**编号：**

**电子科技大学**

**一年级新生课外创新实践项目**

**结题报告**

**选题名称：** 基于 MindSpore 的图像分类设计与开发

**项目负责人：** 张璟哲  **学号：** 2024300905011

**联系电话：** 15388447015 **电子邮箱：** 3127305493@qq.com

**院系及专业：** 格拉斯哥海南学院

**指导教师：** 邓建  **职称：** 讲师

**联系电话：**  **电子邮箱：**

**院系及专业：** 计算机科学与工程学院

**电子科技大学教务处**

**填表日期： 2025 年 9 月 18 日**

**一、项目团队成员**（包括项目负责人、按顺序）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 姓名 | 性别 | 所在院系 | 学号 | 联系电话 | 本人签字 |
| 张璟哲 | 男 | 格拉斯哥海南学院 | 2024300905011 | 15388447015 | descript |
| 曹城玮 | 男 | 格拉斯哥海南学院 | 2024300901016 | 13708032802 | Screenshot_20250918_104839 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

**二、指导教师意见**

|  |
| --- |
| 该组同学认真学习，具有探索和创新精神，完成了新生项目的任务要求，同意结题。  签 名：IMG_256  2025 年 9 月 18 日 |

**三、项目专家组意见**

|  |
| --- |
| 组长签名：（ 学部盖章 ）  年 月 日 |

**四、项目成果**

|  |  |  |
| --- | --- | --- |
| 序号 | 名称 | 说明 |
|  | 基于 MindSpore 的图像分类设计与开发 |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**五、项目研究结题报告**（字数应在3000字左右）

**（一）主要内容：**

1、课题背景

2、课题研究内容与方法

3、研究结果

4、创新点

5、结束语

6、参考文献

**（二）排版要求：**

题名（二号黑体，居中，不超20字）

学院名称 第一作者，第二作者（四号楷体-GB2312，居中）

指导教师姓名及职称（四号楷体-GB2312，居中）

摘 要（五号宋体）：字数一般在300字左右。摘要必须反映全文中心内容，一般包括研究目的、方法、主要观点及结论。写作时，应简写目的，写明采用的具体方法，详细写所得到的结果和结论，要突出反映文章的创新性。要求语言简明、扼要、准确、客观、逻辑性强。总之，摘要应写得内容充实，不要过分抽象或空洞无物，避免使用“对……具有……意义，价值”等评价性用语，避免使用“本文”、“笔者”等第一人称写法。定稿时要注意纠正语病，删减啰唆重复的语句和句子。（五号宋体）

关键词：词1；词2；词3（3-5个反应所研究的领域和关键特征的词，五号宋体）

下接正文（引言、导论性）或一级标题（引言、导论性）。论文Word文档页面设置为A4纸型，页边距各2，文档网格设置为46字×43行，行距16磅，正文用小四号宋体，其中阿拉伯数字、英文用Times New Roman体。论文要求主题明确、数据可靠、逻辑严密、文字精炼，遵守我国著作权法，注意保守国家机密。题名应恰当简明地反映文章的特定内容，要便于编制题录、索引和选定关键词，不宜使用非公知的缩略词、代号等。

标题一（四号黑体，居中，上下各空一行）

下接正文或二级标题。参考其他文献，包括引用原文或参考、综述、评论他人观点，要在文中加引注标记，采用顺序编码制，符号按出现的先后顺序为[1][2]……，用上角标，与文后所列参考文献序号一致。参考文献只列出已经公开出版且在文中加注的文献，著录格式另附。文中图、表应有自明性，且随文出现，须注明图名、表名，按顺序标明序号如表1、表2……、图1、图2……，图名、表名及内容、参考文献均为五号字。请在稿件首页地脚处给出作者简介信息。

1.标题序号（空两格，加黑）

下接正文。我刊论文标题层次采取如下方式：一级标题“一”，二级标题“1.”，三级标题“（1）”，请按此层次顺序设置标题。

参考文献著录原则和方法

[1]作者姓名，作者姓名. 参考文献题目[J]. 期刊或杂志等名称，年份，卷(期数)：文章起-止页码.

