

基于视觉-语言-动作大模型的通用 具身移动操作一体化框架与方法研 究

(清华大学攻读硕士学位研究生论文选题报告)

培养单位：深圳国际研究生院

项目名称：电子与通信工程

申请 人：刘昱杉

学 号：2024214103

指导教师：张晓平 教授

二〇二五年十一月

目 录

| | |
|--------------------------|----|
| 目 录..... | I |
| 第 1 章 课题背景与意义 | 1 |
| 1.1 具身移动操作的研究背景与意义 | 1 |
| 1.2 论文题目的写法 | 2 |
| 1.3 摘要的写法 | 2 |
| 1.4 引言的写法 | 2 |
| 1.5 正文的写法 | 3 |
| 1.6 结论的写法 | 3 |
| 第 2 章 图表示例 | 4 |
| 2.1 插图 | 4 |
| 2.2 表格 | 5 |
| 2.3 算法 | 6 |
| 第 3 章 数学符号和公式 | 7 |
| 3.1 数学符号 | 7 |
| 3.2 数学公式 | 8 |
| 3.3 数学定理 | 8 |
| 第 4 章 引用文献的标注 | 9 |
| 4.1 顺序编码制 | 9 |
| 4.2 著者-出版年制 | 9 |
| 参考文献 | 10 |
| 附录 A 补充内容 | 12 |
| 个人简历、在学期间完成的相关学术成果 | 14 |

第1章 课题背景与意义

1.1 具身移动操作的研究背景与意义

具身移动操作 (Embodied Mobile Manipulation) 的核心目标是在真实或高保真拟真环境中实现“感知—理解—规划—控制”的闭环，使智能体能够依据自然语言或任务指令自主完成各种操作任务。这通常需要机器人具备完备的环境感知与理解，并根据任务类型和操作空间范围实现自主移动与操作规划。在实际应用场景（如家庭、实验室与仓储）中，操作任务并非单一的导航或单一的操作，而是先在未知或半未知空间中移动到合理的语义上下文（例如工作台、目标设备或特定区域），继而完成各种如抓取、开关、插拔、放置等精细动作。因此，面向通用的具身移动操作，更要求机器人在各种复杂场景下，实现稳定的任务执行，并具备跨任务与跨环境的泛化能力、对感知噪声与动态扰动的不敏感与鲁棒性，以及在语义指令驱动下将导航与精细操作一体化闭环（如主动感知、约束安全与失败恢复）的能力。

近年来，基于通用模型 (Foundation Model) 驱动的具身移动和操作分别取得了有效的进展。在移动方面，视觉-语言-导航 (Visual-Language-Navigation, VLN) 在从语言到可达目标的路径规划方面实现了丰富的积累，在仿真和真机中呈现出令人印象深刻的效果；同时，无论是基于模仿学习 (Imitation Learning, IL) 还是离线强化学习 (Offline Reinforcement Learning, RL) 的操作方法，在对象识别、姿态估计、抓取规划等方面都取得了显著的进展，并在真实机器人平台上得到了验证。然而，在移动操作任务中，移动和操作之间存在不可忽视的强耦合：移动阶段所形成的视角、距离与遮挡关系直接决定了后续操作的观测质量与可达性，而操作对目标物体、姿态与约束的先验又会反过来影响移动的策略选择与路径代价。因此，仅在导航或操作单一子任务上取得进展，并不足以保证端到端任务的成功率、稳定性和泛化能力。

与此同时，以视觉-语言-动作 (Visual-Language-Action, VLA) 为代表的大模型在指令理解、情景推理与跨任务迁移上展现出统一表达与泛化能力，通过将开放式语言意图与感知结果对齐为可执行中间表示（如语义子目标、对象一空间一动作三元组），并生成带约束的分层技能序列，从而支撑“从语言到动作”的一体化规划与在动态噪声下的鲁棒重规划，为“从语言到可执行子目标与技能序列”的一体化规划提供了新的可行性。因此，如何将大模型的语义优势落地为低层可控、可验证的移动与操作策略，并在扰动与不确定条件下保持鲁棒，是当前研究的重

