

华南理工大学硕士学位论文

基于渐进式策略的人体运动姿态预测算法

马铁铮

指导教师：聂勇伟 副教授

华南理工大学

2023 年 5 月 20 日

摘 要

3D 人体运动估计 (3D Human motion prediction) 指: 在 3D 空间中, 根据历史人体运动姿态序列, 预测未来的人体姿态运动序列。随着人工智能化浪潮的到来, 该技术被广泛应用于自动驾驶、监控视频异常检测、人体动作捕捉生成等领域中, 有着良好的应用前景和研究价值。例如在自动驾驶算法中, 需要根据行人当前运动轨迹来预测其未来运动趋势, 进而指导自动驾驶程序做出相应处置。

本文提出了一种新颖且高效的 3D 人体运动估计算法, 与现有方法相比, 本方法在预测精确度和运行效率的综合指标上有较大的领先。具体地, 我们在分析总结现有方法优劣的基础上提出了两个改进策略: (1) 3D 空间中的人体运动存在高度的复杂性和不确定性, 给预测过程带来了很大的难度。现有方法往往采用单个网络直接预测。在运动模式简单, 周期性的样本上能达到较高的预测精度。但在处理较为复杂且无明显规律的动作类型时, 往往出现模式坍塌、预测失准等情况。为此我们提出了一种渐进式的网络结构来降低整体的预测难度: 网络由多个阶段构成, 每个阶段不再直接预测最终结果, 而是在上一阶段的基础上完善预测结果。浅层的阶段负责预测运动的大概趋势, 深层阶段则在此基础上完善预测细节, 使预测结果向真值逐渐靠拢, 同时减少各个阶段的预测难度。(2) 3D 人体运动数据同时具有时间和空间两维度, 且人体姿态为无向不规则图结构, 其内涵的空间先验结构信息极为重要。所以, 网络中, 特征提取模块的时空信息提取能力与网络性能密切相关, 现有方法大多使用 *CNN* (卷积神经网络)、*RNN* (循环神经网络)、传统 *GCN* (图卷积网络), 该类方法只适用于处理 2 维空间数据, 难以捕捉时序联系。为此, 我们提出了一种具有时空信息捕捉能力的 *GCN* 模块, 该模块由空间和时间两部分构成, 分别称为 *SD-GCN* (*Spatial Dense Graph Convolution*) 和 *TD-GCN* (*Temporal Dense Graph Convolution*), 两部分串行组合, 当运动序列输入后, 首先由 *SD-GCN* 提取空间信息, 随后送入 *TD-GCN* 提取时间信息, 由此网络间接地捕捉了时空信息, 并具有全局感受野。

在渐进式结构和 *SD-GCN*、*TD-GCN* 这两点改进措施的帮助下, 本方法在 *Human3.6M*、*CMU-MoCap*、*3DPW* 这三个公开数据集上使用公开度量指标, 预测精度较现有方法均有较大提升, 且运行效率无显著落后。

关键词: 3D 人体运动估计、渐进式策略、时空序列、图卷积网络

Abstract

Keywords: 3D Human Motion Prediction、Progressive Learning、Spatial Temporal Sequence、Graph Convolutional Networks

目 录

摘 要	I
Abstract	II
插图目录	VI
表格目录	VIII
第一章 绪论	1
1.1 研究背景和意义	1
1.2 主要研究内容及贡献	2
1.3 论文结构	3
第二章 相关工作	5
2.1 基于循环神经网络	5
2.2 基于卷积神经网络	7
2.3 基于图卷积网络	8
2.4 基于对抗生成网络	9
2.5 基于 Transformer	10
2.6 总结	12
第三章 相关基础知识	13
3.1 图卷积网络	13
3.2 渐进式策略在视觉领域的应用	13
3.3 Transformer 全局注意力机制	13
第四章 基于渐进式策略的多阶段人体运动姿态预测框架	14
4.1 数据描述与问题定义	14
4.1.1 人体运动姿态数据结构	14
4.1.2 人体运动姿态预测问题定义	15
4.2 渐进式人体运动序列预测框架	15
4.2.1 渐进式多阶段预测网络框架	18
4.2.2 基于累积均值平滑的中级监督目标	19
4.2.3 总结	22

第五章 基于时空分离策略的 Non-Local 时空图卷积模块	24
5.1 时空图卷积模块设计思路对比	24
第六章 模板简介	30
6.1 主文件	30
6.2 章节文件	33
第七章 常用环境及参考文献设置	34
7.1 图	34
7.2 表	37
7.3 公式	38
7.4 定理	41
7.5 参考文献	42
第八章 列举环境	57
8.1 调整间距	57
8.1.1 垂直间距	59
8.1.2 水平间距	59
8.2 enumerate 标签样式	61
8.2.1 小括号阿拉伯数字	61
8.2.2 斜体字母	61
8.2.3 大写罗马字母	61
结 论	62
参考文献	63
附 录 1	69
1.1 测试一级标题 section	69
1.1.1 测试二级标题 subsection	69
1.2 测试测试测试	70
1.2.1 测试测试测试	70
附 录 2	72
2.1 测试测试测试	72
2.1.1 测试测试测试	72
攻读博士/硕士学位期间取得的研究成果	75

致 谢	76
-----------	----

插图目录

图 2-1	EDR 网络结构	5
图 2-2	Res. Sup. 网络结构	6
图 2-3	LTD 网络结构	8
图 2-4	Spatial-temporal Transformer ^[52]	11
图 2-5	Spatial-temporal Transformer ^[52]	11
图 4-1	人体运动姿态数据结构	14
图 4-2	LTD 数据填充过程	16
图 4-3	验证实验	16
图 4-4	Coarse To Fine 预测网络	17
图 4-5	渐进式多阶段的预测网络	19
图 4-6	累积均值平滑对比高斯滤波	21
图 5-1	人体运动姿态数据时空结构	25
图 5-2	使用 GCN 同时对时空维度建模	25
图 5-3	LTD 中的图卷积模块	26
图 5-4	ST-GCN 中的图卷积模块	27
图 5-5	STS-GCN 中的图卷积模块	28
图 5-6	基于时空分离策略的 Non-Local 时空图卷积模块	28
图 6-1	TeXstudio 环境	31
图 6-2	TeXstudio 编译选项	31
图 7-1	涵道风扇式无人机	35
图 7-2	T-Hawk	36
图 7-3	GTSpy	36
图 7-4	子图包使用测试	37
图 7-5	常规	44
图 7-6	同步 1	45
图 7-7	同步 2	46
图 7-8	搜索	47
图 7-9	导出	48

图 7-10	引用	49
图 7-11	高级 1	50
图 7-12	高级 2	51
图 7-13	Better BibTeX1	52
图 7-14	Better BibTeX2	53
图 7-15	Better BibTeX3	54
图 7-16	导出文献库	55
图 7-17	导出格式	55
图 7-18	导出文件名	56
图 8-1	enumitem 包对各种间距的定义	58
图 1-1	测试测试测试	70
图 1-2	测试测试测试	70
图 2-1	测试测试测试	73
图 2-2	测试测试测试	73

表格目录

表 7-1	涵道模型参数	38
表 1-1	测试测试测试	71
表 1-2	测试测试测试	71
表 2-1	测试测试测试	74
表 2-2	测试测试测试	74

第一章 绪论

1.1 研究背景和意义

近年来随着人工智能技术和社会经济的高速发展，大量信息化、智能化的新技术渗透到了人们的大众生活中。其中理解和预测人体运动相关研究获得了显著的进展。该技术被广泛应用与自动驾驶、智能机器人、人机交互和多媒体领域。在自动驾驶领域，车载计算机需要预测其他交通参与成员的行动意向和未来位置，并以此来规划车辆未来运行路线。在智能机器人领域，特别是用于协助人类的机器人，如工业机器人、看护机器人等，需要准确地预测人的未来运动来采取对应行动。在人机交互领域，在人口稠密的空间中，机器应准确预测周围的人的动作以安全地穿过人群。在多媒体领域，特别是游戏和影视制作场景中，和通过昂贵的动作捕捉设备获得人体运动模型相比，基于软件的理解和预测人体运动方法更加廉价高效。综上，理解和预测人体运动算法在促进国民经济发展和数字化、智能化转型方面有较高的研究价值。

目前学术界和工业界对该课题进行了较为细致且充分的研究。人体运动预测问题被定义为：在某个三维场景下，已知某个个体的一段历史运动序列，需要根据该段历史运动中包含的趋势或规律，预测该个体在未来的运动序列。该问题的研究重点包含两部分，第一是通过对历史运动序列的理解，提取其中包含的运动信息。例如在观看任意一段运动序列后，人类可以轻易地判别出该序列的运动类型（如行走、拾取物品、舞蹈等）。但对计算机来说，如何理解运动序列中的时序信息是研究的重点。第二是基于对历史运动序列信息的提取，预测未来运动序列。由于人体运动的高度复杂性和不确定性，如何基于有限的运动信息尽可能降低预测过程的不确定性，从而输出准确的未来运动序列，是当前研究的一个主要难点。

在早期的研究中，由于循环神经网络（RNN^[1]）可以利用其内部隐状态（Hidden State）来捕捉输入数据的时间依赖性，对于处理时间序列这类连续数据特别有效，所以 RNN 被用来提取人体运动序列中的时序信息，预测未来的运动序列。

这类方法的主要思想是，每个 RNN Unit 有一个隐状态，可以在每个时间步骤中根据当前的输入的人体运动姿态和以前的隐状态进行更新。这个隐状态是对网络过去所见信息的总结，并被用来对未来进行预测。通过使用以前的隐状态来计算当前的隐状态，RNN 可以捕获输入序列的时间依赖性。这使得 RNN 可以提取关于序列结构和序列元素之间的依赖关系的信息，这对于时间序列预测任务特别有利。

然而，由于 RNN 中每一步输出只与当前输入和上一步隐状态有关，无法对每一步输出进行整体约束。这导致输入序列和预测序列的过渡部分出现不连续的情况。为此现有方法提出了一种有着编码器-解码器结构的序列到序列模型（Sequence-to-Sequence），编码器将输入数据整体映射到隐空间，随后由解码器一次性预测未来运动序列，由此可以对输出进行全局一致性约束。此外，隐状态容量有限，RNN 只能对短期依赖性进行建模，无法处理长距离时序依赖。这导致网络无法完全提取输入序列中的时序信息。为了解决这个问题，出现了长短期记忆（LSTM^[2]）和门控循环单元（GRU^[3]）网络，但它们更加复杂和计算量更大。最后，该类方法通常将一个人体姿态作为一个整体输入 RNN Unit，忽略了人体姿态的空间结构。然后，对于人体运动序列预测问题，人体姿态的空间结构是一个重要的先验信息。这导致基于 RNN 的方法在预测结果真实性和准确性方面有所欠缺。

近年来随着对图卷积网络（GCN^[4]）的深入研究，部分现有方法引入图卷积网络对人体姿态空间结构进行建模。对于人体姿态这类不规则图状数据，图卷积网络有着天然的优势。在这类方法中，人体姿态被视作由一组顶点（或节点）和连接一对顶点的一组边组成的数学结构。通过图卷积网络对复杂的关节点对之间的联系进行建模。但传统图卷积网络主要应用于空间维度，如何设计高效的，具有时空信息提取能力的图卷积网络，对于人体运动序列预测这类涉及到时空序列数据的问题尤为重要，至今也依旧是学术界的一个难点问题。

除了需要尽可能提取输入序列中的时空信息，如何降低预测过程中的不确定性也是需要考虑的问题。在大多数情况下，由于人体运动序列的复杂性，输入序列和未来序列之间存在较大的差异，这导致预测过程存在较大的不确定性和预测歧义。现有方法大多采用单个前馈神经网络，直接接受输入序列，并预测未来运动序列。网络在预测过程中将承受较大的不确定性，预测结果可能出现模式坍塌（Mode collapse）问题。因此，如何设计更高效的预测策略来降低预测过程中的不确定性和歧义，是当前研究中需要重视的问题。

1.2 主要研究内容及贡献

针对上述研究存在的问题和人体运动姿态预测问题的特点，本文主要的研究内容和贡献被总结如下：

1. 基于渐进式策略的人体运动姿态预测算法框架。

- 与现有方法使用单阶段的网络结构不同，我们由浅至深地将预测过程拆分为多个阶段，除开位于网络入口的阶段，其他阶段均在上一步预测基础上进行预测，这将有利于降低每一阶段的预测难度。
- 我们遵循网络由浅至深，预测难度由易到难的原则。浅层阶段只负责预测大致的运动趋势，复杂的运动细节预测则由具有深层语义提取能力的深层阶段负责。
- 为了构建多阶段、渐进式的网络，我们为每个阶段构造对应的中级监督目标（Intermediate target）。具体的，我们设计了一种人体运动轨迹平滑方法，通过平滑关节点运动轨迹的方式，逐步去除运动细节，为每个阶段由深至浅提供不同平滑程度的预测目标。

2. 具有时空信息提取能力的 Spatial-temporal 图卷积网络模块（*SD-GCN* 和 *TD-GCN*）。

- 我们提出了一种新颖的具有时空信息提取能力的图卷积模块，该模块由时间信息提取模块和空间信息提取模块两个独立的图卷积构成。
- 时间信息提取模块称为 *TD-GCN*，输入数据被视为多个关节点轨迹，*TD-GCN* 提取时间维度上的信息。空间信息提取模块称为 *SD-GCN*，输入数据被视为多个人体姿态，*SD-GCN* 提取空间维度上的信息。两个模式以串联的方式构成一个时空图卷积模块，当数据依次通过二者时，网络间接地提取到了时空信息。
- 由于时间和空间信息提取模块相互独立，因此随着输入数据的时间长度和空间复杂度提高，模型空间复杂度只随线性增长而非倍数增长，在保证模型信息提取能力的同时，降低了时间效率。

3. 我们在三个公开数据集上使用通用度量指标，与现有先进方法进行对比。在预测精确性方面大幅领先（Human3.6M 6%-7%，CMU-MoCap 5%-10%，and 3DPW 13%-16%）。并且在时间效率和内存占用指标上我们也处于靠前位置。