[2]刘凡丰. 美国研究型大学本科教育改革透视[J]. 高等教育研究，2003，5(1)：18-19.

没有卷的就直接写2003，（1）（本条为期刊杂志著录格式）

[3]谭丙煜. 怎样撰写科学论文[M]. 沈阳：辽宁人民出版社，1982.5-6.（本条为中文图书著录格式）

[4]作者姓名. 参考文献题目[D]. 南京：南京农业大学，2002.（本条为硕士、博士论文著录格式）

[5]作者姓名. 参考文献题目[N]. 人民日报，2005-06-12（第几版）.（本条为报纸著录格式）

[6]作者姓名. 电子文献题名[EB/OL]．电子文献的出处或可获地址，发表或更新日期.

[7]作者姓名. 参考文献题目[A].主编. 论文集名[C]. 出版地：出版单位，出版年. 起-止页码.（本条为论文集著录格式）

[8]外国作者姓名（作者姓名：姓在前，名在后，姓全拼大写，名缩写，姓与名之间隔半格，作者之间用逗号隔开。）.参考文献题目[M].译者（名字）译. 出版地：出版单位，出版年. 起-止页码.（本条为原著翻译中文的著录格式，多个译者可写为：\*\*\*，\*\*\*，\*\*\*，等译.）

基于MindSpore的图像分类模型优化与端侧部署实践

**学院名称** 格拉斯哥海南学院

**作者**：曹城玮，张璟哲

**指导教师姓名及职称：**邓建 ，讲师

摘要

本项目以图像分类为切入点，基于华为MindSpore深度学习框架，完成了基于MobileNetV2的猫狗分类模型训练与移动端部署实践。项目初期成功实现了基于MobileNetV2的猫狗分类模型训练与移动端部署，中期汇报后重点进行了模型架构改进的理论探索和代码修改尝试。通过深入研究ResNet残差网络和Vision Transformer架构的设计思想，探索了不同网络结构在图像分类任务中的应用特点。同时，项目深入研究了模型轻量化、量化压缩和端侧部署技术，为实际应用提供了技术方案。项目不仅完成了基础的图像分类任务，更重要的是建立了完整的"数据准备-模型训练-模型转换-端侧部署"的深度学习应用开发流程，为后续AI应用开发奠定了坚实基础。

关键词：MindSpore；图像分类；模型优化；端侧部署；深度学习

一、课题背景

1.1 研究现状与发展趋势

当前，图像分类技术作为计算机视觉领域的基础任务，正朝着两个重要方向发展：一是追求更高的识别精度，通过开发更深、更复杂的网络架构如ResNet、Vision Transformer等；二是追求更高的实用性，将强大的AI模型部署到资源受限的移动端或边缘设备上，实现实时、高效的智能应用。

在精度提升方面，从最初的LeNet、AlexNet到后来的VGG、ResNet，再到现在的Vision Transformer，网络架构不断演进。ResNet通过残差连接解决了深度网络的梯度消失问题，使得网络可以训练到100层以上；Vision Transformer则完全摒弃了卷积操作，使用自注意力机制处理图像，在ImageNet等大型数据集上取得了优异的性能。

在实用性方面，轻量级网络架构如MobileNet系列、ShuffleNet系列在保持可接受精度的同时，显著减少了模型参数量和计算复杂度，为AI在端侧设备上的普及奠定了基础。MobileNetV2通过深度可分离卷积和倒残差结构，在保持较高精度的同时大幅降低了计算量；ShuffleNet则通过分组卷积和通道重排技术，进一步提升了计算效率。

1.2 研究意义

本项目以图像分类为切入点，旨在帮助一年级新生了解深度学习的基本概念和一般流程，初步掌握不同神经网络架构的设计思想。通过使用华为自主研发的MindSpore全场景AI框架进行模型训练，并利用MindSpore Lite将模型部署到移动端，获得从理论学习到实际操作的完整体验。

