**第一章 需求分析**

**1.1 项目背景与目的**

随着人类活动的不断扩张和生态环境的恶化，许多鸟类面临着濒危和灭绝的威胁。这些濒危鸟类在生态系统中扮演着重要角色，对维护生态平衡和生物多样性具有至关重要的作用。然而，传统的保护方法往往面临着监测和识别方面的挑战，导致保护工作的效果不佳。

本项目旨在利用图像识别技术与对话语言模型，开发一个基于YOLOv8模型的濒危鸟类图像识别系统与关于濒危鸟类保护的智能对话平台，提高实施濒危鸟类保护工作与宣传保护濒危鸟类相关知识的效率和准确性。通过应用图像识别算法，实现对濒危鸟类的监测、识别，为保护者提供及时的数据和情报，从而更好地制定保护策略和采取行动；当提出关于濒危鸟类的问题时，语言模型可以提供详细的回答，以普及保护濒危鸟类的重要性。

**1.2 设计目标**

* 高准确率：追求高识别准确率，以确保识别结果的可靠性。
* 高效率：优化模型以快速处理大量图像，提高用户体验。
* 用户友好：开发简洁直观的用户界面，使非专业人士也能轻松使用。
* 可扩展性：设计灵活的系统架构，便于未来添加更多濒危鸟类或其他功能。

**第二章 概要设计**

**2.1 模块和框架设计**

B/S架构

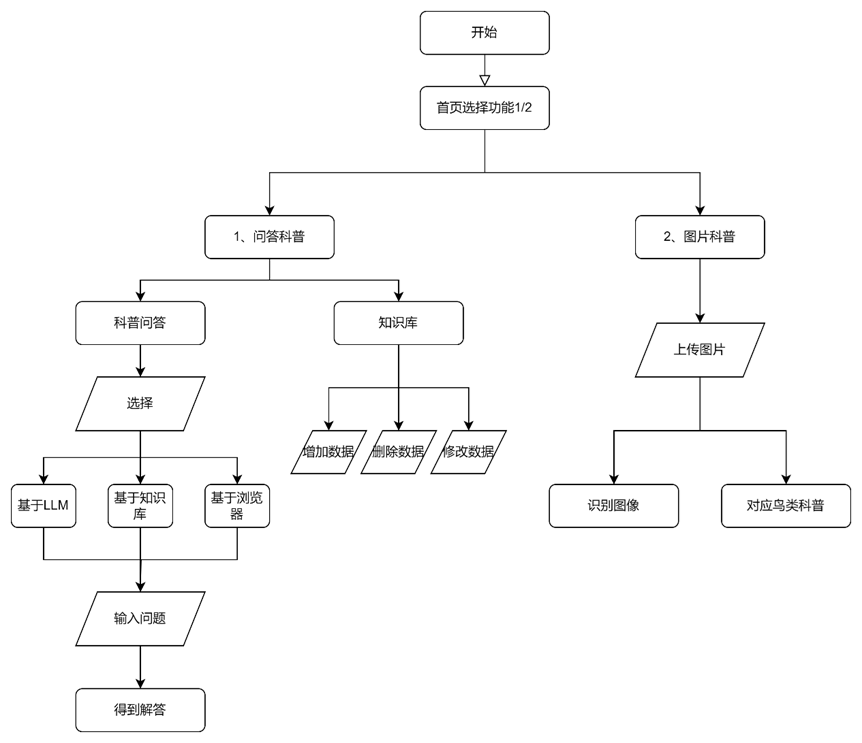
|  |  |  |  |
| --- | --- | --- | --- |
| Server端 | 模块名称 | | 功能描述 |
| 鸟类图像科普 | 图像识别 | 自动识别并展示濒危鸟类的名称、特征等信息 |
| 智能科普 | 基于最新人工智能技术提供动态的濒危鸟类相关知识、保护措施等科普内容 |
| 鸟类综合智能问答 | 基于LLM问答 | 利用LLM处理自然语言查询，提供灵活的问答能力。 |
| 基于知识库问答 | 依托专门构建的鸟类知识库来快速检索信息，对特定、结构化的查询提供精确答案。 |
| 基于浏览器问答 | 维护和更新鸟类知识库，包括添加新的知识点、更新现有信息和优化知识结构。 |
| 知识库管理 | 结合LLM和知识库技术，提供准确、及时的鸟类信息问答服务，支持多语言互动。 |
| Browser端 | 首页 | | 展示项目介绍、团队介绍和用户互动入口。 |
| 关键技术介绍页面 | | 详细介绍项目所采用的关键技术，如图像识别、LLM等。 |
| 开发团队、流程介绍页面 | | 介绍开发团队背景、项目开发流程和未来规划。 |

