

分 类 号 \_\_\_\_\_  
学校代码 10487

学号 M2019xxxxx  
密级 \_\_\_\_\_

華中科技大學

# 硕士学位论文

( 学术型 ☐      专业型 ☐ )

标题：宋体，英文 Times New Roman，一号，加粗，不超 30 字  
(中英文标题、学科专业、导师姓名正确、一致)

学位申请人： XXX

学 科 专 业： XXXXX

指 导 教 师： XXX 教授

答 辩 日 期： 202X 年 X 月 X 日

**A Thesis Submitted in Partial Fulfillment of the Requirements for  
the Master Degree in Engineering**

**English Title, Times New Roman, 小二号, 实词的首字母大写**

**中英文标题、学科专业、导师姓名正确、一致**

**Candidate : xxx (中文习惯, 姓在前且姓全部大写)**

**Major : Control Science and Engineering**

**Supervisor : Prof. Xxxx**

**Huazhong University of Science and Technology**

**Wuhan 430074, P. R. China**

**January, 2026**

## 独创性声明

本人声明所呈交的学位论文是我个人在导师的指导下进行的研究工作及取得的研究成果。尽我所知，除文中已标明引用的内容外，本论文不包含任何其他人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：

日期： 年 月 日

## 学位论文版权使用授权书

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

本论文属于 ☐ 保 密 ☐，在 \_\_\_\_ 年解密后适用本授权书。  
☐ 不保密 ☐。

（请在以上方框内打“√”）

学位论文作者签名：

指导教师签名：

日期： 年 月 日

日期： 年 月 日

## 摘 要

摘要是学位论文极为重要、不可缺少的组成部分，它是论文的窗口，并频繁用于国内外资料交流、情报检索、二次文献编辑等。其性质和要求如下：

[1] 摘要即摘录论文要点，是论文要点不加注释和评论的一篇完整的陈述性短文，具有很强的内含性和独立性，能独立使用和被引用。

[2] 博士学位论文的摘要应包含全文的主要信息，并突出创造性成果。

[3] 内容范围应包含以下基本要素：

(1) 目的：研究、研制、调查等的前提、目的和任务以及所涉及的主题范围。

(2) 方法：所用原理、理论、条件、对象、材料、工艺、手段、装备、程序等。

(3) 结果：实验的、研究的、调查的、观察的结果、数据，被确定的关系，得到的效果、性能等。

(4) 结论：结果的分析、研究、比较、评价、应用；提出的问题，今后的课题，建议，预测等。

(5) 其他：不属于研究、研制、调查的主要目的，但就其见识和情报价值而言也是重要的信息。

[4] 摘要的详简度视论文的内容、性质而定，**硕士学位论文摘要一般为 500 — 600 汉字。**

[5] 摘要及全文中均建议不出现“我们”等字样。摘要中主语（作用）常常省略，因而一般使用被动语态；应使用正确的时态，并要注意主、谓语的一致，必要的冠词不能省略。

[6] 一般不用图、表、化学结构式、计算机程序，不用非公知公用的符号、术语和非法定的计量单位。

[7] 摘要中一般不使用缩写词，若实在需要，在第一次使用前，需给出中文全称（缩写词）；在使用英文缩写词之前，需给出英文全称（英文全称，缩写词），再次出现时可以采用中文或英文缩写词。

[8] **关键词应有 3 至 8 个，另起一行置于摘要下方，领域从大到小排列。关键词之间用分号隔开，最后一个关键词后面无标点。**

# 华 中 科 技 大 学 硕 士 学 位 论 文

---

[9] 摘要、关键词采用中文宋体；英文 Times New Roman；小四号。

[10] 应有与中文摘要和关键词相对应的英文摘要和关键词。英语摘要用词应准确，使用本学科通用的词汇。

**关键词：**关键词 1；关键词 2；关键词 3

## **Abstract**

This is abstract.

英文摘要字体为 Times New Roman，小四，1.5 倍行距。

英文摘要和关键词应与中文相对应。英语摘要用词应准确，使用本学科通用的词汇；摘要中主语（作用）常常省略，因而一般使用被动语态；应使用正确的时态，并注意主、谓语的一致，必要的冠词不能省略。

**Keywords:** Keyword1, Keyword2, Keyword3

## 目 录

摘 要 .....	I
Abstract .....	III
主要符号对照表 .....	VI
<b>1 绪论</b>	
1.1 研究背景与意义 .....	(1)
1.2 国内外研究现状 .....	(3)
1.3 存在的问题 .....	(6)
1.4 本文主要内容 .....	(7)
<b>2 体育动作时空动作检测相关理论和技术</b>	
2.1 引言 .....	(10)
2.2 时空动作检测任务定义 .....	(10)
2.3 体育场景时空动作场景的挑战 .....	(12)
2.4 控制器设计与闭环系统分析（请根据所设计的控制器特点自行拟定具体的题目） .....	(12)
2.5 数值仿真（请拟定具体的题目） .....	(12)
2.6 本章小结 .....	(12)
<b>3 理论/算法类研究类论文</b>	
3.1 引言（引言标题可选） .....	(13)
3.2 ** 理论/算法 .....	(13)
3.3 ** 仿真或算法实现 .....	(13)
3.4 理论/算法准确性的评估 .....	(13)
3.5 分析与讨论 .....	(13)
3.6 本章小结 .....	(13)

#### **4 学位论文写作细则**

4.1 关于图 .....	(14)
4.2 关于表格 .....	(15)
4.3 名词、术语 .....	(16)
4.4 符号、单位的使用 .....	(16)
4.5 数字的使用 .....	(17)
4.6 其它应该注意的问题 .....	(17)

#### **5 总结与展望**

5.1 本文主要内容及结论 .....	(21)
5.2 本文主要创新点 .....	(21)
5.3 展望 .....	(21)

致 谢 .....	(22)
-----------	------

附录 1 攻读硕士学位期间取得的研究成果 .....	(23)
----------------------------	------

附录 2 攻读硕士学位期间参与的科研项目 .....	(24)
----------------------------	------

附录 3 其他附录 .....	(25)
-----------------	------

## 主要符号对照表

xue	我的姓
ruini	我的名
W.M. Zheng	我的老师
Tsinghua	学校名
Long	来个比较长的，看看会出现什么情况。
劝学	君子曰：学不可以已。青，取之于蓝，而青于蓝；冰，水为之，而寒于水。木直中绳。（车柔）以为轮，其曲中规。虽有槁暴，不复挺者，（车柔）使之然也。故木受绳则直，金就砺则利，君子博学而日参省乎己，则知明而行无过矣。吾尝终日而思矣，不如须臾之所学也；吾尝（足齐）而望矣，不如登高之博见也。登高而招，臂非加长也，而见者远；顺风而呼，声非加疾也，而闻者彰。假舆马者，非利足也，而致千里；假舟楫者，非能水也，而绝江河，君子生非异也，善假于物也。积土成山，风雨兴焉；积水成渊，蛟龙生焉；积善成德，而神明自得，圣心备焉。故不积跬步，无以至千里；不积小流，无以成江海。骐骥一跃，不能十步；驽马十驾，功在不舍。锲而舍之，朽木不折；锲而不舍，金石可镂。蚓无爪牙之利，筋骨之强，上食埃土，下饮黄泉，用心一也。蟹六跪而二螯，非蛇鳝之穴无可寄托者，用心躁也。—— 荀况

