

北京工业大学

2022 - 2023 学年 第 二 学期

信息学部 计算机学院

机器学习基础 大作业报告

题 目: 基于 CNN 的扑克牌图像分类

班 级: _____

学生姓名: _____

学 号: _____

2023 年 6 月 1 日

摘 要

通过卷积神经网络进行图像分类任务随着研究的深入被越来越广泛地使用，模型本身也随着使用的需要和研究的基础上有了各种改进和发展。本实验研究的是基于改进的卷积神经网络对图像进行分类任务，取得了良好的分类结果和准确性。我们充分利用残差连接以及 1×1 卷积核的优点，提出基于残差连接和多通道卷积的图像分类网络。在输入层和隐藏层中，我们使用了 7 层卷积层、2 层最大值池化层、6 层批标准化层以及 ReLU 激活函数。在输出层中，根据我们的数据集，我们使用了 1 层全连接层和 Softmax 激活函数输出 53 类的分类概率。我们的创新点在于增加了残差连接，通过引入跨层的直接连接来解决梯度消失和网络退化问题；同时，我们使用尺寸为 1×1 的卷积核替换全连接层，从而加快收敛速度，最终在收敛效果和收敛速度上有了较大提升。最终基于本模型的图像分类准确率达到 90% 以上，各项指标达到理想水平。

关键词：卷积神经网络；图像分类；残差链接；多通道卷积

Abstract

The use of convolutional neural networks for image classification tasks has become increasingly widespread with the deepening of research, and the model itself has undergone various improvements and developments based on the needs of use and research. This experiment studied the classification task of images based on an improved convolutional neural network and achieved good classification results and accuracy. We fully utilized residual connections and 1×1 convolutional kernel and proposed an image classification network based on residual connection and multi-channel convolution. In the input layer and hidden layer, we used 7 convolutional layers, 2 maximum pooling layers, 6 batch normalize layers and ReLU activation function. In the output layer, according to our dataset, we used one fully connection layer with the Softmax activation function to output the classification probability of 53 categories. Our innovation lies in increasing residual connections, which solve the gradient vanishing and network degradation by introducing cross-layer connections. Meanwhile, We use 1×1 convolution kernel to replace the fully connection, so as to accelerate the convergence speed and finally greatly improve the convergence effect and convergence speed. Finally, the accuracy rate of image classification based on this model reaches more than 90%, and each index reaches the ideal level.

Keywords: convolutional neural network; image classification; residual connections; multi-channel convolutio

目录

摘 要.....	2
Abstract.....	3
第 1 章 绪论.....	5
1.研究背景.....	5
2.国内外研究现状.....	5
3.主要研究内容.....	5
第 2 章 研究方法基本原理.....	7
1.残差连接基本原理.....	7
2.多通道卷积基本原理.....	7
第 3 章 基于残差连接与多通道卷积的方案设计.....	9
1.整体网络结构.....	9
2.多通道卷积.....	10
3.残差连接.....	10
4.分类损失.....	11
5.评价指标.....	11
6.核心算法描述.....	12
第 4 章 实验与结论.....	14
1.实验数据与预处理.....	14
2.实验环境设置.....	15
2.1 开发环境.....	15
2.2 实验设置.....	15
2.3 评价指标.....	16
3.训练与验证结果.....	16
4.最终测试结果.....	16
第 5 章 总结.....	18
参考文献.....	19

第 1 章 绪论

1. 研究背景

卷积神经网络 (CNN) 是一种广泛应用于图像分类任务的深度学习模型。卷积神经网络是由纽约大学的 Yann Lecun 于 1998 年提出, 其本质是一个多层感知机, 成功的原因在于其所采用的局部连接和权值共享的方式。CNN 网络主要特点使用卷积层, 这其实是模拟了人的视觉神经, 单个神经元只能对某种特定的图像特征产生响应, 比如横向或者纵向的边缘, 本身是非常简单的, 但是这些简单的神经元构成一层, 在层数足够多后, 就可以获取足够丰富的特征。从机制上讲, 卷积神经网络与人的视觉神经极其相似。该优点在网络的输入是图像时表现的更为明显, 使得图像可以直接作为网络的输入, 避免了传统识别算法中复杂的特征提取和数据重建的过程, 在二维图像的处理过程中有很大的优势。LeNet 模型是在 1998 年提出的, 应用于支票或邮件编码上的手写数字的识别, 为后续 CNN 的发展奠定了基础; AlexNet 经典网络由 Alex Krizhevsky、Hinton 等人提出, 由于受到计算机性能的影响, 虽然 LeNet 在图像分类中取得了较好的成绩, 但是并没有引起很多的关注, 直到 2012 年将卷积神经网络的巨大优势带入人们视野。

