



浙江工业大学

硕士学位论文

人的介入提升深度学习应用精度的方法及其实例研究

作者姓名	花婷婷
指导教师	赵云波 教授
学科专业	控制工程
学位类型	工程硕士
培养类别	全日制专业型硕士
所在学院	信息工程学院

提交日期：2022 年 1 月

Research on the Method of Human Intervention to Improve the Application Accuracy of Deep Learning and Its Applications

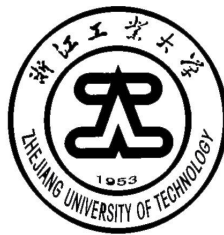
Dissertation Submitted to

Zhejiang University of Technology

in partial fulfillment of the requirement

for the degree of

Master of Engineering



by

Dissertation Supervisor: Yun-Bo Zhao

Jan., 2022

浙江工业大学学位论文原创性声明

本人郑重声明：所提交的学位论文是本人在导师的指导下，独立进行研究工作所取得的研究成果。除文中已经加以标注引用的内容外，本论文不包含其他个人或集体已经发表或撰写过的研究成果，也不含为获得浙江工业大学或其它教育机构的学位证书而使用过的材料。对本文的研究作出重要贡献的个人和集体，均已在文中以明确方式标明。本人承担本声明的法律责任。

作者签名：花婷婷

日期：2021 年 12 月

学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权浙江工业大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于 1、保密□，在一年解密后适用本授权书。

2、保密□，在二年解密后适用本授权书。

3、保密□，在三年解密后适用本授权书。

4、不保密☑。

（请在以上相应方框内打“√”）

作者签名：花婷婷

日期：2021 年 12 月

导师签名：赵云波

日期：2021 年 12 月

中图分类号 TP11

学校代码 10337

UDC 004.5

密级 公开

研究生类别 全日制专业型硕士研究生



浙江工业大学

工程硕士学位论文

人的介入提升深度学习应用精度的方法及其实例研究

Research on the Method of Human Intervention to Improve the
Application Accuracy of Deep Learning and Its Applications

作者姓名 花婷婷

第一导师 赵云波

学位类型 工程硕士

学科专业 控制工程

培养单位 信息工程学院

研究方向 人机融合

答辩日期: 2021 年 12 月 8 日

人的介入提升深度学习应用精度的方法及其实例研究

摘 要

现实中许多智能应用往往并不具备获取足量、高质的训练数据的天然条件，这一定程度上影响了深度学习算法所能达到的精度；而与此同时，众多领域却对其应用精度有着极高的现实要求：这两方面的原因使得在深度学习的基本框架外寻求精度的提升成为近些年的一个重要关注点。

在这一领域中，有研究者发现人的介入可能是提升深度学习应用精度的一种可行途径。该方法借用人机系统中介入思想以解决非人机系统中的相关技术难题，即在机器算法有误时允许人的介入来提升整体系统的应用精度。

本文基于典型实例研究了人的介入提升深度学习应用精度的两种方法，即人的分歧介入方法和人的融合介入方法。前者方法考虑在多个算法产生预测分歧之处，决策权将交由经验丰富的人直接控制；后者则更多的考虑寻求基于普通用户介入下的可操作性，在算法的预测置信度较低时，利用人在感知层面的信息传输实现对算法错误决策的可靠修正。具体研究工作如下：

（1）面对珍珠分拣任务的经济效益与质量需求的高标准，提出了一种人的分歧介入方法，该方法在公开珍珠数据集上，以 4.1% 的额外人工成本提升了近 4% 的珍珠分拣精度，较大程度上解决了现有珍珠分拣精度不高的问题。本实验为珍珠分拣系统提供额外的冗余 AI 算法，通过比对多个算法的预测结果，在产生分歧之处作为仲裁机制来判断系统的预测是否出错，以此驱动珍珠分拣人员的介入。最终使用定义分歧准确指数和额外成本指数这两大性能评价指标来评估设计系统的整体性能。

（2）针对野生鸟类资源信息识别对整个社会生态价值和经济价值的高需求，提出了一种人的融合介入方法，该方法在公开基准 CUB-200-2011 数据集上，以最小化的人工成本将鸟类细粒度图像识别精度提升到 93.8%，实现了目前机器视觉算法在细粒度鸟类图像识别难题上的突破。借用二分类模型评估思想，本实验通过分析算法的 ROC 指标，选取类别预测最优阈值作为仲裁机制，当算法的预测置信度低于阈值条件即作为人的介入时机。并通过将人的视觉信息与嵌入的鸟类知识图谱信息融合为算法提供更具信息性的特征判断实现可靠分类。

关键词：人的介入，人机混合智能系统，深度学习，仲裁机制

RESEARCH ON THE METHOD OF HUMAN INTERVENTION TO IMPROVE THE APPLICATION ACCURACY OF DEEP LEARNING AND ITS APPLICATIONS

ABSTRACT

In reality, many intelligent applications often do not have the natural conditions to obtain sufficient and high-quality training data, which affects the accuracy of deep learning algorithm to a certain extent. At the same time, many fields have very high practical requirements for its application accuracy. These two reasons make it an important concern to seek the improvement of accuracy outside the basic framework of deep learning in recent years.

In this field, some researchers found that human intervention may be a feasible way to improve the application accuracy of deep learning. This method borrows the idea of human-machine system mediation to solve the related technical problems in non-human-machine system. That is, when the machine algorithm is wrong, human intervention is allowed to improve the application accuracy of the overall system.

Based on typical examples, this paper studies two methods of human intervention to improve the application accuracy of deep learning, namely human disagreement intervention method and human fusion intervention method. The former method considers that where multiple algorithms have prediction divergences, the decision-making power will be directly controlled by experienced people. The latter considers the operability based on the intervention of ordinary users. When the prediction confidence of the algorithm is low, the reliable correction of the wrong decision of the algorithm is realized by using human information transmission at the perceptual level. The specific research works are as follows:

(1) Facing the high standard of economic benefit and quality demand of Pearl sorting task, this paper proposes a human disagreement intervention method. This method improves the Pearl sorting accuracy by nearly 4% with 4.1% additional labor cost on the open pearl data set, and solves the problem of low accuracy of the existing pearl sorting to a great extent. The experiment provides additional redundant AI algorithms for the Pearl sorting system. By comparing the prediction results of multiple algorithms, it is used as an arbitration mechanism to judge whether the

prediction of the system is wrong, so as to drive the intervention of Pearl sorting personnel. Finally, two performance evaluation indexes, definition divergence accuracy index and additional cost index, are used to evaluate the overall performance of the design system.

(2) In view of the high demand of wild bird resource information recognition for the ecological and economic value of the whole society, this paper proposes a human fusion intervention method. This method improves the accuracy of bird fine-grained image recognition to 93.8% with minimal labor cost on the open benchmark cub-200-2011 data set. It realizes the breakthrough of machine vision algorithm in fine-grained bird image recognition. Using the evaluation idea of two classification model, this experiment analyzes the ROC index of the algorithm and selects the optimal threshold of category prediction as the arbitration mechanism. When the prediction confidence of the algorithm is lower than the threshold condition, it is regarded as the opportunity of human intervention. By fusing human visual information with the embedded bird knowledge graph information, the algorithm provides more informative feature judgment to realize reliable classification.

KEY WORDS: human intervention, human-machine hybrid intelligent system, deep learning, arbitration mechanism

目 录

摘 要.....	I
ABSTRACT.....	III
插图清单.....	VII
表格清单.....	VIII
第一章 绪 论.....	1
1.1 研究背景和意义.....	1
1.2 国内外研究现状.....	2
1.3 本文研究内容.....	6
1.4 本文组织架构.....	7
1.5 本章小结.....	8
第二章 深度学习技术和人的介入策略方法.....	9
2.1 深度学习相关技术.....	9
2.1.1 ResNet 网络原理.....	9
2.1.2 批归一化.....	10
2.1.3 注意力机制.....	11
2.2 人的介入策略方法.....	12
2.2.1 人机融合的重要性.....	12
2.2.2 人机混合智能系统.....	13
2.2.3 引入人的介入控制策略.....	15
2.2.4 人的介入控制因素.....	15
2.3 本章小结.....	17
第三章 基于人的分歧介入的珍珠自动分拣应用.....	18
3.1 珍珠分拣应用背景.....	18
3.2 方法模块设计.....	19
3.2.1 算法模块设计.....	20
3.2.2 仲裁模块设计.....	21
3.3 实验所用数据集和处理方法.....	22
3.3.1 珍珠分拣数据集.....	22
3.3.2 珍珠分拣数据集的处理.....	23
3.3.3 评价指标.....	23
3.4 基于二算法分歧的实验与分析.....	23

3.4.1 实验设置.....	24
3.4.2 评价指标分析.....	25
3.5 基于多算法分歧的改进优化实验与分析.....	26
3.5.1 提升分歧准确指数.....	26
3.5.2 降低额外成本指数.....	28
3.6 本章小节.....	29
第四章 基于人的融合介入的野生鸟类细粒度图像识别.....	30
4.1 野生鸟类图像识别研究背景.....	30
4.2 系统设计.....	31
4.2.1 方法框架.....	31
4.2.2 技术路线.....	39
4.3 实验所用数据集和阈值设置.....	40
4.3.1 鸟类数据集.....	40
4.3.2 训练参数设定.....	41
4.3.3 阈值设置.....	41
4.4 系统算法与分析.....	43
4.4.1 用户建模.....	43
4.4.2 系统提问设置.....	44
4.4.3 用户响应更新类概率.....	46
4.4.4 结果分析.....	47
4.5 本章小结.....	50
第五章 总结与展望.....	51
5.1 总结.....	51
5.2 展望.....	52
参考文献.....	53
致 谢.....	58
作者简介.....	59
1 作者简历.....	59
2 攻读硕士学位期间发表的学术论文.....	59
3 参与的科研项目及获奖情况.....	59
4 发明专利.....	59
学位论文数据集.....	61

插图清单

图 2-1 残差网络结构示意图.....	9
图 2-2 神经网络其中两层结构图.....	10
图 2-3 神经元模型.....	11
图 2-4 注意力机制分类结构.....	12
图 2-5 近十年关于“human-in-the-loop”研究报告数量.....	13
图 2-6 人机混合智能系统结构图.....	14
图 3-1 人的分歧介入框架图.....	20
图 3-2 分歧仲裁框图.....	21
图 3-3 珍珠七分类视图.....	22
图 3-4 二算法分歧珍珠分类效果图.....	26
图 3-5 基于多算法的人的分歧介入框图.....	27
图 3-6 三算法分歧珍珠分类效果图.....	27
图 3-7 基于算法集成的人的分歧介入框图.....	28
图 4-1 多源信息融合的细粒度图像识别系统框图.....	31
图 4-2 ResNet50 网络结构.....	32
图 4-3 注意力机制模式.....	33
图 4-4 网络局部注意力模块.....	33
图 4-5 模型预测分类热力图.....	34
图 4-6 模型训练阶段准确率和损失曲线图.....	35
图 4-7 基于鸟类的知识图谱.....	35
图 4-8 知识图谱的技术架构图.....	36
图 4-9 鸟类百科网站.....	37
图 4-10 GUI 用户界面.....	38
图 4-11 属性名词参考图.....	39
图 4-12 数据集分布.....	40
图 4-13 类 ROC 曲线绘制.....	42
图 4-14 平均-ROC 曲线绘制.....	43
图 4-15 用户对每种类别介入响应数量.....	49
图 4-16 用户平均回答每题所花时间.....	49
图 4-17 模型预测精度与用户所花时间关系图.....	50

表格清单

表 3-1 ResNet50 + SE-ResNet50 算法分歧准确率.....	24
表 3-2 ResNet50 + Vgg16 算法分歧准确率.....	24
表 3-3 SE-ResNet50 + Vgg16 算法分歧准确率.....	25
表 3-4 二算法分歧性能结果.....	25
表 3-5 多算法分歧性能结果.....	28
表 3-6 三算法集成分歧性能结果.....	28
表 4-1 鸟的部位属性描述.....	36
表 4-2 混淆矩阵.....	42
表 4-3 用户建模伪代码.....	44
表 4-4 问题设计伪代码.....	45
表 4-5 用户响应更新类概率伪代码.....	47
表 4-6 两种阈值设置方法对比.....	47
表 4-7 不同鸟类细粒度图像识别模型对比.....	48

第一章 绪 论

现实中，许多智能应用往往并不具备获取足量、高质的训练数据的天然条件，这一定程度上影响了深度学习算法所能达到的精度；与此同时，众多领域却对应用精度有着极高的现实要求。这就使得在深度学习的基本框架外寻求精度的提升成为了近些年的一个重要关注点。基于此，本章总结和分析了国内外现有的研究方法，并阐述本文的主要研究内容与架构。

1.1 研究背景和意义

近些年得益于互联网/物联网的大数据支持，和以 GPU 运算能力为代表的计算技术的提升，以深度学习为基础的人工智能（Artificial Intelligence，简称 AI）技术蓬勃发展，引发了人工智能发展的一个新浪潮。深度学习的核心思想，是构建能模拟人类大脑进行事物分析和学习的神经网络，从而实现一个更具自动化和便捷的智能时代。目前，基于该技术在图像分类、语音识别和自然语言处理等多个科研领域^[1-4]中获得了突破性进展，对智能交通、智慧医疗、军事、金融等重要应用领域的发展做出了巨大贡献。例如，具有数万亿市场价值的新兴技术领域——自动驾驶技术，依靠的就是以人工智能技术为基础做规模化应用，帮助人类在减少交通拥堵、生命安全、提高生产力等方面产生巨大效益；医疗领域里的疾病诊断技术，也是利用人工智能驱动的图像识别技术取代人力，不仅可以提高准确率和分析速度，而且还能及时诊断出疾病，加快救生药的研发，最大程度地减少患者被误诊、漏诊风险，提高治愈率。从现有的深度学习发展角度来看，未来的人工智能技术会在更多领域展现其极具潜力的应用前景。

众所周知，深度学习的过程就是数据建模的过程，深度学习算法的成功在很大程度上取决于数据的质量和数量^[5,6]。但目前，尽管以深度学习为基础的人工智能技术已经到来，并且还正在进行着迅猛的发展。可我们意识到，现实中许多深度学习应用往往并不具备获取足量、高质的训练数据的天然条件^[7]，或者由于其他的某些原因如现实环境下的不确定性干扰，难以在实用中训练出具有足够精度的算法。而与此同时，众多应用却对算法精度有着极高的现实需求：随着智能应用的不断延伸与落地，人们对自动化系统的高精度要求会越来越严苛，所设计的人工智能系统应具有可靠、精准的决策输出。

这里所谓对高精度的要求，可能是出于对系统应用犯错成本和代价的考虑，也可能只是更高精度会带来更好的收益^[8,9]。例如，表面瑕疵检测是精密零部件质量检测环节极其重要的一环，器件表面出现瑕疵会极大影响到产品的使用性

能、安全和可靠性。深度学习技术可以对零部件表面瑕疵进行检测，大大减少人的劳动力，但如果深度学习技术的准确性达不到所要求的标准，那么反而会造成零部件检测效率降低，带来经济和信誉的双重损失。再如，在农作物育种方面，深度学习技术已经可以帮助筛选更好的基因，提高作物育种的效率和精准度，这是 AI 技术发展带来的优势。通过其对水稻、小麦等作物进行病虫害诊断，可大幅度减少人工对作物进行逐一排查的时间和人力成本。并且，由深度学习驱动的 AI 技术准确性的进一步提升将更有助于降低病虫害带来的损失。

而目前，现有对深度学习算法精度的进一步研究和提升，大多是在深度学习的基本框架内进行。针对数据量少、优质数据难获取等问题，考虑采取无监督学习和弱监督学习等方式减小算法对数据依赖的程度，但相对发展尚不成熟。从另一方面讲，考虑通过训练更大更深的网络从而实现更好的训练效果^[10]，但网络深度过深也会造成模型饱和，从而限制深层网络的学习，同时对计算资源和时间成本的要求也大大增加^[11]。并且现实环境相对复杂，实际使用的智能系统会遭受比原型系统更多的内外部因素干扰，这也给模型算法的训练带来了巨大考验。

因此，在现有的基于深度学习的技术框架下，想要进一步提升深度学习自身算法精度并不容易，这导致其技术限制了应用的广度和深度。至少从有意义的技术研究角度来看，我们需要假设在可预见的未来仍是以弱人工智能的形式存在，这就使得该问题在短期内难以仅在深度学习和自动化领域内解决^[12]。而这也引发了一个新的研究方向，在深度学习的基本框架外寻求精度的提升成为了近些年的一个重要关注点。

1.2 国内外研究现状

总结人类和机器在学习或处理事物方面最大的区别，就在于人类不仅关注其视觉信息，还会考虑从他们的过往经验和对事物的额外信息描述中获取先验知识。这种在人脑中将视觉和知识经验的潜在关联，可以更好地掌握事物的关键信息特征进行决策判断，且对不属于该类事物的表征也能够表现出较好的判别性。

因此，借助人的这一智能优势，将人置于深度学习的基本框架之上成为了提升其精度的可行方案。而这一思想本质源于“Human-in-the-loop”（简称 HITL）方法。HITL 的一般性概括指计算问题的求解需要人的参与或引入人的参与能提升问题求解的效果。而现在更核心的研究想法就是尽可能花少量的人工成本，将人介入建模的某个阶段以实现更加精准的预测系统。这里，人的介入可以为模型应用提供优质的训练数据信息，也能够让模型学习人在处理任务

时的思维方式，还可在机器处理任务过程中出现不确定性或错误决策时，人类及时介入处理。

目前，基于 HITL 的处理方式主要分为三种：一是在数据处理阶段利用人对特定数据做筛选或标签补充；一是在模型学习和训练阶段引入人对任务处理的认知和经验；还有一是在系统执行阶段结合人的智能优势协同决策。具体地，现有对 HITL 的实施主要从以下三个阶段展开：

（1）数据处理阶段

研究发现，相比模型训练和预测工作，大多数研究人员会将 80% 的工作精力花在数据处理阶段，以解决数据量不足、低质等因素导致算法精度提升困难问题。基于此，部分研究学者尝试从数据层面来优化系统性能，使模型能够达到一个更好的预测状态。现阶段，数据处理方式主要有数据标注^[13-15]、构建生成式样本^[16-19]、预训练^[20-22]、数据增强^[23-25]等方法。其中，运用的较为广泛、且直观想到的就是数据标注方法。因为深度学习模型需要从大量有标签数据中进行学习，数据标注越精准、量越大，算法的准确率将会越高^[26]，从而达到令人满意的效果与避免模型的过拟合。为此，数据标注是大多数 AI 算法有效运行的关键技术。可是，标注数据需要花费大量的人工劳动力和时间成本。在大多应用场景下，机器能够获取的通常是大规模的未标注数据集，要得到已标注数据集有难度，并且高质量的标签将会产生昂贵的标注成本^[27]。且随着大数据和 AI 的快速发展，可用的数据集在数据量以及密度方面难以支撑新技术的发展要求。

为此，研究者们开始寻求各种可降低数据标注成本的技术方法。一些研究人员通过生成样本来构建新的数据集，以生成更丰富的样本空间^[28-31]，从而加快模型迭代。还有通过预训练和迁移学习法来实现，如 Transformer^[32]、BERT 和 GPT。这种利用从训练好的数据模型中对特定任务的微调方式发现能够在更多应用中获得可喜的成果。还有为缓解数据量带来的不足问题，部分研究者采用数据增强技术来扩展有标注训练数据集。然而，这些方法的前提仍然需要标注大量数据信息，只是相比前者在数据量上略有减少，并没有从本质上解决问题。另外，针对大量的数据标注问题，还有采用众包^[33]的形式来获取，但这种形式由于人工标注者的经验水平很容易导致标注信息质量层次不齐，影响最终算法的准确性。同时，我们注意到，在众多数据集中，对算法性能的干扰大多取决于一些“关键样本”。而这些“关键样本”足以让模型学习到错误的特征信息。为此如何让模型尽可能避免学习到那些“关键样本”就成为了一个亟待解决的问题。由于模型自身难以对传入的数据进行一个智能的辨别，因此，许多研究者提出了各种基于 HITL 方法进行更快更精准的数据标注。

这些方法与传统的手工标注不同，它更注重人的介入方式和操作体验，以实现一种快速迭代和智能反馈。例如，在自然语言处理应用中，实体抽取^[34]是

一项很重要的任务，它的成功或失败将直接影响自然语言处理技术的性能好坏。因此，Zhang 等人^[35]提出了一种基于 HITL 的实体抽取方法，人类用户通过制定识别实体的正则表达式，并手动标记文档语料库中提及的实体，使得模型能够基于所提供的信息训练出一个高效和精准的分类器。并且惊讶的发现，相比过往研究，该方法不仅所需人工时间成本较少，所得的分类精度也有了新的突破。同样在行人重识别任务中，Liu 等人^[36]对实验数据的处理采用了一种基于强化学习的 HITL 方法。该方法减少了数据预先标注的限制，而是由人类用户动态选择目标训练样本，模型通过逐步收集到的数据进行模型训练，并基于人类对训练样本的正负反馈标记，实现模型参数的微调，最终达到比以往算法更优的识别效果。此外，Butler 等人^[37]设计了一种基于 HITL 的微表情识别任务数据标注系统，该方法为自动数据标注的手动校对提供了一种灵活的界面控制，从而可以基于人的数据标签纠正确保扩展数据的准确性和可用性。借鉴自动标注思想，Adhikari 等人^[38]针对目标检测数据边界框的手动标注任务，融合 HITL 实现了一种半自动的数据标注方法，该方法在小批量有标记图像上迭代训练目标检测器，并学习为下一批无标注图像提供边界框注释，人只需纠正机器标注有误的数据信息即可。结果显示，此方法节省了 75% 的人工标注工作，显著减少了人工劳动时间和成本。Ravanbakhsh 等人^[39]通过对未标注的医学图像分割数据的不确定性度量，将具有高不确定性样本提取交由专家进行注释，最终获得了更优性能的图像分割技术。