1.3 论文结构

本文包括七个主要章节，包含绪论、相关工作、基础知识、渐进式策略算法框架、Spatial-temporal 图卷积网络模块、对这两个模块的进行的实验分析，以及对全文进行总结与展望。其中各个章节的主要内容安排如下：

首先，在第一章，对 3D 人体运动估计问题的研究背景和研究意义进行详细阐述。

其次概述本文的主要研究内容和贡献。

第二章详细介绍了 3D 人体运动估计问题的研究现状和发展历程，对当前研究的参考价值。

第三章主要介绍本文所提出方法中使用到的相关技术和理论基础，为详细叙述新方案打下基础。

第四章提出了渐进式策略算法框架，本文将从实验和直观分析的角度来叙述该设计的合理性和有效性，并且我们还展示了该渐进式策略算法框架具有高度的灵活性，可以与任意网络模块进行组合以适应不同的任务种类。

第五章提出了 Spatial-temporal 图卷积网络模块，在这里我们将详细介绍该模块中的时间和空间部分，以及它们是如何整合为一个整体。

第六章对我们提出的基于渐进式策略的人体运动姿态预测算法在三个公开数据集（Human3.6M、CMU-MoCap、3DPW）上与现有先进方法进行定性和定量的对比。同时，通过消融实验对模型中的各个模块进行定量分析。另外对模型中的一些有趣的细节进行了充分讨论。

最后一章是对本文进行总结与展望，对本文提出的基于渐进式策略的人体运动姿态预测算法进行概略性的总结。同时分析本方法当前的不足，对后续研究提供指导性意见。

第二章 相关工作

近十年, 3D 人体运动姿态预测算法受到了广泛的研究和探讨, 涌现了一大批出色的工作。根据其对人体运动序列的建模方式不同, 现有方法可以分为以下几类: 基于循环神经网络的方法 [5-18]、基于卷积神经网络 (包含 CNN 与 GCN) 的方法 [19-35]、基于对抗生成网络的方法 [9,36-43]。在研究早期, 由于人体运动的序列化特征, 大部分方法使用循环神经网络对输入数据进行建模, 然而循环神经网络的时序记忆能力受限于隐变量的大小, 只能处理短期记忆, 无法处理较长时间的序列。随后出现了一批由卷积神经网络构成的模型, 其中包含 CNN 和 GCN 网络, 前者与 RNN 相比拥有更大的感受野, 这提高了网络的长时序依赖捕捉能力。而 GCN 则更适合处理人体姿态这类不规则的空间数据, 能够感知人体结构先验信息。对抗生成网络近些年也被引入该领域, 对抗生成的策略能够提供在真实性和多样性方面占优的结果, 但网络的训练和最终结果的评估任然有待研究。另外随着近些年 Transformer 在计算机视觉领域的兴起, 部分方法希望凭借其全局感受野的特性来捕捉全局的时序依赖。接下来本文将详细介绍以上四类方法中具有代表性的模型。

2.1 基于循环神经网络

人体运动序列预测问题通常被视为 *seq2seq* 预测任务。RNN 因其在此类任务中的出色表现而得到广泛认可, 这启发了许多研究人员利用基于 RNN 的方法来研究人类运动序列预测任务。EDR [5] 率先将 RNN 引入人体运动序列预测领域, 其结构如图2-1所示。其中 x_t 代表第 t 个时刻的输入的人体姿态, 而 y_t 则代表由 x_t 预测出的未来人体姿

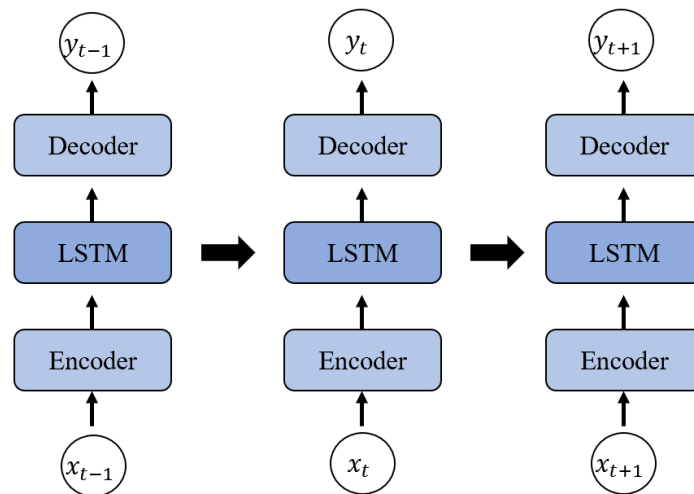


图 2-1 EDR 网络结构

态。网络接受 x 作为每个 RNN 节点的输入，首先输入姿态通过编码器（Encoder）编码到隐空间，随后送入 RNN 层，将时序信息提取并传递给下一个节点。同时通过解码器（Decoder）解码出对应的未来人体姿态作为当前节点的输出。该方法很好地利用了 RNN 的时序数据建模能力，有效提取了输入人体运动序列中的时序信息。但由于当前 RNN 节点是在上一个节点的基础上进行预测，因此容易出现误差累积问题。此外，由于在 EDR 中，未来运动序列被逐时刻、独立地预测，因此在输入序列和预测序列的过渡部分容易出现不连续的现象。Res. Sup.[8] 针对 EDR 中的问题提出了改进措施，如

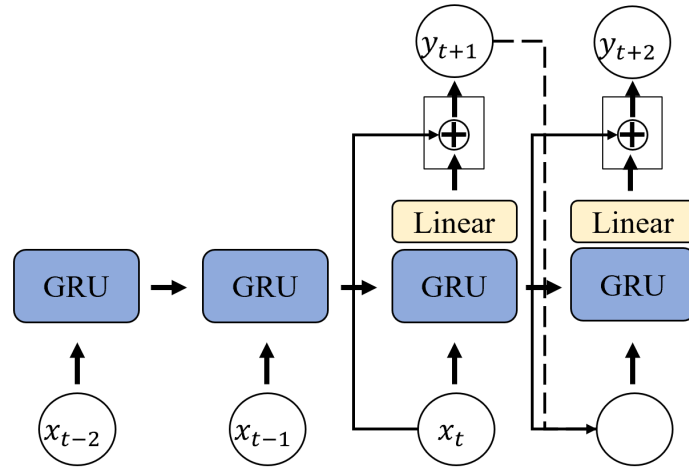


图 2-2 Res. Sup. 网络结构

图2-2所示，Res. Sup. 引入了在自然语言处理领域常用的 *seq2seq* 模型结构，与 EDR 相比，*seq2seq* 统一将输入序列编码到隐空间，此时隐空间包含所有的输入信息，这有助于模型从全局的角度考虑，而不是主要依靠当前时刻的输入。随后通过解码器结构，将隐空间中的信息解码为未来人体运动，当前时刻的输出将作为下一时刻的输入，这有助于保证时序上的连续性，也允许相邻时刻的运动通过残差连接的方式完成一致性约束。除网路结构外，另一些方法从人体运动学入手，通过分析人体运动模式来针对性地设计网络，例如 Tang *et al.*[10] 发现在人体运动中，并非所有关节都处于运动状态。相反只有处于肢体末端的关节位置才会较为频繁地改变。因此，他们提出了针对人体运动模型中频繁运动的关节的方法，称为 HUM。具体的，他们设计了一个新颖的门控单元用来过滤运动幅度小的关节。此外，注意力机制被用来关注具体的运动模式。AHMR[44] 为了捕获更多的长期相关性，在 RNN 单元中，可以同时相邻关节和帧进行编码。此外，它不仅可以同时对本地和全局上下文进行建模，而且还使用了一个注意力模块来帮助更新全局上下文。

虽然上述方法在 EDR 的基础上提出了改进措施,提升了网络性能,但由于 RNN 网络的特性任然无法解决诸如误差累积、过渡部分不连续、训练困难和难以处理长时间依赖关系等问题,这将削弱网络预测的真实性。为此一些新的方法的将目光投向了效率更高,感受野更大的卷积神经网络。

2.2 基于卷积神经网络

人体运动序列数据包含时间和空间两个维度,而卷积神经网络(CNN)在处理空间数据上有天然优势,时序信息也可以由 1D 的 CNN(TCN^[45])进行处理,相比循环神经网络,TCN 更轻量化、推理速度更快、配合空洞卷积^[46]感受野更大。在人体运动预测中,对于模型如何处理空间和时间的依赖关系是一个非常重要的问题。传统的 CNNs 只能捕捉静态图像的空间依赖性,但是在动态场景下,时间信息也是非常关键的。因此,研究人员提出了一些新的 CNN 架构,以处理人体运动预测的时空依赖关系。

在 Butepage *et al.*[33] 中,作者设计了一种新的卷积层来编码不同的时间尺度。这种卷积层可以有效地捕捉局部时间尺度的依赖关系,但是它无法处理长期的时间依赖性。为了解决这个问题,QuaterNet^[47] 引入了扩张卷积,可以在网络中捕捉长期时间依赖关系。该方法在分层输入姿势的情况下表现良好,但仍然无法处理空间依赖性。

为了同时处理空间和时间的依赖性,一些研究人员采用了分层结构的 CNN Li *et al.*[34]。这种 CNN 架构利用卷积结构来捕捉长期隐藏状态,并将其送到解码器中以生成人体姿势。这种方法可以有效地处理时空依赖关系,但是它需要大量的计算资源和训练数据。为了进一步提高模型的性能, Li *et al.*[48] 提出了一种卷积分层自编码器框架,用于表示人体骨骼结构。在这种框架中,分层拓扑被用于表示骨骼结构,并且嵌入了 1D 卷积层来编码每个节点。该框架可以有效地捕捉空间和时间的依赖关系。最近, TrajectoryCNN^[35] 被提出来处理人体运动预测的时空依赖关系。它引入了一种新型的轨迹空间,可以轻松地捕捉各种局部-全局和时空特征。这种框架在许多基准测试中取得了优异的性能。

虽然 CNN 能有效处理时间和空间数据,但 CNN 的规则卷积核决定它适合处理图像或视频这类规则数据。人体姿态属于不规则的无向图结构,人体关节点对应图中的顶点,骨骼对应顶点间的相互关系。这种拓扑结构是极其重要的先验空间信息,能有效辅助模型感知运动模式。而 CNN 的规则卷积核使得它很难利用这类先验信息,因此,在最近的研究中,天然具有拓扑信息处理能力的图卷积网络(GCN)获得了越来越多的关

注。

2.3 基于图卷积网络

GCN 是一种可以处理图形结构数据的神经网络。在 GCN 中，卷积操作是基于邻居节点之间的连接进行计算的，这使得 GCN 可以有效地处理具有不规则连接的数据结构，例如人体关键点。此外，GCN 还可以利用拓扑信息来捕捉节点之间的关系，从而更好地理解图结构数据。该特性对人体运动序列数据处理非常有利。

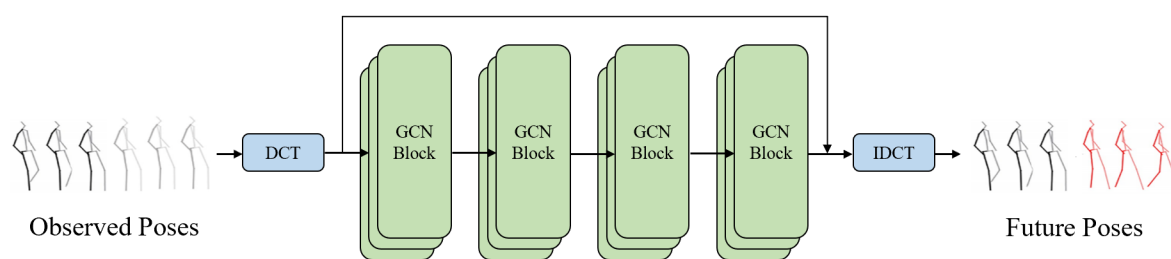


图 2-3 LTD 网络结构

LTD^[20] 率先提出了一种代表性的 GCN 方法（图2-3），使用原始的 GCN 对人体运动序列进行建模。具体的，对于输入的人体运动序列，LTD 将其视作一个不规则的无向图。由于人体运动序列数据包含时间和空间两个维度，而原始的 GCN 只能处理二维平面数据。因此，LTD 将该运动序列中的关节点轨迹视作一个整体，将其放入图结构网络中。即图中的每个节点包含了某个关节点这段时间内的运动轨迹，由此 LTD 实现了使用一个描述平面节点联系的 GCN 来处理时空维度的人体运动序列。在网络结构方面，网络接受历史人体运动序列作为输入，为了保证输入数据和输出数据在时间维度上的一致性，LTD 提出用已知序列的最后一个人体姿态填充输入序列，使其与输出序列时序长度一致。此外，网络输入和输出数据之间的残差连接也得到保证，有助于提高网络的训练效率和预测精确性。完成填充步骤后，输入数据将经过离散余弦变换（DCT）从时域变换到频率域，通过过滤掉低频信息并保留高频信息，可以在降低数据维度的同时，减少噪声。随后，再被传入多个串联的 GCN 模块，将数据映射到隐空间后，提取时空信息，在填充数据的基础上预测未来运动。最后，经过离散余弦逆变换（IDCT）后，输出最终的预测结果。该方法的贡献在于，提出了一种使用原始 GCN 对时序数据进行建模的方式，在最终预测精度上大幅领先基于 RNN 的方法，通过全局的残差连接解决了输入序列和预测序列过渡部分的不连续性。但由于该方法忽略数据的时序特性，仅仅使用 GCN 提取人体姿态的空间结构信息，将关节点轨迹作为一个整体放入图节点中，这导

致该模型对时序运动的感知能力有所欠缺，未来仍然有提升空间。

用于人体姿态提取的方法 ST-GCN[49] 针对 LTD 存在的问题，提出了一种具有时空信息提取能力的 GCN。对于时空人体运动序列数据，一个直观的想法是建立一个跨越时空维度的图，囊括不同时间和空间上的关节点。但由于 GCN 复杂度随着时空维度的增加成倍数上升，这样的图结构数据的复杂度是难以接受的。因此 ST-GCN 提出将时间和空间维度的数据拆分，分别用 TCN[45] 和 GCN 进行处理。具体的，1D 的卷积神经网络 TCN 负责提取各个关节点轨迹中的时序数据，GCN 负责处理人体姿态中的空间结构数据。通过将时空两个维度分为，ST-GCN 将网络的时间复杂度降为线性增长。并且通过实验证明网络时空信息提取能力优于现有方法。但 TCN 为局部算子，感受野被限制在卷积核范围内，导致 ST-GCN 在提取长时依赖上存在缺陷。