2. 国内外研究现状

目前利用卷积神经网络实现图像分类的任务在大数据时代中得到了广泛应用, 同时也发展出创新的研究思路可以应用在不同场景提升分类的准确率和运算速度。he, kaiming 等将残差连接方法应用在卷积神经网络中^[3], 有效地缓解了卷积特征提取的梯度消失问题, 从而使卷积网络的深度大大加深, 学习能力大大增强。美国加州理工学院通过提出两个卷积神经网络模型^[1]: MSLNet 和 HiRISENet, 分别用于对 Mars 图像进行分类和区域分类, 并且使用了 transfer learning 技术来优化模型的性能和泛化能力 (Kiri L. Wagstaff et al); Haojie Ma et al 提出了一种改进的卷积神经网络 Inception V3 架构^[2]通过添加分离和组合层, 改进了 Inception V3 网络, 以更容易处理大型遥感图像。该方法通过在 Yushu 地震后的 0.5 米分辨率航空影像分类中测试, 取得了 90.07% 的测试准确率。Jielu Yan et al 开发了 Multi-Branch-CNN 方法^[4], 这是一种具有多个输入分支的卷积神经网络方法, 用于识别三种不同类型的输入。对于其实际应用, 需要能够识别与训练序列具有高或低相似性的新颖序列的预测模型。实验表明, Multi-Branch-CNN 方法比 13 种传统机器学习算法 (TML13) 表现更好。

3. 主要研究内容

本次研究我们将基于残差连接方法以及多通道卷积构建图像分类网络, 该网络的主要特点使用残差连接方法解决常规卷积网络出现的梯度消失问题, 使我们的网络能够加深, 学习到图像的更多特征。接着我们将使用 1×1 卷积核替换全连接层, 实现在通道维度降维后

输入全连接层，减少了全连接层的参数量，加快了收敛速度。

我们将使用 Kaggle 公开数据集 [Cards Image Dataset-Classification](#) 来训练我们的网络，该数据集包含 53 类扑克牌图像，图像共 8154 张，数据集已划分好训练集、验证集和测试集，验证集和测试集中每类包含 5 张图像，验证集共 265 张图像，测试集共 265 张图像，训练集共 7624 张图像。图像类别分布较为均衡，图像均为 RGB 图像，大小均为 224×224 。图像经过裁剪对齐，大部分图像没有背景，少量图像经过旋转、拉伸、倾斜等处理，出现的背景颜色较为单一。

第 2 章 研究方法基本原理

1. 残差连接基本原理

当神经网络层数比较高时，就很难被训练，容易出现梯度爆炸、梯度消失等问题。残差网络就是一种跳跃连接的网络，它通过将前面的激活值跳过中间的网络层而直接传到更后面的网络中，来避免梯度爆炸和梯度消失，训练更深的网络。

设残差块的输入为 X ，中间网络层的映射为 $H(\cdot)$ ，残差连接的权重为 W ，则残差块的输出表示为

$$F(X) = H(X) + W^T X \quad (2-1)$$

反向传播时，对其求导有

$$\frac{\partial F}{\partial X} = \frac{\partial H}{\partial X} + W^T \quad (2-2)$$

一般情况下，残差连接的权重 W 不全为 0。此时，若中间网络层的梯度几乎为 0，残差块的梯度值总是保留了残差连接的常数项梯度值，该层的梯度值不为 0，使得网络能够持续迭代收敛。