## 1 绪论

### 1.1 研究背景与意义

随着移动设备的普及和互联网技术的快速发展，视频内容数据呈现爆炸式增长。视频由于其直观、生动的表现形式，已成为信息传播和交流的重要媒介。根据中国视听大数据（CVB）<sup>[2]</sup>统计显示，全国卫视频道体育赛事播出总场次 43329 场，其中直播赛事 4901 场，全国累计收视规模超 247.5 亿人次，累计收视时长突破 66.6 亿小时。面对海量的体育场景视频，传统的人工视频分析技术已无法满足人们的需求，针对体育场景的智能化视频内容理解技术已成为研究人员关注的重点。

人工智能技术与体育竞技进行深度融合是行业发展的趋势。根据国家体育总局相关报道，“体育+人工智能”行动将纳入《“十五五”体育科教发展规划》，将汇聚各方科技力量大力推进人工智能在体育行业的应用。在体育运动领域中，基于人工智能的视频分析技术已经在运动员训练辅助、体育赛事分析、赛事转播等多个场景都有丰富的实际应用并发挥着重要作用。在运动员训练中，利用人体三维重建技术，准确分析运动员的动作规范性，辅助动作矫正；在体育赛事分析中，通过视频理解大模型对体育视频进行智能分析，将运动员的各项数据进行量化和可视化，从而提升竞技体育的专业化和智能化水平；在赛事转播中，可以利用计算机视觉和虚拟现实等技术实现智能信息播报和结果预测，增强与线上观众的互动，提升观赛体验。随着相关技术的发展，如 AI 裁判、AI 教练、AI 自动集锦等新兴人工智能技术正逐步渗透到体育运动的各个环节，推动体育产业的蓬勃发展。

时空动作检测作为视频理解领域下的关键技术之一，能在未经剪辑的视频中识别出关注的动作类别、定位动作时间片段的起始帧与结束帧并给出动作主体的空间位置，同时实现时间和空间维度的定位，已在包括智能安防、自动驾驶、虚拟现实等领域有着广泛的应用。近年来，针对体育视频的时空动作检测技术得到了广泛关注，由于其可以从海量未剪辑的体育运动视频中，精确地定位出特定的运动动作片段，并识别出运动员的动作类别，为运动员训练、赛事分析等提供有力的数据支持，极大提升了体育视频分析的效率和精度，为诸多下游应用提供了坚实的技术基础。

然而，不同于常规的安防场景，体育运动场景下的时空动作检测有着多重挑战。

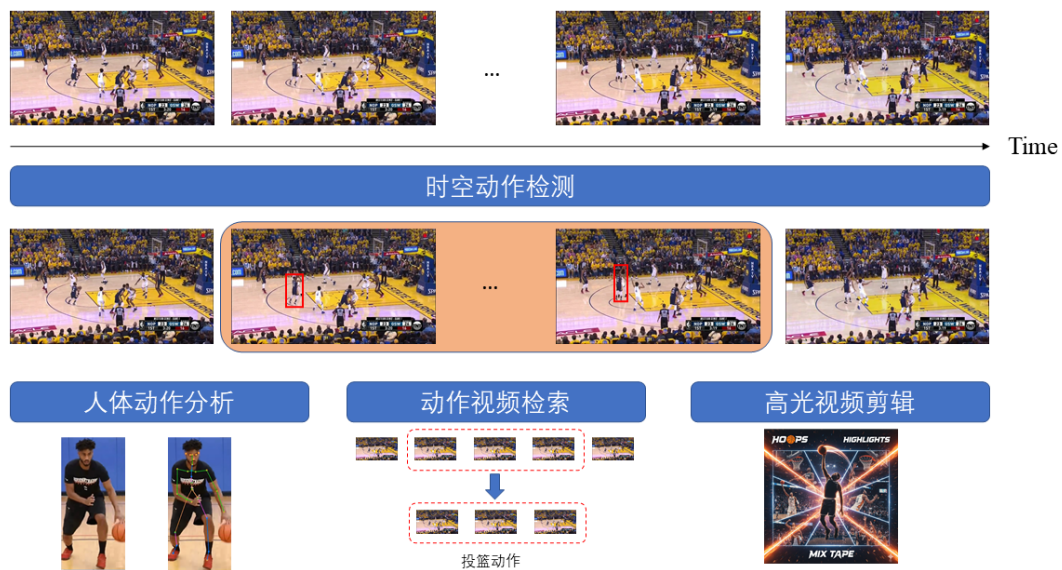


图 1.1 体育时空动作检测的应用

对于体育运动场景，运动的动作通常复杂多变、运动员运动剧烈、速度快，动作过程中伴随大范围的形变和位移，且在运动竞技过程中难免存在运动主体相互遮挡的情况，根据 A 的相关统计，体育动作场景的动作内复杂度和动作间复杂度通常是日常动作的 3-8 倍，为此，如何提高复杂运动情况下的帧间目标一致性是体育视频时空动作检测的一大关键。此外，体育运动视频中动作类别繁多，部分技术动作之间的视觉细微差异也会对时空动作检测带来困难，如足球中的传球和射门、篮球中的二分之一球和三分球，其动作特征十分相似的，主要的区别在于动作发生的位置和动作的意图，这要求模型在理解动作运动特征之外，还需要对全局视频的时空视觉线索进行充分的挖掘，只有对包括场地信息、运动员信息等信息进行充分关系建模，才能实现对细粒度体育动作类别的准确区分。

综上所述，针对体育运动场景的时空动作检测技术的研究，不仅有具体的实用价值，而且具有重要的学术研究意义。在应用方面，体育运动场景的时空动作检测技术可以提升体育视频分析的效率和精度，不仅可以通过对运动人员的动作进行精确的分析，切实地指导训练而提高运动员的竞技能力，也可以为赛事转播提供智能化的技术支持，提升观众的观赛体验，带来丰厚的经济效益；在学术研究方面，深入研究体育视频时空动作检测技术，有助于推动视频理解相关技术在处理复杂运动场景视频的能力。因此，本文针对体育运动场景下的时空动作检测技术展开深入研究。