**2.2 整体框架结构图和整体数据流图**

**第三章 详细设计**

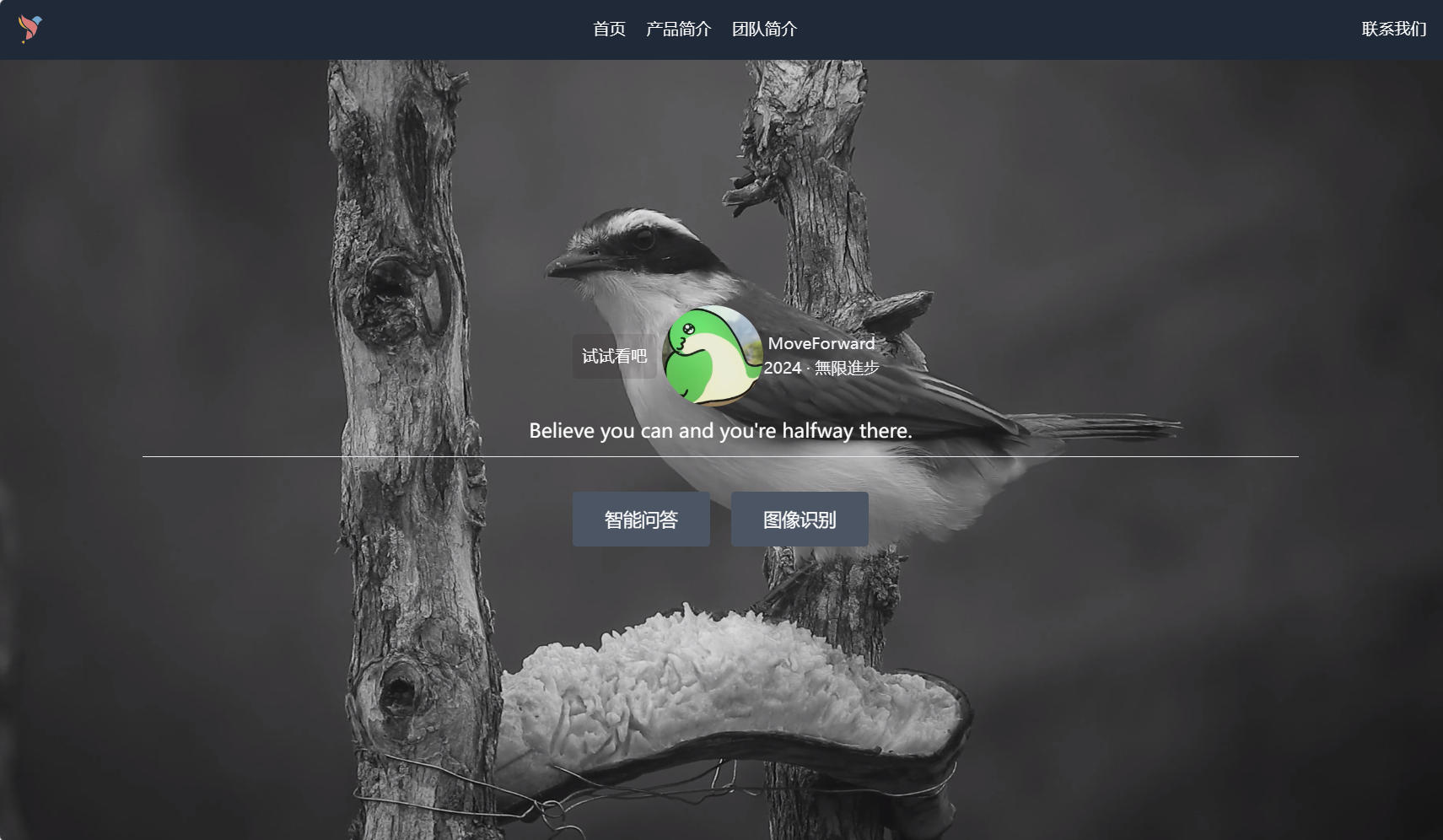
**3.1 用户交互设计**

这个流程不仅简化了用户操作，还确保了信息的准确传递，提高了用户的参与度和满意度。通过集成先进的技术，我们的系统为鸟类保护和科普教育提供了一个强有力的工具，使公众能够更容易地获取有关鸟类的知识，并激发他们对自然保护的兴趣和意识。