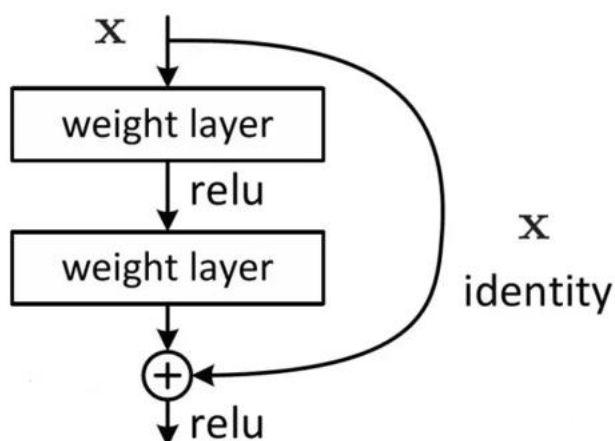


图 2-1 残差连接示意图

2. 多通道卷积基本原理

图像通常具有 RGB 三通道，因此需要一个多通道的卷积层接收图像的多通道信息并进行特征提取或融合。

多通道卷积核进行卷积操作的大致过程如下图所示。对卷积层输入一个多通道图像，卷积核的每一通道对图像的对应通道进行卷积操作，得到该卷积核对该图像进行卷积的多个结果矩阵。接着将这些结果矩阵对位相加后与其他卷积核卷积得到的结果矩阵拼接起来，得到输出图像。

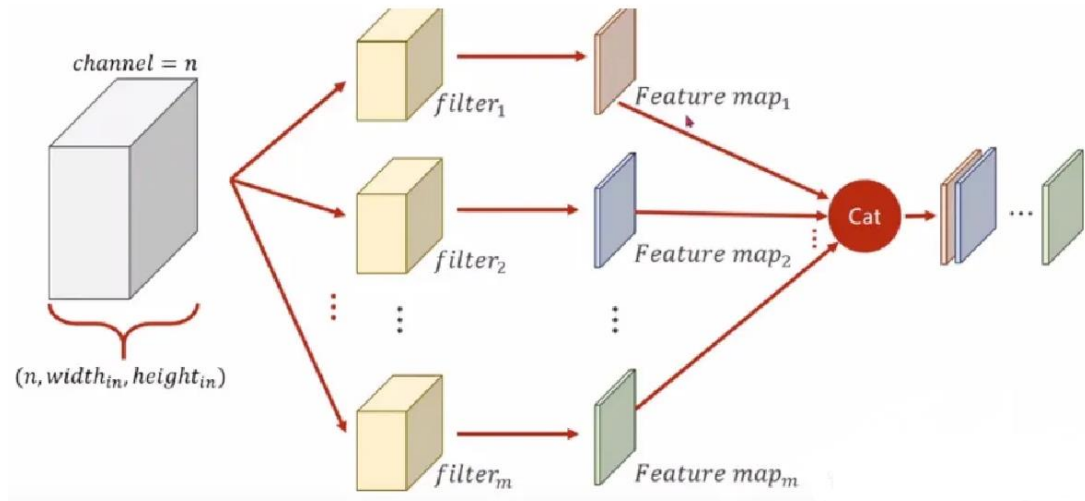


图 2-2 多通道卷积示意图

对于一个多通道图像，记它的宽、高、通道数为一个三元组 (n, w, h) ，该卷积层的输入通道数为 n' ，输出通道数为 m' ，卷积层的输出图像的宽、高、通道数记为 (m, w_1, h_1) ，该卷积层设置填充为 p ，移动步长为 s ，卷积核大小为 $k_1 \times k_2$ ，以上变量具有如下关系。

$$\begin{cases} n = n' \\ m = m' \\ w_1 = \frac{w + 2p - k_1}{s} + 1 \\ h_1 = \frac{w + 2p - k_2}{s} + 1 \end{cases} \quad (2-3)$$

第 3 章 基于残差连接与多通道卷积的方案设计

1. 整体网络结构

在本次图像分类任务中，我们基于以上原理，充分利用残差连接以及 1×1 卷积核的优点改进初始网络，提出基于残差连接和多通道卷积的图像分类网络，如下图所示（重要模块编号区分）。在输入层和隐藏层中，我们使用了 7 层卷积层、2 层窗口大小为 2×2 ，步长为 2 的最大值池化层、6 层批标准化层以及 ReLU 激活函数。在输出层中，我们使用了 1 层全连接层和 Softmax 激活函数输出 53 类的分类概率。

