



上海交通大学学位论文

工业智能模型的模块化方法设计及实现

姓 名：戚大可

学 号：521021910732

导 师：于晗

学 院：计算机学院

专业名称：软件工程

申请学位层次：学士

2025 年 5 月

**A Dissertation Submitted to
Shanghai Jiao Tong University for the Degree of Bachelor**

**DESIGN AND IMPLEMENTATION OF MODULAR
METHODS FOR INDUSTRIAL INTELLIGENCE
MODELS**

Author: Dake Qi

Supervisor: Han Yu

College of Computer Science and Technology

Shanghai Jiao Tong University

Shanghai, P.R. China

May 30th, 2025

上海交通大学

学位论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，已在文中以适当方式予以致谢。若在论文撰写过程中使用了人工智能工具，本人已遵循《上海交通大学关于在教育教学中使用 AI 的规范》，确保人工智能生成内容的应用场景、引用范围及标注方式均符合规定，并杜绝学术不端行为。本人完全知晓本声明的法律后果由本人承担。

学位论文作者签名：戚大可
日期：2025 年 5 月 30 日

上海交通大学

学位论文使用授权书

本人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。

本学位论文属于：

公开论文

内部论文，保密 1 年 / 2 年 / 3 年，过保密期后适用本授权书。

秘密论文，保密 ____ 年（不超过 10 年），过保密期后适用本授权书。

机密论文，保密 ____ 年（不超过 20 年），过保密期后适用本授权书。

（请在以上方框内选择打“√”）

学位论文作者签名：戚大可

日期：2025 年 5 月 30 日

指导教师签名：于玲

日期：2025 年 5 月 30 日

摘要

随着深度学习技术在工业视觉领域的广泛应用，模型的可解释性和复用性问题逐渐凸显。传统的目标检测模型结构复杂，功能模块界限不清，导致模型重组和功能迁移存在较大障碍，难以满足实际场景下的灵活部署和快速迭代需求。

针对上述问题，本文设计并实现了一个面向工业目标检测模型的模型拆分与模块标注平台。该平台集成了模型结构自动拆分、功能语义标注、热力图可视化和知识图谱管理等功能，旨在提升模型的透明度和模块复用效率，为后续的模型优化和应用提供技术支撑。

本文的主要工作包括：(1) 通过基于特征相似度的 K-means 聚类方法，实现目标检测模型的模块化拆分，明确各功能单元的边界；(2) 设计通道筛选与热力图生成流程，结合大模型语义解析，实现模块功能的自动化标注；(3) 开发集成化平台，支持模型管理、参数设置、热力图查看、语义标注和知识图谱存储等操作，提升模块的管理和检索效率。

实验结果表明，所提出的方法能够有效标注模型各模块的功能，并量化其对检测任务的实际贡献。消融实验显示，移除关键模块后，模型在相关任务上的性能明显下降，验证了模块标注的准确性和平台的实用价值。本研究为目标检测模型的模块化、可解释化和高效复用提供了新的解决思路。

关键词：目标检测模型，模块化拆分，功能标注，热力图可视化，知识图谱平台

ABSTRACT

Modularization of industrial object detection models is essential for improving interpretability and reusability. Traditional deep learning models often suffer from unclear module boundaries and limited flexibility, making model adaptation and reuse challenging in practical scenarios.

To address these issues, this thesis develops a platform for modular decomposition and function annotation of object detection models. The platform integrates automated model splitting, semantic annotation, heatmap visualization, and knowledge graph management, aiming to enhance model transparency and module reuse.

The main contributions are as follows: (1) A K-means-based method for modular decomposition of object detection models, clarifying the boundaries of functional units; (2) An automated function annotation process that combines channel selection, heatmap generation, and large model-based semantic analysis; (3) An integrated software platform supporting model management, parameter configuration, heatmap visualization, semantic annotation, and knowledge graph storage.

Experimental results show that the proposed approach can effectively annotate module functions and quantify their impact on detection tasks. Ablation studies demonstrate that removing key modules leads to significant performance drops, confirming the accuracy of the annotation and the practical value of the platform. This work provides a new solution for modular, interpretable, and reusable object detection models.

Key words: object detection, modular decomposition, function annotation, heatmap visualization, knowledge graph platform

目 录

摘要	I
ABSTRACT	II
第一章 绪论	1
1.1 研究背景	1
1.2 研究现状	2
1.2.1 目标检测模型发展	2
1.2.2 模块可解释性方法	2
1.3 研究意义	3
1.4 研究内容	4
1.5 本文组织结构	4
1.6 本章小结	5
第二章 需求分析及系统框架	6
2.1 业务场景描述	6
2.2 功能性需求分析	7
2.3 非功能性需求分析	9
2.4 系统架构设计	9
2.5 本章小结	10
第三章 面向目标检测的模型拆分及模块标注平台实现	11
3.1 目标检测模型的模块化拆分	11
3.1.1 基于层级相似度的 k-means 聚类	11
3.1.2 模块通道选择策略	13
3.1.3 阈值双权融选的通道筛选算法	16
3.2 拆分模块的语义标注	17
3.2.1 通道热力图的可视化	17
3.2.2 模块的标准功能定义	19

3.2.3	大模型提示词设计	20
3.3	功能模块的结构化存储	21
3.3.1	知识图谱构建	21
3.3.2	模块关系可视化	23
3.4	本章小结	24
第四章 原型展示与实验验证		25
4.1	系统实现及部署	25
4.2	界面展示	26
4.2.1	模型管理界面	26
4.2.2	模型拆分参数设置界面	26
4.2.3	模块拆分结果界面	26
4.2.4	热力图查看界面	28
4.2.5	模块语义标注界面	28
4.2.6	知识图谱界面	28
4.3	实验验证	30
4.3.1	工业数据集 MVTEC	30
4.3.2	模块功能验证实验设计	30
4.3.3	实验结果对比分析	31
4.4	本章小结	33
第五章 结论		34
5.1	工作总结	34
5.2	研究展望	34
参 考 文 献		35
致 谢		37

第一章 绪论

1.1 研究背景

随着智能制造技术的迅猛发展，目标检测模型在工业质检、设备巡检等领域的应用日益广泛。以 YOLO 系列^[1-2]为代表的深度学习目标检测模型，虽然在标准测试集上取得了优异的性能，但在实际工业环境中却面临模型复用困难的挑战。由于不同生产线的检测对象和环境条件存在较大差异，已训练好的模型往往难以直接迁移到新的应用场景，导致频繁的模型重训练，造成了大量计算资源的浪费。

在实际应用中，这一问题尤为突出。当检测环境或检测对象发生变化时，原本表现良好的模型常常出现性能下降，不仅增加了模型维护的成本，也影响了生产线的检测效率。其根本原因在于当前主流目标检测模型采用整体式架构设计，深度神经网络各组成部分通过端到端训练紧密耦合，缺乏明确的功能划分^[3-4]。这种“黑箱”特性使得模型调整时难以精准定位需要修改的部分，工程师往往不得不对整个网络进行重新训练，既增加了工作量，也降低了模型迭代效率。此外，整体式架构还限制了模型的针对性优化，当需要改进某一特定功能时，往往需要调整整个模型，造成不必要的计算资源消耗^[5]，严重制约了目标检测模型在工业场景中的灵活应用^[6]。

尽管迁移学习技术能够通过微调参数适应新场景，但本质上仍需对整个模型进行调整^[7-8]。近年来，模块化方法尝试将网络拆分为功能单元^[9]，但在实际应用中仍存在诸多问题^[10]：（1）模块划分标准过于依赖数学特征相似性，导致拆分后的模块缺乏实际语义对应；（2）模块间缺乏明确的协作规则，重组时容易产生特征冲突，影响检测性能。这些问题使得现有模块化方法在实际工业应用中效果有限，难以真正解决模型复用困难的难题，尤其在复杂的工业检测场景下表现得更加明显。

针对上述问题，本文提出结合功能解耦与语义解析的模块化方法。通过分析网络层的特征响应模式，将检测模型拆分为具有明确功能倾向的模块单元，并利用知识图谱建立模块间的协作关系。该方法有望显著提升模型的复用效率，降低训练能耗，为工业智能模型的灵活应用提供坚实的技术支撑。

1.2 研究现状

本课题旨在实现工业智能模型的模块化方法设计与实现，聚焦于目标检测模型的模块拆分及其功能的可解释性标注。因此，本节将从目标检测模型的发展和模块可解释性方法两个方面综述相关研究现状。

1.2.1 目标检测模型发展

目标检测模型的发展历程见证了计算机视觉领域从传统方法向深度学习方法的重大转变。早期目标检测技术主要依赖手工设计的特征提取方法，如 Viola-Jones 检测器利用 Haar-like 特征和级联分类器实现了实时人脸检测，HOG（方向梯度直方图）特征则在行人检测中表现出良好的稳定性^[11]。尽管这些方法为后续研究奠定了基础，但其特征表达能力有限，难以应对复杂场景下的目标检测需求。

随着深度学习的兴起，目标检测进入了两阶段检测器时代。2014 年提出的 R-CNN 首次将卷积神经网络（CNN）与区域建议相结合，通过选择性搜索生成候选框并提取特征，显著提升了检测精度^[12]。Fast R-CNN 通过 RoI 池化实现特征共享，降低了计算冗余；Faster R-CNN 则引入区域建议网络（RPN），实现了端到端的检测框架。这一阶段的模型虽然精度较高，但计算复杂度较大，限制了其实时性应用。

为兼顾速度与精度，单阶段检测器逐渐成为研究热点。YOLO（You Only Look Once）系列模型将检测任务转化为回归问题，实现了单次前向传播即可完成目标定位与分类。SSD（Single Shot MultiBox Detector）则在多尺度特征图上进行预测，进一步提升了小目标检测能力^[13]。这些方法虽然在一定程度上牺牲了精度，但其高效性使其在工业质检、自动驾驶等实时场景中得到广泛应用。

近年来，Transformer 架构的引入为目标检测带来了新的突破^[14]。DETR（Detection Transformer）摒弃了传统锚框机制，通过自注意力机制建模全局上下文关系，实现了端到端的检测流程。后续如 DINO 等改进模型，通过动态标签分配和去噪训练策略，显著提升了小目标检测性能。这类方法在复杂工业场景中展现出更强的适应性，也为模型的可解释性和功能解耦提供了新的思路。

1.2.2 模块可解释性方法

模块可解释性方法作为目标检测领域的重要研究方向，旨在通过结构化技术手段揭示模型内部功能单元的决策逻辑与协作机制。目前主流的模块可解释性方法主要包括特征可视化、梯度分析和注意力机制三大技术路线。

在特征可视化方面，研究者通过提取神经网络中间层的激活特征图，直观展示不同模块对图像区域的关注重点^[15]。例如，卷积模块的可视化研究发现，浅层模块通常呈现边缘检测特征响应，而深层模块则聚焦于复杂纹理和语义理解。这种方法通过热力图形式将模块功能与视觉特征建立关联，但存在解释粒度粗、语义关联弱等问题。近期有学者将语义分割模型与特征可视化结合，通过 DeepLab 生成的语义区域替代传统图像块分割，使模块关注区域与人类认知中的物体边界更为一致。

梯度分析方法则通过反向传播计算模块对输入特征的敏感度^[16]。在双阶段目标检测模型中，该方法被用于分离分类子网络与回归子网络的解释过程：前者通过梯度权重揭示模块对物体类别的判别依据，后者则量化模块对边界框定位的贡献度。例如在医疗影像分析中，该方法验证了肺部定位模块确实关注胸腔区域的解剖学特征，而非通过背景信息进行错误推断。尽管该方法提升了模块功能的可解释性，但对计算资源需求较高，且易受梯度消失问题影响。

