

分类号_____

密级_____

U D C _____

编号_____



博士后研究报告

深度学习在目标检测领域的应用研究与前沿探索

——从数据增强、网络改进到知识要素融合

张正德

工作完成日期 2020 年 7 月— 2022 年 6 月

报告提交日期 2022 年 6 月

上海交通大学（上海）

2022 年 6 月

深度学习在目标检测领域的应用研究与前沿探索

——从数据增强、网络改进到知识要素融合

THE APPLICATION AND FRONTIER OF DEEP LEARNING
ALGORITHM IN THE FIELD OF TARGET DETECTION
—FROM DATA AUGMENTATION, NETWORK IMPROVEMENT TO
KNOWLEDGE ELEMENT FUSION

博 士 后 姓 名 张正德

流动站（一级学科）名称 信息与通信工程

专 业（二级学科）名称 深度学习算法

研究工作起始时间 2020 年 7 月 1 日

研究工作期满时间 2022 年 6 月 30 日

上海交通大学（上海）

2022 年 6 月

摘要

随着交叉熵损失、卷积神经网络、残差连接等技术的突破，深度学习算法已经成为最流行的目标检测算法。在实际应用过程中，复杂的环境条件、有限的计算能力、小样本数据、极低的空间分辨率等问题对实现鲁棒、实时的检测提出了挑战。此外，当代基于连接主义范式设计的神经网络，存在可解释性不足、不完全信息情况下性能大幅度下降等问题，迈向通用智能还需不断地探索和突破。

本文进行了深度学习算法在目标检测领域的应用研究和前沿探索。研究了通用的数据集自适应增强和增广方法；研究了深度学习目标检测、跟踪和分类算法的原理和改进方法，并初步探索了量子力学启发的第三代人工智能算法；开发了标注软件 mmlabelme、AI 算法平台 XAI，实现了深度学习算法从数据标注、算法设计与改进，到优化部署的一站式应用。所提出的数据增强和网络改进方法能在特定任务中有效地提升算法的性能，并广泛应用到该领域的其他任务中。所提出的知识-数据协同驱动的神经网络有望发展为下一代人工智能范式。

具体的工作内容有：

(1) 针对真实复杂场景和有限边缘计算算力场景下，实现鲁棒、实时的斑马线检测的问题，本文提出基于 YOLOv5 的斑马线检测网络(Crosswalk Detection Network, CDNet)，实现车载摄像头视角下快速准确的斑马线检测，并在 Jetson nano 设备上实现实时。强大的卷积神经网络特征提取器用于处理复杂环境，网络中嵌入了 squeeze-and-excitation (SE) 注意力机制模块，使用负样本训练 (NST) 方法提高准确率，区域利用兴趣 (ROI) 算法进一步提高检测速度，提出了一种新的滑动感受野短时向量记忆 (SSVM) 算法来提高车辆过线行为检测精度，使用合成雾增强算法允许模型适应雾天。最后，在 Jetson nano 上以 33.1 FPS 的检测速度，我们在上述复杂场景中获得了 94.83% 的平均 F1 分数。对于晴天和阴天等更好的天气条件，F1 分数超过 98%。该工作为人工神经网络算法在边缘计算设备上具体优化应用提供了参考。源代码、数据集和教程以已开源。

(2) 针对数据量不足和雾天的鲁棒检测问题，我们发布了一个用于绝缘子及其自爆缺陷检测的数据集 (Synthetic Foggy Insulator Detection Dataset, SFID)，并提供了一个检测基准网络 (Foggy Insulator Network, FINet)。在这项工作中，实现并优化了合成雾算法，构建并发布了包含 13000 张图像的 SFID 数据集。通

过引入通道注意力机制将 YOLOv5 网络改进为 SE-YOLOv5，并从零开始训练一个具有 96.2% F1 分数的绝缘子及其缺陷的鲁棒检测模型，并作为基准。本文提出的合成雾算法可广泛用于各种数据集的数据增强。训练好的模型可应用于输电线路检测领域。源代码、数据集和教程已开源。

(3) 针对从具有极低空间分辨率和小样本下鲁棒检测目标的问题，我们探索了从卫星遥感图像中检测绝缘子的可能性。绝缘子直径约为 200 毫米，在最佳卫星遥感图像 (Satellite Remote Sensing Image, SRSI) 中仅占 0.4 个像素。由于极低的空间分辨率和混合像元问题，从 SRSI 中检测绝缘子和自爆缺陷是一项具有挑战性的任务。本文提出了自适应网络和检测网络相结合的两阶段检测方法，采用多重滤波和自适应增强方法实现端到端检测，可以有效提高检测 F1 分数达到 80% 所需的空间分辨率 (Required Spatial Resolution, rSR)。本文提出的想法可以推动从 SRSI 中检测绝缘子和缺陷的任务实现，并且该算法可以广泛应用于低空间分辨率图像的目标检测任务中。此外，针对小样本问题，我们也基于来自云南电网呼马山试验基地及前人的少量、低分辨率输电导线样本，设计使用超分辨率算法、提出基于动态背景和随机变换的离线数据增广算法，生成了清晰的 7000 张输电导线数据集，解决了随机位置、随机角度、随机尺寸和随机背景等数据多样性问题。并且实现了输电导线、散股和断股的实时、鲁棒的检测，平均 F1 分数达 94.7%。该工作中实现的输电导线及其缺陷检测算法能应用于输电线路巡检领域，使巡检更加高效和智能，所提出的超分辨率和数据集增广方法可以广泛应用于各领域的低分辨率图像、小样本数据集中。

(4) 针对不完全信息场景下性能大幅度下降等的难点问题。本文在数据、算法和算力三个要素的基础上，引入知识要素，通过综合利用四大要素，构造更强大的人工智能，以期望解决这一问题。该工作从量子理论出发，提出以量子力学波函数作为知识要素，研究不完全信息场景下的新型检测算法。凝练关键科学问题，针对如何利用机器学习求解基态波函数的问题，研究基于交换不变函数和神经网络的新泛函；针对如何快速求解薛定谔方程的问题，研究基于快速积分、分布式并行和即时编译技术的高性能解算器；针对如何有效利用知识要素的问题，研究基于量子纠缠和量子概率论的知识信息流与检测结果流的双模态融合算法。研究层层递进，最终实现以波函数作为知识要素的知识-数据协同驱动融合检测

算法，解决不完全信息问题。所提出的以量子态、量子测量和量子演化为内涵的知识要素和基于量子纠缠和量子概率论的四大要素融合方法，有望进一步发展为新的人工智能范式。

关键词：人工智能、深度学习、目标检测、量子理论、不完全信息、斑马线检测、边缘计算、绝缘子缺陷检测、卫星遥感图像、动态背景、数据增强和增广

Abstract

With breakthroughs in technologies such as cross-entropy loss, convolutional neural networks, and residual connections, deep learning algorithms have become the most popular object detection algorithms. In the process of practical application, complex environmental conditions, limited computing power, small samples, extremely low spatial resolution and other issues pose challenges to the realization of robust and real-time detection. In addition, contemporary neural networks designed based on the connectionist paradigm have problems such as insufficient interpretability and significant performance degradation in the case of incomplete information. Towards general intelligence, continuous exploration and breakthroughs are required.

In this paper, the application research and frontier exploration of deep learning algorithms in the field of target detection are carried out. We researched the adaptive enhancement and augmentation methods of general data sets; studied the principles and improvement methods of deep learning target detection, tracking and classification algorithms, and initially explored the third-generation artificial intelligence algorithms inspired by quantum mechanics; and developed *mmlabelme* annotation software and the AI algorithm management and deployment platform *XAI*; and realize the one-stop application of deep learning algorithms from data labeling, algorithm design and improvement, to optimized deployment. The proposed method can effectively improve the performance of the algorithm in specific tasks, and it can be widely applied to other tasks in this field. The proposed knowledge-data synergy-driven neural network is expected to develop into the next-generation artificial intelligence paradigm.

The works are:

(1) Aiming at the problem of robust and real-time zebra crossing detection in real complex scenes and limited edge computing power, this paper proposes a crosswalk detection network (CDNet) based on YOLOv5 is proposed to achieve fast and accurate crosswalk detection under the vision of the vehicle-mounted camera, and real-time detection is implemented on Jetson nano device. The powerful convolution neural network feature extractor is used to handle complex environments, the squeeze-and-

excitation (SE) attention mechanism module is embedded into the network, the negative samples training (NST) method is used to improve the accuracy, the region of interest (ROI) algorithm is utilized to further improve the detection speed, and a novel slide receptive field short-term vector memory (SSVM) algorithm is proposed to improve vehicle crossing behavior detection accuracy, the synthetic fog augmentation algorithm is used to allow the model adaptable to foggy scenario. Finally, with a detection speed of 33.1 FPS on Jetson nano, we obtained an average F1 score of 94.83% in the above complex scenarios. For better weather condition such as sunny and cloudy days, the F1 score exceeds 98%. This work provides a reference for the specific application of artificial neural network algorithm optimization methods on edge computing devices. The datasets, tutorials and source codes are available on GitHub.

(2) For the robust detection problem of insufficient data and foggy weather, we released a dataset (Synthetic Foggy Insulator Detection Dataset, SFID) for the detection of insulators and their self-explosion defects, and provided a detection benchmark network (Foggy Insulator Network, FINet). In this work, a synthetic fog algorithm is implemented and optimized. The SFID dataset containing 13000 images was constructed and published. The YOLOv5 network is improved to SE-YOLOv5 by introducing a channel attention mechanism, and a robust detection model for insulators and their defects with 96.2% F1 score is trained from scratch and used as a benchmark. The synthetic fog algorithm proposed in this paper can be widely used for data augmentation of various datasets. The trained model can be applied to the field of transmission line detection. Source code, datasets and tutorials are open sourced.

(3) For the problem of robust detection from extremely low spatial resolution and small samples. We explored the possibility of detecting insulators from satellite remote sensing images, which are about 200 mm in diameter and occupy only 0.4 pixels in the best Satellite Remote Sensing Image (SRSI). Detecting insulator and self-detonation defects from SRSI is a challenging task due to extremely low spatial resolution and mixed pixel issues. In this paper, a two-stage detection method combining adaptive network and detection network is proposed. Multiple filtering and adaptive enhancement method are used to achieve end-to-end detection, which can effectively

improve the required spatial resolution (rSR) to achieve 80% of the detection F1 score). The ideas presented in this paper can advance the goal of detecting insulators and defects from SRSI, and the algorithm can be widely applied to object detection tasks with low spatial resolution images. In addition, based on a small number of low-resolution transmission wire samples from the Yunnan Power Grid Huma Mountain Test Base and predecessors, we used a super-resolution algorithm, and proposed an offline data augmentation algorithm based on dynamic background and random transformation. A dataset of 7000 transmission wires, which solves data diversity problems such as random positions, random angles, random sizes, and random backgrounds. And it realizes real-time and robust detection of transmission lines, loose strands and broken strands, with an average F1 score of 94.7%. The transmission line and its defect detection algorithm implemented in this work can be applied to the field of transmission line inspection, making the inspection more efficient and intelligent. The proposed super-resolution and dataset augmentation methods can be widely used in low-resolution images and small-sample datasets in various fields.

(4) In view of the difficult problems such as the substantial performance drop in incomplete information scenarios. Based on the three elements of data, algorithm and computing power, this paper introduces the element of knowledge, and constructs a more powerful artificial intelligence by comprehensively utilizing the four elements, in order to solve this problem. Starting from quantum theory, this work proposes a new detection algorithm for incomplete information scenarios using wave function as a knowledge element. Condensing key scientific issues, for the problem of how to use machine learning to solve the ground state wave function, we researched new functionals based on exchange invariant functions and neural networks; for the problem of how to quickly solve the Schrodinger equation, we researched the high-performance solver based on fast integration, distributed parallelism for compilation technology; for the problem of how to effectively utilize knowledge elements, the bimodal fusion algorithm of knowledge information flow and detection result flow based on quantum entanglement and quantum probability theory is studied. The research progresses layer by layer, and finally realizes the knowledge-data collaborative driving fusion detection

algorithm with wave function as the knowledge element, and solves the problem of incomplete information. The proposed knowledge elements with quantum state, quantum measurement and quantum evolution as the connotation and the four elements fusion method based on quantum entanglement and quantum probability theory are expected to be further developed into a new artificial intelligence paradigm.

Keywords: artificial intelligence, deep learning, object detection, quantum theory, incomplete information, zebra crossing detection, edge computing, insulator defect detection, satellite remote sensing imagery, dynamic background, data augmentation and enhancement

目 录

目录

摘 要	I
Abstract	IV
目 录	VIII
图 目 录	XII
表 目 录	XV
第 1 章 绪论	1
1.1 人工智能概述	1
1.1.1 人工智能发展简史	1
1.1.2 深度学习工具简介	3
1.1.3 神经网络基本原理	3
1.2 深度学习算法在计算机视觉的应用	4
1.3 研究现状及挑战	5
1.4 迈向第三代人工智能	6
1.4.1 知识-数据协同驱动算法现状	7
1.4.2 量子理论发展现状	8
1.5 研究内容和意义	10
1.6 本报告的结构	11
第 2 章 方法论	13
2.1 第二代人工智能方法	13
2.1.1 基于 YOLOv5 的目标检测算法及改进	13
2.1.1.1 YOLOv5 目标检测算法	13
2.1.1.2 SE-YOLOv5 目标检测算法	14
2.1.2 合成雾/去雾算法	15
2.1.2.1 合成雾原理	15
2.1.2.2 雾化算法的优化	17
2.1.3 离线数据集增强算法	18
2.1.4 在线数据增强算法	18
2.1.5 图像质量提升算法	18
2.1.5.1 超分辨率算法	18
2.1.5.2 高反差保留算法	20
2.1.6 动态背景数据增广算法	20
2.1.6.1 数据标注和定义	20
2.1.6.2 背景图像生成	21
2.1.6.3 随机位置、角度和尺寸方法	24
2.2 第三代人工智能方法	26
2.2.1 量子理论基础知识	26

2.2.2 神经网络泛函方法	27
2.2.2.1 网络结构设计	28
2.2.2.2 KFAC 优化	29
2.2.2.3 形式理论和重整化群	29
2.2.3 高性能计算器实现原理	30
2.2.3.1 离散化	31
2.2.3.2 快速积分	31
2.2.3.3 分布式并行	32
2.2.3.4 即时编译	32
2.2.4 协同驱动神经网络	32
2.2.4.1 神经网络目标检测算法	33
2.2.4.2 构建知识要素	33
2.2.4.3 描述知识要素	34
2.2.4.4 利用知识要素	34
2.3 评价指标	36
2.3.1 混淆矩阵	36
2.3.2 ROC 空间	37
2.3.3 交并比 IoU	39
2.3.4 mAP	39
2.3.5 所需的空间分辨率	41
2.3.6 合成图准确性	41
第 3 章 CDNet：基于 YOLOv5 在 Jetson Nano 上实时和鲁棒的斑马线检测网络 ..	42
3.1 引言	42
3.2 斑马线检测相关研究	44
3.3 方法	45
3.3.1 网络架构	45
3.3.2 NST 负样本训练方法	46
3.3.3 ROI 算法	47
3.3.4 SSVM 算法	48
3.4 消融实验	49
3.4.1 CDNet 数据集	49
3.4.2 检测尺寸选择	50
3.4.3 NST 技术的消融实验	50
3.4.4 ROI 算法的消融实验	51
3.4.5 SSVM 算法的消融实验	51
3.4.6 合成雾增强消融实验	52
3.4.7 与 SOTA 模型的性能对比	52
3.5 结果和分析	53
3.6 总结	55
第 4 章 FINet：基于合成雾增强的绝缘子检测数据集和基准模型 ..	56
4.1 引言	56
4.2 方法	57
4.3 消融实验	57
4.3.1 合成雾算法加速的消融实验	57

4.3.2 神经网络改进的消融实验	58
4.3.3 合成雾数据增强的消融实验	59
4.3.4 合成雾模型在晴天下的实验	60
4.3.5 与 Classic 和 SOTA 模型的比较	61
4.4 结果和分析	62
4.5 总结	63
第 5 章 AIDNet: 基于自适应算法的卫星遥感图像检测绝缘子的探索研究	64
5.1 引言	64
5.2 方法	65
5.2.1 卫星遥感图像模拟算法	65
5.2.2 自适应和目标检测网络	65
5.2.2.1 自适应网络	66
5.2.2.2 检测网络	67
5.2.3 参数设置	68
5.2.4 评价指标	68
5.3 消融实验	68
5.3.1 空间分辨率估计	68
5.3.2 模拟方法和效果	68
5.3.2.1 模拟方法	69
5.3.2.2 不同下采样率下的仿真效果	69
5.3.3 检测性能的消融研究	70
5.3.4 所需的空间分辨率	71
5.4 结果和分析	71
5.5 总结与展望	72
第 6 章 TLDNet: 基于数据增强增广和神经网络的输电导线及其缺陷检测方法	74
6.1 引言	74
6.2 方法	76
6.2.1 图像质量提升算法	76
6.2.2 动态背景数据增广算法	77
6.2.3 输电导线及其缺陷检测网络	77
6.3 消融实验	78
6.3.1 数据集	78
6.3.2 图像质量提升算法消融实验	79
6.3.2.1 定性分析	79
6.3.2.2 定量分析	79
6.3.2 噪声背景消融实验	80
6.3.2.1 定性分析	80
6.3.2.2 定量分析	81
6.3.3 SE 模块消融实验	82
6.4 结果和分析	83
6.4.1 最终性能测试	83
6.4.2 模型的迁移学习应用	84
6.5 总结	87

第 7 章 波函数作为知识要素的第三代人工智能.....	88
7.1 引言	88
7.2 内容	89
7.2.1 神经网络泛函研究	89
7.2.2 高性能解算器研究	90
7.2.3 协同驱动融合目标检测算法研究.....	90
7.3 方法	92
7.4 验证实验.....	95
7.4.1 无限深势阱	95
7.4.2 H 原子体系	97
7.4.2.1 求解径向方程	98
7.4.2.2 求解角向方程	100
7.4.2.3 H 原子体系的定态解.....	102
7.4.2.4 Wavelib 求解 H 原子体系波函数	103
7.4.2.4 精确求解和 Wavelib 库求解结果对比.....	108
7.4.3 谐振子体系	111
7.4.3.1 经典力学描述下的谐振子体系	111
7.4.3.2 量子力学描述下的谐振子体系	112
7.4.3.3 任意项混合高斯函数拟合算法	114
7.4.4 人体关键点组成的多粒子系统	119
7.4.4.1 任务描述	119
7.4.4.2 实现思路	119
7.4.4.3 波函数与关键点检测算法融合	121
7.4.4.4 结果	122
7.4.5 密集遮挡下的跟踪算法	123
7.5 总结	125
第 8 章 软件、平台和库	126
8.1 多模态融合标注软件	126
8.2 人体行为分析算法软件	127
8.3 XAI 算法平台	129
8.4 Python 库：Damei	131
第 9 章 总结与展望	132
9.1 工作总结	132
9.2 工作展望	133
参考文献	134
致 谢	143
个人简历	145
博士后期间的研究成果	146
博士生期间的研究成果	148
永久通信地址	149

图 目 录

图 1.1 人工智能领域关系图	2
图 1.2 神经网络的全连接结构	3
图 2.1 改进的 SE-YOLOv5 目标检测网络	13
图 2.2 从 CSP 到 SE-CSP 的改进。 (a)本地 CSP 网络, (b)嵌入 CSP 网络的 SE 网络	14
图 2.3 不同亮度 L 下的雾化效果。D 为雾厚度。	16
图 2.4 不同雾厚 D 下的雾化效果。L 为亮度。	17
图 2.5 雾化算法的矩阵计算优化	17
图 2.6 超分辨率网络 : (a) 整体网络架构 ; (b) RDB 模块内部	19
图 2.7 输电导线、断股和散股三类目标, 分别用红色、绿色和黄色多边形表示	21
图 2.8 噪声背景合成图样例	22
图 2.9 基于滑动窗口生成的多尺度动态背景 (a) 原始图像 (b-f) 尺度从小到大生成的子背景图	23
图 2.10 神经网络泛函的技术路线	27
图 2.11 Ansatz 泛函神经网络结构	28
图 2.12 高性能解算器技术路线	30
图 2.13 双模态融合目标检测算法的技术路线	33
图 2.14 ROC 空间	38
图 2.15 三种 ROC 曲线和 AUC	38
图 2.16 交并比	39
图 2.17 目标检测的 ROC 曲线 (PR 曲线)	40
图 3.1 真实复杂场景中的斑马线检测示意图。 (a) 清晰可见, (b) 被雨水反射, (c) 被雨刷遮挡, (d) 扭曲, (e) 转向截断, (f) 在夜间, (g) 损坏, (h) 视野被车辆挡住, (i) 眩光, 以及 (j) 部分丢失	43
图 3.2 整体架构示意图。 (a) 脊椎, (b) 颈部, (c) 头部, 和 (d) 斑马线检测和车辆横穿斑马线行为结果	45
图 3.3 典型检测错误 (a) NST 之前的典型错误检测, (b) NST 之后的正确分类	46
图 3.4 ROI 算法的图示。 (a) 原始图像, (b) 通过保持纵横比调整大小, (c) ROI, (d) 填充, (e) 检测到的斑马线, (f) 缩放以匹配原始图像。	48
图 3.5 SSVM 算法说明 : (a) 斑马线检测帧, (b) 滑动感受野, (c) 短期向量记忆, (d) 车辆穿越行为结果	49
图 3.6 F1 分数和 FPS 与检测大小的趋势图	50
图 3.7 使用不同模型的最终检测结果说明。 方框表示检测大小为 640, 而三角形表示检测大小为 288。	55
图 4.1 复杂环境下绝缘子及其缺陷检测结果 (a) 简单背景, (b) 带轻雾的简单背景, (c) 浓雾中的缺陷, (d) 天空背景, (e) 不同尺度, (f) 垂直绝缘子和中雾, (g) 截断绝缘子, (h) 图像模糊, (i) 浓雾中的绝缘子	62

图 5.1 网络架构 (a) 技术路线 ; (b) 自适应网络 ; (c) 检测网络.....	66
图 5.2 不同方法的模拟效果比较。 (a) 原始无人机图像 ; (b) 用插值双三次方法模拟图像 ; (c) 混合像元法模拟图像 ; (d) 真实卫星遥感图像 (SRSI)	69
图 5.3 不同下采样率下的模拟图像, 降采样倍数从 1 到 88x.....	70
图 5.4 模型所需空间分辨率评估结果 : (a) Native YOLOv5 ; (b) AIDNet.....	71
图 6.1 原始输电导线及其缺陷图像	75
图 6.2 数据集生成流程图.....	78
图 6.3 提升图像质量前后对比	79
图 6.4 噪声背景模型 NBM 和动态背景模型 DBM 对比。左侧是真值标注, 右侧是检测结果。(a) NBM 在 NBSet 上的结果 ; (b) NBM 在 DBSet 上的结果 ; (c) DBM 在 DBSet 上的结果.....	81
图 6.5 模型最终检测效果.....	83
图 6.6 Hybrid-M 在额外图像上的检测效果	86
图 7.1 研究内容.....	89
图 7.2 研究目标.....	91
图 7.3 研究方案	92
图 7.4 无限深势阱体系	95
图 7.5 无限深势阱体系结果 (a) 一维无限深势阱波函数 (b) 一维无限深势阱概率密度 (c) 二维无限深势阱波函数.....	96
图 7.6 球坐标系	97
图 7.7 各连带拉盖尔多项式的函数形式	99
图 7.8 H 原子体系的能级	99
图 7.9 各项连带勒让德多项式的函数形式.....	101
图 7.10 前几个连带勒让德多项式图示	101
图 7.11 典型的角量子数和磁量子数下球谐函数的具体表达式.....	102
图 7.12 电子占据轨道的顺序.....	103
图 7.13 不同主量子数、角量子数和磁量子数的电子层一览	103
图 7.14 6-311G 基组及其图示	104
图 7.15 H 原子体系有效势图示	107
图 7.16 精确解法和 Wavelib 库求解的 H 原子体系的前几个 spdf 轨道波函数 (a) 精确解法结果 (b) Wavelib 库求解结果	109
图 7.17 H 原子体系轨道波函数 (a) 直接求解绘制的结果 (b) 本文开发的 Wavelib 库求解的结果	110
图 7.18 精确求解的谐振子体系波函数	113
图 7.19 谐振子体系的波函数 (a-b) 谐振子系统 (c-f) 定态波函数 (G-H) 含时波函数	114
图 7.20 高斯函数图像	114
图 7.21 单中心多轨道高斯函数拟合波函数	116
图 7.22 多中心多轨道高斯函数拟合波函数	117
图 7.23 人体关键点检测和骨架抽象模型.....	119

图 7.24 双空间模型.....	121
图 7.25 MWNet 模型架构.....	123
图 7.26 MWNet 在 MOT20 数据集上检测结果可视化.....	124
图 8.1 mmlableme 多模态融合标注软件.....	126
图 8.2 智慧医疗人体行为分析算法.....	127
图 8.3 智能装备人体行为分析算法.....	128
图 8.4 光流特征.....	128
图 8.5 XAI 算法平台设计.....	129
图 8.6 Damei Python Library 接口文档.....	131

表 目 录

表 2.1 三分类问题的一个混淆矩阵	36
表 3.1 NST 算法的消融结果	50
表 3.2 使用和不使用 ROI 算法的速度和得分比较。	51
表 3.3 SSVM 算法的消融结果	52
表 3.4 合成雾增强的消融结果	52
表 3.5 与以前的目标检测模型的比较。	52
表 3.6 使用差异模型的斑马线检测和车辆横穿行为检测结果。	53
表 4.1 雾化算法的消融实验结果	58
表 4.2 网络改进前后性能对比	59
表 4.3 合成雾数据增强的消融实验结果	59
表 4.4 M2 在 T1 和 T2 测试集上的性 能	60
表 4.5 检测性能比较	61
表 5.1 检测性能对比	70
表 5.2 所需的空间分辨率评估结果	72
表 6.1 低质量数据集模型 LQM 和高质量数据集模型 HQM 性能对比	79
表 6.2 噪声背景模型 NBM 和动态背景模型 DBM 性能对比	81
表 6.3 原生 YOLOv5 和改进 SE-YOLOv5 性能对比	82
表 6.4 最终模型检测结果	84
表 6.5 噪声模型 NBM 和混合模型 Hybrid-M 性能对比	84
表 7.1 MWNet 性能与其他模型对比	124

第1章 绪论

1.1 人工智能概述

人工智能（Artificial Intelligence, AI）是使用机器代替人类实现感知、识别、分析、决策等功能，其本质是对人的意识与思维的模拟。

1.1.1 人工智能发展简史

人工智能自上世纪 50 年代诞生以来，经历了三次“春天”，有两种相互竞争的“范式”。

1955 年，麦卡锡（J. McCarthy）和明斯基（M. L. Minsky）等学者[1]在达特茅斯人工智能夏季研究项目（The Dartmouth Summer Research Project on Artificial Intelligence）中提出：人类思维的很大一部分是按照推理和猜想规则对“词（word）”进行操作所组成的。根据这一思路，人们提出知识与经验的推理模型，即第一代以知识驱动的“符号”主义范式。1956 年，罗森布罗特（Rosenbaltt）等学者[2]认为：感官的刺激不存储在记忆中，而是在神经网络中建立起“刺激-响应”的连接（通道），通过这个“连接”保证智能行为的产生，即第二代数据驱动的“连接”主义范式。当时人们觉得人工智能会在 20 年内改变世界，所有的工作将会被人工智能颠覆。然而，由于当时计算机性能的不足、真实问题的复杂性等原因，科研人员低估了人工智能项目的难度，不仅导致与美国国防高级研究计划署的合作计划失败，而且使得人们对人工智能的前景蒙上了一层阴影。到 1973 年的《莱特希尔报告》明确指出当时的人工智能的任何部分没有达到人们想象的水平，第一个“春天”随之结束，人工智能遭遇长达 6 年的科研深渊。

1980 年，卡内基梅隆大学为数字设备公司设计了一套名为 XCON 的“专家系统”[3]，取得了巨大的成功，人工智能迎来第二次“春天”。“专家系统”是一种采用“知识库+推理机”组合的人工智能程序系统，能为公司每年节省大量的经费，这段时期仅专家系统产业的价值就高达 5 亿美元。然而，在 7 年之后，命运的车轮再次碾过人工智能，苹果和 IBM 生产的台式机性能超过了计算机智能系统，“专家系统”风光不再，人工智能再次陷入低谷。

在困难时期，依旧有科学家坚持做研究。上个世纪 90 年代以来，随着 AI 技

术尤其是神经网络技术的逐步发展，人们开始逐步对 AI 抱有客观理性的认知，人工智能技术进入平稳发展时期。

1997 年，IBM 公司的国际象棋电脑“深蓝”战胜了世界冠军卡斯帕罗夫，人工智能在特定任务中超越人类。

2006 年，Hinton 在深度学习方面取得突破[4]，让人们又一次看到机器超越人类的希望。

2011 年，IBM 公司的人工智能程序“沃森”参加智能问答节目并战胜了 2 位世界冠军。2015-2016 年，Google DeepMind 的人工智能围棋程序 AlphaGo 战胜世界冠军，2017 年，自我博弈的 AlphaZero 又战胜了 AlphaGo。2011 年以来的，人工智能迎来第三个“春天”，人工智能步入快速发展阶段，各个高校、科研院所和公司都开始涌入这个赛道。目前商业上又鼓吹人工智能无所不能，但应看到人工智能依旧有其局限性，我们离“通用人工智能”还有很长的路要走。

在 20 世纪 80 年代之前，符号主义范式一直主导着 AI 的发展，90 年代以来，研究重心逐步转移到连接主义范式上，到 21 世纪进入高潮，大有替代符号主义之势。因此，当代的人工智能技术主要是在连接主义范式下以数据驱动的第二代人工智能。

当代人工智能领域研究的内容包括：计算机视觉、语音技术、自然语言处理、专家系统和机器人等。

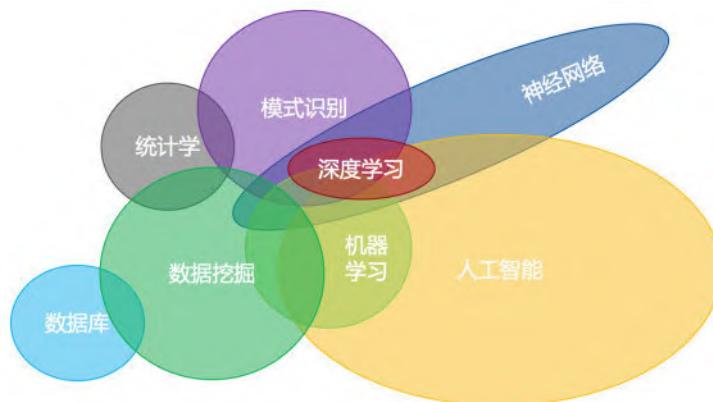


图 1.1 人工智能领域关系图

领域关系图如图 1.1 所示，人工智能包括机器学习、深度学习和模式识别等，而神经网络是机器学习中的一种方法，深度学习是神经网络方法中的一个子集[5]。

1.1.2 深度学习工具简介

深度学习的基础框架的发展，极大地繁荣了该领域。目前主要的深度学习框架有 MXNet、Keras、Tensorflow[6]、JAX[7]，Pytorch[8]，Paddlepaddle[9]等。目前，学术界使用较多是 Facebook 的 Pytorch 框架，工业界比较流行的是 Google 的 Tensorflow，最近，Google 开始逐渐转向 Tensorflow 的简化库 JAX。此外，百度开发的国产化 Paddlepaddle 近年来也发展迅速。

深度学习编程语言方面，目前主流的是 Python 和 C++。Python 易于上手，优秀的社区支持是加速其传播的重要因素，C++在速度方面更有优势，Python 编写的人工智能程序在最终部署前一般会转换为 C++引擎以提升速度。除此之外，Go 语言、R 语言和 Matlab 等也已支持人工智能算法开发。

在开发工具方面，Vscode 和 Pycharm 是主要的 IDE 工具，Jupyter notebook 在代码教学方面具有独特的优势。

1.1.3 神经网络基本原理

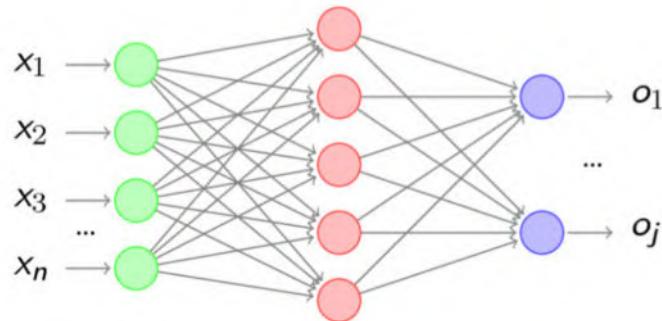


图 1.2 神经网络的全连接结构

一个简单的神经网络的如图 1.2 所示，输入为 x ，输出为 o ，网络从左到右依次是输入层（绿色表示）、隐藏层（红色表示）和输出层（蓝色表示）。一个神经网络的隐藏层可能会有很多。每一层由多个神经元组成（圆圈表示），每个神经元有 1 个或者多个输入，有 1 个输出，输出作为下一层神经元的输入。神经元的输入和输出之间的关系可以表示为：

$$y = \text{Activation} \left(\sum_i (w_i x_i + b_i) \right) \quad (1.1)$$

其中, y 是输出, x_i 是输入, w_i 和 b_i 分别是权重和偏置。一个简化但不确切的学习过程如下: 权重和偏置是需要优化的参数, 网络初始化时候随机赋值。对所有输入与权重相乘并加上偏置后, 采用合适的激活函数进行激活, 得到输出。输入从左向右经过所有神经元, 该过程称为**前向传播**, 最终输出结果是预测值。设计损失函数, 对比预测值和真值的“距离”, 并从右向左计算每一个神经元的权重和偏置的梯度, 该过程称为**后馈传播** (即误差反向传播)。根据设定的学习率和梯度, 更新每一个权重和偏置, 使得预测值和真值不断接近, 该过程称为**梯度下降**。如此往复训练多代后, 模型通过输入的数据和真值“学习”到了从输入到预测值的映射关系。由于激活函数是非线性的, 因此神经网络整体是一个非线性的变换。当然, 这是一个最简化版本的“连接”主义的实现, 实际的神经网络要解决过拟合、凸优化、局部最优和全局最优、梯度消失和梯度爆炸、超参数选取、优化器等诸多问题。

例如, 经典的 ResNet[10]采用残差结构能解决随着网络加深梯度反向传播时消失的问题; 采用卷积操作而不是公式 1.1 所述的全连接操作, 能一定程度防止过拟合, 与池化层结合能不断地增大感受野; 采用规范化层能一定程度防止过拟合的同时防止梯度爆炸; 采用遗传算法或 K 折训练能进行超参数的选择, 等等。

合适的**激活函数**是神经网络获得非线性能力的关键, 经典的激活函数有 Softmax、Sigmoid、ReLU 等。**损失函数**是评估预测值和真值间“距离”的度量, 合适的损失函数对网络训练至关重要, 经典的损失函数有均方误差、交叉熵损失等。针对具体的问题提出新的激活函数或损失函数是有效的提升性能的方案, 例如 LeakyReLU, Generalized Focal Loss[11], Seesaw Loss[12]等。

由于神经网络采用了误差反向传播和自动梯度下降的策略, 使它具有传统的机器学习方法不具备的一个重要优势: **自动特征提取**, 即通过网络的设计我们可以将输入的数据编码到高维空间中, 形成高维张量, 称为特征, 这些特征虽然人类难以理解 (可解释性差), 但是能有效帮助网络完成分类或回归任务。自动特征提取省去了传统算法大量的手动特征工程操作。

1.2 深度学习算法在计算机视觉的应用

计算机视觉是人工智能的一个重要分支, 其核心是使用“机器眼”代替人眼。

计算机视觉系统通过图像/视频采集装置，将图像/视频输入到视觉算法中进行计算，最终得到人类需要的信息。广义上的计算机视觉包含更多种类的传感器，除可见光摄像头外，还有红外摄像头、激光雷达、毫米波雷达等。两种或以上的数据来源可称为多模态，更大范围的多模态感知设备包含声学、GNSS、IMU 和皮肤压力传感器等。

随着深度学习技术的发展、计算能力的提升和视觉数据的增长，计算机视觉在不少应用领域都取得了令人瞩目的成绩。例如，图像/识别的分类、识别、检测、分割、生成、超分辨率、搜索等。这些技术正广泛应用于城市治理、金融、工业和互联网、领域。典型的应用场景有：人脸识别、视频监控分析、工业瑕疵检测、图片识别分析、自动驾驶/辅助驾驶、三维图像数据、医疗影像诊断、光写字符识别等。

1.3 研究现状及挑战

1.3.1 计算机视觉算法应用的挑战

随着交叉熵损失[13]、递归神经网络[14]、卷积神经网络[15]等技术的突破，深度学习算法已成为当前最主流的目标计算机视觉算法。近几年来，经典的卷积神经网络的分类算法有 AlexNet, VGGNet, GoogLeNet, ResNet, ResNeXt, HRNet, EfficientNet, ConvNeXt 等，目标检测算法有 R-CNN[16]、Fast R-CNN[17]、Faster R-CNN[18]、Mask R-CNN[19]、SSD[20]、YOLOv1~v5[21-24]、DETR、YOLOX[25]、Swin-Transformer[26]等，分割算法有 FCN, UNet, SegNet, PSPNet, Deeplabv1~v3, SOLOv1~v2, SCNet, MaskFormer, SegFormer 等，生成式算法有 DCGAN, WGAN-GP, PGGAN, StyleGANv1~v3, CycleGAN 等。改进的方面主要是增强信息的流通、提升特征提取的效率等。

在现有范式下对目标检测算法的发展主要是基于技巧（tricks）的改进如 YOLOX[25]等，我们的工作 CDNet[27]也进行了技巧的改进，通过引入注意力机制、负样本训练技巧、感兴趣区域算法、滑动感受野短时向量记忆算法、合成雾增强算法和使用基于 TensorRT 的剪枝、优化和压缩方法，实现了 F1 分数和速度在特定的斑马线检测任务上比原生 YOLOv5 的明显提升。

最近，在自然语言处理（Natural Language Processing, NLP）领域成熟的

Transformer 模型进军计算机视觉领域，刷新榜单并掀起了一波新的改进网络结构的浪潮，例如 Vit-Transformer、Swin-Transformer、SegFormer、MaskFormer 等等。Transformer 强大的一个重要原因是引入了一种全新的“自注意力”机制，使得模型能在后馈传播的同时“注意”到特征内部之间的关联。

除了不断提升精度的需求外，提升速度也是深度学习算法优化的重要目标，其中，YOLOv5 是在网络设计上对速度优化做的比较好的少数网络之一。知识蒸馏类的网络是转为提升速度而设计的，例如 Localization Distillation, Label Assignment Distillation 等。此外，针对速度的优化更多地涉及算力资源的调度、通信等内容，需要针对硬件进行。因此，现状是网络的剪枝、蒸馏、压缩主要依靠算力提供商或通用的工具进行。例如：微软、亚马逊、Facebook, IBM 等公司共同开发的 ONNX Runtime, 英伟达的 TensorRT, 英特尔的 OpenVINO, 百度飞桨的 PP 和其他国产化平台专用的工具集。

目前，深度学习算法在计算机视觉领域的应用中的挑战主要来自于：复杂的环境条件、有限的计算能力、小样本数据、极低的空间分辨率等问题和鲁棒、实时检测之间的矛盾。

1.3.2 发展通用人工智能的挑战

从发展通用人工智能的角度来看，当代以神经网络为代表的人工智能算法存在依赖大数据、可解释性不足、缺少更多类脑机制等问题。由于设计原理，神经网络需要真值，且大量的数据样本，才能展现其优越性，在很多任务中，数据收集和标注已经逐渐称为其性能的瓶颈。神经网络内部是“黑盒子”，虽然有一些方法例如热力图等能展示中间的特征层是什么样的，但人类依旧无法理解它内部是什么，它有效的真正原因除以解释。神经网络设计上没有类似于人类的利用先验知识的能力和进行推理的能力，使得其有时候犹如人工“智障”。这些问题时当代人工智能研究的前沿问题，不同的学者也从不同方向做出努力和尝试，目前尚未解决。

1.4 迈向第三代人工智能

以连接主义为范式的第二代人工智能使用数据、算法和算力三大要素，解决

了非结构化环境下的不确定性问题，是当代最前沿的技术。我们尝试在目标检测领域进行新的探索，目标检测技术是人工智能最基础、最广泛的应用技术之一。在密集拥挤、时间部分遮挡、传感器死角、雨雪雾天、黑暗等不完全信息场景下，目标检测性能将大幅度下降。不完全信息场景下性能大幅度下降是当前深度学习目标检测技术的难点问题。

张钹院士[28]提出发展下一代人工智能技术解决不完全信息问题，发展思路是：把第一代符号主义 AI 和第二代的连接主义 AI 结合起来，通过同时利用知识、数据、算法和算力四个要素，构造知识-数据协同驱动的更强大的 AI。但如何构建、描述知识要素，如何将四个要素有机融合，目前还不甚清楚。

本文独辟蹊径地提出以波函数作为知识要素，以利用量子理论背后的物理思想和数学工具。量子理论是描述和理解世界的基本理论，波函数是其中最核心的内容，复值表示的波函数可以自然地作为带有先验知识的信息的基础。本文拟通过泛函近似准确求解基态波函数，而后利用量子测量和量子演化的数学形式构建基于波函数的知识要素，依靠高性能解算器实现与神经网络的效率匹配，再借鉴量子复合系统的概念对知识信息流和检测结果流两个模态进行建模，最后利用量子纠缠熵和量子概率论实现有机融合，最终实现一套构建知识要素的工具、描述知识要素的理论和利用知识要素的方法。

由于神经网络更具表现力和灵活性，所研究的神经网络泛函将提供比传统泛函更好的精度和时间行为尺度。由于将采用快速积分、分布式并行和即时编译等技术，所研究的高性能解算器将提供与神经网络效率接近的薛定谔方程解算工具。神经网络泛函和高性能解算器是构建知识要素的基础。知识要素由于采用量子演化和冯诺依纠缠熵描述，能利用检测算法丢失的上下文信息，如：演化信息、前后帧信息、目标与环境交互信息、目标间交互信息等，将提供不完全信息补全的能力。所研究的量子启发的融合算法，由于采用量子纠缠理论和量子概率论，将能解决模态不可分离性和不兼容性问题，有效地融合知识信息模态和检测结果模态，实现数据-知识协同驱动，从而取得检测性能的提升。

1.4.1 知识-数据协同驱动算法现状

在新的知识-数据协同驱动范式方面，针对深度学习面临重要的挑战——不

完全信息、数据或知识匮乏下性能大幅度下降和可解释性差、鲁棒性取决于数据集（小样本）等问题，近两年来，人们越来越关注知识要素。在知识和深度学习算法融合方面，Huang 等人[29]提出了领域知识指导的抽象压缩空间；Song 等人[30]对领域知识的结构进行设计；Shi 等人[31]和 Zhou 等人[32]先后使用隐式变分推断算法来解决维数灾难；Wang 等人[33]提出在泛函空间进行粒子优化的算法，来更好地利用贝叶斯方法学习深度神经网络；Chen 等人[34]提出自适应数据增广流模型，来利用深度神经网络作为非线性函数变换从而定义更加丰富灵活的贝叶斯模型；Gallego 等人[35]通过在变分后验近似中嵌入马尔可夫链采样器，以提高贝叶斯概率模型的推理效率。

以上对新算法的探索都是基于信息领域知识的。从物理的角度，以量子理论为基础，将波函数作为知识要素是一条新的道路。

1.4.2 量子理论发展现状

在量子理论方面，百年来随着量子概念的提出，薛定谔方程的建立，波函数意义的统计诠释，泛函分析方法的发展，量子测量、量子演化、量子干涉、量子纠缠、量子概率论等概念的完善，量子理论已发展为一套理解和描述世界的基本理论。

在材料科学领域，基于密度泛函理论（Density Functional Theory, DFT）的 VASP (Vienna Ab-initio Simulation Package)[36]计算程序可以不依赖任何经验参数从头求解薛定谔方程，利用复值表示的波函数携带的描述多体系统的先验知识，来计算材料的结构参数、状态方程、电子、光学、磁学等各项性质，从模拟的角度完成了许多实验无法完成的分析，预测了许多先于实验的正确结果，取得了很大成功[28-33]。泛函近似和薛定谔方程求解工具是成功的两个重要因素。

在泛函近似方面，经典的泛函是 1996 年 Perdew 等人[37]提出的广义梯度近似泛函。近年也来有所发展，2018 年，Lehtola 等人[38]提供了基于描述交换-关联作用的半局域密度泛函库 Libxc；2020 年，Hyldgaard 等人[39]回顾了范德华密度泛函 Vdw-DF、范德华交换密度泛函 Vdw-DF-xc，并提供了计算和映射特征的工具。以上都是对传统泛函的发展。2020 年末，我们独立地提出可以利用神经网络构建新泛函，并进行前期工作时，2021 年初，DeepMind 发表了他们基于神经

网络开发的新泛函: Ferminet[40], 该泛函在精度和时间尺度上实现了 SOTA, 并开启了以前难以处理的在多电子系统中直接准确地优化波函数的可能性。最近, 该神经网络泛函已被应用于多个发表于 *Science*、*Nature* 子刊等期刊的重要研究中[41-47], 这些研究大都是将泛函应用于量子化学、计算材料科学领域。泛函分析是一种求解非线性方程的通用方法, 是本文构建波函数知识要素的基础工具。本文将继续研究神经网络泛函, 独立自主的神经网络泛函对发展新的知识要素描述方法有重要的意义。

在高效求解工具方面, 经典的方法是基于数值计算的自洽求解方法。VASP 通过在 CPU 上优化和并行, 可以实现上百个原子体系的求解。近年来, 基于 GPU 并行的求解方法也逐步得到实现。2020 年, Smidstrup 等人[48]整合了一套原子尺度的建模工具 QuantumATK, 可以实现快速密度泛函理论或紧束缚模型计算。2020 年, Blaha 等人[49]提出了一种带有并行化特性的 Kohn-Sham 方程求解方法 WIEN2k。但上述工具较为专用, 通用性不够好。2018 年, Sun 等人发布了基于 Libcint[50]的量子化学模拟框架程序 Pyscf[51, 52], 底层采用 C 语言提供高效计算能力, 顶层采用 Python 提供良好的接口, 该算法也成为近期涉及多体薛定谔方程求解的众多研究[46, 53-57]的最流行的库。虽然 Pyscf 有进一步的优化[52], 但其效率依旧无法匹配当前的神经网络, 这些研究中大量的时间开销均在方程求解部分而不是神经网络部分。除了求解工具, 辅助计算工具也有所发展[58, 59]。2020 年, 我们发布了 VaspCZ[60], 提供了更便捷的提交任务和检查结果的命令行接口和高通量计算的底层库。利用开发 VaspCZ 的经验, 目前一个基于 Pyscf 的求解库 Wavelib 已实现, 我们将采用更多技术方案进一步提高它的效率以达到与神经网络相匹配的目的。