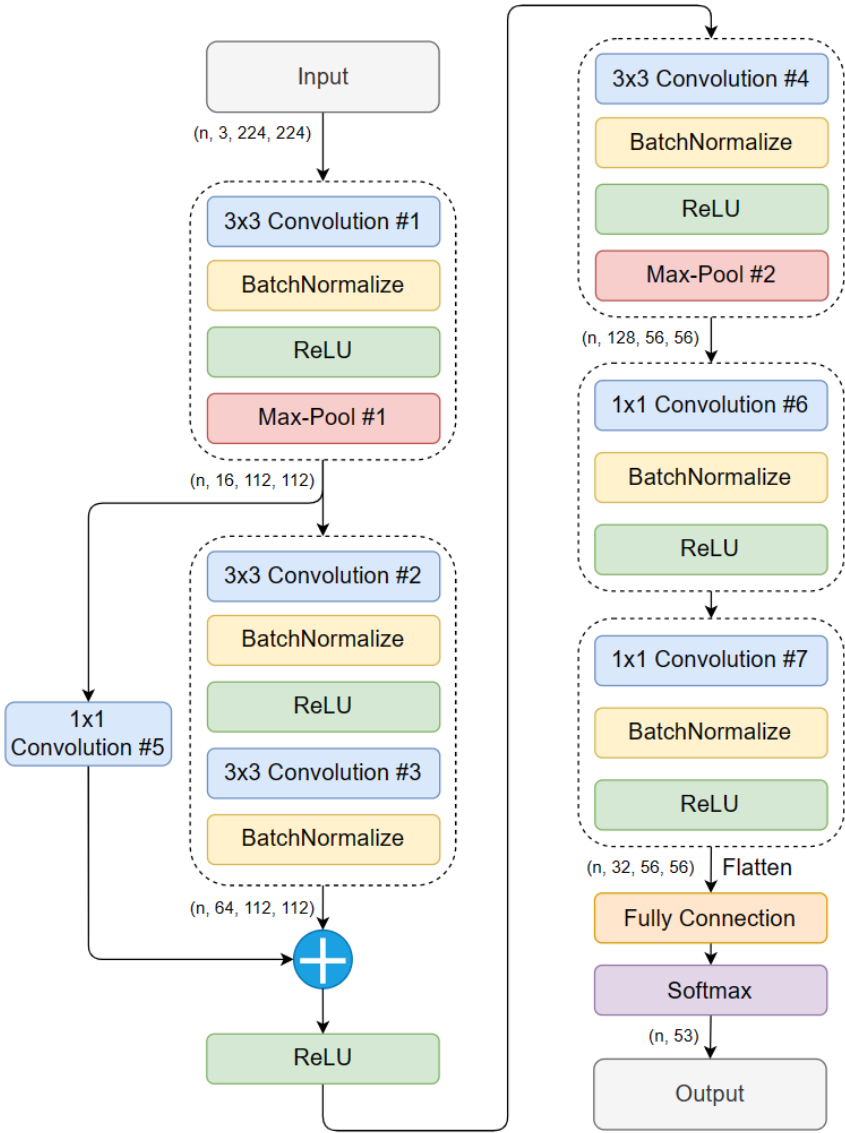


图 3-1 基于残差连接和多通道卷积的网络结构

输入张量的形状在网络中的变化如下表所示（省略不会导致张量形状变化的网络结构，形状中 n 表示当前训练的样本批量）。

表 3-1 张量形状变化表

网络结构	输入形状	输出形状
Convolution #1	(n, 3, 224, 224)	(n, 16, 224, 224)
Max-Pool #1	(n, 16, 224, 224)	(n, 16, 112, 112)
Convolution #2	(n, 16, 112, 112)	(n, 32, 112, 112)
Convolution #3	(n, 32, 112, 112)	(n, 64, 112, 112)
Convolution #4	(n, 64, 112, 112)	(n, 128, 112, 112)
Max-Pool #2	(n, 128, 112, 112)	(n, 128, 56, 56)
Convolution #5	(n, 16, 112, 112)	(n, 64, 112, 112)
Convolution #6	(n, 128, 56, 56)	(n, 64, 56, 56)
Convolution #7	(n, 64, 56, 56)	(n, 32, 56, 56)
Fully Connection	(n, 100,352)	(n, 53)

