLL均为多分类问题常用的损失函数，经实验可以发现，在相同网络以及优化器下，交叉熵和NLL的效果没有太大差异，交叉熵总体上的效果优于NLL，说明softmax函数对提高预测正确率存在一定的作用但并不明显。

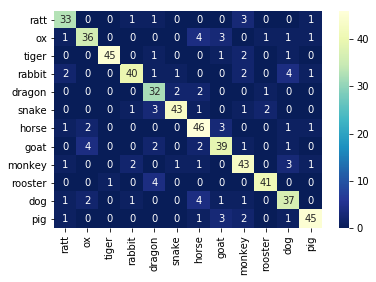
1. **最终最佳测试准确率**

accuracy: 0.833333

error rate: 0.166666

precision: 0.836218

recall: 0.833892

****

最终结果准确率为83.33%，由于使用VGG16网络和训练集较大、输出结果较多导致训练较慢，若需要进一步提升还需要大量时间来进行提升。从混淆矩阵中可以看出一些有意思的东西，老鼠和龙的识别成功率较低，可能训练集中的图像风格不一，同一类别中都有较大的差别。牛也有可能识别为羊，马有可能识别成牛，因为这三类动物相似度较高。识别成功率较高的是猪和鸡，可能因为形态和颜色与其他种类的动物具有较大的差别。

**六、总结与收获**

谈谈学习这门课的心得体会。每个小组成员分别写自己的心得（每人的心得不得少于100字），汇总于此。

学生1姓名：陈增耀

心得：刚开始得知我们在本科大二的时候就要开展“人工智能基础”这门课的时候，我的心情十分忐忑，害怕自己基础薄弱，不能学好这门课。然而后来我在paddlepaddle里学到了许多在线的很好的教学课程，也尝试做了些自己的小项目，在这次图像分类识别的实践中，更是得到了组员朋友们的许多支持，从他们身上我学到了很多，在不断尝试的过程中跌跌撞撞，最后也还是完成了这次实践。

在这次实践中，我们其实尝试了很多超出这次项目实践需求的内容，代码前前后后经过不同同学的手上产生了十多个版本，最后确定下来的这个版本使用了adam的优化器和交叉熵的损失函数，使用了VGG16的神经网络（短期内VGG16训练效果会更好）。我们还曾经试着自己搭建了一个属于我们自己的神经网络mynet，测试效果一般，但搭建这个网络的过程其乐无穷。

学生2姓名：贺思超

心得：最后将项目做出来确实没有那么容易，尽管最后的准确率只达到了83%，但也是尽了最大的努力。这次项目总体思路不难想，而是对所有的细节都要有深层的了解，在课上只是简单地了解了每个超参的含义，但真正实际操作起来需要慢慢去琢磨与分析。像batch\_size和learning rate这种都要根据具体的网络具体的优化器去选择合适的数值更是有所挑战，本来想着尝试RepVGG网络去达到一个更好的效果，结果对这个网络还不够熟悉所以batch\_size和learning rate并没有很好地去考量，而是简单地去与VGG16网络设置的一致，导致最终的效果并不是尽如人意。人工智能这个学科的知识面很广，而且每时每刻都有可能会有更新的知识加入进来，要想跟上人工智能学习的步伐，必须牢牢地打好基础并一直持续不断地给自己灌输新的知识。

学生3姓名：韩熔

心得：在上大学前对人工智能的认识主要来自科幻小说，对这个方向既向往，又好奇，想知道我们现在能做到什么。之后我了解到，图像识别、语义翻译、数据处理，生活中常用的一些工具与AI相关，AI没那么远。后面学到了神经网络，觉得它即复杂又精巧，是很美的构造。

令我印象最深的是，在课上学到AlphaGo Zero用三天时间自己左右互博490万棋局，以100:0的成绩完胜阿尔法狗，很是激动。人类数千年在围棋领域的耕耘，敌不过AlphaGo Zero的3天，在时间尺度上竟有如此巨大的差异。而如果在其他领域也能如此呢？我们的知识、艺术、文化在人工智能的冲击下会有怎样的改变？社会结构又会有怎样的改变？我不恐惧，我期待着那一天。

学生4姓名：何毅

心得：这次的人工智能实训时间虽然有些仓促，但总算是告一段落了。

从刚开始一头雾水，到处查资料，到敲定选题，以及中途修改选题，再到最终得以呈现，整个流程其实是比较充实的。

刚开始确实一直在碰钉子，不知道怎么处理数据，怎么用模型，后来就渐渐的熟悉了起来。

第一次成功的跑出识别结果的那一刻，心里突然就有了一种自豪感，有点不敢相信自己能写出来一个人工智能的程序

上课时讲的那些知识直到真正的用起来时才发现完全是另一种感受，这种能将理论变成实际的感觉很酷。

虽然过程也许不算尽如人意，但最终看到成果的那一刹那，感觉自己的付出终究是有回报的。

学生5姓名：陈杰

心得：在本次项目的完成过程中收获了很多未曾预料到的东西。

曾经在我眼中触不可及的人工智能，却在这段日子里真真切切地切实体会了，也感谢老师能够提供一个这样好的平台能让我有机会去有这样的体验。这门课中让我拓展了一个计算机的全新领域，而最后项目的完成也给了我运用这些专业知识的机会。同时更重要的是培养了自己的自信心，即使是看起来高大上的人工智能，只要肯付出努力，自己同样也能够完成的很棒。

在项目过程中，遇到了困难，也曾和同伴为了项目有过小争执，很感谢我的队友们在项目中的倾力付出，能够和大家从头一起完成一个完整的项目是我的荣幸。

**附：贡献度表格**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 成员 | 组长 | 组员1 | 组员2 | 组员3 | 组员4 |
| 学号 | 20301034 | 20301033 | 20301036 | 20301037 | 20301038 |
| 姓名 | 陈增耀 | 陈杰 | 韩熔 | 贺思超 | 何毅 |
| 贡献度 | 18% | 18% | 18% | 28% | 18% |

***\*注:组员名字按开头字的拼英字母顺序排列***