从量子力学的基本思想, 到泛函近似方法和高效求解工具, 一套基于波函数的描述多体系统的知识体系已经建立, 并在计算材料学领域证明了波函数作为先验知识的可行性。近期的世界前沿也在不断推动量子理论到神经网络的交叉。2021 年, 来自英国开放大学的 Gkoumas 等人[61-65]受量子启发, 借鉴量子测量、量子演化和量子纠缠的概念构建多模态融合模型, 并将其应用于自然语言处理的多模情感分类任务中。目前还还没有将量子理论思想应用于计算机视觉任务的报道。鉴于神经网络和量子理论的发展现状, 我们认为未来三年内在这块量子理论

和人工智能交叉的“处女地”上一定会产出重要的成果，推动新一轮的科技革命。站在巨人的肩膀上，基于前期的工作，有可行的理论、方法和工具来构建、描述和利用波函数知识要素，进而发展出新的目标检测算法，解决不完全信息问题。

1.5 研究内容和意义

本文主要进行深度学习算法在目标检测领域的应用研究和前沿探索。应用研究是基于现有的深度学习算法，针对具体的问题进行网络的改进、数据的增强、提出 *tricks* 等，使得深度学习算法达到精度和速度的需求，赋能应用场景。

本文研究的 CDNet (Crosswalk Detection Network) 在车载摄像头场景下进行鲁棒、实时的斑马线和汽车行为过线分析算法研究，能应用于智能交通管理系统中，所提出的改进方法和技巧对其他神经网络的研究也有借鉴意义。

本文研究的 FINet (Foggy Insulator Network) 在无人机巡检场景下进行绝缘子及其缺陷的检测，能应用于智能电网巡检领域，所提出的合成雾数据增强算法可广泛应用于各种数据集。

本文研究的 AIDNet (Adaptive Insulator Detection Network) 是针对从空间分辨率为 0.5m 的卫星遥感图像中检测直径仅为 0.2m 的绝缘子任务的探索，所提出的网络有效地推进了该任务的实现，所提出的算法可以应用于一般的小目标检测任务中。

本文研究的 TLDNet (Transmission Line Detection Network) 是在无人机巡检领域对输电导线及其散股、断股等缺陷进行检测的网络，所提出的图像质量提升算法、动态背景数据增广算法可以广泛应用于各个领域的低分辨率图像、小样本数据集中。

本文研究的波函数作为知识要素的新型目标检测理论和方法，通过波函数知识要素，发展知识-数据协同驱动的人工智能算法，能解决不完全信息环境下性能大幅度下降的关键难点。综合利用知识、数据、算法和算力四大要素的融合方法，是对下一代人工智能的一种探索，有望进一步发展为新的人工智能范式。基于物理思想和数学表达的通用的理论和方法，在目标检测领域取得成功后，易于推广到诸多其他领域，如融合目标跟踪、带先验知识的语义分割、多模态自然语言处理等，有良好的延伸价值。

1.6 本报告的结构

本文第1章是绪论。其中，1.1节主要介绍人工智能的发展、工具和基础，阐明人工智能发展的三次“春天”和两个范式，并介绍人工智能、机器学习、模式识别、深度学习和神经网络之间的关系，给出简化的实现原理。1.2节简要介绍深度学习算法在计算机视觉领域的应用。1.3节介绍深度学习算法在计算机视觉及子任务目标检测中的研究现状及面临挑战，也引入了当代神经网络与通用人工智能的差距。1.4节介绍了下一代人工智能的发展思路，并在知识-数据协同驱动的思想下，提出使用量子力学波函数作为知识要素发展人工智能的方法。1.5节综述了报告中各项研究的意义。

第2章是方法论。主要分为3个部分：第二代人工智能方法、第三代人工智能方法和评价指标。2.1节介绍在当代人工智能范式下的本文提出或改进的算法。主要是目标检测算法：YOLOv5和引入通道注意力机制改进的SE-YOLOv5，数据相关算法：合成雾/去雾算法、离线数据集增强算法、在线数据集增强算法、图像质量提升算法、动态背景数据集增广算法等。2.2节介绍本文提出的知识-数据协同驱动算法的理论和方法，包括：量子力学基础知识、神经网络泛函的设计与实现方法、高性能解算器的实现原理和协同驱动神经网络的设计与技术路线。2.3节介绍相关评价指标。

第3至6章是在第二代人工智能范式下，将深度神经网络应用在各个具体任务的相关研究，每章均包含：引言、方法、消融实验、结果和分析、总结等几个部分。引言部分描述任务需求，引入科学问题、描述方案和贡献；方法部分描述针对该应用设计和采用的特定方法；消融实验是针对各个算法进行的实验设计以验证有效性。

第7章是波函数作为知识要素的人工智能，主要进行理论推广，阐述如何构建波函数知识要素、如何描述波函数知识要素和如何将波函数知识要素与神经网络结合，分析方法有效的原因。并进行了五个体系的验证实验，给出方案、过程和验证性结果，提出未来的发展路线。

第8章是与本文相关的，在研究过程中设计和实现的软件、平台和Python库。包括多模态融合标注软件、人体行为分析软件、XAI算法平台和Damei Python库。

第 9 章是总结与展望。

第2章 方法论

2.1 第二代人工智能方法

当代人工智能是以连接主义为范式的第二代人工智能。本节主要介绍本文用到的人工智能在目标检测领域的网络结构，并详细描述所提出的网络改进和数据处理算法。

2.1.1 基于YOLOv5的目标检测算法及改进

2.1.1.1 YOLOv5 目标检测算法

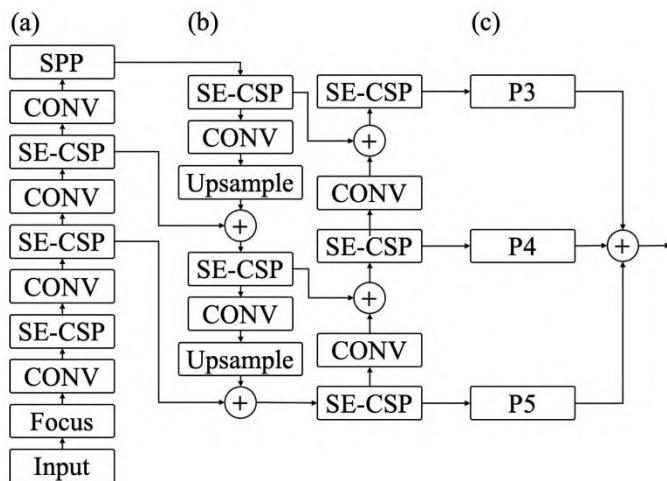


图 2.1 改进的 SE-YOLOv5 目标检测网络

YOLOv5[24]作为2020年SOTA的目标检测算法，由YOLOv3[22]的作者改进并发布，利用各种技术大大提高了检测速度和准确率。在主干部分，Focus结构应用了空间注意力机制，它将输入图像的每个像素上的信息与其周围的信息（右、下和右下像素）连接起来，以增强检测器在空间。基于跨阶段局部网络（Cross Stage Partial Network, CSPNet）[66]的思想，构建了一个bottleneck CSP结构，可以在减少20%的额外计算的情况下提高检测效果。在特征金字塔的最后一层，另一种空间注意机制，即空间金字塔池网络（Spatial Pyramid Pooling Network, SPPNet）[67]，将不同kernel size下最大池化算子计算的特征合并起来，以提升特征提取能力。在颈部部分，YOLOv5利用路径聚合网络（Path Aggregation Network, PANet）[68]通过自下而上的路径增强，在下层使用准确的

定位信号增强整个特征层次来促进信息流, 从而缩短了下层与最上层特征的信息路径。此外, YOLOv5 在训练前采用聚类的方法对训练集的标签进行分析, 自动生成合适的 anchors, 可以加快训练收敛速度, 提高检测精度。

2.1.1.2 SE-YOLOv5 目标检测算法

原生 YOLOv5 中的自动特征提取考虑了每个通道的相等权重。事实上, 每个渠道所包含的信息的重要性是不同的。为了解决这个问题, 主要的改进是将 CSP 改进为 SE-CSP (Squeeze-and-Excitation Cross Stage Partial Network)。如图 2.2 所示, 将通道注意力机制 SE 模块[69]引入 Bottleneck, 在 CSP 内部形成 SE-Bottleneck。

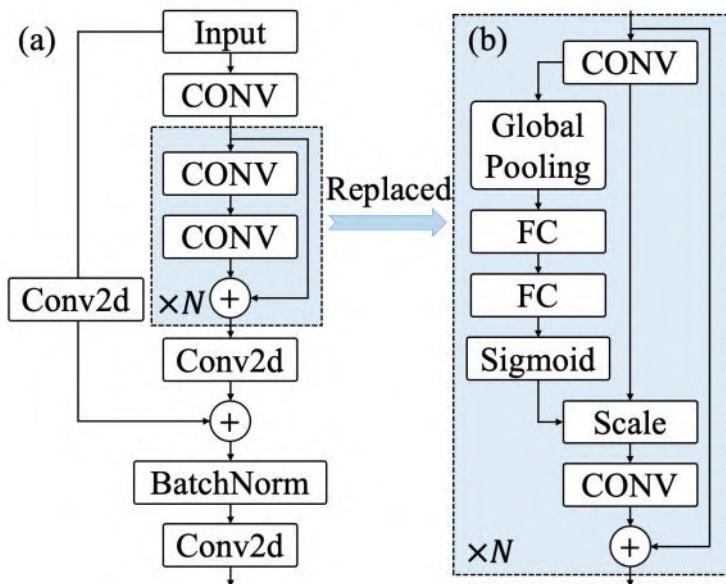


图 2.2 从 CSP 到 SE-CSP 的改进。 (a)本地 CSP 网络, (b)嵌入 CSP 网络的 SE 网络

SE 模块是一种通道注意机制。它在原始残差块的基础上增加了另一条路由, 使用全局池化来获得初始通道权重 (即 Squeeze), 然后使用两个全连接层和 Sigmoid 激活函数来更新每个通道的权重 (即 Excitation), 并且最后将每个通道的权重乘以原始 (即 Scale)。这样, 通过梯度下降来学习每个通道的权重, 以提高网络训练过程中的检测精度。

Squeeze 公式如下:

$$z = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j), z \in R^{H \times W \times C} \quad (2.1)$$

其中 F_{sq} 表示 Squeeze 操作, c 表示每个通道。另外, Excitation 公式如下:

$$s = F_{ex}(z, W) = \sigma\left(W_2 \text{Sigmoid}\left(W_1 z\right)\right) \quad (2.2)$$

其中 F_{ex} 表示 Excitation 操作, $W_1 \in R^{\frac{C}{r} \times C}, W_2 \in R^{C \times \frac{C}{r}}$, r 是超参数, 默认值为 16, 表示第一个全连接层的降维系数。比例尺公式如下:

$$\tilde{u} = F_{scale}(u_c, u_s) = u_c \cdot s_c \quad (2.3)$$

其中 F_{scale} 表示 Scale 操作。

整个 SE 模块实际上学习了每个通道的权重系数, 以便模型更有效地区分特征。

2.1.2 合成雾/去雾算法

使用合成雾算法来扩充数据集, 可以增强模型在有雾情况下的鲁棒性。

2.1.2.1 合成雾原理

在彩色数字图像存储中, 每个像素存储 RGB 三种颜色的值, 值越大, 对应的颜色分量越多。RGB 的三种颜色在灰度图像中均等地合并为一个通道。最大值 (归一化为 1) 表示纯白色, 最小值 0 表示纯黑色。通常, 在彩色图像的大多数非天空区域, 图像像素总是至少有一个值非常低的通道, 称为暗通道先验 [70]。公式如下:

$$J^{dark}(x) = \min_{y \in \Omega(x)} \left(\min_{c \in \{r, g, b\}} J^c(y) \right) \rightarrow 0 \quad (2.4)$$

其中 J_c 表示彩色图像的每个通道, J_{dark} 表示输出灰度图像。使用上述公式计算输出的过程是首先找到每个像素的 RGB 分量的最小值, 然后将其存储在与原始图像大小相同的灰度图像中, 最后进行最小值过滤灰度图像。 $\Omega(x)$ 表示以像素为中心的过滤器窗口。

导致暗通道先验中某个通道的值偏低的三个主要因素。第一个是树叶、树木和岩石等自然景观, 以及汽车、建筑物和城市中玻璃窗的阴影。第二个来自色彩鲜艳的物体或表面。最后一个因素是颜色较深的物体或表面。但是, 雾在原始图像上添加了一个白色蒙版, 使得三个通道的最小值更大。在此基础上, 提出了一种雾像形成模型:

$$I(x) = J(x)t(x) + L(1-t(x)) \quad (2.5)$$

其中 $J(x)$ 表示原始无雾图像, L 表示大气光分量的亮度, $t(x)$ 表示透射率。亮度 L 的取值范围为 0 到 1, 表示添加的雾的颜色, 即灰度值。当 L 分别设置为 0 和 1 时, 将添加纯白雾和黑雾。不同亮度下的起雾效果如图 2.3 所示。亮度 L 越大, 雾越白。

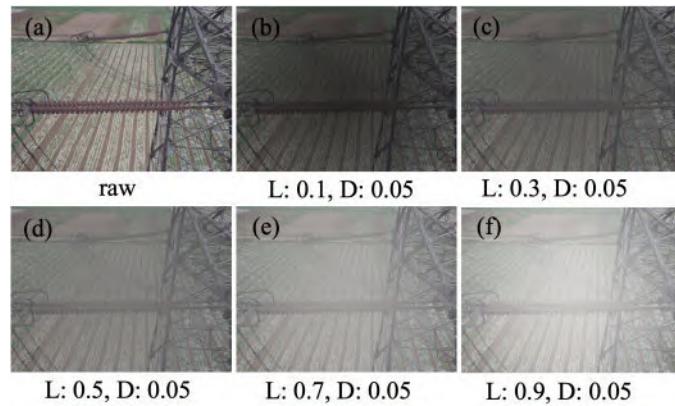
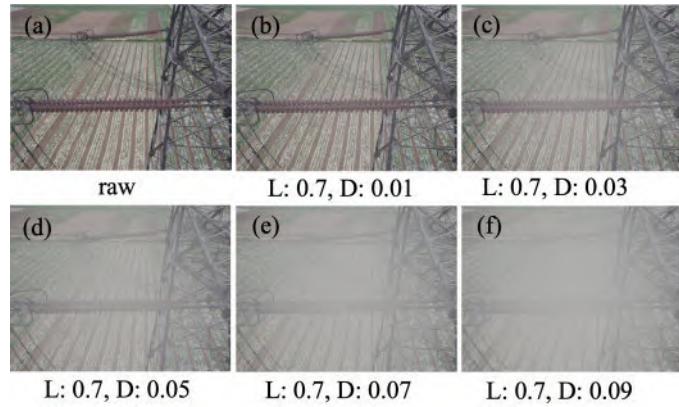


图 2.3 不同亮度 L 下的雾化效果。D 为雾厚度。

公司式 2.5 中, 透光率 $t(x)$ 取值范围为 0~1, 代表输出图像中原始图像和雾的比例。当 $t(x)$ 为 1 时, 表示所有场景都通过了雾, 输出的是纯原始图像。当 $t(x)$ 为 0 时, 任何场景都不能穿过雾, 输出是纯雾图像。因此, 需要为图像中的每个像素设置适当的透射率 $t(x)$ 。公式如下

$$t(x) = \exp \left[-D \left(-0.04 \sqrt{(w-w_c)^2 + (h-h_c)^2} + s \right) \right] \quad (2.6)$$

其中, -0.04 是选定的固定参数。表示图像的中心, 选择雾化中心。 s 表示雾化大小, 即图像宽高最大值的平方根。D 表示雾厚度系数。由式 2.6 可知, D 越大, $t(x)$ 越小, 雾越重, 透光率越小; D 越小, $t(x)$ 越大, 雾越薄, 透光率越大。因此, D 代表雾的厚度。不同雾厚 D 下的雾化效果如图 2.4 所示。

图 2.4 不同雾厚 D 下的雾化效果。L 为亮度。

2.1.2.2 雾化算法的优化

原始雾化算法存在效率低的问题。测试表明，一张 384 万像素的图像起雾时间约为 18 秒。一张原始图像有几百万甚至几千万像素，每个像素都需要循环一次。在每个循环中，都需要创建和分配欧几里得距离和指数透射率等中间变量。这些是效率低下的主要因素。

为了提高雾化算法的速度，采用矩阵计算进行优化。该方法如图 2.5 所示。

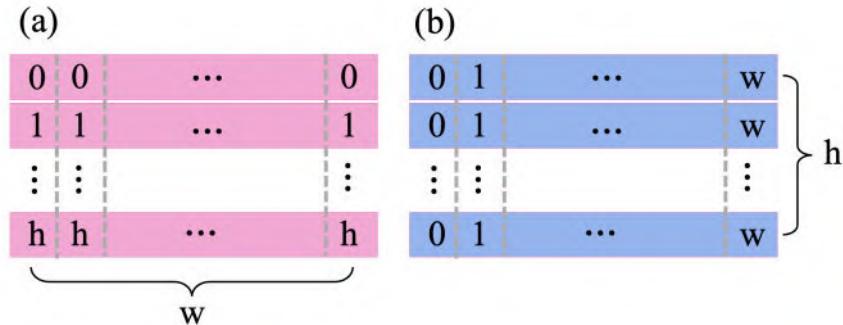


图 2.5 雾化算法的矩阵计算优化

如图，原始图像的宽度和高度分别为 w 和 h 。我们创建一个包含 h 个值（从 0 到 $h-1$ ）的列向量，并沿 1 轴复制 w 次以形成二维矩阵 (w, h) ，如图 2.5 (a) 所示。矩阵的每个元素减去雾化中心的高度 hc 。结果是像素和雾化中心之间的高度差。然后我们创建一个包含 w 个值（从 0 到 $w-1$ 排列）的行向量，并沿 0 轴复制 h 次以形成一个二维矩阵 (w, h) ，如图 2.5 (b) 所示。矩阵的每个元素减去雾化中心的宽度 wc ，结果是像素和雾化中心之间的宽度差。最后，将两个矩阵堆叠形成一个 $(w, h, 2)$ 的数组，并为这个二维数组沿第 2 轴计算 L2 范数。由于该

算法使用矩阵和 L2 范数，省去了每次循环中中间变量的创建和赋值，大大提高了雾化算法的速度。消融实验结果显示优化后速度约为优化前的 40 倍。

2.1.3 离线数据集增强算法

我们还应用了其他数据增强算法，本节将对此进行简要介绍。随机翻转是设置翻转图像的概率。随机仿射变换包括平移、缩放、旋转等。它通过设置变换概率来随机变换或不变换图像。模糊包括随机模糊、运动模糊、中间模糊等。裁剪和填充是对图像中的对象进行裁剪，并将其填充到其他背景上。无损变换是不破坏图像原有空间结构的变换，包括 90、180、270 度旋转，M1、M2、M3、M4 镜像。

在转换过程中，需要对经过真值标注的数据进行相应的变换。我们通过调用 Albumentations API[71]来实现。

2.1.4 在线数据增强算法

为了使神经网络更加鲁棒，在数据加载过程中需要采用一些数据增强算法。我们用于训练的增强算法包括 HSV 色域变换、光照失真、随机缩放、随机左右翻转、随机上下翻转和马赛克数据增强。

2.1.5 图像质量提升算法

2.1.5.1 超分辨率算法

超分辨率 (Super Resolution, SR) 算法，顾名思义，是输入一张低分率 (Low Resolution, LR) 图像，输出一张高分辨率 (High Resolution, HR) 图像的算法。算法能提升图像的分辨率，从而提升品质。

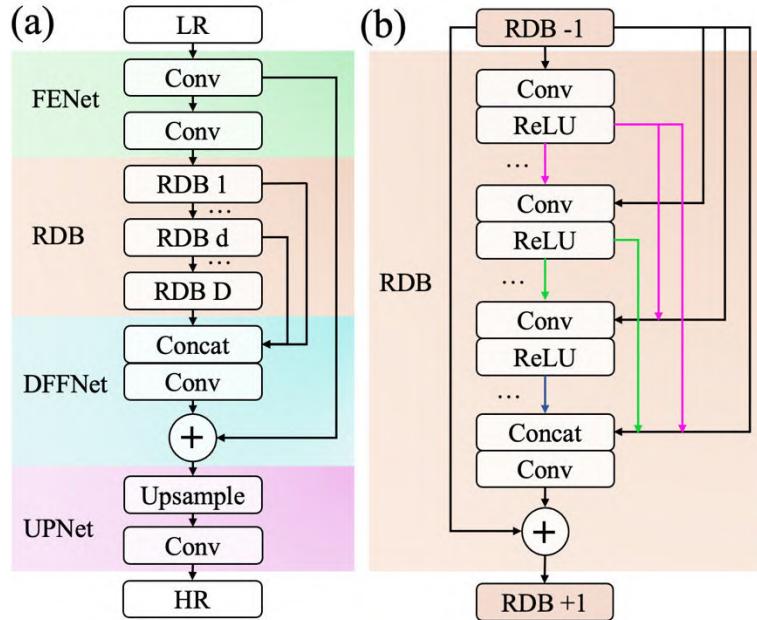


图 2.6 超分辨率网络：(a) 整体网络架构；(b) RDB 模块内部

本文使用的超分辨率算法是基于残差密集网络（Residual Dense Network, RDN）[17]的，网络结构如图 2 所示，整体架构主要分为 4 个模块：特征提取网络（Feature Extraction Network, FENet），残差密集块（Residual Dense Block, RDB），密集特征融合网络（Dense Feature Fusion Network, DFFNet）和采样网络（Upsampling Network, UPNet）。

特征提取网络 FENet 的主要作用是通过两个卷积对输入的低分辨率图像提取特征；残差密集块 RDB 是将经典的残差块（Residual Block）和密集块（Dense Block）进行整合形成的，额外的残差和密集通道能将同时利用前面多个层的特征，有助于提升模型的表达性能，同时能增加网络深度而不会导致梯度消失；密集特征融合网络 DFFNet 使用 Concat、卷积、加权、批归一化将全局特征和全局残差特征进行融合；上采样网络 UPNet 使用上采样和卷积使图像分辨率翻倍。

密集残差块 RDB 是一种连续记忆机制，串联的 RDB 实现了前面所有层到当前层的特征传输。第 c 个卷积层和第 d 个 RDB 层的输出可以表示为：

$$F_{d,c} = \sigma \left(W_{d,c} \left[F_{d-1,0}, F_{d,1}, \dots, F_{d,c-1} \right] \right) \quad (2.7)$$

其中， σ 是 ReLU 激活函数， $W_{d,c}$ 是第 c 个卷积层的权重，所有偏置均被忽略。前面的 RDB 层和后面每个层直接连接到后续层，这样不仅保留了天然的前

馈过程，同时提取了局部的密集特征。

2.1.5.2 高反差保留算法

高斯模糊（Gaussian Blur）是一种图像模糊滤波器，即用正态分布计算图像中每个像素的变换，图像高斯模糊的变换方程为：

$$G(\mu, \nu) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{\mu^2 + \nu^2}{2\sigma^2}\right] \quad (2.8)$$

其中， $r^2 = \mu^2 + \nu^2$ 是模糊半径， σ 是标准差。该方程生成的曲面的等高线是从中心开始呈正态分布的同心圆，每个像素的值都是周围相邻像素值的加权平均。高斯模糊能减少图像噪声和降低细节层次。

高反差保留是指保留图像中反差较大的部分，高反差保留与高斯模糊的关系为：

$$J(x) = I(x) - G(x) \quad (2.9)$$

其中， $J(x)$ 是高反差保留部分， $I(x)$ 是原始图像。对于原始图像，我们进行高反差保留，得到输电导线每股线的边缘信息，然后使用柔光，与原始图像混合，这样就能体现更多的边缘信息，每股线的边缘细节越明显，图像就越清晰。

2.1.6 动态背景数据增广算法

小样本数据增光算法是指将小样本数据离线增广为大数据集的方法。方法概括为：获取图像，标注数据，多尺度背景提取，样本均衡、随机数目、随机位置、随机缩放、随机角度、目标补全生成数据集。本文采用上述方法，使用 13 张小样本数据，可生成了足量标注良好的数据集。

2.1.6.1 数据标注和定义

在进行数据增强之前，需要对数据进行标注和分析。本文使用 Labelme 对提升质量后的 13 张小样本图像进行了标注。本文采用多边形（Polygon）标注方式，而非边界框（Bounding Box, bbox），这样能避免增广变换过程中 bbox 标注包含无用的背景像素，导致的标签歧义问题。

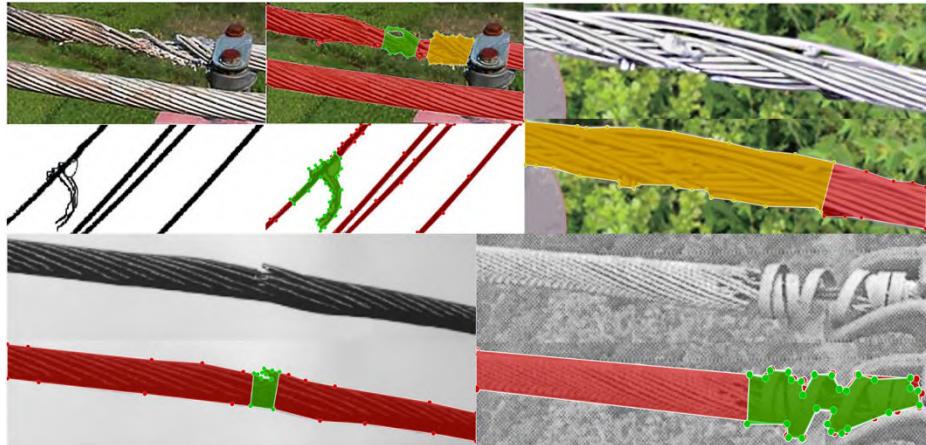


图 2.7 输电导线、断股和散股三类目标，分别用红色、绿色和黄色多边形表示

如图 2.7 所示，我们定义和标注了 3 种目标：输电导线（Transmission Line）、断股（Broken Strand）和散股（Loose Strand）。目标定义如下：

输电导线：包围图像中连续输电导线的最小多边形，用红色多边形标注。不连续的输电导线视为两根导线，如图左上角的上面那个导线，被夹扣分开，则视为两个输电导线目标。

断股：组成输电导线的股线断裂形成的断口的区域，用绿色多边形标注。断口所占像素太小，单独提取断口不易判断是否为断股，因此，本文把断股标注区域扩展为包含与断口区域平行的输电导线的区域。如左侧第二和第三张图，断口仅为垂下的股线和翘起的股线，标注中，将与垂下的股线平行的输电导线和与翘起的股线平行的输电导线也包含进来，形成断股的标注多边形。

散股：组成输电导线的股线散开但未断裂的区域，用红色多边形标注。如图右上角所示，判断散股的方式主要为股线的纹理走向。类似的，散股标注中，也包含了与之平行输电导线。

需要注意的是，三种目标并非完全独立存在，有关联关系。例如，所标注的一根输电导线目标，可能会包含散股、断股目标，后续数据增强中需要考虑。

经过标注，我们在 13 张图像上标注的输电导线、散股和断股目标分别为：22、6 和 11 个。

2.1.6.2 背景图像生成

由于神经网络利用卷积进行特征自动提取，采用单一背景会导致网络收敛到局部最小，而无法寻找到全局最优，从而导致性能不足。因此需要动态生成背景。

(1) 背景生成思想

第一种思想是简单地采用纯黑、纯白或纯灰色(不同灰度)的图像作为背景, 经过测试, 在类似背景的测试集上有较好的性能, 但在彩色背景的测试集上性能大幅下降。第二种思想是采用完全随机的“噪声背景”, 即确定背景图像的宽和高后, 对它的 3 个通道分别随机地赋予 0 到 255, 每个像素点生成随机的 RGB 值。

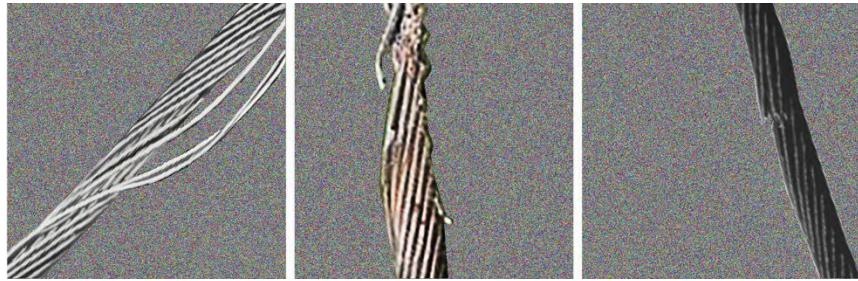


图 2.8 噪声背景合成图样例

噪声背景上添加输电导线目标的合成图样例如图 2.8 所示。虽然理论上每个像素点的 RGB 值是随机的, 神经网络能学习到不同背景信息下的目标。但 4.2 节的实验结果表明, 噪声背景数据集训练的模型, 在只有噪声背景的图像才有良好的检测效果, 当检测“真实背景”的图像时, 性能会大幅度降低。主要原因是“噪声背景”图像和“真实背景”图像的分布不同的。因此, 第二种思想使用 RGB 色彩完全随机“噪声背景”不能得到良好的数据集。

基于上面的分析和实验, 本文采用第三种即基于“真实背景”的动态背景方法。首先, 在网络上寻找合适的背景图像, 如无人机巡检中拍摄到的杆塔、工地、树林、草地等图像, 它的分布是真实的, 去除其中我们关注的目标, 即输电导线、散股、断股等。而后, 我们采用基于动态步长和随机尺度的滑动窗口方法, 把图像切割成不同尺度的背景图。最后, 把基于随机步长和尺度的动态背景与目标合并生成数据集, 标签进行相应变换。

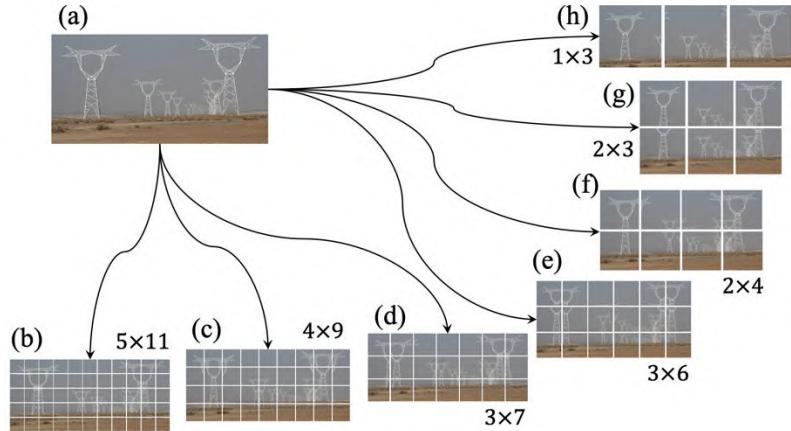


图 2.9 基于滑动窗口生成的多尺度动态背景 (a) 原始图像 (b-f) 尺度从小到大生成的子背景图

我们采用动态背景方法，应用到一张原始图像上，生成的背景图如图 2.9 所示。图中 (a) 是一张 $(3500, 6912, 3)$ 的原始图像，(b-f) 的起始尺寸为 640，公比为 1.25 的等比数列生成的不同尺度下生成的子背景图，每个小的正方形格子的一张子背景。

(2) 动态背景算法

本文设计的基于第三种思想的动态背景生成算法详细描述如下：

朴素的思路是采用滑动窗口方法：设定窗口尺寸为 (a, b) ，正方形窗口 a 和 b 均为 640，从像素坐标 $(0,0)$ 开始，分别沿着 x 轴和 y 轴以步长 $s=640$ 移动，取坐标的窗口图像作为背景。朴素方法存在的问题是：(1) 相邻窗口被完全隔断为不同的背景图，以相邻窗口交接处为主体的背景图实际上是有益的，但被丢弃；(2) 如果原始图像的宽和高不是窗口尺寸的整数倍，原始图像中右侧和下方的有益背景也会被丢弃；(3) 固定的窗口尺寸导致尺度单一，不能利用多尺度信息。我们希望尽可能利用所有有益像素，设计了动态步长和窗口尺寸方法的来构建更多符合真实分布的背景图。

算法如下：

- 1) 设置初始步长 s 为窗口尺寸的 $3/4$ ，即 $s=640*3/4=480$ ，以利用相邻窗口交接处的有益像素。
 - 2) 根据原始图像宽 w 和高 h 修正步长，以充分利用右侧和下方图像信息。
- 设窗口尺寸为 (a, b) ，初始步长 s ，在 x 轴和 y 轴方向上可取的窗口数目 n 分

别为 n_x 和 n_y ，则有关系式：

$$\begin{aligned} w = (n_x - 1)s_x + a &\Rightarrow n_x = \text{round}\left(\frac{w - a}{s_x}\right) + 1 \\ h = (n_y - 1)s_y + b &\Rightarrow n_y = \text{round}\left(\frac{h - b}{s_y}\right) + 1 \end{aligned} \quad (2.10)$$

其中， n 必须为整数，因此，四舍五入对 n 取整，计算新的步长 s' ，有：

$$\begin{aligned} s'_x &= \frac{w - a}{n_x - 1} \\ s'_y &= \frac{h - b}{n_y - 1} \end{aligned} \quad (2.11)$$

3) 设置等比数列滑动窗口尺寸因子，实现多尺度信息提取。

$$\begin{aligned} a_n &= a_0 q^{n-1} \\ b_n &= b_0 q^{n-1} \end{aligned} \quad (2.12)$$

其中，初始窗口尺寸为 (a_0, b_0) ，公比 q 为 1.25。 $n=1,2,3,4\dots$ ，直到窗口尺寸 (a_n, b_n) 大于原始图像尺寸无法取到子图为止。

使用动态背景算法，我们从图 2.9 所示的一张原始背景图中生成了尺度从 640×1.25^0 到 640×1.25^6 ，在不同步长下的子背景图像共计 147 张。

应用到 16 张原始背景图像上，共生成 1271 张子背景图。

2.1.6.3 随机位置、角度和尺寸方法

生成合成图时，为了使得目标位置、角度、尺寸等更加丰富，采用了随机位置、随机尺寸、随机旋转和目标补全等算法。算法如下：

1) 变量定义。把多边形 polygon 标注转换为蒙版 mask，mask 的宽高与背景图像相同（均为 640），通道数为 n ， n 是该张合成图包含的目标数目，每个通道代表一个目标。每个通道的 mask 均是二值化的，即只能取值 0 和 1，0 代表非目标像素点，1 代表目标。相应地定义另一个列表变量 classes，长度为 n ，存储与 mask 的通道对应的目标的类别信息。

2) 随机缩放和旋转。尺寸因子使用以 1.3 为中心，映射到 [0.3, 2.3]，clip 到

[0.5, 2]的随机正态分布获得。旋转因子使用[0, 180°]随机均匀分布获得，旋转中心为目标重心。采用 OpenCV 的旋转缩放变换，输入随机尺寸因子和随机旋转因子获得缩放和旋转矩阵，对蒙版和图像分别执行该变换。

3) 随机位置是在尺寸和旋转变换后，根据目标边界框 `bbox` 的宽和高，采用随机均匀分布计算出 `bbox` 的随机位置，使得目标不出图像的边界范围。对蒙版和图像分别执行平移变换。

4) 处理目标重叠问题。由于每张合成图可能有多个目标，而输电导线目标是“细长”的，随机缩放、旋转和平移后，新目标可能会与已经存在的目标重叠。本文的处理方案是计算新目标与已存在目标的交集百分比 2(Intersection of Percent 2, IOP2)，即新目标和老目标的像素交集与老目标的像素的比值，不断随机取随机因子，直到 IOP2 为 0 时，执行该变换。

5) 目标补全是指多目标的标注进行补全。当我们从样本库随机选择的目标是输电导线时，`mask` 只有 1 通道，即只有一个目标，由于输电导线标注的区域可能会包含散股和断股区域，需要对所包含的散股、断股目标也进行上述缩放、旋转和平移变换变换，将其补全到 `mask` 新的通道上，同时变量 `classes` 进行添加相应元素。

动态背景数据集增广算法已集成到我们开源的 Damei Python 库中，安装后使用 `damei.data.augment` 即可，详细参数可查阅文档[\[72\]](#)。

2.2 第三代人工智能方法

本节描述本文提出的，受量子力学启发，采用波函数作为知识要素的新型人工智能范式，它是一种知识-数据协同驱动的下一代人工智能实现方法。

2.2.1 量子理论基础知识

薛定谔方程：

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + V \right] \Psi(\vec{r}, t) = E \Psi(\vec{r}, t) \quad (2.13)$$

量子理论建立在复值希尔伯特空间上，空间标记为 H 。一个量子态（即波函数）表示为 H 上的一个状态向量，称为右矢 ket，即 $|u\rangle$ ，它的共轭转置向量是左矢 bra，即 $\langle u|$ 。向量 $|u\rangle$ 和向量 $|v\rangle$ 的内积和外积记为 $\langle u|v\rangle$ 和 $|u\rangle\langle v|$ 。

任意系统的纯态波函数 $|u\rangle$ 都可以表示为 n 个基组 $|e_j\rangle$ 的加权和：

$$|u\rangle = \sum_{j=1}^n \sqrt{r_j} e^{i\theta_j} |e_j\rangle \quad (2.14)$$

式中 $\sqrt{r_j} e^{i\theta_j}$ 是复振幅概率， $r_j \in \mathbb{R}$ 是非负标量，满足 $\sum |r_j|^2 = 1$ ， i 表示虚数，

θ 表示相位。

量子混合态波函数 $|\psi\rangle$ 用纯态波函数 $|u\rangle$ 构成的密度矩阵表示，
 $|\psi\rangle = \sum_i P_i |u_i\rangle \langle u_i|$ ，其中 P_i 是组成叠加态的每个纯态向量的概率分布。

从基组到纯态波函数和从纯态波函数到混合态波函数均为量子态叠加。**叠加态**表示量子状态可以以一定的概率同时分布于多个互斥的基态，直到它被测量，测量后坍缩到某一确定的基态上。

多粒子系统的总波函数 $|\psi\rangle$ 可以表达为 N 个粒子的 Slater 行列式的形式：

$$|\psi(\vec{r}_1, \dots, \vec{r}_N)\rangle = |\psi_1 \psi_2 \dots \psi_N\rangle = \begin{vmatrix} \psi_1(\vec{r}_1) & \dots & \psi_N(\vec{r}_1) \\ \vdots & & \vdots \\ \psi_1(\vec{r}_N) & \dots & \psi_N(\vec{r}_N) \end{vmatrix} \quad (2.15)$$

量子演化中, 孤立系统波函数随时间的演化过程是幺正的, 由幺正算子 U 给出。受力场或复杂作用的系统, 演化在哈密顿量下进行, 下一时刻的波函数如下:

$$|\psi(\vec{R}, t+1)\rangle = e^{-\frac{iE_n t}{\hbar}} |\psi(\vec{R}, t)\rangle \quad (2.16)$$

其中, E_n 是哈密顿量的能量本征值, \hbar 是普朗克常量。

量子复合系统是由多个单独的系统组成的复合体系。对于 n_a 和 n_b 维的两个空间 H_A 和 H_B , 一个复合状态向量 $s_c \in H_A \otimes H_B$ 表示为两个空间的任意基 $|i\rangle \otimes |j\rangle$ 的线性组合, 即:

$$|s_c\rangle = \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} \sqrt{c_{i,j}} |i\rangle \otimes |j\rangle \quad (2.17)$$

式中 $\sqrt{c_{i,j}}$ 表示概率幅, 满足 $\sum_{i,j} c_{i,j} = 1$ 。当量子复合系统在包含系统之间相互作用的哈密顿量下演化时, 结果状态不可分离, 即处于纠缠态。

构成复合系统的两个系统的模态可能是兼容的, 也可能不兼容。每个概率事件 A 都有一个投影 Π_A 来描述事件。在量子认知中, 两个事件 A 和 B 的联合概率分布等于两个投影 Π_A 和 Π_B 的乘积, 对应于基态 $A \cap B$ 。如果 $\Pi_A \Pi_B = \Pi_B \Pi_A$, 则称两个事件是兼容的, 否则不兼容。在不相互干扰的情况下, 无法访问两个不兼容事件的联合测量量。经典概率论无法捕获到这种干扰, 总是假设模态是兼容的。量子概率论的数学形式是经典概率论的推广, 同时允许兼容和不兼容的测量。

2.2.2 神经网络泛函方法



图 2.10 神经网络泛函的技术路线

第一部分是神经网络泛函，是构建波函数知识要素的基础。我们将采用网络结构设计、边界条件约束、最小能量优化的技术路线来发展 Ansatz 泛函，采用形式理论和重整化群两种独立的方法来处理微观到宏观的变换。

2.2.2.1 网络结构设计

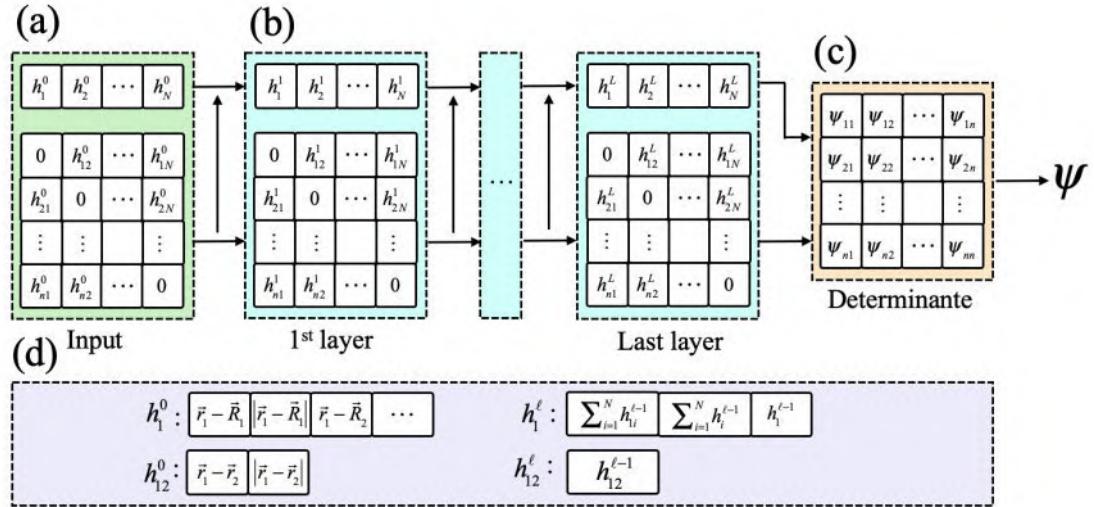


图 2.11 Ansatz 泛函神经网络结构

为了构建能有效表达波函数的神经网络 Ansatz 泛函，我们需要满足公式 2.15 中 Slater 行列式的反对称性，即交换输入变量 \vec{r}_i 和 \vec{r}_j ，只会交换输出矩阵的对应两行或两列，其他部分保持不变。定义交换不换函数为：

$$\psi_i(\vec{r}_j; \{\vec{r}_j\}) = \psi_i(\vec{r}_j; \vec{r}_i, \dots, \vec{r}_{j-1}, \vec{r}_{j+1}, \dots, \vec{r}_N) \quad (2.18)$$

由该交换不变函数构成的单行列式能有效表达任意反对称函数，这种表达方法是构建神经网络 Ansatz 泛函的关键。

拟采用的网络架构如图 2.11 所示。考虑体系具有 N_a 个原子和 N_e 个电子，用 N_k 个行列式来表达交换不变函数，图 2.11 (a) 所示的输入包含单电子流和双电子流，单电子流包含每个电子与核的位置之差 $\vec{r}_i - \vec{R}_i$ 以及欧式距离 $|\vec{r}_i - \vec{R}_i|$ ，双电子流包含电子两两之间的位置之差 $\vec{r}_i - \vec{r}_j$ 和欧式距离 $|\vec{r}_i - \vec{r}_j|$ ，当 $i = j$ 时，矩阵元素为 0，单电子特征的数目是 $4N_a$ 。图 2.11 (b) 所示的中间层采用单粒子流和双粒子流的平均激活合并，并将其添加到单电子流上。单层网络是一个纯线性操作，我们在每一层后添加非线性激活函数后，该网络架构具备了灵活的构建交换

不变函数的能力。在图 2.11 (c) 所示的中间层的最后，采用一个线性变换层作为激活函数，输出被乘以一个指数衰减的波包，用来符合波函数在远离核的时候趋近于 0 的边界条件。总波函数是这些行列式的加权和。

传统的组态相互作用方法如 CI、CISD 等存在尺度一致性的问题，由于神经网络 Ansatz 泛函考虑了每个轨道中两个电子间的任意关联，它有能力学习到尺度一致解。

2.2.2.2 KFAC 优化

与标准的变分蒙特卡洛思想一致，我们采用最小化能量期望值的方法来优化波函数：

$$\mathcal{L} = \frac{\langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle} = \frac{\int \psi^*(\vec{R}) \hat{H} \psi(\vec{R}) d\vec{R}}{\int \psi^*(\vec{R}) \psi(\vec{R}) d\vec{R}} \quad (2.19)$$

式中 \hat{H} 是给定系统的哈密顿量， $\vec{R} = (\vec{r}_1, \dots, \vec{r}_N)$ 是系统的全部电子的位置。

采用 DeepMind 实现的克罗内克因子近似曲率 (Kronecker-Factored Approximate Curvature, KFAC) 作为优化器。KFAC 是适合神经网络的一种自然梯度下降近似方法，它通过如下函数形式来优化 \mathcal{L} ：

$$\delta\theta \propto F^{-1} \nabla_{\theta} \mathcal{L}(\theta) \quad (2.20)$$

其中 F 是费雪信息矩阵 (Fisher Information Matrix, FIM)：

$$F_{ij} = \mathbb{E}_{p(\vec{R})} \left[\frac{\partial \log p(\vec{R})}{\partial \theta_i} \frac{\partial \log p(\vec{R})}{\partial \theta_j} \right] \quad (2.21)$$

KFAC 优化器由于引入了神经网络线性层的梯度结构信息，优化速度优于常用的随机一阶优化方法如 ADAM 等。

上述神经网络泛函设计与 Ferminet[42]相近，但不包含自旋，以满足宏观目标没有磁性的条件。与 Ferminet 使用 Tensorflow 框架不同，我们将在 Pytorch 框架下实现神经网络泛函，以与 Wavelib 库匹配，并利用 Pytorch 1.6 版本以后的复数卷积特性。

2.2.2.3 形式理论和重整化群

本文中我们主要采用形式理论 (Formalism) 来处理微观到宏观的变换。原理

是神经网络泛函的输出概率分布 $P(\vec{X})$ 并不是一个平均场理论的近似解，而是一个 Hartree-Fock 轨道的等效混合，即这是一种数学上的形式表达。利用在 H 空间的一般数学形式，以检测结果和知识信息两个系统构建双模复合希尔伯特空间，神经网络泛函的卷积特征表示为布洛赫球体表面的纯态向量，利用 POV(Projection-Valued Measure) 测量把系统状态从不确定的状态坍缩到精确状态上，根据么正时间演化后的可观测量预测每个目标的位置和类别。该思路 Gkoumas 等人[61, 62]在 NLP 领域采取的形式策略类似，即采用一般的数学形式直接描述宏观概念，跳过微观到宏观的过程，虽然会缺失物理意义(可解释性)，但简单而有效。