2. 多通道卷积

本网络中通过设置多通道卷积核的输入通道数和输出通道数，实现在通道维度上的特征提取和特征融合。各个卷积层的结构参数如下表所示。

在前半部分，我们主要使用 3×3 卷积核进行特征提取，不断提高输出通道数，最终得到通道维数为 128 的特征图。在输入到全连接层进行分类之前，我们使用了 1×1 卷积核在通道维度上进行特征融合，这样能够在保持特征图大小的情况下降低通道维数，使输入到全连接层进行分类的向量长度不会过长，加快收敛速度。此外，由于卷积核参数共享的性质，相比使用全连接层进行降维，使用 1×1 卷积核进行降维所需要的参数量也大大降低，进一步提高收敛速度。

表 3-2 卷积层结构参数表

Convolution	#1	#2	#3	#4	#5	#6	#7
核大小	3×3	3×3	3×3	3×3	1×1	1×1	1×1
步长	1	1	1	1	1	1	1
填充	1	1	1	1	0	0	0
输入通道	3	16	32	64	16	128	64
输出通道	16	32	64	128	64	64	32

3. 残差连接

本网络在 Convolution #2 和 Convolution #3 进行特征提取时容易出现梯度消失问题，因此在此 Max-Pool #1 层输出后使用残差连接将 Convolution #2 和 Convolution #3 短接，将两个输出相加。同时，为了使短路部分的张量形状与 Convolution #3 的输出形状相符，在短路部

分使用 1×1 卷积核进行升维。

使用残差连接可以保证在 Convolution #2 和 Convolution #3 进行特征提取后出现梯度消失的情况下，网络依然能够持续地迭代收敛，有效地解决了梯度消失问题。同时，在残差连接尾部将张量相加也可看做是将两种角度下提取的特征进行融合，提高了网络的学习能力。

4. 分类损失

本网络在最终输出分类概率后，使用多分类的交叉熵损失函数计算分类误差，多分类交叉熵损失函数表示如下。其中 C 为类别数量， y_i 为类别标签 one-hot 编码中的位，当样本属于第 i 类时， $y_i = 1$ ，否则 $y_i = 0$ ， p_i 为样本属于第 i 类的概率。

$$L_{CE} = - \sum_{i=0}^{C-1} y_i \log(p_i) \quad (3-1)$$

单独使用交叉熵损失作为损失函数时，可能会出现过拟合，因此我们增加 L2 正则化项来抑制过拟合。设参数向量为 $\omega = (\omega_0, \omega_1, \omega_2, \dots, \omega_n)$ ，L2 正则化项表示为

$$L_{L2} = \|\omega\|_2^2 = \sqrt{\sum_{i=0}^n \omega_i^2} \quad (3-2)$$

由以上公式，得到本分类模型的损失函数 L ，表示如下。其中 λ 为正则化参数。

$$L = L_{CE} + \lambda L_{L2} \quad (3-3)$$

5. 评价指标

在分类任务中，真实类别与分类结果的对应情况汇总表称为混淆矩阵（Confusion Matrix），设类别数量为 n ，混淆矩阵表示如下。其中的每个元素 a_{ij} 表示真实类别为第 i 类，分类结果为第 j 类的样本数量。

$$Confusion = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \quad (3-4)$$

对于多分类任务中的一个真实类别 C ，将分类结果为 C 的样本数称为真正例数（TP），分类结果不为 C 的样本数称为假负例（FN），分类结果为 C 但真实类别不为 C 的样本数称为假正例（FP），分类结果与真实类别都不为 C 的样本数称为真负例（TN）。

在混淆矩阵中，对于一个真实类别 c ，TP、FN、FP、TN 表示为

$$\begin{cases} TP = a_{cc} \\ FN = \sum_{j=0, j \neq c}^{n-1} a_{cj} \\ FP = \sum_{j=0, j \neq c}^{n-1} a_{jc} \\ TN = \sum_{i=0, i \neq c}^{n-1} \sum_{j=0, j \neq c}^{n-1} a_{ij} \end{cases} \quad (3-5)$$

由此产生了三个指标，分别为准确率(Accuracy)、精确率(Precision)和召回率(Recall)，有如下表示。在多分类任务中，以下指标表示为对于所有类别的相应指标的平均值。

$$\begin{cases} Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \\ Precision = \frac{TP}{TP + FP} \\ Recall = \frac{TP}{TP + FN} \end{cases} \quad (3-6)$$