从上述研究可以看出，随着用户交互信息的不断迭代，模型对数据的参数学习能够不断得到优化，一定程度上提升了深度学习算法对任务处理的效能。

（2）模型训练阶段

但相对而言，前者的研究工作还只是从数据处理的角度出发，并未将人类的经验和知识最大化融入模型，以帮助机器智能体更有效地获取特定任务知识信息。为此，部分研究开始转向模型训练阶段，这一阶段的目标将更加关注算法模型对额外知识引入后的学习情况。

之前已经有部分研究工作采取将先验知识集成到深度学习框架中的案例，如向模型嵌入知识图谱^[40,41]和文本描述^[42,43]。显然，在任务域中引入有效的辅助信息可以提升模型决策的推理能力和置信度。因为通过构建一个任务域的知识库或者知识图谱，可以引导模型从数据中获取所需的信息和知识，从而增强数据的特征表示。如 Marino 等人^[40]基于该思想在视觉算法模型中引入图谱查询神经网络，让模型学习用知识推理关于视觉世界的能力，从而改善深度学习在图像分类任务中的表现。同样，Hu 等人^[41]将各类人们已积累的逻辑规则引入深度神经网络中，利用人类意图和领域知识对神经网络模型的学习进行引导和训练，从而达到对深度学习算法的效果提升。Xu 等人^[44]则利用多模态数据融合的方式，创建了一种视觉-语义模型。其利用非结构化的文本信息和结构化的数据

库信息作为模型识别细粒度图像的先验知识，将挖掘的图像视觉特征映射到丰富的语义空间上，以产生互补的细粒度特征表示，供模型捕获更具判别性的特征进行学习。这种方式一方面能够让模型的识别结果更具可解释性，也能够提升模型算法对细粒度的图像任务识别的准确度^[45]。

总之，通过将常识知识和领域知识集成到深度学习算法中，能够让模型拥有更丰富的先验知识，实现更高层次的语义理解^[46]和更强的推理能力，从而大大降低模型对大规模标注数据的依赖。但是，这类知识信息有其局限。通常知识图谱或者文本描述都是人为构建，只会基于特定的对象信息给予描述，且过于规则化。然而现实任务大多动态复杂，模型需要有更灵活、更快速的思维逻辑，这是模型从上述知识信息中所无法学习到的。

为此，有研究者开始考虑在模型训练阶段引入人的认知思维做进一步的优化，该方法目前已在图像复原^[47]、图像分割^[48]、图像增强^[49,50]等领域中都有突破性的研究进展。如图像复原领域中，Weber 等人^[47]将人类知识嵌入到模型训练过程中，操作人员通过预先设计的用户界面主动对不同数据进行细化，给模型提供一些先验信息引导，从而使图像能够复原到一个理想状态。

其次，在很多应用场景中人具有某种独特的优势，如人能够基于小样本进行有效预测的能力，这相较于深度学习算法有本质的优越性。因此，可将人对数据集结构的一些认知特性引入机器学习算法中，使得模型能够在拟合或寻优的过程中利用这些认知特性，提升最终训练效果。Kumar 等人^[51]就是依据这样的思想，在处理某些自动化应用面临的稀疏数据问题（如在临床诊断和训练数据缺乏的医学领域），通过在训练阶段将医生的认知水平纳入模型学习框架，实现较以往的深度学习算法更高精度的医学诊断。Fong 等人^[52]通过引入人脑在目标识别时的功能性磁共振成像和对其的分析来改造算法分类的损失函数，以提高算法在少量训练样本下的图像分类精度。Taniguchi 等人^[53]通过在朴素贝叶斯方法中添加人的认知特性，一定程度上避免了小样本和不平衡数据在真实环境下不适用带来的负面影响，该方法最终在垃圾邮件分类问题上表现出极高的准确度。

显然，将 HITL 方法框架整合到模型训练阶段，可使得模型能够学习用户推理思维或基于用户反馈的知识信息对算法进行优化，这对于深度学习算法精度的提升呈现了一种新的方式突破。

（3）系统执行阶段

但事实上，让机器去学习人的思维逻辑处理问题是一件非常困难的技术挑战。并且，即使是基于前一小节的训练方法得出的较优算法，也不能表明模型就完全掌握了人在处理目标事物的思维方式。

另一层面，研究发现，许多在封闭环境或特定空间下训练良好的模型，一旦部署到现实自然环境中，就难以实现像实验训练那样完美的性能。这是因为

现实环境因素的不可控性和不确定性，导致随时可能有“新数据”的出现。这里的“新数据”可能是模型在训练阶段未接触过的，也可能是外部环境因素的干扰，也可能是系统获取到质量过低的输入数据，但不管是何种形式的数据，对模型的预测过程都会产生一定的偏差。如果模型对此没有做出鲁棒修正，这种现象会随着时间的推移，将累积更多的误差，最终很可能导致模型出现重大失误。这对于现实应用要求严苛的应用领域将会造成不可挽回的损失。

因此，在系统的执行阶段，需要对模型的算法预测进行可靠监督。而最为直接的监督方式就是进行人工介入。该方法在面对模型自身难以解决的复杂案例时，人的经验知识的加入能够使整个系统决策输出更加精准可靠，并实现在现实场景中的可用性。

目前，在医学图像诊断和自动驾驶等特定领域，人们已经认识到人的介入能够有效克服现有 AI 系统的局限，提升其算法可靠性。例如，2016 年一项关于淋巴结细胞图像癌症检测的研究显示，AI 检测系统和人类专家的结合可将两种方法单独使用时分别为 7.5% 和 3.5% 的错误率降为 0.5%，极大提升了检测精度。针对肺炎诊断技术，采用群体智能平台 Swarm^[54]，通过使群体在模仿生物群体的实时系统中协同工作，从而实现一种人机结合的群体智能诊断。该方式不仅提高了肺炎诊断的准确率，也对未来临床人工智能的部署和实施策略具有重大启示。再比如，完全基于 AI 图像识别的交通标志识别难以应对对抗攻击^[55]，将人类驾驶员保持在驾驶的闭环，随时在自动驾驶系统出错时接管驾驶，成为当前自动驾驶系统的一个必要特征。

可从目前方法中能够发现的是，这种形式需要人参与机器决策的整个过程。这若在面对算法自身有优势处理的地方，产生了非必要的人力成本。我们希望的是，在人工智能技术大力发展的前景下，剖析机器擅长处理的任务（几乎不存在错误输出的情形）并交于其自动执行。而在机器自身处理效果不佳的情况下交于有经验的人介入处理。但现阶段对这一层面的研究工作甚少，可以理解，因为该阶段的工作会涉及到机器与人类用户之间的决策交互，并且现实世界的应用更为复杂，需要考虑应用需求、系统配置和用户体验之间的权衡。因此，一定程度上会对该领域学者存在一些技术挑战。

但总体来说，在这一阶段通过人的介入实现一种人机协作方式，以提升传统单纯依靠机器自动化执行思想，在很大程度上能够开发出更实用可靠的人工智能系统，也有助于提升各类人工智能系统应用的准确性和鲁棒性。

1.3 本文研究内容

从近些年研究现状可知，人的介入是提升深度学习应用精度的一种可行途径。为此，本文将继承人的介入在系统执行阶段思想，进一步探究通过人的介

入策略来提升基于深度学习应用的决策精度。该方法借用人机系统中介入思想以解决非人机系统中的技术难题，即系统在正常运行时由机器的智能自主和机器的自动化共同起作用，而在机器算法有误时允许人的智能介入，以避免机器在此时的自主决策可能导致的严重失误。

由于不同应用场景的特殊性和广泛性，人的介入方式会在交互式界面、算法和人的角色上出现差异。但永恒不变的是其设计理念，通过这一形式既可以提升模型算法性能，也能够减少人在数据处理上的劳动成本。

因此，从人的介入策略角度来看，人的介入时机和人的介入方式是该方法设计的关键。其策略宗旨就是保证机器算法的失误可以被有效捕捉，从而通过介入人的智能实现对整体应用精度的有效提升。基于该思想，本文从系统控制方式以及人的经验考量通过两个实例应用分别设计了人的分歧介入和人的融合介入方法。

人的分歧介入方法更多针对的是在理想情况下人类更为擅长处理的一类任务。该任务大多情况下展现的是专业性较强的应用领域，其所需的人是有该领域较为丰富的经验，这样可使得当机器算法出现问题时，人类可直接接管机器的决策权，这是对系统决策可靠输出的最好表现。对此，本文面向珍珠分拣应用场景，为珍珠分拣系统提供额外的冗余 AI 算法，通过比对多个算法的预测结果，在产生分歧之处作为仲裁机制来判定系统的预测是否出错，以此驱动珍珠分拣人员的介入。最终使用定义分歧准确指数和额外成本指数这两大性能评价指标来评估设计系统的整体性能。

而人的融合介入方法则更多的考虑寻求基于普通用户介入下的可操作性，同样借助人的优异智能，在算法的预测置信度较低时，利用人在感知层面的信息传输辅助机器决策，实现对算法错误决策的可靠修正。为验证该想法，本文基于细粒度鸟类图像识别进行系统设计，通过分析算法模型的 ROC 指标，选取类别预测最优阈值作为仲裁机制，触发阈值条件即认定此处算法识别分类可能有误，可作为人介入系统决策的时机。并通过将人对图像的视觉信息与嵌入的鸟类知识图谱信息相融合，为算法提供更具信息性的特征判断实现精准分类。

1.4 本文组织架构

本文主要利用人的介入方式来提升深度学习应用精度研究，重点对深度学习的算法缺陷、人的介入策略设计等方面进行研究和分析。本文将分为五个章节进行阐述，其组织架构如下：

第一章绪论。本章阐述了在现有的深度学习框架下算法局限对现实应用的高精度要求的影响和意义，并从深度学习建模的各个阶段（数据处理阶段、模型训练阶段和系统执行阶段）分析近年来国内外研究学者们将人置于算法框架

之上实现对整体应用系统性能的提升效果，最后讲述本文的主要研究内容及组织架构。

第二章是对深度学习相关技术和人的介入策略方法进行介绍。首先基于深度学习中运用较为广泛的 ResNet 网络及技术细节进行阐述。随后引出本文的核心方法——人的介入控制策略，介绍了人机混合智能系统的设计理念，然后基于人机混合智能系统对人的介入控制策略方法进行分析。并从设计目标、介入决策和执行成本以及介入时机等因素对系统中人的介入控制设计进行了讨论。

第三章面向珍珠分拣应用，提出了一种通过人的分歧介入提升机器自动分拣准确性的方法。该章介绍了相应的算法模块和仲裁模块，最终通过实验在公开数据集上进行验证，证明了方法的有效性。

第四章针对鸟类细粒度图像识别应用设计了一种多源信息融合的识别系统，系统中采用人的融合介入方法来提升鸟类图像识别精度。本章首先介绍了系统设计的总体框架和实验操作的技术路线，随后根据实验数据集的特性对仲裁阈值进行讨论设置，并描述了基于人的融合介入方法所涉及的算法，最后根据实验结果对整个系统从算法精度和人力成本两个角度综合分析。

第五章是总结及展望。首先对全文工作做梳理总结，然后分析目前工作中存在的不足，并针对性的给予后续性研究展望。

1.5 本章小结

本章首先介绍了基于深度学习应用的高精度现实需求在现有技术算法框架下的研究背景和意义，然后从深度学习建模的三个阶段（数据处理、模型训练和系统执行）分析出发，讲述近年来国内外研究学者们通过将人置于深度学习算法框架之上来实现对系统性能的提升效果，最后描述了本文主要研究内容以及各章节的组成架构。

第二章 深度学习技术和人的介入策略方法

该章节主要介绍深度学习技术和人的介入策略方法相关内容。首先针对深度学习相关技术细节进行细致阐述。然后 2.2 节开始引出本文方法的思想来源，主要对人机融合思想的重要性概述、人机混合智能系统的设计理念以及人在该系统中的作用进行理论分析。其次，重点介绍人机混合智能系统中人的介入控制策略，这也是本文的关键方法。并从设计目标、介入决策和执行成本以及介入时机等因素对系统中人的介入控制设计进行了讨论。

2.1 深度学习相关技术

2.1.1 ResNet 网络原理

残差网络（ResNet），又称深度残差网络。在传统的卷积神经网络（Convolutional Neural Networks, CNN）中，随着网络层数的增加，多层特征会更加丰富。因此，网络层数变深，表明数据处理效果会越好。但是，简单的网络叠加会伴随梯度消失的问题，导致网络的收敛效果变差，从而准确率可能在达到饱和之后出现下降的趋势。

针对上述情况，残差网络采用了跨层恒等连接方式，以解决深度网络难以优化的难题，其结构模块如图 2-1 所示。

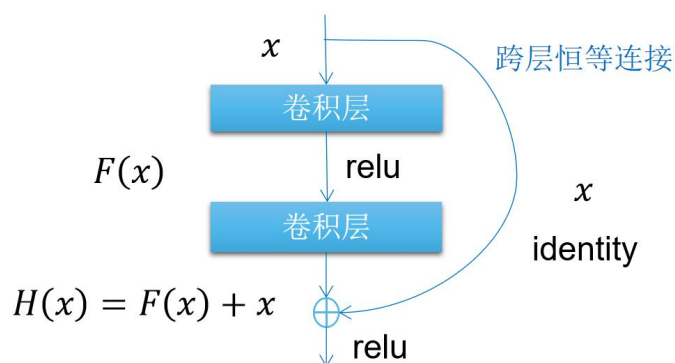


图 2-1 残差网络结构示意图

Fig 2-1. Structure diagram of residual network

假设某段网络的输入数据是 x ，期望输出是 $H(x)$ ，为保证减少来自上层误差的叠加影响，将输入 x 的信息也一并传输到输出端，即 $H(x) = F(x) + x$ ，这样做的好处是采用恒等映射的思想，保证模型的准确率在达到饱和后，能够保持在后面的层次中不会造成精度的下降。于是，残差网络将不再是学习完整的输出，

而是计算目标值 $H(x)$ 和 x 的差值，即所谓的残差 $F(x) = H(x) - x$ 。后期的训练目标则是将残差结果不断逼近于 0，即使网络层数变深，准确率也不会下降。

很明显，如果只是将输入的信息依据传统卷积层或全连接层进行信息传递，会存在一定的信息丢失、误差累积等问题。而 ResNet 则在一定程度上解决了这个难点，通过将输入的信息绕道传输给输出端，一方面能够保证信息传递的完整性，另一方面整个网络也只需要学习输入与输出之间差异的部分，简化了学习目标和难度。

2.1.2 批归一化

批归一化^[56]（Batch Normalization，简称 BN），是神经网络中一种特殊的层。它是深度学习中一项非常重要的技术，不仅可以使深层网络更易于训练，收敛速度加快，而且具备一定的正则化效果，防止模型出现过拟合问题，目前被广泛应用在基于 CNN 网络的分类任务中。

具体的 BN 操作流程如图 2-2 所示，假设一个深层模型的其中两层网络结构如（a）所示，为对输入特征进行归一化处理，需要在每个隐层的神经元激活值后加上一层 BN 层，如图（b）所示。其具体的实现流程如下：

$$\tilde{z}_i = \frac{z_i - E(z_i)}{\sqrt{Var(z_i)}} \quad (2-1)$$

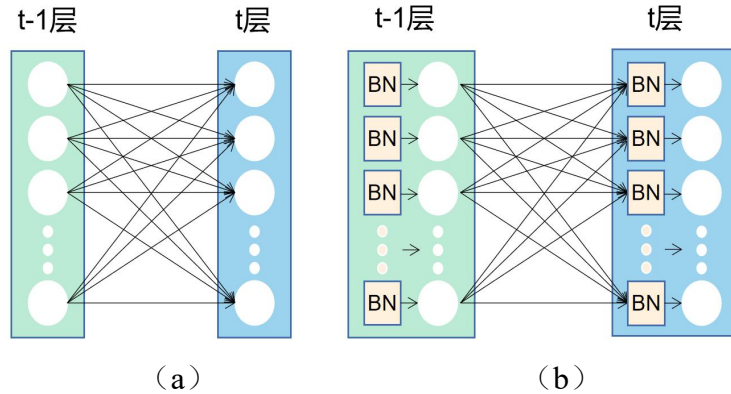


图 2-2 神经网络其中两层结构图；（a）未加 BN 前；（b）加入 BN 后

Fig 2-2. Two layer structure diagram of neural network, (a) Before BN operation, (b) After adding BN operation

图 2-3 为神经元模型，其中 $a = g(z)$ 是激活函数的输出值， g 是激活函数。给定一批输入 $X = [x_1, x_2, \dots, x_m]$ ，在神经网络第 t 层对应的中间隐藏值为 $[z_1, z_2, \dots, z_m]$ 。为实现上述公式操作，需要对其进行均值（公式 2-2）和方差（公式 2-3）的计算，公式 2-4 则表现的就是归一化过程，这一步实现的就是让每一个神经元激活值呈现均值为 0，方差为 1 的正态分布。但考虑到如果让其一直处在这样的分布上，会导致网络表达能力的下降，因此利用公式 2-5 实现

数据分布的平移和缩放，在每个神经元加入两个 γ 和 β 调节参数，以达到实现任意的均值和方差，增强网络的训练能力。

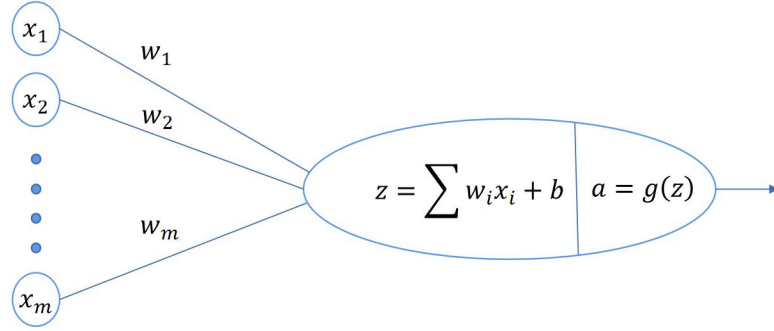


图 2-3 神经元模型

Fig 2-3. Neuron model

$$\mu = \frac{1}{m} \sum_{i=1}^m z_i \quad (2-2)$$

$$\sigma^2 = \frac{1}{m} \sum_{i=1}^m (z_i - \mu)^2 \quad (2-3)$$

$$\tilde{z}_i = \frac{z_i - \mu}{\sqrt{\sigma^2 + \epsilon}} \quad (2-4)$$

$$y_i = \gamma \tilde{z}_i + \beta \quad (2-5)$$

2.1.3 注意力机制

注意力机制（Attention Mechanism），是将模型注意力集中在局部特征信息上的机制，受益于人类视觉的启发，人类在观察一张图片的时候，首先会快速浏览全局图像，然后自然地过滤掉背景信息，将注意力聚焦于图片中的实体对象，并对该实体中最具吸引力的局部部位重点关注。因此，参照人类的视觉注意力模式，基于深度学习的注意力机制将分为两部操作：首先，通过全局扫描找到局部有效特征信息；其次，增强有效特征信息，抑制无用冗余信息。

注意力机制的本质是一种权重的分布，其核心目标，是从大量数据信息中选取对当前任务更为关键的信息。通过将更多的注意力资源投入所需关注目标的细节特征方式，能够极大地提高信息处理的效率和准确性。

而注意力机制的实现通常依附于编码器-解码器（Encoder-Decoder）框架。在机器视觉中，该框架使用掩码（mask）来形成注意力机制^[57]，基于掩码实现权重更新，将图像数据中的关键局部特征表示出来，并加以学习训练，最终使

深度神经网络能够学习到每一数据需着重关注的局部特征。基于这种思想，注意力机制根据不同区域的注意力权重分配方式，可将其概括为柔性注意力和硬性注意力，具体的注意力机制分类如图 2-4 所示。

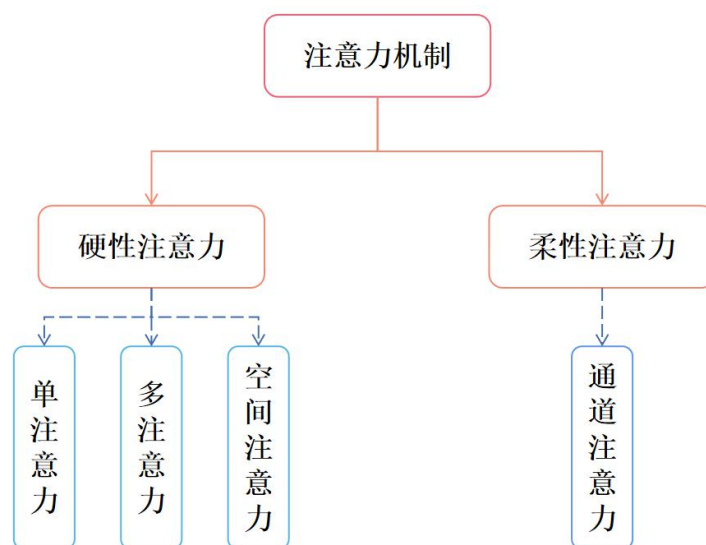


图 2-4 注意力机制分类结构

Fig 2-4. Classification structure of attention mechanism

2.2 人的介入策略方法

2.2.1 人机融合的重要性

将人重新请回控制的闭环本质上源于以深度学习为基础的 AI 算法缺陷问题尚未解决^[12]。事实上，一方面，AI 驱动的机器智能为各个领域的自动化应用带来了巨大的想象空间，越来越多原先并不属于经典自动化范畴的领域也开始成为自动化的目标对象（从工厂流程自动化这一经典领域到无人驾驶这一全新领域）；另一方面，虽然 AI 在很多领域运作良好，但众多的自动化应用对系统稳定可靠运行的现实要求是现有 AI 技术难以保证的。

利用深度学习在人脸识别领域的进展开发出刷脸支付技术是 AI 的一个成功范例，但在可预见的未来，我们仍难以把乘客的性命交给可以识别交通标志的无人驾驶系统，尽管识别有限种类的交通标志在技术上并不比人脸识别更难，但在车辆行驶这样的语境下任何的识别错误（包括 AI 固有的和对它的攻击造成的，比如篡改交通标志）都可能造成难以挽回的巨大损失，因而是难以容许的。这引发的悖论是，只要 AI 技术继续在自动化领域中进一步发展，而强人工智能未能出现，那么人的参与就将变得越来越重要和必不可少。