基于注意力机制的可解释性方法近年来取得显著进展^[17]。通过引入自注意力模块，模型在处理特征时自动生成注意力权重图，直观反映各功能模块在空间维度上的信息聚焦过程。在 Transformer 架构的目标检测模型中，该方法有效揭示了模块间通过注意力交互实现多尺度特征融合的内在机制。有研究团队尝试将注意力权重与知识图谱结合，构建模块协作关系的可视化网络，为工程师调整模块组合提供直观依据。然而，注意力图本身仍需二次解释，且难以区分模块功能的主次关系。

在工业检测场景中，模块可解释性方法正朝着功能解耦与语义标注方向发展。有学者提出采用改进的聚类算法对模型进行模块拆分，并结合大语言模型生成标准化功能描述，使拆分后的边缘检测、纹理分析等模块具备人类可理解的语义标签。这种方法在五金零件检测实验中证明，具有明确功能描述的模块重组效率较传统方法显著提升，为工业场景的快速部署提供了新思路。但该方法仍面临模块接口标准化不足、跨场景兼容性有限等技术瓶颈，需要进一步研究模块间的动态协调机制。

1.3 研究意义

在工业质检场景中，目标检测模型常因生产线环境变化而需要频繁调整，传统方法往往需要对整个模型进行重新训练，导致效率低下、资源浪费。针对这一问题，本文提出的模块化方法通过将复杂模型拆分为多个功能模块，并建立模块间的协作规则，为提升模型复用性和适应性提供了新的解决思路。实际应用中，模块化方法能够显著提升模型的复用效率。例如，当检测对象从金属零件切换为塑料制品时，仅需

替换相关模块（如纹理分析模块），而无需重新训练整个模型。这不仅节省了大量计算资源，还大幅缩短了模型迭代周期，帮助技术人员快速响应不同生产线的需求。同时，模块化设计使得模型维护更加便捷，通过可视化工具，工程师能够直观查看各模块的关注区域，快速定位问题并进行针对性优化，降低了对深度学习专业知识的依赖。因此，模块化方法对于提升工业检测模型的灵活性和实用性具有重要意义。

1.4 研究内容

本研究针对工业场景中目标检测模型复用效率低下的问题，提出了基于模块化拆分与大模型语义标注的解决方案，旨在构建一个涵盖模块拆分、功能标注与知识图谱存储的全流程平台。课题主要研究内容包括：

- (1) 实现了基于特征相似度的模型模块化拆分方法，将复杂的目标检测模型有效划分为功能明确的独立模块，显著提升了模型的复用性和灵活性。
- (2) 构建了模块功能的可视化与自动化标注体系，通过热力图展示和大模型语义解析，实现了对各模块功能的直观展示和标准化描述，增强了模型的可解释性。
- (3) 开发了一体化操作平台，集成了模块拆分、功能标注和知识图谱管理等功能，为工业检测模型的高效管理、快速重组和知识积累提供了有力支撑。

1.5 本文组织结构

本文围绕工业智能模型的模块化方法设计与实现展开研究，全文共分为五章，各章节主要内容如下：

第一章为绪论，介绍了工业场景下目标检测模型复用难题的研究背景，分析了目标检测模型与模块化技术的发展现状，明确了本研究的理论意义与实际价值，并对全文结构进行了总体概述。

第二章为需求分析与系统设计，通过典型工业检测场景的业务需求分析，提炼了系统的功能性与非功能性需求，完成了从数据流到功能模块的整体架构设计，为后续技术实现提供了框架指导。

第三章为核心算法与实现，详细阐述了模型模块化拆分、通道筛选、热力图生成、语义标注及知识图谱构建的完整方法流程。重点介绍了基于特征相似性的模块拆分策略、双权重通道筛选机制，以及结合大模型的功能标注技术。

第四章为系统实现与实验验证，展示了原型系统的主要界面和功能模块，并通过

工业检测数据集对模块化方法的有效性进行了验证，分析了模块拆分精度、语义标注准确率等关键指标，并讨论了实验结果带来的技术启示。

第五章为总结与展望，总结了本研究在模型模块化方法上的创新成果，指出了当前方案在跨场景适配性等方面不足，并对模块接口标准化、动态协调机制等未来研究方向提出了建议。

1.6 本章小结

本章介绍了工业场景中目标检测模型复用难题的研究背景，指出了传统整体式模型存在的“黑箱”特性和复用效率低下等问题。通过分析目标检测模型和模块化技术的发展现状，总结了现有方法在功能解耦与语义解析方面的不足。在此基础上，提出了结合特征分析与知识图谱的模块化方法，并规划了从需求分析到实验验证的技术路线。最后，对全文的章节结构进行了简要说明，为后续研究内容的展开奠定了基础。

第二章 需求分析及系统框架

本章将围绕工业智能模型模块化方法的需求展开分析，并提出相应的系统框架设计。首先结合实际的工业质检场景，描述目标检测模型在应用过程中遇到的问题。随后，从功能性需求和非功能性需求两个方面，对系统应具备的主要功能和性能要求进行梳理。最后，基于需求分析结果，设计出本系统的整体框架结构，为后续的具体实现提供基础和参考。

2.1 业务场景描述

在实际的工业质检过程中，目标检测模型经常需要面对多样化且不断变化的检测任务。例如，在一条生产线上，可能既要检测金属零件表面的划痕，也要检测塑料外壳的装配偏差，甚至还要识别电子元件的焊点缺陷。传统的做法通常是针对每一种缺陷分别训练一个模型，这样不仅模型数量多，维护起来也比较复杂。当生产线更换新产品时，由于材料特性或零件尺寸的变化，原有模型可能会出现漏检或误检，技术人员往往需要重新采集数据并调整模型参数，这一过程既耗时又费力，重复性也很高。

目前常用的整体式目标检测模型，其内部各部分功能紧密耦合。当检测需求发生变化时，工程师很难准确定位到需要优化的具体功能模块，只能通过不断尝试调整网络结构或参数。例如，在金属零件检测中，如果因为光照变化导致纹理识别效果变差，传统方法往往需要重新训练整个模型，而无法单独优化纹理分析部分。这种“牵一发而动全身”的调整方式，不仅增加了维护难度，也影响了模型的迭代效率和适应能力。

随着检测任务的不断丰富，模型的灵活性和可扩展性变得越来越重要。传统整体式模型在面对新的检测需求时，往往难以及时做出调整，导致适应新任务的周期较长。实际操作中，模型结构复杂、功能边界不清晰、调试过程依赖经验等问题也比较突出。

模块化方法为上述问题提供了一种新的解决思路。通过将目标检测模型拆分为边缘提取、纹理分析、形状检测等功能明确的模块，技术人员可以有针对性地替换或优化某些模块。例如，当检测对象从金属零件变为透明塑料制品时，只需替换对材质

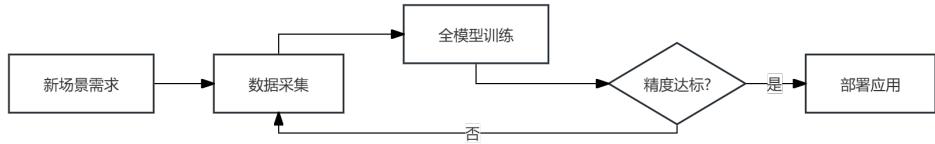


图 2-1 现有业务场景

反光特性敏感的纹理分析模块，而形状检测模块则可以继续使用。模块化设计不仅提升了模型的复用效率，也降低了模型调整和维护的难度。

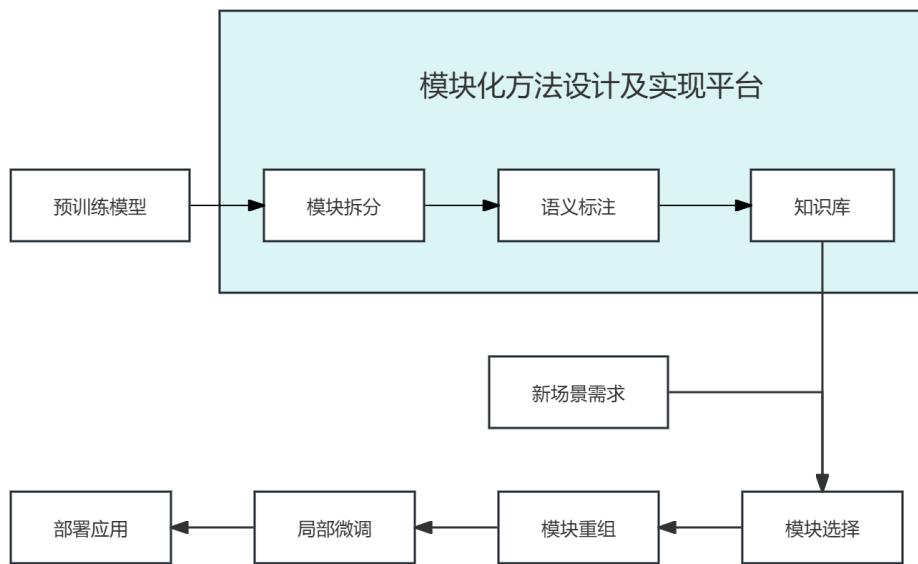


图 2-2 基于模块化方法的业务场景

此外，知识图谱等技术可以用来记录模块之间的功能关系，帮助技术人员理解模块的组合逻辑，实现模块的高效管理和智能推荐。通过“即插即用”的模块化方案，目标检测模型能够更快地适应不同的检测任务，提升了系统的灵活性和可维护性。

2.2 功能性需求分析

本小节结合面向目标检测模型模块化的业务场景，对系统的功能性需求进行分析。平台的用例图如图2-3所示。

本系统的功能性需求包括模型模块化拆分、通道筛选、热力图生成、语义自动化标注、知识图谱构建和可视化平台操作。用户可以上传目标检测模型，平台支持自动

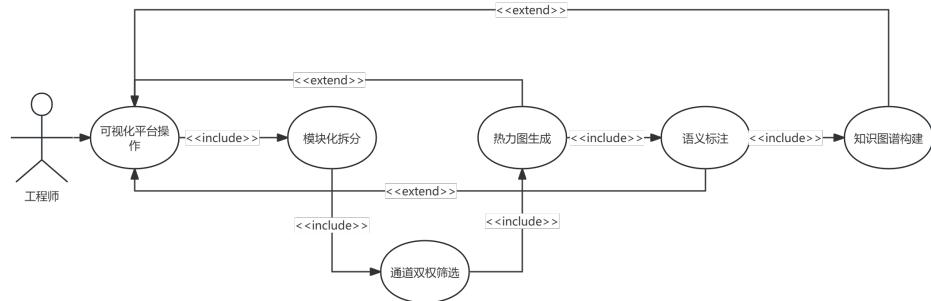


图 2-3 模块化方法设计及实现的平台用例图

将模型划分为如边缘检测、纹理识别等独立功能模块，并可查看和重组各模块。系统通过通道双权筛选，分析通道响应强度和特征重要性，筛选关键通道，为热力图生成和功能分析提供支持。用户能够直观查看各模块生成的热力图，了解其关注区域。平台还利用大语言模型对热力图进行解析，自动生成标准化功能标签，降低理解难度。知识图谱用于存储模块间的功能关联和适配规则，便于查询和重组。所有操作均可通过 Web 界面完成，实现模型和模块的便捷管理。