F1 分数综合考虑了精确率和召回率的结果，表示为精确率和召回率的调和平均数，公式如下。对于多分类任务，F1 分数有两种类型，分别为 Macro-F1 分数和 Micro-F1 分数，Macro-F1 分数表示为每个类别的 F1 分数的平均值，Micro-F1 分数表示为分类的总精确率和总召回率的调和平均值。

$$Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3-7)$$

6. 核心算法描述

基于以上设计，编写网络的训练算法流程如下所示。

算法 3-1：基于残差连接与多通道卷积的图像分类网络训练

输入：扑克图像数据集 $D=\{\text{data}, \text{target}\}$ ，其中 **data** 为图像张量，**target** 为图像标签

输出：网络参数以及图像分类概率

for epoch = 1 to E do

 for batch = 1 to B do

 将训练集的每一批量输入到网络中，输出分类概率

 使用损失函数计算每一批量的损失值

 记录分类结果和损失值

 反向传播并使用优化算法更新网络参数

 end for

 计算本轮训练指标

for batch = 1 to B do

```
    将验证集的每一批输入到网络中，输出分类概率
    使用损失函数计算每一批量的损失值
    记录分类结果和损失值
end for
计算本轮验证指标
end for
```

该算法在每一轮先使用训练集训练网络，更新参数，之后使用验证集对当前网络进行验证，同时计算相应指标，使得结果记录能够跟踪表现网络的训练效果和泛化能力，从而针对性地改进网络。

第4章 实验与结论

1. 实验数据与预处理

本次实验使用的数据集已经划分好训练集、验证集和测试集，图像经过裁剪对齐，大小为 224×224 ，部分图像经过旋转、拉伸、倾斜等处理。部分样本图像如下图所示。

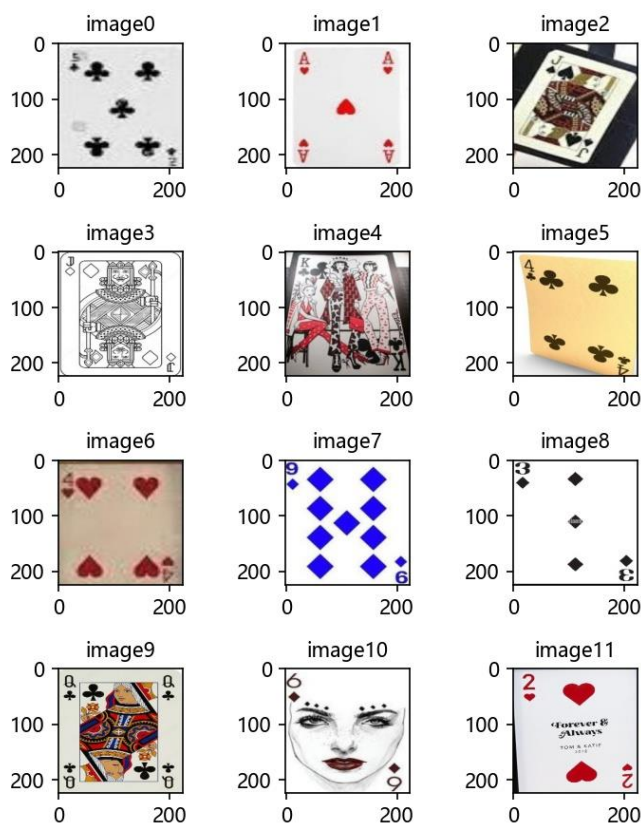


图 4-1 部分样本图像示例

数据集中包含 53 个类别，共 8154 张图像。其中训练集包含 7624 张图像，验证集包含 265 张图像，测试集包含 265 张，在验证集和测试集中，每个类别固定包含 5 张图像。所有图像的分类分布如下图所示。