最近 MSR[28]，更进一步提出了空间层次化的 GCN 网络。它提出了一个类 Unet[50] 网络，编码器部分，逐渐简化人体姿态空间结构，只保留最简洁的空间信息。解码器部分，首先构造空间结构较简单的人体运动序列，随着网络的深入，人体运动序列的空间复杂程度逐渐增加，直到输出具有完整空间结构的数据。具体的 GCN 模块设计上它参考了 LTD，将关节点运动轨迹视作一个整体。该方法提出的空间层次化 Unet 网络，给网络一个渐进式的学习过程，这有利于降低网络的学习难度。但对空间结构进行简化的过程中，破坏了人体结构先验信息，导致网络预测效率相比 LTD 并没有明显提升，某些方面甚至出现了下滑。

由于 GCN 网络对图结构数据中节点关系的处理具有先天的优势，因此 GCN 能够更好地提取人体姿态数据中的结构先验信息。但现阶段的 GCN 对于时空跨维度信息的处理能力任然有待提高，它们或是忽略某一个维度来降低时间复杂度，或是在信息提取能力和时间效率上做出了妥协。因此，如何平衡模型复杂度和时空信息处理能力，将是未来的一个研究重点。

2.4 基于对抗生成网络

人体运动姿态预测算法的一个主要难点在于，预测过程中存在不确定性，这种不确定性是由于输入序列和预测目标序列之前的差异造成的。例如，如果输入序列与预测目标序列关联性强，则预测越简单，反之则越难。针对上述问题，一个解决思路是如上述方法，通过提高网络的时空信息提取能力，尽可能捕捉输入和预测序列间的关联性。另一个思路是引入生成式模型和随机性，生成更真实的运动序列。具体的，近年来由于对

抗生成网络 [51] 的深入研究, GAN 为生成人体运动姿态序列提供了更多新的可能性。

Barsoum *et al.*[36] 率先提出了一种基于 GAN 的 *seq2seq* 人体运动序列预测方法, 它使用改进版的 WGAN-GP 进行训练, 与上述基于 RNN, CNN 或 GCN 的方法不同, 它的网络输入表达为概率密度分布而非固定的人体运动序列。因此, 在预测时可以通过为网络提供不同的随机噪声 z , 来对同一个输入运动序列预测不同的未来运动序列。然而, 虽然该方法在结果真实性方面有所提升, 但由于输入噪声的引入, 预测准确性有所下降。在此基础上, BiHMP-GAN[37] 同样通过在输入序列中添加从固定分布中采样的随机噪声来为预测过程添加随机性。不同的是, BiHMP-GAN 提出了一个双向对抗神经网络来解决预测过程中的模式坍塌问题。与此同时, 受到上述工作的启发, AGED[9] 提出了一种新颖的对抗生成框架, 它具有两个全局的循环鉴别器, 一个鉴别器被用于促进生成序列的保真度, 另一个鉴别器与网络进行联合训练, 保证未来生成序列的连续性。STMI-GAN[38] 也沿用了该思路, 用于处理长时依赖的人体运动序列。Adversarial Refinement Network (ARNet)[42] 设计了一种新的对抗式的误差调整策略, 与上述方法不同的是。判别器不再直接判断生成结果的真实性而是用来估计预测误差, 随后精修模块再根据误差调整预测结果。而 Lyu *et al.*[43] 则利用 GAN 模拟路径积分来解决随机微分方程并预测未来运动轨迹。值得注意的是, 由于 GAN 的对抗训练特性, 想要训练达到平衡状态是非常困难的, Cui *et al.*[41] 提出了一种新的 GAN, 该 GAN 使用了 spectral 归一化, 以避免模式坍塌。还有另一种称为 AMGAN[40] 的策略, 它由复合 GAN 结构设计而成, 包含用于不同低维身体部位的局部 GAN 和用于高维全身的全局 GAN 组成。该方法证明了降维可以有效地提高 GAN 的训练效率。

总而言之, 利用 GAN 的策略主要可分为两类。(1) 被用作学习算法以帮助网络生成更加真实的结果。(2) 利用随机噪声向网络添加随机性, 生成多样化的预测结果。而 GAN 作为一个具有明显优势和劣势的网络, 也会给研究人员的工作带来一定的挑战。

2.5 基于 Transformer

近年来, Transformer 受到了学术界的广泛关注, 它也从自然语言处理 (NLP) 领域被引入到计算机视觉领域, 在诸如图片识别、图片分割等经典问题上大幅领先现有基于卷积神经网络的方法。对于人体运动序列预测问题, 网络需要捕捉长时依赖关系的能力。而 Transformer 的全局感受野特点恰好可以解决该问题。因此出现了一批基于 Transformer 的方法 [52-53]。

Aksan *et al.*[52] 设计了一个包含时间和空间分支的 Transformer 网络，两个分支分别提取输入序列的空间结构信息和时序信息，最后再通过融合模块得到最后的预测结果。

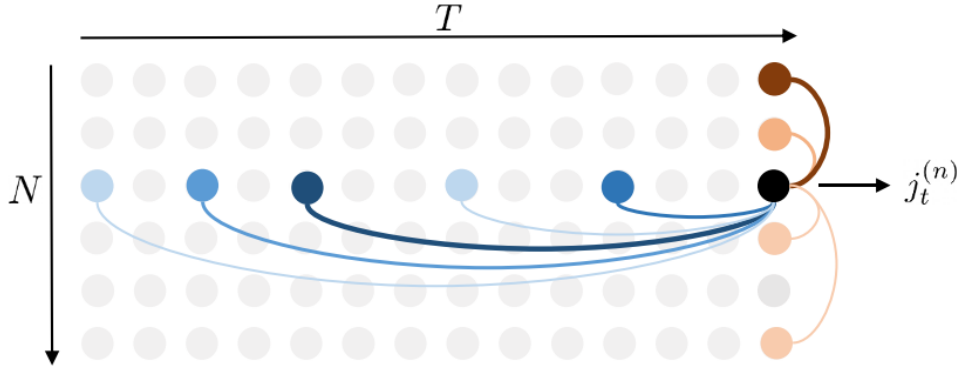


图 2-4 Spatial-temporal Transformer^[52]

其中 spatial-temporal Transformer 原理如图2-4所示， $j_t^{(n)}$ 表示 t 时刻，第 n 个关节点。其中， $j_t^{(n)}$ 只和自己位于同一时间或空间的关节点进行注意力（attention）机制计算，图中颜色的深浅代表关节点之间的关联程度，颜色越深关联性越强，权值也越高，反之则越小。通过分离的时空 transformer，该方法间接地提取了时间和空间信息。

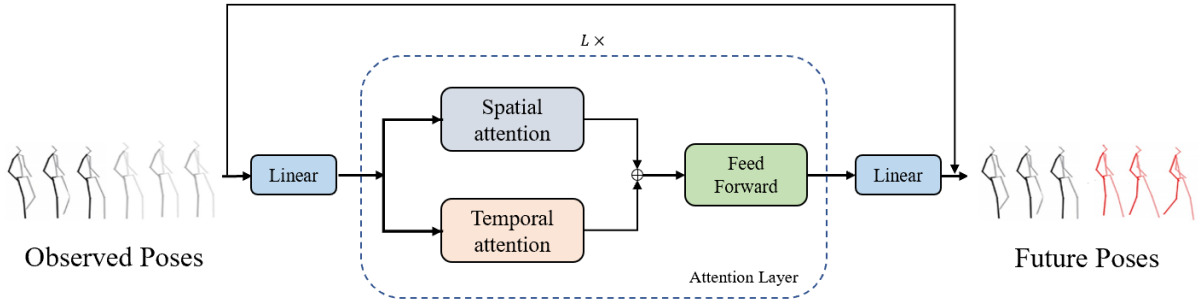


图 2-5 Spatial-temporal Transformer^[52]

完整的网络结构如上图所示，网络由 L 个串联的注意力层构成，每个层包含一个空间维度和时间维度的注意力层，特征被传入注意力层后，分别送往两个分支，用于提取时间和空间信息。提取结束后，空间和时间信息相加，送入前馈神经网络进行特征融合，最终通过线性层输出预测的未来人体运动序列。该方法通过并行的方式分离时间和空间维度，减少了时间复杂度和模型参数量。但分支的方法使得时间和空间维度缺少信息通信手段，导致信息交流受阻，影响最终的模型质量。总的来说，Transformer 高效的全局注意力机制有利于模型捕捉长时序依赖，但 Transformer 的注意力计算模块也导致模型空间的上升和计算开销的增加。此外，时间和空间分支间的通信问题也是未来需要

研究的问题。

2.6 总结

本章，我们对人体运动姿态预测算法的发展做了一个简要的回顾。在初期，研究人员根据循环神经网络在处理时序数据上的优势，设计了 *seq2seq* 的网络模型来对输入序列统一编码后预测未来运动序列。但由于循环神经网络对于时序记忆的能力依赖隐变量的大小，因此难以处理长时依赖。此外，梯度消失等训练问题也困扰着现有方法。随后，研究人员将目光转向了卷积神经网络，特别是图卷积神经网络，它对无结构不规则数据的处理有天然的优势。但如何对传统图卷积网络进行改进，使其拥有时序信息处理能力，任然是当前的研究热点。此外，GAN 机制的引入允许生成更多样化和真实的结果，但其训练过程的不稳定性和噪声对结果准确性的影响有待进一步解决。近些年，Transformer 的兴起给该问题带来了新的解决思路，其全局感受野的特性，允许其捕捉更长范围的全局依赖。但其注意力计算带来的额外计算开销和时空信息间的通信问题，还需要进一步探索。本文希望在现有基于图卷积的方法的基础上，提高模型的时空信息提取能力，并且控制模型的运行开销。

第三章 相关基础知识

3.1 图卷积网络

3.2 渐进式策略在视觉领域的应用

image inpainting, MSR-GCN

3.3 Transformer 全局注意力机制

这章可能要去掉

第四章 基于渐进式策略的多阶段人体运动姿态预测框架

本章节将围绕本文的两个主要贡献点：渐进式的网络学习框架和集成 $SD - GCN$ 和 $TD - GCN$ 的图卷积模块。分别从动机、方案、实现框架和算法细节几个方面对本方法进行详细的阐述。在此之前，我们首先通过数学语言定义人体运动姿态预测问题，并介绍在此过程中使用的相关数据结构，以方便在本文后续章节中进行准确的叙述。

4.1 数据描述与问题定义

4.1.1 人体运动姿态数据结构

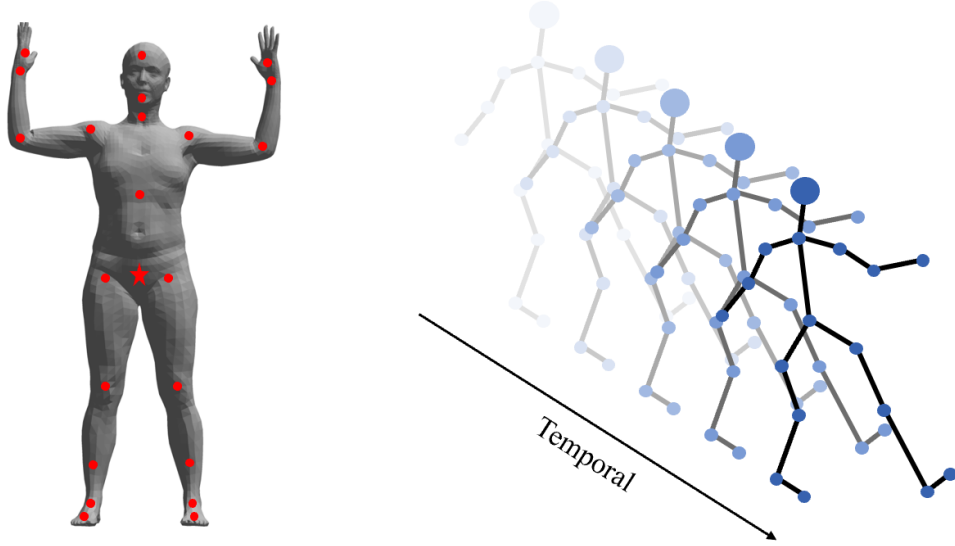


图 4-1 人体运动姿态数据结构

首先我们介绍人体运动姿态预测问题所使用的数据结构。如图4-1左所示，人体运动数据是通过动作捕捉设备，在封闭室内或开放室外场景提取到的人体关键点运动数据，这些数据以 SKT（Skeletal Kinematic Tree）的形式表示和存储。在实际动作捕捉过程中，通常只关注在运动过程中起决定性作用的关节点，例如手肘、肩部、膝盖等。这些关节点在4-1左中以红色标记的形式展现。其中位于胯部的五角星节点被称为根节点，其余节点通过递归的形式计算自身对于根节点的相对位置来得到自身位置。由于本文仅关注三维欧式空间中的人体运动，因此，我们用与根节点的相对 3D 坐标来描述每个关节点空间位置。将独立的关节点按照人体结构连接后，即可得到抽象后人体姿态。由于我们处理的是序列数据，同时包含时间和空间两个维度的关节点，因此在4-1右中，空间维度上描述人体结构信息，时间维度上描述关节点序列运动信息。

4.1.2 人体运动姿态预测问题定义

从数学上，对于一个长度为 T 的人体运动序列，我们将其定义为 $S_{1:T} = \{P_1, P_2, \dots, P_T\}$ ，其中 P_i 当前运动序列中位于 i 时刻的人体姿态。每个人体姿态 P_i 又由若干个关节点组成其中 $P \in \mathbb{R}^V$ ， V 为该人体姿态包含的关节点数量，每个关节点又由一个 D 维的向量描述。