此外，本文将采用重整化群 (Renormalization Group) 方法研究另一种从微观到宏观的变换。基本原理类似于处理伊辛模型的块自旋方法，先让神经网络编码的基本态波函数是微观的基本态，然后对粒子分区，计算每个区的叠加平均波函数，以此来代替区内每个粒子的波函数，如此重复，系统不断被缩放，从而从微观尺度获得宏观尺度的整体行为。近来，Facebook 的 Yann LeCun 团队的科学家也在研究这种基于重整化群的有效理论方法，试图从第一性原理的角度深入理解神经网络[73]。该方法能提供真正意义上的微观到宏观变换，但困难在于多次重整化极大的计算量，是未来量子计算机发展后合适的技术路线。本文研究将涉及重整化群理论，但主要的实现采用形式理论。

2.2.3 高性能计算器实现原理

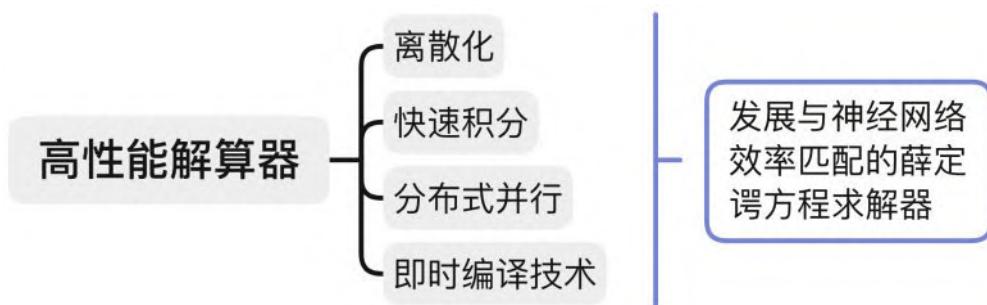


图 2.12 高性能解算器技术路线

第二部分是高性能薛定谔方程解算器，是构建波函数知识要素的重要一环。所采取的技术路线是：利用差分思想，将二阶偏微分物质波场薛定谔方程写成矩

阵形式, 采用离散化数值方法求解; 采用快速积分实现高效计算; 采用 CUDA 编程、GPU 并行和分布式计算利用 GPU 性能; 采用即时编译技术缩短执行时间、减少内存操作、提升可移植性。

2.2.3.1 离散化

采用分离变量法对薛定谔方程 3.1 进行空间和时间变量分离, 得到不含时薛定谔方程, 其是一个二阶偏微分方程, 为了用计算机 (非量子计算机) 求解, 我们采用如下方法对齐离散化。

一维情况下不含时哈密顿量为:

$$\hat{H} = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \quad (2.22)$$

我们将连续空间自变量 x 离散到离散步数 N , 采用极限思想、前馈差分和后馈差分法推导得到二阶微分的矩阵表达如下式第一项所示, 势场 V 可以简单地表达为对角元素矩阵。因此, 哈密顿量的离散化矩阵形式为:

$$\hat{H} = -\frac{\hbar^2}{2m} \begin{bmatrix} -2 & 1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & \cdots & 0 \\ 0 & 1 & -2 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 1 & -2 \end{bmatrix} + \begin{bmatrix} V_1 & & & & \\ & V_2 & & & \\ & & \ddots & & \\ & & & & V_{N-1} \end{bmatrix} \quad (2.23)$$

在离散空间的边缘 (最左侧和最右侧), 该方法给出的数值解是不可靠的, 但考虑到 2.2.2.1 节设计的 Ansatz 泛函中网络的最后一层采用了波包技术使得边界处的波函数趋向于 0, 因此这种不可靠无影响。

2.2.3.2 快速积分

我们通过调用量子化学高效积分库 Libcint 来实现快速积分, 其采用 C 语言实现了标量和旋量的通用积分, 可以评估任意算符形式的积分表达式, 并在 CPU 上达到数倍 I/O 带宽的整体吞吐量。我们的前期工作编写了一个波函数库 Wavelib, 把 Libcint 封装成 Python 函数, 并开发了额外的函数, 方便与神经网络交互。Wavelib 实现了主要的积分有: 重叠积分、电子排斥积分、动能积分、核吸引积分、反库仑相互作用积分等。后来发现, 近年的 Pyscf, Dftatom 和 Ferminet 等涉及第一性原理的代码内部实现也调用了 Libcint 库, 从另一个角度说明该库的有效性。

2.2.3.3 分布式并行

本研究将继续进行 Wavelib 在 GPU 上的实现、并行和分布式计算。在 GPU 上的运行，拟采用 CUDA 编程实现。在单机多卡上的并行计算，拟采用 Nvlink 方法实现卡间通信。在多机上的分布式计算，拟采用 Infinteband 连接方案和 GPU Direct RDMA 技术实现不同主机上的卡间直连，以实现达到甚至超过单机 PCIe 的通信速度。

2.2.3.4 即时编译

本研究还将研究采用即时编译（Just-in-time compilation, JIT）技术提升 Wavelib 的速度。采用 Google 推出的加速线性代数（Accelerated Linear Algebra, XLA）编译器来实现 JIT，JIT 不断分析正在执行的代码，并确定某些部分编译所获得的加速将超过该编译带来的开销，然后使用 XLA 来编译这些代码，以此来加速 Wavelib，从而提升薛定谔方程的求解速度。此外，采用 XLA 也可以加速上节 Asantz 泛函的神经网络。XLA 将神经网络操作符转为计算图，再转为 XLA 高级优化器（XLA HLO），而后编译子图，把多个计算图融合，缩短执行时间；分析内存使用，消除内存操作；减少对自定义操作符的依赖，减少移动端库的大小，提高可移植性。

2.2.4 协同驱动神经网络

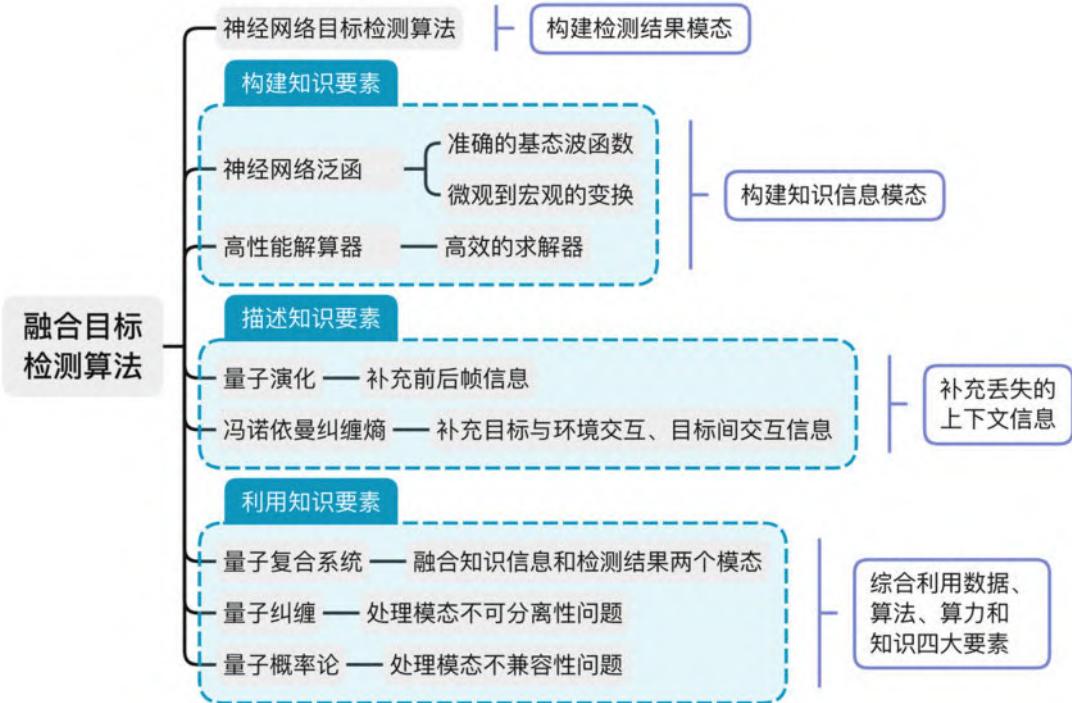


图 2.13 双模态融合目标检测算法的技术路线

第三部分是知识-数据协同驱动融合目标检测算法，是最后也最为关键的一环。所采取的技术路线是：利用神经网络目标检测算法构建检测结果模态；利用神经网络泛函和高性能解算器构建知识要素，利用量子演化和冯诺依曼纠缠熵描述知识要素，利用量子复合系统、量子纠缠和量子概率论利用知识要素，实现双模态融合算法。

2.2.4.1 神经网络目标检测算法

本文的神经网络目标检测算法采用我们在 CDNet[27] 中改进的 SE-YOLOv5，主要使用了空间注意力机制、特征金字塔池化、跨阶段局部网络、特征重用、路径聚合网络、经典的残差网络等技术，改进主要通过通道注意力机制（Squeeze-and-Excitation）实现。

2.2.4.2 构建知识要素

构建知识要素的工具是本文研究的神经网络泛函和高性能解算器。方法是：视目标检测任务中的一个目标为单粒子，多目标检测任务视为对多粒子系统的每个粒子进行位置测量，根据不确定性原理，在使动量分布平均化的过程中，位置测量将有极高的准确率。具体做法是：构建每个粒子的哈密顿量（最简情形是不受任何作用的“自由粒子”，哈密顿量只包含动能项，不包含势能项，对应于经

典牛顿运动定律模型), 采用第一部分研究的神经网络泛函近似获取该哈密顿量下的准确基态波函数(即纯态波函数), 该波函数是由基组 $|e_j\rangle$ 张成的希尔伯特空间下的态矢; 使用神经网络目标检测算法进行检测, 利用反傅里叶变换将检测结果投影到多个互斥的基态上, 并利用 POV 测量使其坍缩到某个确定的基态。

2.2.4.3 描述知识要素

坍缩后的波函数随着时间发生演化, 下一时刻的波函数由公式 2.16 给出, 式中的能量本征值 E_n 通过在神经网络泛函下用高性能解算器求解薛定谔方程获得。这样, 由量子演化描述的波函数知识要素能提供前后帧信息的补充。

在复杂相互作用下, 例如: 单帧内目标-目标间的相互作用、目标与环境的相互作用等, 相互作用的大小可以用冯诺依曼纠缠熵来描述。两体纠缠度可以表示为:

$$V = -\frac{1}{2} \left[\ln \left| \det \left(e^{-N} - 1 \right) \right| + \text{tr} \frac{N}{1 - e^{-N}} \right] \quad (2.24)$$

式中 N 为负厄米矩阵, 可由量子变换理论和波函数密度矩阵求得。这样, 由冯诺依曼纠缠熵描述的波函数知识要素能提供目标-目标间、目标-环境间交互信息的补充。

2.2.4.4 利用知识要素

在下一时刻, 神经网络目标检测算法的结果和波函数的先验知识信息, 分别提供了对目标位置的预测, 前者拥有强大的特征提取和分类能力, 后者包含了前者可能丢失的上下文信息。

我们将检测结果和知识信息视为两个模态: D (Detection)和 K (Knowledge), 将它们预测的多目标位置视为两个多粒子系统, 利用量子复合系统的概念, 进行双模态融合。融合的方法是: 先把模态 D 的结果坍缩到一个新的希尔伯特空间 H_D 里, 然后将 H_D 和 H_K 两个空间组合成一个复合希尔伯特空间 H_{DK} , H_{DK} 中的状态向量如公式 2.17 所示, 新的复合波函数由 POV 测量给出。

当复合系统在包含相互作用的哈密顿量下演化时, 结果状态不可分离, 即处于纠缠态。例如: 在目标忽然被遮挡时, 模态 D 很有可能丢失目标, 人类认知中“猜测”丢失目标的位置, 是基于上下文信息的(例如前一帧目标的位置、速度,

目标所处的环境、目标与其他目标的交互等), 本架构中, 这种先验知识包含在模态 K 内, 这意味着不能单独地考虑一个模态, 两个模态是不可分离的, 即模态的不可分离性。量子纠缠是目前唯一能模拟不可分离性的理论, 我们再次利用冯诺依曼纠缠熵 (公式 2.24), 来度量复合纯态的量子纠缠程度。当模态中有一种出现模棱两可或无信息时, 纠缠程度最高; 当两种模态的上下文信息丰富、明确且同时存在时, 纠缠熵接近于零。纠缠熵为零意味着模态是可分离的, 不需要用量子概率解释, 问题简化为经典概率叠加。这样, 量子启发的双模态融合目标检测算法具备了捕获可分离和不可分离的两个模态之间交互的能力, 是融合算法性能将取得提升的主要原因之一。

另一个问题是两个模态可能存在不兼容性, 即不同的预测顺序可能导致有争议的结果判断。例如: 人类认知中, “先入为主”的信息可能会导致“后入”的正确信息的失效。双模态融合目标检测框架中, 先 D 后 K 和先 K 后 D 可能会产生不同的结果。如 2.2.1 节所述, 对两个事件 A 和 B 的测量量求对易之差, 可以评估模态的兼容程度:

$$E = |\Pi_A \Pi_B - \Pi_B \Pi_A| \quad (2.25)$$

式中, E 是不兼容性, E 为 0 时, 两个模态兼容, 问题简化为经典概率论问题, E 不为 0 时, 模态不兼容, 存在先入为主的判断。我们在 POV 测量中包含不兼容性变量 E, 算法就具备了处理兼容和不兼容的两个模态的能力, 是另一个融合算法性能取得提升的主要原因。

综上所述, 融合目标检测算法包含神经网络和知识信息两个空间。前者主要是数据驱动的检测, 后者则是知识驱动的预测。两个空间交互的工具是神经网络泛函、高性能解算器和量子测量。两个空间信息融合的方法是量子复合系统、量子纠缠和量子概率论。在不完全信息环境下, 如密集拥挤时, 利用纠缠熵补充目标-目标间的交互信息; 雨雪雾天时, 利用纠缠熵补充目标-环境的交互信息; 时间部分遮挡时, 利用量子演化补充的前后帧信息; 传感器死角和黑暗时, 同时利用可见部分波函数坍缩为 0 的先验知识和量子演化补充的前后帧信息, 这些上下文信息与神经网络的有机融合, 是本文所提的量子启发的不完全信息场景下的知识-数据协同驱动的融合目标检测算法性能将取得提升的核心原因。

2.3 评价指标

本节描述本研究中所涉及的性能评价指标。

2.3.1 混淆矩阵

混淆矩阵 (Confusion matrix) [74, 75] 也称误差矩阵, 用 n 行 n 列的矩阵来进行精度评价, 基于混淆矩阵的具体评价指标有准确率、精确率、召回率、特异度和 F1 分数等, 从不同侧面反映分类问题的精度。混淆矩阵由真值和预测值统计得到, TP (True Positive) 代表真值为某类并正确预测为该类的数量, TN (True Negative) 代表真值不为某类并正确预测为其他类的数量, FP (False Positive) 代表真值不为某类但错误地预测为该类的数量 (即虚警或假阳性, I 类错误), FN (False Negative) 代表真值为某类但错误地误测为其他类的数量 (即漏检或假阴性, II 类错误)。

表 2.1 三分类问题的一个混淆矩阵

		Predict Class		
		Cat	Dog	Rabbit
Ground Truth	Cat	6	3	1
	Dog	2	5	1
	Rabbit	0	2	10

一个三分类问题的示例混淆矩阵如表 2.1 所示, 对于猫这个类别, TP 为 6, TN 为 18 ($5+1+2+10$), FP 为 2 (2 只狗被错误地预测为猫), FN 为 4 (3 只猫被错误地预测为狗, 1 只猫被错误地预测为兔子)。基于此, 可以计算反映不同侧面的指标:

$$\begin{aligned}
 ACC &= \frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}} \times 100\% \\
 P &= \frac{N_{TP}}{N_{TP} + N_{FP}} \times 100\% \\
 R &= \frac{N_{TP}}{N_{TP} + N_{FN}} \times 100\% \\
 SPC &= \frac{N_{TN}}{N_{TN} + N_{FP}} \times 100\% \\
 F_{measure} &= \frac{(\alpha^2 + 1) \times P \times R}{\alpha^2(P + R)} \times 100\%
 \end{aligned} \tag{2.26}$$

其中, 准确率 (Accuracy, ACC) 反映总体的预测效果; 精确率(Precision, P)是该类的正确预测占所有预测值为该类的比例, 也称为查准率, 反映预测的精度 (精确率为 1 说明所有预测为该类的结果都是正确的); 召回率 (Recall, R) 是该类的正确预测占所有真值为该类的比例, 也称为查全率, 反映预测的灵敏度 (召回率为 1 说明所有真值中存在的该类都被预测出来了); 特异度(Specificity, SPC) 是其他类的正确预测占所有真值为其他类的比例, 反映预测负样本的灵敏度; $F_{measure}$ 是加权考虑 P 和 R 的综合指标, 最常见的是取参数 $\alpha = 1$ 的 F 分数, 称为 F1 分数, 能同时反映查准率和查全率。

准确率、精确率、召回率、特异度、F1 分数等指标适用于所有分类问题。

2.3.2 ROC 空间

信号检测理论中, 受试者工作特征曲线 (Receiver Operating Characteristic curve, ROC) 是一种坐标图式的分析工具, 又称为感受性曲线(Sensitivity curve), ROC 得名于曲线上各点反映相同的感受性, 它们都是对同一信号刺激的响应, 只不过在不同的判断标准下所有的结果而已。

当信号侦测 (或变量测量) 的结果是一个连续值时, 类与类的边界必须用一个阈值 (即判断标准) 界定, 不同的阈值下 TP、FP、TN 和 FN 会有很大差异。基于混淆矩阵定义真阳性率 (True Positive Rate, TPR) 和假阳性率 (False Positive Rate):

$$TPR = \frac{N_{TP}}{N_{TP} + N_{FN}}$$

$$FPR = \frac{N_{FP}}{N_{FP} + N_{TN}}$$
(2.27)

TPR 代表在所有实际为阳性的样本中, 被正确地判断为阳性的比例, 等价于召回率 R; FPR 代表在所有实际为阴性的样本中, 被错误地判断为阳性的比例, 等价于 1-特异度 SPC。以 FPR 为横坐标, TPR 为纵坐标, 形成 ROC 空间, 对角线把空间划分为两个区域, 左上区域代表较好的分类结果 (胜过随机分类), 右下区域代表差的分类结果 (劣于随机分类), 完美的分类器位于左上角。

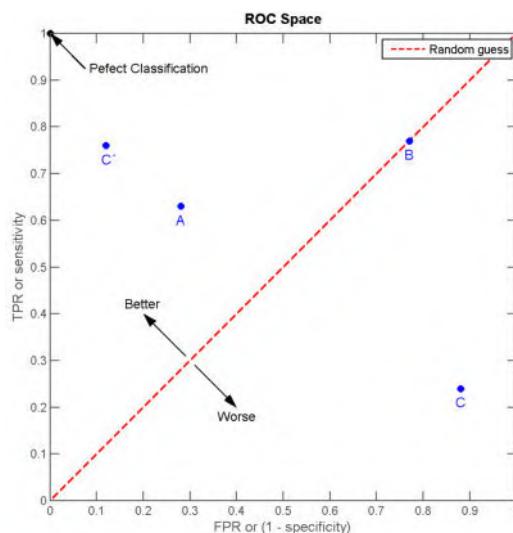


图 2.14 ROC 空间

当一个分类模型和它的阈值确定后, 就能获得空间上的一个点, 确定不同的阈值, 可以获得多个点, 将这些点与(0, 0)点和(1, 1)点连成线, 形成 ROC 曲线, ROC 曲线下的面积是 AUC (Aera Under the Curve)。

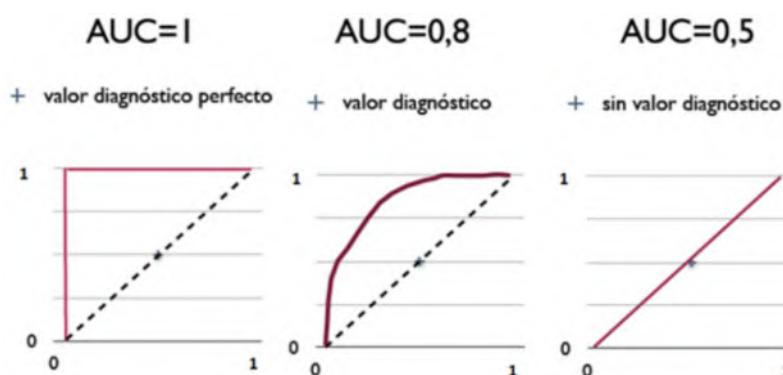


图 2.15 三种 ROC 曲线和 AUC

图 2.15 示例了三种 ROC 曲线和 AUC，最完美的模型的 AUC 为 1，随机分类的 AUC 为 0.5，AUC 衡量了不同阈值下真阳性率和假阳性率，是一种综合指标。

ROC 分析被广泛应用于，医学、无线电、生物学、犯罪心理学领域中，近年来在机器学习和数据挖掘领域也得到了很好的发展。例如目标检测领域中，用精确率 P 和召回率 R 作为横纵坐标构成 ROC 空间，在不同 IoU、不同置信度阈值下，曲线下下的面积 AUC 被定义为 mAP，是一个综合和通用的指标。

2.3.3 交并比 IoU

在描述 mAP 之前，需要先弄清楚 IoU 是什么。

交并比（Intersection of Union, IoU）是衡量两个集合相似度的重要依据，在目标检测领域的被广泛应用于预测和真值的 bounding box 对比，作为一种损失函数，也可以需要作为性能评价的指标。

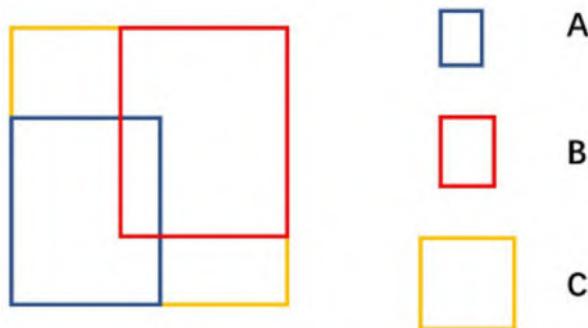


图 2.16 交并比

如图所示，IoU 是预测框 A 和真实框 B 之间的交集和并集之比：

$$IoU = \frac{A \cap B}{A \cup B} \quad (2.28)$$

2.3.4 mAP

目标检测指标是指评估目标检测算法的性能的指标。

指标有：AP、AP@0.5、AP@0.5:0.95、APS、APM、APL、mAP、mAP@0.5、mAP@0.5:0.95、mAPS、mAPM、mAPL 等。

在目标检测任务，会同时输出检测到的目标的位置（如 bounding box）和置

信度 (Confidence)，会存在两个阈值：一是根据预测的 bbox 和真值 bbox 计算得到的交并比 (IoU, Intersection of Union)，二是置信度。不论是 IoU 阈值还是置信度阈值发生变化，在统计混淆矩阵时得到的 TP、TN、FP 和 FN 都会发生变化，最终 F1 分数等指标也会不同。为了更全部地考虑模型在不同阈值下的性能，基于信号检测理论，ROC 曲线 (Receiver Operating Characteristic curve) 下的面积 AUC (Area Under the Curve) 被应用来开发出 AP、mAP 等指标。

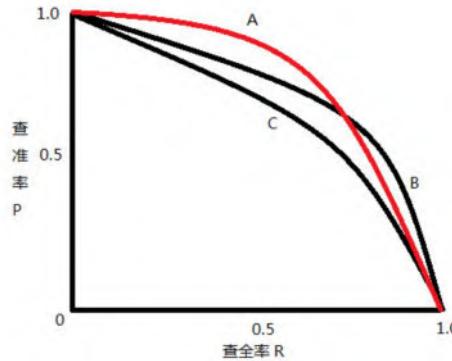


图 2.17 目标检测的 ROC 曲线 (PR 曲线)

平均精度 (AP, Average Precision)，在一个确定的 IoU 下 (一般取 0.5)，取不同的置信度阈值，通常阈值越高，查准率 P 越高，查全率 R 越低，反之阈值越低，查准率 P 越低，查全率 R 越高，在不同置信度下获得一系列 P 和 R，以 P 为纵坐标，R 为横坐标，绘制出多个点，不同的点从左上到右下连线，这样绘制出 ROC 曲线 (即 PR 曲线)，曲线下的面积定义为 AP。

$$AP = \int_{R=0}^1 P dP \quad (2.29)$$

AP 是反映确定 IoU 不同置信度下的综合指标。确切地说，Pascal VOC 数据集的 AP 是指 IoU 为 0.5 时的 AP@0.5。类似地，在不同的 IoU 阈值下，取不同的置信度阈值，可以得到 AP@0.7、AP@0.9 等指标。从 0.5 到 0.95，间隔 0.05 取 10 个 IoU 阈值，得到 10 个 AP，并求和平均得到的指标是 AP@0.5:0.95。AP@0.5:0.95 是反映不同 IoU 不置信度下的综合指标，MS-COCO 数据集兴起后，通常采用的 AP 是 AP@0.5:0.95。

此外，有时需要分别获得对不同尺寸的目标的检测性能，指标 APS, APM, APL 分别对小目标、中目标和大目标进行计算。

上述 AP 系列指标是针对单类目标检测进行的，考虑多类目标时，对每类目

标的 AP 进行求和平均即得到了所有类平均精度 (mAP, mean Average Precision):

$$mAP = \frac{1}{class} \sum_{class} AP \quad (2.30)$$

与 AP 类似, mAP 也有 mAP@0.5, mAP@0.5:0.95, mAPS, mAPM, mAPL 等指标, mAP 系列指标衡量了多目标类别的综合检测能力。

AR 指标是指在设定一个值, 在每张图片中, 在给定值数量的检测结果中计算最大召回率 (max recall), 然后对所有 IoUs 上求平均。mAR 是在所有类别上对 AR 求平均。

2.3.5 所需的空间分辨率 rSR

所需的空间分辨率 (Required Spatial Resolution, rSR) 是在卫星遥感图像中进行目标检测的任务中。rSR@.8F1 是指表示目标检测的 F1 分数至少达到 80% 时所需的空间分辨率, 单位是 mm 等距离单位。

2.3.6 合成图准确性

合成图的准确性是为了确保了增强数据集的准确性和可靠性设计的。它不是一个数值类指标, 而是一个脚本。通过 `check_dataset` 将生成的标注绘制到合成图像中并通过手动采样进行测试的工具。测试工具 API 可在 Damei Python 库文档中找到[\[72\]](#)。

第3章 CDNet：基于YOLOv5在Jetson Nano上实时和鲁棒的斑马线检测网络

在复杂场景和有限计算能力下实现实时、鲁棒的斑马线检测是当前智能交通管理系统（ITMS）的重要难点之一。有限的边缘计算能力和多云、晴天、雨天、雾天和夜间等真实复杂的场景同时对这项任务提出了挑战。本研究提出基于YOLOv5的斑马线检测网络（CDNet），实现车载摄像头视角下快速准确的斑马线检测，并在Jetson Nano设备上实现实时。强大的卷积神经网络特征提取器用于处理复杂环境，网络中嵌入了squeeze-and-excitation（SE）注意力机制模块提升性能，使用负样本训练（NST）方法提高准确率，提出区域利用兴趣（ROI）算法进一步提高检测速度，提出了一种新的滑动感受野短时向量记忆（SSVM）算法来提高车辆过线行为检测精度，使用合成雾增强算法允许模型适应有雾的场景。最后，在Jetson nano上以33.1 FPS的检测速度，我们在上述复杂场景中获得了94.83%的平均F1分数。对于晴天和阴天等更好的天气条件，F1分数超过98%。该工作为人工神经网络算法优化方法在边缘计算设备上的具体应用提供了参考。数据集、教程和源代码可在GitHub上获得。

3.1 引言

斑马线（又称人行横道）是重要的交通安全标志。根据新的交通法规，车辆通过斑马线时应减速，行人通过时停车。由于需要对大量车辆进行持续管理，仅靠交警是不可能维持秩序的。借助人工智能技术的智能交通管理系统（ITMS）有望解决这一问题，这也是当前的发展趋势。

车载摄像头是ITMS的重要传感器之一，基于车载摄像头的斑马线检测信息和车辆横穿行为分析为ITMS提供决策依据。由于边缘计算能力的限制，准确、实时地获取这些信息仍然是一个重要的挑战。

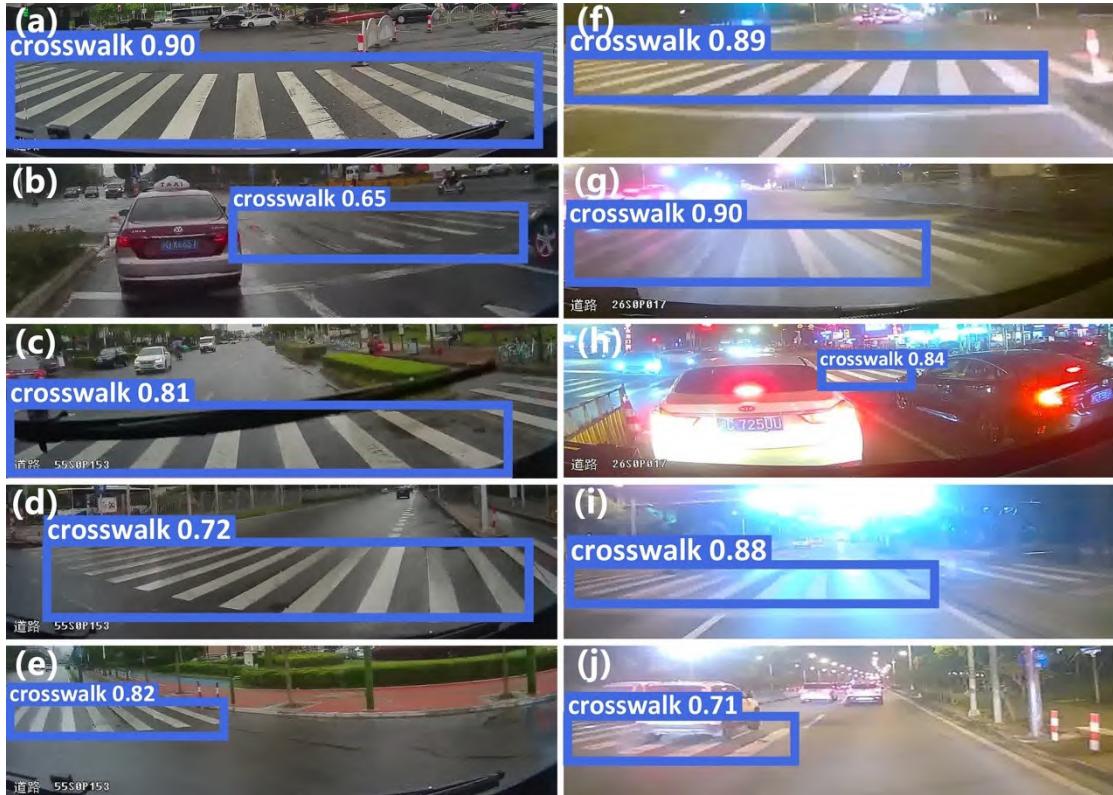


图 3.1 真实复杂场景中的斑马线检测示意图。(a) 清晰可见, (b) 被雨水反射, (c) 被雨刷遮挡, (d) 扭曲, (e) 转向截断, (f) 在夜间, (g) 损坏, (h) 视野被车辆挡住, (i) 眩光, 以及 (j) 部分丢失

如图 3.1 所示,对于真实的车载摄像头图像,传统方法难以处理复杂场景下的斑马线检测,如雨水反射、雨刮遮挡、转向失真、截断、夜间、损坏、视野被车辆遮挡、眩目、部分丢失,以及它们的组合。

本研究围绕在边缘算力设备上实现快速准确的斑马线检测的难题,研究基于 YOLOv5 的针对具体问题的深度学习算法优化方法。我们将 SE 模块插入到 YOLOv5 网络中。以斑马线为对象,创建斑马线数据集,提出提高速度的兴趣区域 (Region of Interest, ROI) 算法,使用负样本训练 (Negative Samples Training, NST) 方法提高准确率,提出滑动感受野短期向量记忆 (Slide receptive field Short-term Vectors Memory, SSVM) 算法将检测任务转移到车辆过线行为检测任务,提高了准确性,并实现了合成雾增强算法,使数据集加倍。最后,在白天、阴天、雨天、雾天和夜间的复杂场景下,在 Jetson nano 上以 33.1 FPS 的检测速度,我们分别实现了斑马线检测和过线行为检测的 81.38% 和 94.83% 的 F1 分数。本

文的贡献总结如下：

- (1) 将 SE 模块插入 YOLOv5，开发出 SE-YOLOv5 模型。
- (2) 应用 NST 算法在不改变检测速度的情况下提高检测精度。
- (3) 提出了 ROI 算法来提高斑马线检测速度。
- (4) 提出了一种 SSVM 算法，将斑马线检测任务转移到车辆横穿斑马线行为检测任务中，提高了检测精度。
- (5) 实现了合成雾增强算法以使数据集加倍并允许模型适应有雾的场景。
- (6) 构建了一个包含 10408 张图像和注释的斑马线检测数据集，提供了车载摄像头采集的共计 3 小时的真实场景视频。

CDNet 的数据集、教程和源代码可在 GitHub[76]上获得，为人工神经网络算法优化方法在边缘计算设备上的具体应用提供参考。

3.2 斑马线检测相关研究

对于斑马线检测，目前多采用传统方法。2000 年，Se[77]提出消失点法，即在图像中寻找两个消失点作为一条线约束，然后使用 RANdom SAmples Consensus (RANSAC) 方法对满足约束条件的线组进行拟合。该算法涉及大量计算并且是不切实际的。2005 年，Uddin 等人[78]提出了双极系数法，利用斑马线与路面的灰度差提取候选区域，然后应用边缘检测、傅里叶变换、特征匹配等操作识别斑马线。由于双极系数法需要对图像进行分块，分块模板的大小会对目标的提取效果产生很大影响，存在人工分割不准确的问题。2011 年，Akinlar 等人[79]提出了边缘绘制 (Edge Drawing, ED) 算法并开发了 EDLines 方法来实现实时线段检测。2016 年，Mascetti 等人[80]建议用于智能手机的斑马线识别软件。为了消除投影失真，他们首先使用手机进行逆投影变换以提高精度，然后他们将 EDLines[79]算法应用到 GPU 中进行并行计算，可以快速识别行人过路处。2017 年，基于双极系数法，Huang 等人[81]改进的斑马线检测算法，可以集成线性判别分析 (LDA) 和 EDLines 方法。白天识别率为 90.3%。但是，在黑暗的环境中会导致错误。2019 年，基于边缘特征和霍夫变换，Chen 等人[82]开发了斑马线识别方法。利用逆透视映射原理，将图像转化为俯视图，然后利用 Sobel 边缘检测模板提取斑马线的边缘信息。此外，该算法在白天光线好的情况下准确率为

97.56%，晚上为 88.67%。但是，测试集中的斑马线基本上是一个大目标，当斑马线到相机距离较远时，没有测试结果。

3.3 方法

3.3.1 网络架构

如图 3.2 所示，对于我们的斑马线检测网络 (CDNet) 的整体架构，主要检测网络基于第 2 章所述的改进的 SE-YOLOv5 算法。针对斑马线检测应用，我们在将图像输入网络之前，首先对 ROI 进行处理。斑马线检测是在网络中进行的 (图 2 (a)、(b) 和 (c))，其中包括三个普通的目标检测器部分，分别是骨干、颈部和头部。然后，如图 2 (d) 所示，利用 SSVM 算法将结果转化为车辆过线行为检测。

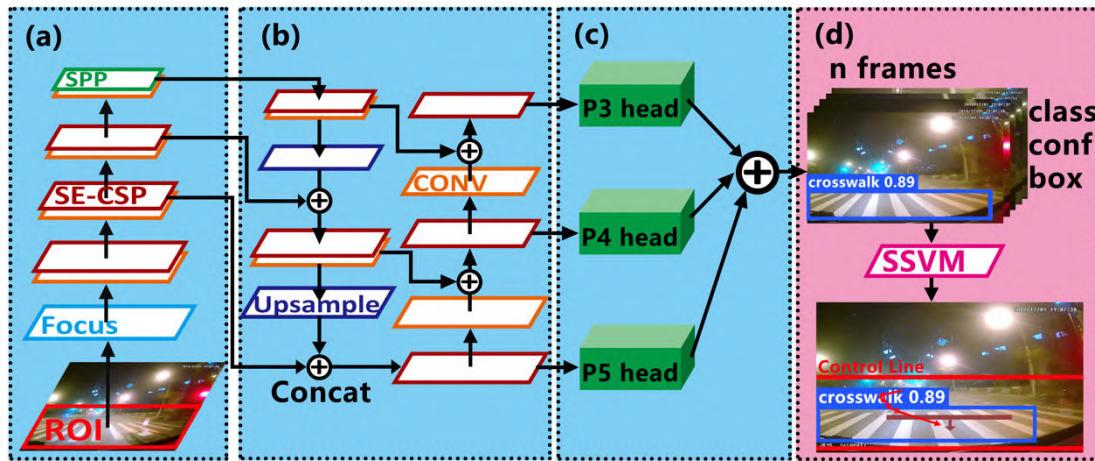


图 3.2 整体架构示意图。 (a) 脊椎, (b) 颈部, (c) 头部, 和 (d) 斑马线检测和车辆横穿斑马线行为结果

我们对 YOLOv5 官方提供的不同尺寸模型 (s、m、l 和 x) 进行测试，最终选择速度和准确率最佳平衡的 YOLOv5m 模型作为基准模型。SE 模型引入后，模型共有 311 层， $2.16E+7$ 个参数。

不同的超参数选择对模型的性能有影响，有些超参数甚至是敏感参数。因此，我们选择默认的超参数作为初始值，并使用遗传算法 (GA) 来进化超参数。最终合适的超参数如下：初始学习率为 0.009；随机梯度下降 (SGD) 算法的动量为 0.875；优化器的权重衰减为 4.2e-4；广义交叉并集 (GIoU) 损失增益为 0.0207；

类损失收益为 0.397；类二元交叉熵损失 (BCELoss) 正权为 2.0；物体损失增益为 1.27；二元交叉熵 BCELoss 正权重为 0.511；锚倍阈值为 3.61，滑动感受野大小为 5。

由于我们的模型是基于 Anchors 的，Anchors 的选择直接影响模型的性能。斑马线一般平坦而长，默认锚点不适合。K-means 聚类方法用于分析数据集并计算合适的锚点。P3、P4 和 P5 头部的合适锚点是 $[(94, 17), (166, 14), (258, 24)]$, $[(365, 31), (404, 53), (356, 77)]$ 和 $[(542, 67), (500, 99), (641, 121)]$ 。

在接下来的几节中，我们将描述我们提出的技术的动机、原理和实现方法。

3.3.2 NST 负样本训练方法

针对斑马线典型的误检，我们设计了 NST 方法减少误检率。



图 3.3 典型检测错误 (a) NST 之前的典型错误检测，(b) NST 之后的正确分类

CNN 使用卷积提取特征并根据特征检测对象，具有相似特征的物体很容易被归为同一类别。斑马线是一系列黑白垂直线。实验表明，大部分斑马线检测误差是由引导箭头的误检测引起的，尤其是三车道引导箭头线（如图 3.3 (a) 所示）。

我们提出了负样本训练 (Negative Samples Training, NST) 方法来解决引导

箭头被误判为斑马线的问题。我们手动将数据集图像中的引导箭头标记为负样本。在训练过程中，斑马线和导向箭头都会被检测到，导向箭头会被排除在最终的检测结果之外。这样，实际上增加了训练中引导箭头的损失函数。

从信息熵理论的角度分析了NST方法的优势。源的信息熵 H 可以表示为：

$$H = -\sum p_i \log P_i \quad (3.1)$$

其中 p_i 表示第 i 个符号出现在源中的概率。信息熵 H 是平均信息量的度量。

应用 NST 方法后，注释文件中可能出现的信号是斑马线、导向箭头和空锚。根据 anchor-based YOLO 原理，每张图像的符号总数为在不同的检测尺寸下，将错误检测的数量转换为概率，这样我们就可以计算出 NST 方法前后的综合信息熵，检测尺寸为 288 时为 0.06305 和 0.06806。显然，增益为 $(0.06806 - 0.06305) \div 0.06305 \approx 7.94\%$ 。

NST 方法的好处如图 3.3 (b) 所示。从第 3.4.2 节的消融实验来看，NST 方法将斑马线检测的 F1 分数提高了 7.69% (从 81.34% 提高到 89.03%)，这与信息熵理论分析的结果一致，且不影响检测速度。

3.3.3 ROI 算法

YOLOv5 集成了模型剪枝、稀疏和半精度浮点 (FP16) 技术，在 GPU 上实现实时检测，但在 Jetson nano 上，测试结果显示原生 YOLOv5m 在输入尺寸为 640 时检测速度为 5.2 FPS，不是实时。为了提升速度，我们提出基于 ROI 的算法。

从 YOLO 技术的原理来看，输入模型的图像尺寸越小，检测头的参数就会以平方倍数减小。而如果图像尺寸太小，则会直接导致图像清晰度不高，特征提取困难。因此，输入模型图像大小的选择是速度和准确性之间的权衡。对于 640 的训练规模，4.2.1 节中进行的实验表明，288 的检测规模可以达到最佳折衷。

一旦选择了尺寸，我们就引入了感兴趣区域 (ROI) 方法来进一步提高检测速度。车载摄像头中可能出现斑马线的区域是固定的 (大约是全图的一半大小)。实现如图 5 所示。原始图像的大小调整为 288×162 ，保持纵横比，并通过应用的 ROI 裁剪为 288×72 。由于 YOLOv5 采用矩形填充原理，最大网络缩放因子

为 32，我们也利用相同的操作将大小填充为 288×96 。在模型预测之后，检测到的斑马线被缩放以匹配原始图像的大小。

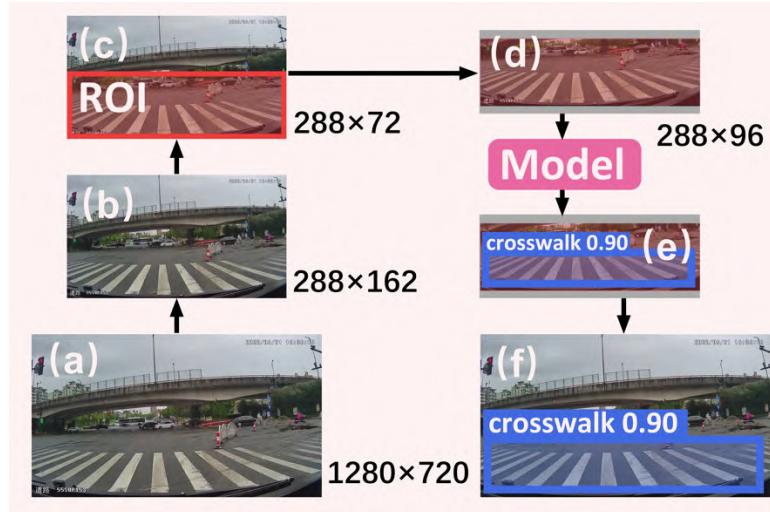


图 3.4 ROI 算法的图示。 (a) 原始图像, (b) 通过保持纵横比调整大小, (c) ROI, (d) 填充, (e) 检测到的斑马线, (f) 缩放以匹配原始图像。

第 3.4.4 节的消融实验结果表明，应用 ROI 算法后，Jetson nano 的检测速度提高了 16.7 FPS，而准确率降低了 5.99 个百分点。

3.3.4 SSVM 算法

车辆通过斑马线的时间以及在整个行程中通过了多少斑马线的车辆横穿行为信息对于 ITMS 决策具有重要意义。在这里，我们提出了滑动感受野短期向量记忆（Slide receptive field Short-term Vectors Memory, SSVM）方法来快速准确地获取这些信息。

如图 3.5 所示，我们构建了一个滑动感受野来分析一系列连续的斑马线检测帧。感受野大小决定了视频中斑马线的连续性。合适感受野大小（图 3.5 (b) 所示为 5）与视频采样的帧率有关。此外，诸如 oi 、 di 、 fo 等信息包含在感受野中的可以用来构造记忆向量。向量的形状为 $(600 \times 5, 2)$ ，其中 600×5 表示向量记忆的最大斑马线数，即 600 秒和 5 个采样频率。每个向量的初始值为 -2，当车辆通过斑马线时，该值将被更新。一般来说，每个向量的方向与像素坐标系的 Y 轴正方向一致。由于检测到的边界框位置的抖动，可能会有一个小的随机方向，尤其是在车辆停止时。如果连续 5 帧（感受野大小）没有检测到斑马线，或者

向量方向与 Y 轴方向相反，则所有短期记忆向量的值将被初始化为-2。当向量的值不是初始值时，表明车辆正准备通过或通过斑马线。最后一个向量的模数与车辆的速度正相关。此外，当从初始点到最后一点的向量第一次与红线相交时，表明车辆刚刚通过斑马线。

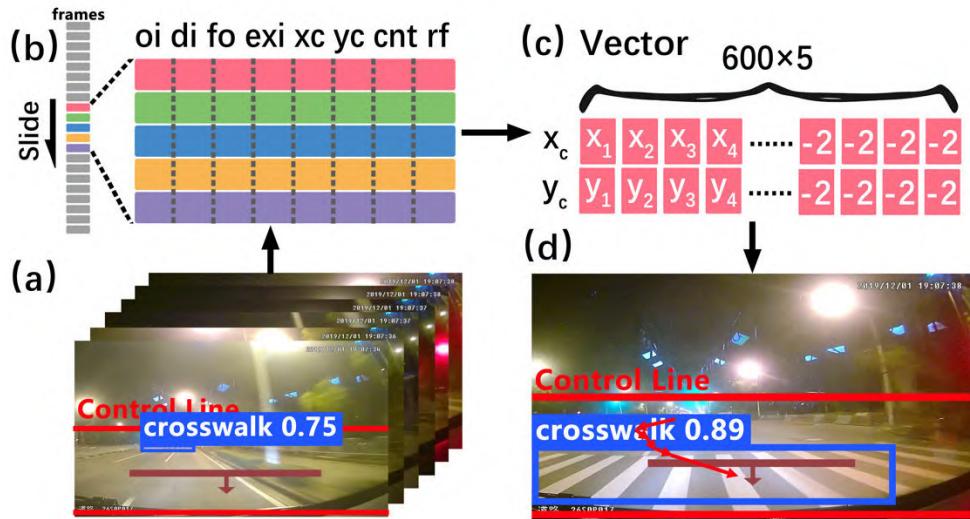


图 3.5 SSVM 算法说明: (a) 斑马线检测帧, (b) 滑动感受野, (c) 短期向量记忆, (d) 车辆穿越行为结果

注: oi: 输出索引, di: 检测到的 id, exi: 斑马线是否存在, xc 和 yc: box 的中心坐标, cnt: 穿越行为计数, rf: 记录标志。