**3.1.1使用流程**

在我们的系统中，鸟类问答科普模块提供了一个易于使用且信息丰富的平台，旨在增强公众对鸟类及其保护的认识。用户首先通过浏览器访问系统首页，这是一个利用传统HTML、CSS、JavaScript以及Tailwind CSS和FontAwesome精心设计的响应式界面。



在选择问答科普模块后，用户将进入一个由streamlit构建的简洁前端界面，在此他们可以键入具体的鸟类相关问题。系统提供了三种问答模式以满足不同的用户需求：基于LLM模型的问答、基于知识库的问答以及基于浏览器的问答。

针对用户的问题，系统采用LangChain框架集成的ChatGLM大预言模型和m3e-base词向量切割模型进行处理。在知识库模式下，问题会被转化为向量并与本地向量数据库中的文档进行匹配，以找到最相关的上下文信息，随后利用LLM模型生成准确的答案。对于浏览器模式，系统通过网络搜索来寻找和展示最相关的信息。不论哪种模式，系统均能在用户界面上直接展示详细的答案或提供链接至答案源。

除了提供深入的问答科普模块外，我们还特别设计了一个鸟类图片科普模块，旨在通过视觉识别技术加深用户对鸟类多样性和特征的了解。用户可通过上传鸟类照片来启动这一模块。系统后台使用经过训练的yolov8模型识别图片中的鸟类种类和位置，并在图片上直接标注。此过程不仅展示了先进的图像识别技术，而且还能即时向用户提供关于所识别鸟类的详细科普信息。

一旦识别出鸟类种类，系统将结合预设的提示词和经过微调的ChatGLM模型，自动生成该鸟类的介绍。这包括鸟类的习性、栖息地、保护状况等有价值的信息，旨在教育用户并提高他们对生物多样性的保护意识。此外，为了优化模型的性能并适应不同的运行环境，我们对yolov8模型进行了减枝操作，包括约束训练、剪枝和回调训练（finetune），以确保即使在资源有限的设备上也能保持高效的识别准确率。

**3.1.2 streamlit**

Streamlit是一个用于快速构建数据应用和可视化界面的开源Python库，该项目团队巧妙地运用其高度灵活且易于使用的特性来开发一款集成了LangChain框架与多个大型语言模型的交互式应用程序。通过Streamlit搭建的应用程序前端界面友好，使得用户能够便捷地进行自然语言处理相关的鸟类知识查询，只需输入问题，即可获得由后端LLM精准生成的专业解答。同时，该应用还利用Streamlit的图形化功能实现了对鸟类图片的上传与实时分析。当用户上传图像时，后台运行的YOLOv8模型会迅速识别出图中鸟类的种类信息，并结合ChatGLM模型生成相应的科普内容，这一过程无缝集成于前端展示环节，确保了从图像识别到文本解释的全过程直观、高效且互动性强。

**3.2 综合智能问答模块（数据流）**

**3.2.1 LLM问答**

**3.2.2 基于知识库的问答**

**3.2.3 基于浏览器问答**

**3.2.4 实时知识库创建**

**3.3 接Langchain和大语言模型和词向量切割模型**

**3.4 鸟类图片科普模块（数据流）**

**Yolov8**

**3.1 模型选择**

**3.1.1 YOLOv8**

图像识别技术在鸟类识别领域的应用面临着多种挑战：鸟类种类繁多，不同种类的鸟在外观上可能仅有细微差别，这要求模型能够捕捉到并识别这些微小的特征差异；鸟类图片通常包含复杂的背景，如树木、天空或水面，这些背景可能会干扰模型对鸟类的识别；同一种鸟在不同的姿态和视角下拍摄的图片可能差异较大，给准确识别带来了额外的挑战；不同的光照条件会影响图片的质量和鸟类的颜色表现，增加识别的难度；在一些图片中，鸟类可能只占很小的一部分，这要求模型能够准确检测到小尺寸目标。

在当前的人工智能领域，YOLO模型因其在处理复杂图像识别任务中表现出的显著优势而受到广泛关注。特别是在鸟类图像识别这一具有挑战性的任务中，YOLO通过其高精度、快速处理能力和优秀的泛化性能，显著提升了识别效果。这些优势使YOLO成为研究和应用中的优选模型，尤其是在需要从复杂背景中准确识别众多相似鸟类种类的场景里。如表1所示。