这样，AI 驱动的机器智能在自动化领域的进程又将人重新引入系统闭环，造成了人与机器关系的根本性变革。过去研究者致力于用机器代替人力，但现

在则必须考虑人与机器在智能层面的共融共存，可又不是基于传统的系统结构范式，即在人的指令下实现机器的自动运行。

从图 2-5 可以看出，近些年研究学者对 HITL 的研究越来越多，这也反映了人在环范式对系统建模的重要性。目前基于人在环的思想已经渐渐渗透到了很多应用研究领域，如安全系统和搜索引擎。由于应用场景的广泛性，不同的人环系统在交互界面、算法和人的角色上存在明显的差距。但设计这些系统的初衷都是希望通过引入少量的人工信息提升整个系统算法的可靠性。

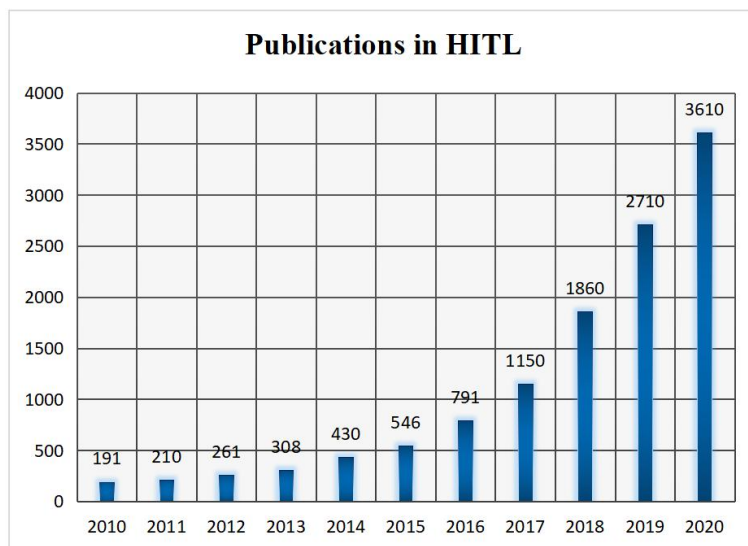


图 2-5 近十年关于“human-in-the-loop”研究报告数量

Fig 2-5. Number of research reports on "human-in-the-loop" in recent ten years

2.2.2 人机混合智能系统

结合人机融合的决策思想，考虑设计一种能面向复杂的自动化应用系统，将其称之为“人机混合智能系统”。它是一类特殊的人机系统，其中机器是具有 AI 赋予的智能和自主性，且机器对象一般为动态自动化系统。图 2-6 中展现的是人机混合智能系统的一个总体结构图，人和 AI 系统都需要根据周围环境情况给出自己的决策判断，系统将其进行融合，以最大化呈现较人或机器单独决策更优的判断，并传给机器让其自动执行。本文基于该系统设计，希望人的工作主要负责最终目标的选择、设定和不确定情况下的临时决断等，以实现人机协同控制。

该文^[12]将人机混合智能系统里的控制策略大致概括为介入控制和共享控制两种。具体定义如下：

(1) 介入控制：人和机器在感知-行动的环路中通过一方干预另一方的方式完成某一动态任务，这一动态任务是人或机器在理想环境下都可以独立完成的。

(2) 共享控制：人和机器在感知-行动的环路中以平等协作的方式完成某一动态任务，这一动态任务是人或机器在理想环境下都可以独立完成的。

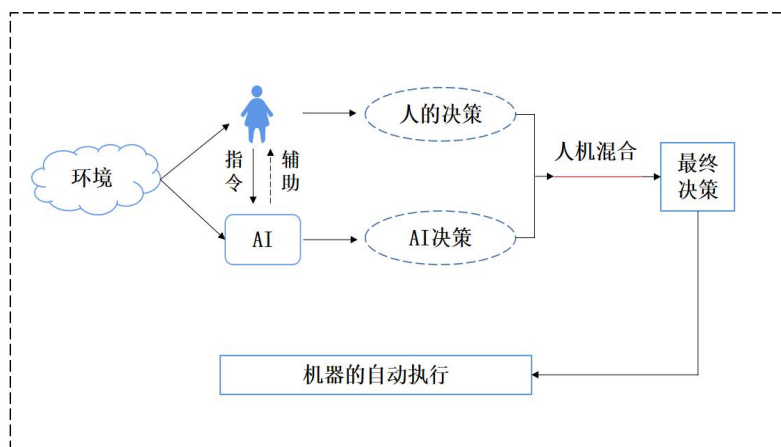


图 2-6 人机混合智能系统结构图

Fig 2-6. Structure diagram of human-machine hybrid intelligent system

根据上述描述可以看出，“动态任务是人或机器在理想环境下都可以独立完成的”是人机混合智能系统中的介入控制和共享控制的一个本质特征。一方面，这一特性意味着无论是人还是机器，在整个系统中所起到的作用都是另一方无法单独取代的，否则便不存在“人机混合”了；另一方面，人和机器在整个系统所能完成的任务情况，极大的与系统所处的环境相关：即在理想环境下，二者都能够独立完成任务，而在非理想情况下，人和机器各有优势，因此需要二者的协同合作。

其次，人机混合智能系统所研究的“动态任务”，其实更多的是对环境的考虑。任务的动态性意味着人机系统将长期暴露在外部环境中。在这一过程中，很难保证人机系统的环境始终处在理想状态，而在非理想状态下，人和机器在完成这一动态任务上各有优势，可促进人和机器之间的相互协作。

再者，针对人机混合智能系统基于动态任务的“感知-行动的环路”设定，其实考量的是感知、决策、行动等不同能力，而人和机器在这些不同能力上各有优势，这成为了考虑人机混合的一项重要原因。

既然在人机混合智能系统中加入了人的决策因素，那么就需要对人在决策过程的地位和作用做明确的分析和说明。一般地，人在系统环路中主要有两大作用。其一，人的引入更多的是处理算法出现错误决策的情形，通过修正、提供反馈、输入更多知识来弥补算法自身的缺陷。其二，希望人的引入能够检测到系统无法直接观察到的事件，或拥有系统内置事件之外的知识。相比前者，该环路人所参与的决策较少，主要通过定期监控机器和环境之间的相互作用，并且仅在认为必要时进行干预。这里，人是否介入干预的一个重要评判标准，就是对机器的信任程度，也就是说，如果人对机器产生正确决策的置信度较低，

则很大概率会进行干预；相反，置信度越高，人越有可能允许机器自主执行决策。

2.2.3 引入人的介入控制策略

其实，在特定的人机混合智能系统中是采取介入控制还是共享控制方式，其决定因素往往并非出于性能考虑，而是由系统的限制性因素所决定的，很多系统可能并不具备进行共享控制的条件，因此本文的重心将放在介入控制研究上。

依据系统的介入控制方式可以分为机器介入和人的介入。机器介入策略是指系统在正常运行时由人的智能自主和机器的自动化共同起作用，而在特定条件下允许机器的智能强制剥夺人的自主性，以避免人在此时的自主控制所导致的严重后果。而人的介入策略是指系统在正常运行时由机器的智能自主和机器的自动化共同起作用，而在特定条件下允许人的智能强制剥夺机器的自主性，以避免机器在此时的自主控制可能导致的严重后果。

由于本文针对的是目前基于深度学习存在的算法缺陷，导致现阶段在一些重要应用领域单纯依靠机器算法无法实现高精度的应用要求，而引入人的介入策略是弥补这一缺陷的可行途径。因此，研究关键是希望能充分利用人所擅长的能力，补足深度学习算法所不擅长的部分，从而提升最终模型算法的精度。可以理解，将人的因素引入到模型算法中，在技术上首要和关键的问题是对机器算法何时犯错给出可信的评估，解决了这一点，系统的设计将会变的简单。但上述关键要求也正是引发方法上困难的地方。因为深度学习本身的缺乏解释性，对其预测结果是否可靠的评估在其本身技术框架内是缺失的。

其次，人的介入方式也要有所考量。为了尽可能利用少量的人力成本，充分发挥人类智能的最大潜力，实现深度学习算法精度的突破性提升，需要针对不同的应用场景的性质、条件、所需成本等多方面进行衡量。就从人类经验角度来看，可大致分为两种。其一，若所涉及的群体在研究领域有较丰富的经验，则考虑在机器智能出现问题时，决策权将直接交由人来控制输出，这是对系统决策可靠性的最好表现。其二，如果是面向普通用户群体，不具备该领域的广泛知识，则考虑利用人在某方面的优异智能辅助机器决策，传输正确的信息引导和修正系统的决策输出。这两种方式也将在本文后续的实例研究中加以考虑和验证。

2.2.4 人的介入控制因素

设计人机混合智能系统时，针对人的介入控制需要考虑几点要素。首先就设计目标而言，设计的总体思想是优化人机混合智能系统的共同目标对象，该目标对象可从系统性能、安全、稳定性以及执行成本等多方面因素去考量。其次，在介入决策和执行的过程中需要成本，因为其过程可能会引发系统运行模态的切变，从而造成系统失衡，且实现介入执行有不可避免的延时问题。另一

方面，为了决定何时介入还需要不得不对另一方及人机系统整体状态进行监控，这相比纯自动化系统执行增加了额外成本。而针对介入时机问题，可依据系统智能程度或对人机赋权考虑在中间某点介入，或者以优化系统性能为标准决定介入时机。

但很显然，以上因素中最难掌控的就是介入时机问题。由前文 2.2.3 节可知，由于对机器决策存在的不确定性评估，难以把握时机的关键点。不同领域、不同级别的智能机器，引发人的介入时机也会有所不同。

研究之所以介入人，是考虑到某些应用场景下人的任务处理优势要比机器好，这也是提出人的介入方法有效的一个本质原因，因此认为在特定场景下人的决策介入是可信且可靠的。那么研究希望，人能够在必要的地方充分发挥自身的智能优势以保证机器智能的顺利进行。而必要的地方也就是人的介入时机点，比如当机器之间发生矛盾时，机器输出结果的置信度较低时，存在突发情况时等，人类都可以介入。并且，实现有机器报警、求助等情况引发的人类介入比人类主动介入的效果要更好。如 Fridman 等人^[58]在处理图像分类任务上，通过比较引入的两个 AI 系统对相同的输入图片给出的预测结果，在决策分类出现矛盾的地方作为人介入的关键节点，也就是认为此时的 AI 系统在分类任务上出现了犯错的可能，人的介入及时减少了系统的决策失误。经过实验验证，该方法将分类错误率由原来的 8.0%降低到 2.8%，确实有效提高了图像分类任务的准确度。同时，他还将此想法运用到了自动驾驶系统设计上，提出了一种人机共驾的实现方案，实验结果显示，在所有人工标注的危险驾驶行为中，有 90.4% 被系统检测出来需要人的介入控制，从而避免了自动驾驶系统可能带来的巨大危险。另外还有 Muthuraman 等人^[59]也是通过观察多个模型输出存在不一致标签信息来检测 CoNLL-2003 语料库中命名分类不正确的实体，经实验验证有 27%的实体被检测出，占整个语料库中总的不正确标签的 64.2%，该方法为机器翻译、信息提取、句法分析等自然语言处理所需的高质量训练数据提供了帮助。

其实，针对机器算法何时犯错，人利用何种方式介入决策这一技术问题，涉及到用户使用和 AI 的结合。至少现阶段，AI 是作为提高人类效率的工具，因此在人的介入设计中应该做到不需要人的时刻紧盯以及频繁的人类反馈，但在有意外情况的时候，应及时提醒反馈给人类无缝接手。因此，目前最好的方式就是在构建好的 AI 系统后建立一个仲裁模块。该模块能够作为人与机器自主决策的切换机制，当所设立的系统切换条件满足时，此时即可作为人的介入时机。一个好的仲裁模块将能够保证机器算法的失误可以被有效捕捉，从而在整个系统中，通过引入少量的人力成本，实现对机器算法性能的有效提升。

2.3 本章小结

本章节首先介绍了有关深度学习相关技术知识，包括对 ResNet 网络、批归一化技术和注意力机制的原理和作用进行细致阐述，为后续的研究和实验奠定了理论基础。其次，通过对人机融合思想以及人机混合智能系统的介绍，引出了本文的核心方法，即通过人的介入策略以最少的成本代价提升模型算法的精度。基于此，本文后续章节将秉承这一理念，并通过两个实例应用验证本文方法的有效性。

第三章 基于人的分歧介入的珍珠自动分拣应用

针对珍珠自动分拣应用，把优等珍珠分成次等品会导致严重利益损失，把有瑕疵珍珠分成上等品会造成产品质量争议和企业信誉受损。因此，为衡量珍珠分拣精度的现实需求和提升精度在现有技术框架下的限制因素，本章有针对性地提出一种通过人的分歧介入提升珍珠自动分拣精度的方法。这是一种在技术上要求较低、在实际中却可以有较大成效的实用方法。

3.1 珍珠分拣应用背景

珍珠是一些瓣鳃纲软体动物（主要是牡蛎）的产物，是重要和常见的珠宝类型。现有珍珠大多通过淡水养殖牡蛎进行生产，我国的产量占世界总产量的80%。珍珠的价值与其外观品相密切相关，光泽、形状、表皮质量、颜色、大小和有无瑕疵等所造成的珍珠优等品和次等品之间价格能差几百上千倍。因而，对于珍珠生产商而言，按照不同的等级将珍珠进行分拣就成为一个必要和重要的步骤。

现有珍珠分拣主要通过人工进行。经验丰富的从业人员通过肉眼观察珍珠的颜色、光泽、形状和质地等多种外观特征，然后依照某种事先定好的等级规则进行分类。容易认识到，对大量外观差距可能并不太大的珍珠进行观察分拣是一项重复单调的劳动；而随着时间的增加，人的疲劳也会大大影响分拣的准确度^[60]。然而，将高等级珍珠误分为低等级固然会错误的降低售卖价值，反过来将低等级珍珠误分为高等级也会影响产品质量和信誉。这样，为了分拣的高效和稳定，就有必要引入自动化的分拣方法。

由于珍珠分拣主要基于珍珠外观进行，自动化的分拣方法因而大多基于机器视觉的方法和技术。例如，论文^[61]利用色调、饱和度和亮度颜色模型转换技术，以及模糊模式识别方法进行珍珠判别归类，论文^[62]利用色调、饱和度和明度颜色模型来提取珍珠颜色信息，论文^[63]提出了基于傅里叶变换的模糊分类方法，论文^[64]通过 Zernike 矩描述珍珠形状特征从而实现分类。为观察珍珠这种细粒度级别的多个视图的多个特征^[65]，在上述使用较少特征的分类方法基础上，论文^[66]使用单目多视角摄像装置，直接获取 5 个不同视角的珍珠表面图像，利用更全面的珍珠检测视角提升了珍珠识别精度。传统计算机视觉方法中珍珠特征的手工标注依赖于专家知识，成本相对较高。近年来，深度学习方法尤其是 CNN 在图像分类领域取得了巨大成功，它可以自动生成有用的特征，从而节省

大量手工标注工作。基于该理论进展，论文^[67]提出了一种多尺度卷积神经网络（multi-scale CNN, MS-CNN）方法，通过充分利用不同观测角度的信息提升自动分类的效果；论文^[68]则提出了一种多视角对抗生成网络（Multiview generative adversarial network, MV-GAN）方法，通过自动扩展标记的多视图珍珠图像实现更可靠的珍珠分类效果。

注意到，尽管上述的各种珍珠分拣的 AI 系统已经在实际中得到了初步应用，但所达到的分拣准确度仍有很大提升空间。例如，2018 年所提出的 MS-CNN 方法只能达到 92.57%^[67]的分拣准确度。而可以理解的是，在此基础上的准确度的进一步提升存在很大的困难。原因一方面是技术上的，这一技术上的限制只有在现有深度学习算法有了本质的改进才有可能得以消除。另一方面的原因是成本上的，如前所述，大量准确标注的珍珠数据集难以获取。上述分析意味着如果需要进一步提升珍珠分拣的准确度，很有可能需要跳出现有的单纯的深度学习技术框架。

另一方面，前面已经指出，对珍珠生产企业而言，1%的分拣准确度提升也有很大的经济价值。这主要是因为珍珠单价昂贵，不同分级之间价格甚至相差几百倍，把优等珍珠分成次等品会导致利益损失严重，而把有瑕疵珍珠分成上等品会造成产品质量争议和企业信誉受损。这就对珍珠分拣准确度的提升提出了现实的需求。

考虑上述两方面的因素，即对珍珠分拣精度的现实需求和提升精度在现有技术框架下的限制，在本章提出一种通过人的分歧介入来提升珍珠分拣精度的方法。该方法的基本思想是找出基于深度学习的自动分拣方法中分错的部分，将这一部分交由经验丰富的人进行进一步的检查判定，通过人机合作达到更高的分拣精度。容易想到，该方法技术上的难点正在于如何准确的识别自动分拣方法的失误，为此使用了一种多个冗余算法分歧的方法，通过引入冗余的自动分拣算法，将多个算法产生判定分歧的珍珠认定为分拣失误，通过在论文^[68]中采集的珍珠数据集上进行实验，证明了方法的有效性。

3.2 方法模块设计

本章所提出的面向珍珠分拣的人的分歧介入方法整体框架如图 3-1 中的下框图所示。与上框图所表示的一般 AI 算法不同，在人的分歧介入方法中，主算法之外另外构建了一个独立的次算法，两个算法独立产生预测输出，输出结果进而交由仲裁模块进行分歧仲裁：如果分歧产生，则认为 AI 系统分类可能有误，此时将交给给人做最终的决策；如果无分歧，则保留主算法的预测结果。

这一方法的思想来源于人机混合智能系统共享控制中的“仲裁”概念：在人机系统中人和机器都可以对系统的运行做出独立的决策，而实际采用的决策通

过对这两个独立决策的“仲裁”得来。类似的，本章针对 AI 系统提供额外的辅助 AI 算法，从两个 AI 算法预测上的分歧来判断 AI 系统的预测是否出错。如果两个独立的 AI 算法都给出了同样结果，有信心认为 AI 系统当前预测是正确的；而如果两个 AI 算法产生了分歧，则有理由认为 AI 系统当前预测的准确性需要慎重评估^[69]。

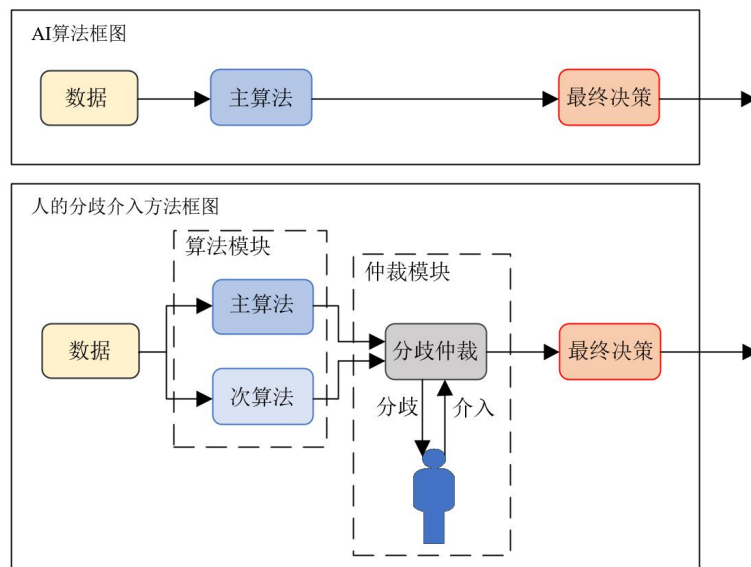


图 3-1 人的分歧介入框架图

Fig 3-1. Frame diagram of human disagreement intervention

下面详细介绍算法模块和仲裁模块的设计。

3.2.1 算法模块设计

算法模块的功能是通过所选择的两种算法产生的预测的分歧来判断算法预测是否有误，因而两种算法选择的核心原则即是，一方面，两种算法本身都应该具有较高的预测准确性，因为否则两种算法的输出判定可能会产生大量无意义的分歧；另一方面，两种算法应保证尽可能大的独立性，否则就难以产生所需要的分歧。另外，所提出的算法仍需要保持某一选择的算法为主，另一算法为次，因为依照该算法结构，系统输出会保持主算法的输出，除非因为分歧引发人的介入，此时的系统输出就成为人的决策。

为了符合上述的选取原则，在实际问题中主次算法的选择可按如下流程进行：首先选取针对当前问题性能最好的算法作为主算法，其次，在所有剩余待选算法中选择与主算法尽可能独立但同时又保证具有较高的自身准确度的算法作为次算法。

采用上述的算法选取流程，关键在于次算法选取中独立性和准确性的折中。如果次算法与主算法的独立性较差，那么主算法预测有误的情形次算法也有很大可能出现同样错误，因而预测错误无法产生分歧，影响所提出的人的分歧介入框架的准确性；相反，如果次算法准确性较差，那么会有很多预测因为次算

法自身的预测失误而产生分歧，无意义的增加了人的介入成本。从上述分析可以看出，在实际的次算法选择中，应首先确保次算法自身的准确性在某一阈值之上，然后再考虑与主算法的独立性。

进一步的，对次算法独立性的确保，可以从训练数据集、训练方式、模型本身等方面的不同来考虑。针对具体问题，需要分析从所有可选算法中进行尽可能最优的选择。

3.2.2 仲裁模块设计

针对珍珠分拣应用在设计仲裁模块时，一是要保证机器算法的失误能够有效检测，二是尽可能通过引入少量的人力成本实现对算法性能的提升。但基于算法模块的设计，考虑通过两个 AI 系统分歧程度驱动人的介入的判断是否准确，在大多分类任务中往往不是问题，只要两个算法的预测类别不同一般便可认为需要人的介入，但在非分类任务中，图 3-1 所示的方法框架就需要通过分歧程度来判断是否需要人的介入，这一判断的准确与否将极大影响方法的准确度和效率。针对这一问题，一般方法是使用某一阈值来判断分歧的程度是否需要人的介入。这样，自然产生的技术问题就是阈值如何选取问题，过大的阈值会将本应交给人的介入的分歧视而不见，影响系统的准确性；过小的阈值则会将无必要的分歧也交由人来处理，增加系统的运行成本。因此本文在实验中将设立分歧准确指数 α 和额外成本指数 β 作为系统分歧的评价指标，通过不断调整其参数来达到最优平衡状态。