第 3.4.5 节的消融研究结果表明, SSVM 算法成功地从斑马线检测结果中获取了车辆穿越行为的信息, 在真实场景中平均 F1 得分达到 93.70%。

3.4 消融实验

3.4.1 CDNet 数据集

基于四种天气环境 (即晴天、阴天、夜间和雨天) 下的真实车载摄像头视频和合成雾算法, 我们构建了一个包含 6868 张图像的数据集用于训练和验证, 一个包含 3540 张图像的数据集用于测试斑马线检测性能, 以及约 3 小时的驾驶视频, 用于测试车辆横穿行为检测性能。

对于训练和验证, 数据集包含 3434 张真实场景图像和 3434 张雾化图像, 训练集和验证集的比例约为 9:1, 即分别为 6160 和 708 张图像。

对于斑马线检测的测试，该数据集包含 3540 张图像，其中一半是雾增强数据，正负样本的比例约为 1:4，即有和没有斑马线的图像分别为 708 和 2832 张。训练和测试数据集没有重叠。

对于车辆穿越行为检测的测试，数据集是 7 个视频，在不同天气条件下的总时长约为 3 小时。在测试中，每条斑马线被认为是一个正样本，两个相邻正样本之间的所有帧都是负样本。

3.4.2 检测尺寸选择

检测大小的选择是速度和分数之间的权衡。一般来说，检测尺寸越大，检测分数越高，但检测速度越慢。

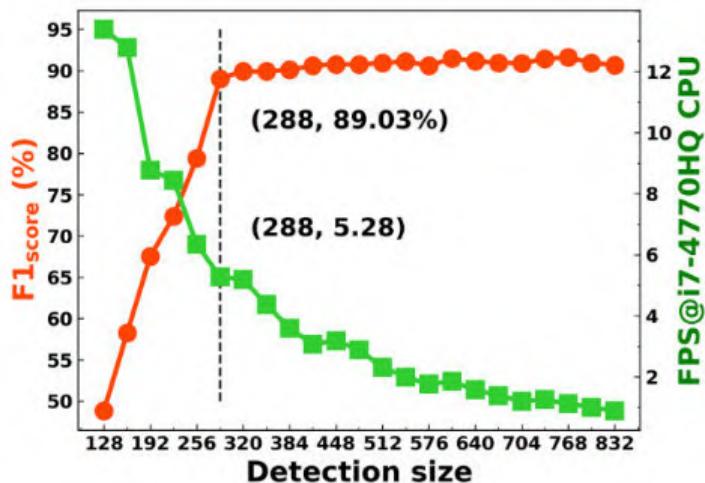


图 3.6 F1 分数和 FPS 与检测大小的趋势图

如图 3.6 中的红点线所示，F1 分数随着检测大小的增加而迅速增加。当检测大小超过 288 时，平台效应使变化更加平滑。如绿色方线所示，检测速度随着检测大小的增加而降低。因此，选择 288 作为检测大小，选择 640 作为基准。

3.4.3 NST 技术的消融实验

使用带和不带引导箭头的数据集作为负样本进行训练，检测大小分别为 288 和 640 用于检测。结果如表 3.1 所示。

表 3.1 NST 算法的消融结果

Detection size	Model	Training set classes	F1 score (%)
----------------	-------	----------------------	--------------

288	YOLOv5	Crosswalk	81.34
	YOLOv5+NST	Crosswalk and guide arrows	89.03
640	YOLOv5	Crosswalk	90.34
	YOLOv5+NST	Crosswalk and guide arrows	91.18

对于 288 的检测大小, 得分增加了 7.69%, 而 640 的得分增加了 0.84%。对于 640 的检测大小, 增量小的原因是模型在大检测大小下可以更好地区分斑马线和引导箭头。

3.4.4 ROI 算法的消融实验

在 288 和 640 的检测尺寸下, 使用和不使用 ROI 算法的结果如表 3.2 所示。

表 3.2 使用和不使用 ROI 算法的速度和得分比较。

Detection size	Model	Detection speed @Jetson nano (FPS)	F1 score (%)
288	YOLOv5+NST	20.0	89.03
	YOLOv5+NST+ROI	38.5	81.27
640	YOLOv5+NST	5.2	91.18
	YOLOv5+NST+ROI	16.7	87.17

当测试大小分别为 288 和 640 时, 应用 ROI 算法后检测速度分别提高了 18.5 和 11.5 FPS。然而, 分数分别降低了 7.76% 和 4.01%。使用 ROI 算法是有益的。

3.4.5 SSVM 算法的消融实验

基于 YOLOv5 结合 NST 和 ROI 技术, 引入 SSVM 算法检测车辆过线行为, 并在测试数据集上进行性能测试。我们共检测到约 3 小时的车载摄像头视频共 53393 帧, 其中斑马线存在约 10000 帧。以检测大小 288 为例, 斑马线正确检测 305 个, 错误检测 3 个, 漏检 38 个, F1 分数达到 93.70%。

在晴天和阴天等较好的条件下, F1 分数可以超过 98%, 说明检测准确率非常高。

在夜间等恶劣条件下, F1 得分最低为 85.28%。如图 3.1 所示, 小雨下 F1 得分较低主要是由于雨后反射和雨刮器遮挡, 而夜间 F1 得分最低主要是由于光

污染和复杂场景的结合。特别是当车辆行驶在路灯不足的乡村道路上时，斑马线的检测更加困难。

如表 3.3 所示，在引入 SSVM 算法后，288 和 640 检测尺寸的平均 F1 分数分别提高了 12.43 和 8.14%。

表 3.3 SSVM 算法的消融结果

Detection size	Model	F1 score (%)		
		Average	Worst	Best
288	YOLOv5+NST+ROI	81.27	-	-
	YOLOv5+NST+ROI+SSVM	93.70	85.28	98.80
640	YOLOv5+NST+ROI	87.17	-	-
	YOLOv5+NST+ROI+SSVM	95.31	91.79	98.82

3.4.6 合成雾增强消融实验

根据参考文献[31]，雾化技术允许模型在任何类型的场景中采用。使用第 3.6 节中描述的合成雾算法来扩充数据集，并获得分别包含 6160 和 3540 图像的训练集和测试集。我们在增强前后使用训练集训练模型，并在雾和非雾测试集上测试性能。

表 3.4 合成雾增强的消融结果

Detection size	Model	F1 score (%)	
		Test set without fog	Test set with fog
640	YOLOv5+NST	91.18	85.83
	YOLOv5+NST+FOG	91.27	91.36

如表 3.4 所示，在没有雾增强的情况下，模型的 F1 分数在遇到雾天时下降了 5.35%，而雾增强数据训练的模型在雾和非雾测试集上都有很好的表现。

3.4.7 与 SOTA 模型的性能对比

我们的模型与经典和 SOTA 模型进行了比较。之前的模型是使用 MMDetection 工具[83]在斑马线数据集（带有负样本和模糊数据）上训练和评估的。结果如表 3.5 所示。

表 3.5 与以前的目标检测模型的比较。

Model	Year	Backbone	mAP@0.5	mAP@0.5:0.95	Speed (FPS)
Faster RCNN[18]	2016	Resnet50	97.7	71.3	22.9
Mask RCNN[84]	2017	Resnet50	97.8	70.0	22.8
YOLOX[25]	2021	Darknet	98.4	73.0	49.8
Swin Transformer[26]	2021	Swin-T	97.9	73.9	18.4
YOLOv5[85]	2021	Darknet	98.2	79.0	87.2
CDNet(Ours)	2021	Darknet	99.0	80.1	61.8

注: 在 NVIDIA RTX 3080 GPU 上测得的检测速度。

结果表明, 声称在 2021 年击败 YOLO 系列的 YOLOX 模型确实比 YOLOv5 具有更高的 mAP@0.5, 但速度略慢, 这与他们论文中的描述一致。2021 年, 号称在检测、分割等主要 CV 任务中名列前茅的 Twin-Transformer, 准确率高, 但速度相对较低, 在他们的论文中与 YOLOv5 没有可比性。我们的模型在超参数、模型设置和锚点选择方面进行了微调, 对于斑马线检测任务具有较高的准确性和速度。

3.5 结果和分析

我们的 CDNet 本质上是一种基于 YOLOv5 的斑马线检测和车辆穿越行为检测算法, 提出并增加了 SE、NST、ROI、SSVM 和 FOG 等技术。基准模型和包含不同技术的模型的性能如表 3.6 所示。为了实现对边缘设备的实时检测, 使用 TensorRT SDK[86]对模型进行剪枝、压缩和转换为引擎。引擎的检测速度在 NVIDIA Jetson Nano 上进行测试。

表 3.6 使用差异模型的斑马线检测和车辆横穿行为检测结果。

Model	288		640	
	Speed (FPS)	F1 score (%)	Speed (FPS)	F1 score (%)
YOLOv5	20.0	81.34	5.2	90.34
YOLOv5+NST	20.0	89.03	5.2	91.18
SE-YOLOv5+NST	18.5	86.41	5.0	92.10
YOLOv5+NST+ROI	38.5	81.27	16.7	87.17
SE-YOLOv5+NST+ROI	33.1	79.33	16.0	86.25
SE-YOLOv5+NST+SSVM	18.5	94.86	5.0	95.47
YOLOv5+NST+ROI+SSVM	38.5	93.70	16.7	95.31
SE-YOLOv5+NST+ROI+SSVM	33.1	94.72	16.0	95.47

SE-YOLOv5+NST+ROI+SSVM+FOG (CDNet)	33.1	94.83	16.0	95.56
---------------------------------------	------	-------	------	-------

注：NST 表示训练过程中包含的负样本。SE 表示将 SE 模块插入到模型中。ROI 表示在检测过程中应用感兴趣区域方法。SSVM 表示在后期处理中应用滑动感受野短期向量记忆方法。FOG 表示在火车期间包括合成雾增强数据。检测大小为 288 和 640。

模型的性能受技术影响。对每种技术的优势及其工作原理进行如下分析：

NST 算法在不改变速度的情况下增加分数。原因是额外的负样本（引导箭头）注释增加了第 3.3 节中描述的信息熵，这增加了斑马线检测的有用信息量。

SE 模块在降低速度的同时略微提高了 F1 分数。分数增加的原因是 SE 模块使网络能够自动学习 3.1 节讨论的不同通道的权重，速度降低的原因是额外的路由增加了模型的参数。

ROI 算法在降低分数的同时大大提高了速度。原因是 ROI 算法在不截断斑马线的情况下将输入减少了一半，从而减少了操作，速度提高了一倍。同时，特征被截断，导致性能下降。

SSVM 算法在不改变速度的情况下大大提高了分数。原因是通过滑动感受野计算的向量使检测器能够获得短期记忆能力，并过滤来自上层的信号。它本质上是一种后处理算法，不需要额外的推理时间，但后处理每张图像大约需要 3 毫秒。

合成雾算法使模型能够适应雾天，同时在其他场景下具有更强的鲁棒性。原因是雾增强扩展了数据集并在一定程度上防止了过度拟合。

此外，模型性能也受参数设置的影响。对于某些参数，微小的变化会引起模型性能的巨大变化，称为灵敏度参数。敏感分析表明，模型分数对锚点、初始学习率和滑动感受野大小等参数敏感，而模型速度与这些参数无关。由于层数和检测大小等参数同时影响分数和速度，因此它们的选择是具有不同灵敏度的权衡问题。敏感分析的适当参数可在第 3.3.1 节中找到。

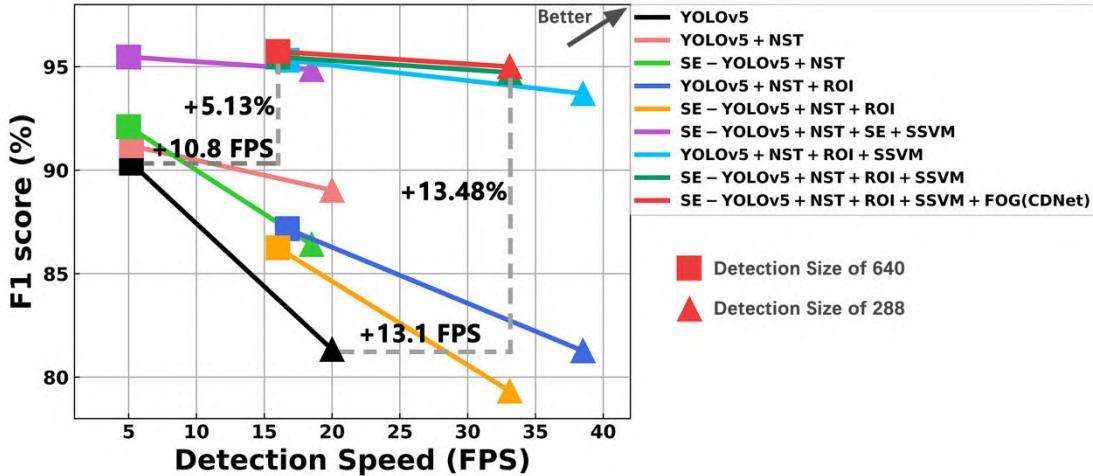


图 3.7 使用不同模型的最终检测结果说明。方框表示检测大小为 640，而三角形表示检测大小为 288。

图 3.7 直观地展示了采用不同技术（不用的颜色）的模型的性能。从基准模型 (YOLOv5) 到提出的 CDNet，对于 288 和 640 的检测大小，分数分别提高了 13.48% 和 5.13%，在 Jetson nano 上速度分别提高了 13.1 和 10.8 FPS。

最后，Jetson nano 上 94.83% 的 F1 分数和 33.1 FPS 的检测速度表明已经实现了一个实时且鲁棒的斑马线检测网络。

3.6 总结

本研究基于 YOLOv5 提出 CDNet，实现车载摄像头视觉下斑马线的快速准确检测。在 Jetson nano 上以 33.1 FPS 的检测速度，我们获得了 94.83% 的 F1 分数。对于更好的天气条件，例如晴天和阴天，F1 分数超过 98%。

本文实现了用于车载摄像头的实时且鲁棒的斑马线检测网络 (CDNet)，将为智能交通管理系统提供核心功能。同时，提出的 SE、NST、ROI、SSVM、合成雾等技术将为人工神经网络优化方法在边缘计算设备上的具体应用提供参考。

CDNet 的数据集、教程和源代码可在 GitHub[76]上获得。

第 4 章 FINet：基于合成雾增强的绝缘子检测数据集和基准模型

绝缘子及其缺陷的检测对于保障电力系统的安全稳定具有重要意义。小样本问题是制约基于神经网络的绝缘子缺陷检测算法性能的主要问题之一。在这项研究中，我们发布了一个用于绝缘子和自爆缺陷检测的数据集，并提供了一个基于改进的 YOLOv5 的基准模型，名为 Foggy Insulator Network (FINet)。在这项工作中，实现并优化了合成雾算法。构建并发布了包含 13000 张图像的绝缘子数据集 (SFID)。通过引入通道注意力机制将 YOLOv5 网络改进为 SE-YOLOv5，并从零开始训练一个具有 96.2% F1 分数的绝缘子及其缺陷的鲁棒检测模型，并作为基准。本文提出的合成雾算法可广泛用于各种数据集的数据增强。训练好的模型可应用于输电线路检测领域。源代码、数据集和教程可在 GitHub 上找到。

4.1 引言

作为输电线路的重要组成部分，绝缘子起到机械支撑和绝缘的作用。由于恶劣的天气条件和其他自然因素，绝缘子可能会爆炸，导致电力中断。严重时会引起各地区电网故障，影响安全稳定用电，造成巨大经济损失[87]。

绝缘子自爆是高压输电线路的常见缺陷。目前，检测绝缘子自爆缺陷的常用方法是利用无人机对绝缘子进行航拍[88]，然后对图像进行分析处理[89, 90]。

绝缘子图像处理方法可分为两类。一种是传统的图像处理方法，它通过空间转换、阈值分割和形态滤波来提取绝缘子[91]。然而，这些方法在处理复杂背景、相机动态大、SNR（信噪比）低等挑战时鲁棒性较差。另一种类型是基于深度学习算法。它将对象检测算法应用于检测绝缘子及其缺陷[92]。

近年来，基于深度神经网络的深度学习算法[93]已成为近年来绝缘子缺陷检测的主流。2021 年，He 等人[94]采用 ResNet101 骨干网络、复制数据增强方法和 Focal Loss 来改进 Faster-RCNN，然后将该模型应用于绝缘子缺陷检测，mAP 最终为 68.3%。Hu 等人[95]提出了一种结合 Faster-RCNN 和 UNet 的通道注意机制来识别绝缘子，mAP 达到 91.9%。Wang 等人[96]提出了一种基于高斯 YOLOv3 的绝缘子检测算法，可以融合高斯分布策略输出对应预测框的坐标均值

和方差, mAP 达到 94.5%。

深度学习算法的优越性需要足够的数据样本。现有的绝缘子数据集存在正负样本极不平衡、数据样本少、数据模糊不清等问题。目前公开的绝缘子数据集中中国电力线绝缘子数据集 (CPLID) [97] 在 2018 年有 600 张正常绝缘子图像和 248 张缺陷绝缘子图像。正常图像是无人机在不同角度拍摄的真实图像, 缺陷图像是将裁剪后的缺陷绝缘子粘贴到背景图像上形成的合成图像。对于通常需要数千个条目的神经网络来说, 848 张图像仍然太少。2020 年, Andrel 等人[98]通过 Albumentations API[71] 中的随机仿射变换、高斯模糊、无损变换等方法实现了 CPLID 的进一步数据增强, 并构建了具有 6860 张训练和测试图像的统一绝缘子公共数据集 (UPID)。

在这项工作中, 合成雾算法使用随机亮度和雾厚度来增强 UPID 数据集, 并构建了一个新的合成雾绝缘子数据集 (SFID), 其中包含 13718 个训练和测试图像。此外, 利用通道注意机制将 YOLOv5 改进为 SE-YOLOv5, 用于检测绝缘子及其缺陷。通过使用增强型 SFID、Squeeze-and-Excitation Cross-Stage Partial (SE-CSP)、Spatial Pyramid Pooling (SPP) 和 Path Aggregation Network (PANet) 等技术, 获得了更加鲁棒和准确的检测模型。

我们的贡献总结如下。

- (1) 开发优化雾化数据增强算法, 提供算法脚本。
- (2) 构建并发布包含 13718 幅绝缘子图像的绝缘子数据集 SFID。
- (3) 通过引入通道注意力机制, 将 YOLOv5 网络改进为 SE-YOLOv5。从头开始训练一个更强大的绝缘子及其缺陷检测模型, 并作为基准。

4.2 方法

本文所使用的方法如第 2 章所述。模型采用 SE-YOLOv5, 采用 2.3 节所述的合成雾算法及其优化算法进行数据集增强, 并采用了 2.4 节所述的多项离线数据集增强算法和 2.5 节所述的多项在线数据集增强算法。

4.3 消融实验

4.3.1 合成雾算法加速的消融实验

雾化算法通过 2.1.2 节描述的方法进行优化。合成效果没有区别，速度大大提高。测试结果如表一所示。对于 1 亿像素的图像，原始雾化算法加雾大约需要 4300 毫秒，优化后大约需要 100 毫秒。速度提高了约 43 倍。对于 3.84 亿像素的图像，原始雾化算法加雾大约需要 17700 毫秒，优化后大约需要 480 毫秒。速度提高了约 37 倍。

表 4.1 雾化算法的消融实验结果

Image size	Pixels(k)	Raw algorithm		Optimized algorithm	
		Time	Speed	Time	Speed
1152*864	1000	~4300 ms	1x	~100 ms	43x
1600*2400	3840	~17700 ms	1x	~470 ms	37x

分析优化原理，优化前后的时间复杂度均为 $O(w \times h)$ 。性能提升的本质是 Numpy 的矩阵计算避免了解释性语言即 python 的低性能。首先，Python 每次取一个对象的值，都会进行索引检查，判断索引是否合理。因此，在 N 个循环中，python 检查索引 N 次，而 Numpy 矩阵检查一次。其次，python 中的伪指针并不指向真正的内存地址，而是对应解释器上的数字。因此，python 中每个循环都需要一个额外的数字索引，而矩阵算法中没有额外的索引。最后，python 中间变量的生命周期只有一个循环。在 python 的每个循环中，中间变量都会被重新创建、分配和释放，而 Numpy 数组中没有中间变量。因此，采用 2.1.2 节中描述的方法创建一个新的数组，然后为每个元素计算相应的雾化算法。起雾效果与生法无异，速度大大提高。

4.3.2 神经网络改进的消融实验

如 2.2.2 节所述，引入注意力机制 SE 模块对 YOLOv5 算法进行改进后，精度略有提升，速度略有下降。使用包含 2743 张图像的测试集。图像共包含 4318 个绝缘子和 760 个自爆缺陷。训练集和测试集之间没有重叠。2.3.1 节所述的混淆矩阵被用作评估方法。由于 F1 分数是一个综合指标，同时考虑了准确率 (Precision, P) 和召回率 (Recall, R)，因此将其作为主要评价指标。

经过充分训练 (300 个 epoch) 后，性能比较如表 4.2 所示。结果表明，模型完全训练后，P、mAP@.5:.95 和 F1 score 这三个指标有显著提升，而 R 和

mAP@.5 这两个指标没有显着提升, 因为它们已经接近 100%。

引入 SE 机制后, 在原有路径的旁边增加了一条新路径, 让网络可以学习不同通道的重要程度, 并专注于更重要的通道。因此, F1 分数一般会有所提高。但是同时网络层数从 266 增加到 311, 相应的网络参数也增加了。因此, 推理时间有所增加。

表 4.2 网络改进前后性能对比

Model	Classes	Num.	P (%)	R (%)	mAP @.5 (%)	mAP @.5:.95 (%)	F1 (%)	Time (ms)
Model without SE module	Insulator	4318	93.8	99.2	99.4	89.2	96.4	
	Defect	760	90.0	99.7	99.5	84.8	94.6	192.6
	Average	5078	91.9	99.5	99.4	87.0	95.5	
Model with SE module	Insulator	4318	94.5	99.4	99.4	89.8	96.9	
	Defect	760	91.7	99.6	99.5	86.7	95.5	200.9
	Average	5078	93.1	99.5	99.5	88.3	96.2	

4.3.3 合成雾数据增强的消融实验

雾化数据增强使模型对检测绝缘子及其在雾天气中的缺陷更加稳健。本节比较了由无雾和含雾训练集训练的模型的性能。无雾训练集来自 UPID 数据集, 包含 5520 张训练图像。含雾训练集来自本文构建的 SFID 数据集, 包含 10975 张图像。测试集是含雾的, 包含 2743 个图像、4318 个绝缘子和 760 个缺陷。这些模型在有限的 100 个 epoch 中进行训练。

测试结果如表 4.3 所示。使用含雾数据集 (M2) 训练的模型比使用无雾数据集 (M1) 训练的模型实现了 P、R、mAP 和 F1 分数的全面提升。主要指标 F1 得分提高 2.7 个百分点。

进一步分析结果表明, 无雾训练集没有看到有雾的数据, 因此在含雾测试集上的测试性能一定不够好。我们还在无雾测试集上测试了 M1 的性能, F1 得分为 92.2%。虽然这个分数高于含雾测试集上 91.1% 的分数, 但仍低于 M2 在含雾测试集上 93.8% 的分数。它说明数据增强模型不仅在雾天得到了改进, 而且整体性能也得到了改进。

表 4.3 合成雾数据增强的消融实验结果

Model	Classes	Object number	P(%)	R(%)	mAP(%)	F1(%)
M1	Insulator	4318	87.7	95.5	96.2	91.4
	Defect	760	87.0	95.0	97.1	90.8
	Average	5078	87.3	95.2	96.6	91.1
M2	Insulator	760	90.9	99.1	99.0	94.8
	Defect	5078	86.7	99.7	99.5	92.8
	Average	5078	88.8	99.4	99.3	93.8

4.3.4 合成雾模型在晴天下的实验

上一节的分析表明，合成雾数据增强不仅使模型在雾天更加鲁棒，而且提高了整体性能。在本节中，我们在 T1 和 T2 测试集上测试 M2 的性能。T1 是来自 UPID 数据集的无雾测试集，其中包含 1339 个图像、2076 个绝缘子和 371 个自爆缺陷。T2 是来自 SFID 数据集的含雾测试集，其中包含 2743 个图像、4318 个绝缘子和 760 个自爆缺陷。这些模型在有限的 100 个 epoch 中进行训练。

测试结果如表 4.4 所示。M2 在 T1 和 T2 上的 R 和 mAP 相同。M2 在 T1 上的 P 和 F1 得分分别达到 89.8% 和 94.3%，分别比 T2 上高 1 和 0.5 个百分点。

与上一节 M1 在 T1 上的 F1 分数为 92.2% 相比，M2 在 T1 上的 F1 分数为 94.3%，提升了 2.1 个百分点。表明合成雾数据增强不仅提高了模型在有雾天气下的鲁棒性，而且提高了模型对无雾图像的鲁棒性。

我们分析了该模型对无雾图像也更稳健的原因。数据、算法和计算能力是深度学习算法的三大载体。对于某个模型，数据对最终的性能有着决定性的影响。一般来说，大模型和小数据的结合会导致过拟合的问题，因为大模型有足够的参数，它可以“记住”每一块训练数据，这说明在训练集上的表现是优秀的，但缺乏在测试集上的鲁棒性。我们使用的深度神经网络有 311 层，是一个非常深的网络。虽然在训练中采用了数据增强、残差网络结构、正则化等方法来防止过拟合，但也需要较大的数据集才能达到更好的效果。一般认为，数万甚至数十万的数据更适合这样的网络。因此，使用合成雾数据增强算法使用数据集训练的模型对雾天和晴天更加鲁棒。

表 4.4 M2 在 T1 和 T2 测试集上的性能

Model	Test set	Classes	Object number	P(%)	R(%)	mAP(%)	F1(%)
T1	Insulator	Insulator	2076	91.1	99.0	99.0	94.9
		Defect	371	88.5	99.7	99.5	93.8
	Average	Average	2447	89.8	99.4	99.3	94.3
M2	Insulator	Insulator	4318	90.9	99.1	99.0	94.8
		Defect	760	86.7	99.7	99.5	92.8
T2	Average	Average	5078	88.8	99.4	99.3	93.8

注: M2 是我们的 SFID 数据集训练的模型。 T1 是来自 UPID 数据集的无雾测试集, T2 是来自 SFID 数据集的含雾测试集。

4.3.5 与 Classic 和 SOTA 模型的比较

本节研究了我们的模型与几个经典和最先进的模型之间的检测性能比较, 这些模型是 Faster RCNN、Mask RCNN、YOLOX、Swin-Transformer 和原生 YOLOv5。前四个模型由 MMDetection 工具运行。所有模型都在建议的基础上进行训练和测试

表 4.5 检测性能比较

Model	Year	Backbone	mAP@.5(%)	mAP@.5:.95(%)
Faster RCNN	2016	Resnet50	98.4	80.1
Mask RCNN	2017	Resnet50	98.3	82.0
YOLOX	2021	Darknet	99.4	86.0
Swin-Transformer	2021	Swin-T	99.0	86.4
YOLOv5	2021	Darknet	99.3	87.0
FINet (ours)	2022	Darknet	99.5	88.3

结果表明, 经典的 Faster RCNN 和 Mask RCNN 的性能相对较差。声称在 2021 年击败 YOLO 系列的 YOLOX 模型确实比 YOLOv5 具有更高的 mAP@.5, 这与他们论文中的描述是一致的。 Swin-Transformer 号称达到了主要 CV 任务的顶峰, 在我们的论文中还不如 YOLOv5。 YOLOv5 和我们改进的 FINet 之所以具有高性能, 是因为它们进行了详细的超参数优化、模型设置和锚点选择。

4.4 结果和分析

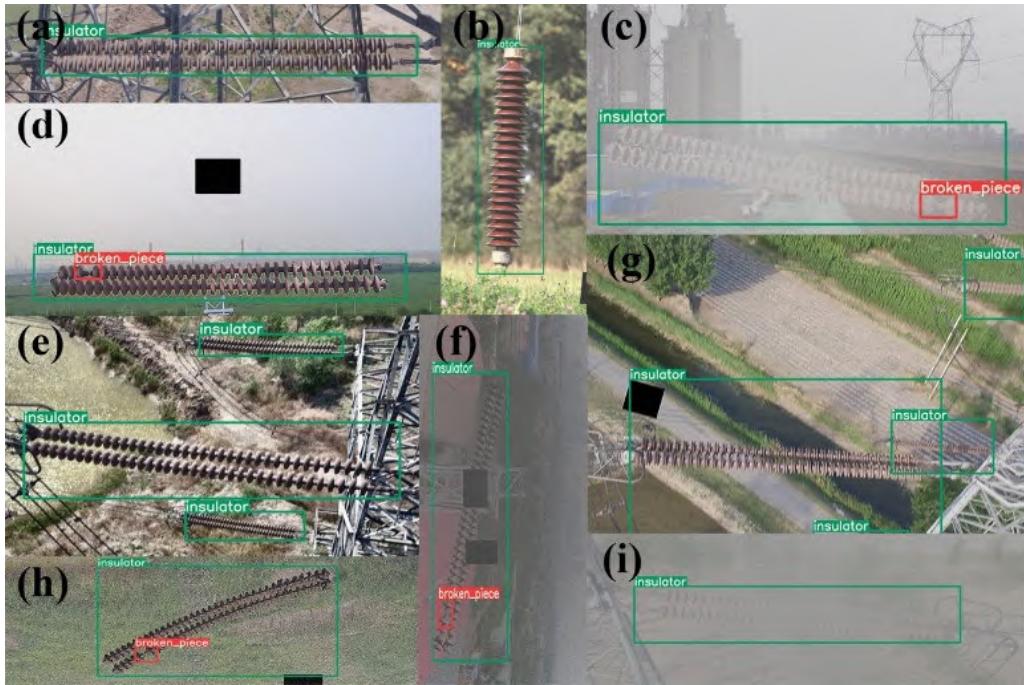


图 4.1 复杂环境下绝缘子及其缺陷检测结果 (a) 简单背景, (b) 带轻雾的简单背景, (c) 浓雾中的缺陷, (d) 天空背景, (e) 不同尺度, (f) 垂直绝缘子和中雾, (g) 截断绝缘子, (h) 图像模糊, (i) 浓雾中的绝缘子

复杂环境下的绝缘子及其缺陷图像和还原结果如图 4.1 所示。由于无人机的拍摄角度, 绝缘子图像的背景主要包括天空、城市、地草、树木和湖泊。图像中绝缘子的尺度不同, 有的像素面积大, 有的像素面积小。一些绝缘子被图像的边缘切断。因为在一些图像中使用马赛克增强, 一些绝缘子被马赛克阻挡。在某些图像中, 绝缘子及其缺陷会被像素模糊。此外, 上述条件会根据不同的雾浓度进行排列组合。绝缘子在浓雾中甚至难以看清, 因为图像中绝缘子的动态范围变化很大。图中绿色框为检测到的绝缘子, 红色框为检测到的缺陷。

根据第 3.4 节的实验, 我们的模型在这些复杂环境中具有良好的检测性能。经过充分训练, 绝缘子和自爆绝缘子的检测 F1 分数分别达到 96.9% 和 95.5%, 综合分数达到 96.2%, 鲁棒性强。在 RTX-3080 GPU 上的进一步测试结果表明, 单张图像检测仅需 16 ms, 达到 62.5 FPS。优化部署后, 可实现对 Jetson nano 等边缘设备的实时检测。

4.5 总结

本文发布了基于提出的合成雾算法的绝缘子和自爆缺陷数据集和基于改进的 YOLOv5 的检测基准。结果如下：

(1) 基于暗通道先验实现合成雾算法，并通过矩阵计算进行优化，使雾化速度提高约 40 倍。

(2) 通过对 UPID 进行扩充，构建了 13000 张图像的绝缘子数据集(SFID)，并在 Github 上发布。

(3) 在 YOLOv5 中引入通道注意力机制，构建绝缘子和自爆缺陷检测网络。该模型经过 SFID 训练，综合 F1 分数达到 96.2%。该模型在晴天和大雾场景下均具有较强的鲁棒性，可应用于输电线路巡检领域。

FINet 的数据集、教程和源代码可在 GitHub[99]上获得。

第 5 章 AIDNet: 基于自适应算法的卫星遥感图像检测绝缘子的探索研究

绝缘子直径约为 200 毫米，在最佳卫星遥感图像（Satellite Remote Sensing Image, SRSI）中仅占 0.4 个像素。由于极低的空间分辨率和混合像元问题，从 SRSI 中检测绝缘子和自爆缺陷是一项具有挑战性的任务。提高所需空间分辨率 (rSR) 的算法设计是推进这项任务的一种方法。本文提出了自适应网络和检测网络相结合的两阶段检测方法，采用多重滤波和自适应增强方法实现端到端检测，可以有效提高 rSR。本章提出的思想可以促进实现从 SRSI 检测绝缘子和缺陷的目标。所提出的算法可以实现空间分辨率相对较差的检测，可应用于一般小目标检测任务中。

5.1 引言

输电线路中的绝缘子是重要的组成部分，对电网的安全起着至关重要的作用，是现代经济社会的基础设施。

目前，大多数关于传输线检测的研究都集中在使用流行的目标检测网络从航空图像中检测绝缘子和缺陷[89, 100-105]。检查任务已从人类转向无人机（Unmanned Aerial Vehicle, UAV）。从卫星遥感图像（Satellite Remote Sensing Image, SRSI）中检测绝缘子的研究很少[106]。随着人工智能和遥感技术的发展，有望实现。

从 SRSI 中检测绝缘子的两个主要问题是空间分辨率 (Spatial Resolution, SR) 差和像素混合。普通相机可以轻松达到毫米级 SR，而目前性能最高的卫星遥感相机只能达到 dm 级，即 0.5 m，相差一百倍。在 SRSI 中，传感器以像素为单位记录地物反射或发射的光谱信息。在多光谱和低分辨率的影响下，一个单元内通常存在不同类型的物体，这就是混合像元问题。

克服这两个问题，实现任务，就像修公路时挖隧道的艰巨任务，需要从“山”的两侧开始。在卫星端，需要采用新技术不断提高图像质量。在检测方面，需要提出新的算法来适应空间分辨率非常低的图像。这项工作是对后一种路线的探索。

本文开发了一种从无人机图像中模拟 SRSI 的合适方法，构建了模拟卫星

遥感数据集，并对当前 SOTA 模型的性能进行了评估。然后设计了两阶段的网络架构，提出了一种可微分网络作为自适应网络，并将 YOLOv5[24]改进为 SE-YOLOv5 作为检测网络，最终提高了有效检测所需的空间分辨率(Required Spatial Resolution, rSR)。

我们的贡献总结如下：

- (1) 提出、实现并测试了从无人机图像中模拟 SRSI 的混合像元方法。
- (2) 在 YOLOv5 中引入 Channel Attention 机制，提高了检测性能。
- (3) 提出了一种端到端的自适应绝缘子检测网络(AIDNet)，有效地提高了 rSR。

5.2 方法

5.2.1 卫星遥感图像模拟算法

我们收集了 10 张带绝缘子的真实 SRSI 图像。初步实验表明，从这些图像中检测绝缘子非常困难，空间分辨率极差，样本很少是主要原因。

为了评估我们距离在真实 SRSI 上检测绝缘子的目标还有多远，需要获取具有不同空间分辨率的 SRSI。提出并测试了插值双三次、插值双线性和混合像元等模拟方法。最后确定效果最好的混合像元法。

混合像元法如下：

$$J_{x,y} = \frac{1}{k^2} \sum_{i=x-\frac{k}{2}}^{x+\frac{k}{2}} \sum_{j=y-\frac{k}{2}}^{y+\frac{k}{2}} I_{x,y}(i,j) \quad (5.1)$$

其中 J 表示输出图像， I 表示原始图像， k 表示内核大小并且等于下采样率。随着内核 k 的增加，一个像素包含更多其他像素的信息，即混合像元。

5.2.2 自适应和目标检测网络

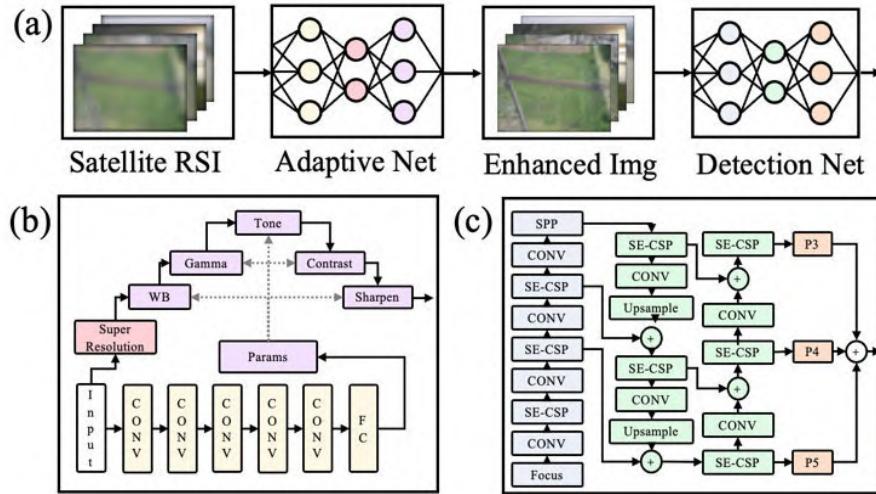


图 5.1 网络架构 (a) 技术路线; (b) 自适应网络; (c) 检测网络

本文的网络架构如图 5.1 所示。

技术路线如图 5.1 (a) 所示, 卫星图像首先进入自适应网络进行图像增强, 然后进入检测网络进行绝缘子和缺陷检测。前者用模拟数据集训练, 后者用普通数据集训练。

5.2.2.1 自适应网络

自适应网络如图 5.1 (b) 所示, 由两部分组成, 一是基于 RDN (Residual Dense Network) [10]的超分辨率算法, 二是基于五种方法的自适应增强算法, 分别是 WB (白平衡)、伽玛、色调、对比度和锐化。

在超分辨率部分 (粉色框), 网络包含四个模块, 共 144 层。第一个特征提取模块通过卷积从输入的低分辨率图像中提取特征。第二个残差模块由经典的 residual block 和 dense block 融合而成, 可以同时利用前几层的特征, 提高模型的表达性能, 避免梯度随着网络的加深而消失。第三个密集特征融合模块使用 Concat、卷积、加权和批量归一化技术来融合全局特征和残差。第四个上采样模块最终通过上采样使图像分辨率加倍, 以提高图像质量。

在自适应增强部分, 有两个模块。第一个自适应模块 (黄色框) 是一个可微的 16 层网络, 具有卷积、批量归一化、Leaky ReLU[107]激活和一个全连接层, 可以提取特征和预测增强参数。第二个增强模块 (紫色框) 根据预测参数执行图像增强过程, 部分解决分辨率差和混合像元问题。过滤器如下:

(a) 白平衡 WB 滤波

WB 滤波器的映射函数如下,

$$J = \left(W_r r_i, W_g g_i, W_b b_i \right) \quad (5.2)$$

其中 W_r, W_g 和 W_b 是要优化的参数, r_i, g_i 和 b_i 是红、绿、蓝三种颜色的值。

(b) 伽玛 Gamma 滤波

Gamma 滤波器的映射函数如下,

$$J = I^G \quad (5.3)$$

其中 Gamma 值 G 将被优化。WB 和 Gamma 是简单的乘法和幂变换, 它们显然是可微的。

(c) Tone 滤波

Tone filter 的映射函数如下,

$$J = \left(L_r(r_i), L_g(g_i), L_b(b_i) \right) \quad (5.4)$$

其中 L 是要优化的单调分段线性函数[108], 表示为 。

(d) 对比度 Contrast

对比度过滤器的映射函数如下,

$$J = \alpha I \times \frac{\frac{1}{2} \left\{ 1 - \cos \left[\left(0.27r_i + 0.67g_i + 0.06b_i \right) \pi \right] \right\}}{0.27r_i + 0.67g_i + 0.06b_i} + (1 - \alpha) I \quad (5.5)$$

其中要优化对比度值 α 。

(e) 锐化 Sharpen

Sharpen filter 的映射函数如下[109],

$$J = I + \lambda [I + Gau(I)] \quad (5.6)$$

其中 $Gau(I)$ 表示高斯滤波器, λ 是要优化的锐化度。

5.2.2.2 检测网络

检测网络如图 5.1 (c) 所示, 我们测试了经典和当前 SOTA 的多个目标检测模型的性能, 选择 YOLOv5 作为基准, 改进后的 SE-YOLOv5 最终用作检测的 AIDNet 网络。

在主干部分 (蓝色框) 中, Focus 模块应用了空间注意机制, 它将输入图像的每个像素的信息与其周围的信息连接起来。CSP (Cross Stage Partial) [66] 模

块使用残差嵌套来增强信息流。SPPNet (Spatial Pyramid-Pooling) [67]模块是另一种空间注意力机制，它通过最大池化算子聚合特征。

在颈部 (绿色框) 中，使用 PANet (路径聚合网络) [68] 架构通过自下而上的路径增强在较低层使用准确的定位信号增强整个特征层次结构来促进信息流，从而缩短低层和顶层特征之间的信息路径。

在头部 (橙色框) 中，卷积用于快速实现特征到结果的映射。

我们所做的改进是将通道注意力机制 SE (Squeeze-and-Excitation) [69] 引入 CSP 模块。改进后网络如第 2.1.1.2 节所述。

5.2.3 参数设置

自适应网络和检测网络分别有 16 层和 311 层。初始学习率为 0.01。随机梯度下降 (SGD) 的动量为 0.937。该模型训练了 300 个 epoch。

5.2.4 评价指标

作为经典的目标检测任务，我们采用了 2.3 节所述的基于混淆矩阵[74]的度量，例如：精度 (P)、召回率 (R)、平均精度 (mAP) 和 F1 分数[75]。

针对本任务，为了评估模型与检测 SRSI 绝缘子的目标之间的距离，使用研究 [8] 中的 rSR 作为主要指标。rSR@.8F1 表示目标检测的 F1 分数至少达到 80% 时所需的空间分辨率。

5.3 消融实验

5.3.1 空间分辨率估计

空间分辨率 (SR) 值越小，图像越好。通过对 UPID 数据集进行采样[98]来估计无人机图像的 SR。结果表明，单个绝缘子的像素范围为 25~40 像素，平均值为 33 像素。由于每个绝缘片的直径约为 0.2 m，因此无人机图像的 SR 粗略计算为 6.0 mm 左右，这与研究[106]中 6.8 mm 的结果一致，远优于 SRSI 的 500 mm。

5.3.2 模拟方法和效果

5.3.2.1 模拟方法

模拟方法选择的目的是开发一种合适的方法, 使模拟图像接近真实的 SRSI 图像。

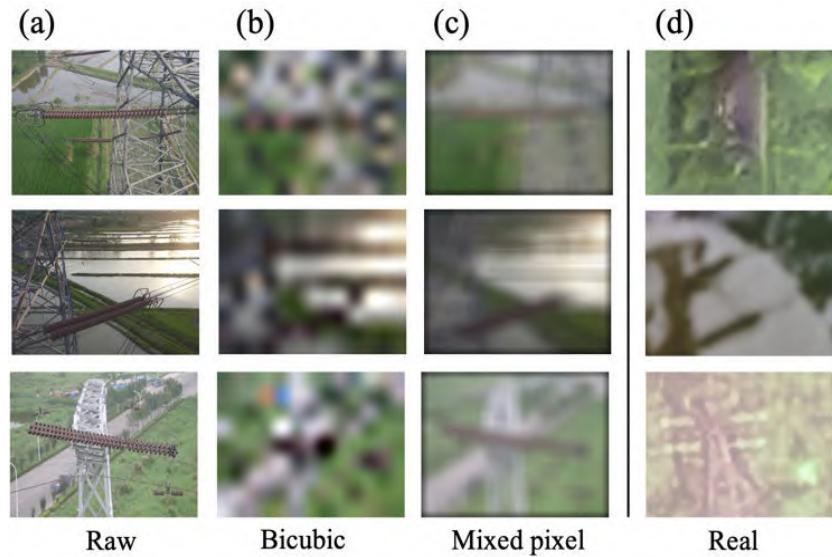


图 5.2 不同方法的模拟效果比较。 (a) 原始无人机图像; (b) 用插值双三次方法模拟图像; (c) 混合像元法模拟图像; (d) 真实卫星遥感图像 (SRSI)

插值双三次、插值双线性和提出的混合像元方法的模拟效果比较如图 5.2 所示。第一列包含 SR 为 6 mm 的原始无人机图像。最后一列包含 SR 为 500 mm 的真实 SRSI 图像。如图 5.2 (b) 所示, 在 88x 下采样率下, 即 528 mm SR 下, 插值双三次方法会使绝缘子形成“气泡”, 这与真实的 SRSI 不一致。Interpolate Bilinear 方法与 Bicubic 的效果相同, 因此不在图上绘制。如图 5.2(c) 所示, 本文提出的混合像元法是本文最终的模拟方法, 因为其效果接近真实 SRSI。

5.3.2.2 不同下采样率下的仿真效果

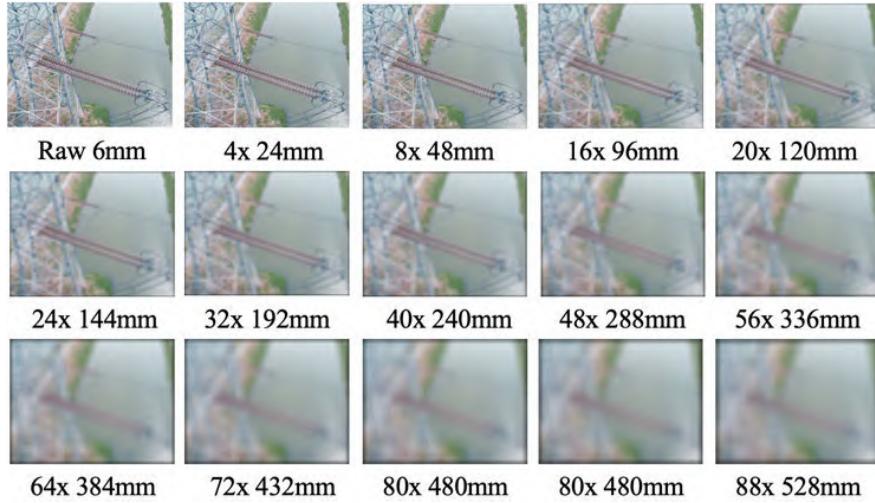


图 5.3 不同下采样率下的模拟图像, 降采样倍数从 1 到 88x

不同下采样率下的模拟图像如图 5.3 所示。下采样率范围为 1 (原始) 到 88x, 对应的 SR 范围为 6 到 528 mm。从图中可以看出, 随着下采样率的提高, 图像变得更加模糊, 每个像素混合了更多来自其他像素的信息。

5.3.3 与 SOTA 模型的性能对比

研究了几个经典和 SOTA 检测模型的绝缘子和缺陷检测性能, 这些模型是 Faster RCNN [18]、Mask RCNN [19]、YOLOX [25]、Swin-Transformer [26] 和 YOLOv5 [24]。这些模型在 UPID 数据集上进行训练和测试, 前四个模型由 MMDetection 工具[83]运行。结果如表 5.1 所示。