对于初学者而言，本项目具有重要的教学价值：首先，通过完整的项目流程，学生可以理解深度学习从数据准备到模型部署的各个环节；其次，通过对比不同网络架构，学生可以深入理解各种设计思想的特点和适用场景；最后，通过端侧部署实践，学生可以了解AI技术在实际应用中的挑战和解决方案。

此外，本项目对于培养实践能力、创新思维，以及未来在人工智能领域进一步学习和探索具有重要的研究意义和实践价值。通过动手实践，学生不仅能够掌握理论知识，更能够培养解决实际问题的能力。

二、课题研究内容与方法

2.1 研究内容

本项目的研究内容主要包括以下几个方面：

2.1.1 深度学习基础与图像分类原理理解

深入学习卷积神经网络（CNN）的基本构成，包括卷积层、池化层、全连接层等核心组件。理解卷积操作的工作原理，包括卷积核的作用、步长和填充的影响等。学习激活函数的作用，如ReLU、Sigmoid、Softmax等，以及它们在网络中的位置和作用。

了解MobileNetV2、ResNet、Vision Transformer等网络的设计特点和关键技术。MobileNetV2采用深度可分离卷积和倒残差结构，在保持精度的同时大幅降低计算量；ResNet通过残差连接解决梯度消失问题，使得网络可以训练得更深；Vision Transformer将图像分割成patches，使用自注意力机制处理，完全摒弃了卷积操作。

2.1.2 MindSpore深度学习框架应用

学习使用MindSpore框架进行数据集的准备与加载，包括数据预处理、数据增强、数据加载器的配置等。掌握网络模型的构建方法，包括层的定义、网络结构的搭建、参数初始化等。学习模型训练的过程，包括损失函数的选择、优化器的配置、学习率调度等。

掌握模型保存与导出的方法，包括checkpoint的保存和加载、MindIR格式的导出等。学习模型评估的方法，包括准确率计算、混淆矩阵分析等。

2.1.3 模型架构改进的理论探索

通过研究ResNet、Transformer等不同架构，探索模型性能提升的方法和策略。深入学习ResNet的设计思想，理解残差连接的作用机制，研究如何通过残差连接解决深度网络的训练问题。学习Vision Transformer的原理，理解自注意力机制在图像处理中的应用，探索Transformer架构在计算机视觉中的优势。

研究不同网络架构的适用场景，分析它们在精度、速度、参数量等方面的权衡。探索如何根据具体任务选择合适的网络架构，以及如何通过架构改进来提升模型性能。

2.1.4 模型轻量化与端侧部署

学习使用MindSpore Lite工具将训练好的模型转换为适用于移动端的.ms格式，实现端侧的图像分类推理。研究模型压缩技术，包括网络剪枝、量化压缩、知识蒸馏等方法。探索端侧部署的优化策略，包括内存优化、计算优化、功耗优化等。

2.2 研究方法

2.2.1 数据准备与预处理

使用猫狗图片数据集，通过preprocessing\_dataset.py脚本对数据进行清洗。数据清洗包括去除损坏的图片文件、统一图片格式、调整图片尺寸等。将数据集按照9:1的比例划分为训练集和评估集，确保训练和评估数据的平衡。

数据预处理包括图像归一化、数据增强等。图像归一化将像素值从[0,255]范围映射到[0,1]范围，并进行标准化处理。数据增强包括随机裁剪、随机翻转、颜色抖动等，以增加数据的多样性，提高模型的泛化能力。

2.2.2 模型训练策略

采用迁移学习的方法，加载预训练权重，冻结骨干网络参数，仅训练顶部分类头。这种方法可以充分利用预训练模型的特征提取能力，同时减少训练时间和计算资源。

使用特征缓存技术，将backbone提取的特征预先存储为numpy数组，这样在训练分类头时就不需要重复进行特征提取，显著加速训练过程。特征缓存技术特别适合迁移学习场景，因为backbone的参数是固定的，特征提取结果不会改变。