表 2-1 模块化方法设计及实现平台用例说明

用例名称	用例说明
模型模块化拆分	基于 K-means 算法将目标检测模型拆分为独立功能模块，解决传统模型局部调整困难的问题
通道双权筛选	从响应强度与特征重要性两个维度筛选关键通道，为模块功能分析提供输入数据
热力图生成	将通道特征响应映射到检测图像形成关注区域热图，直观展示模块的视觉关注重点
语义自动化标注	通过大语言模型解析热力图模式，生成符合实际场景的模块功能描述标签
知识图谱构建	存储模块间的功能关联与适配规则，为模块选择与重组提供知识支撑
可视化平台操作	基于 Web 界面实现模型上传、模块拖拽重组、图谱检索等功能的图形化交互

2.3 非功能性需求分析

系统需要具备良好的流畅性和稳定性，保证模块拆分、热力图生成等主要功能能够在较短时间内完成，避免因等待时间过长影响用户的正常使用。界面应当简洁直观，支持图形化拖拽、参数滑块等交互方式，简化操作流程，并配有必要的操作提示，帮助用户更快地熟悉系统。热力图和知识图谱的可视化展示应当支持层级切换和重点内容的聚焦，确保信息展示清晰明了。

在系统架构方面，需要具备较好的可维护性和扩展性。采用模块化设计有助于后续功能的升级或算法的替换。数据存储要保证模块信息和协作规则的安全与完整，防止因异常操作导致数据丢失或损坏。系统还应兼容常见的目标检测模型格式，能够支持不同规模的网络进行模块化处理，并为后续功能扩展预留技术接口，以适应未来技术的发展需求。

2.4 系统架构设计

本项目采用 MVC (Model-View-Control) 设计模式进行系统架构设计，如图2-4所示。

模型层 (Model) 为系统提供基础数据支持，内容包括工业数据集、目标检测模型库、模块拆分结果、模块语义标注结果以及知识图谱的结构化数据。这些数据为模型的训练、模块的拆分与标注、知识图谱的构建等功能提供了保障。

控制层 (Control) 承担系统的核心服务功能，包含模型拆分模块、语义标注模块和数据存储模块。模型拆分模块实现层级相似度聚类和通道阈值双权融选等算法，用于对模型进行模块化处理。语义标注模块负责热力图的可视化以及基于大语言模型的功能标注，提升模块的可解释性。数据存储模块用于知识图谱的构建和模块关系的可视化，便于后续的查询和管理。

视图层 (View) 为用户提供操作界面，涵盖模型管理、模块拆分参数配置、模块拆分结果查看、通道热力图展示、模块语义标注查看以及知识图谱的浏览。用户可以通过图形化界面对模型和模块进行管理、配置和结果查看，提升了系统的易用性和交互体验。

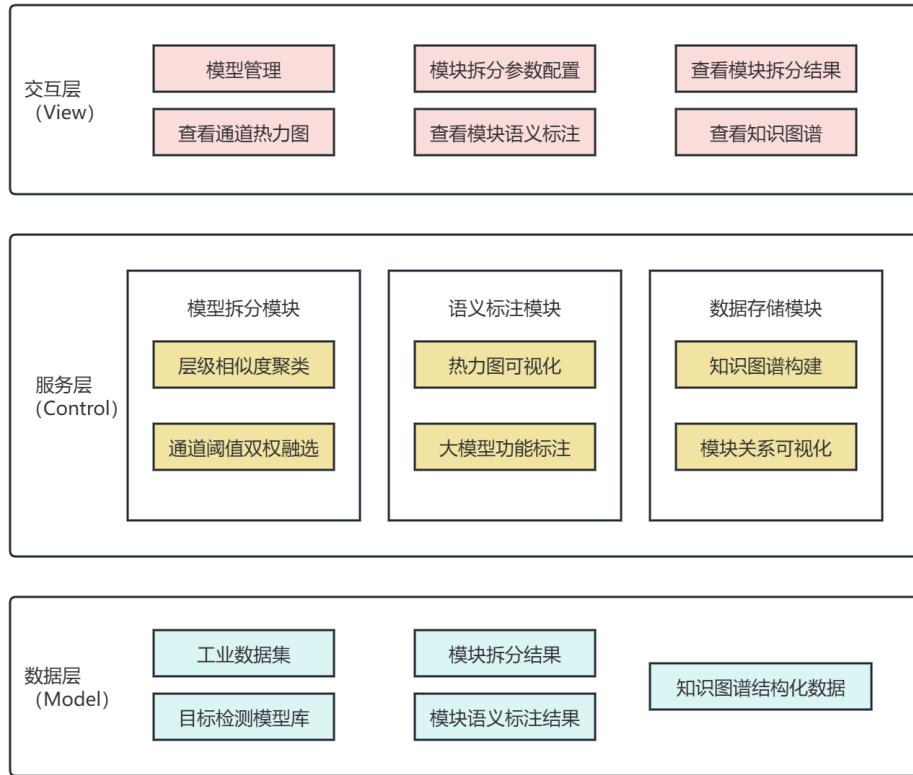


图 2-4 模块化方法设计及实现平台架构图

2.5 本章小结

本章围绕工业智能模型的模块化需求进行了系统设计，分析了传统整体式模型在实际质检场景中存在的复用难题，并提出了模块化方法作为改进思路。内容涵盖了模型拆分、特征可视化、语义标注和知识图谱构建等核心功能需求，同时对系统性能、用户体验、可维护性和兼容性等非功能性需求进行了说明。在此基础上，设计了基于 MVC 架构的系统框架，明确了各层次的功能分工和数据流转关系。通过本章的需求分析与架构设计，为后续模块化方法的具体实现和验证提供了理论基础和技术支持。

第三章 面向目标检测的模型拆分及模块标注平台实现

模块化方法设计及实现平台的基本流程如图3-1所示。用户通过系统界面选择数据集和预训练模型，设置模块拆分参数后，系统会加载模型结构，并利用改进的 K-Means 算法对网络层进行聚类，自动划分出功能明确的子模块。系统对每个模块的激活通道进行筛选，生成热力图以展示模块对输入图像的关注区域，并结合大语言模型自动生成语义功能标签，用户可对标签进行校验和修正。最终，标注后的模块信息被存储到知识图谱数据库，支持模块的查询和管理，为模型的重组和优化提供数据基础，实现了模块的积累与复用。

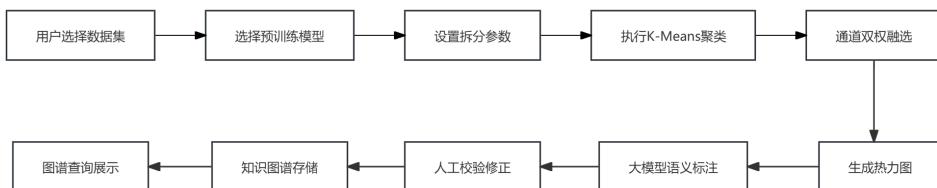


图 3-1 模块化方法设计及实现平台的基本流程

3.1 目标检测模型的模块化拆分

3.1.1 基于层级相似度的 k-means 聚类

为实现网络层的功能解耦，本方法采用特征动态捕获、图像标准化处理和层间相似度计算等步骤对模型进行模块拆分。

特征动态捕获机制能够在卷积层前向传播时自动触发钩子函数，实时获取网络输出的多维特征张量。具体实现时，利用 PyTorch 的前向钩子，将指定卷积层的输出特征从 GPU 显存转移到内存，并按通道维度分离存储。这样可以方便后续对每个通道特征的独立分析，为模块拆分提供细粒度的数据基础。

```

# 钩子函数捕获输出特征
def hook_fn(module, input, output):
    feat = output.detach().cpu().numpy() # 特征张量转 Numpy
    for c in range(feat.shape[1]): # 按通道维度分离
        self.channel_features[c].append(feat[:, c, :, :])
  
```

图像标准化处理用于统一模型的输入格式。原始输入图像会被调整为 320×320 像素，消除分辨率差异带来的影响；BGR 色彩空间会转换为 RGB 格式，以适配预训练模型的输入要求；像素值归一化后，0-255 的整型像素值被映射到 0.0-1.0 的浮点区间，提升数值计算的稳定性。标准化后的三维张量（通道 × 高度 × 宽度）作为模型输入，保证不同来源图像的特征提取具有可比性。

```
processed_img = mmcv.imresize(img, (320, 320)) # 尺寸标准化
processed_img = mmcv.imconvert(processed_img, 'bgr', 'rgb') # BGR 转 RGB
processed_img = processed_img.astype(np.float32) / 255.0 # 归一化
```

层间相似度的计算采用余弦相似度作为主要指标。通过公式 (3-1)，将每层输出的多维特征展平成一维向量，计算特征向量之间的夹角余弦值。

$$\text{Similarity}(f_i, f_j) = \frac{f_i \cdot f_j}{\|f_i\| \|f_j\|} \quad (3-1)$$

其中 f_i 、 f_j 分别表示第 i 、 j 层的特征向量。

实际实现时，采用矩阵运算批量处理特征向量，构建 $N \times N$ 的相似度矩阵。该矩阵可以定量描述各网络层在特征响应模式上的相似程度，相似度越高，说明两层在功能上更为接近。相似度矩阵作为聚类算法的输入，为后续的模块划分提供量化依据。

在模块化分组过程中，还包括特征标准化、质心初始化和聚类后处理等环节。

特征标准化阶段针对网络层特征向量的量纲差异，采用 L2 范数归一化，将原始特征投影到单位超球面空间。这样可以消除特征幅值差异对距离度量的影响，使欧氏距离与余弦相似度等价，保证聚类过程能够准确反映网络层间的功能关系。

$$\hat{f}_i = \frac{f_i}{\max(\|f_i\|_2, \epsilon)}, \quad \epsilon = 10^{-7} \quad (3-2)$$

其中 \hat{f}_i 表示归一化的特征向量， f_i 表示原始特征向量， $\epsilon = 10^{-7}$ 是一个防止数值不稳定的小常数。

归一化过程中引入极小常数作为分母保护项，避免零向量导致的数值异常。

```
norms = np.linalg.norm(features, axis=1, keepdims=True)
normalized_features = features / np.maximum(norms, 1e-7)
```

质心初始化采用 K-Means++ 算法，通过概率密度函数动态选择初始质心。该方法先随机选取第一个质心，后续质心根据样本与已有质心的最小距离平方值按概率分布选取。这样可以让初始质心分布更均匀，减少聚类陷入局部最优的风险。实际操作中，系统会多次独立初始化，选取目标函数值最小的聚类结果作为最终解，以提升聚类的稳定性和可重复性。

聚类后处理阶段包括质心映射和离群点过滤。归一化空间的聚类质心会被反向映射到原始特征空间，恢复特征幅值的实际意义。

$$C_k^{\text{orig}} = c_k^{\text{norm}} \times \frac{1}{N} \sum_{i=1}^N \|f_i\|_2 \quad (3-3)$$

其中 C_k^{orig} 表示原始的聚类中心， c_k^{norm} 表示归一化后的聚类中心， N 表示样本总数。

随后根据通道特征与所属质心的余弦相似度设定过滤阈值，剔除相似度低于 $\theta = 0.65$ 的离群通道。

```
filtered = [c for c in channels if cosine_sim(c, centroid) > 0.65]
```

这种过滤机制有助于消除异常特征对模块划分的干扰，保证同一聚类内的网络层具有较高的一致性。最终聚类结果作为模块划分的依据，为后续的模块重组和功能标注提供结构化数据基础。

3.1.2 模块通道选择策略

模块通道选择主要包括平均值、最大值和中位数三种策略。平均值方法是将所有通道的激活值取算术平均，生成热力图。这种方式能够反映模块整体的关注趋势，但容易掩盖关键通道的显著特征。实际操作中，系统将所有通道的激活矩阵在通道维度上求和并除以通道总数，得到单通道的热力分布图。该方法适合观察整体响应，但在复杂检测任务下，细节特征容易被平均效应削弱。其计算公式为：

$$H_{\text{avg}}(x, y) = \frac{1}{C} \sum_{c=1}^C F_c(x, y) \quad (3-4)$$

其中 $H_{\text{avg}}(x, y)$ 表示位置 (x, y) 处的平均热力图值， C 表示特征通道数， $F_c(x, y)$ 表示第 c 个通道在位置 (x, y) 处的特征响应值。