表 5.1 检测性能对比

Model	Year	Backbone	mAP@.5(%)	mAP@.5:.95(%)	F1(%)
Faster RCNN	2016	Resnet50	97.2	70.1	90.9
Mask RCNN	2017	Resnet50	97.4	69.5	91.1
YOLOX	2021	Darknet	98.4	73.1	94.3
Swin Transformer	2021	Swin-T	97.9	73.6	94.5
YOLOv5	2021	Darknet	98.2	76.5	94.7
SE-YOLOv5 (Ours)	2022	Darknet	98.3	75.8	95.6

结果表明, 在之前的模型中, YOLOX 的 mAP@.5 性能最好, YOLOv5 的 mAP@.5:.95 和 F1 性能最好, Swin-Transformer 综合性能好, 而 Faster RCNN 和

Mask RCNN 表现相对较差。YOLOv5 被选为基准。

5.3.4 所需的空间分辨率

在模拟测试集下评估所需的空间分辨率 (rSR)，该测试集使用上述混合像元方法将图像下采样 1 到 88 倍，模拟空间分辨率为 6 到 528 mm 的 SRSI 图像。

rSR 评估在原生 YOLOv5、改进的 SE-YOLOv5 和两阶段 AIDNet (自适应网络+SE-YOLOv5) 中进行。直观的结果如图 5.4 所示，详细信息如表 5.2 所示。

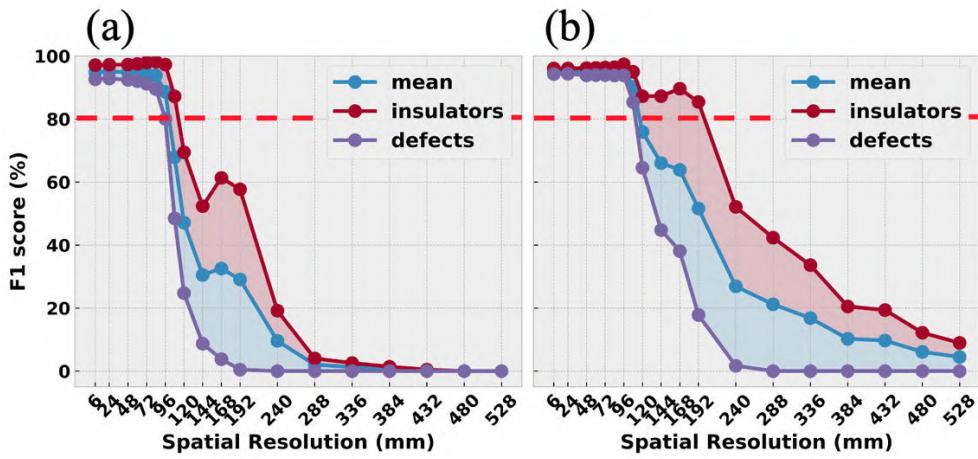


图 5.4 模型所需空间分辨率评估结果： (a) Native YOLOv5; (b) AIDNet

从图 5.4 可以看出，随着空间分辨率 (SR) 变差 (值越大)，F1 得分先小幅上升，然后迅速下降。对于原生 YOLOv5 (图 5.4 (a))，F1 分数在 528 mm SR 处几乎为零。在 16x 下采样率下，平均 F1 得分下降到 84.8%，表明基准模型的 $rSR@.8F1$ 为 96 mm (均值)，优于文献 [8] 中的 68 mm。对于所提出的 AIDNet (图 5.4 (b))，在 528 mm 分辨率下，缺陷的 F1 分数仍然几乎为零，但绝缘子的 F1 分数不再为零。此外，整体下降趋势被延迟，这意味着我们提出的算法有效。

5.4 结果和分析

最终的详细结果如表 5.2 所示。对于原生 YOLOv5，平均值、绝缘子和缺陷的 rSR 分别为 96、108 和 84 mm。引入 SE 机制改进 YOLOv5 后，rSR 仍然为 96、108 和 84 mm，但相应的 F1 分数分别提高了 2.3、4.4 和 2.2 个百分点。

分点（粗体值）。这意味着网络模型本身的改进，虽然提高了检测性能，但对我们改进 rSR 的目标的用处有限。

采用我们提出的两阶段模型 AIDNet(即自适应网络 + SE-YOLOv5)，均值、绝缘子和缺陷的 rSR 分别达到 108、192 和 108 mm，有显着改进。具体来说，绝缘子检测的 rSR 有最大的改进。

分析原因，自适应网络中的超分辨率算法提高了图像质量，WB、Gamma、Contrast 和 Tone 等逐像素滤波器可以看作是像素注意机制，Sharpen 滤波器可以突出图像细节。

需要注意的是，由于 rSR 还没有达到 500 mm，AIDNet 仍然无法实现直接从 SRSI 中检测绝缘子的目标。未来需要继续深入研究算法和硬件。

表 5.2 所需的空间分辨率评估结果

Down Samples Rate	Mean F1 (%)				Insulators F1 (%)				Defects F1 (%)			
	SR (mm)	YO LOv 5	SE-YOLOv5	AIDNe t(Ours)	YO LOv 5	SE-YOLOv5	AIDNe t(Ours)	YO LOv 5	SE-YOLOv5	AIDNe t(Ours)	YO LOv 5	AIDNe t(Ours)
1x	6	93.5	95.1	95.1	96.2	97.4	96.0	90.7	92.8	94.3		
4x	24	93.8	94.9	95.2	96.8	97.0	96.0	90.7	92.8	94.3		
8x	48	93.8	94.9	95.0	97.2	97.8	96.2	90.0	91.9	93.9		
12x	72	93.0	94.7	95.2	96.8	98.5	96.4	89.1	90.8	94.0		
14x	84	91.7	93.7	95.1	96.7	98.5	96.5	86.7	88.9	93.8		
16x	96	84.8	87.1	95.6	93.8	97.4	97.4	75.9	76.9	93.9		
18x	108	62.7	66.4	90.2	80.7	85.1	95.0	44.6	47.6	85.4		
20x	120	42.8	44.7	75.9	68.0	65.5	87.3	17.6	24.0	64.5		
24x	144	33.9	26.7	66.0	59.6	48.1	87.3	8.2	5.4	44.8		
28x	168	37.2	28.7	63.8	68.0	56.0	89.6	6.3	1.4	38.1		
32x	192	31.8	26.6	51.6	61.5	53.1	85.4	2.1	0.2	17.8		
40x	240	14.3	7.4	26.9	28.6	14.6	52.1	0	0.2	1.7		
rSR (mm)	96	96		108	108	108	192	84	84	108		

5.5 总结与展望

在本文中，我们探索了从卫星遥感图像 (SRSI) 中检测绝缘子及其缺陷的任务的可能性，由于空间分辨率极差和混合像元问题，目前这几乎是不可能的。但我们采用自适应网络+检测网络的 AIDNet，推进了这项任务的实现进程。

结果如下：

(1) 实验表明, 我们提出的混合像元方法是从无人机图像中模拟 SRSI 的合适方法。

(2) 结果表明, 当前 SOTA 检测模型 YOLOv5 所需的 SR (rSR@.8F1) 对绝缘子和缺陷的要求分别为 96 和 84 mm, 而我们提出的 AIDNet 则达到了 192 和 108 mm。

(3) 由于自适应算法带来的益处, 自适应网络加检测网络的两阶段架构可能是未来实现这一目标的途径。

展望如下:

(1) 本文通过超分辨率和常用的自适应滤波器实现对 rSR 的改进。针对混合像元等问题设计新的自适应算法 (如基于非线性变换、高维希尔伯特空间等的算法) 应该会进一步提高 rSR。

(2) 本文利用无人机可见光图像模拟 SRSI, 训练自适应网络。获取训练集的更好方法应该来自无人机的遥感图像, 这与 SRSI 同源。

(3) 此外, 在卫星端, 利用高光谱等多种方法来提高 SRSI 的质量, 是从硬件层面实现这一目标的另一个方面。

第 6 章 TLDNet: 基于数据增强增广和神经网络的输电导线及其缺陷检测方法

本文基于来自云南电网呼马山试验基地及前人的少量、低分辨率输电导线样本, 使用超分辨率算法、提出基于动态背景和随机变换的离线数据增广算法, 生成了清晰的 7000 张输电导线数据集, 解决了数据多样性中的随机位置、随机角度、随机尺寸和随机背景等问题。同时, 本文改进了 YOLOv5 目标检测算法, 实现了输电导线、散股和断股的实时、鲁棒的检测, 平均 F1 分数达 94.7%。本文提出的超分辨率和数据增广方法可以广泛应用于各领域的低分辨率图像、小样本数据集中, 本文实现的输电导线及其缺陷检测算法能应用于输电线路巡检领域, 使巡检更加高效和智能。

6.1 引言

输电导线是电力系统中的重要组成部分, 由于地理分布位置特殊, 除了承受电力负荷和机械载荷的内部压力外, 还长期经受风吹日晒、雨雪风沙、电闪雷击等恶劣环境的侵害, 极易出现如断股、散股等缺陷, 如果不及时处理常常会引发恶性电力事故。

输电导线故障巡检分为人工巡检和无人机巡检。人工巡检检测效率低、工作强度大、检测速度慢; 无人机巡检可以有效减少人力、物力的投入, 而且能够排除人为因素的干扰, 评估导线状况, 但其性能依赖于算法。输电导线故障检测算法是关键技术。

在这一领域, 前人进行了一些研究。例如, 2006 年, 吴功平等人[110]使用小波分析方法, 用红外和电磁传感器对铝绞线和钢芯的断股故障进行了检测。2015 年, 王万国等人[111]使用水平和垂直方向的梯度算子, 对线结构进行感知, 实现了基于无人机图像的输电导线断股及异物缺陷检测。2017 年, 齐国顺等人[112]使用图像处理技术和图像识别技术, 采用设定好的交叉点结构匹配模板, 对导线断股进行了检测。2020 年, 黄新波等人[113]提出一种基于加权色差法的分割方法实现对不同背景条件下输电导线的分割提取, 同时, 结合纹理特征和神经网络实现 2 类导线缺陷的检测与定位。2021 年, 罗永超等人提出了一种基于

MATLAB 图像识别的输电线路断股散股检测方法。2021 年, 孙日高等人[114]为了降低输电导线故障检测中的冗余计算量, 使用了水平校正、图像区域分割等技术实现了断股故障和异物悬挂故障的区域定位。2021 年, 李飞伟等人[115]使用全卷积神经网络 (Fully Convolutional Network, FCN) 实现了图像分割, 实现了输电导线目标分割, 而后采用滑动窗口的思想对分割结果进行特征提取, 实现断股检测。此外, 他们还实现了基于 UDP (User Datagram Protocol) 协议的 WIFI 图像传输, 形成了输电导线断股缺陷检测系统。在调研的算法中, 每种算法均针对特定的一张或几张图片进行检测, 在输入分布完全不同时, 难以有效检测。

深度学习算法[93]已成为当前最为主流的目标检测、缺陷检测算法。使用深度学习对输电导线及其缺陷检测有巨大的优势, 例如: 不论输电导线在图像的什么位置, 导线走向, 尺度大小, 正光逆光, 都可以实现自动导线提取、自动特征提取和缺陷所在位置及类型检测, 在不同复杂背景下依然奏效。

当前的神经网络算法是数据驱动的, 庞大的、标注良好的数据集是影响模型性能的直接因素。目前我们没有看到关于输电导线及其缺陷检测的公开的良好的数据集, 而收集、整理、特别的精细标注大量数据是一个繁琐、枯燥的工作。本文希望通过设计算法, 实现低分辨、小样本数据集上鲁棒的检测。

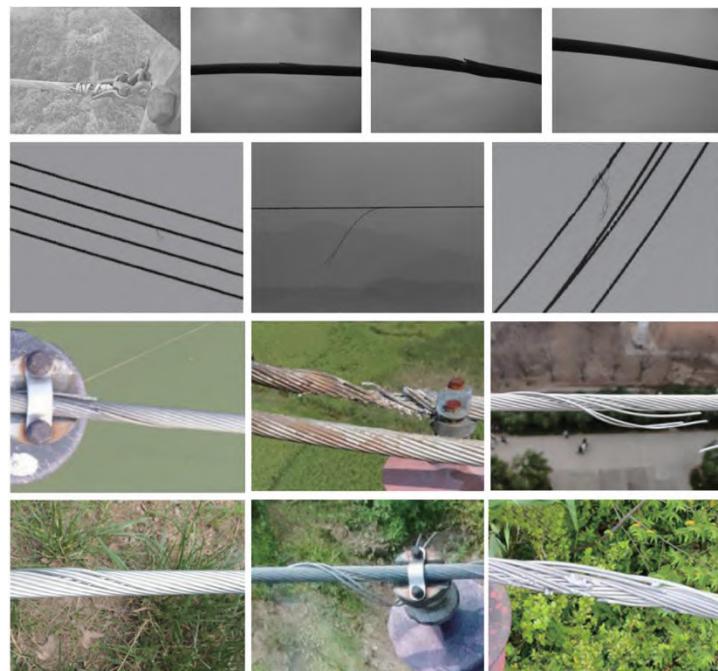


图 6.1 原始输电导线及其缺陷图像

我们首先在云南电网呼马山试验基地采集了输电导线图像，并提取了文献内的图像，形成了如图 6.1 所示的原始输电导线样本。分析发现，图像存在背景动态范围大，背景复杂，环境光复杂，图像分辨率低，输电导线不清晰等问题。

在本文中，为了解决图像动态范围大、背景复杂、环境光复杂的问题，我们提出使用基于 YOLO 的改进目标检测算法 SE-YOLOv5 来提高鲁棒性；为了解决分辨率低的问题，我们提出使用基于 RDN（Residual Dense Network）[116]的超分辨率算法来提升图像质量；为了解决小样本问题，我们提出基于动态背景的随机数目、随机位置、随机缩放和随机角度方法来生成数据集。最终，我们获得了复杂背景、复杂环境光、任意导线尺度、走向的输电导线、散股和断股的鲁棒检测方法，检测综合分数达到 94.7%。

本文贡献如下：

- (1) 实现了基于 RDN 的超分辨率算法，应用到输电导线数据中，形成了高质量图像。
- (2) 实现了基于动态背景的数据集生成方法，从 13 张原始图像中，生成了 7000 张共计约 15000 样本的数据集。
- (3) 实现了基于改进 SE-YOLO 的输电导线、散股和断股的实时、鲁棒检测算法，检测 F1 分数达到 94.7%。

6.2 方法

6.2.1 图像质量提升算法

本文采用的如 2.1.5 节所述的图像质量提升算法，包含超分辨率算法和高反差保留算法。

高分辨率算法的参数设置如下：原始图像尺寸不一，最小的分辨率为(730, 534)共 39 万像素，最大的为(1518, 810)共 122 万像素，平均为(1065, 744)共 79 万像素。我们采用 Zhang[116]等人训练好的 PSNR-larger 模型来进行超分辨，模型权重大小为 66.1Mb。对每张图像视情况进行 1 到 3 次增强，每增强一次，分辨率翻倍，增强后平均分辨率为(4597, 3132)共 1440 万像素。

高反差保留中高斯模糊半径为 5 像素。

6.2.2 动态背景数据增广算法

目前无良好的公开的输电导线及其缺陷检测数据集, 采用 2.7 节所述的动态背景数据增广算法

6.2.3 输电导线及其缺陷检测网络

本文采用如 2.1.1 节所述的改进的 SE-YOLOv5 目标检测算法作为输电导线及其缺陷的检测网络。

网络参数设置如下: 原生模型有 263 层 $2.15E7$ 个参数, 引入 SE 模块后, SE-YOLOv5m 有 311 层 $2.16E7$ 个参数。主要参数设置如下: 初始学习率为 0.01, 梯度下降动量为 0.937, GIoU 损失增益为 0.05, 分类损失增益为 0.5, 目标 BCE 损失权重为 1.0, IoU 阈值为 0.5, 多锚阈值为 4.0, 训练代数为 300 代, 批大小为 32, 输入网络的尺寸为 $(640, 640)$ 。基于 YOLO 的检测算法预定义的锚框对收敛速度和最终精度均有较大影响, 我们使用 K-means 聚类方法分析了数据集并计算了合适的锚框, P3, P4 和 P5 对应的锚框分别是: $[(10, 13), (16, 30), (33, 23)]$, $[(30, 61), (62, 45), (59, 119)]$ 和 $[(116, 90), (156, 198), (373, 326)]$ 。

本文采用的评价指标是 2.3.1 节所述的目标检测领域常用的基于混淆矩阵^[74]的指标, 具体是: 查准率 P、查全率 R、平均检测精度 mAP@.5:.95 和 F1 分数。采用综合考虑查准率和查全率的 F1 分数作为主要指标。

6.3 消融实验

6.3.1 数据集

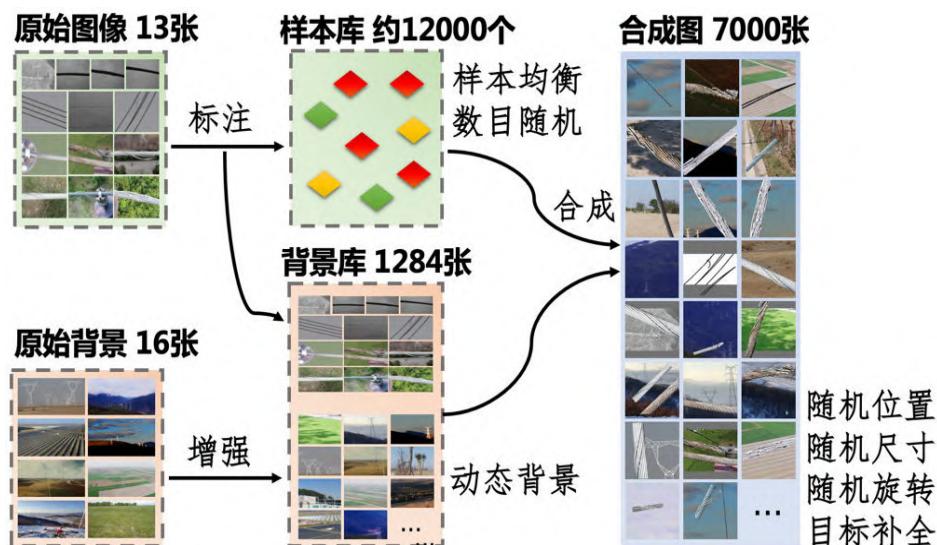


图 6.2 数据集生成流程图

数据集生成流程如图 6.2 所示。基本思路是：(1) 标注原始图像获取小样本；(2) 随机选择获取样本库；(3) 动态背景方法生成背景库；(4) 合成为图像，图像划分为训练集和测试集。

首先，采用数据标注方法对 13 张原始图像进行标注，获得了小样本目标，输电导线、散股和断股目标分别为 22、6 和 11 个，共计 39 个。根据最终生成的合成图像数目为 7000 张，平均每张图约 1.7 个目标（该均值根据原始图像的标注得到），计算总共所需的样本数目约为 15000 个。计算中每张图样本数目采用随机正态分布生成。考虑 2.1 节所述，目标样本之间存在关联关系，即一个输电导线标注区域，可能会包含散股、断股区域等，程序中采用样本均衡系数为 8:1:1，最终生成的约 15000 个样本中，三类目标的比例约为 5:3:2。

另一方面，采用如 2.1.6 节所述的动态背景算法，对 16 张高清背景大图进行多尺度动态背景处理，共生成 1271 张额外的背景图像，加原始图像的 13 张，背景库共计 1284 张。这样做确保了背景库的图像是真实的分布。

最后，我们采用多项随机方法将样本库和目标库进行合成，生成了 7000 张数据集，划分为 6000 张训练集和 1000 张测试集。

6.3.2 图像质量提升算法消融实验

本节分析所提出的质量提升算法的效果。

6.3.2.1 定性分析

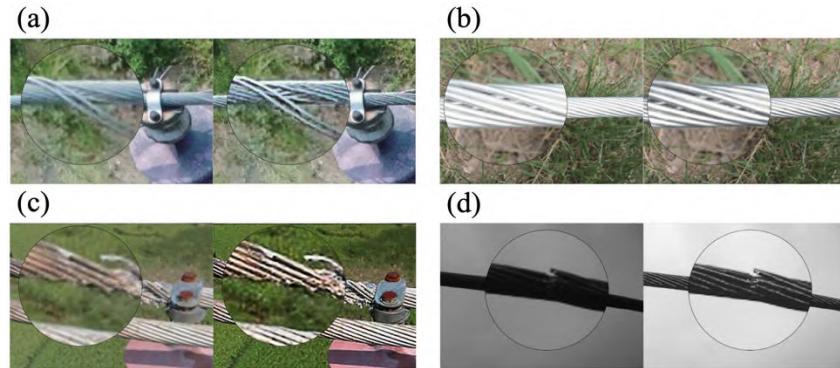


图 6.3 提升图像质量前后对比

应用第 2.1.5 节所述的图像提升质量算法, 结果如图 6.3 所示, 图中显示了 4 张不同背景下的输电导线图像, 每张图左侧是原始的未做任何操作的低分辨率图像, 右侧是应用超分辨率算法和高反差保留算法后的最终结果。处理后的图像不仅有用更高的分辨率, 而且在输电导线细节表现上更加清晰。

处理后的图像分辨率更高的原因是: 每次应用超分辨率算法, 图像的分辨率增加为原来的 2 倍, 应用 n 次后, 分辨率增加为原来的 2^n 倍。

输电导线细节更加清晰的原因是: 在设置合理的高斯半径下, 高反差保留相当于进行了去高斯模糊, 保留了更多的边缘细节, 对于输电导线这种由多根股线组成的图像, 每根股线的边缘更加清晰。

6.3.2.2 定量分析

我们分别使用未经质量提升的原始图像和经过质量提升的图像, 均进行数据增广后, 分别生成低质量数据集 LQ(Low Quality)Set 和高质量数据集 HQ(High Quality)Set。均在所改进的 SE-YOLOv5 网络上训练, 分别得低质量数据集模型 LQM 和高质量数据模型 HQM。测试评估结果如表 1 所示:

表 6.1 低质量数据集模型 LQM 和高质量数据集模型 HQM 性能对比

模型	测试集	目标数	P(%)	R(%)	mAP(%)	F1(%)
LQM	LQSet	1640	87.1	98.9	94.4	92.6
LQM	HQSet	1704	67.5	70.4	53.7	68.2

HQM	HQSet	1704	90.8	99.0	95.0	94.7
-----	-------	------	------	------	------	------

结果表明：LQM 在 LQSet 测试集上的 F1 分数为 92.8%，具有良好的效果，但其 HQSet 测试集上，F1 分数仅达为 68.2%。分析原因，这与训练数据集中图像模式（pattern）与测试集差异交大有关。在公平测试下，即两个模型分别在对应的测试集上测试。结果表明 LQM 的 F1 分数为 92.6%，而 HQM 达到 94.7%，说明本文提出的图像质量增强算法，不仅在直观上更加清晰地展示了图像细节，而且对模型性有 2.1 个点的提升。

6.3.2 噪声背景消融实验

本节分析动态背景算法的效果。

6.3.2.1 定性分析

噪声背景是随机背景的一种。我们采用 0~255 完全随机生成 RGA 图像得到“随机”的背景图来生成数据集，生成的图像如图 6.4 (a) 所示。由于效果检测效果不理想，我们开发了 2.2 节所述的动态背景算法，生成图像如图 6.4 (b-c) 所示。

为了对比噪声背景和动态背景算法的差异，我们分别生成了噪声背景数据集 NB(Noise Background)Set 和动态背景数据集 DB(Dynamic Background)Set，并分别训练了噪声模型 NBM 和动态模型 DBM。原始图像均采用提升分辨率后的图像，网络均采用改进的 SE-YOLOv5。

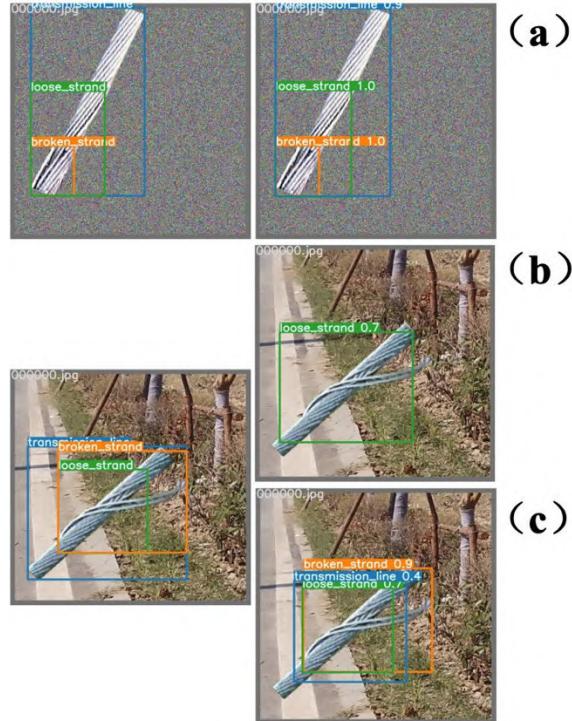


图 6.4 噪声背景模型 NBM 和动态背景模型 DBM 对比。左侧是真值标注，右侧是检测结果。 (a) NBM 在 NBSet 上的结果； (b) NBM 在 DBSet 上的结果； (c) DBM 在 DBSet 上的结果

检测效果直观对比如图 6.4 所示。与左侧的真值标注相比，NBM 在 NBSet 上成功地检测出输电导线、散股和断股，但在 DBSet 上只检测出散股，输电导线和断股均漏检；DBM 在 DBSet 上能成功检测出输电导线、散股和断股。

6.3.2.2 定量分析

NBM 和 DBM 的测试评估结果如表 6.2 所示。

表 6.2 噪声背景模型 NBM 和动态背景模型 DBM 性能对比

模型	测试集	目标数	P(%)	R(%)	mAP(%)	F1(%)
NBM	NBSet	1650	89.2	99.5	94.7	94.0
NBM	DBSet	1704	22.4	40.0	16.4	27.8
DBM	DBSet	1704	90.8	99.0	95.0	94.7

结果表明：与图 6.4 的结果一致，NBM 在 NBSet 上的 F1 分数达 94.0%，说明在噪声背景训练集上训练的模型，对于噪声背景的目标有很好的检测性能。但是，NBM 在 DBSet 上的检测精度大幅度降低，F1 分数仅为 27.5%，分析表明这是由于噪声背景和真实背景在分布上的差异，显著地影响了模型的性能。而 DBM

在 DBSet 上的 F1 分数达到了优秀的 94.7%。

在实际应用中，待检测图像均为真实背景的，所以完全随机噪声背景生成数据集训练的模型是极不鲁棒的。因此，本文提出的动态背景算法，能产生真实的数据的背景多样性，是模型取得高性能的重要原因。

6.3.3 SE 模块消融实验

本节分析引入 SE 注意力机制改进网络的效果。

我们使用 DBSet，在其他参数相同的情况下，分别对原生的 YOLOv5 网络和改进的 SE-YOLOv5 网络进行了模型和性能测试，结果如表 6.3 所示。

表 6.3 原生 YOLOv5 和改进 SE-YOLOv5 性能对比

模型	测试集	目标数	P(%)	R(%)	mAP(%)	F1(%)	Time (ms)
YOLOv5m	DBSet	1704	89.1	98.8	95.4	93.7	4.2
SE-YOLOv5m	DBSet	1704	90.8	99.0	95.0	94.7	4.3

Note: 测试时间是在 RTX 3080 上，批大小为 16，输入尺寸为 640 时测试的推理和非极大值抑制总时间。

结果表明：原生的 YOLOv5 模型能达到 93.7% 的 F1 分数，引入 SE 模块后，模型在测试集上的查准率、查全率和 F1 分数均有所上升，F1 分数提升了 1 个百分点，mAP@.5:.95 有所下降，检测速度略微降低。性能提升的主要原因是 SE 注意力机制能学习各个通道的重要程度，自动给予重要通道的特征更多的权重，速度略微下降的原因是额外的路线增了模型的层数和参数数目。

6.4 结果和分析

6.4.1 最终性能测试

利用本文所提出的图像质量提升算法来获得高质量的原始图像,而后采用所提出的数据增广算法来生成动态背景数据集,在改进的 SE-YOLOv5 模型上训练模型,训练好的模型在测试集上的测试结果如下:

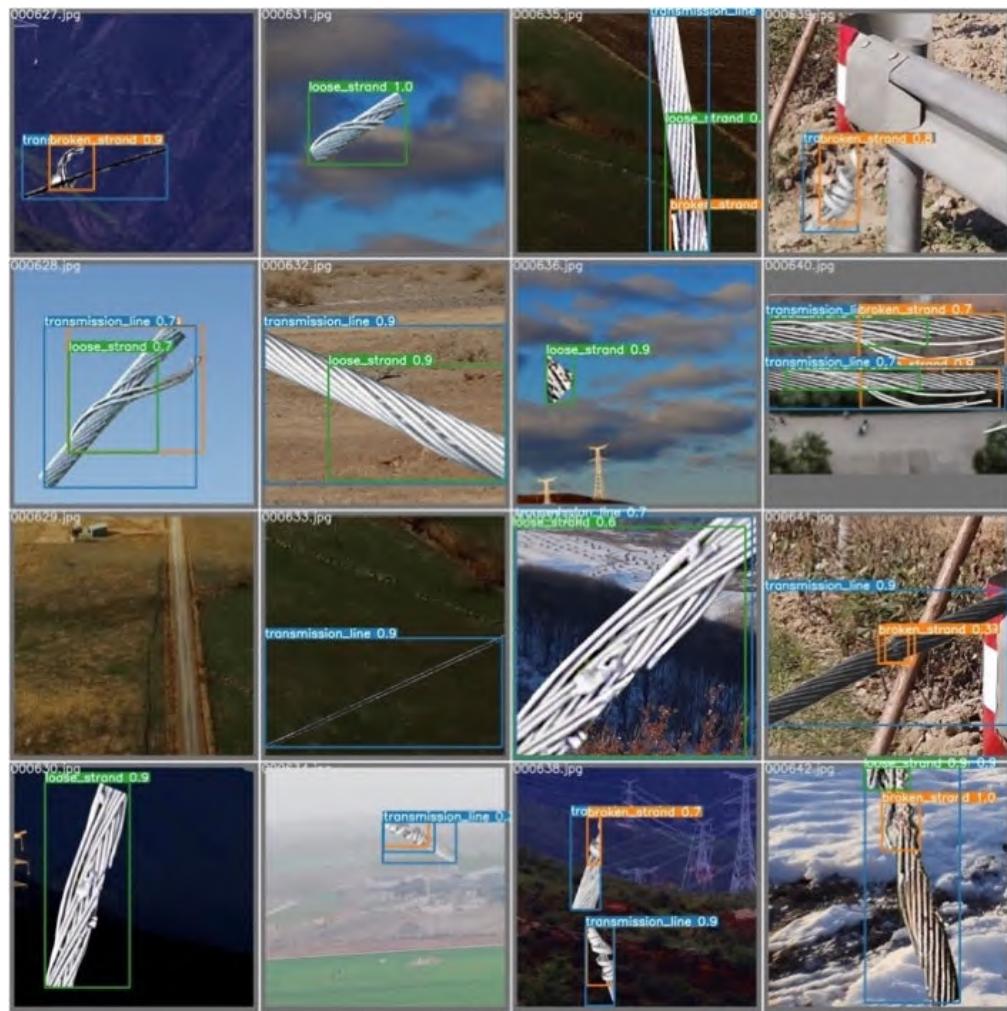


图 6.5 模型最终检测效果

随机抽取一个 batch 的 16 张图检测效果如图 6.5 所示,结果显示,不论输电导线在图像的什么位置,导线走向如何,尺度大小如何,正光还是逆光,本文提出的算法均能准确地检测出输电导线、散股和断股,本文提出的输电导线及其缺陷目标检测算法有较强的泛化能力和鲁棒性。

详细的不同类别的检测性能如表 6.4 所示。共检测 1000 张图像,输电导线、

散股和断股目标数目分别为 847、493 和 364，共计 1704 个，三个类别的检测 F1 分数均大于 90%，平均检测 F1 分数达到 94.7%。

表 6.4 最终模型检测结果

目标类别	目标数	P(%)	R(%)	mAP(%)	F1(%)
平均	1704	90.8	99.0	95.0	94.7
输电导线	847	88.8	97.8	93.0	93.0
散股	493	88.4	99.2	94.7	93.5
断股	364	95.1	99.9	97.3	97.5

速度方面，在 RTX3080 GPU 上，输入尺寸为 640 时，单张图检测的推断和非极大值抑制的总时间约 16.1 ms，即 62.1 FPS。实际离线处理过程中，可多张图合并成 1 个 batch 进行检测，批尺寸为 16 时平均单图检测时间为 4.3ms，约 233FPS。

最终，本文的模型的检测速度为 62.1FPS（单图），F1 分数为 94.7%。

6.4.2 模型的迁移学习应用

需要注意的是，上文训练好的模型虽然能实现随机位置、随机角度、随机尺寸和随机背景的鲁棒检测，但当出现被检测目标具有不同纹理（如图 6.6 中的输电导线、散股和断股不同于原标注的 13 张图像）时，性能会大大折扣，即该模型对不同的位置、角度、尺寸、背景具有鲁棒性，但纹理鲁棒性不强。

为了进行纹理鲁棒性实验和改进，我们额外采集了 64 张全新的输电导线和缺陷图像，进行了相关测试，结果如表 6.5 所示。

表 6.5 噪声模型 NBM 和混合模型 Hybrid-M 性能对比

模型	测试集	目标数	P(%)	R(%)	mAP(%)	F1(%)
NBM	Additional-Set	162	84.3	60.9	58.7	68.5
Hybrid-M	Hybrid-Set	3152	91.7	98.3	94.4	94.9
Hybrid-M	DBSet	1704	91.1	98.8	94.4	94.8

纹理鲁棒性实验结果如表 6.5 的第一行所示，直接用上文的动态背景模型 DBM(13 张原始图像增广后的数据集训练的)对额外图像进行检测的结果。结果表明，模型虽然有 84.3% 的查准率，查全率仅 60.9%，综合 F1 分数为 68.5%。这表明模型有一定性能（高于完全随机基准），但鲁棒性不足。

解决模型的纹路鲁棒性的有效途径是使用迁移学习特性进行模型性能的改进。方法是：对额外采集的 64 张图像采取上文所述相同的流程，先用 labelme 标注，然后用图像质量提升算法、数据增广算法生成数据集 7000 张（增广是必要的，否则会造成样本不均衡），并与原来 7000 张 DBSet 合并形成混合数据集 Hybrid-Set；而后采用迁移学习的方法，用 Hybrid-Set 训练集对 DBM 模型进行迁移训练，得到混合模型 Hybrid-M。

对迁移学习详细描述如下：第一步加载与原模型完全相同的网络结构，即 SE-YOLOv5；第二步加载原来训练好的权重文件，即.pt 文件，加载过程中，读取训练好的每一个卷积层、全连接层等的权重并进行自动设置；第三步，将混合数据集作为输入，对模型进行训练，训练结束后保存最优性能模型。由于原模型已经具有一定检测能力，将其性能“迁移”到新数据集上，因此称为“迁移学习”。由于新增了数据集，有时也被称为“增量学习”。由于通常会对新的训练过程设置较小的学习率和较小的训练代数，有时也被称为“模型微调”。

迁移学习后的性能如表 5 第二行所示，Hybrid-M 在 Hybrid-Set 测试集（测试集与训练集数据无交叉）上的综合 F1 分数达到了 94.9%，具有很好的鲁棒性。注意，迁移后的模型仅对这 77 种的纹理和接近的纹理具有鲁棒性，对于全新的纹理，可以继续使用本文提出的方法进行迁移学习。本提出的数据增广方法有价值的前提是纹理多样性远小于数据多样性。由于数据多样性包含：纹理多样性、位置多样性、角度多样性、尺寸多样性和背景多样性等，这显然是成立的。

为了证明混合模型对原 DBSet 也有良好的效果，我们也进行了测试，结果如表 5 第三行所示，综合分数为 94.8%，比 DBM 的 94.7% 提升了 0.1%，说明采集更多数据并再次进行标注、增强、增广，进行迁移训练，是解决纹理多样性问题的有效途径。更多的数据，会带来模型好的通用性。

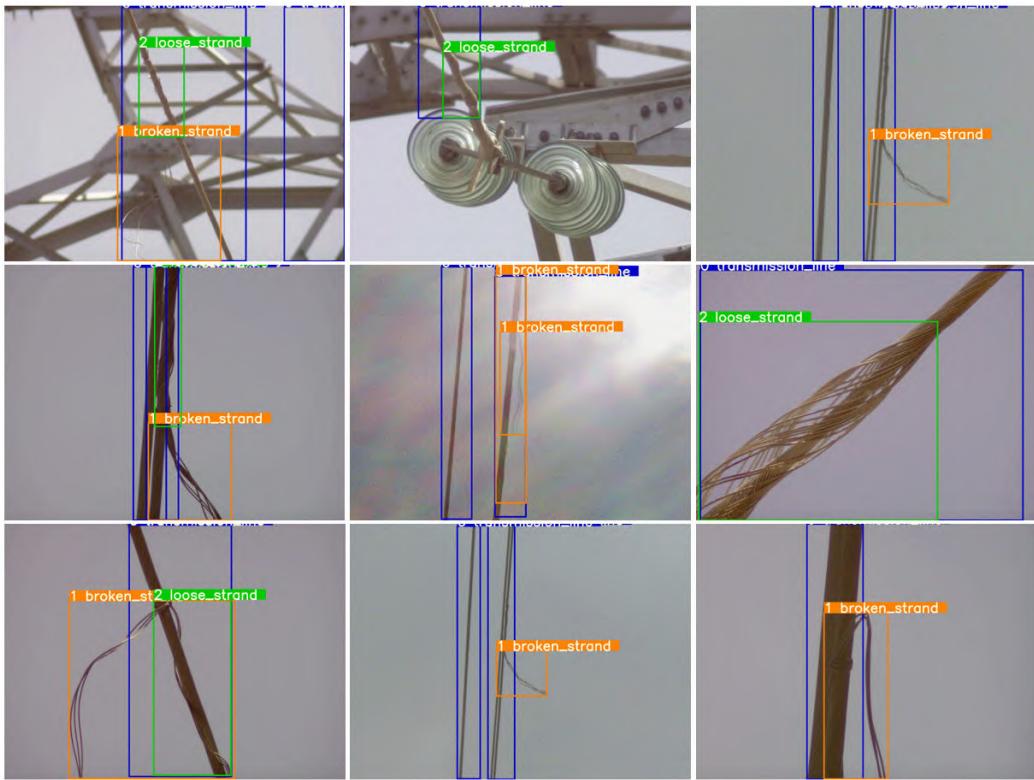


图 6. 6 Hybrid-M 在额外图像上的检测效果

混合数据训练的模型 Hybrid-M 在额外图像上的检测效果直观地展示在图 11 中, 由于被测图像不包含在训练数据集中, 模型表现出的良好性能是通过“学习”得到的。

6.5 总结

为了在低分辨率、小样本下实现动态范围大、环境光多样等复杂背景下的输电导线及其缺陷的实时和鲁棒检测。本文从来自云南呼马山试验基地和文献的 13 张原始输电导线图像出发, 进行了如下工作:

- (1) 实现了超分辨率算法, 解决原始图像分辨率低的问题;
- (2) 实现了高反差保留算法, 解决原始图像细节不清晰的问题;
- (3) 提出和实现了动态背景算法, 生成了多尺度、真实的背景图像;
- (4) 采用样本均衡、随机样本数目、随机位置、随机缩放、随机旋转、目标补全等方法生成了 7000 张高质量输电导线数据集。
- (5) 引入通道注意力机制, 改进网络形成了 SE-YOLOv5 目标检测算法, 经过训练, 实现实时、鲁棒的输电导线、散股和断股三类目标检测。

本文提出的上述数据集生成方法, 可以广泛的应用在各种领域的目标检测任务中, 提供了一种通用的应对低分辨率图像、小样本数据集的方法; 本文生成的输电导线数据集, 能推动基于深度学习的人工智能算法在该领域的应用; 本文训练好的输电导线及其缺陷检测模型, 能应用到无人机输电导线巡检领域, 提高电路系统运行稳定性和安全性; 此外, 本文提出的算法是可以不断迁移学习进行演进的, 提供更丰富的纹理, 能持续提升模型的通用性。

本文算法存在的不足是: 相比于分割的算法, 本算法无法给出图像哪些像素的是目标的信息; 相比与电磁等缺陷检测方法, 本文算法无法检测输电导线内部的缺陷。此外, 本文提出的动态背景数据集增广方法无法提供纹理的多样性, 为了获得更强的鲁棒性, 在使用时需要尽量在增广前的原始标注数据中提供尽可能多纹理种类的目标样本。

第7章 波函数作为知识要素的第三代人工智能

本章在知识-数据协同驱动的新范式下，探索量子力学波函数作为知识要素的第三代人工智能的可行性，进行了理论推广、方案设计，初步进行了验证实验，提出新的技术路线。

7.1 引言

快速发展的人工智能深度学习方法成为主流的目标检测算法，不完全信息场景下性能大幅度下降的问题为目标感知带来了新的发展机遇和挑战，谁能有机融合知识、数据、算法和算力四个要素，解决不完全信息问题，谁就拥有了下一代人工智能技术的主动权。量子理论是一套理解和描述世界的基本理论，复值表示的波函数已经在计算材料学领域被证明了其描述先验知识的可行性。受量子理论启发，本文首次提出以波函数作为知识要素，从学术研究和应用的角度看，本研究需要解决以下挑战：

(1) 在泛函近似方面，如何开发独立自主的神经网络泛函和如何在此基础上处理从微观到宏观的变换是一个挑战。

(2) 在高效求解工具方面，开发效率与神经网络匹配的高性能薛定谔方程解算器是一个挑战。

(3) 从构建、描述、利用知识要素的角度来看，构建知识要素的工具可采用泛函近似和高性能解算器，描述知识要素的方法可采用多体理论，如何利用波函数先验知识，即如何将波函数知识要素与其他三大要素有机融合是一个挑战。

本文有如下意义：

(1) 本文将提出一种全新的构建知识要素的方法，对发展新的知识要素有重要的探索意义。

(2) 本文将利用波函数知识要素，发展新的知识-数据协同驱动目标检测算法，解决不完全信息环境下性能大幅度下降的关键难点，有重要的实践意义。

(3) 本文将发展综合利用知识、数据、算法和算力四大要素的融合方法，是对下一代人工智能的一种探索，有望进一步发展为新的人工智能范式，有重要的开拓意义。

(4) 本文是一套基于物理思想和数学表达的通用的理论和方法，在目标检

测领域取得成功后，易于推广到诸多其他领域，如融合目标跟踪、带先验知识的语义分割、多模态自然语言处理等，有良好的延伸价值。

7.2 内容

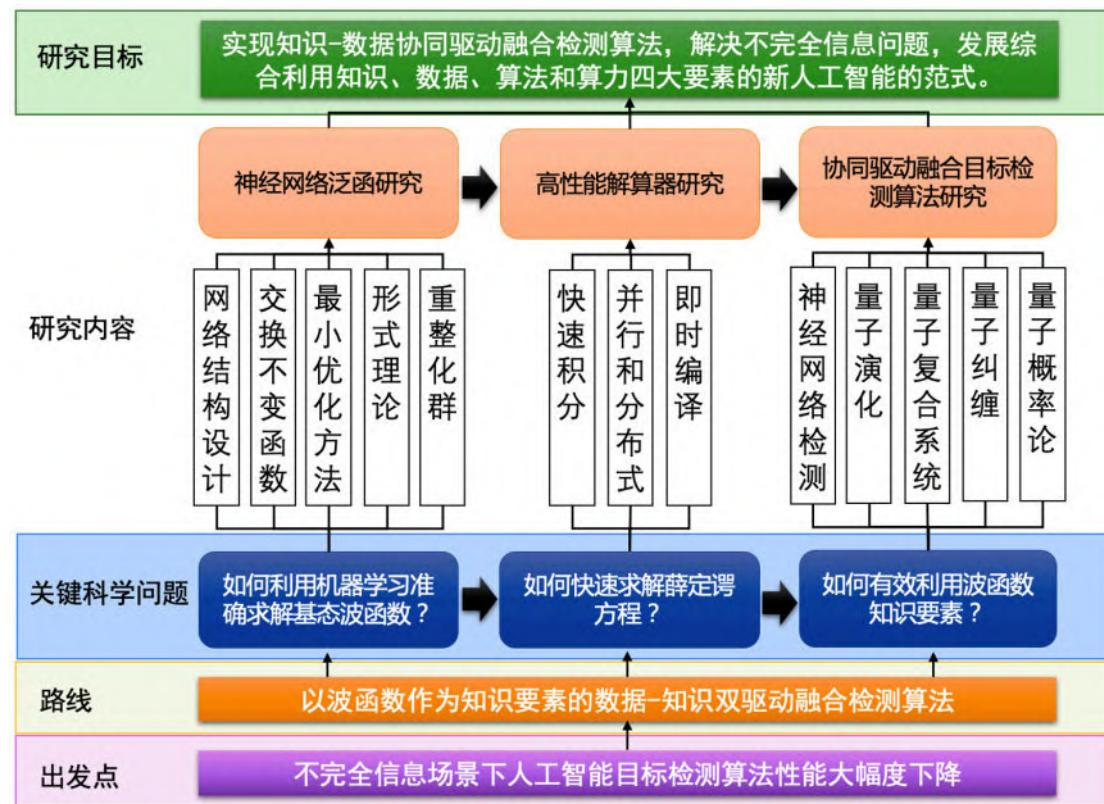


图 7.1 研究内容

本项目拟开展三项研究：神经网络泛函、高性能解算器和协同驱动融合目标检测算法。

7.2.1 神经网络泛函研究

第一部分将采用泛函近似方法研究神经网络泛函。利用神经网络构建满足交换反对称性和边界条件的交换不变函数，然后使用自然梯度下降优化网络，将该网络作为一种 Ansatz 泛函提供准确求解基态波函数的能力。

基于该泛函，研究两个独立的方法处理微观到宏观的变换：一是形式理论 (Formalism)，采用数学形式策略直接描述宏观概念，跳过微观到宏观的过程，虽缺失物理意义，但简单而有效。二是重整化群 (Renormalization Group, RG)，

利用有效理论不断对微观重整合化，得到宏观尺度的整体行为。

第一部分是构建波函数知识要素的基础。

7.2.2 高性能解算器研究

第二部分将采用多项技术研究求解薛定谔方程的高性能解算器，以前期开发的 **Wavelib** 库为基础，在已实现的基于分离变量法、极限思想、前馈差分和后馈差分的离散化和基于 Libcint 的快速积分上，进一步采用 CUDA 编程、Nvlink 和 Direct RDMA 技术实现 **Wavelib** 在 GPU 上的运行、并行和分布式并行，采用加速线性代数(XLA)编译器来实现 **Wavelib** 库的即时编译(Just-in-time compilation)，进一步提升效率，以达到与神经网络效率匹配的目的。