## 1.2 国内外研究现状

### 1.2.1 时空动作检测研究现状

时空动作检测是动作识别任务和时序动作检测任务的延伸。动作识别旨在对一段视频片段中发生的动作进行分类，仅关注这段视频发生了什么动作，属于基础的视频分类任务；时序动作检测则需要在动作识别的基础上，确定动作在视频中发生的时间区间，同时实现动作分类和时间维度的定位；而时空动作检测不仅仅需要知道视频动作的类别和发生区间，还需要在空间维度上对动作目标主体进行定位，给出动作发生过程中在视频画面中的位置，从而实现时间维度和空间维度的定位。

近年来，随着深度学习技术的快速发展，时空动作检测技术取得了显著进展。现有的时空动作检测方法主要根据模型结构范式可以分为双阶段方法和单阶段方法两大类，本节将分别介绍双阶段和单阶段时空动作检测方法的研究现状。

双阶段时空动作检测方法通常会依赖于额外的目标检测器或 RPN 网络，预先生成的候选动作主体的 ROI 区域，再通过对 ROI 区域进行时空特征提取和环节关系建模完成时空动作检测任务。

(1) **双阶段时空动作检测**根据所使用的 ROI 区域的特点，可以分为帧级的方法和片段级的方法。前者通常会利用关键帧上的所产生的 ROI 区域尝试与全局视觉特征进行交互融合，从而提升模型的环境理解能力；后者则是更加关注于如何生成高质量的动作管级别 3D 候选区域，以提升模型的动作理解能力。

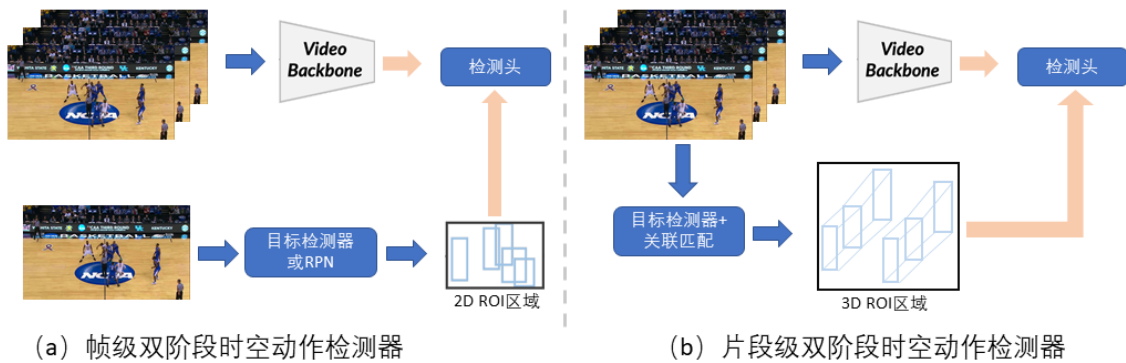


图 1.2 双阶段时空动作检测算法可分为 (a) 帧级检测方法，(b) 片段级检测方法

帧级的双阶段方法受益于目标检测任务的发展，SAHA 等人首次在时空动作检测任务中引入了基于 Fast R-CNN 的 RPN 网络来代替传统的无监督区域生成算法，

实现了对动作主体的高效定位，他们将 RGB 图像和光流图像分别输入到两个独立的区域提议网络，以输出检测框和动作类别得分。ACAM 通过设计一种关系建模模块，对由 RPN 生成的候选区域特征与全局特征图进行交互建模，提升了模型对环境相关动作的检测能力。MRSN 则是基于 Vision transformer 架构，将全局视频特征与局部候选区域特征进行 Patch 化处理，通过 attention 机制进行运动员特征与全局视觉特征的交互融合，增强了模型的场景理解能力。HIT 额外引入了人体关键点和手部区域信息，通过融合人体-手部-关键点三重特征，提升了模型对细粒度动作的识别能力。EVAD 通过设计了基于关键帧的 Token dropout 机制，来提升模型的推理效率。

片段级的双阶段方法则是更加关注于如何在进行关系建模之前进行帧间目标关联，以生成高质量的动作管级别 3D 候选区域。CFAD 提出了一种从粗到精获取 3D ROI 的新范式，通过粗略模块对长时域信息进行参数化建模，先从视频流中估算出初步的动作管，随后再利用精细模块，在关键时间戳的引导下，对初步估算的管柱位置进行选择性和细化。TrAD 则是通过利用额外的目标跟踪器生成 TOI 区域获得高质量的动作管候选区域，为动作分类提供了高质量的空间信息先验。ART 则是参考了目标跟踪中的相关性学习方法，通过计算帧间的 query 相似度，设计了 query 层级的匹配机制，进而生成 3D 片段级的 query，隐式地实现了帧间目标关联和动作管生成。

## （2）单阶段时空动作检测

近年来单阶段目标检测器的快速发展，如 YOLO、centerNet、Sparse R-CNN、DETR 等范式通过端到端的模型设计，高效地实现了视觉定位任务。时空动作检测方法受此启发，也产生了很多单阶段时空动作检测方法，直接对视频帧进行时空特征提取和动作检测，无需额外的候选区域生成步骤。

YOWO 是参考 YOLO 首次将单阶段目标检测器思想引入时空动作检测任务的方法，其设计了双分支视觉特征提取网络，分别使用一个 2D backbone 和一个 3D backbone 来提取空间和时序特征，并通过信息融合模块将两者进行融合，最终通过单个检测头实现动作分类和位置信息的回归。MOC 则是参考 centerNet 的思想，设计了一种自上而下的基于中心点时空动作检测方法，通过预测输入片段关键帧的中心点和其它帧中心点的相对偏移量，以实现了对目标运动信息的提取并直接输出片段级的预测结果。STAN 是一个低功耗、高性能的即插即用模块，它通过增强模型对视角和几何变换的适应能力，让现有的视频识别算法变得更加鲁棒。Tuber 则是首

个将 DETR 范式引入时空动作检测任务的方法，通过设计时空动作查询机制，利用 Transformer 的自注意力机制对时空特征进行建模，实现了对动作类别和位置的直接预测。在此基础上，STAR 则是受益于新型的视频预训练模型 Vivit，并设计了时空解耦的时空动作查询并解耦时空注意力计算模块，在相关基线上取得了显著提升。STMixer 通过设计自适应采样模块和双分支时空特征融合模块，在高效地提取时空特征的同时，实现了对时空特征的充分融合，高效地实现了单阶段时空动作检测。SPDet 则是的核心在于使用可学习的管柱查询直接进行时空动作检测，它彻底舍弃了传统方法中繁琐的手工预设锚点和低效的帧间链接操作，通过直接对可以在更长的窗口（如 64 帧）内进行全局回归，这种长时域建模能力使得模型能够更精确地利用长期信息，并显式预测动作的时间边界。