具体的分歧仲裁框图如图 3-2 所示。

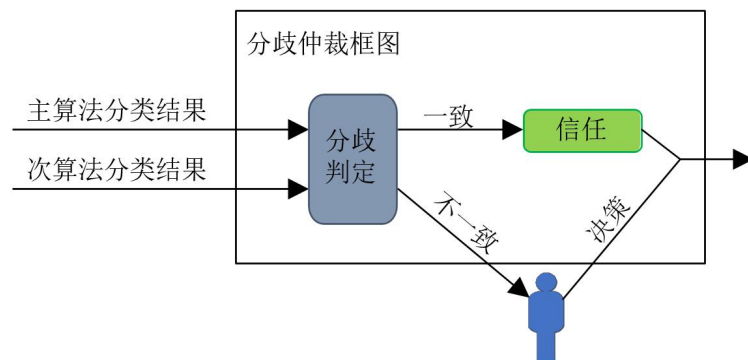


图 3-2 分歧仲裁框图

Fig 3-2. Frame diagram of disagree arbitration

在实验中，由两种 AI 算法对同一珍珠进行分类，如果两个独立的 AI 算法都给出了相同的分类结果，则有信心认为 AI 系统当前的分类结果是正确的；如果两个 AI 算法产生分歧给出了不同的分类结果，则有理由认为 AI 系统当前的分类结果需要进行慎重评估，进而将该珍珠交由有经验的珍珠分拣人员进行检查判定，并假定凡是交由人做进一步判定的珍珠最终的分类结果都是正确的。

因为考虑到训练有素的人在非疲劳情况下，对珍珠种类判断的准确度近乎 100%，这符合实际情况，也是所提出人的介入方法有效的一个本质原因。

3.3 实验所用数据集和处理方法

3.3.1 珍珠分拣数据集

本文在 Xuan 等人^[68]提出的数据集上对所提出的方法进行评估。该珍珠数据集利用单目多视角摄像装置拍摄来自浙江诸暨的大量珍珠的主、后、左、右和上 5 个不同视角的表面图像做了预处理，并让有经验的专业人员对图像进行了标记，如图 3-3 所示。数据集包含有 10500 颗珍珠拍摄而成的 52500 张带有标签信息的图像，52500 张图像按 6: 2: 2 的结构分成训练集、验证集和测试集。从而，本数据集共收录训练集图像 31500 张、验证集图像 10500 张和测试集图像 10500 张。

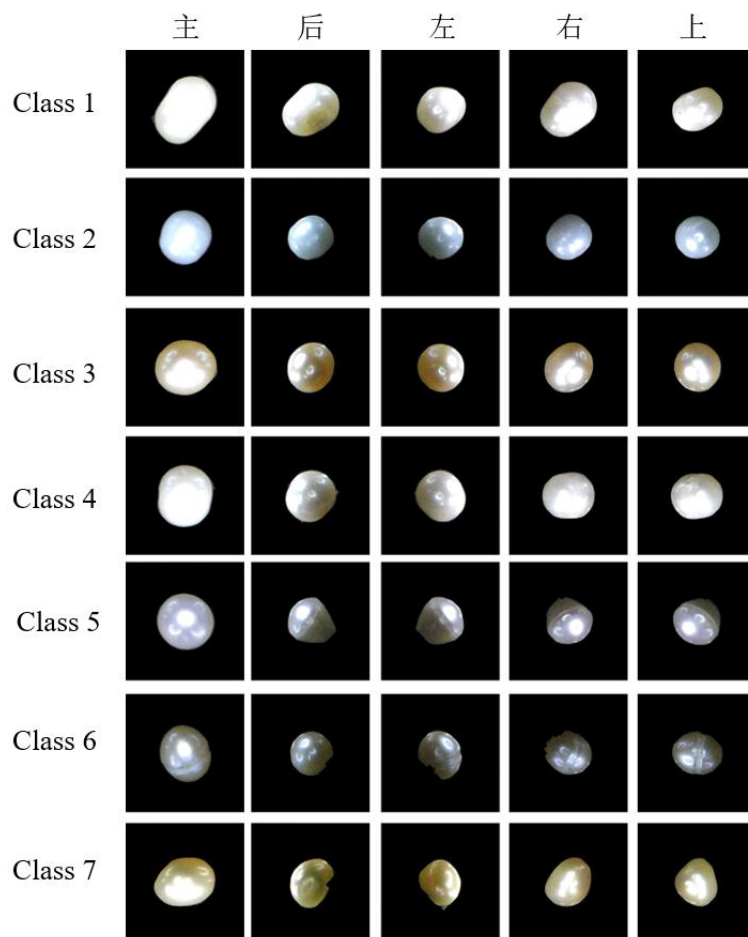


图 3-3 珍珠七分类视图

Figure 3-3. Pearl seven category view

珍珠按粗糙规则分类可大致分为两类：1) 一类为价值较低的扁平型或深瑕疵珍珠；2) 另一类为瑕疵程度较小的珍珠。前者可进一步细分为三类：1) 具

有多个平面的珍珠；2) 形状对称的珍珠；3) 其他珍珠。后者也可进一步细分为四类：1) 短半径与长半径之比大约大于 0.7 的珍珠；2) 颜色浅淡的珍珠；3) 隐含斑点的珍珠；4) 其余珍珠。

将上述珍珠类别分别记为 Class1-7，并给每张图片制定好标签，在模型训练阶段采用训练集和验证集，在测试阶段采用测试集评估训练模型的能力。

3.3.2 珍珠分拣数据集的处理

本文采用 Ubuntu16.04 服务器训练珍珠分拣模型，算法使用 Pytorch 实现，具体设置如下。训练模型时设置训练最大次数为 10 个 Epoch，将训练和测试时的单次处理样本数 Batch 设置为 32，学习率为 0.0001，采用 Adam 优化器，并将 step-size 设置为 5，gamma 设置为 0.1，学习率将在训练过程中经过 5 次变化逐渐变成 0.1。在训练阶段，本文会对图片进行预处理，包括随机二维裁剪、给定概率水平翻转等。将在每一个训练 Epoch 后进行一次验证，最后选取最优模型用于仲裁系统。

3.3.3 评价指标

在实验验证阶段，使用分歧准确指数 α 和额外成本指数 β 作为系统分歧的评价指标。分歧准确指数 α 指的是主算法分类错误后被纠正的图片数目占该模型分类错误的图片总数目的比值，用于评价该方法找出分类错误的能力。额外成本指数 β 指的是主算法分类正确但交于人进行判断的图片数目占最优模型分类正确的图片总数目的比值，用于表示该方法消耗的无效人力成本^[60]。

分歧准确指数越高越好，而额外成本指数越低越好，当两者的数值均为最优值时，说明分歧部分均为主算法出错的部分，是最优结果，但实际中几乎不可能达到。对于主算法分类出错交于人进行分类的那部分人力成本是算法的必要成本，因此在成本估算上更有意义的是所需的额外成本。

分歧准确指数 α 计算公式如下：

$$\alpha = \frac{F}{E} \times 100\% \quad (3-1)$$

额外成本指数 β 计算公式如下：

$$\beta = \frac{P}{N - E} \times 100\% \quad (3-2)$$

其中 N 为测试集数目， E 为主算法分类错误的图片总数目， F 为主算法分类错误后被纠正的图片数目， P 为主算法分类正确但交于人进行判断的图片数目。通常评价指标还包括准确度。

3.4 基于二算法分歧的实验与分析

为检验方法有效性，本节基于二算法分歧的珍珠分拣进行验证，并通过分歧准确指数和额外成本指数这两个性能指标的显示，来评估系统的整体性能。

3.4.1 实验设置

本文选取了目前比较主流先进的几个算法，包括 ResNet50、SE-ResNet50、Vgg16 网络模型进行对比评估。

如表 3-1、3-2 和 3-3 所示，本文分别对 ResNet50、SE-ResNet50、Vgg16 算法模型进行训练，再两两搭配成三组，对比准确率结果。为对比本文提出的方法而选用的集成学习法（Ensemble Method），将若干个模型的输出结果概率进行平均得到最终的预测输出，其最终预测为：

$$H(x) = \sum_{i=1}^T w_i h_i(x) \quad (3-3)$$

其中 $h_i(x)$ 是第 i 个模型的概率输出结果， T 是模型的个数， w_i 是单个模型 h_i 的权重，且

$$w_i \geq 0, \sum_{i=1}^T w_i = 1 \quad (3-4)$$

为了简便起见，这里取 $w_i = \frac{1}{T}$ 。

此外，为了验证两个算法的分歧确实找出了 AI 系统分类错误的部分，还提出了一种基线方法，随机选取测试集中人的分歧介入方法产生的等量珍珠交于人进行决策判断，并称之为随机监督方法（Random Supervision Method），并在实验过程中认为交于人的这部分最终分类结果都是正确的。

表 3-1 ResNet50 + SE-ResNet50 算法分歧准确率

Table 3-1. Disagreement accuracy of ResNet50 + SE-ResNet50 algorithms

模型/精度	单个模型(%)	集成方法(%)	随机监督方法(%)	本文方法(%)
ResNet50	91.84	93.36	92.87	96.27
SE-ResNet50	92.25			

表 3-2 ResNet50 + Vgg16 算法分歧准确率

Table 3-2. Disagreement accuracy of ResNet50 + Vgg16 algorithms

模型/精度	单个模型(%)	集成方法(%)	随机监督方法(%)	本文方法(%)
ResNet50	91.84	92.12	92.73	96.15
Vgg16	91.29			

表 3-3 SE-ResNet50 + Vgg16 算法分歧准确率

Table 3-3. Disagreement accuracy of SE-ResNet50 + Vgg16 algorithms

模型/精度	单个模型(%)	集成方法(%)	随机监督方法(%)	本文方法(%)
SE-ResNet50	92.25	93.23	93.03	96.16
Vgg16	91.29			

由以上三个表实验结果可知, 本文提出的方法是有优势的, 比之单独的模型算法的准确率^[70]提升了将近 4%, 且较两个算法进行集成后的准确率也有明显提升。即使因为人的参与最终分类结果准确率必定有所提升, 但相较于随机监督方法的实验结果可以发现准确率也有明显提升, 这表示了本文提出的方法确实可以在一定程度上提升珍珠分拣的精度。

3.4.2 评价指标分析

为进一步说明本文方法的有效性, 本文分别从分歧准确指数和额外成本指数两个性能指标进行分析, 来评估系统的整体性能。二算法分歧性能结果如表 3-4 所示。

由表 3-4 可以看出, 本文分歧准确指数在 51%左右, 相较于随机监督方法的性能结果有明显提升, 这表示了本文提出的方法确实更有效地找到 AI 系统分类出错的部分, 此外额外成本指数也有相对减少, 这表示本文提出的人的分歧介入方法用更少的人力来提升珍珠分拣的精度。

表 3-4 二算法分歧性能结果

Table 3-4. Disagreement performance results of two algorithms

模型/性能	分歧准确指数 α (%)		额外成本指数 β (%)	
	随机监督方法	本文方法	随机监督方法	本文方法
ResNet50+ SE-ResNet50	7.6	51.8	7.8	4.1
ResNet50+ Vgg16	11.0	52.9	8.3	4.6
SE-ResNet50 +Vgg16	10.1	50.5	7.9	4.5

另外, 为了能更清晰地展示本实验的实验效果, 在以 ResNet50 和 SE-ResNet50 两种算法分歧的情形为例, 挑选了几张图片进行分析, 这两种算法中 SE-ResNet50 算法更优, 如图 3-4 所示。图 3-4 中分别展示了四种情形。1) 一类分歧: 两种算法分类均错误但仍分歧的情形, 这也是研究最期望的结果; 2) 二类分歧: 主算法分类错误而次算法分类正确导致分歧的情形, 这 and 一类分歧

共同决定了分歧准确指数的大小；3）三类分歧：主算法分类正确而次算法分类错误导致分歧的情形，这是额外成本指数的来源；4）不分歧：两种算法分类均错误但不分歧的情形，这是导致研究最终准确率提升不上去的原因。其中算法对错误分类结果的概率并不是很高，这也是未来研究可以切入的点，利用这一优势在减少人力成本的基础上提升珍珠分拣精度。

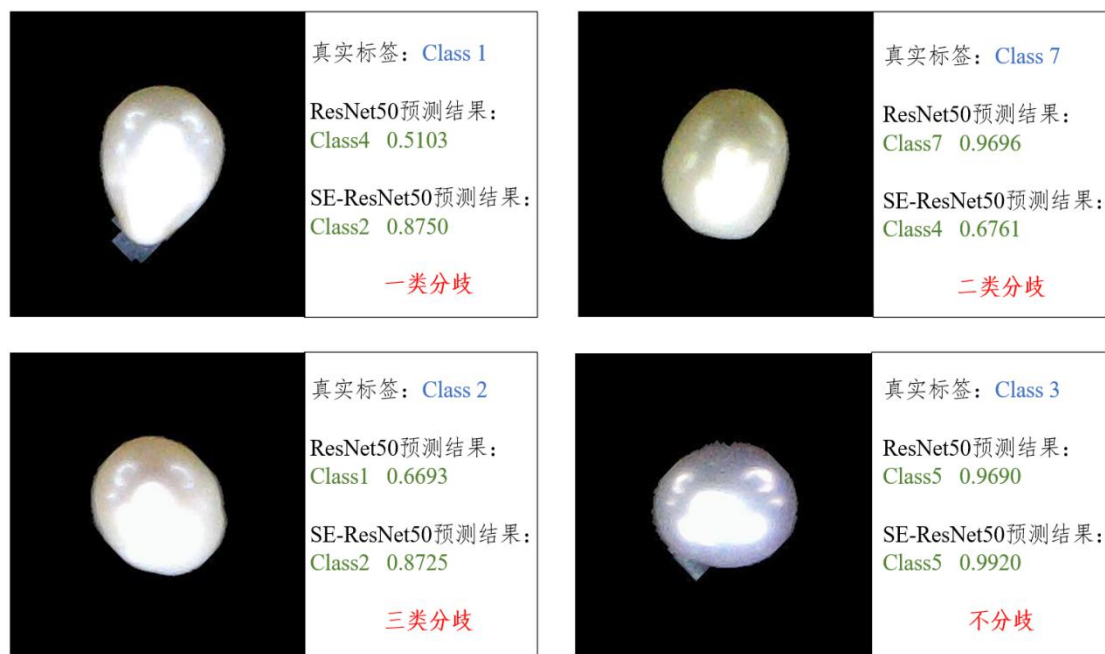


图 3-4 二算法分歧珍珠分类效果图

Fig 3-4. Disagreement of two algorithms for pearl classification

3.5 基于多算法分歧的改进优化实验与分析

本节在二算法的基础上进行改进，实现基于三算法甚至多算法的分歧介入。算法设计的目的主要是考虑提升分歧准确指数和降低额外成本指数。

3.5.1 提升分歧准确指数

在面向对成本相对不敏感但对分拣结果准确率有更高要求的应用场景，多算法的人的分歧介入框图如图 3-5 所示，在主算法的基础上设计多个次算法，其性能结果如表 3-5 所示。

图 3-6 展示了三算法分歧的珍珠分类效果图，其中 SE-ResNet50 为主算法。如图 3-6 所示，依次展示四种情形，1）一类分歧：三种算法分类均错误但仍分歧的情形，是此次研究最期望的结果；2）二类分歧：一种算法分类正确其余两种算法分类错误也导致分歧的情形；3）三类分歧：两种算法分类正确而另一算法分类错误导致分歧的情形，相对于二算法这导致人工总工作量增加；4）不分歧：三种算法分类均错误但不分歧的情形，这是研究中最不愿看到的。

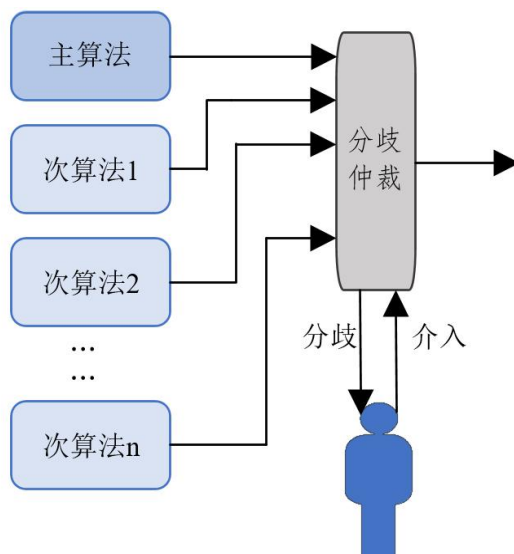


图 3-5 基于多算法的人的分歧介入框图

Fig 3-5. Frame diagram of human disagreement intervention based on multi-algorithms

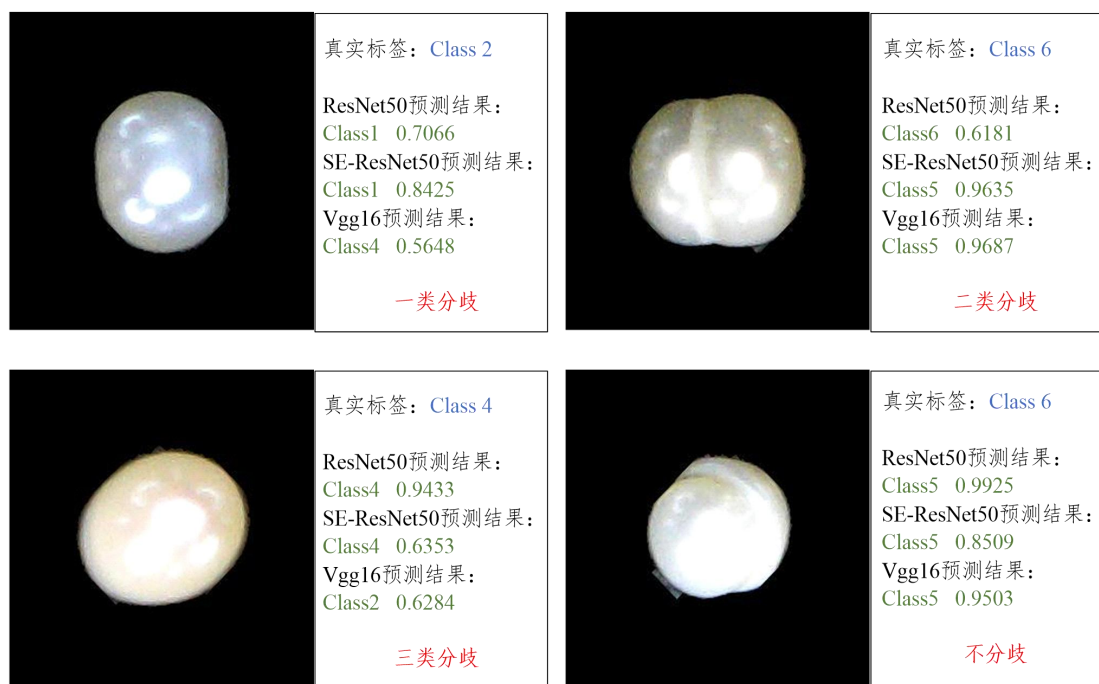


图 3-6 三算法分歧珍珠分类效果图

Fig 3-6. Disagreement of three algorithms for pearl classification

从表 3-5 中可以看出使用多个算法确实可以提升系统最终分类结果的准确性，分歧准确指数也有明显的升高。但同时随着算法个数的增加，额外成本也在逐渐增加，而准确率的提升幅度却明显减少，有一定的限度。因此所需的成本和准确率的提升之间需要一个衡量，在实际中要因事制宜。

表 3-5 多算法分歧性能结果

Table 3-5. Disagreement performance results of multi-algorithms

方法	$\alpha(\%)$	$\beta(\%)$	准确度(%)
单算法	--	--	92.25
二算法	52.9	4.6	96.25
三算法	67.8	7.2	97.50
四算法	74.1	9.4	97.99

3.5.2 降低额外成本指数

针对珍珠分拣准确性已满足要求，但需要降低人工成本的应用场景，可以在三算法的基础上加入一个集成的概念，将三算法集成为二算法，其他部分同二算法方法，如图 3-7 所示。

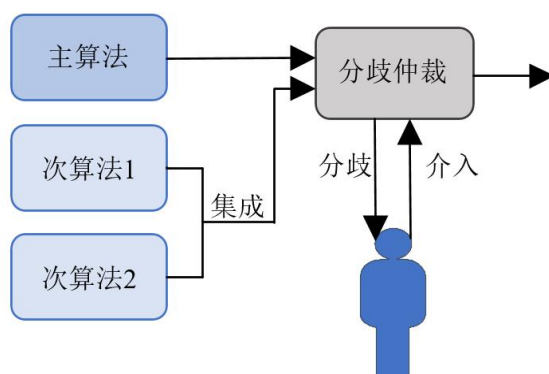


图 3-7 基于算法集成的人的分歧介入框图

Fig 3-7. Frame diagram of human disagreement intervention based on ensemble algorithms

选用 ResNet50、SE-ResNet50、Vgg16 算法，在表 3-1、3-2、3-3 的集成结果基础上进行实验结果分析，如表 3-6 所示。

表 3-6 三算法集成分歧性能结果

Table 3-6. Disagreement performance results of ensemble three algorithms

模型	$\alpha(\%)$	$\beta(\%)$	准确度(%)
ResNet50+ SE-ResNet50	54.1	2.7	96.24
ResNet50+ Vgg16	51.6	2.8	96.24
SE-ResNet50+Vgg16	56.6	2.5	96.22