要方向。

1.2 论文题目的写法

论文题目应简明扼要地反映论文工作的主要内容，力求精炼、准确，切忌笼统。论文题目是对研究对象的准确、具体描述，一般要在一定程度上体现研究结论，因此，论文题目不仅应告诉读者这本论文研究了什么问题，更要告诉读者这个研究得出的结论。例如：“在事实与虚构之间：梅乐、卡彭特、沃尔夫的新闻观”就比“三个美国作家的新闻观研究”更专业、更准确。

1.3 摘要的写法

论文摘要是对论文研究内容的高度概括，应具有独立性和自含性，即应是一篇简短但意义完整的文章。通过阅读论文摘要，读者应该能够对论文的研究方法及结论有一个整体性的了解，因此摘要的写法应力求精确简明。论文摘要应包括对问题及研究目的的描述、对使用的方法和研究过程进行的简要介绍、对研究结论的高度凝练等，重点是结果和结论。

论文摘要切忌写成全文的提纲，尤其要避免“第1章……；第2章……；……”这样的陈述方式。

1.4 引言的写法

一篇学位论文的引言大致包含以下几个部分：1、问题的提出；2、选题背景及意义；3、文献综述；4、研究方法；5、论文结构安排。

- 问题的提出：要清晰地阐述所要研究的问题“是什么”。^①
- 选题背景及意义：论述清楚为什么选择这个题目来研究，即阐述该研究对学科发展的贡献、对国计民生的理论与现实意义等。
- 文献综述：对本研究主题范围内的文献进行详尽的综合述评，“述”的同时一定要有“评”，指出现有研究状态，仍存在哪些尚待解决的问题，讲出自己的研究有哪些探索性内容。
- 研究方法：讲清论文所使用的学术研究方法。
- 论文结构安排：介绍本论文的写作结构安排。

^① 选题时切记要有“问题意识”，不要选不是问题的问题来研究。

1.5 正文的写法

本部分是论文作者的研究内容，不能将他人研究成果不加区分地掺和进来。已经在引言的文献综述部分讲过的内容，这里不需要再重复。各章之间要存在有机联系，符合逻辑顺序。

1.6 结论的写法

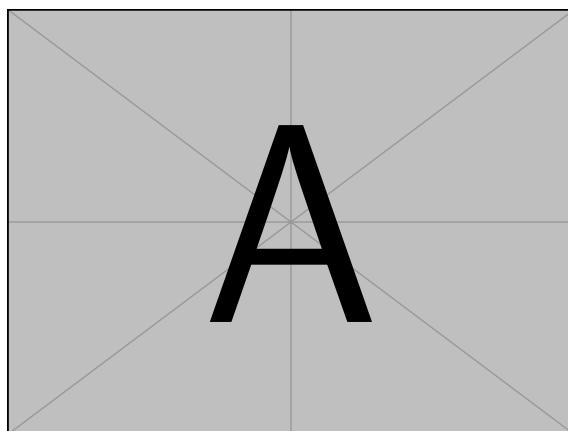
结论是对论文主要研究结果、论点的提炼与概括，应精炼、准确、完整，使读者看后能全面了解论文的意义、目的和工作内容。结论是最终的、总体的结论，不是正文各章小结的简单重复。结论应包括论文的核心观点，主要阐述作者的创造性工作及所取得的研究成果在本领域中的地位、作用和意义，交代研究工作的局限，提出未来工作的意见或建议。同时，要严格区分自己取得的成果与指导教师及他人的学术成果。

在评价自己的研究成果时，要实事求是，除非有足够的证据表明自己的研究是“首次”、“领先”、“填补空白”的，否则应避免使用这些或类似词语。

第 2 章 图表示例

2.1 插图

图片通常在 `figure` 环境中使用 `\includegraphics` 插入，如图 2.1 的源代码。建议矢量图片使用 PDF 格式，比如数据可视化的绘图；照片应使用 JPG 格式；其他的栅格图应使用无损的 PNG 格式。注意，LaTeX 不支持 TIFF 格式；EPS 格式已经过时。