对于人体运动姿态预测问题，网络 Φ 接收已知输入序列 $S_{1:T_h} = \{P_1, P_2, \dots, P_{T_h}\}$ 作为输入，预测未来运动序列 $S_{T_h+1:T_h+T_f} = \{P_{T_h+1}, P_{T_h+2}, \dots, P_{T_h+T_f}\}$ ，这一过程的数学描述如公式4-1所示。其中 θ 为可训练的网络参数。

$$S_{T_h+1:T_h+T_f} = \Phi(S_{1:T_h}, \theta) \quad (4-1)$$

4.2 渐进式人体运动序列预测框架

正如1.1节所提到的，预测过程中的不确定性是影响预测精确度进一步提升的关键因素，而这种不确定性来自输入运动序列和待预测序列之间的差异。简而言之，由于输入运动序列和预测运动序列之间存在较大差异（例如，输入运动和待预测运动的运动模式有较大差异），网络无法根据输入序列中的信息来准确推测未来运动，导致预测结果脱离真实情况。因此，如何降低预测过程中的不确定性成为了当务之急。

在调研过程中，我们注意到 LTD[20] 提出的一项改进使得预测精度相较于现有方法得到了极大提升。在早期的方法中，如公式4-1所示，输入运动序列长度 $1:T_h$ 与预测序列长度 $T_h+1:T_h+T_f$ 通常存在差异，通常预测序列的长度要远远长于输入序列（例如，输入 10 帧预测 25 帧），这使得网络需要在毫无参考基础的情况下去构造未来运动序列。这通常会导致预测结果不连续，与真值出现较大偏差。针对该现象，LTD[20] 提出公式4-2，通过使用输入序列的最后一帧来填充输入序列，使得输入序列的长度和预测序列保持一致。

$$\begin{aligned} \tilde{S}_{1:T_h} &= [S_{1:T_h}, \{P_{T_h}^{T_h+1}, \dots, P_{T_h}^{T_h+T_f}\}] \\ S_{T_h+1:T_h+T_f} &= \Phi(\tilde{S}_{1:T_h}, \theta) \end{aligned} \quad (4-2)$$

从图4-2可以看到，经过填充后的输入数据对网络有两点促进作用，第一，输入和输出维度一致，避免了数据维度变换过程中的不确定性。第二，网络在预测未来序列时可以在已知部分最后一帧基础上进行预测，降低了预测的难度。然而这种粗糙的填补方法也有其固有缺陷。首先整个填充过程不区分时序距离，全部使用最后一帧进行填充，

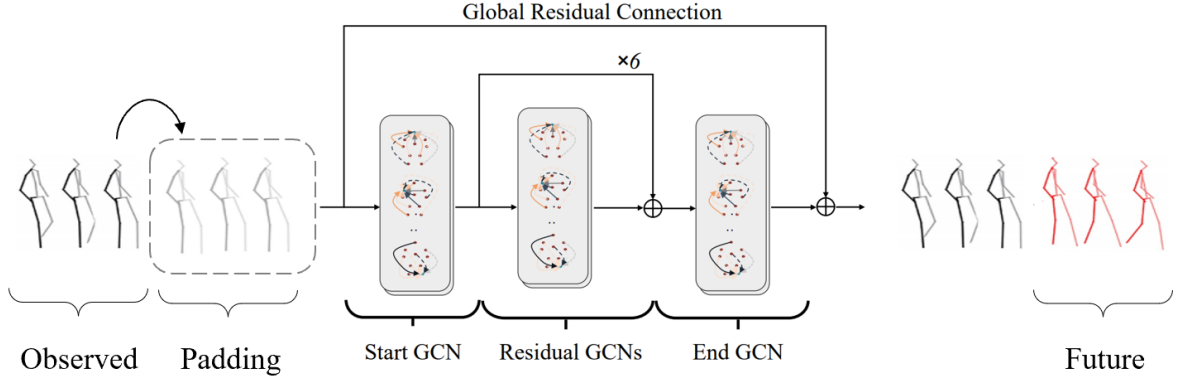


图 4-2 LTD 数据填充过程

对于离 P_{T_h} 较近的未来运动， P_{T_h} 还能提供一定的参考。随着时间向前，未来帧与 P_{T_h} 的关联越来越弱，其提供的参考价值也越来越低，预测的不确定性也逐渐增加。因此该填充方法无法缓解较远距离的预测不确定问题。

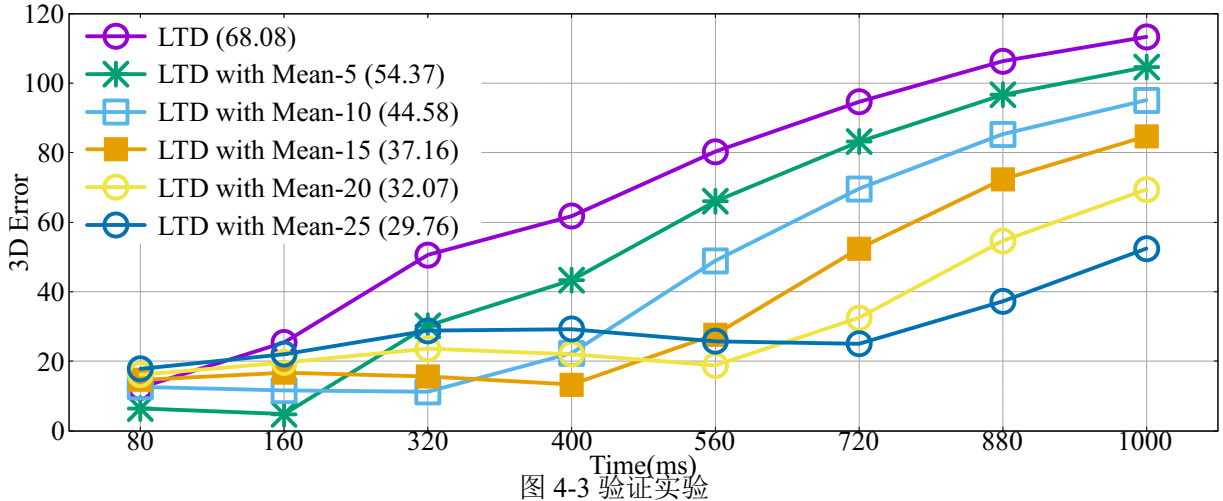


图 4-3 验证实验

LTD 证明了通过缩小输入数据和预测目标之间的维度差异，可以有效降低预测不确定性。而我们希望验证，缩小二者在内容上的差异也能达到同样的效果。因此我们以 LTD 为基准，设计了图4-3中所示的验证实验。在实验中，预测目标保持不变，输入数据的填充内容被替换为部分预测目标的均值，例如图中所表示的前 L 帧的均值（Mean- L ）。从直观感觉上，填入均值相当于向输入数据泄露了部分预测目标的信息，间接缩短了二者在内容上的差距，预测难度也会随之降低。图4-3展示的结果也验证了我们的猜想，混入的平均预测目标信息有效缩短了输入数据和预测目标之间的差距，整体预测精度相较于 LTD 有了明显提升，且混入的信息越多对应位置和整体上预测精度提升越多。

在分析上述方法得失后，我们认为，缩小输入序列和待预测序列的差异（维度差异、内容差异等）可以有效降低预测过程中的不确定性，使得预测结果向我们期望的方

向趋近。但单纯地填充输入序列最后一帧所带来的性能提升还有一定的上涨空间，验证实验为我们提供了一个思路，可以通过缩小输入数据和预测目标内容上的差异来进一步提升预测精度。

由于在实际预测场景中，无法将预测目标信息泄露给输入数据，因此我们考虑通过降低预测目标的难度来缩小二者之间的内容差异。受到最近被广泛应用的由粗糙到精细（Coarse To Fine）策略的启发，我们设计了一个简易实验来验证我们的想法。Coarse to fine 策略与大多数一步到位的方法不同，预测被分为了两个阶段。位于网络浅层的阶段被称为粗糙（Coarse）网络，它接收原始的输入，并输出一个较为粗糙的结果，虽然该结果离最终的目标存在一定的偏差，但与最初的输入信息相比，它已经包含了目标的绝大部分信息。随后该粗糙结果被送入精修（Fine）网络，精修网络将在粗糙网络的基础上进一步完善预测细节。该策略被广泛应用于图片修复（Image Inpainting）[54-55] 领域，原始缺失图片通常由粗糙网络生成一个低分辨率较模糊的修复版本，此次修复的目的是修复图片内容的结构。随后，粗修版本被送入精修网络提高分辨率并进一步完善细节，最后输出高质量的修复结果。参考该思路，我们设计了一个 Coarse To Fine 人体运动序列预测网络。

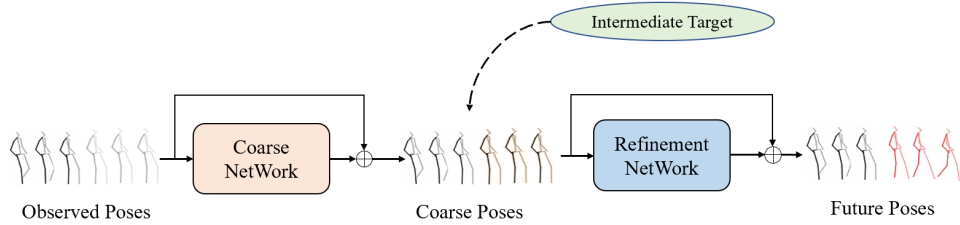


图 4-4 Coarse To Fine 预测网络

在最初的粗修阶段，我们仍然保留了 LTD 中基于最后一帧的填充步骤，因为通过此步骤可以保证网络输入输出维度一致，减少预测维度上的不确定性。经过填充后的输入序列经过粗修网络，预测得到一个较为粗糙的结果，该结果受到中级监督目标的监督。该中级监督目标相比较最终的预测目标，去除了部分运动细节，只保留了运动大致趋势，通过缩小输入信息与预测目标之间的差距，降低了预测过程的不确定性。随后，粗糙的预测结果被送入后续的精修网络，进一步丰富动作的细节。此时，预测结果受到原始真值的监督，以期望获得与真值一致的结果。该模型通过两阶段的结构，将预测过程拆分为两个部分，每个部分的预测不可确定性得到了减少。预测精确性与单阶段的 LTD 网络相比有显著提升，这一结果在实验部分有所证明。

4.2.1 渐进式多阶段预测网络框架

随后我们进一步拓展了该过程，将一个两阶段的 Coarse To Fine 网络拓展为多阶段的网络。通过对预测过程的进一步细分，每个阶段的预测难度被进一步降低，网络也容易做出准确的预测。我们将多阶段的网络模型定义如下。

$$\begin{aligned}\hat{S}_{1:L}^1 &= \Phi^1([S_{1:T_h}; P_{T_h}, \dots, P_{T_h}]), \\ \hat{S}_{1:L}^i &= \Phi^i([S_{1:T_h}; \hat{S}_{T_h+1:L}^{i-1}]), i = 2, 3, \dots, T,\end{aligned}\tag{4-3}$$

公式4-3中，现有方法中单阶段的预测过程 Φ 被拆分为多个阶段 $\Phi = \{\Phi_1, \dots, \Phi_T\}$ ，每个阶段在上一个阶段的预测基础上，不断完善预测精度，使得预测精确度不断稳步提升，相比较现有的单阶段网络，多阶段网络预测过程更加可控，每阶段的预测结构都受到与之对应的中级监督目标的监督，此外预测任务的细分也使得每阶段网络的预测不确定性进一步降低，预测难度也进一步降低。为此，我们设计了一个多阶段的网络，其网络结构图如图4-5所示。

多阶段的网络设计包含两点，第一是每个阶段（Stage）的子网络设计，在这里我们最初使用了 LTD 中提出的 GCN 模块（在文章后续内容中我们提出了具有更强时空信息提取能力的 GCN 模块用以替换）。每个子网络内部为一个解码器编码器网络（Encoder-Decoder），编码器解码器结构对称。内部由多个 GCB（GCN Block）构成，每个 GCB 又由两个 GCL（GCN Layer）构成，GCL 是网络的最基本的构成结构，其具体结构如图4-3右下角所示，GCL 由一个 GCN、BatchNorm、Tanh 和 DropOut 组成，可以完成最基本的人体运动数据时空信息提取功能。值得注意的是，网络的规模并不随着阶段数量的增加而提升，在我们的设计中，网络中的 GCB 数量固定，随着阶段的增加，每个阶段包含的 GCB 数量成比例下降。网络参数量的增加在一定范围下提高了网络的容量和表达能力，但当突破某一阈值的时候网络复杂度增加带来的计算开销负担抵消了这一优势。因此我们的多阶段网络，在提升网络性能的同时，并没有提高网络的复杂程度和计算开销。第二是，中级监督目标的设计，中级监督目标必须具有层次化特点，遵循由简单到复杂的原则，浅层网络负责较为简单的处理大致框架的任务，而深层网络负责较为复杂的细节完善任务。在其他类型任务（图像等规则数据）的中级监督目标设计中，可以简单地通过降低分辨率、模糊处理和提取边缘特征等方式降低图片的细节和提取内容的结构信息来构造中级监督目标。但人体运动姿态数据是不规则的图状数据，并且已经高度抽象。无法通过寻常的降维手段提取结构信息或降低运动复杂度。虽然当前