2.2.3 模型架构改进探索

ResNet**架构研究**：深入学习ResNet残差网络的设计思想，理解残差连接的作用机制。残差连接通过跳跃连接将输入直接加到输出上，使得网络可以学习残差映射，从而解决深度网络的梯度消失问题。研究如何将MobileNetV2替换为ResNet18/ResNet50，探索残差连接对模型表达能力的提升作用。

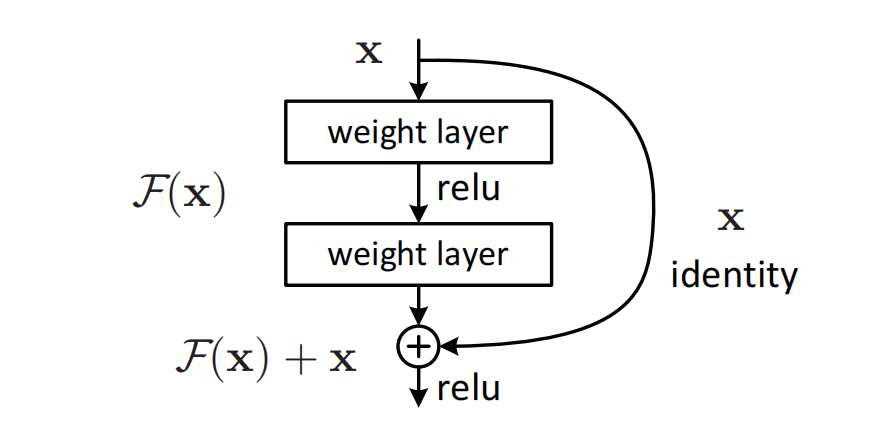


图1：残差学习的一个构建模块

Transformer**架构研究**：研究Vision Transformer在图像分类任务中的应用，探索自注意力机制在计算机视觉中的优势。Vision Transformer将图像分割成固定大小的patches，然后将每个patch线性投影为embedding，加上位置编码后输入到Transformer编码器中。自注意力机制可以捕获图像中任意两个位置之间的关系，具有全局感受野。

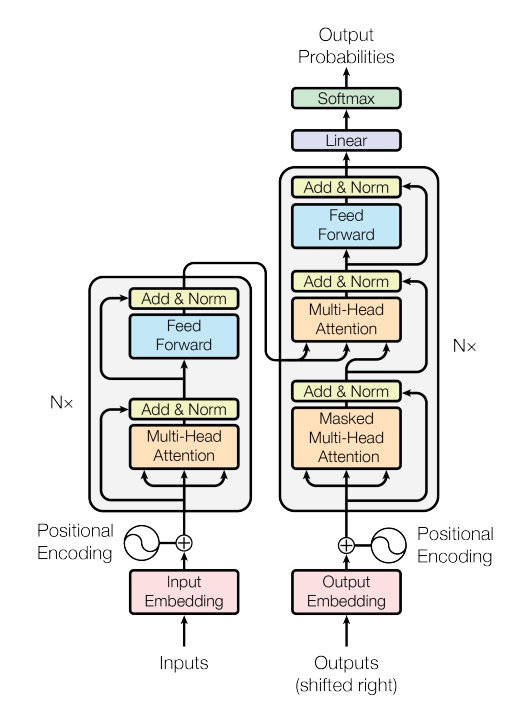


图2：Transformer 模型架构

**代码修改尝试**：通过修改src/models.py中的define\_net函数，尝试集成mindcv库中的不同模型架构。mindcv是MindSpore的计算机视觉库，提供了大量预训练的模型。通过调用mindcv.create\_model函数，可以方便地创建各种预训练模型。

2.2.4 模型优化技术研究

**网络剪枝理论**：学习通过L1/L2正则化对BatchNorm的γ参数进行剪枝的方法。网络剪枝通过移除不重要的连接或通道来减少模型大小和计算量。BatchNorm的γ参数可以反映通道的重要性，通过正则化这些参数，可以识别出不重要的通道并将其移除。