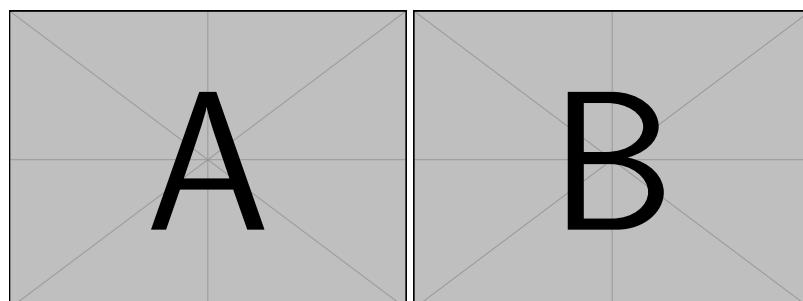


国外的期刊习惯将图表的标题和说明文字写成一段，需要改写为标题只含图表的名称，其他说明文字以注释方式写在图表下方，或者写在正文中。

图 2.1 示例图片标题

若图或表中有附注，采用英文小写字母顺序编号，附注写在图或表的下方。国外的期刊习惯将图表的标题和说明文字写成一段，需要改写为标题只含图表的名称，其他说明文字以注释方式写在图表下方，或者写在正文中。

如果一个图由两个或两个以上分图组成时，各分图分别以 (a)、(b)、(c)…… 作为图序，并须有分图题。推荐使用 `subcaption` 宏包来处理，比如图 2.2(a) 和图 2.2(b)。



(a) 分图 A

(b) 分图 B

图 2.2 多个分图的示例

2.2 表格

表应具有自明性。表中参数应标明量和单位的符号。为使表格简洁易读，均采用三线表（例如表 2.1）。必要时可加辅助线，三线表无法清晰表达时可采用其他格式。

表序与表题置于表的上方。表单元格中的文字一般应居中书写（上下居中，左右居中），不宜左右居中书写的，可采取两端对齐的方式书写。

表 2.1 三线表示例

| 文件名 | 描述 |
|-----------------|------------------|
| thuthesis.dtx | 模板的源文件，包括文档和注释 |
| thuthesis.cls | 模板文件 |
| thuthesis-*.bst | BibTeX 参考文献表样式文件 |

若表中有附注，采用英文小写字母顺序编号，附注写在表的下方。推荐使用 **threeparttable** 宏包。

表 2.2 带附注的表格示例

| 文件名 | 描述 |
|----------------------------|------------------|
| thuthesis.dtx ^a | 模板的源文件，包括文档和注释 |
| thuthesis.cls ^b | 模板文件 |
| thuthesis-*.bst | BibTeX 参考文献表样式文件 |

^a 可以通过 `xelatex` 编译生成模板的使用说明文档；使用 `xetex` 编译 `thuthesis.ins` 时则会从 `.dtx` 中去除掉文档和注释，得到精简的 `.cls` 文件。

^b 更新模板时，一定要记得编译生成 `.cls` 文件，否则编译论文时载入的依然是旧版的模板。

如某个表需要转页接排，可以“续表”的形式另页打印，格式同前，只需在每页表序前加“续”字即可。续表均应重复表头。推荐使用 **longtable** 宏包。

表 2.3 跨页长表格的表题

| 表头 1 | 表头 2 | 表头 3 | 表头 4 |
|-------|------|------|------|
| Row 1 | | | |
| Row 2 | | | |
| Row 3 | | | |
| Row 4 | | | |

续表 2.3 跨页长表格的表题

| 表头 1 | 表头 2 | 表头 3 | 表头 4 |
|--------|------|------|------|
| Row 5 | | | |
| Row 6 | | | |
| Row 7 | | | |
| Row 8 | | | |
| Row 9 | | | |
| Row 10 | | | |