从表 3-6 可以看出，与二算法的性能结果相比，提升相同的准确率所需要的人工监督成本有所减少，当准确率均达到 96%左右，分歧准确指数为 51%左

右时，额外成本指数却降低了约 2%，用更少的人力来提升珍珠分拣准确性。

3.6 本章小节

面向珍珠分拣应用场景，本章提出了一种通过人的分歧介入提升机器自动分拣准确性的方法。其中，设计了相应的算法模块和仲裁模块，以找出基于深度学习的自动分拣技术中分错的部分，并将这一部分交由专业人员进行进一步的分类，通过人机合作达到更高的分拣准确度。该方法技术上的难点在于如何准确的识别自动分拣方法的失误，为此使用了一种多个冗余算法分歧的方法，通过引入冗余的自动分拣算法，将多个算法产生判定分歧的珍珠认定为分拣失误。文章中定义了包括分歧准确指数和额外成本指数在内的性能评价指标，以维系在较少的人力成本下达到对机器算法可靠性的提升。在公开珍珠数据集上，研究提出的方法以 4.1% 的额外人工成本提升了近 4% 的珍珠分拣精度，验证了方法的有效性，也实现了该方法的设计初衷。

第四章 基于人的融合介入的野生鸟类细粒度图像识别

由于野生鸟类资源对整个社会具有极高的生态价值和经济价值，鸟类图像识别技术作为野生鸟类信息管理系统中的关键环节，应该具备精准、可靠的信息输出。然而，面对物种多样的鸟类群体，其呈现的细粒度图像特征存在高度的类间相似性及类内差异性，使得现有的自动识别系统还不能实现分类更优的性能。为了应对这一挑战，本章利用一种人的融合介入方式来增强鸟类细粒度图像识别精度，通过嵌入外部但易获取的信息进一步提升系统分类的准确性。

4.1 野生鸟类图像识别研究背景

野生鸟类作为生态系统组成中的一部分，在维持生态平衡中起着至关重要的作用。鸟类对生态环境的变化非常敏感，它们能在第一时间做出反应，因此科学家们把鸟类作为衡量一个地区生态环境是否合格的指标者。但随着国家经济的高速发展，生态资源却在一步步地遭受破坏，其中影响较大的就是鸟类物种的发展。据统计，目前全世界每种鸟类数量低于一万只的约有 1500 种，其中包括有近 400 种鸟类濒临灭绝，700 种鸟类也已非常罕见。对此，世界各国都非常重视鸟类的保护工作，自然保护区、湿地公园等地方管理部门、科学研究者、鸟类爱好者等都纷纷加入其中，有关鸟类识别与分布动态监测、信息管理系统的搭建，已成为各地区的常态化工作。

但目前国内外对野生鸟类的识别和管理主要依靠人工和经验积累，相对效率低下，且人工成本高昂。基于 AI 技术的大力发展，以及计算机视觉技术的提高，越来越多地区通过利用鸟类信息化数据库构建起野生鸟类图像自动识别系统，从而提高监测效率和识别准确度。但研究发现，鸟类物种识别与一般的图像识别任务不同，其属于细粒度图像领域范畴，内部存在着高度的类间相似性以及类内差异性问题^[71-73]，这对于计算机视觉算法提出了巨大挑战。

现阶段针对鸟类细粒度图像识别工作主要集中于如何定位最具区别的图像区域并依靠其提高算法捕获细微差异的能力展开研究，目前该研究方法大致分为两类：基于定位-分类方法和基于端到端视觉特征编码方法。基于定位-分类方法仿照人类思维方式，先通过快速扫描的方式，找到目标对象具有区分性的部位，然后对该部位进行仔细比对识别，其方法主要分成两个部分，即辨别性部位定位和定位部位的细粒度学习。对于辨别性部位定位技术，一类方法是利用鸟类关键部位的标注信息，比如 Mask-CNN^[74]，这种强监督方式的缺点是依赖标注数据，这在现实应用中会产生过高的标注成本，并且人工标注信息不一

定是最具辨别性区域，这完全依赖于标注者的认知水平；还有一种是基于弱监督，比如引入视觉注意力机制^[75]，但这类方法的缺点是，引入的辅助模块会带来更大的计算代价。其算法研究通过独立的检测鸟类的多个关键性区域，而忽略了物体的多个关键性区域的内在关联。第二种基于端到端的视觉特征编码方法，以增强图像的全局视觉信息，来直接学习到更具辨别力的全局表征。如利用卷积特征的高阶统计编码^[76,77]方式，增强 CNN 对视觉图像的中层特征学习，但相对统计特征维度过高，计算开销大，不利于向大规模数据扩展。

由于目前多数鸟类种群数量少、密度低，其数据存在的量少、特征细微等因素，导致对鸟类识别算法提出了更大挑战。但迫切的现实需求又使得鸟类信息管理与监测需对野生鸟类有较高的识别性能，以防止某种鸟类因环境、外来物种入侵或生态资源破坏等因素处于濒危状态。为此，本章提出了一种人的融合介入方法，以最小化的人工成本提升鸟类细粒度图像识别精度，实现目前机器视觉算法在细粒度鸟类图像识别难题上的突破。

4.2 系统设计

本节针对细粒度鸟类图像分类任务存在的高度相似性难点设计了一种多源信息融合的细粒度图像识别系统，其设计思想就是利用人的融合介入方式增强鸟类图像识别精度。

4.2.1 方法框架

整个系统设计框架如图 4-1 所示，主要分成三个部分，一个是针对计算机视觉识别鸟类图像的分类算法，一个是基于鸟类图谱的构建，描述类-属性之间的关联，还有一个是基于人在环的问答系统，主要用来提供人的视觉信息，以帮助计算机实现高精度的识别分类。

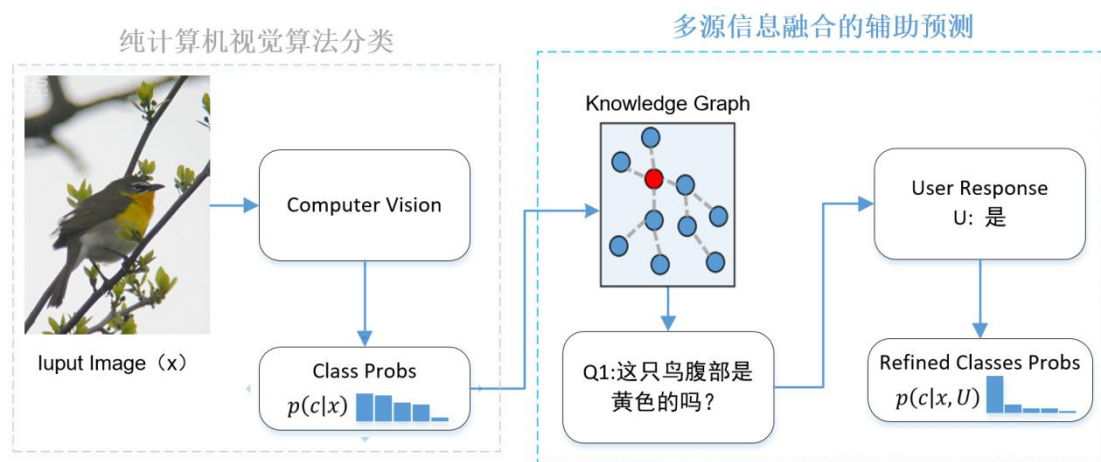


图 4-1 多源信息融合的细粒度图像识别系统框图

Fig 4-1. Block diagram of fine-grained image recognition system based on multi-source information fusion

下面将详细介绍设计系统的三个模块。

(1) Resnet50 网络

深度卷积神经网络广泛应用于细粒度图像任务中，常见的深度卷积神经网络包括 VGGNet、ResNet 和 AlexNet 等。在 2015 年的 ILSVRC 竞赛中，ResNet 网络在图像分类和物体识别方面取得了绝对优势，该模型的提出有效缓解了网络深度增加导致的训练误差增大问题，且收敛速度加快。目前 ResNet 有 18、34、50、101、152 等多种不同深度的网络结构层数。其中最常用的就是 ResNet50 网络，它兼顾了模型的准确性和运算量。图 4-2 展示的就是 ResNet50 网络结构，第一阶段主要是对输入图像进行卷积、正则化、激活函数、最大池化的计算，第二、三、四、五阶段都设有残差块，每个残差块中都有三层卷积，通过前五个阶段的卷积计算输出，并进行最大池化操作将其转化成一个特征向量，最后分类器可通过对这个特征向量的计算输出类别概率。

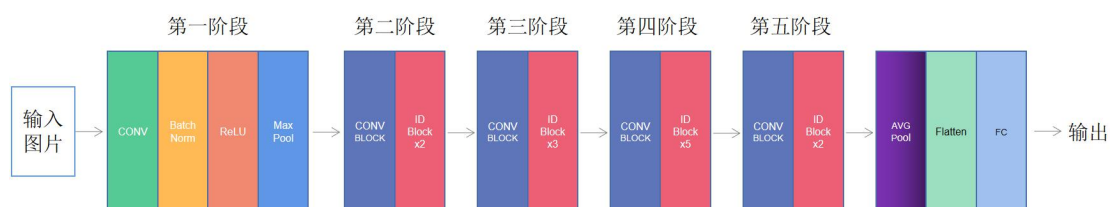


图 4-2 ResNet50 网络结构

Fig 4-2. ResNet50 network structure

本章将采用 ResNet50 网络作为鸟类细粒度图像识别的基准模型，根据传入的图像识别鸟的具体类别。这一阶段会采取一种端到端的方式对模型进行训练，系统所接受到的信息仅为图像及其标签，无其他人工标注信息。这一方式一方面可以减少多个子模块之间产生的累计误差，另一方面也降低了工程设计的复杂度，减少了模型训练所需的人工标注成本。

然而，该模型训练方式也存在一个弊端。原先通过设计多个子模块来完成模型对各个目标的训练，以达到对关键特征的学习，但端到端的模型训练方式可能会对细粒度图像来说产生大量非必要的计算，因为区分不同类别之间的信息往往反映在某些局部特征上，而这种方式可能会将许多无关信息也考虑在内，影响了模型的训练速度，也会对最终的预测结果产生一定的干扰。因此为解决上述所带来的训练缺陷，本文在 ResNet50 网络模型中，添加了注意力机制模块。该模块一方面可以让模型完成对前景对象的检测定位，排除掉背景噪声的干扰，另一方面，还可以让网络模型自主地发现该类图像中与其他类别图像有显著差异的区域，从而更细致地学习该区域具有辨别性的特征，以达到提高分类精度的目的^[57]。

如图 4-3 所示就是一种基于注意力机制的训练模式，通过对输入信息由粗粒度的鸟类原始图像，不断聚焦到更具辨别性的区域实现图像的分类任务。在

网络结构中添加视觉注意力机制，可以让深度学习模型获得更多关键部位的特征信息，从而减小类内差异的影响。图 4-4 是网络局部注意力模块图，实验选择在 ResNet50 网络的每个残差块之后都加入注意力模块，这样可根据不同特征图对识别任务的贡献程度进行赋权，将特征能力提取强的特征图赋予较大的权重，以快速地从大量信息中定位有价值的目标区域，最终提升算法的学习效率和准确性。

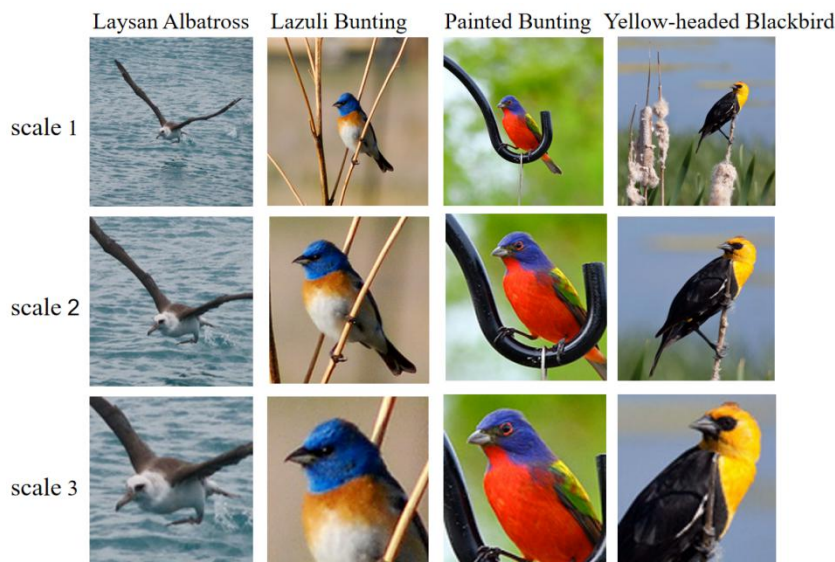


图 4-3 注意力机制模式

Fig 4-3. Attention mechanism model

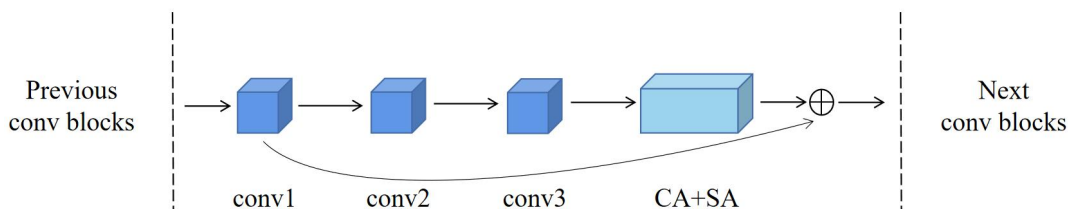


图 4-4 网络局部注意力模块

Fig 4-4. Network local attention module

为了查看注意力机制的引入对模型训练产生的影响，使用热力图可视化方法来了解模型的学习情况。本章采用 Grad-CAM (Gradient-weighted Class Activation Map) [78] 算法，它能对输入图像生成类激活的热力图。Grad-CAM 是与目标任务类别有关的二维特征分数网络，网络的每个位置表示该类别的重要程度，颜色越深的地方表示原始图片对应区域对网络的响应越高、贡献越大，网络模型对其局部特征学习到的权重越大。Grad-CAM 使用 ResNet50 网络中最后一层卷积层的梯度数据为每个神经元分配重要值，以做特定的目标注意决策。公式 (4-1) 表示的是类别 c 中，第 k 个特征图的权重大小，公式 (4-2) 则是为

了获取类别判别图 Grad-CAM，计算权重矩阵和梯度相对于激活函数的连续矩阵乘积，以获得特征图 k 对目标类别 c 的“重要”程度。

图 4-5 显示了模型对新测试图像的预测分类结果，其中，图的上半部分是模型在训练阶段未加注意力机制情况下的预测结果，下半部分则是加入注意力机制后呈现的预测结果。从热力图可以较明显地看出，未加注意力机制的模型没有抓取到真实类“Orange crowned Warbler”的关键特征，使其分类结果错误。

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (4-1)$$

$$L_{Grad-CAM}^c = ReLU \left(\sum_k \alpha_k^c A^k \right) \quad (4-2)$$

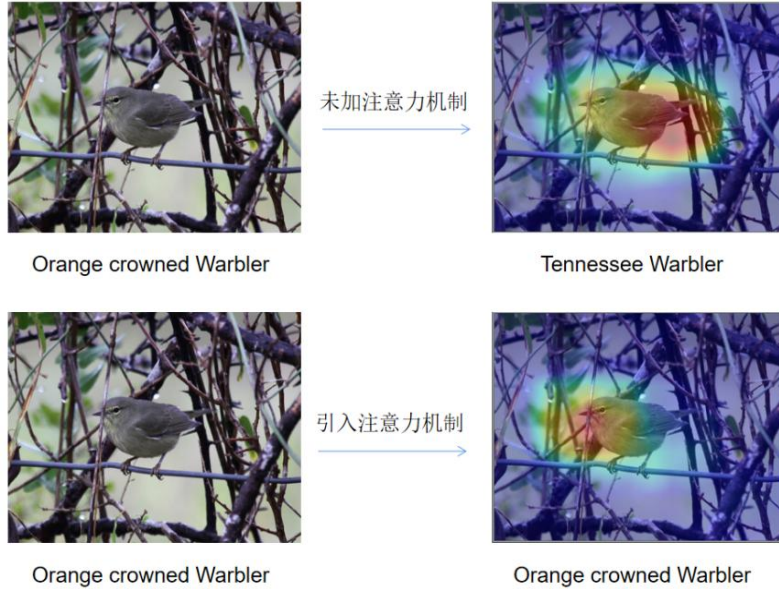


图 4-5 模型预测分类热力图

Fig 4-5. Model prediction classification thermodynamic diagram

基于训练好的 ResNet50 网络模型，图 4-6 展示了模型训练阶段的准确率曲线和损失曲线图效果。其中，左侧是模型训练阶段的准确率曲线图，右侧是模型损失曲线图。从中可知，基于注意力机制的算法模型在鸟类细粒度图像分类任务中已达到 87.1% 的精度，超越了大多数算法在该类任务上的精度表现。

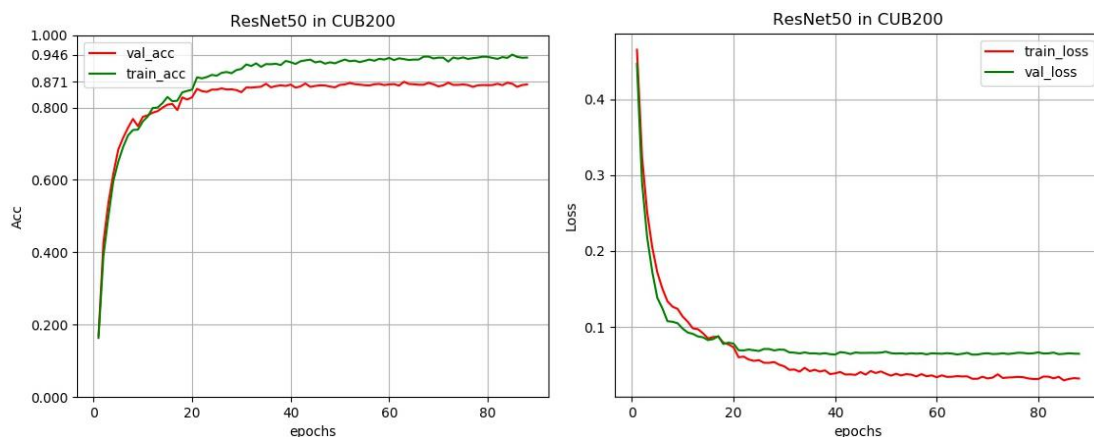


图 4-6 模型训练阶段准确率和损失曲线图

Fig 4-6. Accuracy and loss curve of model training stage

(2) 知识图谱构建

由于计算机视觉算法存在的技术局限，本文需要对模型预测不准确的鸟类图像进行更加细致的判断，介入额外的辅助信息帮助机器实现可靠的分类结果。在本章实验中，引入知识图谱这种具有结构化大数据的知识库信息，提供更加细节的特征描述帮助模型提高预测结果的准确性。

知识图谱(Knowledge Graph)，是以现实世界中存在的实体或概念为节点，形成实体或概念之间的语义关系的知识库。它可以通过查询百科网站等半结构化数据库，从中提取研究所需的实体和模式信息。如图 4-7 所示，本文构建了一个基于鸟类的知识图谱，其中每一鸟类周围都包含其所属的属性特征节点，不同类别之间会存在相同属性情况，在图谱中会共享一个节点进行连接。

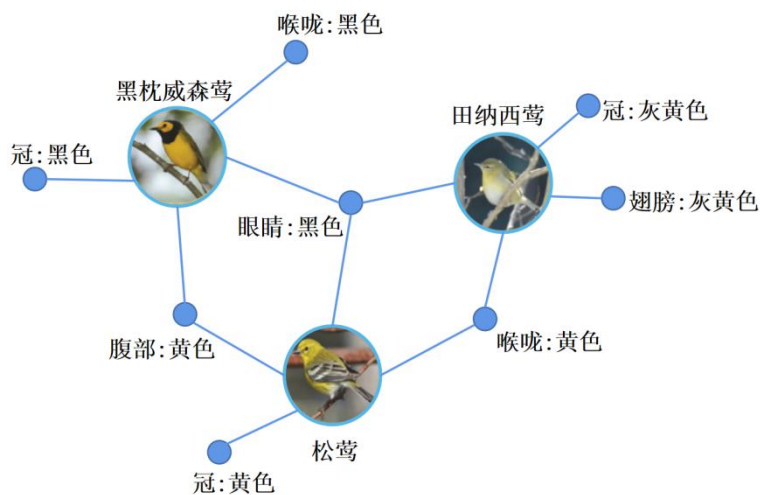


图 4-7 基于鸟类的知识图谱

Fig 4-7. Knowledge graph based on birds

图 4-8 呈现的是知识图谱构建的技术架构，其中蓝色框标识为知识图谱的构建过程。依据该知识图谱架构，本章采用 Neo4j 开源图数据库对鸟类数据集的类别及其属性关系做存储。在鸟类数据集的知识提取阶段，主要数据来源于

官网提供的人工标注信息，其有对每一类的部位属性进行特征描述，从而得到每一类别之间的属性信息。最终通过属性对齐可以收集到每种鸟类基于 15 个部位的 28 组属性描述，其具体细节如表 4-1 所示。