实际实验中，平均值策略生成的热力图整体较为均匀，缺乏明显的重要区域，如图3-2所示。

最大值策略则关注高响应通道的特征表现。系统会计算每个通道的全局平均响应值，按从高到低排序，选取前 5 个通道，将这些通道的激活值加权叠加生成热力图。

```
selected = sorted_indices[:max(5, len(channel_means)//20)] # 取 TOP5
```

这种方法能够突出模块对显著特征的关注区域，但有时会忽略一些次要但重要的细节。实验中，热力分布容易过度集中，整张热力图颜色偏红，如图3-3所示。

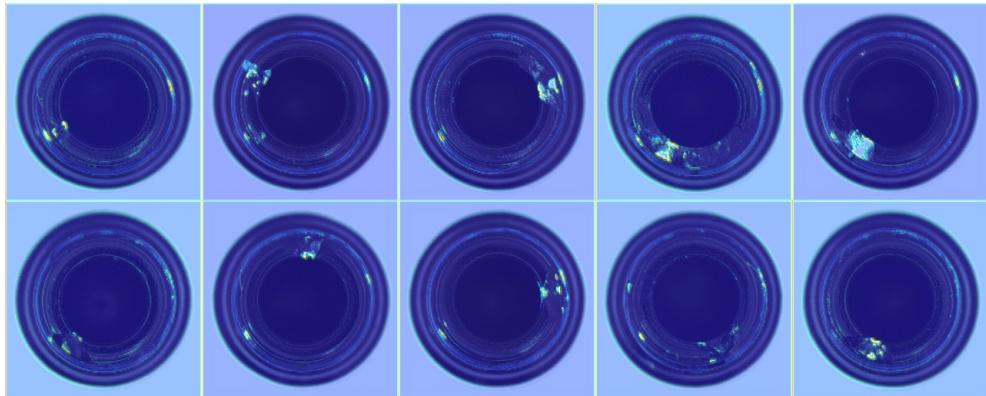


图 3-2 平均值策略的热力图

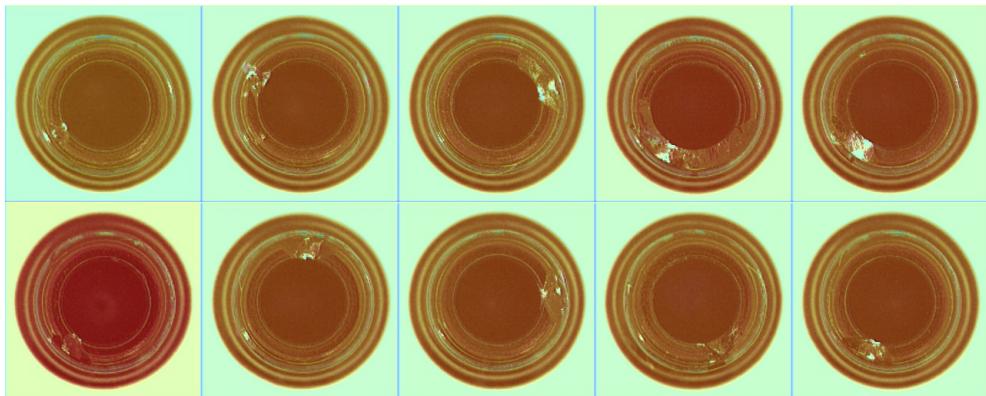


图 3-3 最大值策略的热力图

中位数策略是一种折中方法，选取响应强度处于中间区段的通道。系统对通道响应值排序后，截取排序中间位置附近的 5 个通道，取其激活值的平均值生成热力图。

```
mid = len(sorted_indices)
selected = sorted_indices[max(0,mid-2):min(len(sorted_indices),mid+3)] # 取中间 5 通道
```

这种方式可以排除极端高或低响应通道的影响，平衡特征表达的显著性和完整性，具有一定的抗噪能力。但有时会丢失部分关键特征。实际效果一般，重点区域不够明显，如图3-4所示。

三种策略的对比实验显示，不同方法在特征保留和抗干扰能力上各有优缺点，为后续的双权融选策略提供了参考。

此外，SHAP 值分析可以量化通道特征对检测结果的贡献度，为通道筛选提供理论依据。该方法基于 Shapley 值理论，将每个通道视为参与者，通过边际贡献计算其

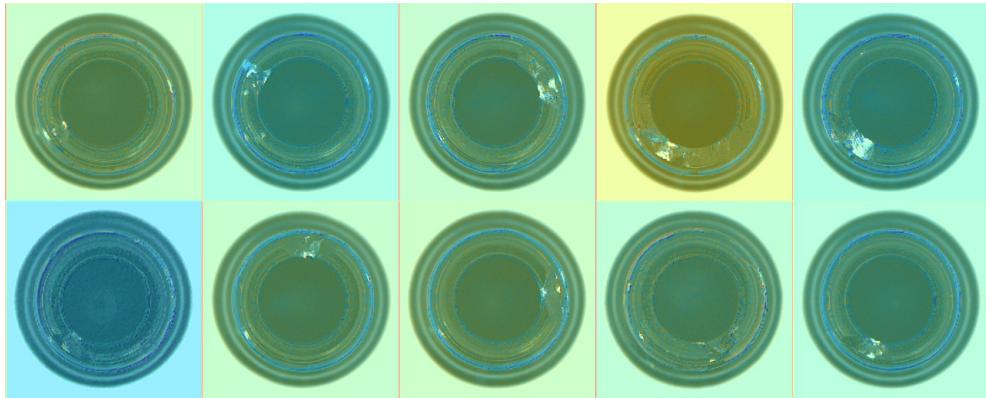


图 3-4 中位值策略的热力图

重要性。通道 c 的 SHAP 值定义为：

$$\phi_c = \frac{1}{N} \sum_{i=1}^N (f(S_i \cup \{c\}) - f(S_i)) \quad (3-5)$$

其中 S_i 为随机采样的通道子集， $f(\cdot)$ 为模型输出响应函数。

实际实现时，系统会随机采样多个通道子集，分别计算包含和不包含目标通道时的模型输出差异，多次采样平均后得到通道的 SHAP 值。虽然这种方法能较为客观地反映通道的稳定贡献，但计算量较大，通常需要用梯度积分法加速。

在通道重要性计算阶段，系统对输入图像进行多次特征扰动，记录各通道 SHAP 值的绝对响应均值，并整合空间维度信息，得到通道的全局重要性评分。该评分可以克服单次分析的随机波动，更稳定地反映通道在不同场景下的平均贡献。重要性评分高的通道通常具有较强的特征判别能力。

```
shap_values = explainer.shap_values(input_img, nsamples=50) # 采样 50 次
channel_importance = np.mean(np.abs(shap_values), axis=(1, 2, 3)) # 维度整合
```

该指标通过平均绝对 SHAP 响应值反映通道的稳定贡献，克服了单样本分析的随机性。

动态通道筛选策略能够根据通道的重要性分布自动选择参与热力图生成的通道。系统会将所有通道按照重要性得分从高到低排序，选取排名前五的通道用于后续的热力图融合，这样既能突出关键特征，又能减少无关信息的干扰。在实际实验过程中，基于 SHAP 值得到的热力图通常呈现出非常鲜明的色彩分布，红色和蓝色区域界限分明，对比度较强，如图3-5所示。

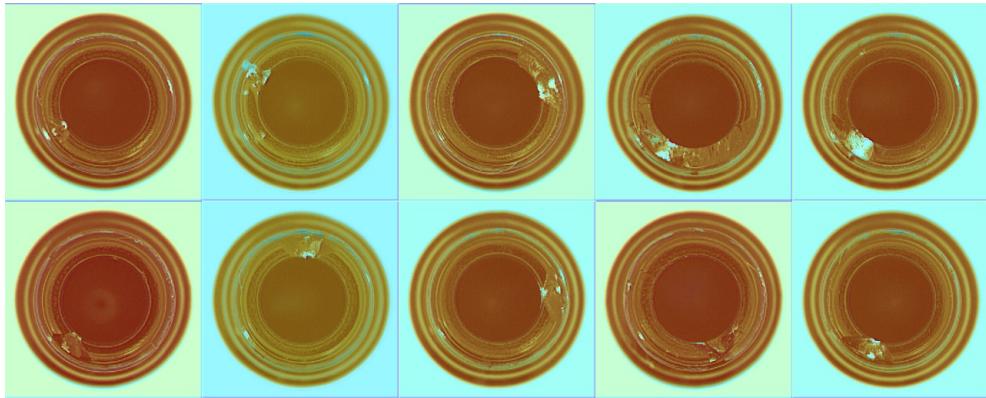


图 3-5 SHAP 值策略的热力图

3.1.3 阈值双权融选的通道筛选算法

本方法采用分阶段的筛选流程，以提升通道选择的准确性和有效性。在初步筛选阶段，依据通道响应值的统计分布，设置动态阈值区间。具体做法是计算所有通道响应均值的百分位数，去除响应最低的 30% 和最高的 10% 通道。响应较低的通道通常包含较多噪声信息，而响应过高的通道可能导致特征过曝。通过这一处理，可以缩小候选通道的范围，为后续筛选提供更可靠的基础。过滤条件如下：

$$\text{Percentile}(R_i, 30\%) < R_c < \text{Percentile}(R_i, 90\%) \quad (3-6)$$

其中 R_c 表示通道 c 的响应均值。

```
low_cutoff = int(len(channels) * 0.3)      # 去掉最低 30%
high_cutoff = int(len(channels) * 0.9)       # 去掉最高 10%
valid_indices = sorted_indices[low_cutoff:high_cutoff]
```

在综合评分阶段，采用双权融合机制，将通道的响应强度和特征重要性（SHAP 值）结合。两项指标均经过最小-最大归一化，消除量纲影响。最终的综合评分由归一化响应值和归一化 SHAP 值各占一半权重，既能突出细节特征，也能兼顾全局判别能力。综合得分的计算方式如下：

$$S_c = 0.5 \cdot \tilde{R}_c + 0.5 \cdot \tilde{H}_c \quad (3-7)$$

其中 \tilde{R}_c 为归一化响应值， \tilde{H}_c 为归一化 SHAP 值。

```
norm_response = (response - min_r) / (max_r - min_r + 1e-8)    # 响应归一化
norm_shap = (shap - min_h) / (max_h - min_h + 1e-8)           # SHAP 归一化
combined = 0.5 * norm_response + 0.5 * norm_shap             # 双权融合
```

在优选阶段，系统根据综合评分对通道进行排序，自动选取得分最高的 5 个通道作为最终结果。对于小目标检测等特殊场景，可以适当调整通道数量，以获得更好的细节表现。

```
top_indices = np.argsort(combined_scores)[-n_channels:][::-1] # 降序选取
selected_channels = valid_indices[top_indices] # 映射回原始通道索引
```