2.3 算法

算法环境可以使用 `algorithms` 或者 `algorithm2e` 宏包。

算法 2.1 Calculate $y = x^n$

输入: $n \geq 0$

输出: $y = x^n$

```

 $y \leftarrow 1$ 
 $X \leftarrow x$ 
 $N \leftarrow n$ 
while  $N \neq 0$  do
    if  $N$  is even then
         $X \leftarrow X \times X$ 
         $N \leftarrow N/2$ 
    else { $N$  is odd}
         $y \leftarrow y \times X$ 
         $N \leftarrow N - 1$ 
    end if
end while

```

第3章 数学符号和公式

3.1 数学符号

中文论文的数学符号默认遵循 GB/T 3102.11—1993《物理科学和技术中使用的数学符号》^①。该标准参照采纳 ISO 31-11:1992^②，但是与 \TeX 默认的美国数学学会（AMS）的符号习惯有所区别。具体地来说主要有以下差异：

1. 大写希腊字母默认为斜体，如

$$\Gamma \Delta \Theta \Lambda \Xi \Pi \Sigma \Upsilon \Phi \Psi \Omega.$$

注意有限增量符号 Δ 固定使用正体，模板提供了 `\increment` 命令。

2. 小于等于号和大于等于号使用倾斜的字形 \leqslant 、 \geqslant 。
3. 积分号使用正体，比如 \int 、 \oint 。
4. 偏微分符号 ∂ 使用正体。
5. 省略号 `\dots` 按照中文的习惯固定居中，比如

$$1, 2, \dots, n \quad 1 + 2 + \dots + n.$$

6. 实部 Re 和虚部 Im 的字体使用罗马体。

以上数学符号样式的差异可以在模板中统一设置。另外国标还有一些与 AMS 不同的符号使用习惯，需要用户在写作时进行处理：

1. 数学常数和特殊函数名用正体，如

$$\pi = 3.14 \dots; \quad i^2 = -1; \quad e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n.$$

2. 微分号使用正体，比如 dy/dx 。
3. 向量、矩阵和张量用粗斜体（`\mathbf`），如 \mathbf{x} 、 $\boldsymbol{\Sigma}$ 、 \boldsymbol{T} 。
4. 自然对数用 $\ln x$ 不用 $\log x$ 。

英文论文的数学符号使用 \TeX 默认的样式。如果有必要，也可以通过设置 `math-style` 选择数学符号样式。

关于量和单位推荐使用 `siunitx` 宏包，可以方便地处理希腊字母以及数字与单位之间的空白，比如： $6.4 \times 10^6 \text{ m}$, $9 \mu\text{m}$, kg m s^{-1} , $10^\circ\text{C} \sim 20^\circ\text{C}$ 。

^① 原 GB 3102.11—1993，自 2017 年 3 月 23 日起，该标准转为推荐性标准。

^② 目前已更新为 ISO 80000-2:2019。

3.2 数学公式

数学公式可以使用 `equation` 和 `equation*` 环境。注意数学公式的引用应前后带括号，通常使用 `\eqref` 命令，比如式 (3.1)。

$$\frac{1}{2\pi i} \int_{\gamma} f = \sum_{k=1}^m n(\gamma; a_k) \mathcal{R}(f; a_k). \quad (3.1)$$

多行公式尽可能在“=”处对齐，推荐使用 `align` 环境。

$$a = b + c + d + e \quad (3.2)$$

$$= f + g \quad (3.3)$$

3.3 数学定理

定理环境的格式可以使用 `amsthm` 或者 `ntheorem` 宏包配置。用户在导言区载入这两者之一后，模板会自动配置 `theorem`、`proof` 等环境。

定理 3.1 (Lindeberg–Lévy 中心极限定理): 设随机变量 X_1, X_2, \dots, X_n 独立同分布，且具有期望 μ 和有限的方差 $\sigma^2 \neq 0$ ，记 $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ ，则

$$\lim_{n \rightarrow \infty} P\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq z\right) = \Phi(z), \quad (3.4)$$