### 1.2.2 体育视频时空动作检测研究现状

体育场景的视频理解算法研究不仅可以辅助运动员训练，提高观众的观赛体验，带来明显的经济效益，而且由于体育运动的特殊性，相比于一般的视频分析任务更具挑战性，因此近年来受到了广泛关注。

Khurram Soomro 等人最先关注到体育视频研究价值，根据 UCF101 数据集衍生出了针对体育场景的时空动作检测数据集 UCF Sports，该数据集包含了 10 类体育运动动作，并提供了每个动作的时空标注信息，主要研究了基于传统手工特征的体育视频动作识别方法。随着相关研究的深入，越来越多更具挑战性的体育视频数据集被提出，如 Sports-1M、FineGym、SoccerNet 等，涵盖了包括篮球、足球、体操等多种体育运动场景，为体育视频理解算法的研究提供了丰富的数据资源。

当前针对体育视频理解分析的研究多种多样，如体育视频质量评估、体育视频问答大模型、AI 足球裁判等，极大推动了人工智能技术在体育视频理解方面的应用。在足球运动视频中，JaiiHeild 等人提出了视频助理裁判系统，可以自动化足球决策。该系统利用多视图视频分析的最新发现，向裁判提供实时反馈，并帮助他们做出可能影响比赛结果的明智决策。Hassan 等人提出了针对足球比赛一的种密集视频字幕的工作，重点关注生成以单个时间戳为锚定的文本评论，能够生成满足足球比赛视频中的带时间戳的评论。Saikat 等人提出双重交互强化学习代理来联合执行基于传球计数的控球统计生成的球队控球统计数据。Honda 等人使用多模态的方法，结合了视觉信息和轨迹数据，为传球接球预测提供了新的方法。

针对体育场景下的时空动作检测任务研究,近些年也取得了显著的进展。TAAD 针对体育视频场景下的时空动作检测任务,基于当前先进的目标跟踪器 YOLOv5-DeepSort,并使用预训练的 reID 基线模型 OsNet-x0-25 作为特征提取器进一步提高对运动员的外观判别能力,在进行时空动作检测前便实现了对运动员的高质量跟踪,从而提升了模型在复杂多人运动场景下的动作检测能力。PoSTAL 则是创新性地引入了 BLIP 多模态预训练模型,通过对体育视频中运动员的衣物颜色和号码等文本信息进行提取和融合,并通过设计的基于 Prompt 的动作编码器和动作管解码器直接预测动作管,提高了模型对细粒度篮球动作类别的区分能力。HHIDet 则是针对篮球场景和网球场景中运动员交互关系复杂的问题,使用人体检测器获取视频中所有的人体提议,并提出了显式的交互提议生成和提议间的信息交换机制,以更好地捕捉主体和客体之间的空间几何关系。

## 1.3 存在的问题

虽然目前针对体育视频的时空动作检测技术已经取得了一定的进展,但仍然存在一些亟待解决的关键问题:

(1) 复杂运动状态的帧间信息关联。对于体育场景下动作的运动特点,通常存在高速运动、巨大形变、相互遮挡等复杂情况,现有的时空动作检测方法大多依赖于预训练的目标检测器或 RPN 网络生成候选区域在时间维度进行复制,难以应对大位移、强遮挡等情况,甚至会带来噪声干扰。虽然近年来有一些方法尝试通过目标跟踪等手段进行实例级别帧间关联,但大多难以适应体育场景下复杂的运动情况,导致生成的动作管质量不高,影响了后续的动作检测性能,且对于遮挡情况,3D ROI 由于几何约束的限制,十分容易引入干扰信息。

(2) 多人体育动作情况的关系信息建模。体育运动视频中动作类别繁多,部分技术动作之间的视觉细微差异也会对时空动作检测带来困难,很多动作只有结合场景信息才能进行准确识别,尤其是在多人运动场景下,运动员与运动员之间、运动员与环境直接的交互关系复杂。现在的方法大多以运动员为中心,与环境进行注意力计算以实现交互建模,或通过图神经网络等方式对视频中关键元素的 RPN 特征进行关系建模但是此类方法仅局限于局部区域的信息交互,难以实现对全局时空视觉信息的充分挖掘,限制了模型对细粒度动作类别的区分能力,并且在复杂多人运动

场景下，遮挡和对抗导致基于 ROI 的交互关系也难以被充分建模。

(3) 体育运动视频的时空特征建模。现有的时空动作检测方法大多采用 3D CNN 或 Transformer 等网络结构进行视觉特征提取，且现有的框架大多依赖于视频提取视觉特征，或辅以光流或人体骨骼关键点补充运动信息，但是对于最为关键的视觉特征对时序维度和空间维度进行统一建模，导致时序特征和空间特征存在严重信息耦合。而体育运动视频中，时序特征和空间特征的差异性较大，时序特征主要体现在动作的运动轨迹和速度变化上，而空间特征则更多体现在动作发生的具体位置和环境背景上，二者的差异性使得统一建模难以充分挖掘各自的特征信息，限制了模型对细粒度动作类别的区分能力。

## 1.4 本文主要内容

针对以上时空动作检测技术在体育场景下存在的关键问题，本文提出了两种新型的时空动作检测算法，并分别从以上三个方面进行了改进和提升。具体工作内容如下：

(1) 基于时空动作主题引导的时空动作检测方法。针对复杂运动状态下的帧间信息关联问题，本算法提出了通过设计了一种时空动作主题预测模块 TAM (Topic Aware Module) 对输入视频片段进行全局主题特征压缩和提取，并通过主题特征引导的帧间信息关联模块 TIG (Topic Induced Grouping) 实现对复杂运动状态下的帧间信息关联。此外，为了减少 3D ROI 引入的噪声干扰，本文设计了一种自适应特征时空特征采样策略和可变形时空注意力模块，以提升模型对动作主体的特征提取能力。由此实现特征级的帧间信息关联，提升了模型在复杂运动状态下的动作检测能力。

(2) 基于多模特征解耦的体育视频时空动作检测方法。针对多人体育动作场景下的关系信息建模问题，本算法利用当前具备强大场景理解能力和多模态特征提取能力的预训练模型 Qwen-VL，根据关注的动作运动特点构建多模态特征库，借用多模态大模型的强理解能力，实现对包括场地信息、运动员信息等多模态特征的提取和融合，以加强对细粒度动作的理解能力。除此之外，本算法考虑到体育运动视频中时序特征和空间特征差异性，针对基于 Qwen-VL 所提取到的多模态特征，设计了一种双分支时空特征解耦模块，以分别对时序特征和空间特征进行独立对齐，显著提高了算法对细粒度动作的检测能力。