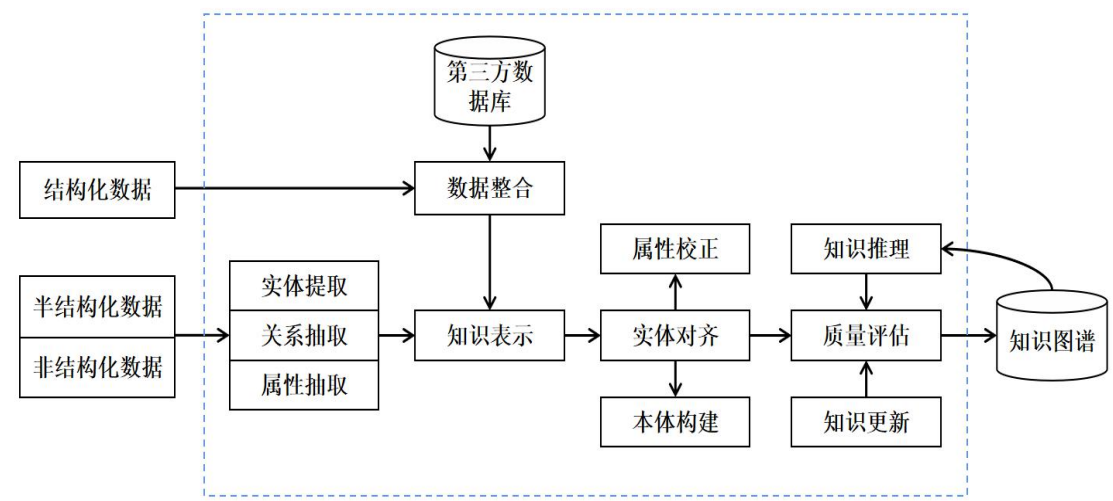


图 4-8 知识图谱的技术架构图

Fig 4-8. Technical architecture of knowledge graph

表 4-1 鸟的部位属性描述

Table 4-1. Part attribute description of bird

部位	属性	部位	属性	部位	属性
喙	喙的形状	肩羽	羽的颜色	身体	上半部分颜色
	喙的长度		羽的形状		下半部分颜色
	喙的颜色		羽的图案		主要颜色
腹	腹部图案	背	背部颜色	胸	胸部图案
	腹部颜色		背部图案		胸部颜色
前额	前额颜色	颈	颈部颜色	头	头的图案
尾	尾上/下覆羽颜色	眼睛	眼睛颜色	腿	腿的颜色
	尾部形状	冠	鸟冠颜色	鸟的	尺寸
	尾部图案	喉	喉咙颜色	整体	形状

后期为扩增实验数据类别，验证系统的可扩展应用性能，针对数据集以外的类别信息可通过相关百科类网站（如图 4-9）进行知识获取。得益于大数据技术的发展，目前有众多如 WhatBird、All About Birds 等鸟类网站，其网站中有许多鸟类的描述信息。利用爬虫技术可将所需要的鸟类信息进行信息抽取，包括从网页信息中抽取目标鸟类实体信息，采集特定实体的属性信息，并从相关语料中提取出实体之间的关联关系，最终形成网状的知识结构。知识图谱可根

据类别新增或是原有数据知识缺陷补充更新知识库，从而构建一个大规模、高质量的结构化知识库。

引入知识图谱这一辅助信息，可利用结构化的知识库找寻不同类别之间相同的信息，也可以查找到区别不同类别之间的关键信息。本章的实验目标是最大化找到能够区分所关注的几类相似物种间的属性特征。依据该思想，本文采取的技术方法是计算每一属性节点的信息增益值，因为信息增益是特征选择的一个重要指标，用来衡量一个属性区分数据样本的能力。在本实验中代表在给定一个特征属性的条件下，使得初始分类预测概率的不确定性减少的程度。最终希望遍历知识图谱上所有可能的属性节点，以获得信息增益值最大的节点特征。本章希望结合知识图谱数据，利用这种方法来有效收集最具代表性和最有价值的人类反馈。

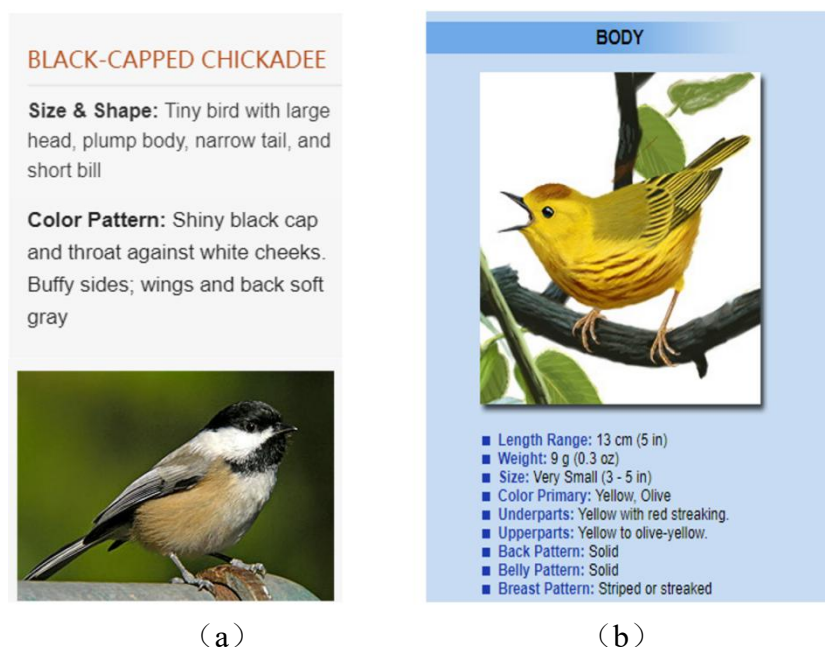


图 4-9 鸟类百科网站；（a）All About Birds 网站；（b）WhatBird 网站

Fig 4-9. Bird encyclopedia website; (a) All About Birds website, (b) WhatBird website

（3）GUI 人机问答模式设计

因为本章实现基于人的融合介入的智能系统，人可以通过修正、提供反馈和传入更多相关知识来弥补机器算法自身的缺陷。因此，本章利用 PyQt 设计了一个 GUI 人机问答模式界面。本章对图形用户界面的设计宗旨是不需要一定是具有相关领域经验的人员才可使用，简单易懂的问答方式普通用户也可轻易上手，开拓用户人群的使用性，提高人机交互的高效性。

细粒度视觉分类任务对于人和机器来说都很困难。因此，一种交互式算法，在计算机视觉难题尚未解决的过渡时期，通过这种方式帮助人类提供可用的智能服务，这比自动化但容易出错的算法更可取。那在本章中对人的介入就要考

量准确率与时间的关系：通过花少量的用户反馈时间提供更多可用信息，以实现可靠精准的输出。这不仅可以获得更好的用户体验，而且可以最大化减小最终预测的不确定性。

因此，针对上述想法的考量，本章设计了图 4-10 所示的人机问答界面。用户会从显示的 UI 界面中看到一张被判为系统识别有误的鸟类图像，一个需要用户回答的局部属性问题，包含一个由问题引出的鸟类部位参考图（图 4-11 所示），以方便用户做出正确的视觉判断。用户需要做的就是针对问题给出自己的视觉信息，这里的用户信息传递参数有两个值，一个是基于二进制问题的回答，包含“是”或“否”，考虑到鸟类姿态多变，拍摄的图像可能会出现局部部位遮挡或未涉及的情况，所以还设置了“看不见”的情况；另一个是用户基于回答情况的置信度评估，用户需要根据自己的视觉能力给出较为准确的判断准则，因为本文是希望发挥出人类的视觉优势，将系统从原先的机器视觉局限中挣脱出来，以提供更具信息性的特征描述引导其做出更为可靠的决策。



图 4-10 GUI 用户界面

Fig 4-10. GUI user interface

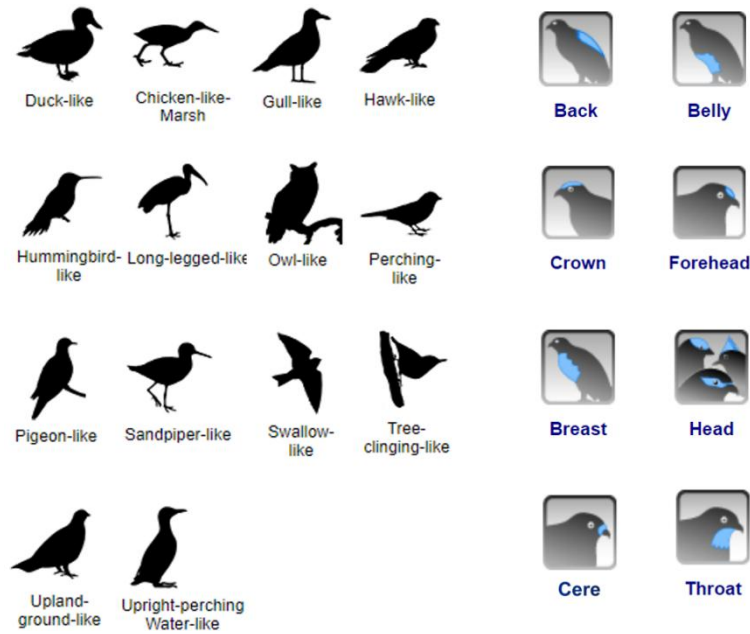


图 4-11 属性名词参考图

Fig 4-11. Attribute noun reference diagram

4.2.2 技术路线

为了验证本章提出的方法框架对鸟类数据集的识别效果，将进行以下实验操作：

- 1) 首先选取鸟类细粒度图像中具有代表性的公开数据集作为本实验的验证对象；
- 2) 针对上述所选数据集构建类-属性特征关联的知识图谱；
- 3) 将数据集按一定比例进行训练集、验证集和测试集划分，利用训练集和验证集对 ResNet50 网络进行训练和测试，通过不断调整可调参数（如学习率等），直至测试精度达到一个稳定的最高值，保存为最优模型；
- 4) 分析最优分类模型给出的预测概率结果 $p(c|x)$ ，选取合适的阈值，以便将计算机识别有误的情况最大化筛选出来另作辅助预测；阈值之上的则以计算机视觉算法预测结果为准^[79]；
- 5) 同时对用户响应进行建模，利用训练集图像数据让系统了解用户对类别属性的感知程度；
- 6) 建模完后，开始使用测试集对本章设计的系统进行实验测试，一旦触发仲裁机制则表明该图像需要进一步的预测分类。此时，提取计算机视觉算法给出的前五个类概率信息作为基准信息，并在事先规划好的知识图谱上获取这五个类别及其属性节点另连接成一个子图；
- 7) 基于构建的子图，计算所有特征节点的信息增益值，并检索最具信息性的特征属性作为问题向人类用户进行询问；
- 8) 人类用户介入系统问题，根据自己的视觉感知信息传输给计算机；
- 9) 最终计算机结合嵌入的多源信息（知识图谱、人的视觉感知），重新做

概率预测 $p(c|x, U)$ ，并输出分类结果。

4.3 实验所用数据集和阈值设置

4.3.1 鸟类数据集

实验所选数据集是 CUB-200-2011，它是美国加州理工学院公开的数据库，同时也是细粒度识别研究的基准数据集。该数据集内含 200 种鸟类物种，共 11788 张图像，其中训练集 5994 张，测试集 5794 张。该数据集提供的数据样本姿态、光照和尺寸规模变化大，样本来源也较为广泛，非常贴近真实场景下的图像细粒度分类任务。本实验将测试集按 5: 5 从中划分出 2916 张数据图像作为验证集，以用于查看模型训练效果。图 4-12 显示了实验所用数据集的数据分布情况，很明显，图（a）所示的每种鸟类数据量在 40-60 张图片不等。为确保模型接收的数据样本均衡性，每种鸟类的训练图像约有 30 张，如图（b）所示。

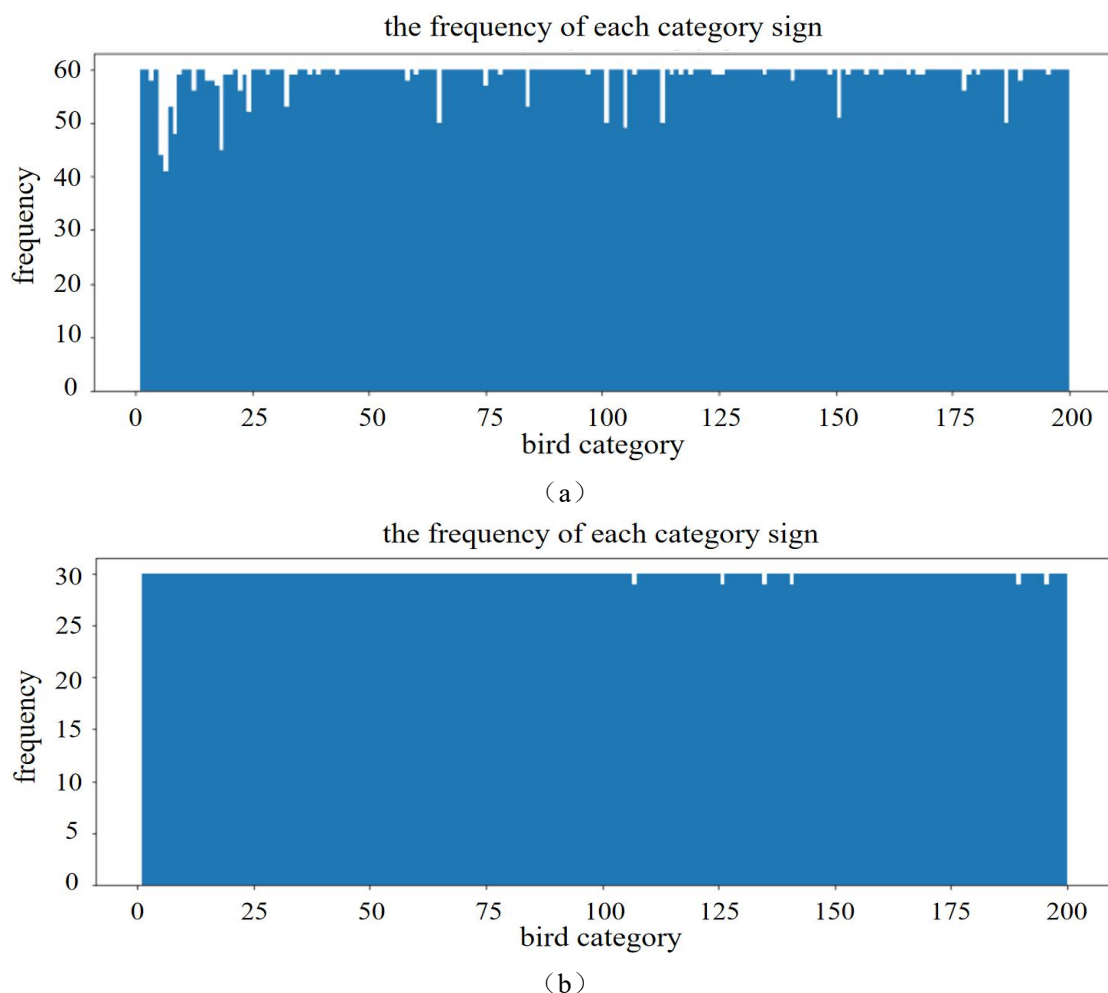


图 4-12 数据集分布；（a）CUB 数据集；（b）训练集

Fig 4-12. Data set distribution, (a) CUB dataset, (b) Training set

4.3.2 训练参数设定

本章实验所有预训练和微调任务采用动量为 0.9 的随机梯度下降算法进行网络的权重更新训练。在预训练阶段，使用在 ImageNet 数据集上预训练好的 ResNet50 网络做基准模型进行参数初始化，图像预处理到 448x448，并使用 BN 作为正则化器。微调阶段，设定 0.001 的初始学习率，0.00005 的权值衰减，使得一定程度上减少模型过拟合的问题。在第 20 个周期后，对学习率进行衰减，设为 0.0005，到达第 30 周期之后，改为 0.0002，最后在到达第 50 周期后，将学习率衰减到 0.0001。由于 CUB-200-2011 数据集中每类物种训练集的图像数量过少，采取在已有数据的基础上进行数据扩增，包括对图像进行随机翻转，旋转，剪裁，缩放等几何变换操作，以提高模型的泛化能力。

4.3.3 阈值设置

与第三章设计宗旨一样，在进行细粒度鸟类图像识别系统设计时，也要确保机器算法的失误能够被有效检测，并且尽可能通过较少的人力介入实现对算法性能的提升。因此本章通过设立一个阈值作为仲裁机制，阈值之下的图像需要借助本章设计系统进一步预测，阈值之上的仍以计算机视觉算法预测结果为准。需要注意的是，计算机视觉算法的预测结果输出以概率最高的类别（也叫 Top1 分类）为准，包括和阈值的对比也是看 Top1 的预测概率值大小。

这里设定当模型预测概率高于阈值则为正类，即视为算法分类正确，反之则视为分类错误。当模型预测分类的置信度低于阈值时，机器的决策将变得不可信，需要人的决策介入。实验希望的是，被视为正类的图像预测结果尽可能正确，同时能够最大程度地提取出机器识别有误的类别图像，这样可以减少人的介入成本，也可以提高最终系统的预测准确性。基于该思想，参考二分类方法中对模型性能的评价指标 ROC，本实验也绘制基于多分类的 ROC 曲线，从而选取最优阈值作为仲裁机制的衡量指标。

在绘制 ROC 曲线之前，需要对模型预测可能给出的几种情况进行分析，可通过表 4-2 查看。其中 TP 表示模型预测正类，且就为目标类别；FP 表示模型预测正类，但不是目标类别；FN 表示模型预测负类，但为目标类别；TN 则表示模型预测负类，且不是目标类别。ROC 曲线的绘制其实基于一系列 TPR 和 FPR 的数值，设置不同的阈值，分类结果会有所改变，TPR 和 FPR 的数值也会发生一系列变化。具体的 TPR、FPR 计算如下：

$$TPR = \frac{TP}{TP + FN} \quad (4-3)$$

$$FPR = \frac{FP}{TN + FP} \quad (4-4)$$

表 4-2 混淆矩阵

Table 4-2. Confusion matrix

	预测正类	预测负类
目标类别	TP	FN
其他类别	FP	TN

实验利用验证集依据上述方法对每一类别绘制 ROC 曲线，如图 4-13 所示。由于该数据集有 200 个类别，为简化阅读效率，实验将 200 个类别的曲线放置在一张图上。从图中可以看出，前期模型训练成效较好，对大多数鸟类都能表现出良好的分类性能。且曲线越靠近理想数值 (0, 1) 点，表明模型预测准确性越高。

基于曲线绘制情况，寻找最佳界值点，即曲线上最靠近 (0, 1) 的点，此时 TPR 和 FPR 均达到一个平衡点，模型表现整体最优。由于每一类别都绘制了 ROC 曲线，通过上述方法将会对应产生一最佳界值点，而该最佳界值点将作为本实验的仲裁阈值，触发阈值条件则作为人介入时机的关键点。

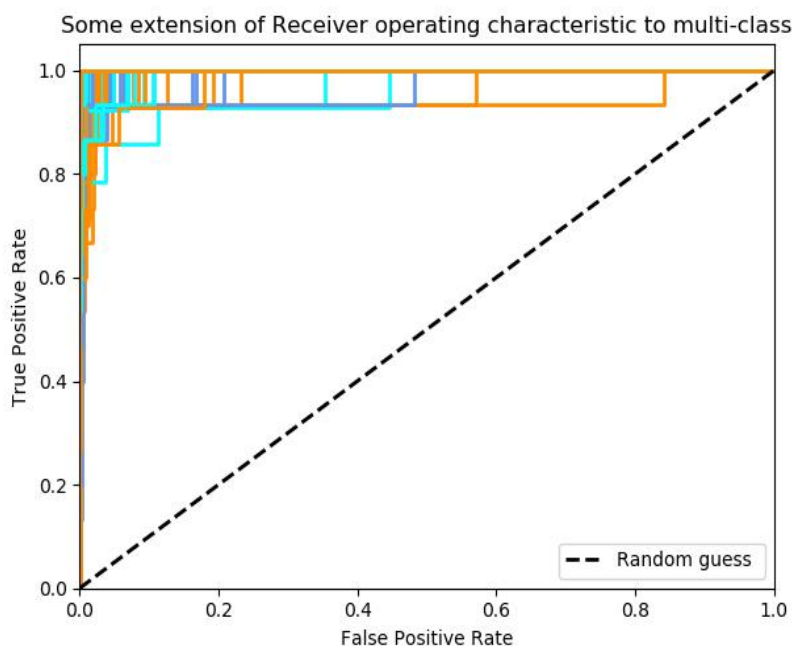


图 4-13 类 ROC 曲线绘制

Fig 4-13. ROC curve of class drawing

为了解模型对该鸟类细粒度图像的整体分类性能，本实验还绘制了一张平均 ROC 曲线（如图 4-14），同样也基于该曲线求取最佳阈值，针对两种阈值选取情况本章将做对比实验，主要从人力成本和系统准确率之间进行讨论。

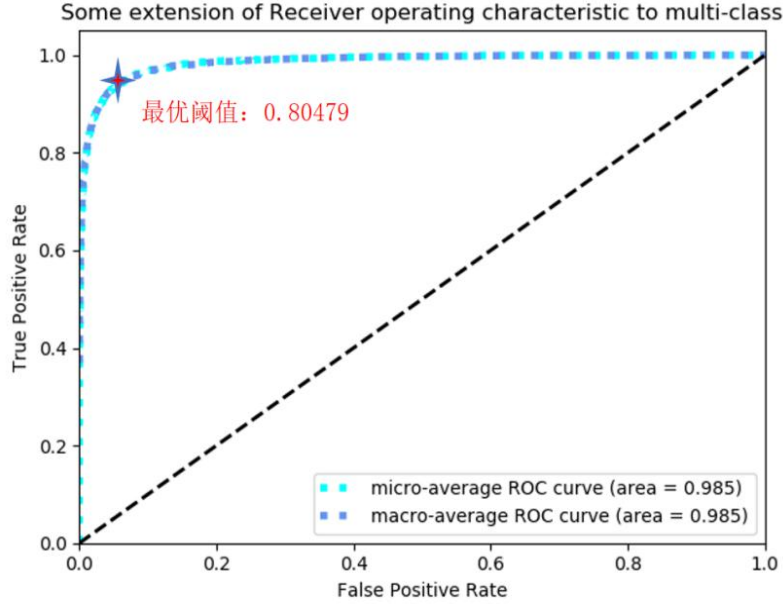


图 4-14 平均 ROC 曲线绘制

Fig 4-14. Average ROC curve drawing

4.4 系统算法与分析

4.4.1 用户建模

涉及到该实验中对人的鸟类识别经验要求较低，为达到更好的模型性能，本文需要设计和进行更严格的深入用户研究，以评估人在环框架的有效性和鲁棒性。在系统测试之前，由于存在人的介入决策信息，实验需要事先对用户建模，即将人的认知与处理能力引入数据分析的过程中。因此设计算法 4-1（如表 4-3 所示）让系统了解用户对类别属性的感知程度。因为用户可能会存在主观因素或是对某一属性错误判断的情况，实验需要考虑这方面因素的干扰，以便最终在测试阶段减少来自人类决策判断的失误影响。