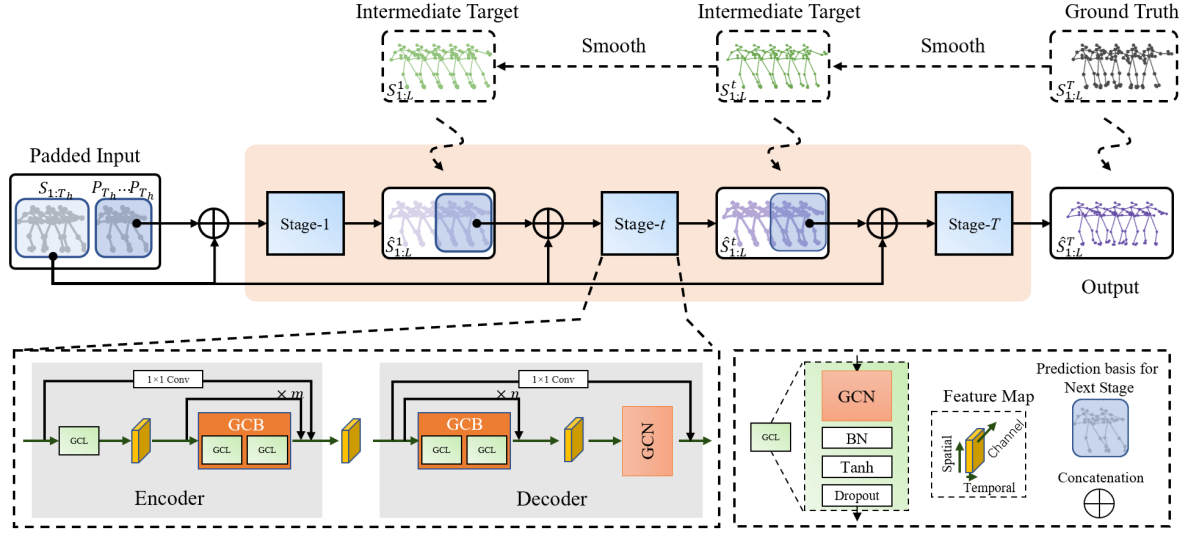


图 4-5 渐进式多阶段的预测网络

也有一些方法，如 MSR[28] 提出在空间维度上对人体结构进行降维，合并运动模型类似的相邻关节来减少图中的关节数量。该方法建立了空间层次化的中级监督目标，但该方法破坏了重要的人体结构先验信息，导致渐进式的策略发挥了极其有效的作用。

4.2.2 基于累积均值平滑的中级监督目标

因此我们设计了一种适用于人体运动姿态时空数据的中级监督目标构建方法，该方法可以在时间维度上降低数据的运动复杂度，同时保持原有的空间结构，以便后续网络提取其中的人体结构先验信息。具体的，我们在不同阶段对关节点轨迹施加不同强度的平滑算法。网络浅层因为表达能力不足，因此其对应的中级监督目标被施加了更强的平滑来降低运动的复杂度和预测难度。随着网络深入，网络的表达能力增加，能够承担更复杂的预测任务，此时，对于中级监督目标的平滑程度就会被削弱，以帮助其在之前阶段的预测基础上，丰富结果的细节。为此，我们设计了一种名为累积均值平滑 (Accumulated Average Smoothing) 的方法构造用于人体运动姿态序列中级监督目标。

在介绍累积均值平滑算法之前，为了方便叙述，我们首先给出人体运动模型中关节点轨迹的数学定义。假设每个姿态包含 V 个关节点，每个关节点由 D 维的向量描述。一个人体运动姿态序列 $S_{1:L}$ 包含 $V \times D$ 条轨迹： $\{T_j | j \in [1, M \times D]\}$ ，每条轨迹 T_j 由同一个关节点的某位维度上的运动组成： $T_j = \{x_j^i | i \in [1, L]\}$ 。由于所有轨迹都由同样的平滑方法处理，因此在下面的叙述中我们忽略了不同轨迹的区别，统一用 T 代称 T_j 。

T 由两部分组成：已知的运动序列 $\{x^i | i \in [1, T_h]\}$ 和待预测的运动序列 $\{x^i | i \in [T_h + 1, T_h + T_f]\}$ 。由于已知的运动序列属于模型的输入数据，不需要预测，所以不需

要经过累积平滑算法处理。待预测的运动序列是网络需要预测的部分，需要用累积平滑算法调节该部分数据的预测难度。目前被广泛使用的平滑方式是基于高斯卷积核的滤波器。高斯滤波器被广泛应用于图像平滑领域，它是一种常见的线性滤波器。它的原理是将一个二维高斯分布函数应用于图像的每一个像素，使得该像素周围的像素加权平均起来，从而达到平滑图像的目的。高斯滤波器的核心是高斯核（Gaussian kernel），也称为卷积核（Convolution Kernel）或滤波器（Filter）。高斯核是一个二维高斯分布函数，它的中心是图像上的当前像素点。高斯核中的每个元素表示该位置的权重，越靠近中心位置的像素权重越高，越远离中心位置的像素权重越低。通常情况下，高斯核是一个奇数 \times 奇数的矩阵，这样可以保证中心像素的位置。在轨迹平滑算法中，2D 的高斯卷积核退化为一维，其卷积核权重计算公式为：

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \quad (4-4)$$

其中 x 表示当前位置相对于卷积核中心的距离， σ 表示标准差，标准差越大越靠近卷积核中心的权重越高，反之权重分布越平均。

$$G = \frac{1}{\sqrt{2\pi}\sigma} \begin{bmatrix} e^{-\frac{((N-1)/2)^2}{2\sigma^2}} & \dots & e^{-\frac{((N-1)/2)^2}{2\sigma^2}} \\ \vdots & \ddots & \vdots \\ e^{-\frac{0^2}{2\sigma^2}} & \dots & e^{-\frac{0^2}{2\sigma^2}} \\ \vdots & \ddots & \vdots \\ e^{-\frac{((N-1)/2)^2}{2\sigma^2}} & \dots & e^{-\frac{((N-1)/2)^2}{2\sigma^2}} \end{bmatrix} \quad (4-5)$$

如果需要生成一个 $N \times 1$ 的高斯卷积核，可以将上述公式4-4代入矩阵形式，得到卷积核矩阵4-5，其中， G 是 $N \times 1$ 的高斯卷积核矩阵， N 表示卷积核的大小。

由于在本问题中，只需要对一条轨迹待预测部分进行平滑，已知的运动序列保持不变即可。而高斯滤波器在对轨迹进行处理时，不可必要的需要计算卷积核范围内的所有阶段数据，因此在已知运动序列和待预测部分的过渡部分会出现跳跃（Jump）现象，这是由于高斯滤波器在计算过渡部分的平滑值时，将卷积核范围内已知运动序列纳入计算。而已知运动序列并没有进行平滑操作，所以导致计算结果在该处出现了跳跃现象。此外，由于高斯滤波器在计算时，其节点原始值的权重较高，导致平滑力度不足，难以构造层次化的中级监督目标。

因此，我们提出了累积均值平滑算法来解决以上两个问题。我们将该算法定义如下：

$$\bar{x}^i = \frac{1}{i - T_h} \sum_{k=T_h+1}^i x^k, \forall i \in [T_h + 1, T_h + T_f]. \quad (4-6)$$

待预测部分的某个节点的平滑值由其之前所有节点的累计平均值得到。首先平滑值计算过程只涉及待预测部分，不涉及已知部分，这就避免了出现过渡部分的跳跃问题。其次，该节点的平滑值是由前面所有节点的平均值计算得出的，这种方法消除了原始节点权重过高影响平滑结果的问题。这意味着每个节点在平滑过程中都具有相等的影响力，没有任何一个节点能够主导结果。随着离已知部分的距离增加，该节点参与计算的节点数量增加，平滑力度也随之增强。这是因为距离已知部分较远的节点预测难度更大，需要更强的平滑力度来降低预测难度。

相反，靠近已知部分的节点保留了更多的原始信息，这是因为这些节点的不确定性较低，预测难度也相对较低，因此不需要进行过度平滑。因此，该算法能够基于每个节点的不确定性程度和预测难度来适当地平衡平滑力度和原始信息保留。这使得该算法可以在处理各种复杂的预测问题时，提供精确的结果，并减少预测误差。总之，累积均值平滑算法相比传统的高斯滤波算法能够生成过渡更平滑，难度梯度更合理的中级监督目标，能有效地帮助网络建立一个渐进式的学习框架，降低预测难度，提高网络学习效率。

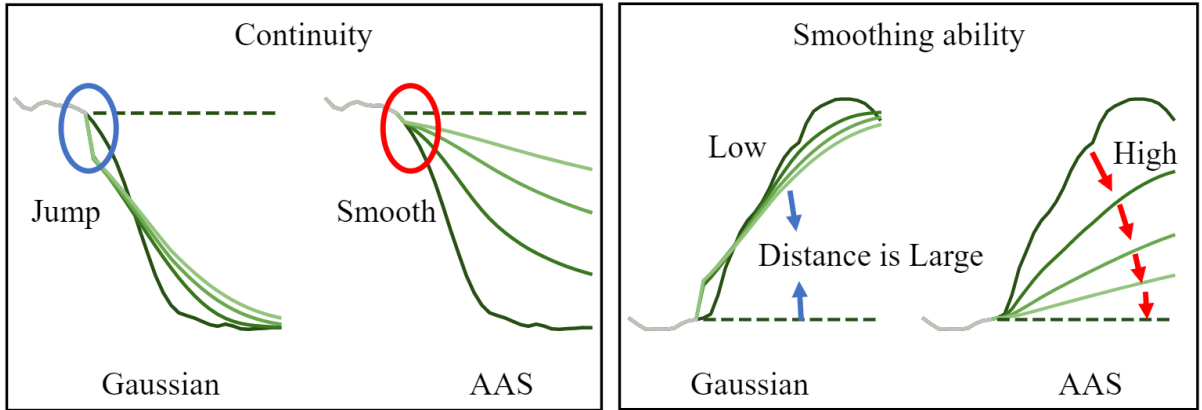


图 4-6 累积均值平滑对比高斯滤波

上图展示了累积均值平滑算法（AAS）和高斯滤波算法（Gaussian）的结果对比。其中灰色实线是输入模型的已知运动序列，深绿色虚线是使用已知运动序列最后一帧构成的填充序列。接下来由深到浅的绿色实现是不同阶段的平滑结果。拥有最深颜色的曲线是待预测的原始运动序列，稍浅一点的为经过一次平滑的结果，后续逐渐变浅的线段是经过多次平滑的结果，颜色越浅则经历的平滑次数越多。

图4-6左展示了高斯滤波出现的过渡部分跳跃问题。从图中我们可以看到，待预测的原始轨迹与已知序列的过渡部分是平滑的。然而在经过高斯滤波处理后，平滑后的待预测轨迹与已知部分的连接处，出现了明显的跳跃现象，这是由于在计算该部分平滑值时，卷积核窗口包含了已知序列和待预测序列两部分的信息。而我们提出的累积均值平滑算法解决了这一问题，在对某个节点进行平滑处理时，纳入计算的运动数据只涉及当前节点之前的待预测运动序列，不包含已知运动序列，平滑的强度随时间增加而线性增长。例如，当计算待预测序列的第一个节点的平滑值时，由于该节点已经位于待预测序列的顶端，因此，计算该节点的平滑值只需将该节点代入公式4-6，得到的结果也既是它本身。由于平滑节点数据未发生突变，因此可以和已知部分平滑过渡。

图4-6右，展示了累积均值平滑算法和高斯滤波算法的平滑力度对比。高斯滤波算法的计算过程中，在卷积核窗口中给予了原始节点过高的权值，导致平滑结果与原始节点的差异较小。最终，即使在经过迭代后的多次平滑步骤。最终结果的平滑程度仍然较低，无法满足构造层次化的中级监督目标的要求。而在累积均值平滑算法中，采用了累积的策略，越远离已知部分的节点受到的平滑力度越大，其次，使用均值的计算方法而不是加权平均的算法，平滑结果受到原始数据的影响更小，平滑的结果也更加明显。累积均值平滑算法的平滑结果表明，该算法能够在去除信号噪声的同时，保留数据序列的趋势特征。与其他常见的平滑算法相比，累积均值平滑算法在平滑强度方面表现更加优秀。这种层次化的特点使得累积均值平滑算法的平滑结果更加符合渐进式策略中预测目标由易到难的变化趋势，为多阶段预测网络模型提供了更好的中间监督目标。这种中间监督目标可以有效地降低预测难度，并提高预测的准确性和稳定性。因此，基于累积均值平滑算法的平滑结果是一种有价值的中间监督目标，它可以帮助多阶段预测网络模型实现更加准确、稳定和可靠的预测结果。

4.2.3 总结

本章介绍了基于渐进式策略的多阶段人体运动姿态预测网络，以及对应渐进式多阶段网络结构的中级监督目标构造方法。首先我们从对现有方法的分析入手，发现现阶段人体运动姿态预测问题的难点在于输入数据和待预测数据之间的差异过大，导致网络预测过程存在不确定性。LTD 提出使用已知部分来填充空白维度，使得输出输出数据达成维度上的统一，消除由于维度差异带来的不确定性。受此启发，我们希望更进一步，通过降低二者在内容上的差异性来消除预测不确定性。为此我们首先设计了一个 Coarse to fine 二阶段实验网络，人体运动姿态预测被分为两个部分，第一个阶段网络的预测目

标是较简单的大致的运动趋势，第二阶段的网络在上一步的基础上完善复杂的运动细节，使最终的神经网络输出与真值一致。在两阶段的网络中，我们在不增加网络参数数量的前提下，通过分解任务的方式将缩小了各个阶段中，输入与输出间内容上的差距，从而降低了预测的不确定性。在最终的设计中，我们将两阶段的网络推广到多阶段的网络，进一步体现渐进式策略的优势。

我们的另一个贡献点是提出了一种名为累积均值平滑的中级监督目标构造方法，该方法相比较常用的高斯滤波平滑方法，可以避免平滑后运动序列的过渡部分出现跳跃现象。此外由于其累积平滑的特性，拥有更强的平滑能力，相比高斯滤波平滑能够生成更具层次化的中级监督目标，辅助多阶段网络构建一个从难到易的网络预测框架。

除了通过基于渐进式策略的多阶段网络结构来降低输入数据和预测目标内容上的差距，我们还期望提高网络中的基础图卷积模块的时空信息提取能力，来降低预测过程的不确定性。现有图卷积模块或缺失了对时序信息的建模，或感受野范围受限于传统算子的卷积核大小，又或是带来了不可接受的网络规模膨胀。而我们希望提出一种在不显著提升时空开销的前提下，拥有对时空信息高效建模能力的 GCN 模块。我们将在下一节详细阐述该方案。