其中 $\Phi(z)$ 是标准正态分布的分布函数。

证明 Trivial. ■

同时模板还提供了 `assumption`、`definition`、`proposition`、`lemma`、`theorem`、`axiom`、`corollary`、`exercise`、`example`、`remark`、`problem`、`conjecture` 这些相关的环境。

第 4 章 引用文献的标注

模板支持 BibTeX 和 BibLaTeX 两种方式处理参考文献。下文主要介绍 BibTeX 配合 `natbib` 宏包的主要使用方法。

4.1 顺序编码制

在顺序编码制下，默认的 `\cite` 命令同 `\citet` 一样，序号置于方括号中，引文页码会放在括号外。统一处引用的连续序号会自动用短横线连接。

```
\cite{zhangkun1994}           ⇒ [1]
\citet{zhangkun1994}          ⇒ 张昆等[1]
\citetp{zhangkun1994}         ⇒ [1]
\cite[42]{zhangkun1994}        ⇒ [1]42
\cite{zhangkun1994, zhukezhen1973} ⇒ [1-2]
```

也可以取消上标格式，将数字序号作为文字的一部分。建议全文统一使用相同的格式。

```
\cite{zhangkun1994}           ⇒ [1]
\citet{zhangkun1994}          ⇒ 张昆等 [1]
\citetp{zhangkun1994}         ⇒ [1]
\cite[42]{zhangkun1994}        ⇒ [1]42
\cite{zhangkun1994, zhukezhen1973} ⇒ [1-2]
```

4.2 著者-出版年制

著者-出版年制下的 `\cite` 跟 `\citet` 一样。

```
\cite{zhangkun1994}           ⇒ 张昆等 (1994)
\citet{zhangkun1994}          ⇒ 张昆等 (1994)
\citetp{zhangkun1994}         ⇒ (张昆等, 1994)
\cite[42]{zhangkun1994}        ⇒ (张昆等, 1994)42
\citetp{zhangkun1994, zhukezhen1973} ⇒ (张昆等, 1994; 竺可桢, 1973)
```

注意，引文参考文献的每条都要在正文中标注^[1-34]。

参考文献

- [1] 张昆, 冯立群, 余昌钰, 等. 机器人柔性手腕的球面齿轮设计研究[J]. 清华大学学报: 自然科学版, 1994, 34(2): 1-7.
- [2] 竺可桢. 物理学论[M]. 北京: 科学出版社, 1973: 56-60.
- [3] Dupont B. Bone marrow transplantation in severe combined immunodeficiency with an unrelated mlc compatible donor[C]//White H J, Smith R. Proceedings of the third annual meeting of the International Society for Experimental Hematology. Houston: International Society for Experimental Hematology, 1974: 44-46.
- [4] 郑开青. 通讯系统模拟及软件[D]. 北京: 清华大学无线电系, 1987.
- [5] 姜锡洲. 一种温热外敷药制备方案: 中国, 88105607.3[P]. 1980-07-26.
- [6] 中华人民共和国国家技术监督局. GB3100-3102. 中华人民共和国国家标准-量与单位[S]. 北京: 中国标准出版社, 1994.
- [7] Merkt F, Mackenzie S R, Softley T P. Rotational autoionization dynamics in high rydberg states of nitrogen[J]. J Chem Phys, 1995, 103: 4509-4518.
- [8] Mellinger A, Vidal C R, Jungen C. Laser reduced fluorescence study of the carbon monoxide and triplet rydberg series - experimental results and multichannel quantum defect analysis[J]. J Chem Phys, 1996, 104: 8913-8921.
- [9] Bixon M, Jortner J. The dynamics of predissociating high Rydberg states of NO[J]. J Chem Phys, 1996, 105: 1363-1382.
- [10] 马辉, 李俭, 刘耀明, 等. 利用 REMPI 方法测量 BaF 高里德堡系列光谱[J]. 化学物理学报, 1995, 8: 308-311.
- [11] Carlson N W, Taylor A J, Jones K M, et al. Two-step polarization-labeling spectroscopy of excited states of Na2[J]. Phys Rev A, 1981, 24: 822-834.
- [12] Taylor A J, Jones K M, Schawlow A L. Scanning pulsed-polarization spectrometer applied to Na2[J]. J Opt Soc Am, 1983, 73: 994-998.
- [13] Taylor A J, Jones K M, Schawlow A L. A study of the excited $1\Sigma g^+$ states in Na2[J]. Opt Commun, 1981, 39: 47-50.
- [14] Shimizu K, Shimizu F. Laser induced fluorescence spectra of the a $3\Pi u$ -X $1\Sigma g^+$ band of Na2 by molecular beam[J]. J Chem Phys, 1983, 78: 1126-1131.
- [15] Atkinson J B, Becker J, Demtröder W. Experimental observation of the a $3\Pi u$ state of Na2[J]. Chem Phys Lett, 1982, 87: 92-97.
- [16] Kusch P, Hessel M M. Perturbations in the a $1\Sigma u^+$ state of Na2[J]. J Chem Phys, 1975, 63: 4087-4088.
- [17] 广西壮族自治区林业厅. 广西自然保护区[M]. 北京: 中国林业出版社, 1993.
- [18] 霍斯尼. 谷物科学与工艺学原理[M]. 李庆龙, 译. 2 版. 北京: 中国食品出版社, 1989: 15-20.
- [19] 王夫之. 宋论[M]. 刻本. 金陵: 曾氏, 1865 (清同治四年).