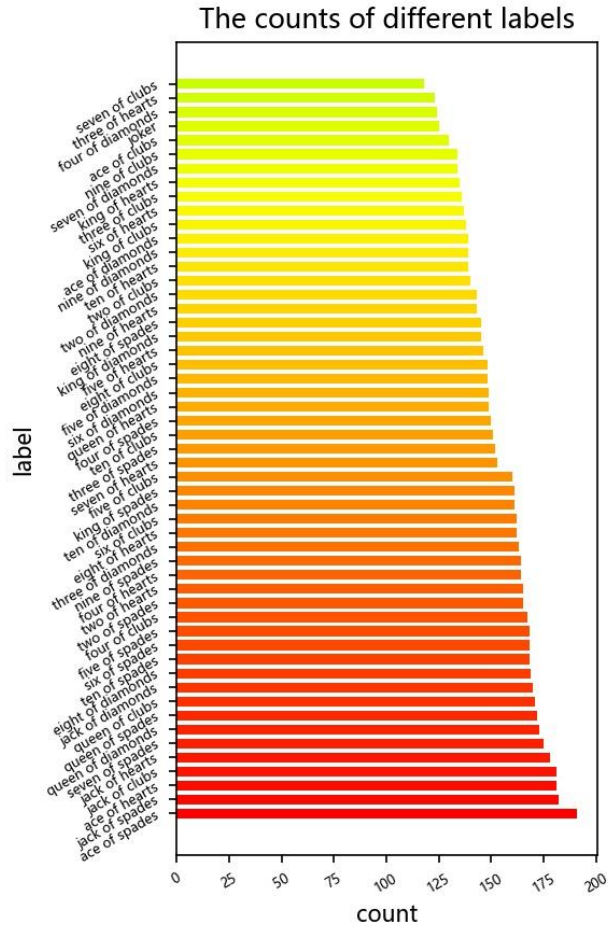


图 4-2 数据集类别分布图

图像为 RGB 三通道，原始张量形状为(224, 224, 3)，每个元素值为图像在对应通道上的色彩强度值，范围为 0 到 255，将图像张量进行归一化，将强度值压缩为 0 到 1 之间，之后进行转置，使张量形状为(3, 224, 224)。

2. 实验环境设置

2.1 开发环境

本实验使用 AutoDL 平台上的一台 GPU 型号为 RTX 3090，具有 24G 显存的主机，通过 SSH 连接本地 PyCharm 开发环境进行开发。深度学习模型基于开源框架 PyTorch 开发，同时集成 TorchMetrics 指标计算库、Matplotlib 可视化库等进行实验过程处理。

2.2 实验设置

本实验在模型训练时使用 Adam 优化器进行最小化损失，总训练轮次为 120 次，训练批量为 64，学习率为 1×10^{-4} ，正则化参数 λ 为 8.4×10^{-3} 。

2.3 评价指标

本实验在模型训练验证阶段使用准确率和 Macro-F1 分数两个指标评估模型的分类效果，绘制模型随着训练轮次增加的准确率曲线和 F1 分数曲线，同时绘制模型的损失曲线观察模型的收敛情况。