第五章 基于时空分离策略的 Non-Local 时空图卷积模块

在第4章中提到，设计合理的网络架构和学习策略，进而减少预测过程中的不确定性，是提高人体运动姿态预测精度的一个有效方法。但更多的方法，着眼于提升网络特征提取能力，通过从输入数据中获得更多的信息来降低预测过程中的不确定性。早期的方法借助循环神经网络在处理序列化数据上的优势，设计了基于 RNN 的人体运动姿态运动预测算法。但这类方法只考虑了数据的序列化特性，忽略了对重要的人体空间结构信息进行建模。随后的方法，注意到了图卷积网络在对不规则数据进行结构建模的能力。设计了使用 GCN 对人体结构信息进行建模的方法。但由于人体运动姿态数据包含时间和空间两部分，使用传统的图卷积网络进行建模时，邻接矩阵的规模将随着时空维度倍数增加。因此，有部分方法提出将时空信息的处理步骤分离，空间维度由 GCN 处理，时间维度则由 TCN^[45]处理。但 TCN 的卷积核大小限制了其感受野范围，不利于捕捉时间序列中长时依赖。鉴于现有方法存在的种种不足，我们提出了一种基于时空分离策略的 Non-Local 时空图卷积模块。该方法通过分离时空维度降低了网络的时间复杂度。在时间和空间维度均使用 Non-Local 的 GCN 算子，赋予网络全局感知能力。接下来，我们将首先介绍现有时空图卷积模块和本方案的设计思路以及优劣对比，随后再详细叙述基于时空分离策略的 Non-Local 时空图卷积模块的构造细节。

5.1 时空图卷积模块设计思路对比

再展示各个时空图卷积模块设计思路之前，我们首先展示人体运动姿态数据的时空结构特点。

如图5-1所示，图中的彩色节点代表关节点，同一种颜色的节点代表某个关节点在时序上的运动轨迹（Trajectory）。同一时间点所有不同颜色的节点构成一个人体姿态。在数学上，我们将上述人体运动姿态序列定义为一个时空无向图（spatialtemporal graph） $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ 。其中， \mathcal{V} 是对应人体运动姿态序列中所有关节点的节点集合。而 \mathcal{E} 是描述这些关节点对间联系的边集合。图5-1展示了一个长度为 4（ $T=4$ ），每个姿态包含 3 个关节点（ $V=3$ ）的人体运动序列。而时空图卷积模块的任务就是借助邻接矩阵 $\mathbf{A}^{ST} \in \mathbb{R}^{VT \times VT}$ 对包含时间和空间两个维度的人体运动姿态序列进行建模。通过如下递归的时空图卷积网络可以提取到运动序列中的高阶语义信息。

$$\mathbf{H}^{(l+1)} = \mathbf{A}_{ST}^{(l)} \mathbf{H}^{(l)} \mathbf{W}^{(l)}, \quad (5-1)$$

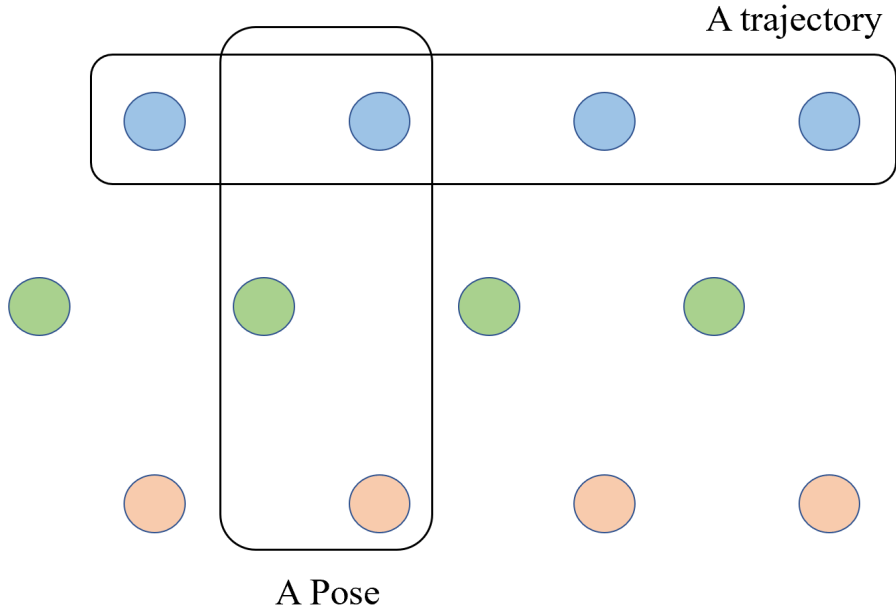


图 5-1 人体运动姿态数据时空结构

其中, l 指第 l 时空图卷积层, $\mathbf{A}_{ST}^{(l)} \in \mathbb{R}^{VT \times VT}$ 则是该层描述时空依赖关系的邻接矩阵。 $\mathbf{H}^{(l)} \in \mathbb{R}^{VT \times F^l}$ 是该层待处理的特征图, 特征空间等于时空维度。 $\mathbf{W}^{(l)} \in \mathbb{R}^{F^l \times F^{l+1}}$ 是特征映射权值矩阵, 负责将关节特征在特征空间上进行变换。最终, $\mathbf{H}^{(l+1)}$ 作为 $\mathbb{R}^{VT \times F^{l+1}}$ 空间中的特征图被送往下一个阶段进行进一步处理。

在处理过程中, 由于数据包含时间和空间两个维度, 每个维度上的数据增加都会导致节点数量成倍膨胀, 因此要求时空图卷积模块具有高效率。同时, 为了对连续的时序数据和不规则的空间结构进行建模, 时空图卷积模块需要具备较强的时空信息提取能力。而现有方法很难同时兼顾着两点要求。

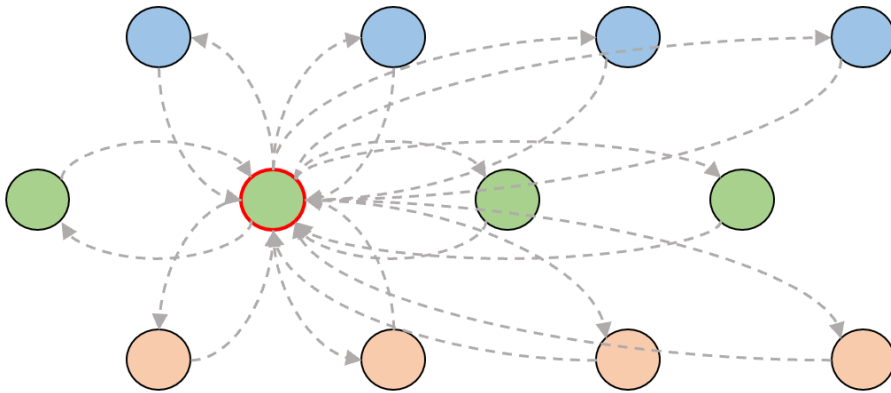


图 5-2 使用 GCN 同时对时空维度建模

最直接的思路是如图5-2所示使用一个 GCN 同时对时空两个维度进行建模, 图中的

虚线表示由邻接矩阵描述的关节点对间关系，图中红圈关节点的邻接关系规模为 \mathbb{R}^{VT} 。虽然该 GCN 可以通过邻接矩阵学习，不同空间位置、不同时间点的关节点对的联系。但这意味着在时空卷积层中，邻接矩阵的规模与数据时空维度相关 $\mathbf{A}^{ST} \in \mathbb{R}^{VT \times VT}$ ，这将导致模型规模剧烈膨胀，严重降低模型的运行效率。同时，如此规模的网络也使得训练过程更加困难，容易出现欠拟合现象。因此该设计并没有被实际应用。

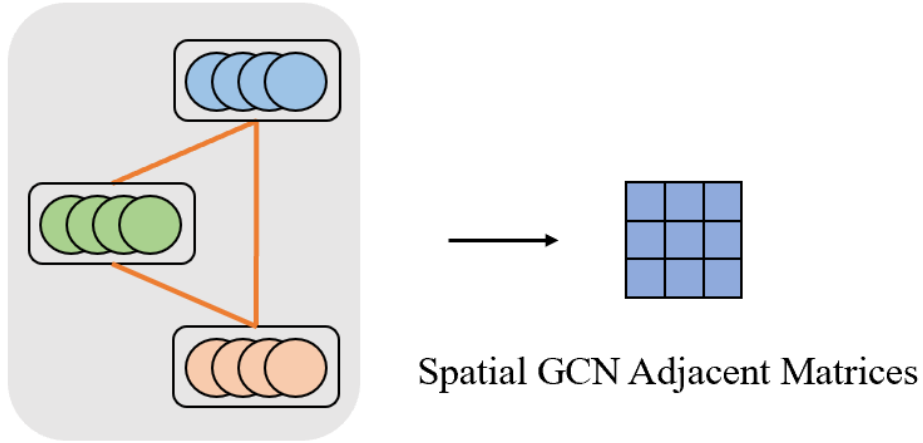


图 5-3 LTD 中的图卷积模块

作为率先使用 GCN 的方法，LTD 提出：忽略在时序空间上对输入数据的时序结构进行建模，而是将关节点运动序列看作图卷积网络中的节点值，在特征空间上通过权值矩阵 W 对齐进行变换。具体 GCN 结构由图5-3可见，其中由多个圆形组成的序列代表关节点运动序列，而它们被当作节点放入无向图中。该设计中，由于将关节点运动序列看作节点值 $\mathcal{V}_S = \{v | v \in \mathbb{R}^T\}$ ，时序数据被组织为一个节点值为关节点运动序列的空间无向图。避免了对时间节点间的邻接关系进行建模，因此如公式5-2所示，只需要一个空间上的 GCN 对人体空间结构进行处理，大大降低模型的空间复杂度。

$$\mathbf{H}^{(l+1)} = \mathbf{A}_S^{(l)} \mathbf{H}^{(l)} \mathbf{W}^{(l)}, \quad (5-2)$$

然而，该方法仅仅对人体空间结构进行建模，忽略了数据时序上的联系，时空信息提取能力并不完备，制约了模型对输入信息的提取与利用。

和 LTD 同期的工作 ST-GCN^[49]提出了不同的思路，它使用了分离输入数据时间维度和空间维度的策略，分别设计算子对时间和空间信息进行提取。如图5-4所示。图左部分，为空间维度特征提取步骤，ST-GCN 根据人体空间结构特征使用 GCN 进行建模，且通过参数共享机制降低模型的参数量，同一个运动序列，位于不同时间点的人体姿态

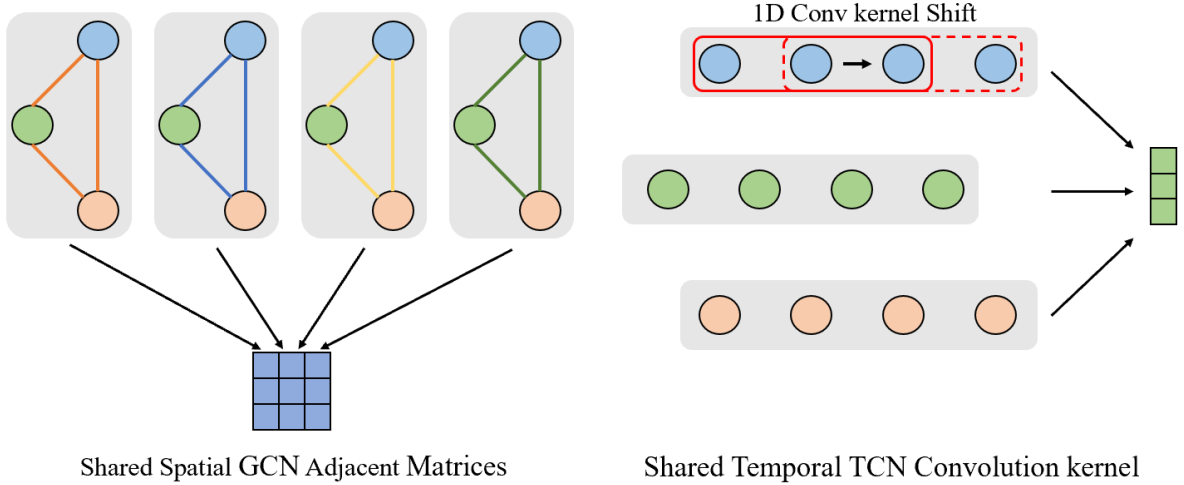


图 5-4 ST-GCN 中的图卷积模块

共享一个图卷积网络，因为人体空间结构先验信息是普适通用的，不会随着时间位置的不同而变化，因此对位于不同时刻的人体姿态，可以使用同一个 GCN 处理。图右展示了时间维度特征提取步骤，不同于人体空间结构属于不规则无向图，时序上的关节点轨迹可以看作一条 1D 的规则数据，因此可以使用适用于 1D 结构数据建模的 TCN^[45]算子，TCN 是一种继承自 CNN 的模型，其空间不变性特点得以保留。在进行空间特征提取时，TCN 采用了参数共享的思路，以便将不同空间位置的关节点运动轨迹共享一个 TCN。与 CNN 类似，TCN 的卷积核也沿时间维度滑动，以提取时间信息。然而，这种设计导致模型在时间维度上的感受野受限于卷积核的尺度，从而限制了其长时依赖捕捉能力。综上，其卷积模块以公式 5-3 形式表示，输入数据在经过 GCN 提取空间信息后，再由 TCN 提取时间信息，最终间接地捕捉到了数据中的时空信息。虽然 ST-GCN 较 LTD，补全了模型在时序信息提取能力上的缺陷，但受限于是感受野的局部性，在长时依赖捕捉能力上任然有所欠缺。

$$\mathbf{H}^{(l+1)} = TCN(\mathbf{A}_S^{(l)} \mathbf{H}^{(l)} \mathbf{W}^{(l)}), \quad (5-3)$$

近来，STS-GCN^[56]提出了一种在时空维度均使用图卷积网络的方法，它扩展了图卷积网络（GCN）并添加了捕捉时间序列上人体姿势演变的时间卷积。与 ST-GCN 类似，它同样将时间和空间维度拆分。在空间维度上，它使用传统的图卷积网络对人体结构信息进行建模。在时间维度，与 ST-GCN 将关节点运动序列看作规则的 1D 数据不同，它选择用图结构来定义时序上关节点对的连接关系。具体的，它将关节点运动序列也视为一个无向不规则的图结构，图中的关节点不光被允许与邻接节点产生联系，还可以和