- [20] 赵耀东. 新时代的工业工程师[M/OL]. 台北: 天下文化出版社, 1998[1998-09-26]. <http://www.ie.nthu.edu.tw/info/ie.newie.htm>.
- [21] 全国信息与文献工作标准化技术委员会出版物格式分委员会. GB/T 12450-2001 图书书名页[S]. 北京: 中国标准出版社, 2002: 1.
- [22] 全国出版专业职业资格考试办公室. 全国出版专业职业资格考试辅导教材: 出版专业理论与实务·中级[M]. 2014 版. 上海: 上海辞书出版社, 2004: 299-307.
- [23] World Health Organization. Factors regulating the immune response: Report of WHO Scientific Group[R]. Geneva: WHO, 1970.
- [24] Peebles P Z, Jr. Probability, random variables, and random signal principles[M]. 4th ed. New York: McGraw Hill, 2001.
- [25] 白书农. 植物开花研究[M]//李承森. 植物科学进展. 北京: 高等教育出版社, 1998: 146-163.
- [26] Weinstein L, Swertz M N. Pathogenic properties of invading microorganism[M]//Sodeman W A, Jr, Sodeman W A. Pathologic physiology: mechanisms of disease. Philadelphia: Saunders, 1974: 745-772.
- [27] 韩吉人. 论职工教育的特点[C]//中国职工教育研究会. 职工教育研究论文集. 北京: 人民教育出版社, 1985: 90-99.
- [28] 中国地质学会. 地质评论[J]. 1936, 1(1)–. 北京: 地质出版社, 1936–.
- [29] 中国图书馆学会. 图书馆学通讯[J]. 1957(1)–1990(4). 北京: 北京图书馆, 1957–1990.
- [30] American Association for the Advancement of Science. Science[J]. 1883, 1(1)–. Washington, D.C.: American Association for the Advancement of Science, 1883–.
- [31] 傅刚, 赵承, 李佳路. 大风沙过后的思考[N/OL]. 北京青年报, 2000-04-12(14)[2002-03-06]. <http://www.bjyouth.com.cn/Bqb/20000412/B/4216%5ED0412B1401.htm>.
- [32] 萧钰. 出版业信息化迈入快车道[EB/OL]. (2001-12-19)[2002-04-15]. <http://www.creader.com/news/20011219/200112190019.htm>.
- [33] Online Computer Library Center, Inc. About OCLC: History of cooperation[EB/OL]. 2000 [2000-01-08]. <http://www.oclc.org/about/cooperation.en.htm>.
- [34] Scitor Corporation. Project scheduler[CP/DK]. Sunnyvale, Calif.: Scitor Corporation, 1983.