实验选取 20 名无鸟类识别经验的用户作为实验者进行建模，其中男性 10 名，女性 10 名，所使用的建模数据为模型训练阶段的 5994 张训练集，按类别每人分配 300 张左右图像进行观察。用户需要通过给定的图像信息，根据该数据集中包含的每一属性值，判断图像 x_j 中是否存在一 a_i 属性特征。当所有用户操作完成后，系统计算出用户对特定属性值感知判断的概率，以及针对每一类别下用户对其属性值的感知概率判断。具体计算如公式（4-5）、（4-6）所示：

$$p_H(\hat{a}_i) = \frac{\sum_j \hat{a}_i^j}{n} \quad (4-5)$$

$$p_H(\hat{a}_i|c) = \frac{\sum_j \hat{a}_i^j 1[c_j = c]}{\sum_j 1[c_j = c]} \quad (4-6)$$

其中, a_i 是一个二元属性, 表示某一属性存在与否的状态, 由 0 或 1 来表示, 即 0 表示该属性存在, 1 表示该属性不存在。 $\hat{a}_i$ 是用户对 a_i 进行预测的随机变量。 $\hat{p}_i$ 表示用户对第 i 个属性判断存在的概率, $\hat{p}_i^c$ 表示用户判断所有属于类别 c 的图片中第 i 个属性存在的概率。这里 $1[\dots]$ 是一个指示函数, 当满足图片 x_j 为类别 c_j 时输出为 1, 否则为 0。

表 4-3 用户建模伪代码
Table 4-3. User modeling pseudo code

算法 4-1: 用户建模

输入: 训练集中的图片 x_j , 类别标签 c_j 及其属性值 a_i 。

输出: 用户响应概率 $\hat{p}_i$ 、 $\hat{p}_i^c$ 。

- 1) 创建二元属性值 a_i 的用户感知集合: $\hat{a}_i = \{ \}$
- 2) **for** $j = 1$ to n **do**
- 3) 计算属性 a_i 存在在图片 x_j 中的用户感知: $\hat{a}_i^j = p(a_i^j)$
- 4) **end for**
- 5) **Return** $\hat{p}_i = p_H(\hat{a}_i)$, $\hat{p}_i^c = p_H(\hat{a}_i|c)$

4.4.2 系统提问设置

对于一张需要进一步做分析判断的图像, 根据机器视觉模型对其给出的预测概率结果, 获取预测前五的类别及概率信息, 并在知识图谱中提取这五个类别及其相连属性节点, 构建图谱子图。接下来所有关于该图像的计算和节点查询都建立在该子图之上。详细的系统问题设计如表 4-4 所示:

设子图谱 $N = (V, E)$ 是一连通网, V 是顶点集合, 包含类别名称和属性特征; E 是边的集合, 即属性描述。其中 Q^t 是一问题集合, 包含对知识图谱子图中的所有节点进行逐一属性提问, 例如一属性“翅膀是黑色的”的问题就是“该鸟的翅膀是黑色的吗?”, 而 AQ^t 将会存储系统提问过的问题集合。为选取对最终决策结果信息帮助最大的属性特征, 需要对其中的每一属性进行信息增益计算。

信息增益的求取由两部分组成, 首先需要计算信息熵, 它代表随机变量的不确定性度量, 熵值越小表明数据的不确定性越低。如 (4-7) 公式展示, 它是对前期计算机视觉分类预测的不确定性评估。其次还需要计算其条件熵, (4-8) 公式所示, 在已知某一属性 a_i 条件下, 再次评估其分类预测的不确定性。信息增益的计算, 其实就是信息熵与条件熵之间的差值。通过遍历所有的属性节点 V^c , 条件熵值越小, 对最终分类预测的不确定性程度越小, 信息增益值将会越

大，系统给出的决策结果也将越可靠。最终系统通过选取最大信息增益值对应的属性节点，并以问题的形式向人类用户寻求回答。

$$H(c|x) = - \sum_{j=1}^n p(c|x) \log p(c|x) \quad (4-7)$$

$$H(c|x, a_i) = - \sum_{j=1}^n p(c|x, a_i) \log p(c|x, a_i) \quad (4-8)$$

表 4-4 问题设计伪代码

Table 4-4. Problem design pseudo code

算法 4-2：问题设计

输入： 机器视觉预测概率前五的类别概率 $p(c_k|x)$ 及其属性节点 v_i^c 。

输出： 提问最具信息性的属性问题 q_i 。

- 1) $AQ^t = \{\}$. If $t = 0, Q^t = \{q_1, \dots, q_m\}$, else $Q^t = Q^{t-1}$
 - 2) 用户响应集合 $U^{t-1} = \{u^1, \dots, u^{t-1}\}$
 - 3) **for** $i \in 1 \dots \text{size}(V^c)$ **do**
 - 4) $QIG[i] = [False, \emptyset, \emptyset]$
 - 5) 计算 $IG(c; u^t|x, U^{t-1})$
 - 6) $MIG(i) = \max_i IG(c; u^t|x, U^{t-1})$
 - 7) **if** $QIG[i][2] \leq MIG(i)$ **then**
 - 8) $QIG[i] = [True, q_i, MIG(i)]$
 - 9) **for** $QIG[i][0] == True$ **do**
 - 10) $AQ^t.add(QIG[i][1]), Q^t.remove(QIG[i][1])$
 - 11) **end for**
 - 12) **Return** q_i
-

为保证针对同一图像所提出的问题不具重复性，在每一阶段计算属性节点信息增益时，被提问过的属性将不再被计算，即不会被选择作为问题提出。由于问题设置具有迭代性，直到系统对某一类别预测概率高于阈值时才会结束提问，因此公式（4-9）是在前一条件熵的基础上增加了新一阶段用户响应给出的属性条件。

$$H(c|x, U^{t-1}) = - \sum_{c_k} p(c|x, U^{t-1}) \log p(c|x, U^{t-1}) \quad (4-9)$$

其中，

$$p(c|x, U^{t-1}) = \frac{p(U^{t-1}|c, x)p(c|x)}{\sum_{c_k} p(U^{t-1}|c_k) p(c_k|x)} \quad (4-10)$$

这里 $p(U^{t-1}|c, x) = p(U^{t-1}|c)$ ，因为实验中对用户属性感知的建模仅取决于对象类别 c 而非图像。考虑实验中的用户响应回答是独立的，即用户每次的回答不会受先前问题信息的影响，因此 $p(U^{t-1}|c) = \prod_1^{t-1} p(u^m|c)$ 。

$$p(u^t|x, U^{t-1}) = \sum_{c_k} p(u^t|c_k) p(c_k|x, U^{t-1}) \quad (4-11)$$

最终每一阶段的信息增益计算可由公式（4-12）可得，

$$IG(c; u^t|x, U^{t-1}) = p(u^t|x, U^{t-1}) \left(H(c|x, U^{t-1}) - H(c|x, u^t \cup U^{t-1}) \right) \quad (4-12)$$

4.4.3 用户响应更新类概率

针对上述系统提出的特征属性问题，人类给予“是”或“否”的回答，“是”代表该特征属性存在，此时 $\hat{a}_i = 1$ ；“否”代表该特征属性不存在，此时 $\hat{a}_i = 0$ 。基于人类视觉的信息传输用 u^t 来表示，即 $u^t = (\hat{a}_i, r_i)$ ，其中 r_i 表示用户回答的置信度，实验设置“猜测，可能，确定”这三个参数作为人类决策过程的置信度判断。因此，针对用户响应的概率计算如下：

$$p(u^t|c) = p(\hat{a}_i, r_i|c) = \alpha_{r_i} \cdot p(\hat{a}_i|c) \quad (4-13)$$

其中， α_{r_i} 参数用于改变用户响应的权重，根据用户给出的响应置信度对响应进行缩放，以确保用户响应的有效贡献。明显 $\alpha_{\text{猜测}} < \alpha_{\text{可能}} < \alpha_{\text{确定}}$ ，由于会出现图像像素质量不高、目标物体局部部位遮挡等因素干扰，导致用户存在视觉信息的不确定性，因此针对用户较为明确的信息判断，应给予更高的置信权重。这里，经过系统多次调参后，选取 $\alpha_{\text{猜测}} = 0.4$ ， $\alpha_{\text{可能}} = 0.72$ ， $\alpha_{\text{确定}} = 0.95$ 时，模型性能表现最优。

在每一阶段，算法会根据用户响应传输重新做概率预测，整个类概率更新计算如表 4-5 所示。

$$\begin{aligned} p(c|x, U^t) &= \frac{p(U^t|c) \cdot p(c|x)}{\sum_{c_k} p(U^t|c_k) \cdot p(c_k|x)} \\ &= \frac{p(U^{t-1}|c) \cdot p(u^t|c) \cdot p(c|x)}{\sum_{c_k} p(U^{t-1}|c_k) \cdot p(u^t|c_k) \cdot p(c_k|x)} \end{aligned} \quad (4-14)$$

表 4-5 用户响应更新类概率伪代码

Table 4-5. User response update probability pseudo code

算法 4-3: 用户响应更新类概率

输入: 用户对 q_i 问题的响应回答 u^t 。

输出: 结合用户反馈响应重新做预测分类 $p(c|x, U^t)$ 。

- 1) $u^t = (\hat{a}_i, r_i), r_i \in \{\text{猜测}, \text{可能}, \text{确定}\}$
- 2) $U^t = U^{t-1} \cup u^t$
- 3) **for** $c_k \in \text{the top5 } C \text{ of probability}$ **do**
- 4) $p(u^t|c_k) = \alpha_{r_i} \cdot p(\hat{a}_i|c_k)$
- 5) **if** $\hat{a}_i = 1$ **then**
- 6) $p(\hat{a}_i|c_k) = \hat{p}_i^c$
- 7) **else**
- 8) $p(\hat{a}_i|c_k) = 1 - \hat{p}_i^c$
- 9) 计算 $p(c_k|x, U^t)$
- 10) **end for**
- 11) **Return** $\text{argmax}_c p(c|x, U^t)$

4.4.4 结果分析

表 4-6 是本章对两种阈值设置方法的对比, 主要是为了最大化找出模型自主预测过程中存在的错误判断, 将其视为负类交由人进一步判断。第一种是多类阈值法, 即每一类都会根据模型预测能力设定一个阈值, 其优点是有针对性的了解模型在各类别预测分类中的置信度, 以便更好地找出模型错误分类情况, 缺点就是一旦图像质量不高, 可能会有损模型判断标准, 这对于前期以高概率预测的某类别来说, 即使最终模型仍判断正确但依旧被视为负类, 增加了人力成本。第二种是平均阈值法, 它是基于对每一类别的 ROC 曲线取平均, 绘制一条综合 ROC 曲线得来, 即最终的仲裁阈值唯一, 优点是能够减少系统中混入预测精度不高但模型预测正确的图像, 但缺点就是会扩大正类中错误分类的比例。

表 4-6 两种阈值设置方法对比

Table 4-6. Comparison of two threshold setting methods

	负类图像	TP	FN	FP	TN	TPR	FPR
多类阈值法	449	2406	134	61	315	0.9472	0.1622
平均阈值法	402	2425	115	89	287	0.9547	0.2367

为选取更优的阈值，将两种阈值设定在验证集上展示效果。从表中可以看出，多类阈值法相较于平均阈值法能检测出更多的负类图像，占整个验证集错误率的 83.8%。但同时，也增加了不必要交给人类判断的正类图像数量。为权衡人力成本和预测精度，分别对两种方法求取 TPR 和 FPR 值，并计算与 (0, 1) 点之间的距离。结果显示，多类阈值法对整个系统的性能表现更优。因此，在最终的测试阶段，系统将基于该方法作为仲裁机制的评判标准。当然，若是对人力成本要求不是很高的应用场景，可以通过平均阈值法增大阈值范围，将更多可能的负类图像交由本章设计的系统进行再判断。

通过测试集对本章设计的系统进行验证，表 4-7 呈现了不同鸟类细粒度图像识别模型在 CUB 数据集上的识别精度对比。这里所选取的算法模型有一个共性，即都是运用易获取的外部辅助信息（网络数据、文本描述、用户信息等）帮助模型提升鸟类图像识别精度。但相对来说，这样的方式仍然解决不了模型算法的本质局限。模型在复杂环境下的自主执行能力时好时坏，相对不稳定，一方面是由于训练阶段无法模拟现实场景下所有可能发生的案例情况，另一方面也是由于模型对自身任务的学习掌握程度不够“深”，存在某些方面的智能仍未达到人类水平。因此，在未完美解决上述问题的情形下，本文借助人机智能优势，在机器智能决策有误的地方，实现人机的协同合作，提升最终系统的可靠输出。

表 4-7 不同鸟类细粒度图像识别模型对比

Table 4-7. Comparison of different bird fine-grained image recognition models

方法	基准模型	外部信息	识别精度
Multi-grained	VGG-19	WordNet	83.0%
AP-RCNN	CaffeNet	Web data	84.6%
CVL	VGG-16	Text	85.6%
KERL	VGG-16	Attributes	87.0%
MetaFGNet	ResNet-34	Auxiliary data	87.6%
TA-FGVC	ResNet	Text	88.1%
Bi-Model	ResNet-50	Text	88.7%
RIBQ	ResNet-50	User information	90.1%
Ours	ResNet-50	Human-in-the-loop	91.7%

除了实验精度分析以外，本节还针对测试阶段用户对系统回答的题目个数以及所花时间进行了图示分析，如图 4-15、4-16 所示。从图中可知，用户通过使用本章设计的鸟类细粒度图像分类系统，对机器分类有误的图像的介入响应问题数量均在 5 个以内，且平均每题所花时间在 3-5 秒之间。相对来说，本章的系统设计尽可能考虑用户体验的便利性，以花少量的问题数量提问来获取用户最大的信息反馈。同时，考虑到用户的经验储备，在用户界面提供相应的部位参考以减轻用户的使用负担。

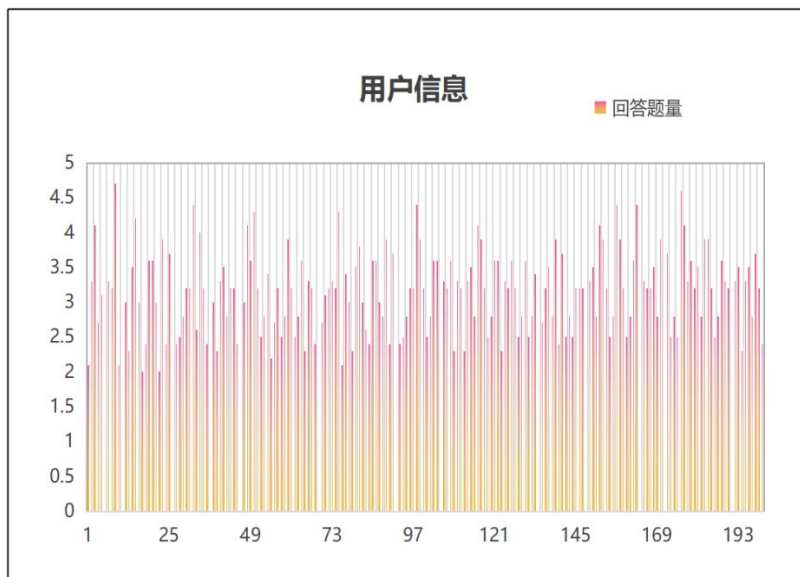


图 4-15 用户对每种类别介入响应数量

Fig 4-15. Number of user intervention responses for each class

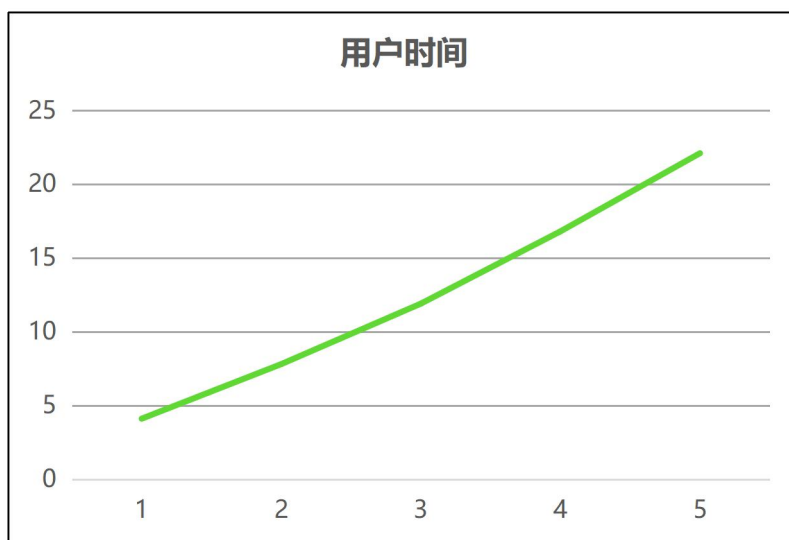


图 4-16 用户平均回答每题所花时间

Fig 4-16. Average time spent by users answering each question

由于本章在知识图谱的查询阶段只考虑了前 TOP-5 类别情况，这在一定程度上遗漏了某些概率极小的正确类别，导致后续的工作达不到更高的精度效果。

因此，本节在这一因素考虑上扩大类别查询，并考虑 TOP-K ($K \geq 5$) 在分类精度与人介入决策所需时间之间的关系，如图 4-17 所示。从图中可以明显看出，随着 K 值的增大，系统的分类精度不断上升，但相对人花的时间成本也有所增加。并且随着 K 值的增大，系统分类精度提升速率也将达到一个饱和状态，而人介入时间会持续增加，因此本章不再对后续的 K 值进行实验比较。这一方法的算法精度与人力成本之间的权衡就要根据具体的应用场景有所选择。

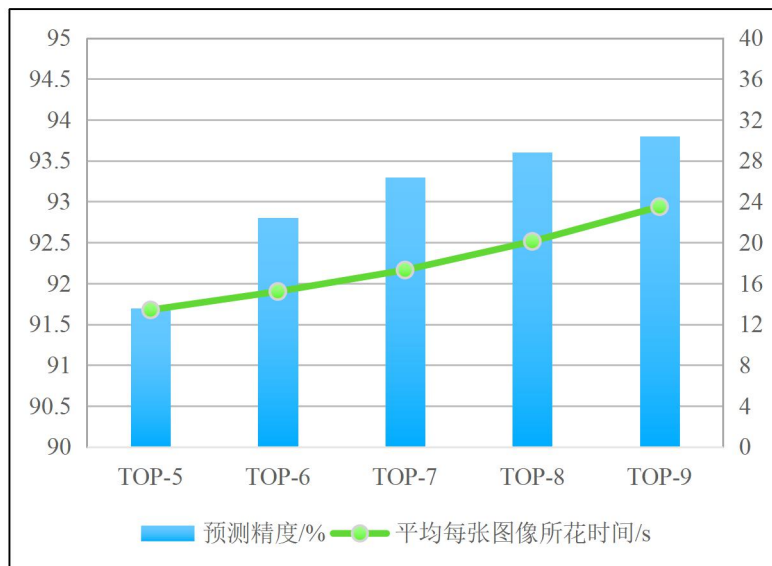


图 4-17 模型预测精度与用户所花时间关系图

Fig 4-17. Relationship between model prediction accuracy and user time

4.5 本章小结

本章设计了一种多源信息融合的细粒度图像识别系统，利用人的融合介入方式针对计算机易混淆的细粒度鸟类图像做进一步预测分类，通过嵌入多源信息（知识图谱和人的视觉信息）向算法提供更具信息性的特征描述引导计算机实现高精度的分类。经实验验证，与过往的算法模型相比，本系统在识别精度有明显提升，且决策输出更为可靠。

第五章 总结与展望

5.1 总结

基于深度学习的不断发展，如今的 AI 技术有了突飞猛进的提升。目前众多领域借助成熟的深度学习技术，以节省大量的人类劳动成本而大力发展人工智能。但随着越来越多的智能机器融于人类社会，人们对系统的应用性能要求也随之提高。在现有的基于深度学习的技术框架下，许多领域对智能系统算法的高精度要求并不容易满足。这是因为其准确度的提升很大程度上取决于大量优质训练数据的获取和高效算法的开发。但目前众多人工智能系统可能并不具备足量、高质的训练数据的天然条件，优质数据的难得和标注数据的高昂成本也让模型算法的提升增添了难度。这就使得在深度学习的基本框架外寻求精度的提升成为近些年的一个重要关注点。

在这一领域中，有研究者发现人的介入可能是提升深度学习应用精度的一种可行途径。该方法借用人机系统中介入思想以解决非人机系统中的相关技术难题，即系统在正常运行时由机器的智能自主和机器的自动化共同起作用，而在机器算法有误时允许人的介入来提升整体系统的决策精度。

本文基于典型实例研究了人的介入提升深度学习应用精度的两种方法，即人的分歧介入方法和人的融合介入方法。前者方法考虑在多个算法产生预测分歧之处，决策权将交由经验丰富的人直接控制；后者则更多的考虑寻求基于普通用户介入下的可操作性，在算法的预测置信度较低时，利用人在感知层面的信息传输实现对算法错误决策的可靠修正。具体研究工作如下：

(1) 面对珍珠分拣任务的经济效益与质量需求的高标准，提出了一种人的分歧介入方法，该方法在公开珍珠数据集上，以 4.1% 的额外人工成本提升了近 4% 的珍珠分拣精度，较大程度上解决了现有珍珠分拣精度不高的问题。本实验为珍珠分拣系统提供额外的冗余 AI 算法，通过比对多个算法的预测结果，在产生分歧之处作为仲裁机制来判断系统的预测是否出错，以此驱动珍珠分拣人员的介入。最终使用定义分歧准确指数和额外成本指数这两大性能评价指标来评估设计系统的整体性能。

(2) 针对野生鸟类资源信息识别对整个社会生态价值和经济价值的高需求，提出了一种人的融合介入方法，该方法在公开基准 CUB-200-2011 数据集上，以最小化的人工成本将鸟类细粒度图像识别精度提升到 93.8%，实现了目前机器视觉算法在细粒度鸟类图像识别难题上的突破。借用二分类模型评估思想，