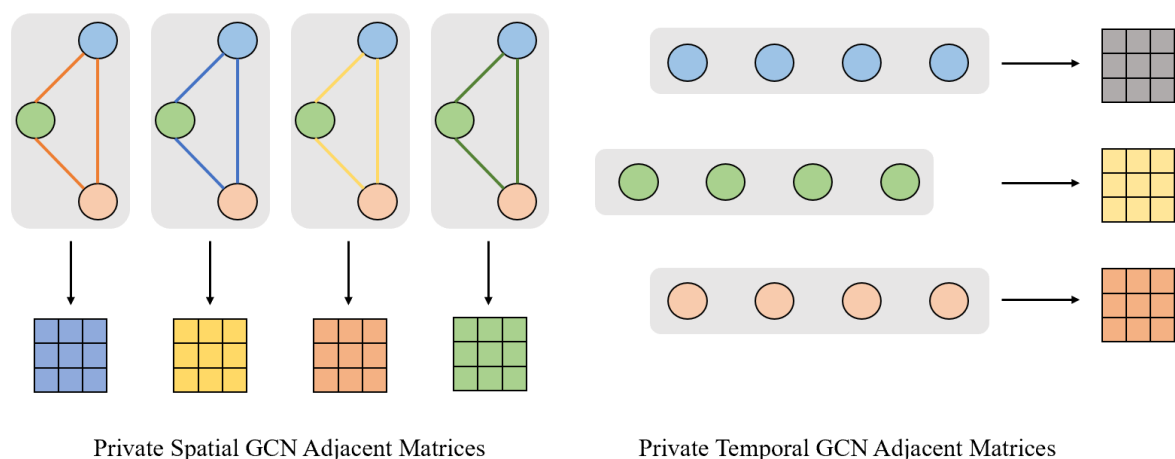


图 5-5 STS-GCN 中的图卷积模块

序列中任意一个关节点建立邻接关系。这使得网络拥有了全局的时序信息感知能力，能捕捉更长时间范围内的运动信息。这将有助于网络从整体上理解输入运动序列，从而对未来运动序列做出更准确的预测。然而，与 ST-GCN 中同一时间维度或空间维度中的数据共享同一个特征提取模块不同，STS-GCN 为不同时刻的人体姿态和不同空间位置的关节点运动轨迹设计了私有的特征提取模块。如图5-5中所示。对于不同时刻的人体姿态，STS-GCN 为每一个都分配了一个不共享空间图卷积。对于不同空间位置的关节点序列，则由对应的、私有的时间图卷积提取特征。这虽然在一定程度上提高了网络的冗余度。但由于不同的人体姿态和关节点轨迹在空间或时间上都具有相同的空间结构或时序位置，它们对网络来说是平等的，这种平等性的存在使得网络可以更加普适地处理不同的人体姿态和关节点轨迹，而无需为每种情况设计特定的特征提取模块。

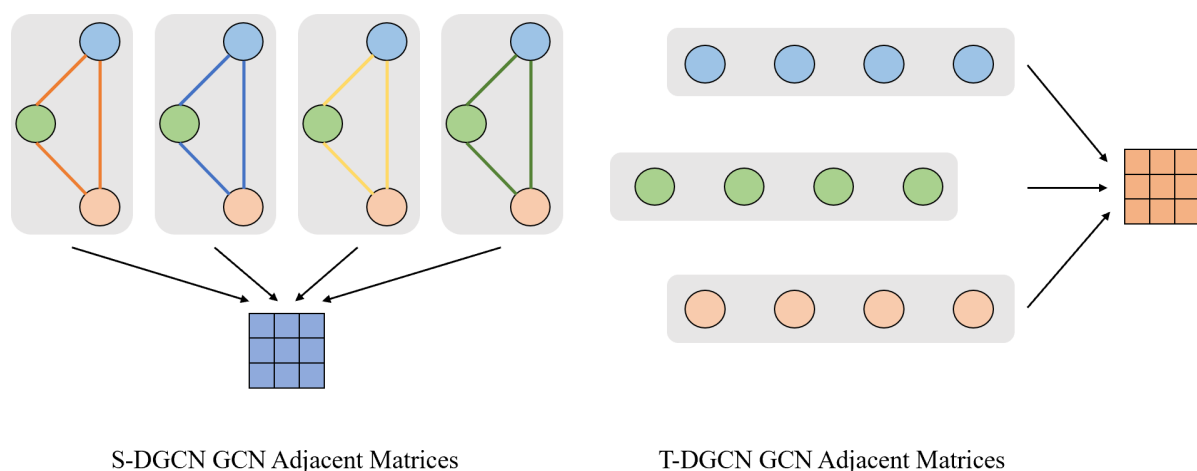


图 5-6 基于时空分离策略的 Non-Local 时空图卷积模块

我们则基于对以上方法的分析提出了一种基于时空分离策略的 Non-Local 时空图卷积模块。

第六章 模板简介

与很多外文杂志社不同，大部分中文期刊都不提供 L^AT_EX 模板给投稿者使用，也很少有学校给学生提供官方的毕业论文模板。目前 github 上的大部分模板都是由学生发起的非官方模板。在此感谢 Shun Xu 以及 yecfly 等人的工作，他们的无私奉献使得华南理工大学硕博士毕业论文也可以使用 L^AT_EX 撰写。

本模板是直接修改前人的模板得到的，更详细的介绍可到 [57-58] 下载。本章仅从用户的角度简要介绍模板的使用，而尽量避免涉及 L^AT_EX 的模板制作细节（实际上是因为本人也不会）。正如我们使用手机并不需要了解麦克斯韦方程组，使用 L^AT_EX 写作也无需了解模板是如何制作的。

L^AT_EX 的源代码保存在后缀名为 .tex 的文件中。当编写长篇文档时，例如当编写书籍、毕业论文时，单个源文件会使修改、校对变得十分困难。将源文件分割成若干个文件，例如将每章内容单独写在一个文件中，会大大简化修改和校对的工作。为方便，本文将 scutthesis.tex 文件称为主文件，而将 abstract.tex、chapter0x.tex、conclusion.tex 等文件称为章节文件。

值得注意的是，要每次编译时都更新参考文献著录，TeXstudio 软件的选项-> 设置中的构建并查看、编译器需要设置成如图6-1、6-2所示。此时只需在任意一个文件中点击构建并查看按钮即可编译文档。每次编译都更新参考文献会使得编译时间很长。

6.1 主文件

scutthesis.tex 文件相当于主函数，调用各章的内容。L^AT_EX 源代码以一个 \documentclass 命令作为开头，它指定了文档使用的文档类。文档类规定了 L^AT_EX 源代码所要生成的文档的性质——普通文章、书籍、演示文稿、个人简历等等。

```
\documentclass[ options ]{ class-name }
```

其中 class-name 为文档类的名称，如 L^AT_EX 提供的 article, book, report，可在其基础上派生的一些文档类或者有其它功能的一些文档类。L^AT_EX 提供的基础文档类见文献 [59]。还可以自定义文档类，如华南理工大学硕博士论文文档类 scutthesis，其实现保存在后缀名为 .cls 的文件中。可选参数 options 为文档类指定选项。

document 环境当中的内容是文档正文：

```
\begin{document}
正文内容
\end{document}
```

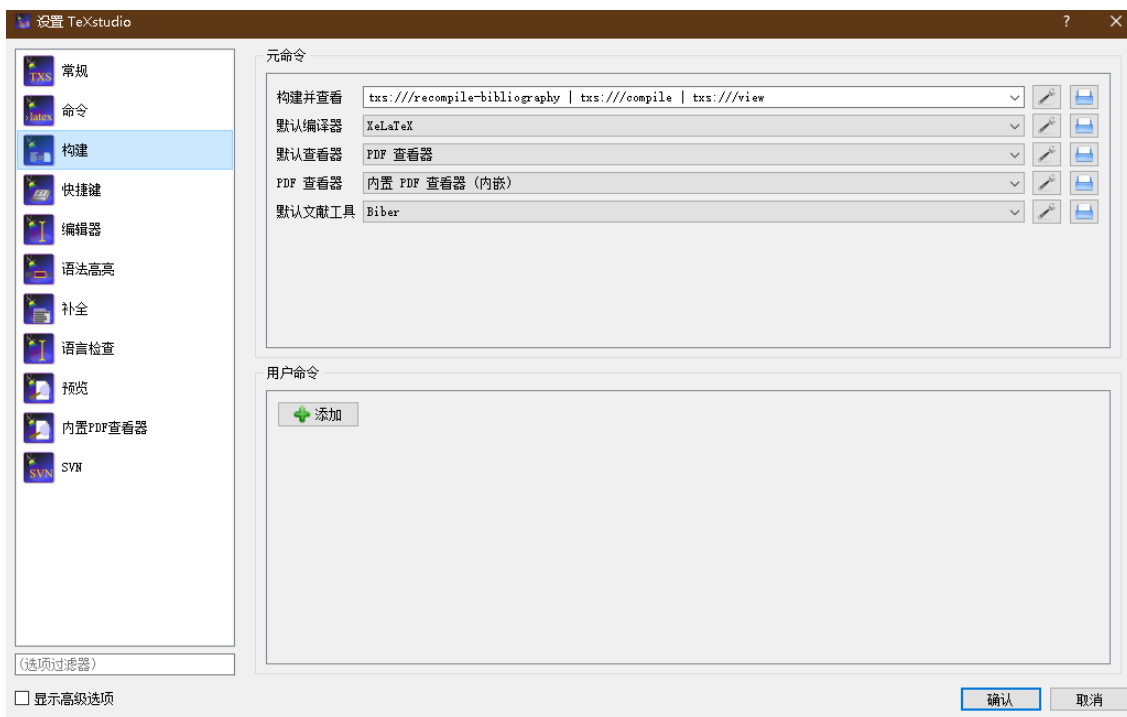



图 6-1 TeXstudio 环境

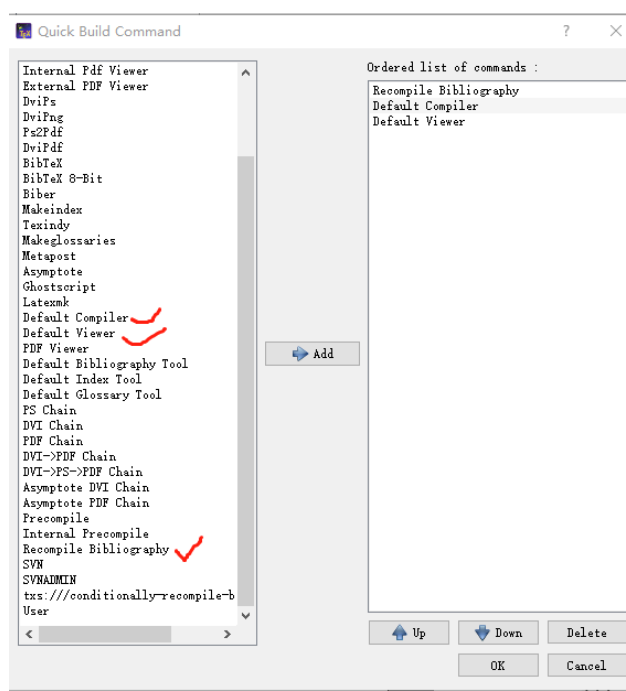


图 6-2 TeXstudio 编译选项

正文中包含各章节内容:

```
\include{abstract} % 中英文摘要
\tableofcontents   % 目录
\listoftables      % 表格目录 (可选)
\listoffigures     % 插图目录 (可选)
```



```
\include{symbols} % 符号对照表(可选)
\include{abbreviation} % 缩略词
...
\include{chapter01} % 第一章
\include{chapter02} % 第二章
\include{chapter03} % 第三章
% 自行根据需要添加章节。
...
\include{conclusion} % 结论
...
\printbibliography % 参考文献著录
\include{appendix} % 附录
\include{pub} % 成果
\include{ack} % 致谢
```

其中 % 之后的内容为注释, ... 表示省略其他代码, 仅保留论文内容主体部分。
`\include{xxx}` 指令用于包含 `xxx.tex` 文件的内容, 各章节的内容主要在 `xxx.tex` 中保存。在 `\documentclass` 和 `\begin{document}` 之间的位置称为导言区。在导言区中一般会使用 `\usepackage` 调用宏包, 以及会进行对文档的全局设置。本模板的导言区除调用所需的宏包外, 还进行了页眉页脚的设置。有的模板会把所有调用宏包的指令放到一个 `.sty` 宏包文件中, 页面的设置放在文档类文件 `.cls` 文件中。因本人时间有限, 就不做整理, 欢迎有志之士加入完善。使用本模板并不需要了解导言区的指令, 在需要时额外添加即可(要注意宏包冲突)。特别地, `\includeonly{xxx}` 指令用于使文档仅编译 `xxx.tex` 文件的内容, 这就是分章节包含 (`include`) 的好处, 可大大减少编译时间。

将封面打印保存为 `thesis_cover.pdf` 文件, 硕士使用 `master_cover.docx`, 博士使用 `doctor_cover.doc`。如果有更新版本的封面, 可自行替换。文档类默认是博士论文, 下面指令将控制添加封面与否:

```
\documentclass[unicode, master, pdfcover]{scutthesis} % 使用pdf文件封面的 硕士模板
\documentclass[unicode, master]{scutthesis} % 不使用pdf文件封面的 硕士模板
\documentclass[unicode, pdfcover]{scutthesis} % 使用pdf文件封面的博士模板
\documentclass[unicode]{scutthesis} % 不使用pdf文件封面的博士模板
```

不使用 `thesis_cover.pdf` 文件指定的封面时, 将使用草稿封面。草稿封面也可以减少编译时间, 因此可以在最终提交论文时再使用论文封面。草稿封面用以下指令设置:

```
%%%%%%%%%%%%草稿封面设置%%%%%%%%%%%%
\title{LaTeX模板}
\author{蒙超恒}
\supervisor{指导教师: 裴海龙\ 教授}
\institute{华南理工大学}
\date{2020年5月20日}
%%%%%%%%%%%%
```

6.2 章节文件

章节文件如 `chapter0x.tex` 等，其内容由 `\chapter{章名}` 开头。新建一章可新建一个文件并由 `\chapter{新建章名}` 开头填写内容即可。节及小节分别用 `\section{新建节名}`、`\subsection{新建小节名}` 命令。