第二部分是波函数知识要素与神经网络协同的关键。**Wavelib** 库可在我们开发的 Damei Python 库中找到[\[72\]](#)。

7.2.3 协同驱动融合目标检测算法研究

第三部分将采用神经网络和量子理论研究不完全信息场景下的融合目标检测算法。

一侧采用“经典”的神经网络目标检测算法，使用残差网络、空间金字塔池化、跨阶段局部网络、密集连接、空间注意力机制、通道注意力机制、路径聚合网络、特征金字塔网络等技术，进行 One-shot 目标检测，输出视为一个模态：检测结果模态。另一侧采用量子启发的波函数知识要素。利用 POV (Projection-Valued Measure) 测量将某一时刻的检测结果坍缩到某一特定的基态波函数上，利用幺正变换使系统在哈密顿量上演化，为下一时刻提供先验知识，构建第二个模态：知识信息模态。下一时刻，采用量子复合系统建模由上一时刻演化得到的波函数和当前测量得到的波函数，融合的波函数表示为由两个空间的任意基的线性组合构成的复合希尔伯特空间下的态矢。融合过程中，双模态的不可分离性由量子纠缠理论处理，不兼容性由量子概率论处理。

神经网络目标检测算法是数据驱动的，知识信息模态是知识驱动的。该知识的本质是 One-shot 目标检测算法丢失的、由量子演化和冯诺依曼纠缠熵“找回”的上下文信息，如：系统状态随时间演化的前后帧信息、目标-目标之间的交互信

息、目标与环境的交互信息等。其中，前后帧信息也可以用基于跟踪或基于循环神经网络的算法处理，后两项的处理方法还未见报道。

第三部分的知识-数据协同驱动融合目标检测算法是描述、利用波函数知识要素的具体实践，是实现和验证四要素范式神经网络的开创性内容。

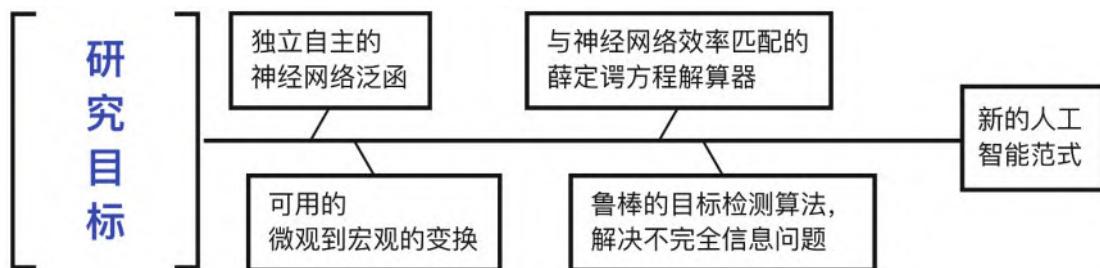


图 7.2 研究目标

本项目从不完全信息场景下人工智能目标检测算法性能大幅度下降的难点问题出发，独辟蹊径提出以量子力学波函数作为知识要素，研究知识-数据协同驱动融合目标检测算法。

凝练关键科学问题，针对如何利用机器学习求解基态波函数的问题，研究神经网络泛函，目标是利用神经网络的表现力和灵活性，开发出独立自主的、更精确的 Ansatz 泛函，并在此基础上，利用形式理论和重整化群方法，实现可用的从微观到宏观的变换。

针对如何快速求解薛定谔方程的问题，研究高性能解算器，目标是基于快速积分、分布式并行和即时编译技术，发展出一套快速的、与神经网络效率匹配的薛定谔方程求解工具。

针对如何有效利用波函数知识要素的问题，研究协同驱动融合检测算法，目标是利用由量子演化和冯诺依曼纠缠熵补充的上下文信息，发展出更鲁棒的目标检测算法。

研究层层递进，项目目标是实现以波函数作为知识要素的融合目标检测算法，解决不完全信息问题。最终目标是综合利用知识、数据、算法和算力四大要素，发展新的人工智能范式。

7.3 方法



图 7.3 研究方案

本文拟采用理论推广、工具实现和实际案例相结合的研究方案并循序渐进地展开。

首先，将量子理论推广到人工智能感知系统中，推导物质波场薛定谔方程的表达形式，基于交换不变函数和 KFAC 优化方法，研究神经网络泛函，来准确地构建基态波函数，并在此基础上，研究基于形式理论或重整化群的微观到宏观的变换。

而后，以离散化、快速积分为基础，基于分布式并行和即时编译技术，研究求解薛定谔方程的高性能解算器。

最后，在不完全信息场景下，设计双模态融合算法，采用神经网络泛函、高性能解算器和量子测量方法来构建知识要素，形成知识要素模态；采用量子演化、冯诺依曼纠缠熵描述来知识要素，补充上下文信息；采用量子复合系统、量子纠缠和量子概率论来利用知识要素，融合知识信息模态和检测结果模态。

最终实现鲁棒的目标检测算法，解决不完全信息问题，验证方法的有效性，实现构建、描述和利用波函数知识要素的理论、方法和工具。

本文采用的方法如 2.2 节所述，基于量子力学的基础知识，实现神经网络泛函和高性能解算器，最终开发知识和数学协同驱动的算法。

量子力学描述一个物理系统有三个基本的要素：**量子态**、**可观测量**和**动力学表达**（随时间演化规律）。

量子态是体系的状态，即波函数，波函数的模方是粒子出现的概率，波函数

不是唯一的，且是一个平方可积的线性空间，一个平方可积的空间是一个希尔伯特空间，希尔伯特空间中的向量就是波函数。一个量子态是一个态矢量，在希尔伯特空间中选定一组基矢（即选定表象）后，态矢量可以用波函数在这组基矢上的投影表示。数学上，可以不引入坐标系，直接用符号表示矢量。类似地，量子力学中的态矢量也可以不引入表象，直接用矢量符号表示，即狄拉克算符。狄拉克算符把希尔伯特空间分为左矢空间和右矢空间，左矢 $\langle\psi|$ 和右矢 $|\psi\rangle$ 如2.2.1节所述。

任意可观测量由算符 \hat{O} 表示。任意外力学量（位置、动量等）的期望值可以用狄拉克符号表示为：

$$\langle O \rangle = \langle \psi | \hat{O} | \psi \rangle \quad (7.1)$$

例如，位置算符 $\hat{x} = x$ ，动量算符 $\hat{p} = -i\hbar\nabla$ 。

波函数随时间的变化即动力学表达，薛定谔方程左侧是关于时间的一阶偏导，右侧是关于位置的二阶偏导，求解过程中使用变分原理将变量分离，求得含时项的解为： $e^{-iE_n t/\hbar}$ ，因此总波函数表达为定态波函数与含时项的积：

$$\Psi(\vec{r}, t) = \psi_n(\vec{r}) e^{-iE_n t/\hbar} \quad (7.2)$$

求解多体波函数的问题归结为求解公式2.13式的薛定谔方程，其在一维情况的形式如下：

$$i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2 \Psi}{\partial x^2} + V\Psi \quad (7.3)$$

采用分离变量法把时间和空间部分分离： $\Psi(x, t) = \psi(x)\varphi(t)$ ，得到：

$$\left\{ \begin{array}{l} \frac{d\varphi}{dt} = -\frac{iE}{\hbar}\varphi \\ -\frac{\hbar^2}{2m} \frac{d^2\psi}{dx^2} + V\psi = E\psi \end{array} \right. \quad (7.4)$$

其中，第二式为定态薛定谔方程。公式7.4的解为：

$$\begin{cases} \varphi(t) = e^{-iEt/\hbar} \\ \psi(x) = \sum_{n=1}^{\infty} c_n \psi_n(x) \end{cases} \quad (7.5)$$

定态解 $\{\psi_n(x)\}$ 中每个解对应一个分离的能量 $\{E_n\}$ ，即允许的能量是量子化的。

体系波函数的通解为：

$$\Psi(x, t) = \sum_{n=1}^{\infty} c_n \psi_n(x) e^{-iE_n t/\hbar} = \sum_{n=1}^{\infty} c_n \Psi_n(x, t) \quad (7.6)$$

其中， $|\psi(x, t)|^2$ 是在位置 x 和时间 t 找到粒子的概率。 $|c_n|^2$ 是一次测量得体系的能量为 E_n 的概率。能量的期望值是 $\langle H \rangle = \sum_{n=1}^{\infty} |c_n|^2 E_n$ 。

7.4 验证实验

7.4.1 无限深势阱

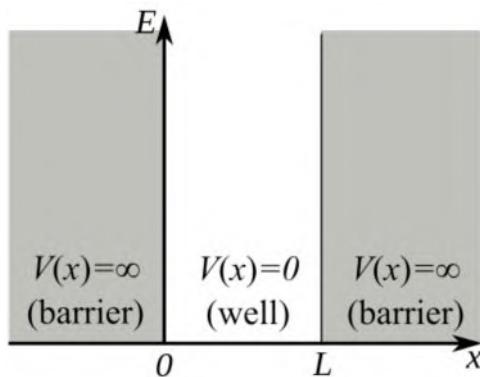


图 7.4 无限深势阱体系

无限深势阱体系如图 7.4 所示，是理想化的简单体系，用于基础的验证。势阱长度为 L ，在势阱内部，粒子受到的势场为 0，在其他区域势阱无穷大，即：

$$V(x) = \begin{cases} 0, & 0 \leq x \leq L \\ \infty, & \text{otherwise} \end{cases} \quad (7.7)$$

将其代入薛定谔方程 7.4 式，解为：

$$\begin{cases} \psi_n = \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi x}{L}\right) \\ E_n = \frac{n^2\pi^2\hbar^2}{2ma^2} \end{cases} \quad (7.8)$$

用我们开发的 Wavelib 库求解无限深势阱体系，结果如图 7.5 所示。

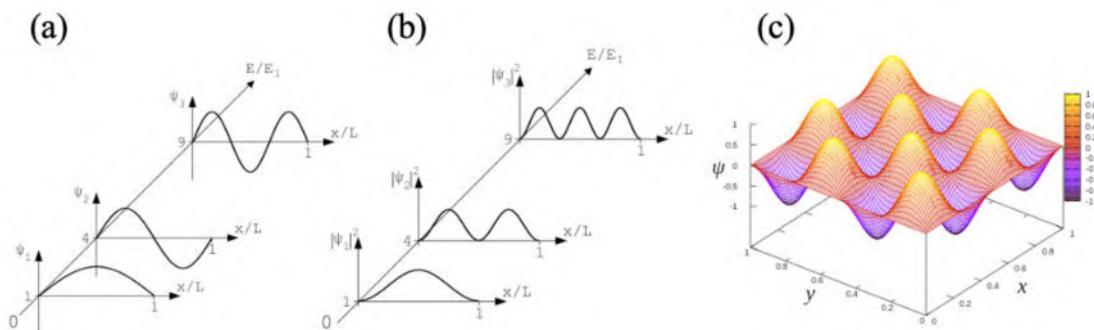


图 7.5 无限深势阱体系结果 (a) 一维无限深势阱波函数 (b) 一维无限深势阱概率密度
(c) 二维无限深势阱波函数

无限深势阱是为数不多的可直接求解出的体系, 图中所示采用 `Wavelib` 库的数值解法求解的结果与直接求解的结果相同。

7.4.2 H 原子体系

H 原子体系是可以直接求解的真实系统之一。体系由一个 H 原子核和一个核外电子组成，电子感受到的势场是中心势场：

$$V(\vec{r}) \rightarrow V(r) = -\frac{e^2}{4\pi\epsilon_0 r} \quad (7.8)$$

球坐标系如图 7.6 所示：

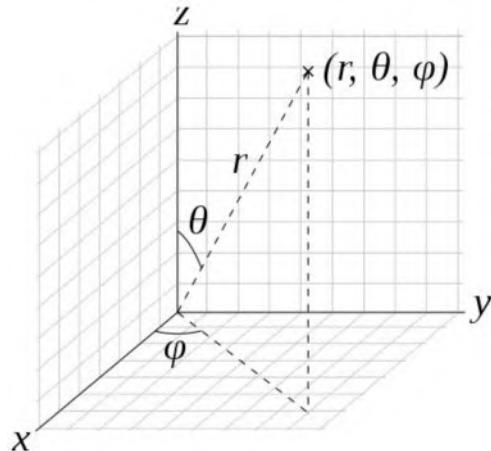


图 7.6 球坐标系

球坐标系下， r 是径向距离 (Radial Distance)，表示球面坐标点到球心的距离。 θ 是极角 (Polar Angle)，表示 z 轴与 r 的夹角，取值范围是 $[0, \pi]$ 。 ϕ 是方位角 (Azimuth Angle)，表示赤道面 (x 轴与 y 轴确定的平面) 上起始于 x 轴，沿着逆时针方向量出的角度。注意此处的 ϕ 有时也表示为 ϕ ，它与波函数的符号有重复，但表达的意义不同，本文采用学界一贯以来的表达习惯，未加以区分，读者应自行注意。公式 2.13 的薛定谔方程中各项变换如下：

$$\begin{aligned} r &= \sqrt{x^2 + y^2 + z^2} \\ \theta &= \arccos\left(\frac{z}{r}\right) = \arcsin\left(\frac{\sqrt{x^2 + y^2}}{r}\right) = \arctan\left(\frac{\sqrt{x^2 + y^2}}{z}\right) \\ \phi &= \arccos\left(\frac{x}{r \sin \theta}\right) = \arcsin\left(\frac{y}{r \sin \theta}\right) = \arctan\left(\frac{y}{x}\right) \\ \nabla^2 &\equiv \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \left(\frac{\partial^2}{\partial \phi^2} \right) \end{aligned} \quad (7.9)$$

同样,采用分离变量法对径向部分和角向部分分离, $\psi(r,\theta,\phi) = R(r)Y(\theta,\phi)$,

得:

$$\begin{cases} \frac{1}{R} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) - \frac{2mr^2}{\hbar^2} [V(r) - E] = \ell(\ell+1) \\ \frac{1}{Y} \left\{ \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial Y}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2 Y}{\partial \phi^2} \right\} = -\ell(\ell+1) \end{cases} \quad (7.10)$$

其中, 第一式为径向方程 (Radial Equation), 第二式为角向方程 (Angular Equation)。

7.4.2.1 求解径向方程

求解径向方程, 将式 7.8 的单电子感受到的势场代入, 有:

$$-\frac{\hbar^2}{2m_e} \frac{d^2 u}{dr^2} + \left[-\frac{e^2}{4\pi\epsilon_0 r} \frac{1}{r} + \frac{\hbar^2}{2m_e} \frac{\ell(\ell+1)}{r^2} \right] u = Eu \quad (7.11)$$

其中, 有效势为: $V_{eff} = -\frac{e^2}{4\pi\epsilon_0 r} \frac{1}{r} + \frac{\hbar^2}{2m_e} \frac{\ell(\ell+1)}{r^2}$, 方程的解为:

$$\begin{cases} R_{n\ell} = \sqrt{\left(\frac{2}{na}\right)^3 \frac{(n-\ell-1)!}{2n(n+\ell)}} e^{-\frac{r}{na}} \left(\frac{2r}{na}\right)^\ell \left[L_{n-\ell-1}^{2\ell+1} \left(\frac{2r}{na}\right) \right] \\ E_n = -\left[\frac{m_e}{2\hbar^2} \left(\frac{e^2}{4\pi\epsilon_0} \right)^2 \right] \frac{1}{n^2}, n = 1, 2, 3, \dots \end{cases} \quad (7.12)$$

其中, 主量子数 $n = 1, 2, 3, \dots$, 波尔半径 $a \equiv \frac{4\pi\epsilon_0 \hbar^2}{m_e e^2} = 0.529 \times 10^{-10} m$, 连带拉盖尔多项式为:

$$L_q^p(x) \equiv (-1)^p \left(\frac{d}{dx} \right)^p L_{p+q}(x) \quad (7.13)$$

拉盖尔多项式为:

$$L_q(x) \equiv \frac{e^x}{q!} \left(\frac{d}{dx} \right)^q (e^{-x} x^q) \quad (7.14)$$

各连带拉盖尔多项式 $L_{n-\ell-1}^{2\ell+1}$ 为:

$L_0^0(x) = 1$	$L_0^2(x) = 1$
$L_1^0(x) = -x + 1$	$L_1^2(x) = -x + 3$
$L_2^0(x) = \frac{1}{2}x^2 - 2x + 1$	$L_2^2(x) = \frac{1}{2}x^2 - 4x + 6$
$L_0^1(x) = 1$	$L_0^3(x) = 1$
$L_1^1(x) = -x + 2$	$L_1^3(x) = -x + 4$
$L_2^1(x) = \frac{1}{2}x^2 - 3x + 3$	$L_2^3(x) = \frac{1}{2}x^2 - 5x + 10$

图 7.7 各连带拉盖尔多项式的函数形式

7.11 式的解 7.12 式中, 能量 $E_1 = -13.6\text{ eV}, E_2 = -3.40\text{ eV}, \dots$, 如图 7.8 所示。

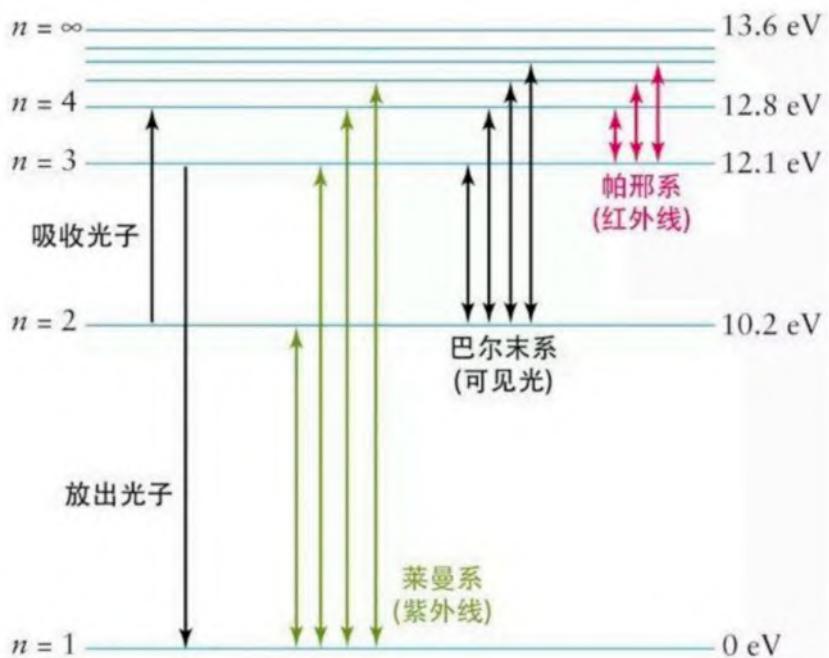


图 7.8 H 原子体系的能级

图 7.8 即量子力学中经典的 H 原子体系的能力, 能量是分裂的, 电子从低能级跃迁到高能级需要吸收光子, 从高能级跃迁到低能级会放出光子, 释放或者吸收的能量与能级间的能量差直接相关, 其中莱曼系对应紫外线, 巴耳末系对应可见光, 帕邢系对应红外线。

7.4.2.2 求解角向方程

求解公式 7.10 中的角向方程，再次分离变量： $Y(\theta, \phi) = \Theta(\theta)\Phi(\phi)$ ，有：

$$\left\{ \begin{array}{l} \frac{1}{\Phi} \frac{d^2 \Phi}{d\phi^2} = -m^2 \\ \frac{1}{\Theta} \left[\sin \theta \frac{d}{d\theta} \left(\sin \theta \frac{d\Theta}{d\theta} \right) \right] + \ell(\ell+1) \sin^2 \theta = m^2 \end{array} \right. \quad (7.15)$$

上式的解为：

$$\left\{ \begin{array}{l} \Phi(\phi) = e^{im\phi} \\ \Theta(\theta) = AP_\ell^m(\cos \theta) \end{array} \right. \quad (7.16)$$

其中，方位角是解是周期性是，即：

$$\begin{aligned} e^{im(\phi+2\pi)} &= e^{im\phi} \\ e^{2\pi im} &= 1 \\ m &= 0, \pm 1, \pm 2, \dots \end{aligned} \quad (7.17)$$

式中， m 是磁量子数。

解 7.16 式中的连带勒让德函数（Associated Legendre Function）为：

$$P_\ell^m \equiv (-1)^m (1-x^2)^{\frac{m}{2}} \left(\frac{d}{dx} \right)^m P_\ell(x) \quad (7.18)$$

而勒让德多项式（Legendre Polynomial）由罗德里格斯公式（Rodrigues Formula）给出：

$$P_\ell(x) \equiv \frac{1}{2^\ell \ell!} \left(\frac{d}{dx} \right)^\ell (x^2 - 1)^\ell \quad (7.19)$$

其中，角量子数 ℓ 必须是非负整数，且 $m > \ell$ 时， $P_\ell^m = 0$ ，即：

$$\ell = 0, 1, 2; m = -\ell, -\ell + 1, \dots, 0, \dots, \ell - 1, \ell \quad (7.20)$$

各项连带勒让德多项式的函数形式为：

$P_0^0 = 1$	$P_2^0 = \frac{1}{2}(3\cos^2 \theta - 1)$
$P_1^1 = -\sin \theta$	$P_3^3 = -15\sin \theta(1 - \cos^2 \theta)$
$P_1^0 = \cos \theta$	$P_3^2 = 15\sin^2 \theta \cos \theta$
$P_2^2 = 3\sin^2 \theta$	$P_3^1 = -\frac{3}{2}\sin \theta(5\cos^2 \theta - 1)$
$P_2^1 = -3\sin \theta \cos \theta$	$P_3^0 = \frac{1}{2}(5\cos^3 \theta - 3\cos \theta)$

图 7.9 各项连带勒让德多项式的函数形式

前几个连带勒让德多项式的几个直观图示如图 7.10 所示。

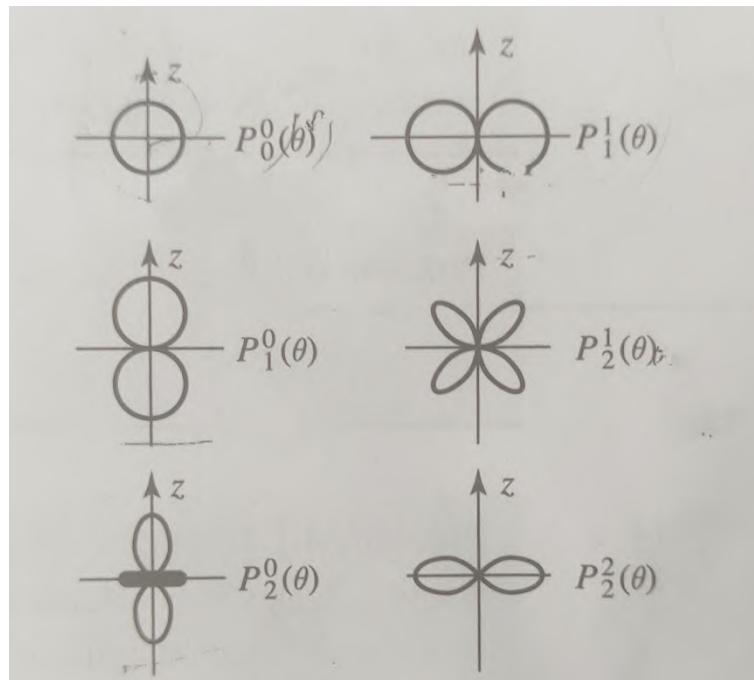


图 7.10 前几个连带勒让德多项式图示

为确定公式 7.16 中的系数 A ，我们对空间波函数进行归一化，有：

$$\left\{ \begin{array}{l} \int_0^\infty |R|^2 r^2 dr = 1 \\ \int_0^\pi \int_0^{2\pi} |Y|^2 \sin \theta d\theta d\phi = 1 \end{array} \right. \quad (7.21)$$

解得：

$$A = \sqrt{\frac{(2\ell+1)}{4\pi} \frac{(\ell-m)!}{(\ell+m)!}} \quad (7.22)$$

将 7.16 和 7.22 式整合, 得角向方程的解, 称为球谐函数 $Y_\ell^m(\theta, \phi)$, 表达式

为:

$$Y_\ell^m(\theta, \phi) = \sqrt{\frac{(2\ell+1)}{4\pi} \frac{(\ell-m)!}{(\ell+m)!}} e^{im\phi} P_\ell^m(\cos \theta) \quad (7.23)$$

球谐函数与角量子数 ℓ 和磁量子数 m 直接相关, 典型的角量子数和磁量子数取值下的具体表示式如图 7.11 所示。

ℓ	m	$\Phi(\varphi)$	$\Theta(\theta)$	极坐标中的表达式	直角坐标中的表达式	量子力学中的记号
0	0	$\frac{1}{\sqrt{2\pi}}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2\sqrt{\pi}}$	$\frac{1}{2\sqrt{\pi}}$	s
1	0	$\frac{1}{\sqrt{2\pi}}$	$\sqrt{\frac{3}{2}} \cos \theta$	$\frac{1}{2} \sqrt{\frac{3}{\pi}} \cos \theta$	$\frac{1}{2} \sqrt{\frac{3}{\pi}} \frac{z}{r}$	p_z
1	+1	$\frac{1}{\sqrt{2\pi}} \exp(i\varphi)$	$\frac{\sqrt{3}}{2} \sin \theta$	$\frac{1}{2} \sqrt{\frac{3}{\pi}} \sin \theta \cos \varphi$	$\frac{1}{2} \sqrt{\frac{3}{\pi}} \frac{x}{r}$	p_x
	-1	$\frac{1}{\sqrt{2\pi}} \exp(-i\varphi)$	$\frac{\sqrt{3}}{2} \sin \theta$	$\frac{1}{2} \sqrt{\frac{3}{\pi}} \sin \theta \sin \varphi$	$\frac{1}{2} \sqrt{\frac{3}{\pi}} \frac{y}{r}$	p_y
2	0	$\frac{1}{\sqrt{2\pi}}$	$\frac{1}{2} \sqrt{\frac{5}{2}} (3 \cos^2 \theta - 1)$	$\frac{1}{4} \sqrt{\frac{5}{\pi}} (3 \cos^2 \theta - 1)$	$\frac{1}{4} \sqrt{\frac{5}{\pi}} \frac{2z^2 - x^2 - y^2}{r^2}$	$d_{3z^2 - r^2}$
2	+1	$\frac{1}{\sqrt{2\pi}} \exp(i\varphi)$	$\frac{\sqrt{15}}{2} \sin \theta \cos \theta$	$\frac{1}{2} \sqrt{\frac{15}{\pi}} \sin \theta \cos \theta \cos \varphi$	$\frac{1}{2} \sqrt{\frac{15}{\pi}} \frac{zx}{r^2}$	d_{zx}
	-1	$\frac{1}{\sqrt{2\pi}} \exp(-i\varphi)$	$\frac{\sqrt{15}}{2} \sin \theta \cos \theta$	$\frac{1}{2} \sqrt{\frac{15}{\pi}} \sin \theta \cos \theta \sin \varphi$	$\frac{1}{2} \sqrt{\frac{15}{\pi}} \frac{yz}{r^2}$	d_{yz}
2	+2	$\frac{1}{\sqrt{2\pi}} \exp(2i\varphi)$	$\frac{\sqrt{15}}{4} \sin^2 \theta$	$\frac{1}{4} \sqrt{\frac{15}{\pi}} \sin^2 \theta \cos 2\varphi$	$\frac{1}{4} \sqrt{\frac{15}{\pi}} \frac{x^2 - y^2}{r^2}$	$d_{x^2 - y^2}$
	-2	$\frac{1}{\sqrt{2\pi}} \exp(-2i\varphi)$	$\frac{\sqrt{15}}{4} \sin^2 \theta$	$\frac{1}{4} \sqrt{\frac{15}{\pi}} \sin^2 \theta \sin 2\varphi$	$\frac{1}{2} \sqrt{\frac{15}{\pi}} \frac{xy}{r^2}$	d_{xy}
3	0	$\frac{1}{\sqrt{2\pi}}$	$\frac{1}{2} \sqrt{\frac{7}{2}} (5 \cos^3 \theta - 3 \cos \theta)$	$\frac{1}{4} \sqrt{\frac{7}{\pi}} (5 \cos^3 \theta - 3 \cos \theta)$	$\frac{1}{4} \sqrt{\frac{7}{\pi}} \frac{z(2z^2 - 3x^2 - 3y^2)}{r^3}$	$f_{z(5z^2 - 3r^2)}$

图 7.11 典型的角量子数和磁量子数下球谐函数的具体表达式

7.4.2.3 H 原子体系的定态解

将径向方程的解 7.12 式和角向方程的解 7.23 式联合, 得到不氢原子体系的定态波函数 (不含时) 的解为:

$$\psi_{nlm} = R_{nl}(r) Y_\ell^m(\theta, \phi) = \sqrt{\left(\frac{2}{na}\right)^3 \frac{(n-\ell-1)!}{2n(n+\ell)}} e^{-\frac{r}{na}} \left(\frac{2r}{na}\right)^\ell \left[L_{n-\ell-1}^{2\ell+1} \left(\frac{2r}{na}\right)\right] Y_\ell^m(\theta, \phi) \quad (7.24)$$

其中, 主量子数 n 决定了能量, 对应电子所能存在的不同壳层。角量子数 ℓ 决定了电子亚层, 决定了轨道类型, 如 s 、 p 、 d 、 f 轨道。磁量子数 m 决定了轨道的形状, 如 p 轨道分为 p_x, p_y, p_z 轨道。

在多电子体系中，电子优先占据低能级的轨道，占据顺序如图 7.12 所示。

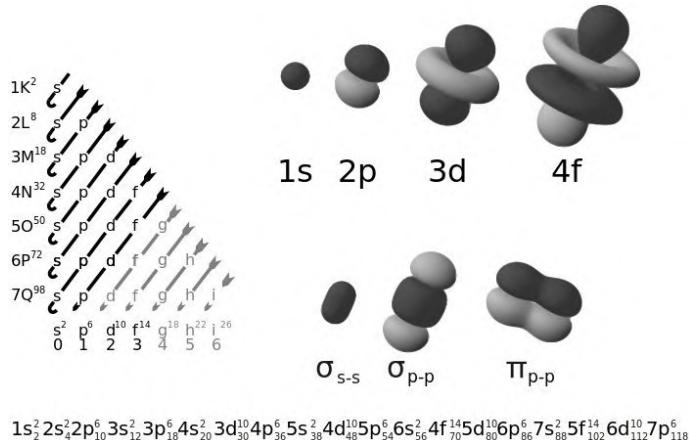


图 7.12 电子占据轨道的顺序

不同的主量子数、角量子数和磁量子数下的电子层一览如图 7.13 所示。

s ($l=0$)		p ($l=1$)		d ($l=2$)						f ($l=3$)					
$m=0$	$m=0$	$m=\pm 1$	$m=0$	$m=\pm 1$		$m=\pm 2$		$m=0$	$m=\pm 1$	$m=\pm 2$	$m=\pm 3$	$m=0$	$m=\pm 1$	$m=\pm 2$	$m=\pm 3$
s	p_z	p_x	p_y	d_{z^2}	d_{xz}	d_{yz}	d_{xy}	$d_{x^2-y^2}$	f_{z^3}	f_{xz^2}	f_{yz^2}	f_{xyz}	$f_{z(x^2-y^2)}$	$f_{x(x^2-3y^2)}$	$f_{y(3x^2-y^2)}$
$n=1$	•														
$n=2$	•	●	●	●											
$n=3$	•	●	●	●	●	●	●	●	●	●	●				
$n=4$	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
$n=5$	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
$n=6$	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
$n=7$	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●

图 7.13 不同主量子数、角量子数和磁量子数的电子层一览

7.2.2.4 Wavelib 求解 H 原子体系波函数

采用 Wavelib 求解 H 原子体系的波函数，设置 H 原子位置为原点，基组选取 6-311G，如下：

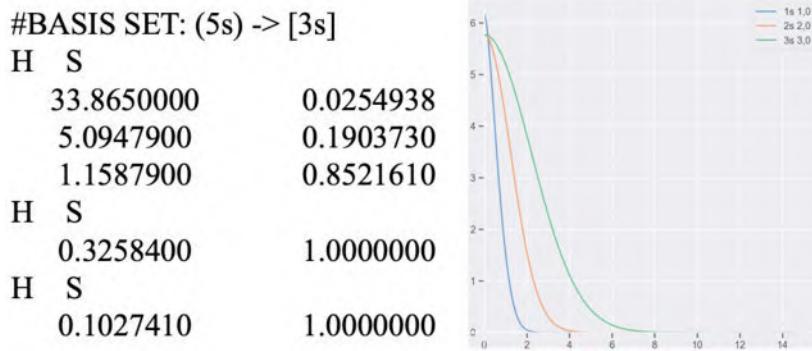


图 7.14 6-311G 基组及其图示

(1) 求解密度矩阵 D

使用最小轨道 (MINAO) 方法初猜密度矩阵 D ，H 原子有 1 个电子，占据数为 $[1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0]$ ，总共 $s*1+1p*3+1d*5+1f*7=16$ 个轨道。

读取 ANO 基组，将 ANO 基组的密度矩阵投影到 6-311G 基组的密度矩阵中。设 H 原子在 ANO 基组下的密度矩阵为 D_1 ，在 6-311G 基组下的密度矩阵为 D_2 。H 原子的电子倾向于占据能量最低的轨道，密度矩阵为占据数对角化矩阵：

$$D_1 = \begin{bmatrix} 1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix} \quad (7.25)$$

形状为 $(16, 16)$ 。总能找到一个变换矩阵 P ，使得

$$D_2 = P D_1 P \quad (7.26)$$

其中：

$$P = S_{22}^{-1} \langle S_{22} | S_{11} \rangle = S_{22}^{-1} S_{21} \quad (7.27)$$

其中， S_{11} 和 S_{22} 分别为 ANO 和 6-311G 轨道的自重叠矩阵， S_{21} 为交叉重叠矩阵。

设基组的指数系数分别为 $\alpha, \beta, \gamma, \delta$ ，四个积分中心分别为 A, B, C, D ，则重叠积分公式为：

$$I^{overlap} = \int e^{-\alpha|r-A|^2} e^{-\gamma|r-C|^2} dr \quad (7.28)$$

代入高斯基组，计算出 6-311G 基组的重叠矩阵，矩阵元为：

$$S_{ij} = \langle \mu_i | \mu_j \rangle = \tilde{N}_i \tilde{N}_j \int_0^\infty r^{2\ell+2} e^{-\alpha_i |r-R_A|^2} e^{-\alpha_j |r-R_C|^2} dr \quad (7.29)$$

类似地, 计算出 ANO 的重叠矩阵 S_{II} , 得到形状为 (16, 16) 的单位矩阵。

对于交叉矩阵 S_{2I} , 交叉积分为:

$$\langle \mu | \text{int} \sigma | \nu \rangle, \mu \in \text{basis1}, \nu \in \text{basis2} \quad (7.30)$$

矩阵元 $S_{ij} = \langle \mu_i | \nu_j \rangle$ 。

根据式 7.27 求得 P 矩阵, 并代入式 7.26 即可求得最终得到 6-311G 基组下的密度矩阵为 D_2 , 值为:

$$D_2 = \begin{bmatrix} 0.096 & 0.155 & 0.1 \\ 0.155 & 0.25 & 0.162 \\ 0.1 & 0.162 & 0.105 \end{bmatrix}$$

(2) 计算哈密顿量

哈密顿量包含核哈密顿量, 有效势。

核哈密顿量记为 h_{1e} ,

$$h_{1e} = \langle \mu | \hat{T} | \nu \rangle + \langle \mu | \hat{V}_{nuc} | \nu \rangle \quad (7.31)$$

式中, 第一项为动能项, 第二项为核吸引项。

动能积分为:

$$I^{kinetic} = \int e^{-\alpha |r-A|^2} \left[-\frac{1}{2} \nabla^2 \right] e^{-\gamma |r-C|^2} dr \quad (7.32)$$

根据 7.32 式可计算出动能矩阵 $h_{kinetic}$ 。

核吸引积分为:

$$I^{nuclear_attraction} = \int e^{-\alpha |r-A|^2} e^{-\beta |r-B|^2} \frac{1}{|r-C|} dr \quad (7.33)$$

根据 7.33 式可计算出核吸引矩阵 $h_{nuclear}$ 。

有效势记为 V_{eff} , 包含库仑项和交换项。

是电子 (记为电子 2) 占据某个轨道 χ_j 后, 对我们关注的电子 1 所产生的库仑排斥作用。定义库仑算符来表示空间一点 \vec{x}_i 感受到的某个轨道上的 χ_j 电子的

(平均) 库仑势, 表达式为:

$$\hat{J}_j(\vec{x}_1) = \int \left| \chi_j(\vec{x}_2) \right|^2 \frac{1}{r_{12}} d\vec{x}_2 \quad (7.34)$$

则该点受到的其他 $N-1$ 个电子的总库仑势, 即库仑项为: $\sum_j J_j(\vec{x}_1)$ 。

交换项是交换轨道带来的, 与库仑算符类似, 交换算符表达式为:

$$\hat{K}_j(\vec{x}_1) = \int \chi_j^*(\vec{x}_2) \frac{1}{r_{12}} \chi_j(\vec{x}_2) d\vec{x}_2 \quad (7.35)$$

类似的, 交换势为: $\sum_j K_j(\vec{x}_1)$ 。

计算库仑项和交换项的涉及两个电子的积分, 双电子积分的基本形式:

$$I = \left\langle \chi_1(x_1) \chi_2(x_2) \left| \frac{1}{r_{12}} \right| \chi_1(x_1) \chi_2(x_2) \right\rangle = \int \chi_1^*(x_1) \chi_2^*(x_2) \frac{1}{r_{12}} \chi_1(x_1) \chi_2(x_2) dx_1 dx_2 \quad (7.36)$$

其中 $\chi_1(x_1)$ 和 $\chi_2(x_2)$ 是参与积分的单电子波函数, x_1 和 x_2 是电子坐标, 包含三

个方向的笛卡尔坐标和一个自旋坐标, $\frac{1}{r_{12}}$ 是 SI 单位制下表示两电子间库仑排斥力的双电子算子。

电子排斥积分为:

$$I^{electron_repulsion} = \left\langle ik \left| jl \right. \right\rangle = \iint e^{-\alpha|r_1-A|^2} e^{-\beta|r_1-B|^2} \frac{1}{|r_1-r_2|} e^{-\gamma|r_2-C|^2} e^{-\delta|r_2-D|^2} dr_1 dr_2 \quad (7.37)$$

对于该体系, 我们计算的无对称性的双电子排斥积分形状为(3,3,3,3), 8-fold 双电子排斥积分形状为(21,)。

综合式子 7.34 带 7.37, 库仑势和有效势可表达为:

$$\begin{aligned} v_{jl}^J &= \sum_{ik} \left\langle ik \left| jl \right. \right\rangle D_{ik} \\ v_{il}^K &= \sum_{kj} \left\langle ik \left| jl \right. \right\rangle D_{kj} \end{aligned} \quad (7.38)$$

因此, 最终的有效势 V_{eff} 为:

$$v_{eff} = v^J - \frac{1}{2} v^K \quad (7.39)$$

有效势如下图所示:

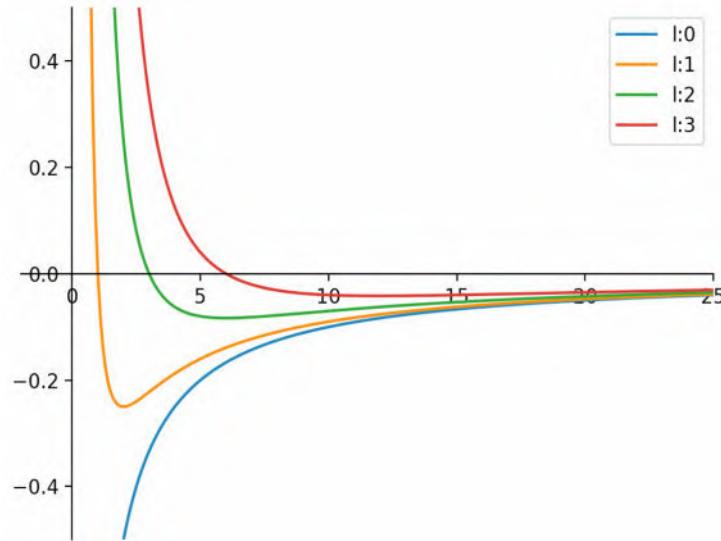


图 7.15 H 原子体系有效势图示

(3) 计算总能和轨道系数

有了密度矩阵 D 和哈密顿量 H 后，通过矩阵对角化即可计算总能。该处的总能称为 Hartree-Fock 能量，等于核排斥部分与电子部分的和。

核排斥能量也称为静态库仑能，表达式为：

$$E_{nuc} = \sum_i \sum_j \text{charges}[i] \frac{1}{rr} \text{charges}[j] \quad (7.40)$$

式中，Charges 是每个原子的电子数目组成的列表， rr 是原子间坐标间隔的矩阵，当只有 1 个原子时，核排除能量 $E_{nuc} = 0$ 。

电子能量为：

$$E_{electron} = \sum_{ij} h_{ij} D_{ji} + \frac{1}{2} \sum_{ijkl} D_{ji} D_{lk} \langle ik | jl \rangle = \sum_{ij} h_{ij} D_{ji} + \frac{1}{2} \sum_{ij} v_{ij}^{eff} D_{ji} \quad (7.41)$$

我们计算出的 H 原子体系核排除能 $e_1 = -0.4936$ ，电子能量 $e_{coul} = 0.1710$ ，总能 $e_{tot} = -0.3226$ 。

对角化过程中同时计算体系的轨道系数矩阵 C ，表达式如下：

$$C = \begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1K} \\ C_{21} & C_{22} & \cdots & C_{2K} \\ \vdots & \vdots & & \vdots \\ C_{K1} & C_{K2} & \cdots & C_{KK} \end{pmatrix} \quad (7.42)$$

轨道系数矩阵 C 与基组可计算出轨道波函数（纯态波函数），即 2.2.1 节中的 2.14 式，其矩阵表达式如下：

$$\begin{bmatrix} \chi_1 \\ \chi_2 \\ \vdots \\ \chi_K \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1K} \\ C_{21} & C_{22} & \cdots & C_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ C_{K1} & C_{K2} & \cdots & C_{KK} \end{bmatrix}^T \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_K \end{bmatrix} \quad (7.43)$$

纯态波函数对应于电子所能占据的轨道，体系总的波函数表达为 N 个粒子的 Slater 行列式的形式，即 2.15 式，矩形形式如下：

$$\Psi(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_i(\vec{x}_1) & \chi_j(\vec{x}_1) & \cdots & \chi_k(\vec{x}_1) \\ \chi_i(\vec{x}_2) & \chi_j(\vec{x}_2) & \cdots & \chi_k(\vec{x}_2) \\ \vdots & \vdots & & \vdots \\ \chi_i(\vec{x}_N) & \chi_j(\vec{x}_N) & \cdots & \chi_k(\vec{x}_N) \end{vmatrix} \quad (7.44)$$

7.2.4.4 精确求解和 Wavelib 库求解结果对比

我们分别实现了基于物理学家推导的直接精确求解方法和基于所开发的 Wavelib 库离散数值解法，计算了 H 原子体系的波函数，结果对比如图 7.16 和图 7.17 所示。

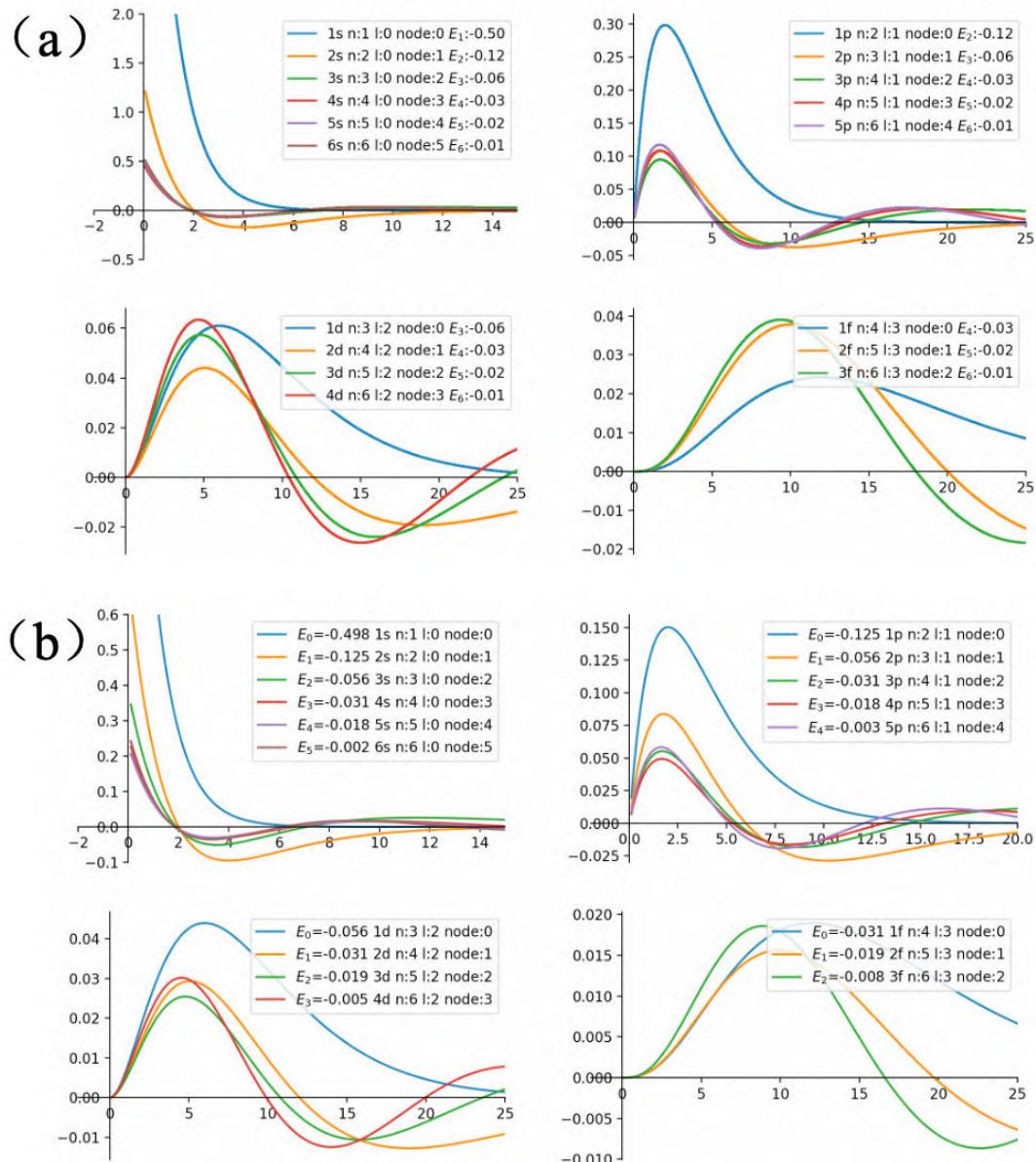


图 7.16 精确解法和 Wavelib 库求解的 H 原子体系的前几个 spdf 轨道波函数 (a) 精确解法

结果 (b) Wavelib 库求解结果

从轨道波函数的对比结果来看, Wavelib 库是有效的。

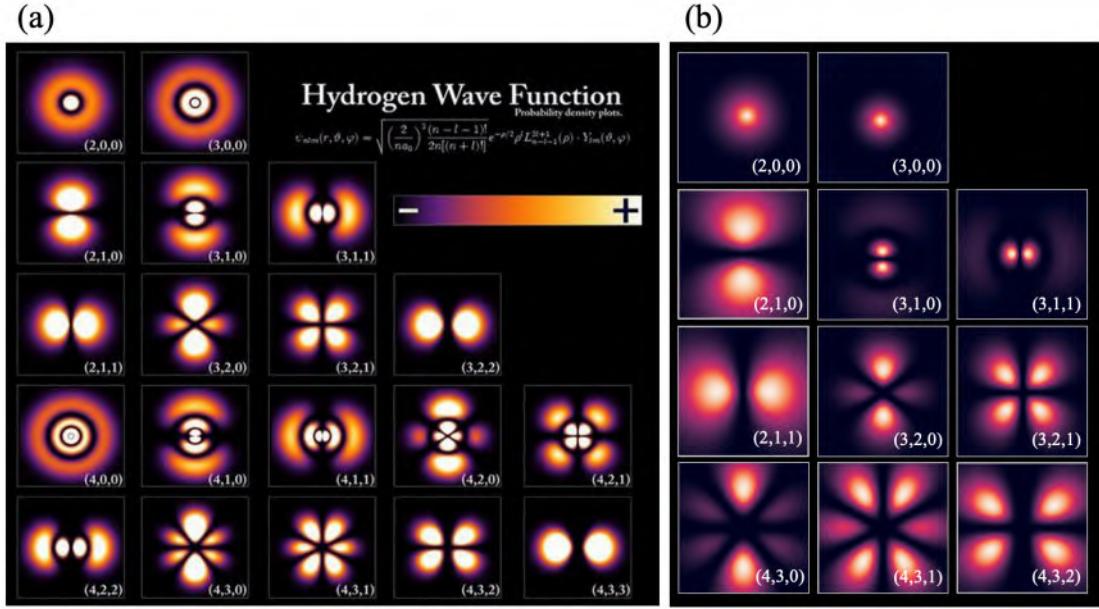


图 7.17 H 原子体系轨道波函数 (a) 直接求解绘制的结果 (b) 本文开发的 Wavelib 库求解的结果

我们绘制了 H 原子体系的轨道波函数的热力图, 如图 7.17 所示, (a) 由直接求解薛定谔方程的结果绘制而成 (b) 由本文开发的 Wavelib 库求解结果绘制而成。轨道波函数图根据电子出现的概率密度绘制, 电子波函数的模方 (即出现的概率) 越大的地方, 越亮。电子概率密度的三维空间的, 展示的图均是其在 (x, y) 平面上的切片。图中可以看出, 本文开发的解算器, 能准确得求解 H 原子体系的波函数。