**量化压缩技术**：研究后训练量化（PTQ）和量化感知训练（QAT）技术。量化将模型参数从32位浮点数转换为8位整数，可以大幅减少模型大小和推理时间。PTQ在训练完成后进行量化，而QAT在训练过程中就考虑量化的影响。

**知识蒸馏方法**：学习Teacher-Student蒸馏方法的原理和应用。知识蒸馏使用一个大的教师网络来指导一个小的学生网络学习，学生网络可以学习到教师网络的知识，同时保持较小的模型大小。

三、研究结果

3.1 项目完成情况

根据中期报告，项目已成功完成以下工作：

3.1.1 MobileNetV2基础实现

成功搭建了基于MindSpore的MobileNetV2模型，完成了猫狗分类任务的训练和部署验证。在模型搭建过程中，深入理解了MobileNetV2的设计思想，包括深度可分离卷积、倒残差结构等关键技术。

深度可分离卷积将标准卷积分解为深度卷积和点卷积两个步骤，可以大幅减少计算量。倒残差结构在残差块中先进行1x1卷积扩展通道数，然后进行3x3深度卷积，最后进行1x1卷积压缩通道数，这种结构在保持精度的同时提高了计算效率。

在训练过程中，使用了迁移学习的方法，加载了ImageNet预训练权重，冻结了backbone参数，只训练分类头。这种方法不仅减少了训练时间，还提高了模型的性能。

3.1.2 模型架构改进探索

**ResNet架构学习**：深入学习了ResNet残差网络的设计思想和数学原理。ResNet通过残差连接解决了深度网络的梯度消失问题，使得网络可以训练到100层以上。残差连接的核心思想是让网络学习残差映射，即H(x) = F(x) + x，其中F(x)是残差函数，x是恒等映射。

通过研究ResNet的代码实现，理解了残差块的结构和实现方法。每个残差块包含两个3x3卷积层，中间有BatchNorm和ReLU激活函数。残差连接将输入直接加到输出上，使得梯度可以直接传播到前面的层。

**Vision Transformer研究**：研究了Vision Transformer在图像分类任务中的应用机制。ViT将图像分割成16x16的patches，每个patch被线性投影为embedding，加上位置编码后输入到Transformer编码器中。Transformer编码器使用多头自注意力机制和前馈网络，可以捕获图像中任意两个位置之间的关系。

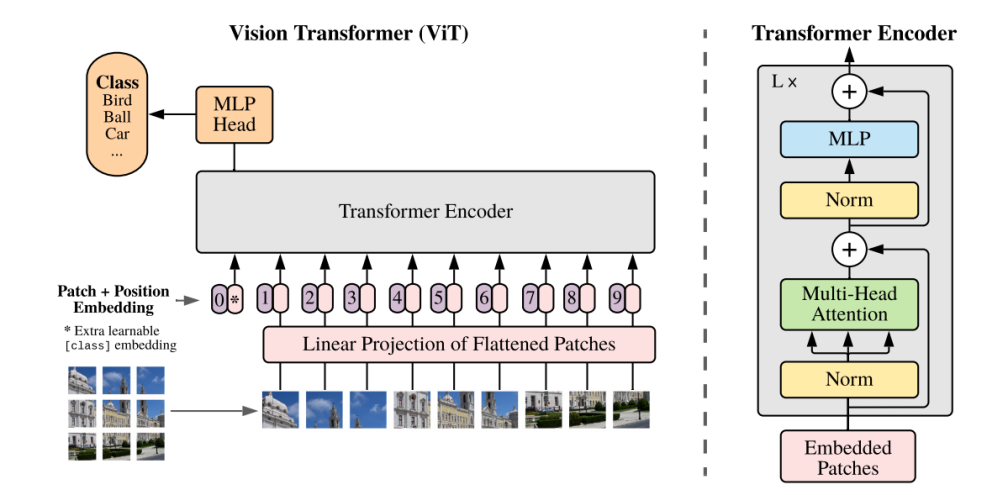


图3：ViT工作示意图

通过研究ViT的代码实现，理解了patch embedding、位置编码、多头自注意力等关键组件。自注意力机制可以计算每个patch与其他所有patch的关系，具有全局感受野，这是ViT相比CNN的一个重要优势。