本文共分为 5 章内容，章节内容之间的关系图如图 1.3 所示。

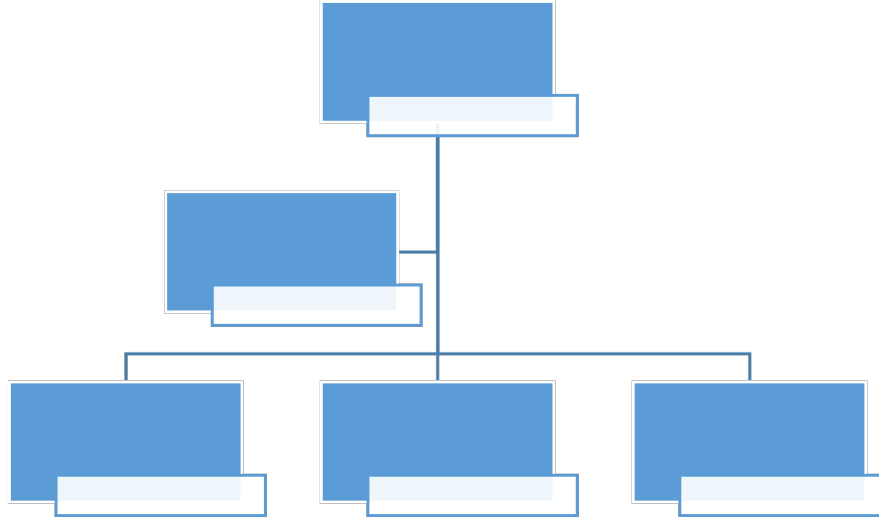


图 1.3 组织结构图

第一章：绪论。在本章中，首先介绍了体育视频时空动作检测技术的研究背景与意义。其次，综述了国内外的体育运动场景下时空动作检测领域的研究现状，并分析和总结了现有相关研究中仍然存在的 key 问题。最后针对相关 key 问题，阐述了本文的研究内容，包括基于时空动作主题引导的时空动作检测算法和基于多模特征解耦的体育视频时空动作检测算法。

第二章：时空动作检测相关理论和技术。本章详细介绍了与本文工作相关的时空动作检测相关理论和技术，包括的基于查询的集合预测网络、视频理解大模型、多模态信息知识库等技术。然后，我们通过对现有体育场景下时空动作检测数据集进行详细讨论，细致说明体育场景动作检测的特点。最后，介绍了时空动作检测的性能评价指标。

第三章：基于时空动作主题引导的时空动作检测方法。针对复杂运动状态的帧间空间信息关联，本章提出了基于时空动作主题特征引导策略，并结合自适应时空动作采样和可变形时空注意力灵活地进行帧间信息关联，以解决常规 ROI 几何约束引入噪声的问题。时空动作主题特征引导策略通过一个时空动作主题特征感知模块对动作主题特征进行提取，并通过指导时空特征采样来引导特征级别信息关联。自适应时空动作采样和可变形时空注意力提升了模型对动作主体的特征提取能力，进而提升了复杂运动状态下的时空动作检测性能。最后，通过对比实验验证了所提方

法的优越性和有效性。

第四章：基于多模特征解耦的体育视频时空动作检测方法。针对多人体育动作情况的关系信息挖掘和体育运动视频的细粒度时空特征建模，本章提出了一种利用多模态大模型提取多模态特征构建多模态特征库的方法，以提升模型对复杂体育动作的理解能力。随后，为了使多模态特征可以更好地适配时空动作检测任务，设计了一种双分支时空特征解耦模块，以分别空间分支通过空间编码器提取关键帧的空间特征，时序分支通过时序编码器提取时序特征，并分别通过时间对齐和空间对齐模块进行特征对齐，显著提升了模型对细粒度动作的检测能力。最后，通过大量实验验证了所提方法在体育视频时空动作检测任务中的有效性和优越性。

第五章：总结与展望。本章总结了论文的主要内容和创新点，指出了当前工作的不足之处，并提出了未来研究的方向。

## 2 体育动作时空动作检测相关理论和技术

### 2.1 引言

本章首先明确了时空动作检测的任务定义，说明其相比于动作检测和时序动作检测任务的优势与挑战，阐释了其在视频中实现分类、时序定位和空间定位的任务目标（第 2.1 节）。随后，本章分析了现有视频特征提取方法的常用基础模型，并说明了其各自优缺点（第 2.2 节）。然后，基于现有的检测方法范式，本章分析了双阶段检测范式和单阶段检测范式各自的优势，并重点说明了基于查询机制的动作检测器的理论基础和其在时空动作检测任务中的使用（第 2.3 节）。最后，本章对相关领域的基准数据集和评价指标进行了介绍（第 2.4 节），并具体分析了体育场景下的时空动作检测的难点（第 2.5 节），并对本章内容进行了总结（第 2.6 节）。

### 2.2 时空动作检测任务定义

从任务目标来看，时空动作检测是动作分类和时序动作检测的进一步延伸。动作检测任务主要关注对一段指定视频数据中所包含的动作类别进行分类，通常假设视频数据已经经过剪辑处理，确保视频中仅包含单一动作类别的信息，因此动作检测任务的主要目标是识别视频中所包含的动作类别。时序动作检测任务则进一步要求模型不仅能够识别视频中的动作类别，还需要对动作发生的时间段进行定位，即确定动作的起始时间和结束时间。然而，时空动作检测任务在两者的基础上更进一步，通常处理的是未经剪辑的长视频数据，视频中可能包含多个动作类别，并且这些动作可能在时间和空间上交织在一起。因此，时空动作检测任务的目标不仅包括动作分类和时序定位，还需要实现对动作在空间维度的定位出特定的运动动作片段。

具体来说，为了统一表达，假设视频序列为  $V = \{I_t\}_{t=1}^T$ ，动作类别集合为  $\mathcal{C}$ 。对于动作检测来说动作识别任务目标是判定整段（或已剪辑好的）视频所属的类别，不涉及具体的时间定位和空间定位。

$$\Phi(V) = c_i \quad (2.1)$$

其中  $c_i \in \mathcal{C}$ 。对于时序动作检测任务目标是确定输入视频片段中动作发生的时间范



图 2.1 时空动作检测任务目标：从未剪辑视频中获取完成动作分类、时间定位和空间定位围（何时发生）以及动作类别。

$$\Phi(V) = (c_i, t_b, t_e) \quad (2.2)$$

其中  $t_b$  为开始帧， $t_e$  为结束帧。时空动作检测根据图中的定义，该任务需要同时确定动作的类别、时间范围以及每一帧中的空间位置（何处发生）。

$$\Phi(V) = (c_i, \{R_t^i\}_{t=t_b}^{t_e}) \quad (2.3)$$

$$\{R_t^i\}_{t=t_b}^{t_e} = \{(x_{min}, y_{min}, x_{max}, y_{max})_t \mid t \in [t_b, t_e]\} \quad (2.4)$$