最终选中的通道通过加权叠加生成热力图。实验结果表明，双权融合策略能够更清晰地突出关注区域的边界，特征聚焦也更加合理，验证了该方法的有效性，如图3-6所示。

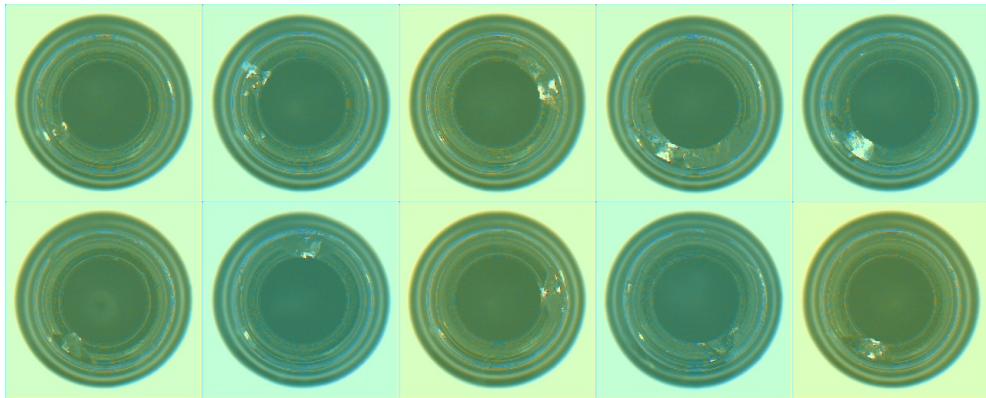


图 3-6 双权融选策略的热力图

3.2 拆分模块的语义标注

3.2.1 通道热力图的可视化

特征加权融合阶段通过对筛选得到的通道激活值进行线性叠加，能够增强关键区域的视觉显著性。具体做法是将选定通道的二维特征图在通道维度上直接求和，得到原始热力分布矩阵：

$$H(x, y) = \sum_{c \in V} F_c(x, y) \quad (3-8)$$

其中 $H(x, y)$ 表示位置 (x, y) 处的热力图值， V 表示选定的特征通道集合， $F_c(x, y)$ 表示第 c 个通道在位置 (x, y) 处的特征响应值。

采用这种不加权重的融合方式，可以保持特征响应的原始空间分布，避免人为设定权重带来的主观影响，使热力图能够真实反映模块的注意力分布。对于多通道显著

响应的复杂场景，叠加后的热力矩阵能够融合不同通道的关注点，形成更全面的可视化效果。

极差归一化映射用于提升热力图的对比度和可视化效果。通过计算热力矩阵的全局最小值和最大值，将原始数值线性映射到 0-255 的标准灰度范围：

$$\hat{H} = 255 \times \frac{H - H_{\min}}{H_{\max} - H_{\min} + \epsilon} \quad (3-9)$$

其中 \hat{H} 表示归一化后的热力图， H 表示原始热力图， H_{\max} 和 H_{\min} 分别表示热力图中的最大值和最小值， $\epsilon = 1 \times 10^{-8}$ 用于防止分母为零。

归一化处理不仅能让低响应区域保持可辨识的灰度差异，还能避免高响应区域出现过饱和。归一化后的热力矩阵可以直接与原始输入图像叠加，通过颜色映射算法生成红蓝渐变的热力图。红色区域表示模块高度关注的位置，蓝色区域则代表低响应区域。标准化处理为不同模块热力图的对比分析提供了统一的基准。

为实现热力图与原始检测图像的协同可视化，本方法采用多步处理流程。首先，应用 JET 色标进行颜色空间映射，将单通道热力灰度值转换为伪彩色编码。颜色映射函数定义为：

$$C(h) = \text{JET}(|255 \times h|) \quad (3-10)$$

其中 $C(h)$ 表示热力图的彩色可视化结果， h 表示输入的热力图值。

JET 色标通过蓝-绿-黄-红的渐变色谱对应 0-255 的强度区间，红色表示高响应，蓝色表示低响应。这样可以显著提升人眼对响应强度差异的辨识度，使模块关注区域的分布特征更加直观。

针对热力图与原始图像分辨率不一致的问题，采用双线性插值算法进行尺寸对齐。插值过程中，热力图在上采样到原始图像尺寸时，像素值通过与周围像素的加权平均获得，能够保证边缘的平滑性和空间连续性。插值公式如下：

$$I'(x, y) = \sum_{i,j} w_{ij} I(\lfloor x \cdot s \rfloor + i, \lfloor y \cdot s \rfloor + j) \quad (3-11)$$

其中 $I'(x, y)$ 表示插值后图像在位置 (x, y) 处的像素值， $I(x, y)$ 表示原始图像在位置 (x, y) 处的像素值， s 表示缩放因子， w_{ij} 表示插值权重系数。

该方法能够有效消除特征图下采样带来的分辨率损失，使热力图与原始图像的细节特征精确对齐。

最终，采用透明度混合技术实现热力图与原始图像的叠加显示。热力图层设置为 60% 的透明度，原始图像保留 40% 的可见度。每个像素的 RGB 通道分别进行加权求

和，生成既包含语义信息又保留场景细节的融合图像。该方法能够突出模块的注意力分布，同时保留原始图像内容，有助于分析特征响应与实际检测目标之间的关系，为后续的功能分析提供直观的视觉依据。

$$\text{Blended} = 0.6 \times \text{Original} + 0.4 \times \text{Heatmap} \quad (3-12)$$

3.2.2 模块的标准功能定义

为提升大模型在语义标注任务中的准确性，本研究结合工业检测场景的实际需求，制定了 12 类模块的标准功能类别（见表3-1）。这些类别涵盖了目标检测模型的主要功能层次。基础特征模块主要负责底层视觉特征的提取，包括边缘检测、纹理识别、颜色感知和光照适应等功能；目标解析模块侧重于目标的定位和几何分析，涉及物体检测、边界框回归、关键点检测和姿态估计等任务；语义理解模块则用于更高层次的推理，包括语义分割、场景分类、关系建模和多模态融合等能力。

本研究通过规范术语体系，减少了语义上的歧义。例如，将“边缘附近的高频信号处理”统一归为“边缘检测”，避免出现“轮廓提取”“边界感知”等不统一的表述。结构化的分类体系有助于大模型更快地定位模块的核心功能属性。例如，当模块输出特征与纹理模式密切相关时，可以直接归类为“纹理识别”，而不是模糊地描述为“特征分析”。

此外，功能层级的划分有助于大模型理解各模块之间的协作关系。基础特征模块为上层模块提供几何和材质信息，目标解析模块依赖这些信息实现精确定位，语义理解模块则整合多层次特征进行最终决策。通过这种层级化的体系，大模型能够建立“边缘检测 → 物体检测 → 异常识别”的逻辑链条，更准确地描述各模块在检测流程中的具体作用，从而提升功能解释的可靠性。

表 3-1 模块标准功能

边缘与基础感知	目标检测与结构分析	语义与高层推理
边缘检测	物体检测	语义分割
纹理识别	边界框回归	场景分类
颜色感知	关键点检测	关系建模
光照适应	姿态估计	多模态融合

3.2.3 大模型提示词设计

本系统在进行语义标注时，采用了分步引导和层级化的提示词设计。通过结构化输入，可以有效提升大语言模型对模块功能的解析能力。提示词会先给出模块的基本信息，包括所属的目标检测模型类型、处理的图像类别以及通道的平均响应强度等参数，帮助模型建立必要的上下文。

在核心功能标注环节，提示词限定了 12 种标准功能选项，要求模型只能在如”边缘检测””纹理识别”等预设类别中选择最合适的功能标签。这样可以避免模型给出过于宽泛或不准确的答案，保证标注结果的规范性。

判断依据部分，提示词强调结合可观测的视觉特征进行分析。模型需要关注热力图中高响应区域的空间分布，以及在不同检测场景下的响应稳定性。通过空间分布和统计规律的结合，模型能够将抽象的特征表达转化为具体的视觉证据，减少主观判断的影响。

```
% prompt 模版
这是xx模型的xx模块
请根据这10张热力图，分别对应类别xxx, xxx, ……，通道平均响应度为xx, xx, ……
执行以下操作：

1. 核心功能标注
为每个模块生成1个主要功能标签，格式：
"[类别]_[功能]_模块"
功能只能从12个标准功能里选择

2. 判断依据要求
- 关键关注区域（如：人/交通工具/户外物品）
- 跨图像响应模式（如：在8/10图中对动物有强响应）

3. 输出格式
- 功能标签：[__]_[__]_模块
- 置信度：(0-1)
```

输出格式采用模板化设计，确保标注结果的统一性。功能标签采用三段式命名，例如”苹果_颜色感知_模块”，依次包含检测对象、功能类别和模块类型。置信度分數则要求模型根据特征匹配程度给出 0 到 1 之间的量化值。标准化的输出不仅便于后续知识图谱的自动构建和模块检索，也有助于形成完整的语义标注流程。

1. 类别: apple

• 热力图观察 :

- 热力图显示了多个场景中的苹果，苹果区域被显著激活（蓝色高亮）。
- 苹果的形状和颜色在图像中被清晰识别。
- 关键关注区域是苹果本身及其表面纹理。

• 功能判断依据 :

- 边缘检测：苹果的轮廓边界被清晰地识别出来。
- 纹理识别：苹果表面的纹理特征（如斑点和凹凸）被有效捕捉。
- 颜色感知：苹果的颜色（通常是红色或绿色）在图像中被准确感知。
- 物体检测：苹果作为一个独立的目标被成功检测出来。

• 主要功能选择 :

- 根据上述分析，最核心的功能是物体检测，因为模块能够准确识别并定位苹果这一目标。

• 置信度评估 :

- 响应度为0.8218，表明模块对苹果的检测非常敏感且稳定。
- 跨图像响应模式一致，几乎所有图像中的苹果都被正确识别。

• 功能标签 :

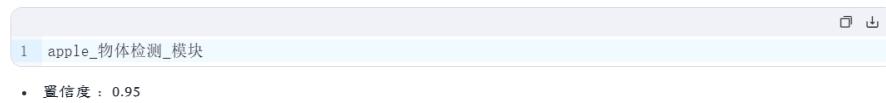


图 3-7 大模型标注结果示例

3.3 功能模块的结构化存储

3.3.1 知识图谱构建

本课题构建的知识图谱整体结构如图3-8所示，包含17个模型节点和69个模块节点。知识图谱中还包括类别节点和功能节点，四类主要实体节点共同构成了完整的知识体系。模型节点用于记录目标检测框架的基本信息（如YOLOv3），模块节点表示模型中经过拆分的功能单元，类别节点对应检测目标的语义分类（如“bottle”、“grid”），功能节点则用于标注模块的标准功能类型。系统采用动态节点管理机制，在新建节点时会自动检查缓存区是否已存在同名节点，若已存在则直接复用，避免重复创建带来的数据冗余。这种设计不仅节省了数据库空间，也保证了节点语义的一致性。

知识图谱中的关系建模采用三层结构。模型节点与模块节点之间通过“包含”关系相连，反映模型的组成结构；模块节点与类别节点之间通过“响应值”关系关联，记录模块对不同检测目标的激活强度；类别节点与功能节点之间则通过“置信度”关系连接，表示语义标注的可信度。以模块节点为例，其与不同类别节点之间的响应值可以反映该模块在不同检测任务中的表现差异。例如，某模块对“bottle”类别的响应值