**表1 YOLO优势**

|  |  |
| --- | --- |
| 难点 | YOLO的优势 |
| 多样性和相似性 | YOLO模型凭借其深层的神经网络可以区分细微的特征，即使是外表相似的鸟类也能被准确识别。 |
| 复杂背景 | 采用先进算法优化背景抑制，减少背景噪声干扰，提高对鸟类目标的识别准确性。 |
| 姿态和视角变化 | 模型能够学习鸟类在不同姿态和视角下的表现，提高识别的灵活性和准确性。 |
| 光照条件变化 | YOLO模型能够适应不同光照条件下的图像识别，保持稳定的性能。 |
| 小目标检测 | YOLO模型在提升了速度的同时，也针对小目标检测进行了优化，有助于识别图像中的小型鸟类。 |

与传统的卷积神经网络（CNN）模型相比，YOLO在鸟类图像识别上的表现同样值得关注。传统CNN模型虽然在图像分类等任务中取得了显著成果，但在进行复杂的目标检测和分类任务时，往往需要更多步骤和复杂的处理流程。相比之下，YOLO作为一种端到端的目标检测模型，不仅提高了处理速度，而且通过一次性推理即可完成目标的检测和分类，极大提高了效率和准确性。如表2所示。

然而，随着计算机视觉领域的快速发展，YOLO系列模型也在不断进化。每一次迭代，都带来了新的改进和特性，以适应日益复杂的应用需求。YOLOv8作为该系列中的最新成员，继承了YOLO模型高速和高精度的核心特点，同时引入了创新的技术，以提升模型性能和效率。

|  |  |  |  |
| --- | --- | --- | --- |
| 特征 | YOLOv8 | YOLOv5 | CNN模型 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**3.1.2 ChatGLM**

ChatGLM-6B是一个开放源代码的双语对话语言模型，基于General Language Model(GLM)架构开发，拥有62亿个参数。该模型采用了与ChatGPT类似的技术，并专门针对中文问答与对话场景进行了优化。通过接受大约1T中英文标识符的训练，并结合了监督微调、自我反馈以及基于人类反馈的强化学习技术，ChatGLM-6B展现出了生成符合人类偏好回答的强大能力。

**表4 ChatGLM优势**

|  |  |
| --- | --- |
| 优势 | 描述 |
| 对话优化 | 特别为中文问答和对话场景进行了优化，确保ChatGLM可以理解并生成符合语境和文化相关性的回复 |
| 大规模训练数据 | 经过大约1T标识符的中英文训练，ChatGLM展现了对语言细微差别的广泛理解，这对于生成连贯多样的对话至关重要 |
| 技术先进 | 集成了监督微调、人类反馈强化学习等技术，使ChatGLM能够提供与人类偏好更为一致的回应 |
| 可扩展性和集成性 | 凭借先进的语言模型能力，ChatGLM可以扩展到各种应用程序，从客户服务自动化到互动教育平台 |
| 类人交互 | 采用了基于人类反馈的强化学习，意味着该模型经过训练能生成更具吸引力和类似人类的回复，提升用户体验 |

**3.1.3 m3e-base**

m3e-base主要通过其先进的文本向量化技术支持问答科普和知识库管理功能。m3e-base能够将自然语言文档和用户的查询问题转换成数学向量，这一过程对于理解和处理大量自然语言信息至关重要。特别是在进行问答时，m3e-base通过文本和问句的向量化，以及后续的向量匹配，确保了系统能够快速而准确地找到与用户查询最相关的信息，为大型语言模型（LLM）生成回答提供了准确的上下文。m3e-base的加入，极大提升了问答系统的效率和准确性，使得用户能够获得更为精准和丰富的鸟类保护知识。

|  |  |
| --- | --- |
| m3e-base优势 | 说明 |
| 高效率 | 快速处理文本向量化，确保实时响应用户查询 |
| 精确匹配 | 精准的向量匹配技术提高问答系统的准确性 |
| 适应性强 | 支持多种本地向量数据库，灵活适应不同的存储和检索需求 |
| 增强问答质量 | 为LLM提供准确上下文，生成更贴近用户查询的回答 |

**3.5 数据集制作及模型训练过程**

**3.5.1 鸟类图片数据集制作**

**3.5.2 LLM微调数据集制作**

**3.7 创新点**

**3.7.1 版本控制**

**3.7.2 Langchain**

**3.7.3**

**第四章 测试报告**

**第五章 安装及使用**

**第六章 项目总结**

**参考文献**