其中  $c_i \in \mathcal{C}$  是动作标签， $\{R_t^i\}$  是从开始时间  $t_b$  到结束时间  $t_e$  每一帧图像  $I_t$  中对应的边界框或区域集合。

在图2.1 具体展示了时空动作检测相比于动作检测和时序动作检测的任务目标的区别。根据任务目标的差异，可以看出时空动作检测相较于动作检测和时序动作检测具有更高的复杂性和挑战性，需要模型具备更强的时空理解能力和精确的定位能力，而由于其能够提供更丰富的动作信息，因此在实际应用中具有更广泛的适用性和价值。表2.1 对比了三种任务类型及其典型应用场景。

表 2.1 动作检测任务类型及其适用场景对比

任务类型	任务目标	典型应用场景
动作识别	动作类别	视频分类、内容审核
时序动作检测	动作类别 + 时序定位	视频检索、录像回溯
时空动作检测	动作类别 + 时序定位 + 空间定位	高光视频生成、运动竞技分析

## 2.3 视频特征提取方法

动作识别、时序动作检测、时空动作检测这三者的核心均依赖于对视频数据中时空特征的提取与建模。时空特征融合了视频的空间信息（如目标姿态、场景布局）与时间信息（如运动轨迹、动态演变），为不同任务提供基础表征。尽管任务的目标各异，这些方法常共享相同的骨干网络，本小节将从 3D CNN 和 Vision Transformer 两大主流框架介绍视频特征提取方法的具体原理。

## 2.4 时空动作检测方法范式

### 2.4.1 双阶段时空动作检测方法

控制器设计的详细叙述……

### 2.4.2 单阶段时空动作检测方法

稳定性分……

### 2.4.3 基于查询的时空动作检测方法

## 2.5 相关数据集和评价指标

### 2.5.1 相关数据集

### 2.5.2 评价指标

## 2.6 本章小结

本章主要介绍系统与控制理论类论文正文章节的框架结构。在每章的最后，都需要对该章的内容进行小结，不宜太长，建议 1/2-2/3 页版面较好。主要小结一下本章用什么理论或方法、做了什么事、得到的重要结果或结论。

### 3 理论/算法类研究类论文

#### 3.1 引言（引言标题可选）

无论什么类型的论文，在章节的标题下，都需要简要说明本章研究的背景或动机，以起到承上启下的作用。若少于 1 个页面时，建议省略标题“引言”，直接在章的标题下写上几段话即可。

对于非会议、期刊的信息来源，若为网址，请在当页中脚注中加以标注\*。

理论研究类论文，一般包括原理介绍、理论推导/或算法设计思想，再通过模拟仿真给出结果。该类论文若提出的是新理论或算法，一般应与现有理论/算法比较。当然也可以通过实验加以验证，以评估其准确性。具体内容应包括以下几个部分：

#### 3.2 \*\* 理论/算法

#### 3.3 \*\* 仿真或算法实现

#### 3.4 理论/算法准确性的评估

#### 3.5 分析与讨论

#### 3.6 本章小结

本章简要给出理论/算法类研究论文的基本框架。在每章的最后，都需要对该章的内容进行小结，不宜太长，建议 1/2-2/3 页版面。主要小结一下本章用什么理论或方法、做了什么事、得到的重要结果或结论。

## 4 学位论文写作细则

学位论文很多的错误源于凌乱的格式。

为规范学位论文写作，本章结合工科学位论文的特点，参照学术出版规范，就图、表制作，名词、术语、单位、符号、数量等的使用规范化进行了说明。

（论文中所有图表清晰，字体、格式一致，图标题在图下方，表标题在表上方。  
图表引用和图表不跨页，确有需要跨页不超过 1 页）

### 4.1 关于图

参照 CY/T 170-2019《学术出版规范插图》标准执行。

论文中的图一般居中放置，大小要合适，宽度不得超过文字边缘，图中文字一般不大于五号；图像分辨率应尽量不低于 300 dpi；若为自画图，尽量采用中文标识，图中文字经缩放后，字号不得大于五号；图注也只需加注中文，宋体，五号。若引用文献中的图，应在图注最后标注引用，图注后无需加标点符号。文中所有的图都应予以编号，图序号按“章”进行编号，如图4.1所示：制作图时请注意：

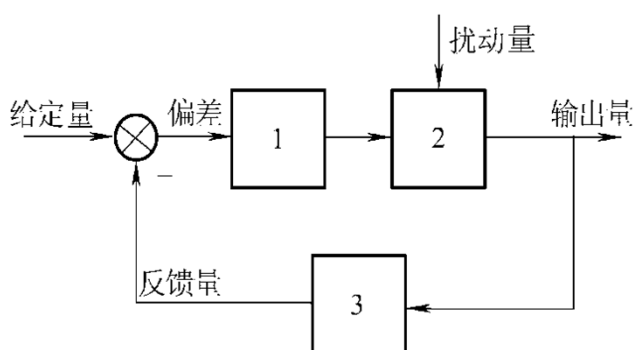


图 4.1 闭环系统示意图

- （1）内容与形式应力求统一。风格、体例应一致。线条图应清晰，线型选用、线条粗细应规范，色调准确，图形布局合理、大小适当。连续色调图应清晰，层次感和饱和度适当。
- （2）每幅都得有图题。图题应准确、简明地阐释该图内容。

- (3) 图应与正文内容相关。应选择能有效传达关键信息的图，应具有自明性、简明性、科学性和艺术性。
- (4) 结构示意图、原理示意图和流程图的设计制作应符合现行的国家标准或行业标准。
- (5) 坐标曲线图的坐标轴、标值线的画法应规范，标目、标值、坐标原点应标注完整、规范、统一。
- (6) 图中涉及标志用图形符号、设备用图形符号和技术文件用图形符号应符合现行的国家标准。图中的术语、数值、符号等应与正文其它图中的表述一致。
- (7) 图应尽量与相应的文字靠近，根据排版可放在文字的下方或上方。各章节不能因为图的原因出现大段留白的现象。
- (8) 图中应有对于需特别说明问题部分的标识说明，如图像中需给出比例尺，在彩图中给出 Color bar 等。说明字体与比例尺的字体应至少比正文小一号。

## 4.2 关于表格

学位论文中的表格要使用三线格，居中放置。表格题目应居中放置于表格顶端。仅提供中文表头，表头及表中文字应小于正文字号，一般为宋体，五号；若其中含英文或数字时，字体为 Times New Roman，五号。表格题目与表格尽量不要分页。实在需要分页时，请在下一页续表。若该表完全源于文献，应在表格题目最后标注引用。标头最后不加标点。如表4.1所示：

表 4.1 表格示例

Aaa	Bbb	Ccc	Ddd	Eee
1-1	1-2	1-3	1-4	1-5
2-1	2-2	2-3	2-4	2-5
3-1	3-2	3-3	3-4	3-5
4-1	4-2	4-3	4-4	4-5