7.4.3 谐振子体系

上述实验中采用了现有的基组，应用到目标检测任务中，我们需要开发出弹性的，可自动学习到的基组。本文在谐振子体系下做了初步尝试。

谐振子体系也是理想化的能直接精确求解的几个体系之一，一个简单的谐振子体系是在弹性限度内，由真空中自由悬挂在弹簧下的重物与弹簧组成的系统，弹簧的一端固定，重物仅在有限的位移内运动。

谐振子体系中，应力与应变成线性关系，满足胡克定律，即弹簧的弹力 F 和弹簧的长度变化量 x 的关系为：

$$F = -kx \quad (7.45)$$

其中， k 是劲度系数。符号代表弹簧的弹力与 x 的伸长（或压缩）方向相反。

通过变形储存在弹簧中的弹性势能为：

$$U = \frac{1}{2}kx^2 \quad (7.46)$$

势能函数在 U - x 平面内是一段抛物线。随着弹簧沿 x 方向变形（无论是拉伸还是压缩），势能相应增加。

如果将一块质量为 m 的物体悬挂在弹簧的末端，然后对其施加一个轴向扰动（敲打或拉开一段距离突然松手），物体和弹簧组成的系统将会以一个固有角频率（共振角频率）开始振动：

$$\omega = \sqrt{\frac{k}{m}} \quad (7.47)$$

7.4.3.1 经典力学描述下的谐振子体系

经典力学下的谐振子运动方程如下：

$$F = m \frac{d^2x}{dt^2} \quad (7.48)$$

联立式 7.47 和 7.48，得：

$$\frac{d^2x}{dt^2} + \omega^2 x = 0 \quad (7.49)$$

上式是一个常微分方程，解为：

$$x = A \cos(\omega t + \varphi_0) \quad (7.50)$$

其中，积分常数 A 代表振幅， φ_0 代表相位。

速度和加速度均随时间周期变化如下：

$$\begin{aligned} v &= \frac{dx}{dt} = -A\omega \sin(\omega t + \varphi_0) \\ a &= \frac{d^2x}{dt^2} = -A\omega^2 \cos(\omega t + \varphi_0) \end{aligned} \quad (7.51)$$

7.4.3.2 量子力学描述下的谐振子体系

量子力学描述下，简谐系统中粒子的势能项为：

$$V = \frac{1}{2}m\omega^2 x^2 \quad (7.52)$$

代入 2.13 式子的薛定谔方程，推导出精确解的为：

$$\begin{aligned} E_n &= \left(n + \frac{1}{2} \right) \hbar\omega \\ \psi_0(x) &= \left(\frac{m\omega}{\pi\hbar} \right)^{\frac{1}{4}} \exp \left[-\frac{m\omega}{2\hbar} x^2 \right] \\ \psi_n(x) &= \frac{1}{\sqrt{n!}} (\hat{a}_+)^n \psi_0(x) \end{aligned} \quad (7.53)$$

其中，

$$\begin{aligned} \hat{a}_+ \psi_n &= \sqrt{n-1} \psi_{n-1} \\ \hat{a}_- \psi_n &= \sqrt{n} \psi_{n-1} \end{aligned} \quad (7.54)$$

初态可由下式确定：

$$\psi_0 = \sqrt{\Delta x} \left(\frac{m\omega}{\pi\hbar} \right)^{\frac{1}{4}} \exp \left[-\frac{m\omega}{2\hbar} (x - a_0)^2 \right] \quad (7.55)$$

体系总的波函数为：

$$\Psi = \sum_n c_n \psi_n \quad (7.56)$$

其中，系数可由下式确定：

$$c_n = [\psi]_{N \times N} [\psi_0]_N \quad (7.57)$$

计算总能有两种方法，是等效的：

$$E = \sum_n c_n^* c_n E_n$$

$$E = \sum_n \psi_0^* H \psi_0$$
(7.58)

将时间项的解, 即公式 7.5 中第一式包含进来, 可得波函数的动力学变化表达式为:

$$\Psi(x, t) = \sum_n c_n \psi_n(x) \exp\left[-\frac{iE_n t}{\hbar}\right]$$
(7.59)

直接精确求解得到的谐振子体系波函数如下:

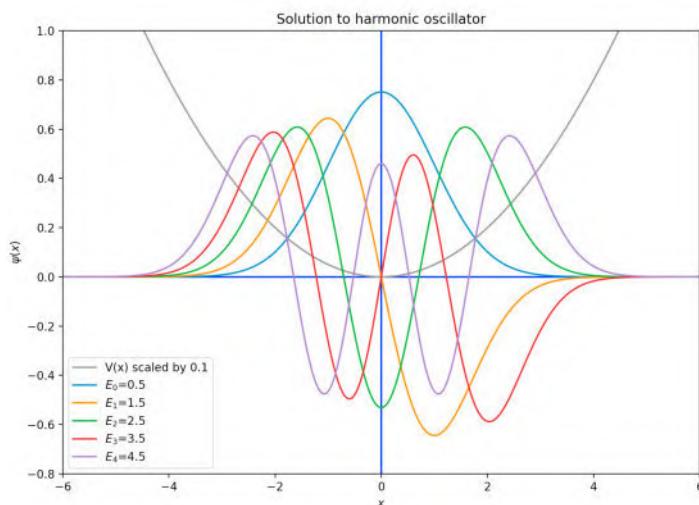


图 7.18 精确求解的谐振子体系波函数

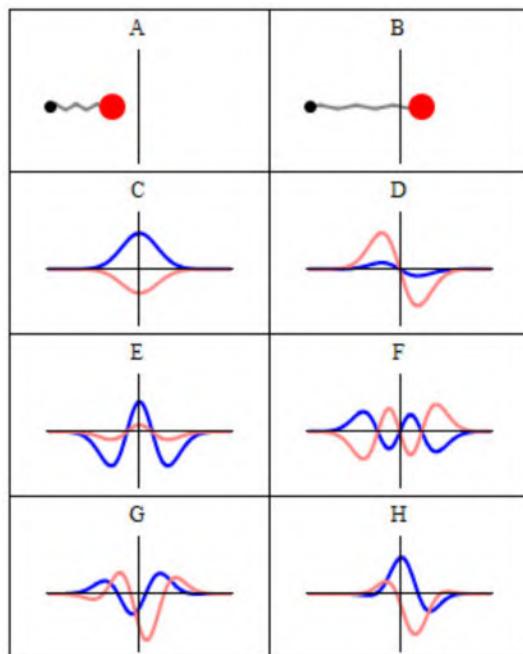


图 7.19 谐振子体系的波函数 (a-b) 谐振子系统 (c-f) 定态波函数 (G-H) 含时波函数

图 7.18 和图 7.19 展示了精确求解的谐振子体系的波函数。基态的波函数可以作为视为高斯基函数的展开。本文探索了 Pytorch 训练高斯基函数展开下的纯态波函数，实现任意项混合高斯函数拟合算法。

7.4.3.3 任意项混合高斯函数拟合算法

(1) 高斯基函数

高斯基函数 (Gaussian Type Orbital, GTO) 是以高斯函数为基组的一种方法，高斯函数可表示为：

$$f(x) = ae^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (7.60)$$

其中， a 是幅值， μ 是均值， σ 是标准差。不同幅值、均值、标准差下的高斯函数图像如图 7.20 所示。

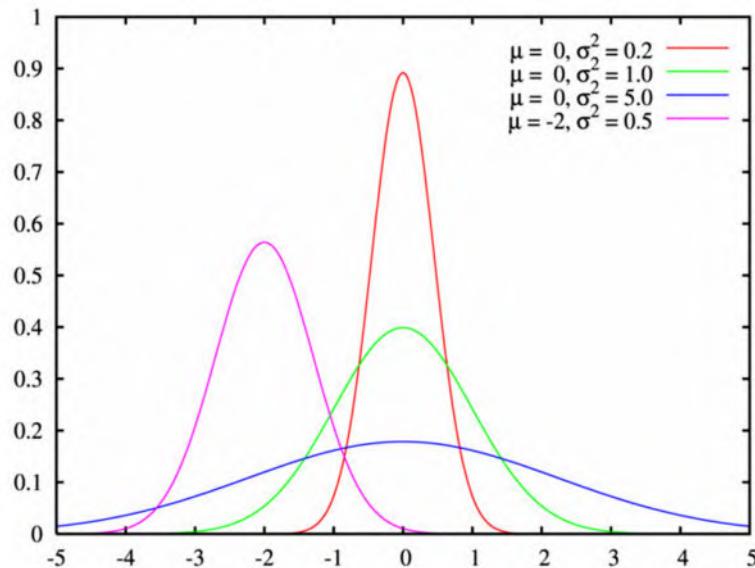


图 7.20 高斯函数图像

图中，半峰全宽 $FWHM = 2\sqrt{2\ln 2}\sigma \approx 2.355\sigma$ ，全空间的高斯积分可以快速计算：

$$\int_{-\infty}^{\infty} ae^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = a\sigma \cdot \sqrt{2\pi} \quad (7.61)$$

在球坐标下，用于拟合原子轨道的 GTO 称为球谐高斯函数 (Spherical

Harmonic Gaussians)，也称为原始高斯函数 (Primitive Gaussian Function, PFG)：

$$\phi^{PGF}(\alpha, n, \ell, m, r) = \tilde{N}(n, \alpha) Y_\ell^m r^\ell e^{-\alpha r^2} \quad (7.62)$$

其中，归一化系数为：

$$\tilde{N}(n, \alpha) = \left[\frac{(2n+2)!\pi^{1/2}}{2^{2n+3}(n+1)!\alpha^{n+3/2}} \right]^{-1/2} \quad (7.63)$$

为了获得更接近真实波函数，实际上会将多个 PGF 加和起来 (加和系数从此固定不变)，构成一个形状更好收缩高斯函数 (Contracted Gaussian Function, CGF)：

$$\phi^{CGF}(\alpha, n, \ell, m, r) = \sum_{k=1}^K D_k \tilde{N}(n, \alpha) Y_\ell^m r^\ell e^{-\alpha r^2} \quad (7.64)$$

确定原子轨道后，分子轨道可根据公式 7.43 展开为原子轨道的线性组合。

(2) 拟合实验

设置数值解法的矩阵尺寸为 2014 个，选取前 6 个轨道拟合 1s-6s 轨道，每个轨道都是一系列 x 的值对应于 y 的值，不同的轨道指数系数 α 相同，但收缩系数 D_k 不同。

定义收缩高斯函数模型，设置可训练参数为指数 α 和收缩系数 D_k ，进行一个高斯函数的拟合：输入的 $x: (1000,)$ 0 到 10，间隔 0.01。真值 y 为: $(1000,)$ 对应于 x 的高斯函数的值。优化算法采用 SGD 或 Adam。

训练过程中，出现梯度爆炸问题，改进方法是真值和预测值都要进行归一化。当只有 1 个高斯函数时，因为进行了归一化，收缩系数对函数无影响。

采用矩阵方法快速实现收缩高斯函数的最后一下，即 r 列向量叉乘 α 行向量，表达式为：

$$\begin{bmatrix} r_1^2 \\ r_2^2 \\ \vdots \\ r_n^2 \end{bmatrix} \begin{bmatrix} -\alpha_1 & -\alpha_2 & \cdots & -\alpha_k \end{bmatrix} = \begin{bmatrix} -\alpha_1 r_1^2 & -\alpha_2 r_1^2 & \cdots & -\alpha_k r_1^2 \\ -\alpha_1 r_2^2 & -\alpha_2 r_2^2 & \cdots & -\alpha_k r_2^2 \\ \vdots & \vdots & & \vdots \\ -\alpha_1 r_n^2 & -\alpha_2 r_n^2 & \cdots & -\alpha_k r_n^2 \end{bmatrix} \quad (7.65)$$

对每个元素取 e 指数，而后乘以行向量 $D[D1, D2, \dots, Dm]$ ，对每行求和。

训练中出现问题：谐振子的质量和圆频率均取 1 时，本征波函数的取值大约

是-4~4，输入的 x 是[-10, 10]均匀取 2014 个值，由于输入的 x 绝对值远大于 1，而波函数经过了归一化，幅值很小，导致拟合高斯函数过程中梯度消失($1e-30 \sim 1e-120$ 次方)，从而训练能跑，但权重并不更新。

解决方案是：自动获取本征波函数 x 的取值范围，缩放到-1~1。暂时只实现了[-10, 10]的标准化，缩放到[-1.8, 1.8]。

训练中出现的问题 2 是：高斯函数对关于 y 轴对称的波函数拟合效果较好，但无法拟合关于原点对称的波函数。如所图 7.21 示。

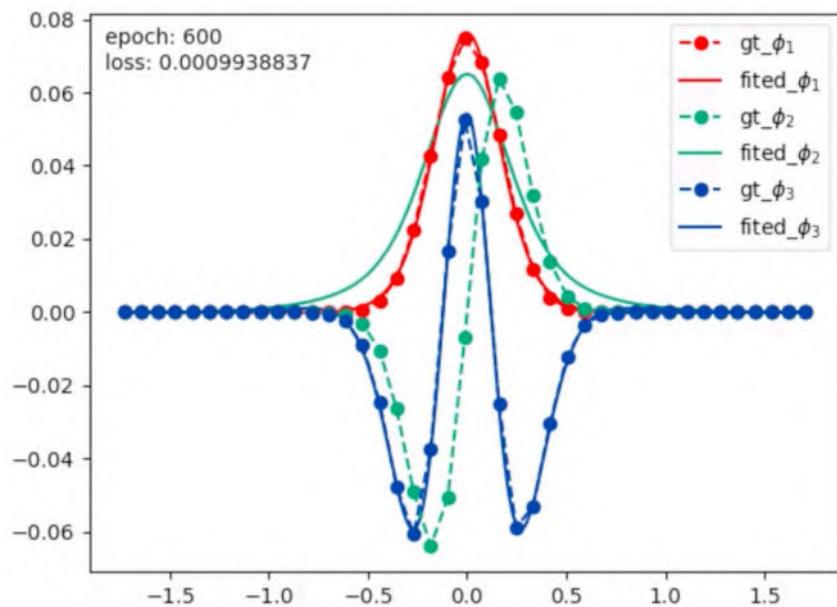


图 7.21 单中心多轨道高斯函数拟合波函数

图中，红色、绿色、蓝色点虚线分别是离散法求得的前三个本征波函数，相应实线分别是高斯函数拟合得到的波函数曲线。可以看出，第一个波函数（红色点虚线）和第三个波函数（蓝色点虚线）关于 y 轴对称，拟合得的高斯函数（红色实线和蓝色实线）与离散解高度重合，拟合效果较好。但拟合得的第二个波函数（绿色实线）和离散解波函数（绿色点虚线）相差很大。总损失是 $1e-3$ 数量级。分析原因：第二个波函数关于原点对称，为奇函数，而高斯函数均为偶函数，多个中心均为原点的高斯函数收缩成的 函数也必定为偶函数。

解决方案是：中心不重合的两个高斯函数可以收缩成 1 个偶函数。定义第三种可学习的参数：高斯中心 R 。高斯中心与高斯指数系数一一对应，只设置一半高斯中心为 0 且不可学习（向下取整），其余高斯中心设置为可学习参数。

经过改进后，7.65 式变为：

$$\begin{bmatrix} (r_1 - R_1)^2 & (r_1 - R_2)^2 & \cdots & (r_1 - R_k)^2 \\ (r_2 - R_1)^2 & (r_2 - R_2)^2 & \cdots & (r_2 - R_k)^2 \\ \vdots & \vdots & \ddots & \vdots \\ (r_n - R_1)^2 & (r_n - R_2)^2 & \cdots & (r_n - R_k)^2 \end{bmatrix} \begin{bmatrix} -\alpha_1 \\ -\alpha_2 \\ \vdots \\ -\alpha_k \end{bmatrix} = \begin{bmatrix} -\alpha_1(r_1 - R_1)^2 & -\alpha_2(r_1 - R_2)^2 & \cdots & -\alpha_k(r_1 - R_k)^2 \\ -\alpha_1(r_2 - R_1)^2 & -\alpha_2(r_2 - R_2)^2 & \cdots & -\alpha_k(r_2 - R_k)^2 \\ \vdots & \vdots & \ddots & \vdots \\ -\alpha_1(r_n - R_1)^2 & -\alpha_2(r_n - R_2)^2 & \cdots & -\alpha_k(r_n - R_k)^2 \end{bmatrix} \quad (7.66)$$

改进后的训练结果如图 7.22 所示。

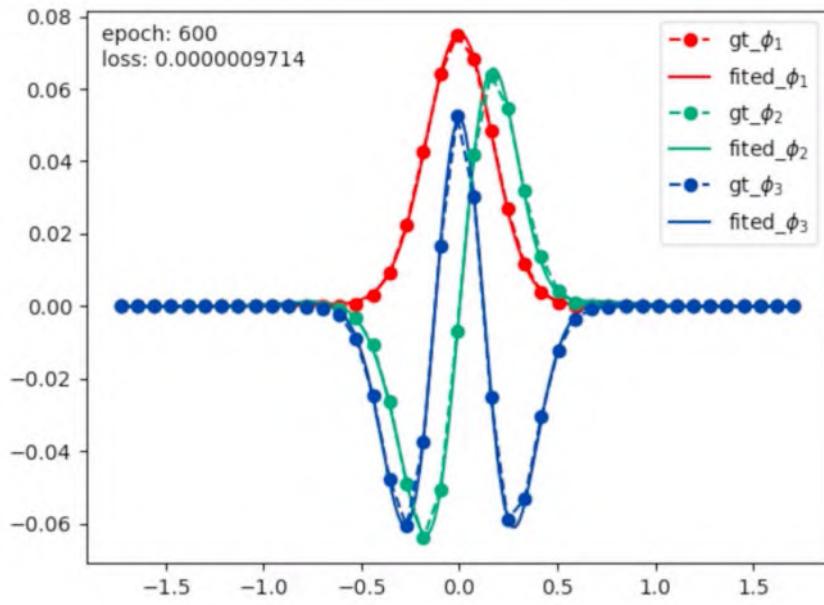


图 7.22 多中心多轨道高斯函数拟合波函数

如上图所示，改进后，对于关于原点对称的第二个波函数（绿色），高斯函数拟合结果（实线）也能很好地与真值（点虚线）重合，总损失小于 $1e-6$ 。

(3) 收缩系数数目选取

收缩系数数目应为被拟合的波函数（基组）数目 N 的 $2N+1$ 倍。原理：第 1、3、5 波函数关于 y 轴对称，极值点有 1、3、5 个；第 2、4、6 波函数关于原点对称，极值点有 2、4、6 个。第奇数个波函数对于高斯函数数目没有要求，第偶数个波函数要求至少有极值点数目个中心不重合的高斯函数数目。即要拟合前 N 个波函数，至少要求 N 个中心不重合的高斯函数。考虑算法中共有收缩系数数目的一半个（向下取整）中心重合的高斯函数，再考虑多加 1 个中心不重合高斯函数，因此，设置 $2N+1$ 个收缩系数数目，对应于设置 N 个中心重合的高斯函数和 $N+1$ 个中心不重合的高斯函数。如下式：

$$\begin{bmatrix} D_{11}e^{-\alpha_1(r_1-R_1)^2} & D_{21}e^{-\alpha_2(r_1-R_2)^2} & \dots & D_{k1}e^{-\alpha_k(r_1-R_k)^2} \\ D_{11}e^{-\alpha_1(r_2-R_1)^2} & D_{21}e^{-\alpha_2(r_2-R_2)^2} & \dots & D_{k1}e^{-\alpha_k(r_2-R_k)^2} \\ \vdots & \vdots & & \vdots \\ D_{11}e^{-\alpha_1(r_n-R_1)^2} & D_{21}e^{-\alpha_2(r_n-R_2)^2} & \dots & D_{k1}e^{-\alpha_k(r_n-R_k)^2} \end{bmatrix} \quad (7.67)$$

上式矩阵为 (n, k) , 对每行求和, 得到 $(n,)$ 就是第 1 个收缩系数对应的基组。同理对 m 个进行操作, 得到 (m, n) , 即所有的基组。

7.4.4 人体关键点组成的多粒子系统

7.4.4.1 任务描述

本文基于上述经过验证的波函数求解方法和基组构建方法, 初步探索了波函数在人体关键点检测任务中的应用方法。

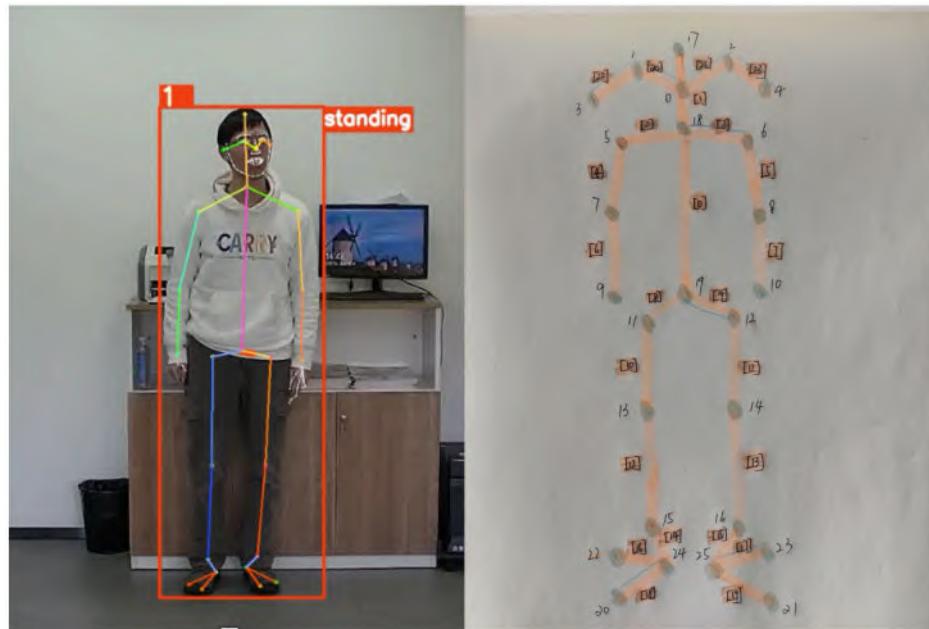


图 7.23 人体关键点检测和骨架抽象模型

如图 7.23 所示, 人体关键点检测的任务是通过单帧图像自动检测出人体的骨架关键点, 抽象模型如右图所示, 人体骨架可以简化为 26 个关键的关节连接而成。当前已经有 OpenPose, AlphaPose 等骨架检测的深度神经网络算法可以实现该功能, 但在遮挡情况下, 关键点会丢失。我们希望采用波函数编码, 而后通过量子演化, 然后解码 (量子测量) 的方式补全未检测到或丢失的关键点信息。

7.4.4.2 实现思路

将人体视为由 26 个粒子组成的物理系统, 每个粒子即一个关节 (关键点)。类比于原子核-电子系统, 该系统不存在原子核, 粒子运动在其他粒子构成的势场中, 哈密顿量仅包含动能和相互作用势能两项:

$$H = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} \quad (7.68)$$

用开发的 Wavelib 库求解这个哈密顿算符下的薛定谔方程, 可得到粒子的波

函数和本征值。

N 个粒子体系的波函数的形式可以表达为 Slater 行列式的形式, 即公式 2.15。Slater 行列式具有的性质有: 交换两个粒子, 相当于交换两行, 行列式正负变号; 当粒子占据同一轨道式, 相当于有两列相等, 行列式为 0。

Slater 行列式内的矩阵元是纯态波函数 (即分子轨道波函数), 而纯态波函数由基组 (即所有原子轨道波函数) 的线性组合构成, 也就是公式 7.43。

在 7.68 式的哈密顿量下, 由于存在粒子之间的相互作用, 薛定谔方程变得相当复杂, 我们做一些简化。

根据变分原理 (Variation Principle), 最接近真实体系的波函数应给出最低的能量, 有:

$$E_0 = \langle \Psi_0 | H | \Psi_0 \rangle \quad (7.69)$$

采用拉格朗日乘数法, 对函数 $L = E_0[\{\chi_i\}] - \sum_{a,b}^N \varepsilon_{ab} (\langle \chi_a | \chi_b \rangle - \delta_{ab})$ 变分求极

值, 可将多体薛定谔方程转化为单体薛定谔方程 (Hartree-Fock 方程):

$$f(i)\chi(\vec{x}_i) = \varepsilon\chi(\vec{x}_i) \quad (7.70)$$

其中, $f(i)$ 是单粒子算符, 即 Fock 算符, 形式为:

$$f(i) = -\frac{1}{2}\nabla_i^2 + v^{HF}(i) \quad (7.71)$$

其中, 第一项是动能项, 第二项具有复杂的形式, 但与 H 原子体系的 7.39 式类似包含库仑项和交换性, 代表第 i 个粒子所受到的所有其它粒子的势场的影响。

近似的精髓是, 将复杂的多体粒子问题转化为简单的单粒子问题。

第二项的势场来自“其它粒子”, 故其值本身又依赖于“其它粒子”的轨道解。即, Fock 算符的获取其实又依赖于其自身的本征值 ε (然而 ε 的值一开始并不知道)。所以, HF 方程是非线性的, 通常需要迭代求解。

迭代求解的方法称为迭代求解方法为自洽场方法, 来自于自洽场理论 (Self-Consistent-Field, SCF)。流程为:

(1) 初猜一套轨道 χ , 从而可以构造如上第二项的势场, 进而构造 Fock 算

符。

(2) 以初猜的 Fock 算符求解 HF 方程, 得到一套新的本征函数(即轨道 χ)。

(3) 采用新的轨道构造新的平均势场及 Fock 算符, 即回到步骤 1。

(4) 重复以上步骤知道体系达到收敛(consistency)。

收敛的判定标准, 如: 前后两步的轨道几乎不再有区别, 或本征值的差极小。

求解的波函数用公式 7.64 展开, 称为基组展开原理。基组展开原理如下:

HF 方程原则上无穷多个解 (无穷多个轨道), 实际中只能取有限个轨道, 称为基函数 (basis function) 或基组(basis set)。假设有 K 个空间轨道基组, 则有 $2K$ 个自旋轨道基组。以这些基组去解 N 粒子体系的 HF 方程, 得到一系列正交轨道, 其中能量最低的 N 个自旋轨道是充满的, 以这 N 个轨道构成的 Slater 行列式代表了基态波函数; 其余 $2K-N$ 个为空轨道。

基函数越多、越完备, 体系的计算结果就越精确, 能量也越低 (越接近精确值)。

基组展开用原子轨道线性组合 (Linear Combination of Atomic Orbitals, LCAO) 方法, 即采用我们在 7.4.3.3 节中实现的专用基函数 (任意项混合高斯基函数) 展开。

7.4.4.3 波函数与关键点检测算法融合

基于上述思想, 我们开发了相应的求解算法, 并设计的图 7.24 的双空间模型实现波函数和神经网络构成的双空间模型, 来预测的关键点。

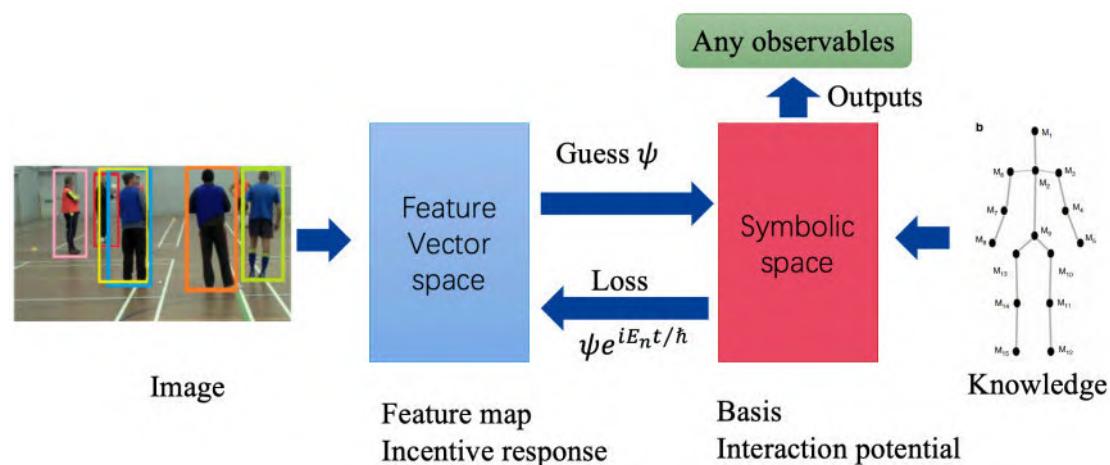


图 7.24 双空间模型

如图 7.24 所示, 我们设计的双空间模型的左侧是神经网络, 包含图像输入

到检测网络、跟踪网络和 AlphaPose 关键点检测网络，检测的结果组成了特征空间，即由由刺激到相应的特征图空间。右侧是波函数知识要素，设定体系具有的粒子数 $N=26$ ，由基组和相互作用势能组成符号空间。特征空间预测的关键点，用来猜测波函数，即对波函数进行编码，从而与符号空间通信。符号空间内的波函数作为作为一种损失函数，为特征空间提供额外的损失项，称为物质湮灭损失，即关键点从已检测到忽然漏检时，提供物质湮灭损失很大的损失值，当关键点检测保持原来的状态（一直检测到或一直不存在时），物质湮灭损失为 0。另外，符号空间内的波函数经过含时演化后，在有限时间内的波函数是可信的，用于提供演化的损失。最终的输出结果从符号空间中根据波函数到可观测量的计算得到，即公式 7.1。

7.4.4.4 结果

我们实现了一套基于上述方法的双空间人体关键点检测算法，并在开放的数据集中进行训练。实验结果表明，在单张 GTX3080 显卡上，训练一代的时间约 3 天多，训练完需要一年多，没有完成最终的实验。

分析原因，一是硬件算力不足；二是波函数求解过程中没有经过优化，只采用了基本的离散数值求解方法，这也是后来本文提出需要研究基于快速积分、分布式并行和即时编译的高性能解算器的原因；三是在设计上，该思路采用了“经典”的量子力学求解方法，默认采用了自治场方法和重整化群，在宏观系统上，该思路可能需要在量子计算机发展起来以后才有望实现，这是后来本文提出采用“形式理论”方案的主要原因，即直接利用量子力学的数学表达形式，利用量子启发的方案完成波函数与神经网络的融合。

7.4.5 密集遮挡下的跟踪算法

在上述工作的基础上，我们采用了“形式理论”，即量子力学启发的方案，设计了物质波神经网络（Matter Wave Network, MWNet），用于频繁遮挡场景的目标跟踪。

当前基于检测的跟踪的主要问题之一方法是目标湮灭问题，即目标在其中丢失或由目标探测器的错误导致的更多帧、尤其是在恶劣天气或遮挡场景中。在这项工作中，为了提高跟踪的鲁棒性，提出了基于量子理论的物质波网络（MWNET）。目标通过量子测量将检测结果编码到物质波中，以利用复值的特征。通过量子演化从湮灭中再生出目标，以修复检测错误。网络可以改进前后框架之间的信息流以及跟踪轨迹的连续性。最后，在 MOT20 测试集实现了 73.8% 的 MOTA 和 75.1% 的 IDF1 分数。我们的工作证明了量子演化在重建信息方面的有效性，并且该模型可以在频繁遮挡场景中实现对多个目标的连续跟踪。

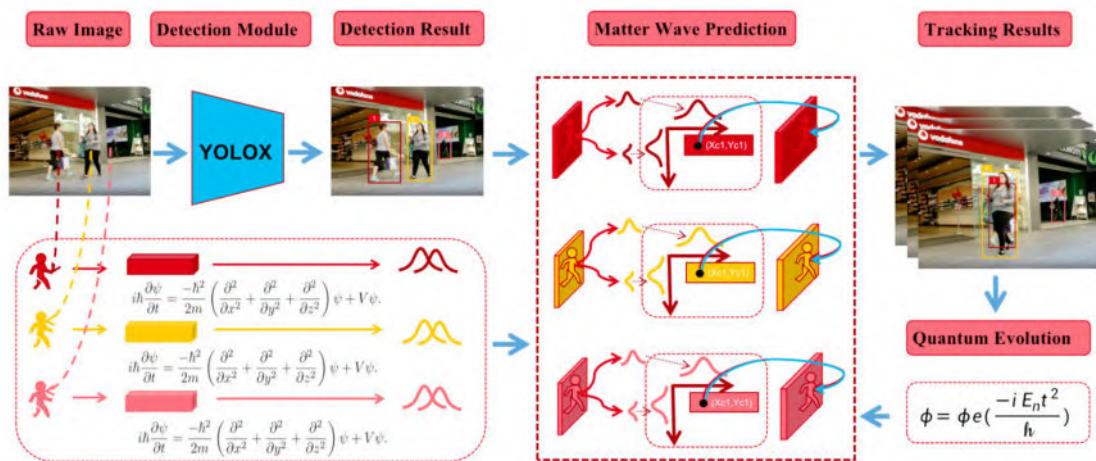


图 7.25 MWNet 模型架构

我们设计的 MWNet 模型架构如图 7.25 所示，通过目标检测模块将其分为多条路径，每条路径表示不同目标的物质波函数的计算。多条路径计算通过共享同一事物，在不同的时间拥有不同的基组。同时，将目标的外观特征与物质波相关联，实现多目标跟踪。



图 7.26 MWNet 在 MOT20 数据集上检测结果可视化

MWNet 在 MOT20 数据集上可视化检测结果如图 7.26 所示，在密集遮挡场景下，实现了性能较好的跟踪效果。

检测性能如所示，结果显示，MWNet 有较好的跟踪效果，特别是 ID switch 指标有明显的提升。

表 7.1 MWNet 性能与其他模型对比

Method	MOTA \uparrow	IDF1 \uparrow	MT \uparrow	ML \downarrow	IDs \downarrow
CorrTracker[25]	65.2	69.1	66.4	8.9	5,183
JDMOTGNN[24]	67.1	67.5	53.1	13.2	3,131
FairMOT[23]	61.8	67.3	68.8	7.6	5,243
SORT[11]	42.7	45.1	16.7	26.2	4,470
LPCMOT[26]	56.3	62.5	34.1	25.2	1,562
MWNNet(ours)	73.8	75.1	69.4	7.1	1,046

7.5 总结

本章描述了我们把量子力学的波函数引入神经网络中,发展下一代人工智能的初步探索。

我们梳理了量子力学的基础知识,对计算理论和方法进行了推广,并完成了离散求解算法的实现,设计了量子力学和人工智能结合的方法。

我们先在无限深势阱、H 原子体系和谐振子体系上进行了验证,结果显示我们的求解方法是准确的。

而后,我们将计算方法推广到人工智能领域,与深度算法结合进行探索研究。在人体关键点检测的实验中,设计了双空间模型,并跑通了训练,但由于训练速度问题,未完成实验。

最后,我们改进方案,采用量子力学启发的思想,设计了基于物质波的跟踪算法,用于密集遮挡下的跟踪,对原有的算法进行了改进,获得了很好的效果。

最近(2021-2022 年),如 1.4.1 节所述 DeepMind 发布了基于神经网络的新的泛函 Ferminet,在精度和时间尺度上实现了 SOTA,并开启了以前难以处理的在多电子系统中直接准确地优化波函数的可能性,马上被应用于多个发表于 *Science*、*Nature* 子刊等期刊的重要研究中。英国开放大学的 Gkoumas 等人也受量子启发,借鉴量子测量、量子演化和量子纠缠的概念构建多模态融合模型,并将其应用于自然语言处理的多模情感分类任务中,实现了 SOTA。虽然我们是独立提出的这种思想,但已经看到这些团队走在了“前头”,他们的方案也给予了我们很大的启发。虽然不清楚真正的第三代人工智能什么时候到来,但我们相信未来几年内在这块量子理论和人工智能交叉的“处女地”上一定会产出重要的成果,推动新一轮的科技革命。

第8章 软件、平台和库

本章简要介绍作者在博士后期间设计和实现的软件、算法平台 XAI 和 Damei Python 库。

8.1 多模态融合标注软件

为了满足算法实现和项目的需求，我们基于 `labelme` 开发了多模态融合半自动标注软件 `mmlabelme`，这种一项共性技术。可实现多种模态数据的联合标注，实现目标类别、目标跟踪 ID、目标状态等目标信息的标注，可实现矩形、多边形、点、线、线带、圆等目标类型的标注，拥有继承功能，并且集成了检测、跟踪和状态分类等 AI 算法，可实现半自动标注。

开发完成的软件界面如图 8.1 所示。

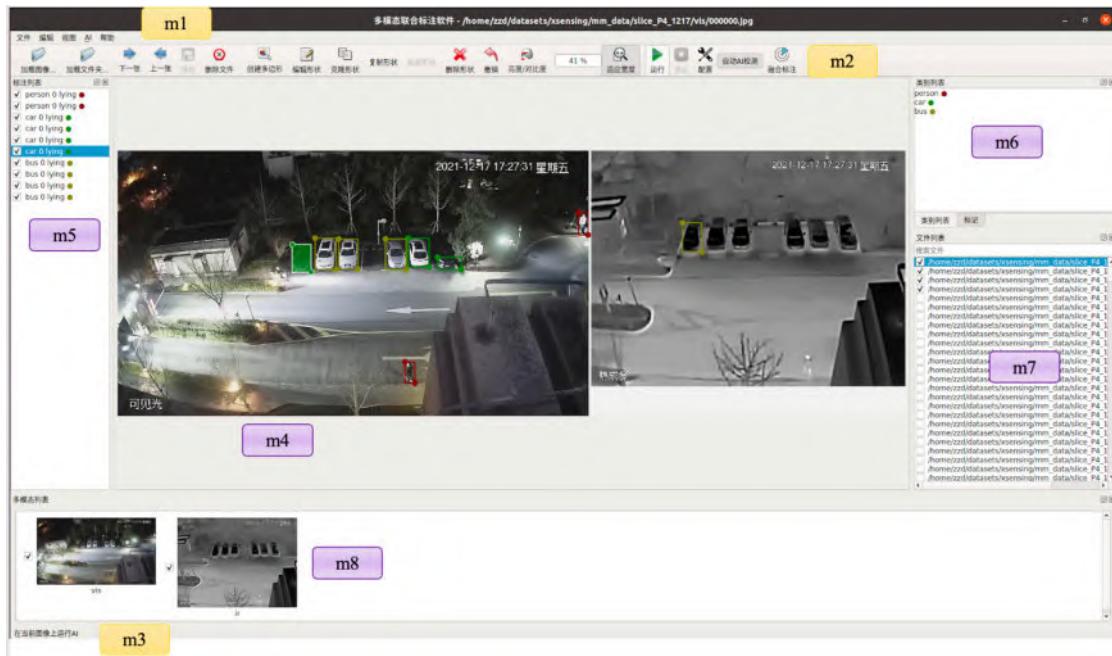


图 8.1 `mmlableme` 多模态融合标注软件

如上图所示，`mmlabelme` 分为 8 个部分，`m1` 到 `m8` 分别是菜单栏、工具栏、状态栏、画布、标注坞、类别坞和文件坞。

菜单栏包含文件、编辑、视图、AI 和帮助菜单，提供全部功能。

工具栏是常用菜单栏的快捷图标，包含加载图像、加载文件夹、上一张、下一张、保存、删除文件、创建多边形、编辑形状、克隆形状、复制形状、粘贴形

状、删除形状、撤销、亮度/对比度、适应窗口、AI 算法的运行、停止、配置、自动运行和多模态融合标注等功能。

状态栏用于显示程序状态或进行操作提示。

画布展示一个条目的数据，对于多模态数据，自动适应单/多模态的选择和显示，并且提供与键盘、鼠标动态交互的功能。

标注坞用于展示当前标注的目标、类型、跟踪 ID 和状态等具体信息。

类别坞用于展示去重的标注类别，用于快速选择。

文件坞用于展示所有的数据条目，开头打钩表示已标注的数据，并提供快速搜索、切换数据条目等功能。

8.2 人体行为分析算法软件

人体行为分析软件是为了工程应用所开发的算法程序，并进行了商业部署。



图 8.2 智慧医疗人体行为分析算法

图 8.2 展示了我们开发的用于智慧医疗的人体行为分析算法，其包括目标检测算法、多目标跟踪算法、人体关键点检测算法和人体行为分类算法，实现了端到端的检测，在 GTX 3090 服务器上进行了商业化部署，可实现实时的检测。算法程序与应用前端通过 Flask 通信，对原有的智慧医疗系统进行了赋能。



图 8.3 智能装备人体行为分析算法

图 8.3 展示了我们开发的针对智能状态的人体行为分析算法。为了在有限的算力上实现实时和鲁棒的检测，我们对模型进行的优化，采用了目标检测和基于光流的人体行为分类算法，实现端到端检测和分析，有效提升了速度。

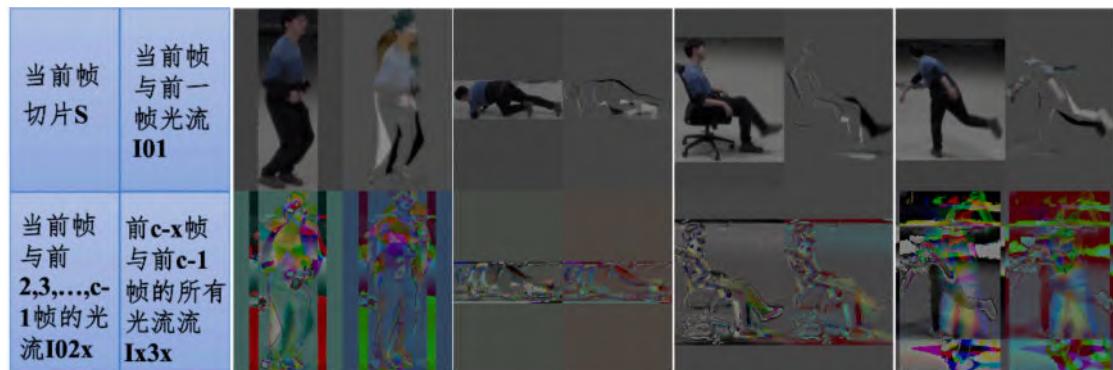


图 8.4 光流特征

我们设计的全新的光流特征如图 8.4 所示。特征中不仅包含了当前帧目标切片（纹理、姿态、类型等信息），而且包含了当前帧与前 $c-1$ 帧的（例如 c 为 5）光流信息，这些信息携带了可分离的状态信息，提升了分类器针对更多类别的分类性能。

该算法实现了 14 种人体行为的检测：行走、跳跃、跑步、坐下、躺下、趴下、打架、带来、安装/卸载、带走、遗留、投掷、持枪和射击，目前在有限的数据集上实现了 99% 的 F1 分数，在 Jetson NX 上速度为 12 FPS。

8.3 XAI 算法平台

Xsensing AI 算法平台：XAI-platform，简称 XAI，为了将来敏捷开发、快速部署的需求而设计的一个柔性可扩展的人工智能算法平台。XAI 向后管理算法，向前提供核心的 AI 算法能力，为后续人工智能应用提供强大的后台支撑。面向模块开发者，提供开发文档，统一接口；面向使用者，提供 SDK 和跨语言调用。

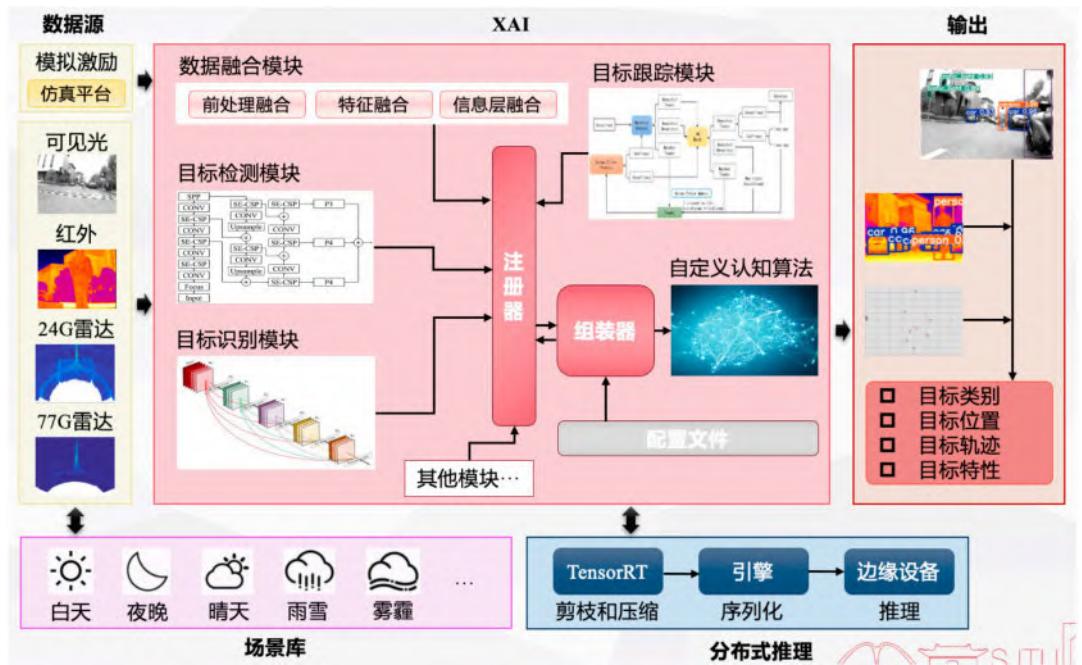


图 8.5 XAI 算法平台设计

XAI 算法平台的架构设计如图 8.5 所示，支持白天、夜晚、晴天、雨雪、雾霾等多种场景下的多模模块数据输入，例如红外、可见光、雷达和模拟激励数据。算法平台通过模块注册和组装实现对算法模块的管理，可实现分布式的推理部署，便捷的自定义算法流程和运行。