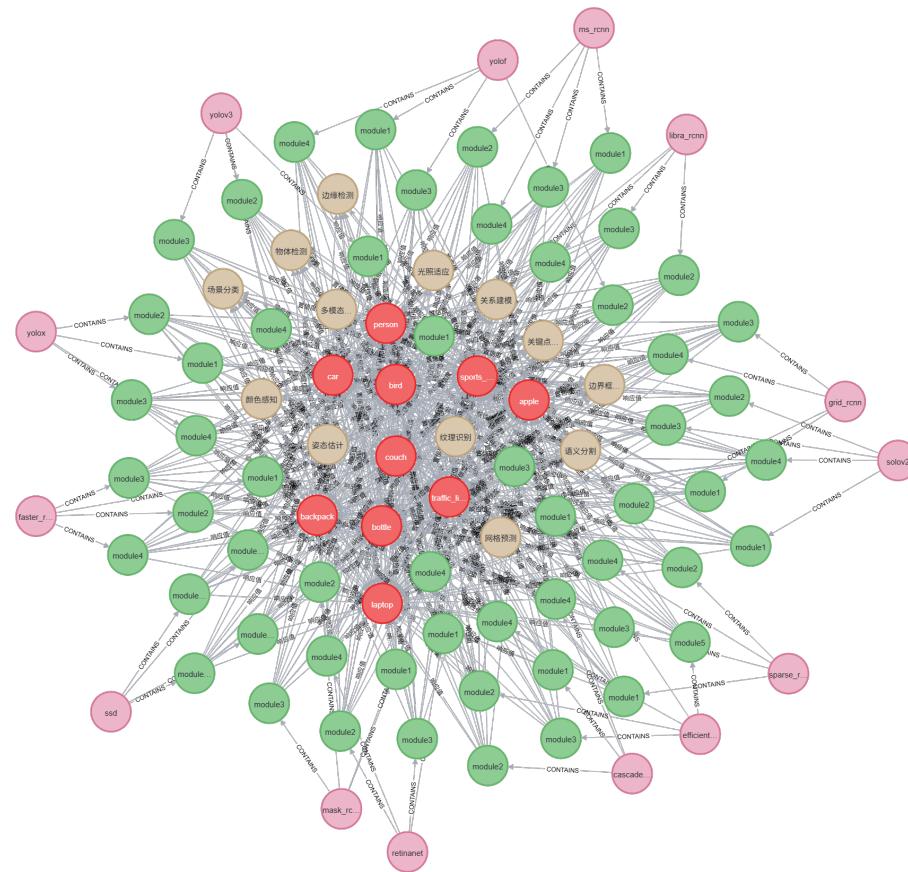


图 3-8 模块化方法设计及实现的知识图谱

表 3-2 知识图谱节点与关系类型说明

类型	名称	说明
节点	模型 (Model)	表示目标检测框架的基本信息，如 YOLOv3 等
节点	模块 (Module)	表示模型中拆分得到的功能单元
节点	类别 (Category)	对应检测目标的语义分类，如”bottle”、”grid”等
节点	功能 (Function)	标注模块的标准功能类型，如”边缘检测”、”纹理识别”等
关系	包含 (CONTAINS)	连接模型与模块，表示模块属于某一模型
关系	响应值 (response_level)	连接模块与类别，记录模块对不同检测目标的激活强度
关系	置信度 (confidence)	连接类别与功能，表示语义标注的可信度

为 0.85，说明其在该类别检测中具有较高的特征敏感性。

```
# 建立模型与模块的包含关系
graph.create(Relationship(model_node, "CONTAINS", module_node))
# 记录模块对类别的响应值
graph.create(Relationship(module_node, "响应值", category_node, response_level=0.85))
# 记录类别与功能的置信度
graph.create(Relationship(category_node, "置信度", function_node, confidence=0.92))
```

为提升节点检索效率，系统采用内存字典作为缓存。在节点创建和查询过程中，优先访问内存缓存而非直接操作数据库，将节点检索的时间复杂度从线性降低到常数级。

表3-3展示了部分模块在不同检测类别下的响应强度。例如，某模块对”bottle”类别的响应为 0.85，而对”grid”类别仅为 0.72。这种量化差异为模块功能分析提供了客观依据。结构化的存储方式使得模块与功能、检测目标之间的关联关系更加清晰和可追溯，为后续模型管理和分析提供了数据基础。

表 3-3 功能模块的结构化存储

model name	module name	module hierarchy	category name	response level	function name	confidence
yolov3	module2	2.1-2.2	bottle	0.6017	边缘检测	0.85
yolov3	module2	2.1-2.2	carpet	0.5774	纹理识别	0.78
yolov3	module2	2.1-2.2	grid	0.5831	物体检测	0.72

3.3.2 模块关系可视化

本系统采用动态视觉编码方法，突出知识图谱中关键的关联关系。在模块与检测类别的响应强度分析中，系统利用图数据库查询语言，自动识别每个检测类别中响应值最高的模块。具体实现时，系统检索所有模块与检测类别之间的响应强度数据，通过聚合运算筛选出每个类别对应的最大响应值。对于这些高响应关系，系统在可视化时将连接线条设置为红色并加粗，便于用户一目了然地发现对特定检测目标影响显著的模块。

```
MATCH (m:Module)-[r:响应值]->(c:Category)
WITH c, MAX(r.response_level) AS max_val    # 获取最大响应值
SET r.highlight='red', r.line_width=3          # 设置红色视觉特征
```

在功能标注置信度的可视化方面，系统会自动遍历模块与功能类别之间的置信度关系。通过排序筛选出每个功能类别中置信度评分最高的路径，并采用绿色渐变色

进行高亮显示。此类高置信度关系的线条宽度也会加粗至 3 像素，与普通关系的浅灰色细线形成明显对比。通过这种颜色和线宽的区分，知识图谱中的可靠功能路径能够更加直观地呈现，方便用户在人工核查时快速定位重点信息。

```
MATCH (c:Category)-[r:置信度]->(f:Function)
WITH c, MAX(r.confidence) AS max_conf      # 获取最大置信度
SET r.highlight='green', r.line_width=3    # 设置绿色视觉特征
```

3.4 本章小结

本章围绕目标检测任务，介绍了模型拆分与模块标注平台的具体实现方法。内容涵盖了从用户配置、算法处理到模块存储的完整流程。模型拆分部分，采用了基于层级相似度的 k-means 聚类方法，并结合通道筛选和阈值双权融合策略，实现了对目标检测模型的有效分解。在模块语义标注方面，利用热力图可视化技术直观展现了各模块的功能特性，结合标准功能定义和大模型提示词，实现了模块功能的自动标注。知识图谱部分，构建了模块功能的结构化存储体系，通过实体节点、关系建模和视觉编码，形成了模块、功能与应用场景之间的可视化知识网络。整体流程为后续模型的优化和适配提供了理论基础和方法支持。

第四章 原型展示与实验验证

本章围绕所开发的目标检测模型模块化管理与知识图谱系统，展开原型功能的展示，并结合实验对系统性能进行验证。内容涵盖各主要功能界面的实际运行效果，以及在典型数据集和模型上的实验分析。通过系统展示与实验结果，能够直观体现平台的应用能力和性能表现。

4.1 系统实现及部署

本系统采用前后端分离的架构设计。前端部分使用 JavaScript 语言，基于 React 框架开发，并集成了 Ant Design 组件库对界面进行美化和布局优化。Antd 提供了丰富的 UI 组件，使得系统界面更加美观、交互更加友好。用户可以通过前端页面进行模型配置、模块管理、热力图查看等操作。

后端部分采用 Python 语言，基于 Flask 框架实现，主要负责处理前端请求、业务逻辑和数据接口。后端集成了 PyTorch 和 MMDetection 框架，支持目标检测模型的加载、推理和模块拆分等功能。PyTorch 作为主流的深度学习框架，提供了灵活的模型开发和训练环境；MMDetection 则为目标检测任务提供了丰富的模型库和工具支持。后端还实现了与数据库的数据交互，为前端提供统一的数据访问接口。

知识图谱数据库采用 Neo4j 进行存储和管理。Neo4j 作为图数据库，能够高效地表示和查询实体之间的复杂关系，适合用于存储模块、功能、类别等多种节点及其关联。系统通过后端与 Neo4j 进行数据交互，实现了知识图谱的动态更新和可视化查询。

由于条件限制，系统的前端、后端以及数据库均部署在本地计算机上。各部分通过本地网络进行通信，保证了系统的基本功能和数据交互的顺利进行。整体架构具有良好的扩展性和可维护性，能够满足目标检测模型模块化管理和知识图谱构建的需求。

4.2 界面展示

4.2.1 模型管理界面

模型管理界面是系统的重要组成部分，主要用于展示和管理当前可用的目标检测模型及数据集。界面中会列出所有已集成的模型，包括模型名称、简介和参数量等信息，方便用户了解和对比不同模型。同时，用户还可以查看和选择可用的数据集，并支持上传本地模型文件和数据集。用户在界面上选择目标模型和数据集后，可直接点击“开始拆分”按钮，进入模型拆分与模块分析流程，整个操作过程简洁高效，提升了系统的易用性。

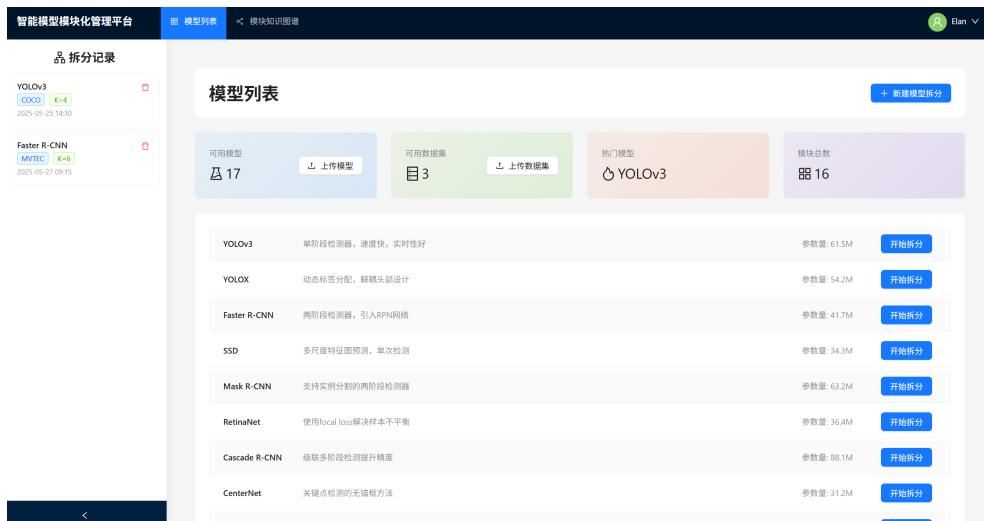


图 4-1 模型管理界面

4.2.2 模型拆分参数设置界面

在用户选择好目标模型和数据集后，系统会进入模型拆分参数设置界面。该界面主要用于配置模型拆分过程中所需的关键参数，包括分区数量 K 、运行次数 n_init 以及最大迭代次数 max_iter 。用户可以根据实际需求调整这些参数，以获得更合适的拆分效果。参数设置完成后，点击“开始拆分”按钮即可启动模型拆分流程。界面设计简洁明了，方便用户快速完成参数配置和操作。

4.2.3 模块拆分结果界面

模块拆分结果界面用于展示模型拆分后的各个模块信息。界面中会列出每个模块的编号、对应的层级信息以及相关操作选项。用户可以在操作栏中选择“生成热力



图 4-2 模型拆分参数设置界面

图”，以可视化方式查看某个模块的热力图特征，也可以点击“语义标注”对指定模块进行功能标注。此外，界面右上方还提供“一键语义标注”功能，支持对当前模型下所有模块进行批量标注，并集中展示标注结果。该界面设计便于用户直观管理和分析模块拆分的具体情况。



图 4-3 模块拆分结果界面

4.2.4 热力图查看界面

热力图查看界面为用户提供了模块响应特征的可视化展示。在该界面中，用户可以查看当前模块对每个数据集类别的响应度。点击某一类别的“查看热力图”按钮后，界面下方会显示该类别对应的热力图，帮助用户直观了解模块在不同类别上的激活情况。该功能有助于分析模块的特征分布和作用效果。