一般情况下，表应放在文字的下面，有时依据排版的需要，也可以放在文字上方。但无论如何，图应尽量与文字靠近。

- (1) 表格序号按“章”进行编号，如第3章第1个表为“表3-1”，以此类推。

- (2) 表格内容与正文配合应相得益彰，内容适合用表表达。
- (3) 表格应具有自明性和简明性，栏目设置应科学。
- (4) 表格中的数据应具有完整性和准确性。
- (5) 表格中连续数的分组应科学，不得重叠和遗漏。
- (6) 表格中的数值修约和极原数值的书写应符合 GB/T 8170 的规定。
- (7) 表格中的量和单位的名称、符号及书写应符合国家标准 GB3100~3102—93 量和单位的系列标准和相关行业标准。
- (8) 表格中数字形式的使用应符合 GB/T 15835 规定。
- (9) 表格中的科学技术名词应符合 CVT 119 的规定。
- (10) 表格中的术语、数值、符号等应与正文以及同一文本中其他表格中的表述一致。
- (11) 全书的表格的表号、表题，表头、表身、表注的格式应统一。

## 4.3 名词、术语

学位论文存在大量的专业名词与专业术语，针对同一内容的名字、术语有多种表达方式时，原则上以关系最为密切的学科为准，并尽量符合最新的国家或行业规程、规范。若属全国科学技术名词审定委员会最新公布各类学科的“名词”，则须严格执行，且在全文中统一。

外国专有名称在释文中首次出现时，应附原文和简称。例如：“美国垦务局（United States Bureau of Reclamati, USBR）”、“美国大坝委员会（United States Committee On Large Dams, USCOLD）”。在同一条目中再次出现时可采用中文或英文缩写。

## 4.4 符号、单位的使用

标点符号的使用，应符合国家标准 GB / T 15834—2011《标点符号用法》。

- (1) 论文中的计量单位应采用法定计量单位（简称法定单位），应符合国家标准 GB 3100~3102—93 量和单位的系列标准和相关行业标准。
- (2) 表格及插图中，使用单位符号，不使用单位名称和单位中文符号。叙述性文字中，优先使用单位符号。必要时，可使用单位名称，但不可使用单位中文

符号。如：流量为 11400

- (3) 两个物理量（量值加单位）在表示范围时，两个量值用波浪线“~”连接后使用一个计量单位，如：应写作 800~1500 m<sup>3</sup>/s，而不写作 800m<sup>3</sup>/s~1500 m<sup>3</sup>/s。

## 4.5 数字的使用

数字的使用应符合国家标准 GB/T15838—2011《出版物上数字用法》，同时应符合相关行业标准。

以下情况应使用阿拉伯数字：

- (1) 统计表中的数值，如：正负整数、小数、百分数、分数、比例。
- (2) 物理量量值，如：150 m<sup>3</sup>/s，200 kg，注意数量与单位之间应有空格。
- (3) 非物理量量值。如：21.35 元，480 人。
- (4) 当阿拉伯数字与汉字数字混用时，要顾及上下行文的协调一致。
- (5) 两个百分数表示范围时，要使用两个百分号，如 15
- (6) 专业性科技出版物上的多位数字，应从小数点算起，每三位留空半个数码位置，不采用传统的以千位撇“，”分节的办法，如 3800000 应写成 3 800 000，或写成 380 万，而不要写成 3，800，000。

以下情况应使用汉字数字：

- (1) 定型的词、词组、成语、惯用语或具有修辞色彩的词语中作为词素的数字。  
如：星期六、四氧化三铁、五省一市、“八五”计划、第三季度等。
- (2) 相邻两个数字并列连用表示概数，连用的两个数字间不得用顿号“、”隔开。  
如：二三米、十三四吨、一千七八百元。
- (3) 不是出现在具有统计意义的一组数字中的整数一至十。如：一个人、三本书、五个百分点等。
- (4) 带有“几”的数字，表示概数。如：十几天、几千年等。

## 4.6 其它应该注意的问题

### 4.6.1 关于文献引用

在学位论文撰写中，凡是文字、图、表来自于参考文献的，必须要加注文献信息。国际惯例：与他人文章有 20 个字连续雷同的就是抄袭。抄袭是必须严禁的行为，

否则害人害己。建议大家最好用自己的理解来撰写相关内容，特别是第一章绪论中更应注意。

在引用参考文献时，不要将文献标识直接写在某一小节的标题上，这将被视为整节内容都是引用；如有某几句话完全来自文献的，则在这部分内容的最后一句的结束加文献标注。

在课题组内部，经常出现多位同学做相近的课题，或课题的关联性较大。每个学生介绍研究背景或实验室材料与方法时，请尽可能用自己理解后的语句去撰写，以避免雷同，千万不要图省事将前一届学生的论文大段地照搬过来。有些图若是能自己画的，最好自己重新画一遍，以示区分。

硕士学位申请人的文献阅读量一般不少于 40 篇，其中外文文献一般不少于 1/3；近五年的论文一般不少于 1/3；绪论部分应对所读文献加以分析和综合。

## 参考文献格式

中文书刊：作者按中文写法，姓在前、名在后；英文书刊：作者按英文习惯写法，如名在前、姓在后，名用首字母缩写、姓用全称。一般 6 人以内须列出全部作者，6 人以上写 6 人再加“等”（英文加“et al”）。每个参考文献的最后不加标点符号。

(1) 图书：最多列出 6 个作者，作者与作者之间用逗号分隔. 书名. 版本（第 × 版）. 译者. 出版地: 出版者, 出版年. 起页-止页（可选）

(2) 期刊：最多列出 6 个作者，作者与作者之间用逗号分隔. 文章名. 期刊名（全称）. 年号, 卷号（期号）: 起页-止页或论文编号

(3) 会议论文集：最多列出 6 个作者，作者之间用逗号分隔. 文章名. 见（英文用“in”）: 会议名称（或论文集）. 会议城市, 国家, 会议时间, 出版者, 出版年: 起页-止页

(4) 专利：专利申请者. 专利题名. 专利国别, 专利文献种类, 专利号, 出版年

(5) 学位论文：作者. 题名: [博士（或硕士）学位论文]. 保存地点: 保存单位（如华中科技大学, 年份）

## 参考文献（举例）

[1] 闫明礼, 张东刚. CFG 桩复合地基技术及工程实践（第二版）. 北京: 中国水利水电出版社, 2006

[2] M. Chalfie, S. R. Kain. Green fluorescent protein: properties, applications, and protocols. Hoboken, New Jersey: Wiley-interscience, 1998