算法平台采用算法流、管道和算法模块的设计理念，算法每个对数据池有输出的算法组合称为一个算法流，每个算法流 Stream 输入一种或多种特定类型的数据，每个算法流 Stream 包含至少 1 个管道 Pipeline，每个管道 Pipeline 由输入模块 Input+算法模块 Module+输出模块 Exporter 顺序连接而成，一个流包含多个算法时，前一个管道的输出模块是后一个管道的输入模块。每个算法模块 Module 提供特定的功能，内部包含数据前处理 Pre-process、算法模型 Model 和数据后处理 Post-process 等。

算法平台有如下特性：

- 提供可见光、红外、雷达三个模态的相关 ai 算法的统一调用接口，柔性可扩展。
- 提供算法模型的推理、训练、评估、可视化等功能，可实现算法模型自学习和自演进。
- 提供算法模块处理流程自定义、参数配置自定义。
- 提供 XAI 作为 Python 库使用的 SDK 接口。
- 提供 XAI 作为微服务部署到中心平台的 gRPC 接口，可实现分布式部署，支持跨系统、跨语言的调用。

8.4 Python 库：Damei

Damei 是一个人工智能和量子力学交叉 Python 库，是作者在博士后期间将各种设计神经网络的通用函数、类进行封装，提供便捷实现的功能型库。其中，包含了第 7 章用到的求解薛定谔方程、波函数投影、演化等的库 Wavelib。

使用 Damei 库，仅需 pip install damei 即可。功能如下图：

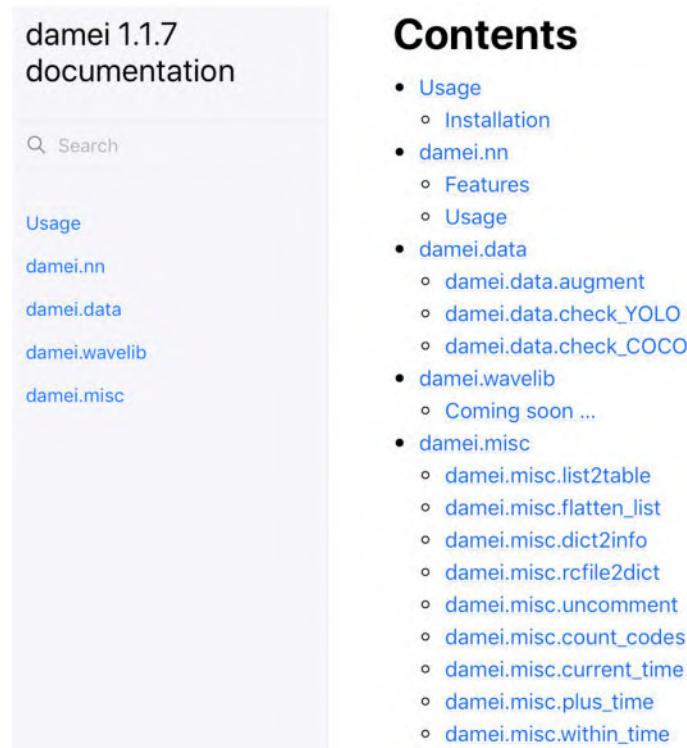


图 8.6 Damei Python Library 接口文档

Damei Python 库的接口文档（部分）如图 8.6 所示，目前文档中包含神经网络模块（nn）、数据相关模块（data）、波函数库（wavelib）和其他杂项（misc），还有一些功能正在测试中，文档将不断完善。

开源部分代码见 Github 仓库，<https://github.com/zhangzhengde0225/damei>，详细的接口文档链接见仓库的 Readme.md。

第9章 总结与展望

9.1 工作总结

本文进行了深度学习算法在目标检测领域的应用研究与前沿探索。

一条主线是抓住人工智能深度学习算法在目标检测领域的应用点，在实际的应用过程中，提出数据增强和网络改进的相关方法，进行相关的研究工作。另一条主线是基于对当代人工智能的认识和数理基础，在连接主义和符号主义融合的思想下，提出以波函数作为知识要素的一种第三代人工智能范式，进行相关理论推广和探索工作。第一条主线是第二条主线提供基础，第二条主线为改进第一条提供指导，二者相辅相成。

在应用研究方面，本文针对复杂的环境条件、有限的计算能力、小样本数据、极低的空间分辨率等问题，提出了基于通道注意力机制 SE 的方法改进网络，提出了负样本训练方法（NST）提升精度，提出了感兴趣区域（ROI）算法提高检测速度，提出了滑动接收场短时向量记忆方法（SSVM）提高精度，提出了合成雾算法并进行了速度改进使模型适应雾天，提出了混合像元方法从无人机图像中模拟卫星遥感图像，提出了基于白平衡、Gamma、Tone、Contrast 和锐化等多级滤波的自适应网络以提高检测所需的空间分辨率，提出了基于超分辨率算法和高反差保留的算法提升图像质量，提出了基于动态背景、随机尺寸、随机位置、随机角度的离线数据集增广算法大大降低了大量数据的获取难度。该条主线的相关研究不仅为具体的任务场景提供解决方案，而且所提出的算法、思想能应用于相关的其他领域中。

在前沿探索方面，本文针对不完全信息、可解释性差、鲁棒性取决于数据集等问题，提出了波函数作为知识要素的知识-数据协同驱动的一种第三代人工智能实现方法。对量子力学基本理论和第一性原理方法进行了推广，提出了新的技术路线是：神经网络泛函方法、高性能解算器和协同驱动神经网络。本文的方案是在不断尝试中发展的，在无限深势阱体系、H 原子体系和振子体系验证了离散数值解法和任意项混合高斯函数拟合算法的准确性和有效性；在人体关键点检测的应用中，设计了双空间模型用于融合知识要素和神经网络，由于采用了自洽场和重整化群方案，在当前算力下无法有效训练而以失败告终；在密集遮挡场景

下跟踪算法的应用中，我们采用量子启发和形式理论的方案，设计了由波函数编码、量子演化和波函数解码（量子测量）为核心的跟踪算法改进方案，有效提升了模型性能，初步证明了量子力学的波函数作为一种先验知识的表达，确实一些任务中可能发挥作用。该条主线是独辟蹊径的一种尝试，未来有可能发展为新的人工智能范式。

此外，在研究的过程中，本文发布了用于车载摄像头下斑马线检测的数据集 CDNet，开源了相关代码[76]；发布了用于无人机视角下绝缘子检测的数据集 SFID，开源的测试基准[99]；开发了多模态半自动标注软件 mmlabelme；开发了 2 款赋能产业的人体行为分析算法软件；开发了柔性可扩展的人工智能算法管理和应用平台 XAI；开发和开源了量子力学和人工智能交叉的 Python 库 Damei[72]。

9.2 工作展望

未来的工作将主要围绕波函数作为知识要素的知识-数据协同驱动的下一代人工智能算法展开。目前已有世界顶尖的团队在进行相关的研究工作，紧跟科研前沿，争取做出重要成果。

但正如 Arikan 教授所说，绝妙的主意往往是海市蜃楼，需要在接下来的 2-3 年沉下心来，确保这个 idea 是真实的。在这个过程中，团队的合作、前沿的跟踪、个人的努力和良好的环境都很重要。

重大的发现很少，人的一生中能得到一个，就是幸运的。为了平衡这一点，致力于基础研究的同时，也需要研究一些小问题，以确保会有结果，正应本文标题中的“应用研究”和“前沿探索”两条主线。

望未来，能做到既“仰望星空”，又“脚踏实地”，乘风破浪，一路向前。

参考文献

1. McCarthy, J., M.L. Minsky, and C.E. Shannon, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence* [J]. *Ai Magazine*, 1955, **27**(4): 12-14. DOI.
2. Rosenblatt, F., *The Perception - A Probabilistic Model for Information-Storage and Organization in the Brain* [J]. *Psychological Review*, 1958, **65**(6): 386-408. DOI: <https://doi.org/10.1037/h0042519>.
3. Zhang, Y.D., L.N. Wu, and S.H. Wang, *Survey on Development of Expert System* [J]. *Computer Engineering and Applications*, 2010, **46**(19): 43-47. DOI: <https://doi.org/10.3778/j.issn.1002-8331.2010.19.012>.
4. Hinton, G.E., S. Osindero, and Y.W. Teh, *A Fast Learning Algorithm for Deep Belief Nets* [J]. *Neural Computation*, 2006, **18**(7): 1527-1554. DOI: <https://doi.org/10.1162/neco.2006.18.7.1527>.
5. Wei, H.X., M. Tu, and X.P. Zhang, *Deep Learning and Image Recognition Principle and Practice*. 1st ed. 2019, Beijing: China Machine Press. 1-2.
6. Google. *End to end Open Source Machine Learning Platform*. 2022. Access date. Available from: <https://tensorflow.google.cn>.
7. Google. *JAX: Composable Transformations of Python+Numpy Programs*. 2022. Access date. Available from: <https://github.com/google/jax>.
8. Community. *Pytorch*. 2022. Access date: [6-15]. Available from: <https://pytorch.org>.
9. Baidu. *Paddlepaddle*. 2022. Access date: [6-15]. Available from: <https://www.paddlepaddle.org.cn>.
10. He, K., et al. *Deep Residual Learning for Image Recognition* [C]. in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016. Las Vegas, NV, USA. DOI: <https://doi.org/10.1109/CVPR.2016.90>.
11. Li, X., et al., Generalized Focal Loss: Learning Qualified and Distributed Bounding Boxes for Dense Object Detection. *ArXiv*, 2020. DOI: <https://arxiv.org/abs/2006.04388>.
12. Wang, J.Q., et al. *Seesaw Loss for Long-Tailed Instance Segmentation* [C]. in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021. DOI.
13. Janocha, K. and W.M. Czarnecki, *On Loss Functions for Deep Neural Networks in Classification* [J]. *Schedae Informaticae*, 2017, **25**(1): 45-49. DOI: <https://doi.org/10.4467/20838476SI.16.004.6185>.
14. Lipton, Z.C., J. Berkowitz, and C. Elkan, A Critical Review of Recurrent Neural Networks for Sequence Learning. *ArXiv*, 2015. 1-1. DOI: <https://arxiv.org/abs/1506.00019>.
15. Derevyanko, G., et al., *Deep Convolutional Networks for Quality Assessment of Protein Folds* [J]. *Bioinformatics*, 2018, **34**(23): 4046-4053. DOI: <https://doi.org/10.1093/bioinformatics/bty494>.
16. Girshick, R., et al. *Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation* [C]. in *IEEE Conference on Computer Vision and*

- Pattern Recognition (CVPR)*. 2014. Columbus, OH, USA. DOI: <https://doi.org/10.1109/cvpr.2014.81>.
17. Girshick, R. *Fast R-CNN* [C]. in *IEEE International Conference on Computer Vision (ICCV)*. 2015. Santiago, Chile. DOI: <https://doi.org/10.1109/iccv.2015.169>.
18. Ren, S.Q., et al., *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks* [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **39**(6): 1137-1149. DOI: <https://doi.org/10.1109/tpami.2016.2577031>.
19. He, K.M., et al. *Mask R-CNN* [C]. in *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017. Venice, Italy. DOI: <https://doi.org/10.1109/iccv.2017.322>.
20. Liu, W., et al. *SSD: Single Shot MultiBox Detector* [C]. in *European Conference on Computer Vision (ECCV)*. 2016. Amsterdam, Netherlands: Springer. DOI: https://doi.org/10.1007/978-3-319-46448-0_2.
21. Redmon, J., et al. *You Only Look Once: Unified, Real-Time Object Detection* [C]. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016. Las Vegas, NV, USA: IEEE. DOI: <https://doi.org/10.1109/cvpr.2016.91>.
22. Redmon, J. and A. Farhadi, *YOLOv3: An Incremental Improvement*. ArXiv, 2018. 1-1. DOI.
23. Bochkovskiy, A., C.Y. Wang, and H.Y.M. Liao, *YOLOv4: Optimal Speed and Accuracy of Object Detection*. ArXiv, 2020. 1-1. DOI: <https://arxiv.org/abs/2004.10934v1>.
24. Ultralytics. *YOLOv5*. 2020. Access date: [2022-02-01]. Available from: <https://github.com/ultralytics/yolov5>.
25. Ge, Z., et al., *YOLOX: Exceeding YOLO Series in 2021*. ArXiv, 2021. 1-1. DOI: <https://arxiv.org/abs/2107.08430>.
26. Liu, Z., et al., *Swin Transformer: Hierarchical Vision Transformer using Shifted Windows*. ArXiv, 2021. 1-1. DOI: <https://arxiv.org/abs/2103.14030>.
27. Zhang, Z.D., et al., *CDNet: A Real-Time and Robust Crosswalk Detection Network on Jetson Nano Based on YOLOv5* [J]. *Neural Computing & Applications*, 2022, **1**(1): 1-1. DOI: <https://doi.org/10.1007/s00521-022-07007-9>.
28. Zhang, B., J. Zhu, and H. Su, *Towards the Third Generation of Artificial Intelligence* [J]. *SCIENTIA SINICA Informationis*, 2020, **50**(9): 1281-1302. DOI: <https://doi.org/10.1360/SSI-2020-0204>.
29. Huang, S., et al. *Combo-Action: Training Agent For FPS Game with Auxiliary Tasks* [C]. in *the AAAI Conference on Artificial Intelligence (AAAI)*. 2019. DOI: <https://doi.org/10.1609/aaai.v33i01.3301954>.
30. Song, S.H., et al. *Playing FPS Games With Environment-Aware Hierarchical Reinforcement Learning* [C]. in *the International Joint Conference on Artificial Intelligence (IJCAI)*. 2019. DOI: <https://doi.org/10.24963/ijcai.2019/482>.
31. Shi, J.X., S.Y. Sun, and J. Zhu, *A Spectral Approach to Gradient Estimation for*

- Implicit Distributions*, in the International Conference on Machine Learning (ICML). 2018. p. 4633-4653.
32. Zhou, Y., J.X. Shi, and J. Zhu, *Nonparametric Score Estimators*, in the International Conference on Machine Learning (ICML). 2020. p. 1-1.
33. Wang, Z.Y., et al., *Function Space Particle Optimization for Bayesian Neural Networks*, in the International Conference on Learning Representations (ICLR). 2019. p. 1-12.
34. Chen, J.F., et al., *VFlow: More Expressive Generative Flows with Variational Data Augmentation*, in the International Conference on Machine Learning (ICML). 2020. p. 1-1.
35. Gallego, V. and D.R. Insua, *Variationally Inferred Sampling through a Refined Bound* [J]. Entropy, 2021, **23**(1): 1-1. DOI: <https://doi.org/10.3390/e23010123>.
36. Kresse, G. and J. Furthmuller, *Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set* [J]. Physical Review B, 1996, **54**(16): 11169-11186. DOI: <https://doi.org/10.1103/PhysRevB.54.11169>.
37. Perdew, J.P., K. Burke, and M. Ernzerhof, *Generalized Gradient Approximation Made Simple* [J]. Physical Review Letters, 1996, **77**(18): 3865-3868. DOI: <https://doi.org/10.1103/PhysRevLett.77.3865>.
38. Lehtola, S., et al., *Recent Developments in LIBXC - A Comprehensive Library of Functionals for Density Functional Theory* [J]. Softwarex, 2018, **7**(1): 1-5. DOI: <https://doi.org/10.1016/j.softx.2017.11.002>.
39. Hyldgaard, P., Y. Jiao, and V. Shukla, *Screening Nature of The van der Waals Density Functional Method: A Review and Analysis of The Many-Body Physics Foundation* [J]. Journal of Physics-Condensed Matter, 2020, **32**(39): 1-1. DOI: <https://doi.org/10.1088/1361-648X/ab8250>.
40. Pfau, D., et al., *Ab Initio Solution of the Many-electron Schrodinger Equation with Deep Neural Networks* [J]. Physical Review Research, 2020, **2**(3): 1. DOI: <https://doi.org/10.1103/PhysRevResearch.2.033429>.
41. Atz, K., F. Grisoni, and G. Schneider, *Geometric Deep Learning on Molecular Representations* [J]. Nature Machine Intelligence, 2021, **3**(12): 1023-1032. DOI: <https://doi.org/10.1038/s42256-021-00418-8>.
42. Li, X., et al., *Fermionic Neural Network with Effective Core Potential* [J]. Physical Review Research, 2022, **4**(1): 1. DOI: <https://doi.org/10.1103/PhysRevResearch.4.013021>.
43. Westermayr, J. and P. Marquetand, *Machine Learning for Electronically Excited States of Molecules* [J]. Chemical Reviews, 2021, **121**(16): 9873-9926. DOI: <https://doi.org/10.1021/acs.chemrev.0c00749>.
44. Li, L., et al., *Kohn-Sham Equations as Regularizer: Building Prior Knowledge into Machine-Learned Physics* [J]. Physical Review Letters, 2021, **126**(3). DOI: <https://doi.org/10.1103/PhysRevLett.126.036401>.
45. Adams, C., et al., *Variational Monte Carlo Calculations of A <= 4 Nuclei with an Artificial Neural-Network Correlator Ansatz* [J]. Physical Review Letters, 2021, **127**(2). DOI: <https://doi.org/10.1103/PhysRevLett.127.022502>.
46. Kirkpatrick, J., et al., *Pushing the Frontiers of Density Functionals by Solving*

- the Fractional Electron Problem* [J]. *Science*, 2021, **374**(6573): 1385-+. DOI: <https://doi.org/10.1126/science.abj6511>.
47. Unke, O.T., et al., *SpookyNet: Learning Force Fields with Electronic Degrees of Freedom and Nonlocal Effects* [J]. *Nature Communications*, 2021, **12**(1): 1. DOI: <https://doi.org/10.1038/s41467-021-27504-0>.
48. Smidstrup, S., et al., *QuantumATK: An Integrated Platform of Electronic and Atomic-Scale Modelling Tools* [J]. *Journal of Physics-Condensed Matter*, 2020, **32**(1): 1-1. DOI: <https://doi.org/10.1088/1361-648X/ab4007>.
49. Blaha, P., et al., *WIEN2k: An APW+lo Program for Calculating the Properties of Solids* [J]. *Journal of Chemical Physics*, 2020, **152**(7): 1-1. DOI: <https://doi.org/10.1063/1.5143061>.
50. Sun, Q.M., *Libcint: An Efficient General Integral Library for Gaussian Basis Functions* [J]. *Journal of Computational Chemistry*, 2015, **36**(22): 1664-1671. DOI: <https://doi.org/10.1002/jcc.23981>.
51. Sun, Q.M., et al., *PYSCF: The Python-based Simulations of Chemistry Framework* [J]. *Wiley Interdisciplinary Reviews-Computational Molecular Science*, 2018, **8**(1): 1-1. DOI: <https://doi.org/10.1002/wcms.1340>.
52. Sun, Q.M., et al., *Recent Developments in The PySCF Program Package* [J]. *Journal of Chemical Physics*, 2020, **153**(2): 1-1. DOI: <https://doi.org/10.1063/5.0006074>.
53. Bogdanov, N.A., et al., *Enhancement of Superexchange Due to Synergetic Breathing and Hopping in Corner-sharing Cuprates* [J]. *Nature Physics*, 2022, **18**(2): 190. DOI: <https://doi.org/10.1038/s41567-021-01439-1>.
54. Hermann, J., Z. Schatzle, and F. Noe, *Deep-neural-network Solution of The Electronic Schrodinger Equation* [J]. *Nature Chemistry*, 2020, **12**(10): 891. DOI: <https://doi.org/10.1038/s41557-020-0544-y>.
55. Meister, R., S.C. Benjamin, and E.T. Campbell, *Tailoring Term Truncations for Electronic Structure Calculations Using a Linear Combination of Unitaries* [J]. *Quantum*, 2022, **6**(1): 1-1. DOI: <https://doi.org/10.22331/q-2022-02-02-637>.
56. Gunther, J.M., et al., *Improving Readout in Quantum Simulations with Repetition Codes* [J]. *Quantum Science and Technology*, 2022, **7**(1): 1-1. DOI: <https://doi.org/10.1088/2058-9565/ac3386>.
57. Nigam, J., M.J. Willatt, and M. Ceriotti, *Equivariant Representations for Molecular Hamiltonians and N-center Atomic-scale Properties* [J]. *Journal of Chemical Physics*, 2022, **156**(1): 1-1. DOI: <https://doi.org/10.1063/5.0072784>.
58. Stoliaroff, A., S. Jobic, and C. Latouche, *PyDEF 2.0: An Easy to Use Post-Treatment Software for Publishable Charts Featuring a Graphical User Interface* [J]. *Journal of Computational Chemistry*, 2018, **39**(26): 2251-2261. DOI: <https://doi.org/10.1002/jcc.25543>.
59. Kundu, S., et al., *PASTA: Python Algorithms for Searching Transition stAtes* [J]. *Computer Physics Communications*, 2018, **233**(1): 261-268. DOI: <https://doi.org/10.1016/j.cpc.2018.06.026>.
60. Zhang, Z.D., et al., *VaspCZ: An Efficient VASP Computation Assistant Program* [J]. *Nuclear Techniques*, 2020, **43**(3): 30501-30501. DOI: <https://doi.org/10.1088/1674-1015/ab9301>.

- <https://doi.org/10.11889/j.0253-3219.2020.hjs.43.030501>.
61. Gkoumas, D., et al. *An Entanglement-driven Fusion Neural Network for Video Sentiment Analysis* [C]. in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*. 2021. Montreal: International Joint Conferences on Artificial Intelligence Organization. DOI: <https://doi.org/10.24963/ijcai.2021/239>.
 62. Gkoumas, D., et al., Quantum Cognitively Motivated Decision Fusion for Video Sentiment Analysis. ArXiv, 2021. 827-835. DOI.
 63. Li, Q.C., et al. *Quantum-inspired Neural Network for Conversational Emotion Recognition* [C]. in *AAAI Conference on Artificial Intelligence*. 2021. Electr Network. DOI.
 64. Liu, Y., et al. *What Does Your Smile Mean? Jointly Detecting Multi-Modal Sarcasm and Sentiment Using Quantum Probability* [C]. 2021. Punta Cana, Dominican Republic: Association for Computational Linguistics. DOI: <https://doi.org/10.18653/v1/2021.findings-emnlp.74>.
 65. Zhang, Y.Z., et al., *A Quantum-like Multimodal Network Framework for Modeling Interaction Dynamics in Multiparty Conversational Sentiment Analysis* [J]. *Information Fusion*, 2020, **62**(1): 14-31. DOI: <https://doi.org/10.1016/j.inffus.2020.04.003>.
 66. Wang, C., et al. *CSPNet: A New Backbone that can Enhance Learning Capability of CNN* [C]. in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2020. Seattle, WA, USA. DOI: <https://doi.org/10.1109/CVPRW50498.2020.00203>.
 67. He, K., et al., *Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition* [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, **37**(9): 1904-1916. DOI: <https://doi.org/10.1109/TPAMI.2015.2389824>.
 68. Liu, S., et al. *Path Aggregation Network for Instance Segmentation* [C]. in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018. Salt Lake City, UT, USA: IEEE. DOI: <https://doi.org/10.1109/CVPR.2018.00913>.
 69. Hu, J., et al., Squeeze-and-Excitation Networks. ArXiv, 2019. 1-1. DOI: <https://arxiv.org/abs/1709.01507>.
 70. He, K.M., J. Sun, and X.O. Tang, *Single Image Haze Removal Using Dark Channel Prior* [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(12): 2341-2353. DOI: <https://doi.org/10.1109/tpami.2010.168>.
 71. Buslaev, A., et al., *Albumentations: Fast and Flexible Image Augmentations* [J]. *Information*, 2020, **11**(2): 1-1. DOI: <https://doi.org/10.3390/info11020125>.
 72. Zhang, Z.D. *Damei Python Library*. 2022. Access date: [2022-06-15]. Available from: <https://github.com/zhangzhengde0225/damei>.
 73. Daiel, A.R. and S. Y., *The Principles of Deep Learning Theory*. 1st ed. An Effective Theory Approach to Understanding Neural Networks. 2021, arXiv: arXiv.
 74. Stehman, S.V., *Selecting and Interpreting Measures of Thematic Classification*

- Accuracy [J]. Remote Sensing of Environment, 1997, **62**(1): 77-89. DOI: [https://doi.org/10.1016/s0034-4257\(97\)00083-7](https://doi.org/10.1016/s0034-4257(97)00083-7).
75. Chicco, D. and G. Jurman, *The Advantages of the Matthews Correlation Coefficient (MCC) over F1 Score and Accuracy in Binary Classification Evaluation* [J]. Bmc Genomics, 2020, **21**(1): 1-1. DOI: <https://doi.org/10.1186/s12864-019-6413-7>.
76. Zhang, Z.D. *The Dataset, Demo and Source Codes of Crosswalk Detection Network (CDNet)*. 2022. Access date: [2022-02-25]. Available from: <https://github.com/zhangzhengde0225/CDNet>.
77. Se, S. *Zebra-crossing Detection for the Partially Sighted* [C]. in *2000 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2000. DOI: <https://doi.org/10.1109/CVPR.2000.854787>.
78. Uddin, M.S. and T. Shioyama, *Detection of pedestrian crossing using bipolarity feature-an image-based technique* [J]. IEEE Transactions on Intelligent Transportation Systems, 2005, **6**(4): 439-445. DOI: <https://doi.org/10.1109/TITS.2005.858787>.
79. Akinlar, C. and C. Topal. *Edlines: Real-time Line Segment Detection by Edge Drawing* [C]. in *2011 IEEE International Conference on Image Processing*. 2011. DOI: <https://doi.org/10.1109/ICIP.2011.6116138>.
80. Mascetti, S., et al., *ZebraRecognizer: Pedestrian Crossing Recognition for People with Visual Impairment or Blindness* [J]. Pattern Recognition, 2016, **60**(1): 405-419. DOI: <https://doi.org/10.1016/j.patcog.2016.05.002>.
81. Huang, X. and Q. Lin, *An Improved Method of Zebra Crossing Detection Based on Bipolarity* [J]. Computer Applications and Software, 2017, **34**(12): 202-205. DOI: <https://doi.org/10.3969/j.issn.1000-386x.2017.12.038>.
82. Chen, N., F. Hong, and B. Bai, *Zebra Crossing Recognition Method Based on Edge Feature and Hough Transform* [J]. Journal of Zhejiang University of Science and Technology, 2019, **31**(06): 476-483. DOI: <https://doi.org/10.3969/j.issn.1671-8798.2019.06.008>.
83. Chen, K., et al. *MMDetection: Open MMLab Detection Toolbox and Benchmark*. 2019. Access date: [2021-10-10]. Available from: <https://github.com/open-mmlab/mmdetection>.
84. He, K.M., et al. *Mask R-CNN* [C]. in *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017. DOI: <https://doi.org/10.1109/iccv.2017.322>.
85. Ultralytics. *YOLOv5*. 2020. Access date: [2021-02-01]. Available from: <https://github.com/ultralytics/yolov5/tree/v4.0>.
86. NVIDIA. *TensorRT Open Source Software*. 2019. Access date: [2021-10-10]. Available from: <https://github.com/NVIDIA/TensorRT>.
87. Yang, L., et al., *A Review on State-of-the-Art Power Line Inspection Techniques* [J]. IEEE Transactions on Instrumentation and Measurement, 2020, **69**(12): 9350-9365. DOI: <https://doi.org/10.1109/tim.2020.3031194>.
88. Liu, C.Y., et al., *Insulator Faults Detection in Aerial Images from High-Voltage Transmission Lines Based on Deep Learning Model* [J]. Applied Sciences-Basel, 2021, **11**(10): 1-1. DOI: <https://doi.org/10.3390/app11104647>.

89. Ma, Y.P., et al., *Real-Time Detection and Spatial Localization of Insulators for UAV Inspection Based on Binocular Stereo Vision* [J]. *Remote Sensing*, 2021, **13**(2): 1-1. DOI: <https://doi.org/10.3390/rs13020230>.
90. Li, Y.C., et al., *A Method for Autonomous Navigation and Positioning of UAV Based on Electric Field Array Detection* [J]. *Sensors*, 2021, **21**(4): 1-1. DOI: <https://doi.org/10.3390/s21041146>.
91. Yin, J.G., et al., *Edge Detection of High-Voltage Porcelain Insulators in Infrared Image Using Dual Parity Morphological Gradients* [J]. *IEEE Access*, 2019, **7**(1): 32728-32734. DOI: <https://doi.org/10.1109/access.2019.2900658>.
92. Nguyen, V.N., R. Jenssen, and D. Roverso, *Automatic Autonomous Vision-based Power Line Inspection: A Review of Current Status and the Potential Role of Deep Learning* [J]. *International Journal of Electrical Power & Energy Systems*, 2018, **99**(1): 107-120. DOI: <https://doi.org/10.1016/j.ijepes.2017.12.016>.
93. LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning* [J]. *Nature*, 2015, **521**(7553): 436-444. DOI: <https://doi.org/10.1038/nature14539>.
94. He, W.Q., J. Li, and W.H. Zhao, *Insulator Anomaly Detection Based on an Improved Faster RCNN* [J]. *Zhejiang Electric Power*, 2021, **40**(8): 40-46. DOI: <https://doi.org/10.19585/j.zjdl.202108006>.
95. Hu, X. and Y.N. Li, *Hybrid Model Insulator Fault Detection Based on Faster R-CNN and U-net* [J]. *Video Engineering*, 2021, **45**(5): 125-130. DOI: <https://doi.org/10.16280/j.videoe.2021.05.034>.
96. Wang, Q. and B.S. Yi, *Insulator Defect Recognition in Aerial Images Based on Gaussian YOLOv3* [J]. *Laser & Optoelectronics Progress*, 2021, **58**(12): 254-260. DOI: <https://doi.org/CNKI:SUN:JGDJ.0.2021-12-025>.
97. Wang, Z.H. *Insulator Data Set - Chinese Power Line Insulator Dataset (CPLID)*. 2018. Access date: [2021-09-14]. Available from: <https://github.com/InsulatorData/InsulatorDataSet>.
98. Andrel, V.S., et al. *Unifying Public Datasets for Insulator Detection and Fault Classification in Electrical Power Lines*. 2020. Access date: [2021-09-14]. Available from: <https://github.com/heitorcfelix/public-insulator-datasets>.
99. Zhang, Z.D. *FINet*. 2021. Access date: [2022-2-25]. Available from: <https://github.com/zhangzhengde0225/FINet>.
100. Gu, Z.C., et al. *Railway Insulator Defect Detection with Deep Convolutional Neural Networks* [C]. in *12th International Conference on Digital Image Processing (ICDIP)*. 2020. Electr Network: Proceedings of SPIE. DOI: <https://doi.org/10.1117/12.2572918>.
101. Tao, X., et al., *Detection of Power Line Insulator Defects Using Aerial Images Analyzed With Convolutional Neural Networks* [J]. *IEEE Transactions on Systems Man Cybernetics-Systems*, 2020, **50**(4): 1486-1498. DOI: <https://doi.org/10.1109/tsmc.2018.2871750>.
102. Gao, Z.S., et al., *Novel Feature Fusion Module-Based Detector for Small Insulator Defect Detection* [J]. *IEEE Sensors Journal*, 2021, **21**(15): 16807-16814. DOI: <https://doi.org/10.1109/jsen.2021.3073422>.

103. Shi, C.X. and Y.P. Huang, *Cap-Count Guided Weakly Supervised Insulator Cap Missing Detection in Aerial Images* [J]. IEEE Sensors Journal, 2021, **21**(1): 685-691. DOI: <https://doi.org/10.1109/jsen.2020.3012780>.
104. Ling, Z.N., et al., *An Accurate and Real-time Method of Self-blast Glass Insulator Location Based on Faster R-CNN and U-net with Aerial Images* [J]. Csee Journal of Power and Energy Systems, 2019, **5**(4): 474-482. DOI: <https://doi.org/10.17775/cseejpes.2019.00460>.
105. Chen, S.Q., et al., *Learning Slimming SAR Ship Object Detector Through Network Pruning and Knowledge Distillation* [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, **14**(1): 1267-1282. DOI: <https://doi.org/10.1109/jstars.2020.3041783>.
106. Zheng, Z.Z., *Research on Insulator Super-resolution Sub-pixel Target Reconstruction Technology Based on Satellite Time Series Images*. 2021, University of Electronic Science and Technology of China: Hunan, China. p. 1-152. DOI.
107. Xu, B., et al., Empirical Evaluation of Rectified Activations in Convolutional Network. ArXiv, 2015. 1-1. DOI: <https://arxiv.org/abs/10.48550/arXiv.1505.00853>.
108. Hu, Y.M., et al., *Exposure: A White-Box Photo Post-Processing Framework* [J]. Acm Transactions on Graphics, 2018, **37**(2): 1-1. DOI: <https://doi.org/10.1145/3181974>.
109. Polesel, A., G. Ramponi, and V.J. Mathews, *Image Enhancement via Adaptive Unsharp Masking* [J]. IEEE Transactions on Image Processing, 2000, **9**(3): 505-510. DOI: <https://doi.org/10.1109/83.826787>.
110. 吴功平, 肖晓晖, and 申瑞刚, *高压输电导线的损伤检测与故障诊断* [J]. 仪器仪表学报, 2006(10): 1204-1208. DOI: <https://doi.org/10.19650/j.cnki.cjsi.2006.10.008>.
111. 王万国, et al., *基于无人机图像的输电线断股与异物缺陷检测方法* [J]. 计算机应用, 2015, **35**(08): 2404-2408. DOI: <https://doi.org/10.11772/j.issn.1001-9081.2015.08.2404>.
112. 齐国顺, et al., *基于图像处理技术的飞机巡线导线断股缺陷检测方法* [J]. 黑龙江电力, 2017, **39**(06): 522-526. DOI: <https://doi.org/10.13625/j.cnki.hljep.2017.06.011>.
113. 黄新波, et al., *基于径向基概率神经网络的输电导线缺陷状态识别* [J]. 电力系统自动化, 2020, **44**(03): 201-210. DOI: <https://doi.org/10.7500/AEPS20190122004>.

114. 孙日高, et al., 基于图像区域分割的输电导线可疑区域定位方法 [J]. 化工自动化及仪表, 2021, **48**(01): 29-33. DOI: <https://doi.org/10.3969/j.issn.1000-3932.2021.01.006>.
115. 李飞伟, et al., 基于无人机的配电导线断股缺陷检测系统设计与实现 [J]. 测绘地理信息, 2021, **46**(04): 132-135. DOI: <https://doi.org/10.14188/j.2095-6045.2020579>.
116. Zhang, Y.L., et al. *Residual Dense Network for Image Super-Resolution* [C]. in *2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018. Salt Lake City, UT, USA. DOI: <https://doi.org/10.1109/cvpr.2018.00262>.

致 谢

回首过去的两年，对我来说意义非凡，成功经历了从粒子物理与原子核物理到深度学习专业的转变。在研究报告结尾之际，必须感谢那些一路走来给予我帮我的朋友们和家人们。

首先，感谢我的合作导师郁文贤教授。郁老师和蔼的性格、渊博的知识和极具前瞻性的眼光对我产生了深远的影响。特别是，郁老师投身国防科技建设 20 余载，科研成果硕果累累，一手打造了北斗产业西虹桥基地，是全国“最美退役军人”，这种对产学研事业孜孜不倦的精神，是我今后工作的榜样。在此，对郁老师致以诚挚的谢意！

感谢裴凌教授在研究上给予的帮助。裴老师严谨的治学态度和认真的科学作风，使我受益匪浅。特别是，在具体的研究方向上给予的鼓励和帮助，正是在与裴老师的沟通中，才逐步萌发量子力学与人工智能交叉的研究思路，在裴老师的鼓励中，才进行了理论推广和一些验证工作，是我未来前进的方向。在此，对裴老师致以衷心的感谢！

感谢李东瀛副教授在工作上给予的帮助。特别是在智能装备 XAI 算法平台的设计、实现等方面专业的指导，这使得我既能有机会看到顶层的设计，又能进行核心算法的研发，对我今后的发展多有益处。

感谢刘海春老师的工作上的指导，特别是生活上的关心和未来发展的规划。

感谢实验室龚玲、潘常春、邹丹平、张增辉、陶卫、吕娜等老师的帮助和指导。感谢储磊博士、郭炜炜师兄诸多信息的沟通。感谢徐成、吴书宇、张博、周嘉琪、李季蹊、王奕人、李涛、王若晨、吴奇、夏宋、李琦等同学平日的支持和帮助。

感谢北航吕广宏、乐小云、赵润东老师在交叉研究方面提供的指导。

感谢黄真、兰志才、王朝锟、赵耀、戴晟、朱玲玲、宗忆陈、崔学智、康健、卢敏、李扬、李岚臻、孙德君、朱逸嘉等同事在工作上的帮助。

感谢郁宏伟、万松、江婷婷、朱亚夫、施秋蓉、李慧、展山山等同事的帮助，特别是良好的办公环境。

此外，感谢我的家人一直以来的对我生活、学业和工作上的包容、理解和支持的。感谢父母的养育之恩，没有父母的爱护和教诲，我不可能从云南的偏远边

境小镇走到现在有望为国家贡献力量。感谢丈母爹、丈母娘，没有他们的大力帮助和支持，我应该本科毕业后就止步了，不会探索科研的高峰。还要特别感谢他们帮带孩子，让孩子健康、茁壮地成长，使我和谈蒙露能安心地进行博士后工作和博士学业。感谢姐姐、姐夫一家人的帮助和支持，特别是姐姐在为人处理方面给予的榜样和指导。

感谢叶翠翠、叶莎莎、张佳洁等要好的玩伴，使我在紧张的科研工作中有一个疏解压力和焦虑的地方。

最后，特别感谢给到我的老婆谈蒙露和孩子张本之。谈蒙露是我背后那个优秀的女人，从大学以来陪伴、理解和支持我，不仅在生活上照料我，而且在工作上帮助我。特别是，遇到工作上的困难和生活上的苦难，都能很好地沟通，排解焦虑、规划未来、探讨人生的意义，共同在修身齐家治国，实现中华民族伟大复兴的征程上前进！

本工作得到了国家自然科学基金、上海北斗导航研发与转化功能平台、上海市科学技术委员会、装备预研领域基金会、上海交通大学 HPC 计算中心、南方电网公司科技重大专项的资金和算力支持，在此一并表示感谢。

张正德

2022 年 6 月 14 日

个人简历

张正德，男，1992年2月出生，中共党员。教育和博后经历如下：

2011年9月-2015年6月，在北京航空航天大学物理科学与核能工程学院攻读本科，专业：应用物理，获得理学学士学位。

2015年9月-2020年6月，在中国科学院上海应用物理研究所硕博连读，专业：粒子物理与原子核物理，获理学博士学位。

2020年7月-2022年6月，在上海交通大学电子信息与电气工程研究院信息与通信工程从事博士后研究工作，方向：深度学习算法研究。

目前主要从事人工智能深度神经网络算法的研究和应用工作，提出和实现了多项神经网络改进 tricks。

博士后期间，参与百万级 ZF 预研项目 1 项，横向项目 4 项，十万级横向项目 1 项。以一作或通讯身份在 NEURAL COMPUT APPL, IEEE T INSTRUM MEAS 等国内外重要期刊上发表上发表论文 7 篇，以合作作者身份发表论文 5 篇。

其中，神经网络赋能智慧交通领域的论文 CDNet 引起较高关注，2022 年 3 月在 Github 开源以来获得 85stars。VASP 高通量计算辅助程序 VaspCZ 开源以来收获 73 stars。

基于所研发的深度学习算法，获得专利授权 3 项，软著授权 1 项，其中软著进行了技术成果转移（165 万）。

提出量子启发的深度神经网络，在数据驱动的基础上，引入波函数知识要素，用量子演化补充前后帧信息，用冯诺依曼纠缠熵补充目标与环境、目标之间的交互信息，用量子复合系统对多模态进行建模，用量子纠缠处理模态的不可分离性，用量子概率论处理模态的不兼容性，是一种全新的第三代人工智能范式，有望获得原创性技术突破。

博士后期间的研究成果

一作及通讯作者论文：

[1] **Zhang, Z.D.***, Tan M.L. et al., *CDNet: A Real-Time and Robust Crosswalk Detection Network on Jetson Nano Based on YOLOv5* [J]. *Neural Computing & Applications*, 2022, 1(1): 1-1. (SCI, 中科院 Q2)

DOI: <https://doi.org/10.1007/s00521-022-07007-9>.

开源地址: <https://github.com/zhangzhengde0225/CDNet> (85 颗星)

[2] **Zhang Z.D.***, Zhang Bo et al., *FINet: An Insulator Dataset and Detection Benchmark Based on Synthetic Fog and improved YOLOv5* [J]. *IEEE Transactions on Instrumentation and Measurement*, 2022. (SCI, 中科院 Q2, 二审中)

开源地址: <https://github.com/zhangzhengde0225/FINet> (16 颗星)

[3] Zhou F.R., **Zhang Z.D.*** et al., *AIDNet: Detecting Insulators and Defects from Satellite Remote Sensing Images - An Exploration* [C]. *ICCNC 2022 International Conference on Computer Networks and Communications*. Hangzhou, China, Jul. 2022. (EI 会议, 已录用)

[4] Zhou F.R., **Zhang Z.D.*** et. al. *TLDNet: 一种基于数据增强增广和神经网络的输电导线及其缺陷检测方法* [J]. *南方电网技术*, 2022. (中文核心, 二审中)

[5] Xu, C., Pei, L* and **Zhang Z.D.***. *MWNet: A Tracking Method for Frequently Occluded Scenes Based on Matter Wave*. *ICIP 2022 International Conference on Image Processing*, Bordeaux, France. Oct. 2022. (国际会议, 二审中)

合作作者论文：

[6] Tan, M.L., Zhu G.F.*, **Zhang Z.D.**, Zou Y., Yu X.H.*, Yu C.G. Dai Y. and Yan R. *Burnup Optimization of Once-through Molten Salt Reactors Using Enriched Uranium and Thorium* [J]. *Nuclear Science and Techniques*, 2022, 33(1): 1-1. (SCI, 中科院 Q2)

DOI: <https://doi.org/10.1007/s41365-022-00995-2>.

[7] Ji, Z.H., Zhang L.L., Tang, D.M., Chen, C.M. Nordling, T.E.M., **Zhang Z.D.**, Ren C.L., Da B., Li X., Guo S.Y., Liu C. and Cheng H.M. *High-throughput Screening and Machine Learning for the Efficient Growth of High-quality Single-wall Carbon*

Nanotubes [J]. Nano Research, 2021, 14(12): 4610-4615. (SCI, 中科院 Q1)

DOI: <https://doi.org/10.1007/s12274-021-3387-y>.

专利和软著：

[8] 2022. 专利. 基于深度学习的可见光、红外和雷达融合目标检测方法。已授权。

专利号：CN202111440787.X

[9] 2021. 专利. 一种基于深度学习的输电线路绝缘子缺陷检测方法。已授权。专

利号：CN202110253313.8

[10] 2021. 专利. 一种基于神经网络的自然灾害后建筑损伤评估方法。已授权。

专利号：CN202111384342.4

[11] 2020. 软著. AILabelImage: 目标自动检测、跟踪和半自动状态标注软件。已

授权。登记号：2021SR0416473

[12] 2021. 技术成果转移. 基于深度学习的目标自动检测、跟踪和半自动状态标

注软件。2021SR0416473, 共 165 万, 1/3

算法软件：

[13] 多模态标注软件 mmlabelme

[14] 人体行为分析软件 2 个

[15] Xsensing Artificial Intelligence (XAI) 算法平台

[16] Damei Python Library. 开源地址: <https://github.com/zhangzhengde0225/damei>

博士生期间的研究成果

一作论文：

[1] **Z.D. Zhang**, C.L. Ren*, M.L. Tan, Y.Q. Yang, Y.R. Yin, C.Y. Wang, H. Han, P. Huai*. *Migration behavior of tellurium in bcc iron against typical alloying elements: A first-principles study* [J]. Computational Materials Sciences. 2020, 179(1): 109571. (SCI, 中科院 Q3)

DOI: <https://doi.org/10.1016/j.commatsci.2020.109571>

[2] **Z.D. Zhang**, M.L. Tan, R.L. Ren and P. Huai*. VaspCZ: 一个提高效率的 VASP 计算辅助程序[J]. 核技术. 2020, 43(3): 030501. (中文核心)

DOI: <https://doi.org/10.11889/j.0253-3219.2020.hjs.43.030501>

开源地址: <https://github.com/zhangzhengde0225/VaspCZ> (73 颗星)

合作作者论文：

[3] J.B. Hu, J.P. Hu, **Z.D. Zhang**, K.C. Shen, Z.F. Liang, H. Zhang, Q.W. Tian, P. Wang, Z. Jiang, H. Huang, J.W. Wells, F. Song. *Ullmann coupling of 2, 7-dibromopyrene on Au(111) assisted by surface adatoms* [J]. Applied Surface Science. 2020, 513(1): 145797. (SCI, 中科院 Q1)

DOI: <https://doi.org/10.1016/j.apsusc.2020.145797>

[4] J.B. Hu, F.S. Strand, R.K. Chellappan, **Z.D. Zhang**, K.C. Shen, J.P. Hu, G.W. Ji, P. Huai, H. Huang, P. Wang. Z.S. Li, Z. Jiang, J.W. Wells, F. Song. *Direct Synthesis of Semimetal Phthalocyanines on a Surface with Insights into Interfacial Properties* [J]. Journal of Physical Chemistry C. 2020, 124(15): 8247-8256. (SCI, 中科院 Q2)

DOI: <https://doi.org/10.1021/acs.jpcc.0c00895>

[6] L.X. Liao, X. Zhang, C.L. Ren, **Z.D. Zhang**, H.F. Huang, G.H. Ma, P. Huai. *First-principles study of helium behavior in nickel with noble gas incorporation* [J]. Journal of Applied Physics. 2020, 127. (SCI, 中科院 Q3)

DOI: <https://doi.org/10.1063/1.5145016>

永久通信地址

永久邮箱: drivener@163.com

GITHUB 主页: <https://github.com/zhangzhengde0225>

ORCID 主页: <https://orcid.org/0000-0002-6542-052X>