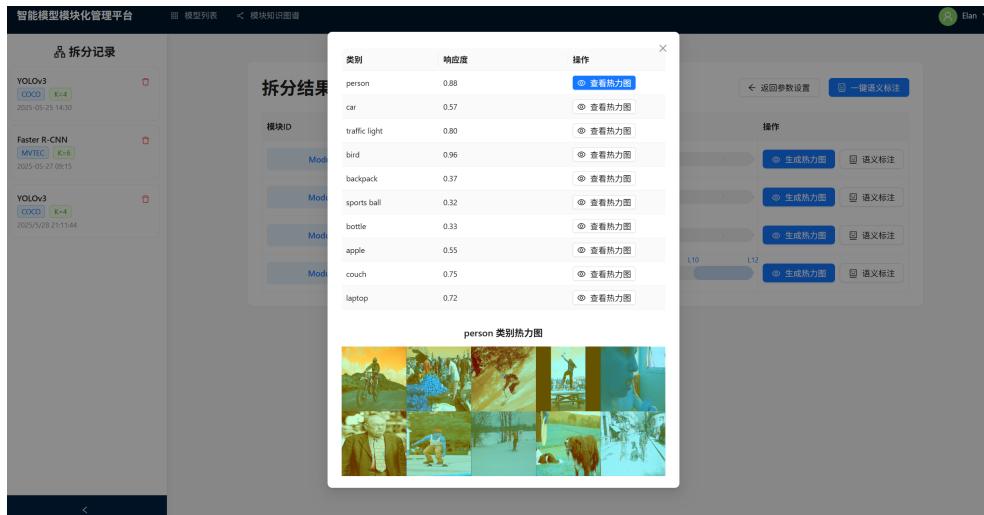


图 4-4 热力图查看界面

4.2.5 模块语义标注界面

模块语义标注界面用于展示模块的功能标注结果。图 4-5 显示的是单个模块的标注结果页面，左上角展示了该模块的基本信息，下方依次列出了各类别的响应度、对应的功能标签以及置信度分数，便于用户详细了解该模块在不同类别下的表现和功能归属。

图 4-6 展示的是整个模型所有模块的标注结果页面。界面中为每个模块选取了置信度最高的功能进行集中展示，使用户能够快速把握模型整体的功能分布情况。

4.2.6 知识图谱界面

知识图谱界面为用户提供了模块标注结果的可视化展示方式。用户可以在该界面选择感兴趣的模型，系统会以知识图谱的形式呈现该模型下各模块的功能标注结果。通过节点和关系的图形化展示，用户能够直观地了解模块、功能、类别等实体之间的关联结构，便于对模型的整体功能分布和模块间关系进行分析。

Module 2 功能分析

层级信息: Layer 2.1 - 2.3

类别	响应度	功能	置信度
person	0.85	目标定位	0.88
car	0.57	特征提取	0.72
traffic light	0.89	特征提取	0.74
bird	0.80	实例分割	0.79
backpack	0.74	特征提取	0.90
sports ball	0.55	语义分割	0.78
bottle	0.23	实例分割	0.89
apple	0.80	目标定位	0.96
couch	0.39	边缘检测	0.96
lantern	0.40	特征提取	0.81

图 4-5 单个模块语义标注界面

模块语义标注

YOLOv3 + COCO (K=4)

使用大模型对所有拆分模块进行语义理解和功能标注，生成易于理解的命名和描述。

模块ID	功能类型	功能描述	语义化命名	置信度
Module 1	边缘检测	负责图像的边缘检测和轮廓提取...	yolov3_module1_边缘检测	77%
Module 2	特征提取	从图像中提取关键特征，包括...	yolov3_module2_特征提取	81%
Module 3	目标定位	根据提取的特征确定目标位置...	yolov3_module3_目标定位	71%
Module 4	分类决策	对检测到的目标进行分类，并...	yolov3_module4_分类决策	89%

存入数据库并查看知识图谱

图 4-6 全部模块语义标注界面

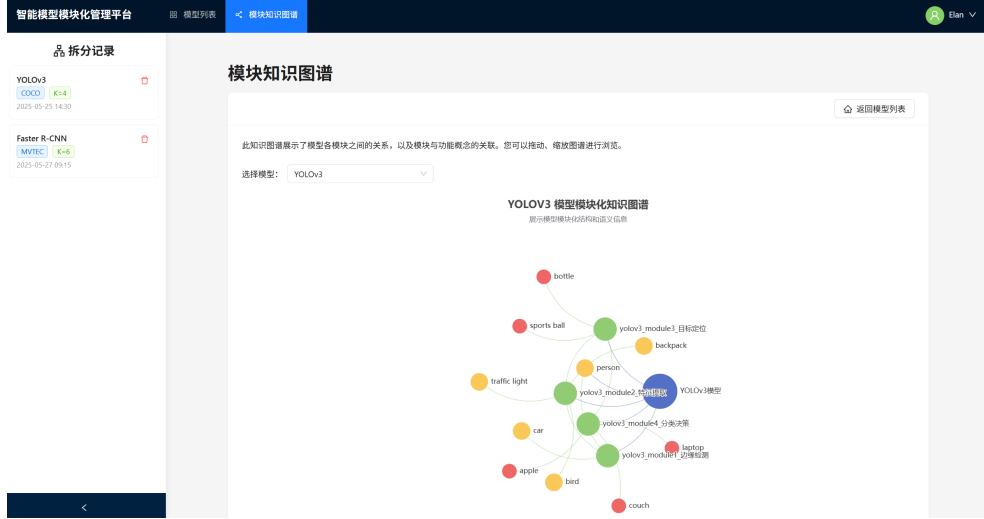


图 4-7 知识图谱界面

4.3 实验验证

4.3.1 工业数据集 MVTEC

MVTEC 数据集是一套专为工业视觉检测任务设计的数据集，涵盖了多种常见工业产品类别，如金属零件、纺织品和电子元件等。该数据集包含了丰富的正常样本和多种类型的缺陷样本，能够较好地反映实际工业场景中的检测需求。图像分辨率较高，细节信息充足，适合用于评估目标检测和缺陷识别等相关算法的性能。

4.3.2 模块功能验证实验设计

本实验围绕模块语义标注功能展开。根据知识图谱，“module1”的语义标注为“边缘检测”，因此选用 MVTEC 数据集中的 bottle 类图片，重点考察该模块对边缘检测任务的实际作用。实验设计分为两部分：一方面，通过缺陷检测数量和平均置信度的变化，评估去除 module1 后模型整体检测性能的变化；另一方面，利用 Canny 算法对检测结果的边缘区域进行可视化，分析模型在边缘特征提取方面的能力变化。

检测数量差异 ΔN 用于衡量模型在去除模块前后，检出缺陷目标数量的变化，定义如下：

$$\Delta N = N_{\text{orig}} - N_{\text{ablate}}$$

其中， N_{orig} 表示原始模型检测到的缺陷数， N_{ablate} 表示消融后模型的检测数。

平均置信度差异 ΔC 用于反映模型对检测结果的置信度变化，计算方式为：

$$\Delta C = \frac{1}{K} \sum_{k=1}^K \left(C_{\text{orig}}^{(k)} - C_{\text{ablate}}^{(k)} \right)$$

其中, K 为两次检测中都被识别出的缺陷目标数量, $C_{\text{orig}}^{(k)}$ 和 $C_{\text{ablate}}^{(k)}$ 分别为第 k 个目标在原始模型和消融模型中的置信度。

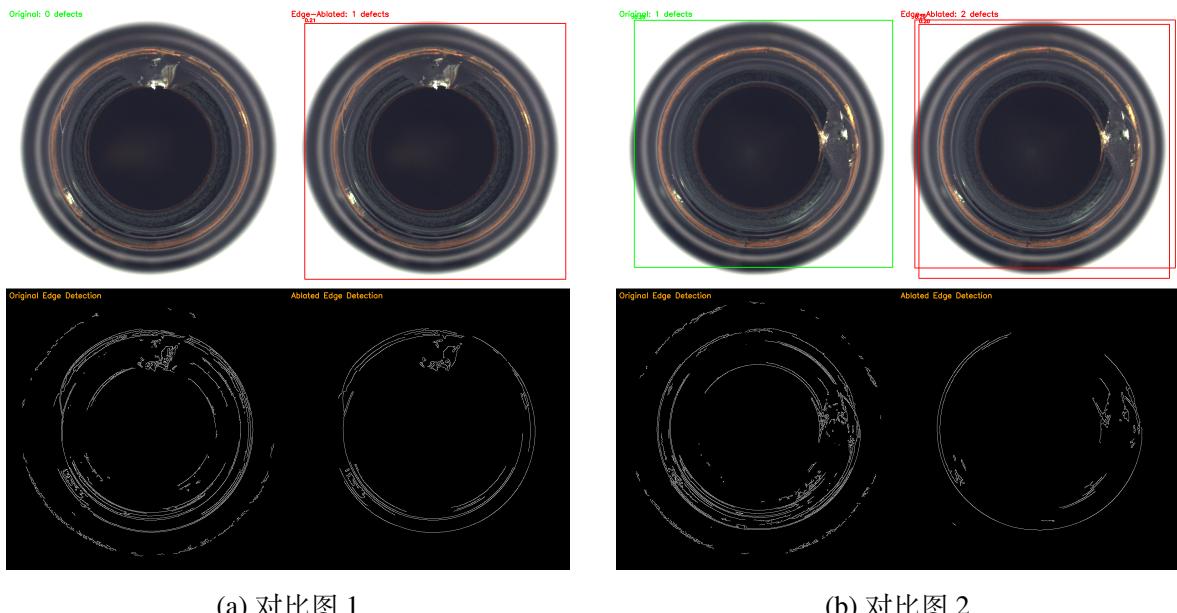
实验过程中, 采用绿色和红色框体叠加的方式展示检测结果, 便于直观比较模型在缺陷检测和边缘提取方面的表现差异。通过上述定量和可视化手段, 能够较为全面地分析 module1 对模型整体功能和边缘检测能力的具体影响。

4.3.3 实验结果对比分析

本节对 10 个模块的消融实验结果进行了详细分析。实验发现, 绝大多数模块在被移除后, 模型在标注功能的识别能力和置信度方面都出现了不同程度的下降。这一现象表明, 各模块在模型整体性能中发挥着不可忽视的作用。部分模块消融后置信度略有提升, 主要原因在于边缘特征减弱后, 模型对缺陷区域的细节捕捉能力下降, 导致缺陷区域在视觉上变得更加突出, 从而提升了置信度分数。整体来看, 模块的存在有助于模型更全面地理解和表达图像中的细节信息, 缺失某些模块会影响模型对特定特征的把握。

以 module1 为例, 消融实验进一步揭示了该模块在边缘检测任务中的重要性。去除 module1 后, 模型对缺陷区域的整体识别能力变化不大, 主要缺陷依然能够被检测出来, 说明模型的基础检测能力尚可维持。然而, 在边缘细节的表现上, 消融模型与原始模型存在明显差异。具体表现为, 消融模型的边缘轮廓变得模糊, 部分细小结构和裂纹无法完整还原, 边缘的连续性和精细度也有所降低。由此可见, module1 对于缺陷边缘的细致提取具有重要意义。图 4-8 展示了原始模型和消融模型在同一缺陷样本上的边缘检测效果, 从视觉上直观反映了两者在边缘特征提取方面的差异。

表 4-1 总结了 10 个模块消融实验的具体结果。从表中可以看出, 不同模块的消融对模型性能的影响存在差异, 但整体趋势较为一致。去除功能模块后, 模型在对应标注功能上的能力普遍出现明显下降, 这一结果进一步验证了语义标注的准确性。消融分析不仅帮助理解各模块在模型中的具体作用, 也为后续模型结构的优化和完善提供了有价值的参考。