**代码修改尝试**：通过mindcv库尝试集成ShuffleNetV1模型，修改了相关代码。mindcv提供了丰富的预训练模型，包括ResNet、MobileNet、ShuffleNet、ViT等。通过调用mindcv.create\_model函数，可以方便地创建这些模型。

在代码修改过程中，遇到了参数不匹配、接口不一致等问题。通过查阅文档、分析代码、调试程序等方式，逐步解决了这些问题，积累了宝贵的编程经验。

3.1.3 技术方案研究

**模型轻量化研究**：深入研究了模型轻量化的理论基础，包括网络剪枝、量化压缩、知识蒸馏等方法。网络剪枝通过移除不重要的连接或通道来减少模型大小；量化压缩将模型参数从32位浮点数转换为8位整数；知识蒸馏使用大模型指导小模型学习。

**端侧部署研究**：探索了端侧部署的优化策略和技术方案。端侧部署面临内存限制、计算能力限制、功耗限制等挑战。通过模型压缩、算子优化、内存管理等方法，可以在保证精度的同时提高推理效率。

**技术流程建立**：建立了完整的模型训练到部署的技术流程，包括数据准备、模型训练、特征缓存、模型转换、端侧部署等环节。每个环节都有详细的技术方案和实现方法。

3.2 技术实现成果

3.2.1 完整开发流程

建立了从数据预处理、模型训练、特征缓存、模型转换到端侧部署的完整技术链路。数据预处理包括图像清洗、格式统一、尺寸调整等；模型训练包括网络搭建、参数初始化、损失函数配置、优化器设置等；特征缓存将backbone特征预先存储，加速训练过程；模型转换将训练好的模型转换为MindIR格式；端侧部署将模型部署到移动设备上进行推理。

3.2.2 代码框架搭建

实现了基于MobileNetV2的完整代码框架，包括数据加载、模型定义、训练循环、模型保存等模块。代码框架具有良好的模块化设计，便于后续扩展和修改。通过配置文件管理超参数，提高了代码的可维护性。

3.2.3 工程化实现

成功将训练好的模型转换为MindIR格式，并通过MindSpore Lite部署到移动端设备。MindIR是MindSpore的中间表示格式，可以跨平台使用。MindSpore Lite是轻量级的推理引擎，专门用于端侧部署。

3.3 学习成果

3.3.1 理论知识掌握

深入理解了不同网络架构的设计原理，包括深度可分离卷积、残差连接、自注意力机制等关键技术。深度可分离卷积将标准卷积分解为深度卷积和点卷积，可以大幅减少计算量；残差连接通过跳跃连接解决梯度消失问题；自注意力机制可以捕获全局依赖关系。

3.3.2 实践能力提升

掌握了MindSpore框架的使用方法，具备了独立进行深度学习项目开发的能力。学会了数据预处理、模型搭建、训练配置、模型保存等基本操作，能够独立完成一个完整的深度学习项目。

3.3.3 问题解决能力

在项目过程中遇到的技术问题，通过查阅文档、分析代码、讨论交流等方式得到了有效解决。例如，在集成mindcv库时遇到了版本兼容性问题，通过查阅官方文档和社区讨论，找到了解决方案。

3.3.4 代码修改经验

通过尝试修改代码集成不同模型架构，积累了宝贵的编程实践经验和问题调试能力。学会了如何阅读和理解复杂的代码结构，如何定位和解决代码中的问题，如何调试和优化程序性能。