本实验通过分析算法的 ROC 指标，选取类别预测最优阈值作为仲裁机制，当算法的预测置信度低于阈值条件即作为人的介入时机。并通过将人的视觉信息与嵌入的鸟类知识图谱信息融合为系统提供更具信息性的特征判断实现可靠分类。

5.2 展望

本文提出的核心方法在两个实例应用中都得到了有效验证，所设计的人机混合智能系统也具有一定的实用性，但仍有一些地方需要改进和完善。

（1）在对珍珠分拣进行算法设计时，其中的主次算法的选择尚缺少系统性可操作的方法，仍需针对不同应用进行尝试，这一定程度上影响了所提出方法的广泛使用性。

（2）在鸟类细粒度图像识别系统设计上，阈值的设置方法较依赖模型性能的好坏。若所选取的机器视觉模型精度不够乐观，一方面被视为正类预测的图像中会涵盖许多错误分类的图像，另一方面也会增加人类的劳动负担，将与本文研究的核心思想有所出入。

（3）对于查找知识图谱中的关键特征信息节点，本文采用的是计算相关节点的信息增益方法，这种方式存在的缺点就是对图谱中连接边数较多的属性节点有所偏好，这会增加用户被提问的次数，影响用户体验。因此，在后续研究过程中，预想在信息增益的基础上乘上一个惩罚函数，这个惩罚函数具体应与每一节点属性值所属类别的权重相关，通过这种方式以最大化查找出极具辨别性特征，同时降低计算“偏好”带来的影响。

参考文献

- [1] Gjoreski M, Janko V, Slapnicar G, et al. Classical and deep learning methods for recognizing human activities and modes of transportation with smartphone sensors[J]. Information Fusion, 2020, 62(1): 47-62.
- [2] Zhang Q, Yang L T, Chen Z, et al. A survey on deep learning for big data[J]. Information Fusion, 2018, 42(1): 146-157.
- [3] Khan Z Y, Niu Z, Sandiwarno S, et al. Deep learning techniques for rating prediction: a survey of the state-of-the-art[J]. Artificial Intelligence Review, 2021, 54(3): 95-135.
- [4] Dong S, Wang P, Abbas K. A survey on deep learning and its applications[J]. Computer Science Review, 2021, 40(1): 100379-100400.
- [5] Wang K, Zhang D, Li Y, et al. Cost-Effective Active Learning for Deep Image Classification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 27(12): 2591-2600.
- [6] Ghiassi A, Younesian T, Zhao Z, et al. Robust (Deep) Learning Framework Against Dirty Labels and Beyond[C]. //2019 First IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications. Los Angeles, US: IEEE, 2019: 236-244.
- [7] Chen Z, Wang K, Wang X, et al. Deep Co-Space: Sample Mining Across Feature Transformation for Semi-Supervised Learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 28(10): 2667-2678.
- [8] Li K, Wang X, Ji L. Application of multi-scale feature fusion and deep learning in detection of steel strip surface defect[C]. //2019 International Conference on Artificial Intelligence and Advanced Manufacturing. Dublin, Ireland: AIAM, 2019: 656-661.
- [9] He Z, Liu Q. Deep regression neural network for industrial surface defect detection[J]. IEEE Access, 2020, 8(1): 35583-35591.
- [10] Eldan R, Shamir O. The Power of Depth for Feedforward Neural Networks[J]. Computer Science, 2015, 49(1): 907-940.
- [11] Huang G, Liu Z, Maaten L, et al. Densely Connected Convolutional Networks[C]. //2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 2261-2269.
- [12] 赵云波, 康宇, 朱进. 人机混合智能系统自主性理论和方法[M]. 北京: 科学出版社, 2021..
- [13] Ling H, Gao J, Kar A, et al. Fast interactive object annotation with Curve-GCN[C]. //In Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Long Beach, CA, USA: IEEE, 2019: 5252-5261.
- [14] Egorow O, Lotz A, Siegert I, et al. Accelerating manual annotation of filled pauses by automatic pre-selection[C]. //2017 International Conference on Companion Technology. Ulm, Germany: IEEE, 2017: 1-6.
- [15] Alonso O. Challenges with label quality for supervised learning[J]. Journal of Data and Information Quality, 2015, 6(1): 1-3.
- [16] Niu S, Li B, Wang X, et al. Defect image sample generation with gan for improving defect recognition[J]. IEEE Transactions on Automation Science and Engineering, 2020, 17(3): 1611-1622.

- [17] Shen H T, Zhu X, Zhang Z, et al. Heterogeneous data fusion for predicting mild cognitive impairment conversion[J]. *Information Fusion*, 2021, 66(1): 54-63.
- [18] Li J, Yang J, Hertzmann A, et al. LayoutGAN: Synthesizing Graphic Layouts with Vector-Wireframe Adversarial Networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(7): 2388-2399.
- [19] Zhang X, Zhao J, Lecun Y. Character-level Convolutional Networks for Text Classification[C]. //Neural Information Processing Systems. Canada: MIT Press, 2015: 649-657.
- [20] Bahrami S, Dornaika F, Bosaghzadeh A. Joint auto-weighted graph fusion and scalable semi-supervised learning[J]. *Information Fusion*, 2021, 66(1): 213-228.
- [21] Chouhan V, Singh S K, Khamparia A, et al. A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-ray Images[J]. *Applied Sciences*, 2020, 10(2): 5591-5617.
- [22] Cherti M, Jitsev J. Effect of large-scale pre-training on full and few-shot transfer learning for natural and medical images[J]. *Pattern Recognition*, 2021, 113(1): 107826.
- [23] Yun S, Han D, Oh S J, et al. CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features[C]. //2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, South Korea: IEEE, 2019: 6022-6031.
- [24] 蒋芸, 张海, 陈莉, 等. 基于卷积神经网络的图像数据增强算法[J]. *计算机工程与科学*, 2019, 41(11): 2007-2016.
- [25] Cubuk E, Zoph B, Mane D, et al. Auto-Augment: Learning Augmentation Policies from Data[C]. //2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019: 113-123.
- [26] 蔡莉, 王淑婷, 刘俊晖, 等. 数据标注研究综述[J]. *软件学报*, 2020, 31(2): 302-320.
- [27] 范举, 陈跃国, 杜小勇. 人在回路的数据准备技术研究进展[J]. *大数据*, 2019, 005(006): 1-18.
- [28] Xie Q, Dai Z, Hovy E, et al. Unsupervised Data Augmentation for Consistency Training[C]. //In Neural Information Processing Systems (NIPS). Vancouver, Canada: Curran Associates, 2019: 6256-6268.
- [29] Ledig C, Theis L, Huszar F, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network[C]. //2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 105-114.
- [30] Kupyn O, Budzan V, Mykhailych M, et al. DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks[C]. //2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, UT, USA: IEEE, 2018: 8183-8192.
- [31] Nazeri K, Ng E, Joseph T, et al. Edgeconnect: Generative image inpainting with adversarial edge learning[C]. //2019 IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019.
- [32] Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need[C]. //Proceeding of the 31st International Conference on Neural Information Processing Systems. Long Beach, CA, USA: IEEE, 2017: 6000-6010.
- [33] Willis C G, Law E, Williams A C, et al. CrowdCurio: An online crowdsourcing platform to facilitate climate change studies using herbarium specimens[J]. *New Phytologist*, 2017, 215(1): 479-488.
- [34] Martinez-Rodriguez J L, Hogan A, Lopez-Arevalo I. Information Extraction meets the Semantic Web: A Survey[J]. *Semantic Web*, 2020, 11(2): 255-335.

- [35] Zhang S, He L, Dragut E, et al. How to Invest my Time: Lessons from Human-in-the-Loop Entity Extraction[C]. //In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage, AK, USA: ACM, 2019: 2305-2313.
- [36] Liu Z, Wang J, Gong S, et al. Deep Reinforcement Active Learning for Human-in-the-Loop Person Re-Identification[C]. //2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, South Korea: IEEE, 2019: 6122-6131.
- [37] Butler C, Oster H, Togelius J. Human-in-the-Loop AI for Analysis of Free Response Facial Expression Label Sets[C]. //In Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents. Glasgow, UK: ACM, 2020: 1-8.
- [38] Adhikari B, Huttunen H. Iterative Bounding Box Annotation for Object Detection[C]. //In Proceedings of the 25th International Conference on Pattern Recognition (ICPR). Milan, Italy: IEEE, 2021: 4040-4046.
- [39] Ravanbakhsh M, Tschernezki V, Last F, et al. Human-machine collaboration for medical image segmentation[C]. //IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Barcelona, Spain: IEEE, 2020: 1040-1044.
- [40] Marino K, Salakhutdinov R, Gupta A. The More You Know: Using Knowledge Graphs for Image Classification[C]. //2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 20-28.
- [41] Hu Z, Ma X, Liu Z, et al. Harnessing Deep Neural Networks with Logic Rules[C]. //Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany: ACL, 2016: 2410-2420.
- [42] He X, Peng Y. Fine-grained Image Classification via Combining Vision and Language[C]. //2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 7332-7340.
- [43] Song K, Wei X S, Shu X B, et al. Bi-Modal Progressive Mask Attention for Fine-Grained Recognition[J]. IEEE Transactions on Image Processing, 2020, 29(1): 7006-7018.
- [44] Xu H, Qi G, Li J, et al. Fine-grained image classification by visual-semantic embedding[C]. //Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI'18). Stockholm, Sweden: IJCAI, 2018: 1043-1049.
- [45] Liu Z. Entity-Duet Neural Ranking: Understanding the Role of Knowledge Graph Semantics in Neural Information Retrieval[C]. //Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne, Australia: ACL, 2018: 2395-2405.
- [46] Chen X, Li L J, Li F F, et al. Iterative Visual Reasoning Beyond Convolutions[C]. //2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 7239-7248.
- [47] Weber T, Hußmann H, Han Z, et al. Draw with me: Human-in-the-loop for image restoration[C]. //Proceedings of the 25th International Conference on Intelligent User Interfaces. Cagliari, Italy: ACM, 2020: 243-253.
- [48] Ravanbakhsh M, Tschernezki V, Last F, et al. Human-machine collaboration for medical image segmentation[C]. //IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Barcelona, Spain: IEEE, 2020: 1040-1044.
- [49] Murata Y, Dobashi Y. Automatic image enhancement taking into account user preference[C]. //2019 International Conference on Cyberworlds (CW). Singapore: CW, 2019: 374-377.
- [50] Fischer M, Kobs K, Hotho A. Nicer: Aesthetic image enhancement with humans in the loop[C]. //The Thirteenth International Conference on Advances in Computer-Human Interactions. Valencia, Spain: IARIA, 2020: 357-362.

- [51] Kumar V, Smith-Renner A, Findlater L, et al. Why didnt you listen to me? Comparing user control of human-in-the-loop topic models[C]. //Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: ACL, 2019: 6323-6330.
- [52] Fong R C, Scheirer W J, Cox D D. Using human brain activity to guide machine learning[J]. Scientific Reports, 2018, 8(1): 5397-5406.
- [53] Taniguchi H, Sato H, Shirakawa T. A machine learning model with human cognitive biases capable of learning from small and biased datasets[J]. Scientific Reports, 2018, 8(1): 7397-7409.
- [54] Patel B N, Rosenberg L, Willcox G, et al. Human-machine partnership with artificial intelligence for chest radiograph diagnosis[J]. npj Digital Medicine, 2019, 2(1): 111-129.
- [55] Eykholt K, Evtimov I, Fernandes E, et al. Robust physical-world attacks on deep learning models[C]. //2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 1625-1634.
- [56] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. Proceedings of the 32nd International Conference on International Conference on Machine Learning, 2015, 37(1): 448-456.
- [57] 李旭. 基于注意力机制的细粒度图像分类方法研究[D]. 河北: 河北农业大学, 2020.
- [58] Fridman A, Ding L, Jenik B, et al. Arguing machines: human supervision of black box AI systems that make life-critical decisions[C]. //2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Long Beach, CA, USA: IEEE, 2019: 1335-1343.
- [59] Muthuraman K, Reiss F, Xu H, et al. Data Cleaning Tools for Token Classification Tasks[C]. //Proceedings of the Second Workshop on Data Science with Human in the Loop: Language Advances. Online: Association for Computational Linguistics, 2021: 59-61.
- [60] 赵云波, 花婷婷, 赵丽丽, 等. 一种基于多算法集成的分歧介入珍珠分拣方法[P]. CN202110386522.X, 2021-07-23.
- [61] 李革, 李斌, 王莹, 等. 基于 HSL 模型的珍珠光泽度分选方法[J]. 农业机械学报, 2008, 39(6): 113-117.
- [62] Tian C Y. A Computer Vision-Based Classification Method for Pearl Quality Assessment [C]. //2009 International Conference on Computer Technology and Development. Kota Kinabalu, Malaysia: IEEE, 2009: 73-76.
- [63] Li G, Li B, Wang Y, et al. Pearl Shape Recognition Based on Computer Vision[J]. Transactions of the Chinese Society of Agricultural Machinery, 2008, 39(7): 129-132.
- [64] Huang S L, Xu Z, Tao D C, et al. Part-stacked CNN for Fine-Grained Visual Categorization[C]. //2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 1173-1182.
- [65] Zhang H, Xu T, et al. SPDA-CNN: Unifying Semantic Part Detection and Abstraction for Fine-Grained Recognition[C]. //2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 1143-1152.
- [66] 汤一平, 夏少杰, 李陈荣, 等. 基于单目多视角视觉的珍珠品质检测[J]. 农业机械学报, 2014, 45(4): 276-283.
- [67] Xuan Q, Fang W B, et al. Automatic Pearl Classification Machine Based on a Multistream Convolutional Neural Network[J]. IEEE Transactions on Industrial Electronics, 2018 ,65(8): 6538-6547.
- [68] Xuan Q, Chen Z Z, Liu Y, et al. Multiview Generative Adversarial Network and Its

- Application in Pearl Classification[J]. IEEE Transactions on Industrial Electronics, 2019, 66(10): 8244-8252.
- [69] 赵云波, 花婷婷, 王岭人, 等. 一种基于人的分歧介入的珍珠分拣方法[P]. CN202011210928.4. 2021-02-19.
- [70] Songpan W. Improved Skin Lesion Image Classification Using Clustering with Local-GLCM Normalization[C]. //2018 2nd European Conference on Electrical Engineering and Computer Science (EECS). Switzerland, Bern: IEEE, 2018: 206-210.
- [71] 罗建豪, 吴建鑫. 基于深度卷积特征的细粒度图像分类研究综述[J]. 自动化学报, 2017, 43(008):1306-1318.
- [72] Chen T, Lin L, Chen R, et al. Knowledge-Embedded Representation Learning for Fine-Grained Image Recognition[C]. //Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm, Sweden: IJCAI, 2018: 627-634.
- [73] Branson S, Van Horn G, Wah C, et al. The Ignorant Led by the Blind: A Hybrid Human-Machine Vision System for Fine-Grained Categorization[J]. International Journal of Computer Vision, 2014, 108(2): 3-29.
- [74] Wei X S, Xie C W, Wu J X, et al. Mask-CNN: Localizing parts and selecting descriptors for fine-grained bird species categorization[J]. Pattern Recognition, 2018, 76(1): 704-714.
- [75] Zheng H, Fu J, Zha Z J, et al. Learning Rich Part Hierarchies With Progressive Attention Networks for Fine-Grained Image Recognition[J]. IEEE Transactions on Image Processing, 2019, 29(99): 476-488.
- [76] Yu C, Zhao X, Zheng Q, et al. Hierarchical Bilinear Pooling for Fine-Grained Visual Recognition[C]. //Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, et al. Computer Vision - ECCV 2018. Cham: Springer International Publishing, 2018: 595-610.
- [77] Xing W, Yue Z, Gong Y, et al. Grassmann Pooling as Compact Homogeneous Bilinear Pooling for Fine-Grained Visual Classification[C]. //15th European Conference. Munich, Germany: ECCV, 2018: 365-380.
- [78] Selvaraju R R, Cogswell M, Das A, et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization[J]. International Journal of Computer Vision, 2020, 128(2): 336-359.
- [79] 赵云波, 花婷婷, 宋涛, 等. 一种基于多源信息融合的细粒度图像分类方法[P]. CN202110393741.0[P]. 2021-10-19.

致 谢

时间过的很快，转眼间，我的研究生之旅已经接近尾声。我还清晰的记得，当时刚进入校园时对研究生生活的憧憬，对舍友来自何方的好奇，对实验室师兄师姐科研实力的钦佩。现在细细回味研究生这两年多的时光，发现自己也有了许多成长，从最初的科研小白到现在已经拥有自己的论文和专利，还有今年8月出版的人机项目专著，很荣幸能够参与到书本编写的人员名单中。但这些收获不是我一个人的努力，还有身边的导师、同学、家人和朋友的共同帮助，在这里，我由衷地向每一位帮助过我的人表示诚挚地感谢！

首先，我要感谢一直为我的科研工作尽心尽职的导师，赵云波教授。赵老师是我迈向科研道路的指路人，在我研究进展受阻时能够给我提供一些可能性的解决思路，让我能有效地将自己的科研工作进展下去。虽然每周的组会使我惧怕，担心自己的研究成果达不到老师的预期，但赵老师并没有因实验不理想、论文结构混乱而指责我们，而是科学分析可能是哪个环节或哪个想法有问题，及时帮我们改正。同时，针对我这次的毕业论文撰写情况，赵老师也花了很多时间帮我审阅，及时给我指出论文中存在的问题，并提出指导性的意见与建议。如果没有赵老师对我学习生涯的指导与帮助，我也不能顺利完成研究生阶段的科研工作。

其次，我要感谢我实验室的师兄师姐们，在我刚进入实验室这个大家庭时，是他们主动帮助我尽快地融入大家，让我能结识到一群可爱的同学们。他们也会在我遇到科研上的难题时，帮助我一同攻克，提供一些指导意见。清晰地记得，刚开始做实验每次遇到程序 bug 时，师兄们都会不厌其烦地带着我去解决。

再者，我还要感谢一直陪伴着我的朋友们：赵丽丽、袁李婷、杨冰冰、崔奇、宋涛等，感谢你们对我生活上的照顾，也让我的学习生涯增添了不少欢乐。特别是我的两个舍友丽丽和李婷，在我每次感受到科研压力的时候，都会带着我出去散散心，吃吃饭，逛逛街，从而缓解我心中的忧愁，让我梳理好心情再次投入到科研工作中。

最后也是最重要的，非常感谢家人对我默默地支持与陪伴，让我能够安心地学习。也是你们在我最难过的时候，能够立刻给我温暖与鼓励，真心由衷地感谢你们，也希望你们能身体健康，换我陪伴你们长久。

再次，也非常感谢各位评审专家、老师和学者们，在百忙之中审阅我的论文！

作者简介

1 作者简历

1997年02月出生于江苏省淮安市。

2015年9月——2019年6月，宿迁学院机电工程学院自动化专业学习，获得工学学士学位。

2019年9月——2022年1月，浙江工业大学信息工程学院控制工程专业学习，攻读工程硕士学位。

2 攻读硕士学位期间发表的学术论文

[1] 花婷婷,王岭人,赵云波. 利用人的分歧介入增强珍珠自动分拣可靠性研究[J]. 计算机测量与控制, 2021.

3 参与的科研项目及获奖情况

[1] 人机系统中人与机器的自主性边界及其切换策略研究. 科技创新 2030-“新一代人工智能”重大专项课题(2018AAA0100801).

[2] 复杂环境下非完全信息博弈决策的智能基础模型研究. 科技创新 2030-“新一代人工智能”重大专项课题(2018AAA0100801).

4 发明专利

[1] 赵云波,花婷婷,王岭人,唐敏.一种基于人的分歧介入的珍珠分拣方法.中国,CN112381127A[P]. 2021-02-19.

[2] 赵云波,花婷婷,赵丽丽,崔奇.一种基于多算法集成的分歧介入珍珠分拣方法.中国,CN113159150A[P]. 2021-07-23.

[3] 赵云波,花婷婷,宋涛,王岭人.一种基于多源信息融合的细粒度图像分类方法.中国,CN113516156A[P]. 2021-10-19.

[4] 赵云波,唐敏,花婷婷,赵丽丽.一种基于多粒度特征与中层特征的分心驾驶检测方法.中国,CN111695535A[P]. 2020-09-22.

[5] 赵云波,赵丽丽,花婷婷,苏振岭.一种基于人的生理认知特点对苹果精细化分拣的分类方法.中国,CN112668645A[P]. 2021-04-16.

[6] 赵云波,赵丽丽,花婷婷,苏振岭.一种基于人的心理认知模型的青菜病害分类检测方法.中国,CN113283470A[P]. 2021-08-20.

学位论文数据集

密 级*	中图分类号*	UDC*	论文资助
公开	TP11	004.5	
学位授予单位名称	学位授予单位代码	学位类型*	学位级别*
浙江工业大学	10337	工程硕士	全日制专业型硕士
论文题名*	人的介入提升深度学习应用精度的方法及其实例研究		
关键词*	人的介入, 人机混合智能系统, 深度学习, 仲裁机制		论文语种*
并列题名*	无		中文
作者姓名*	花婷婷	学 号*	2111903177
培养单位名称*	培养单位代码*	培养单位地址	邮政编码
浙江工业大学 信息工程学院	10337	杭州市潮王路 18 号	310032
学科专业*	研究方向*	学 制*	学位授予年*
控制工程	人机融合	2.5 年	2022 年
论文提交日期*	2022 年 01 月		
导师姓名*	赵云波	职 称*	教授
评阅人	答辩委员会主席*	答辩委员会成员	
盲评	杨东勇	石崇源、洪榛、阮中远、施朝霞	
电子版论文提交格式: 文本 (☒) 图像 (☐) 视频 (☐) 音频 (☐) 多媒体 (☐) 其他 (☐)			
电子版论文出版 (发布) 者	电子版论文出版 (发布) 地	版权声明	
论文总页数*	61 页		
注: 共 33 项, 其中带*为必填数据, 为 22 项。			