(a) 对比图 1

(b) 对比图 2

图 4-8 边缘检测模块消融实验对比图

表 4-1 边缘检测模块消融实验结果

Module	Original	EdgeAblated	CountDiff	PercentageChange	ConfidenceDiff
1	6	9	-3	-50.00%	-0.0482
2	11	7	4	36.36%	0.0327
3	7	5	2	28.57%	0.0153
4	4	7	-3	-75.00%	-0.0413
5	9	11	-2	-22.22%	-0.0312
6	12	8	4	33.33%	0.0553
7	7	4	3	42.86%	0.0284
8	5	8	-3	-60.00%	-0.0197
9	10	8	2	20.00%	0.0223
10	6	10	-4	-66.67%	0.0381

4.4 本章小结

本章详细介绍了模型拆分与模块标注平台的系统实现过程以及相关实验验证。内容包括系统的整体架构设计和本地部署方式，平台采用前后端分离的结构，增强了各功能模块的灵活性。通过对模型管理、参数设置、拆分结果、热力图、语义标注和知识图谱等六个主要界面的展示，较为全面地呈现了平台的操作流程和功能特点。在实验部分，基于工业视觉数据集 MVTEC 进行了模块消融实验，对模块拆分方法的有效性进行了定量分析。实验结果表明，移除特定模块会对检测性能产生一定影响，说明模块化方法在实际检测任务中具有一定的应用价值，也为后续模型的优化和迭代提供了数据支持和技术基础。

第五章 结论

5.1 工作总结

本研究针对工业场景下目标检测模型复用难的问题，提出并实现了一套基于功能解耦与语义解析的模块化方法。主要工作和创新点归纳如下：

- (1) 实现了基于特征相似度的模型模块化拆分方法。通过 K-means 聚类对网络层特征进行分析，实现了模型的自适应分组和功能模块划分，提升了模型的可解释性和局部优化能力。
- (2) 构建了模块功能的自动语义标注流程。结合通道筛选、热力图生成和大模型语义解析，能够为每个模块生成标准化、可理解的功能标签，增强了模型的透明度和可复用性。
- (3) 开发了集成化的软件平台。平台集成了模型管理、参数设置、热力图可视化、语义标注和知识图谱等功能，支持模块的高效存储、检索与操作。消融实验验证了模块拆分和标注方法的有效性。

综上，本文提出的模块化方法为目标检测模型的灵活重组和高效复用提供了技术基础，也为后续模型优化和维护提供了新的思路和工具。

5.2 研究展望

基于本研究已有的研究成果和实验结果，对未来研究进行以下的展望：

- (1) 目前系统适配的模型类型仅限于目标检测模型，尚未覆盖其他主流的大型视觉模型，如 ResNet 等图像识别网络。未来希望能够扩展平台的适用范围，支持更多类型的深度学习模型，实现更广泛的模块化管理和功能标注。
- (2) 实验过程中所用数据集规模较小，验证时每类仅选取了 10 张图片，样本数量有限。后续工作中计划引入更大规模的数据集，提升系统对大批量数据的处理能力和实验结果的代表性。
- (3) 当前采用的语义标注策略较为粗略，依赖大模型生成的标签在准确率上尚未达到理想水平。未来希望结合更精细的算法和优化方法，进一步提升模块语义标注的准确性和可靠性。

参 考 文 献

- [1] HUSSAIN M. Yolov1 to v8: Unveiling each variant—a comprehensive review of yolo[J]. IEEE Access, 2024, 12: 42816-42833.
- [2] HUSSAIN M. Yolo-v5 variant selection algorithm coupled with representative augmentations for modelling production-based variance in automated lightweight pallet racking inspection[J]. Big Data and Cognitive Computing, 2023, 7(2): 120.
- [3] XU R, LIN H, LU K, et al. A forest fire detection system based on ensemble learning[J]. Forests, 2021, 12(2): 217.
- [4] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020: 390-391.
- [5] CHENG Y, WANG D, ZHOU P, et al. A Survey of Model Compression and Acceleration for Deep Neural Networks[J]. IEEE Signal Processing Magazine, 2022, 39(1): 101-116.
- [6] ZHANG X, REN M, URTASUN R, et al. TinyNAS: A Differentiable Neural Architecture Search for Tiny Neural Networks[J]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022: 11416-11425.
- [7] HU E J, SHEN Y, WALLIS P, et al. LoRA: Low-Rank Adaptation of Large Language Models[J]. Advances in Neural Information Processing Systems (NeurIPS), 2022.
- [8] HOULSBY N, GIURGIU A, JASTRZEBSKI S, et al. Parameter-Efficient Transfer Learning for NLP[J]. Proceedings of the 36th International Conference on Machine Learning (ICML), 2019: 2790-2799.
- [9] YE H, LI G Y, JUANG B H. Deep Learning for MIMO Detection[J]. IEEE Transactions on Signal Processing, 2021, 69: 2766-2781.
- [10] WANG T, LIU M, ZHU J, et al. Semantic-Aware Visual Decomposition for Image Coding[J]. International Journal of Computer Vision, 2023.
- [11] RAHMAD C, ASMARA R A, PUTRA D R H, et al. Comparison of Viola-Jones Haar Cascade Classifier and Histogram of Oriented Gradients (HOG) for face detection[C]//IOP Conference Series: Materials Science and Engineering: vol. 732: 1. 2020: 012038.
- [12] DING X, XIE E, WANG W, et al. Dynamic R-CNN: Towards High Quality Object Detection via Dynamic Label Assignment[J]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 11068-11077.
- [13] LI Y, CHEN X, ZHU Y, et al. Small Object Detection: A Survey[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(3): 2547-2555.
- [14] ZHU X, SU W, LU L, et al. Deformable DETR: Deformable Transformers for End-to-End Object Detection[J]. International Conference on Learning Representations (ICLR), 2021.
- [15] CHEFER H, GUR S, WOLF L. Transformer Interpretability Beyond Attention Visualization[J]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 782-791.
- [16] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization[J]. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017: 618-626.
- [17] PANDEY L N, VASHISHT R, RAMASWAMY H G. On the Interpretability of Attention Networks

[J]. Proceedings of The 14th Asian Conference on Machine Learning, 2023, 189: 832-847.

致 谢

衷心感谢我的导师于晗老师在论文选题、研究方法和写作过程中给予的悉心指导和耐心帮助。感谢软件学院的各位老师在学习和科研道路上的教诲与支持。感谢实验室的学长们在实验和技术细节上的无私帮助。感谢我的父母一直以来的理解和鼓励，为我提供了坚实的后盾。感谢学校和国家为我提供良好的学习和研究环境。以上所有支持和帮助都让我顺利完成本论文，在此一并致以诚挚的谢意。

DESIGN AND IMPLEMENTATION OF MODULAR METHODS FOR INDUSTRIAL INTELLIGENCE MODELS

With the rapid advancement of intelligent manufacturing, object detection models have become indispensable in industrial quality inspection, equipment monitoring, and related fields. Despite the impressive performance of deep learning-based models such as the YOLO series on standard benchmarks, their practical deployment in industrial scenarios is often limited by poor interpretability, low reusability, and high maintenance costs. The tightly coupled, end-to-end architecture of mainstream object detection models leads to unclear functional boundaries, making it difficult to adapt, optimize, or reuse specific components for new tasks or environments. This "black box" nature not only increases the workload for engineers but also restricts the flexibility and efficiency of model deployment and iteration.

To address these challenges, this thesis proposes and implements a comprehensive platform for modular decomposition and function annotation of object detection models. The platform is designed to clarify the internal structure of deep learning models, support fine-grained module management, and facilitate the reuse and optimization of model components. The system integrates several key technologies, including automated model splitting based on feature similarity, semantic annotation of functional modules, heatmap visualization for interpretability, and knowledge graph-based storage and management. By combining these techniques, the platform aims to provide a systematic solution for the modularization and explainability of object detection models in industrial applications.

The research begins with the design of a K-means-based modular decomposition method, which analyzes the feature similarity between network layers to adaptively group them into functional modules. This approach enables the clear separation of different functional units within the model, laying the foundation for targeted optimization and flexible reuse. To further enhance interpretability, the thesis develops an automated function annotation process that leverages channel selection strategies, heatmap generation, and large model-based semantic analysis. By integrating channel response values and SHAP values, the system generates intuitive heatmaps that highlight the focus areas of each module. These heatmaps are then interpreted using large language models to produce standardized, human-understandable

function labels for each module.

For efficient management and retrieval of module information, a knowledge graph-based system is constructed using Neo4j, which organizes the relationships between modules, their functions, and application scenarios. This knowledge graph enables engineers to quickly locate and recombine modules for new tasks, supporting the flexible deployment of object detection models in diverse industrial settings. The entire workflow is encapsulated in an integrated software platform, featuring a user-friendly interface built with React and a robust backend powered by Flask. The platform supports model management, parameter configuration, module splitting, heatmap visualization, semantic annotation, and knowledge graph operations, streamlining the process of model modularization and function annotation.

The thesis also details the technical challenges encountered during the development of the platform, such as the selection of appropriate clustering algorithms for model decomposition, the integration of interpretability techniques for deep neural networks, and the design of efficient data structures for knowledge graph storage. Solutions to these challenges are discussed, including the use of feature normalization, centroid initialization strategies, and the adoption of dual-weight fusion for channel selection. The semantic annotation process is further refined by defining standard function categories and prompt templates, which help large language models generate more accurate and human-readable module descriptions.

Extensive experimental validation is conducted on representative datasets, including MVTEC for defect detection and COCO for general object detection. The experiments demonstrate that the proposed modular decomposition and annotation methods can accurately identify and label the functions of model modules. Ablation studies further show that removing key modules, as identified by the annotation process, leads to significant drops in detection performance, confirming the effectiveness and reliability of the approach. The platform's scalability and adaptability are also discussed, highlighting its potential to support a wider range of deep learning models and larger datasets in the future. Limitations of the current work, such as the relatively small scale of validation data and the need for more precise semantic annotation algorithms, are acknowledged, providing directions for future research and development.

In addition, the thesis explores the broader implications of modularization and explainability for the development and maintenance of industrial AI systems. By enabling engineers

to better understand and manage the internal structure of object detection models, the platform not only improves the efficiency of model adaptation and reuse but also supports more transparent and trustworthy AI solutions. The knowledge graph-based management system further facilitates collaboration and knowledge sharing among engineers, promoting the accumulation and transfer of expertise within organizations.

Furthermore, the thesis discusses the potential for extending the platform to support a wider variety of deep learning models beyond object detection, such as image classification and segmentation networks. The modular approach outlined in this work lays the groundwork for more flexible and scalable AI systems, where components can be easily updated, replaced, or repurposed as technology evolves or as new industrial requirements emerge. The integration of knowledge graphs not only aids in technical management but also supports organizational memory, allowing best practices and module performance data to be accumulated and referenced over time.

The research also highlights the importance of user experience in the adoption of AI tools in industry. The platform's interface and workflow are designed to lower the barrier for engineers and practitioners, making advanced model management and interpretability techniques accessible even to those without deep expertise in machine learning. This democratization of AI technology is crucial for accelerating the digital transformation of traditional industries.

In summary, this thesis presents a novel and practical solution to the challenges of interpretability and reusability in industrial object detection models. By enabling modular decomposition, automated function annotation, and knowledge graph-based management, the proposed platform enhances the transparency, flexibility, and efficiency of model development and deployment. The research outcomes contribute valuable insights and tools for the advancement of modular, explainable, and reusable deep learning models in industrial applications, and lay a solid foundation for future work in this important area.