四、创新点

4.1 技术方法创新

4.1.1 多架构对比研究

系统性地研究了MobileNetV2、ResNet、Vision Transformer等不同架构的设计思想和应用特点，为模型选择提供了理论依据。通过对比分析，深入理解了各种架构的优缺点和适用场景。

MobileNetV2适合资源受限的场景，具有较小的模型大小和较快的推理速度；ResNet适合精度要求较高的场景，具有强大的特征提取能力；Vision Transformer适合大规模数据集的场景，具有全局建模能力。

4.1.2 端到端优化流程

建立了完整的"数据准备-模型训练-模型转换-端侧部署"的深度学习应用开发流程，形成了可复制的技术方案。该流程涵盖了从原始数据到最终应用的各个环节，为后续项目提供了参考模板。

4.1.3 多层次优化策略

结合网络剪枝、量化压缩、知识蒸馏等多种优化技术，探索了模型精度与效率的平衡方法。通过多种优化技术的组合使用，可以在保证精度的同时大幅减少模型大小和推理时间。

4.2 实践应用创新

4.2.1 教学实践价值

项目为深度学习初学者提供了完整的实践案例，从理论到应用形成了闭环。通过动手实践，学生可以深入理解深度学习的原理和应用，培养解决实际问题的能力。

4.2.2 工程化实现

成功将学术研究成果转化为可实际部署的工程应用，验证了技术的实用性。通过端侧部署，验证了模型在实际应用中的性能和稳定性。

4.2.3 跨平台兼容

基于MindSpore框架实现了CPU、GPU、NPU等多平台的无缝切换，体现了框架的优越性。MindSpore的跨平台特性使得模型可以在不同硬件平台上运行，提高了应用的灵活性。

五、结束语

本项目成功完成了基于MindSpore框架的图像分类模型训练与端侧部署实践。通过系统性的理论学习和实践探索，我们不仅掌握了MobileNetV2网络的设计原理，更重要的是建立了完整的深度学习应用开发流程。

项目的主要贡献包括：1）成功实现了基于MobileNetV2的猫狗分类模型训练与移动端部署，验证了MindSpore框架的实用性；2）深入研究了ResNet、Transformer等不同网络架构的设计思想，为后续学习奠定了坚实的理论基础；3）通过代码修改尝试，积累了宝贵的编程实践经验和问题解决能力；4）建立了完整的深度学习应用开发流程，为后续项目提供了参考模板。

通过本项目，我们深入理解了深度学习的基本原理和工程实践，掌握了模型训练、优化和部署的完整技能。虽然在架构改进方面还需要进一步的学习和实践，但这个过程让我们对深度学习有了更深入的认识，也培养了我们的学习能力和实践能力。

在项目过程中，我们遇到了许多技术难题，如环境配置、代码调试、模型部署等。通过查阅文档、分析代码、讨论交流等方式，我们逐步解决了这些问题，不仅提高了技术水平，也培养了解决问题的能力和团队合作精神。

未来，我们将继续深入学习更先进的模型架构和优化技术，如神经架构搜索（NAS）、联邦学习、多模态学习等。通过更多的实践来提升自己的技术能力，为深度学习技术的普及和应用贡献自己的力量。同时，我们也会将学到的知识和经验分享给其他同学，促进共同进步。

六、参考文献

[1] Howard A G, Zhu M, Chen B, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications[J]. arXiv preprint arXiv:1704.04861, 2017.

[2] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 770-778.

[3] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale[J]. arXiv preprint arXiv:2010.11929, 2020.

[4] 华为技术有限公司. MindSpore深度学习框架技术白皮书[R]. 2020.

[5] Sandler M, Howard A, Zhu M, et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 4510-4520.

[6] Zhang X, Zhou X, Lin M, et al. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 6848-6856.

[7] Liu Z, Lin Y, Cao Y, et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012-10022.

[8] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J]. arXiv preprint arXiv:1512.03385, 2015.

[9] Vaswani A, Shazeer N, Parmar N, et al. Attention is All You Need[C]. Advances in Neural Information Processing Systems, 2017: 5998-6008.