附录 A 补充内容

附录是与论文内容密切相关、但编入正文又影响整篇论文编排的条理和逻辑性的资料，例如某些重要的数据表格、计算程序、统计表等，是论文主体的补充内容，可根据需要设置。

附录中的图、表、数学表达式、参考文献等另行编序号，与正文分开，一律用阿拉伯数字编码，但在数码前冠以附录的序号，例如“图 A.1”，“表 A.1”，“式 (A.1)”等。

A.1 插图

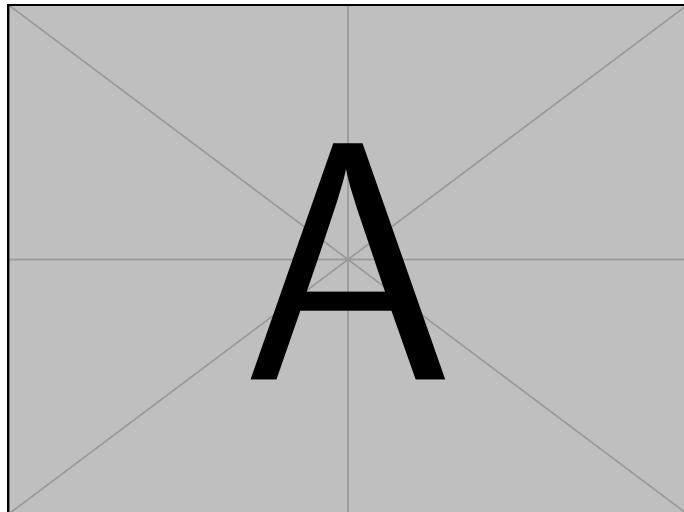


图 A.1 附录中的图片示例

A.2 表格

A.3 数学表达式

$$\frac{1}{2\pi i} \int_{\gamma} f = \sum_{k=1}^m n(\gamma; a_k) \mathcal{R}(f; a_k) \quad (\text{A.1})$$

A.4 文献引用

附录[A.1] 中的参考文献引用[A.2]示例[A.1-A.2]。

表 A.1 附录中的表格示例

| 文件名 | 描述 |
|-----------------|--------------------|
| thuthesis.dtx | 模板的源文件，包括文档和注释 |
| thuthesis.cls | 模板文件 |
| thuthesis-*.bst | BibTeX 参考文献表样式文件 |
| thuthesis-*.bbx | BibLaTeX 参考文献表样式文件 |
| thuthesis-*.cbx | BibLaTeX 引用样式文件 |

参考文献

- [A.1] Dupont B. Bone marrow transplantation in severe combined immunodeficiency with an unrelated mlc compatible donor[C]//White H J, Smith R. Proceedings of the third annual meeting of the International Society for Experimental Hematology. Houston: International Society for Experimental Hematology, 1974: 44-46.
- [A.2] 郑开青. 通讯系统模拟及软件[D]. 北京: 清华大学无线电系, 1987.

个人简历、在学期间完成的相关学术成果

个人简历

2003 年 9 月 11 日出生于吉林省蛟河市。

2020 年 9 月考入吉林大学电子科学与工程学院电子信息科学与技术专业,2024 年 6 月本科毕业并获得理学学士学位。

2024 年 9 月考入清华大学深圳国际研究生院, 攻读电子与通信工程硕士至今。

在学期间完成的相关学术成果

学术论文：

- [1] **Liu Y**, Mu S, Chao X, et al. AVR: Active Vision-Driven Precise Robot Manipulation with Viewpoint and Focal Length Optimization.
- [2] Chao X, Mu S, **Liu Y**, et al. Exo-ViHa: A Cross-Platform Exoskeleton System with Visual and Haptic Feedback for Efficient Dexterous Skill Learning.
- [3] Li S, **Liu Y**, Chao X, et al. ALARMbot: Autonomous Laboratory Safety Inspection and Operable Hazard Intervention Robot Enabled by Foundation Models.