在模型测试阶段，我们使用准确率、精确率、召回率和 Macro-F1 分数四个指标评估模型的总体最终性能，同时绘制了混淆矩阵，观察模型在具体类别上的分类情况。

3. 训练与验证结果

模型在训练验证阶段的损失曲线、准确率曲线和 F1 分数曲线如下图所示。

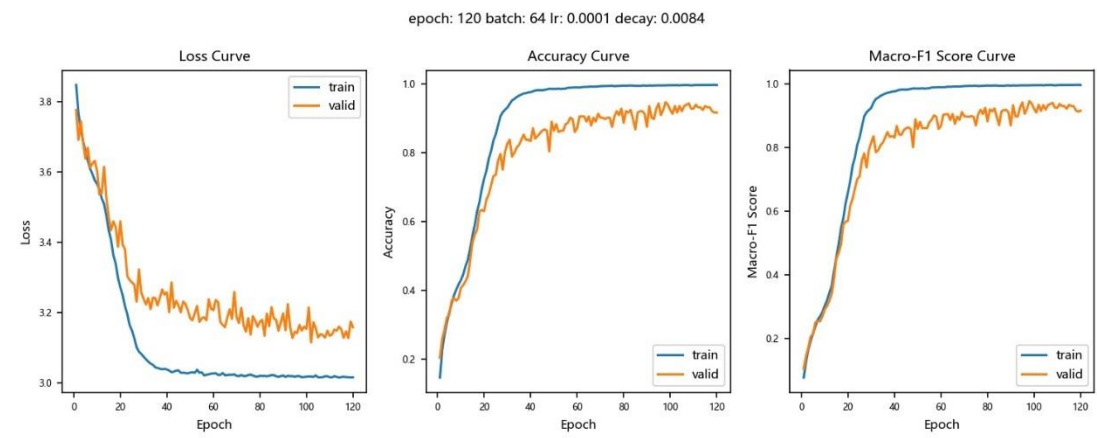


图 4-3 模型训练验证结果曲线图

我们取最后 10 轮的准确率和 F1 分数平均值作为模型在训练验证阶段的最终性能，如下表所示。模型的收敛速度和分类指标等均达到理想水平，可以进行最终性能测试。

表 4-1 训练验证指标表

	准确率	Macro-F1 分数
训练集	99.7%	92.8%
验证集	0.997	0.926

4. 最终测试结果

模型在测试阶段的各项指标如表 4-2 所示，混淆矩阵如图 4-4 所示。模型的各项指标达到理想水平，混淆矩阵中样本数量集中在主对角线上，说明绝大部分类别，模型分类结果与真实类别相同，模型分类效果达到理想水平。

表 4-2 最终性能指标表

准确率	精确率	召回率	Macro-F1 分数
88.3%	90.6%	88.3%	0.881

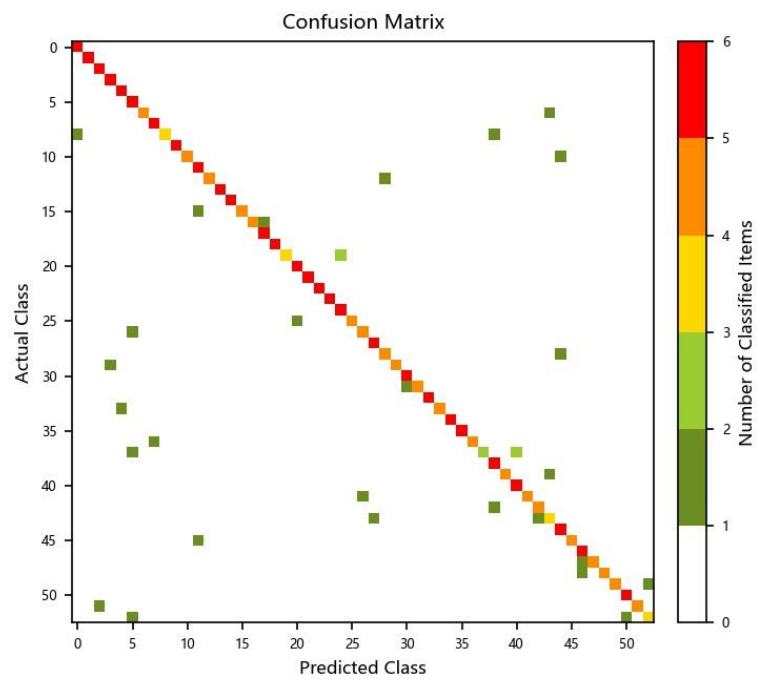


图 4-4 最终测试混淆矩阵

第5章 总结

在本次机器学习大作业中，我首先构建了常规的三层卷积网络，在数据集上表现不佳。之后通过查阅文献资料，使用了残差连接和 1×1 卷积核来解决梯度消失和收敛速度过慢的问题，最后通过调整参数得到了表现较好的模型。在这个过程中，我不仅对卷积神经网络本身有了更深的认识，而且对于模型训练的多种优化方法、思想和数据集的各种处理方法也有了广泛的了解和尝试，对机器学习这个方向有了更深的理解。

参考文献

- [1] Kiri L W, You L, Alice S, Kevin G, Thamme G, Jordan P, et al. Deep Mars: CNN Classification of Mars Imagery for the PDS Imaging Atlas.[C], AAAI Conference on Artificial Intelligence, 2018: 7867-7872.
- [2] Haojie M, Yalan L, Yuhuan R, Dacheng W, Linjun Y, Jingxian Y, et al. Improved CNN Classification Method for Groups of Buildings Damaged by Earthquake, Based on High Resolution Remote Sensing Images.[J], Remote sensing, 2020, 12(2): 260.
- [3] kaiming H, xiangyu Z, shaoqing R, jian S, et al. Deep Residual Learning for Image Recognition[C], Computer Vision and Pattern Recognition, 2016, abs/1512.03385(1): 770-778.
- [4] Jielu Y, Bob Z, Mingliang Z, Hang F K, Shirley W I S, et al. Multi-Branch-CNN: Classification of ion channel interacting peptides using multi-branch convolutional neural network[J], Computers in Biology and Medicine, 2022, 147: 105717.