- [3] 詹向红, 李德新. 中医药防治阿尔茨海默病实验研究述要. 中华中医药学刊, 2004, 22(11): 2094-2096
- [4] E. S. Lein, M. J. Hawrylycz, N. Ao, M. Ayres, A. Bensinger, A. Bernard, et al. Genome-wide atlas of gene expression in the adult mouse brain. Nature, 2007, 445(7124): 168-176
- [5] M. L. Bouxsein, S. K. Boyd, B. A. Christiansen, R. E. Guldberg, K. J. Jepsen, R. Müller. Guidelines for assessment of bone microstructure in rodents using micro-computed tomography. Journal of Bone and Mineral Research, 2010, 25(7): 1468-1486
- [6] Y. Shunsuke, A. Masahide, K. Masayuki, M. Yoshizawa. Performance evaluation of phase-only correlation functions from the viewpoint of correlation Filters, in: 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Honolulu, HI, USA, 12-15 Nov. 2018, Proceedings of the IEEE, 2019: 1361-1364
- [7] T. Yao, J. Wan, P. Huang, X. He, F. Wu, C. Xie. Building efficient key-value stores via a lightweight compaction tree. ACM Transactions on Storage, 2017, 13(4): 1-28
- [8] 刘加林. 多功能一次性压舌板: 中国, ZL92214985. 2 [P]. 1993
- [9] 李清泉. 基于混合数据结构的三维 GIS 数据模型与空间分析研究 [博士学位论文]. 武汉: 武汉测绘科技大学, 1998

注释体例的基本内容、结构与位置

(1) 基本内容与结构

“注释体例”含“资料性注释”和“内容性注释”两方面，合一编排。

(2) 位置

正文内需注释之处依次排注号，释文于当页下部逐条依次编排。可在正文和页下注之间划一道分隔线，或通过不同的字体将二者区分开来。

(3) 排版

【字体】中文：小五，宋体；英文：times new roman 9 号字体；

【行距】单倍行距；

【段落】顶格写，无首行缩进，也无左缩进；

【序号】用 □ 这种格式，序号后空一个字符，每页重新编序；

【页码】中文：第 x-x 页，如第 16-17 页。英文：pp.x x，如 pp.5 8, 单页用 px，如 p19.

【标点符号】中文使用中文状态下标点符号，英文使用英文状态下标点符号，切忌混用。

以下是几个例子, 注意, 中文参考文献需要额外增加一个 lang="chinese"

期刊论文举例:

Pan 等人提出了一种全新的算法<sup>[? ? ?]</sup>。

会议论文举例:

XXXXXX<sup>[?]</sup>

图书举例:

XXXXXX<sup>[? ?]</sup>

学位论文举例:

XXXXXX<sup>[? ?]</sup>

## 4.6.2 公式

公式中主要字母的字号与正文一样（即小四号），尽量避免图片转贴过来的公式。公式编号根据章节按顺序进行编排，如第 2 章第 3 个公式，标注为 2-3，将公式编号以右对齐方式排列，但注意公式是以居中的格式排列的（类似教材中的通用格式）。

## 4.6.3 本章小结

本章主要介绍学位论文写作的规范化要求，包括图、表制作及其与正文的对应关系；专业名词、术语、单位、符号及数字的使用等。

## 5 总结与展望

### 5.1 本文主要内容及结论

对全文进行全面地总结，并根据各章节归纳出若干有机联系的论点。按正文的内容分段描述，包括本研究“做了什么（提出\*\*新理论/算法、设计或研发\*\*工艺/仪器）、获取什么结果、得出什么结论”。

请特别注意，全文总结与摘要及各章的小节要有所区分，不能简单的拷贝。这里的重点是结论，结论应该准确、完整、明确、精练。

### 5.2 本文主要创新点

通常情况下，学位论文的创新点应放在最后一章。

创新点要凝炼，表述要清晰明了，如提出了什么创新的思路，主要特点是什么，相比现有理论或技术的提高是什么、或者有什么新的发现，是否具有重要的科学意义或应用前景。既不能过于简单，也不要太细。

硕士学位论文创新点不宜太多，一般为2个左右即可，要注意归纳创新点，千万不要以为越多越好。论文的创新不以创新点的多少来评定的，而以其创新性的价值来评定。几章的工作合在一起凝炼成一个创新点也不是不可以的。

### 5.3 展望

对本研究成果的意义、推广应用的现实性或可能性加以论述。同时，描述本文研究中尚存在的不足，或因时间尚未完成但又必须继续的工作，对进一步的工作进行展望。

## 致 谢

对在课题研究及论文写作过程中给予指导和帮助的导师、校内外专家、实验技术人员、同学等表示感谢。

在致谢时建议具体，不同的人如何助力完成你的论文，都需要特别注明。如导师、其他老师或实验技术人员、以及同学对你论文的贡献是不一样的，有指引课题方向、修改论文，也有具体教会实验操作，也有协助你做了哪方向的实验，或者给你精神安慰、陪你度过紧张的研究生生涯。

越具体越能表达你真实的感受，否则就是毫无意义的套话。

## 附录 1 攻读硕士学位期间取得的研究成果

### 发表与接收论文

- [1] Linqiang Pan, **Lianghao Li**, Ran Cheng, Cheng He, Kay Chen Tan.[J]. IEEE Transactions on Cybernetics, vol. 58, no. 6, pp. 3325-3337, 2019. (SCI 源刊; IF:11.448; 署名单位: 华中科技大学)
- [2] 参照参考文献列出学术论文相关信息（含期刊、会议、或参编书稿），但无论有多少个作者，都必须列出全部作者名；若为英文论文，则名在前、姓在后，姓名均为全称；在本人的名字加粗，以示区别（若为第一作者，则需在最后特别注明署名华中科技大学是否为第一单位）
- [3] 若已发表，按参考文献给出页码；若只是 online, 给出链接；若接受或修改或投稿或拟投，也必须分别注明
- [4] 一般情况，一作或重要的论文放在前面

### 专 利

- [1] 全部作者的姓名全称，本人的名字加粗. 专利题名. 专利国别，专利文献种类，专利号或申请号

### 标 准

- [1] 全部作者的姓名全称，本人的名字加粗. 标准题名. 哪种层次的标准，发表年

### 科技奖励

- [1] 全部作者的姓名全称，本人的名字加粗. 题目. 国家级/省部级科技类奖，获奖年
- [2] 全部作者的姓名全称，本人的名字加粗. 题目. 国际/国内竞赛类奖，获奖年

## 附录 2 攻读硕士学位期间参与的科研项目

### 1. 项目类型

项目名称: 项目名称

项目编号: No. 88888888

起止时间: 2018 年 8 月至 2018 年 8 月

担任角色: 担任角色

### 附录 3 其他附录

可包括详细的公式推导、实验数据、计算程序、援引他人的原始资料、数据及其设备条件等。