正文的的书写和 `txt` 文本文件的书写类似。 \LaTeX 源代码中，空格键和 `Tab` 键输入的空白字符视为“空格”。连续的若干个空白字符视为一个空格。一行开头的空格忽略不计。行末的回车视为一个空格；但连续两个回车，也就是空行，会将文字分段。多个空行被视为一个空行。也可以在行末使用 `\par` 命令分段。在本模板中，英文之间的空格被保留，中文之间的空格被忽略。特别地，摘要，附录，结论等两个字的大纲级别为章的章名，中间使用空格隔开。对此论文撰写规范并没有明文要求，只是为了美观。也可以全部不加空格。一般情况下，在文本文字中添加空格使用 `\quad` 命令，但由于文献 [60] 所述原因，直接使用 `\quad` 命令会报警，因而使用 `\texorpdfstring{\quad}{}{}`，其中最后一个 `{}` 里面可以加一个空格，不影响使用。目录二字之间添加空格在 `scutthesis.cls` 文件 317 行设置。

正文本环境中使用公式，即行内公式，需要用两个 `$` 包围，如源码：`$a+b=c$` 显示为 $a + b = c$ 。使用其他字符可自行百度或阅读参考文献。再次提醒，使用 \LaTeX 撰写论文不需要研究其原理，在达到某种效果（图文显示、公式显示效果）时百度或查书寻找其代码即可。

综上，论文撰写只需要将自己的文本（包含行内公式）放到相应的章节处，并添加行间公式、图表环境并填写图表即可。行间公式、图表将在下一章介绍。

第七章 常用环境及参考文献设置

强烈建议在使用公式、表格、定理环境时进行百度，没必要研究各种用法，只需要知道自己需要什么。因本人的论文所用表格较少，因而对表格不是很熟悉，本章对表格的介绍相应的较少。本章仅介绍本人在论文撰写过程中常用的环境以及参考文献设置。

7.1 图

图的导入需要提前准备好图片文件，最好是.png、.eps、.pdf 或.jpg 文件。另外，如果是从 matlab 导出图片文件，可使用 print 函数或手动导出，print 函数的使用可参考 ICGNC2020plot.m 以及 PlotToFileColorPDF.m 文件等。手动导出（matlab 的 figure 界面的“文件”->“导出设置”设置好大小、分辨率和线宽等然后点击“应用于图窗”）主要用于观察效果，可设置某种样式名称后保存该样式，下次使用时加载，具体可百度“matlab 导出高清图片”。需要特别注意的是一定要 1:1 导入 matlab 生成的图片，并且图中文字设置好字体字号。否则缩放之后，图片的字号就变了，盲审老师一眼就能看出来字号不对，就很麻烦。这就是为什么要在 matlab 点击“应用于图窗”进行预览，观测效果后再 1:1 使用图片。

使用如下代码放置独立成行的图片，效果如图7-1所示

```
\begin{figure}[htbp]
% 图片居中（列居中对齐）
\centering
% 包含当前路径下的Fig文件夹的图片文件DFUAV_f31.png
\includegraphics[scale=1]{Fig/DFUAV_f31.png}
% 添加标签one_DFUAV以及图标题“涵道风扇式无人机”，引用某图时使用\ref{xxx}，其中xxx就是
% 标签，图编号是自动生成的。
\caption{\label{one_DFUAV}涵道风扇式无人机}
\end{figure}
```

其中 figure 为环境名，[htbp] 表示将图片设置为浮动体，实际上这在.cls 文件已经设置过，因而可以省略。[scale=1] 表示安装 1:1 的比例导入图片，还可以按其他方式导入，需要时可自行百度。

使用如下代码划分页面并排放置图7-2、图7-3

```
\begin{figure}[htbp]
\centering
\begin{minipage}[c]{0.5\textwidth} % minipage将页面划分为0.5\textwidth
\centering
\includegraphics[width=6cm,height=6cm]{Fig/honeywell_t-hawk.jpg}
\caption{\label{Hawk}T-Hawk}
\end{minipage}%
\begin{minipage}[c]{0.5\textwidth}
\centering
```



图 7-1 涵道风扇式无人机

```
\includegraphics[width=6cm,height=6cm]{Fig/GTSpy.jpg}
\caption{\label{GTSpy}GTSpy}
\end{minipage}
\end{figure}
```

其中 [c] 表示行居中对齐。当图片大小不一但又需要 1:1 导入时，图标题可能行不对齐，因此可以改为如下指令：

```
\begin{figure}[htbp]
\centering
\begin{minipage}[c]{0.5\textwidth}
\centering
\includegraphics[scale=1]{Fig/honeywell_t-hawk.jpg} %1:1导入
\end{minipage}%
\begin{minipage}[c]{0.5\textwidth}
\centering
\includegraphics[scale=1]{Fig/GTSpy.jpg}
\end{minipage}\\[1pt]
\begin{minipage}[t]{0.5\textwidth} % 以下为新添加页面划分，[t]表示行顶部对齐
\caption{\label{Hawk}T-Hawk}
\end{minipage}%
\begin{minipage}[t]{0.5\textwidth}
\caption{\label{GTSpy}GTSpy}
\end{minipage}%
\end{figure}
```

通常一个 figure 内含有其他小的 figure，可以使用一些宏包，但最初本着简单的原则，本模板并没有使用这些子图包。后来应同学们要求在，把子图的功能加上，主要是修改了模板文件（scutthesis.cls 文件）的功能包参数。注意，很多网上拿到的代码不一定可以精确的调子图标题字体字号，因为此模板的子图标题字体字号是利用 subfig 宏包的选项进行设置的（在 scutthesis.cls 文件的“图表环境”中），而有些教程使用 subcaption



图 7-2 T-Hawk

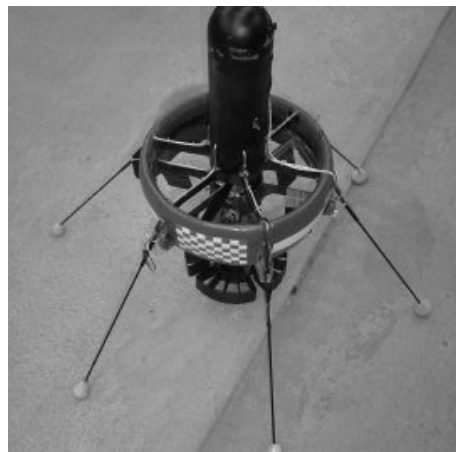


图 7-3 GTSpy

进行同样的设置，还需进一步验证可行性。另外图的排版方法很多，有些宏包已经被弃用，所以尽量使用本文给出的案例的格式进行排版图片。

常见的子图包有 `subfigure` 和 `subfig`。`subfigure` 是比较老的了，这里使用 `subfig` 包。两个包在使用的时候用法不同，千万不要混淆了，不然可能会报错。`subfig` 包的命令是 `\subfloat`。这里给出一种使用 `subfig` 包的常用排版，如图 7-4 的子图 7-4 b)，其中 a) 的试验并不好（这里测试了交叉引用 `\subref{xxx}` 和 `\subref*{xxx}`）。必要时也可以排版多行多列的图、调整图之间的间距，具体可百度。

```
\begin{figure}[!h]
  \centering
  \subfloat[不合理的轨迹]{\includegraphics[width=6cm,height=6cm]{Fig/Figure_1.png}%
    \label{Fig:1:a}}
  \subfloat[优化的轨迹]{\includegraphics[width=6cm,height=6cm]{Fig/Figure_2.png}
    \label{Fig:1:b}}
  \\ % 用 \\ 换行，也可以此处空一行进行换行，只有两个图的话下面就不需要了。
  \subfloat[不合理的轨迹]{\includegraphics[width=6cm,height=6cm]{Fig/Figure_1.png}%
    \label{Fig:1:c}}
  \subfloat[优化的轨迹]{\includegraphics[width=6cm,height=6cm]{Fig/Figure_2.png}%
    \label{Fig:1:d}}
  \caption{子图包使用测试}\label{Fig:1}
\end{figure}
```

% 引用某子图时使用 `\subref{xxx}`，其中 xxx 就是标签 Fig:1:a

子图的引用比较特殊，命令有：`\subref{xxx}` 和 `\subref*{xxx}`

注：在 `subfig` 包使用说明中，`\subref{xxx}` 和 `\subref*{xxx}` 分别由参数 `listofformat` 和 `subrefformat` 控制，

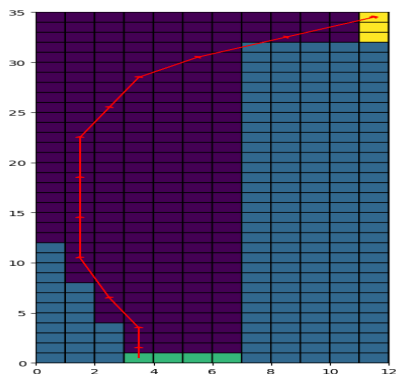
并由如下定义，根据撰写规范需要定义为：

```
\DeclareSubrefFormat{empty}{}
\DeclareSubrefFormat{simple}{#1#2}
\DeclareSubrefFormat{parens}{#1 #2)}
\DeclareSubrefFormat{subsimple}{#2}
\DeclareSubrefFormat{subparens}{ #2)}
```

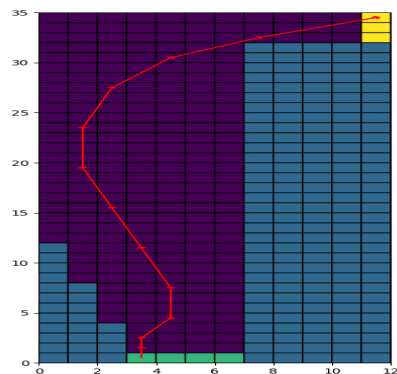
和

```
\DeclareCaptionListOfFormat{empty}{}
\DeclareCaptionListOfFormat{simple}{#1#2}
```

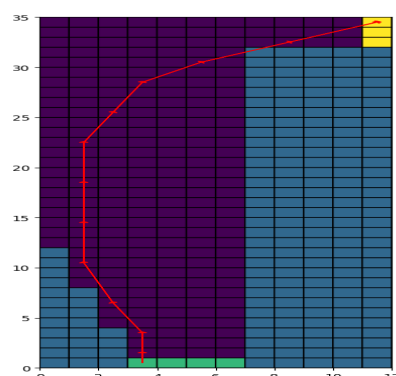
```
\DeclareCaptionListFormat{parens}{#1 #2)}
\DeclareCaptionListFormat{subsimple}{#2}
\DeclareCaptionListFormat{subparens}{ #2)}
```



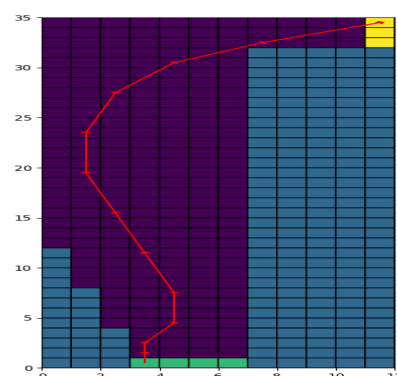
a) 不合理的轨迹



b) 优化的轨迹



c) 不合理的轨迹



d) 优化的轨迹

图 7-4 子图包使用测试

7.2 表

本节仅展示使用常见的三线表

```
\begin{table}
\caption{\label{TDF_para}涵道模型参数} %表题在上
\centering % 表居中
\small % 表内字体小一号（即设置成和表题字号一致）
\begin{tabular}{cccc} % cccc表示4列并居中，若列之间需要分隔符则设置为|c|c|c|c|
\hline % \hline表示横线。列之间的元素用&分隔，\tabularnewline表示换行
参数符号 & 数值 & 参数符号 & 数值 \tabularnewline
\hline
$I_x$ & $0.054593$ & $I_y$ & $0.017045$ \tabularnewline
$l_1$ & $0.0808$, \text{m} & $l_2$ & $0.175$, \text{m} \tabularnewline
$l_4$ & $0.2415$, \text{m} & $l_5$ & $0.1085$, \text{m} \tabularnewline
\hline
\end{tabular}
```

```
\end{tabular}
\end{table}
```

表 7-1 涵道模型参数

参数符号	数值	参数符号	数值
I_x	054593	I_y	0.017045
l_1	0.0808 m	l_2	0.175 m
l_4	0.2415 m	l_5	0.1085 m

7.3 公式

除了前面讲行内公式，常用的还有行间公式。公式中的数学符号可自行百度，本章仅介绍常用的几种公式环境。

单独成行的行间公式在 \LaTeX 里由 `equation` 环境包裹。`equation` 环境为公式自动生成一个编号，这个编号可以用 `\label` 和 `\ref` 生成交叉引用，`amsmath` 宏包的 `\eqref` 可为引用自动加上圆括号；如式(7-1)所示。

```
\begin{equation}
a+b=c \quad \label{eq_1}
\end{equation}
```

$$a + b = c \quad (7-1)$$

若不需要编号则加星号，改为

```
\begin{equation*}
a+b=c
\end{equation*}
```

其他环境类似。当使用 `$` 开启行内公式输入，或是使用 `equation` 环境时， \LaTeX 就进入了数学模式。数学模式相比于文本模式有以下特点：

1. 数学模式中输入的空格被忽略。数学符号的间距默认由符号的性质（关系符号、运算符等）决定。需要人为引入间距时，使用 `\quad` 和 `\qquad` 等命令。
2. 不允许有空行（分段）。行间公式中也无法用 `\\` 命令手动换行。排版多行公式需要用到其他各种环境。
3. 所有的字母被当作数学公式中的变量处理，字母间距与文本模式不一致，也无法生成单词之间的空格。如果想在数学公式中输入正体的文本，简单情况下可用

`\mathrm` 命令。或者用 `amsmath` 提供的 `\text` 命令（仅适合在公式中穿插少量文字。如果你的情况正好相反，需要在许多文字中穿插使用公式，则应该像正常的行内公式那样用，而不是滥用 `\text` 命令）。

实际上更常用的的是多行公式，不需要对齐的公式组可以使用 `gather` 环境，需要对齐的公式组用 `align` 环境。长公式内可用 `\\` 换行。

如果需要罗列一系列公式，并令其按照等号对齐，可用 `align` 环境，它将公式用 `&` 隔为两部分并对齐。分隔符通常放在等号左边：

```
\begin{align}
  a &= b + c \\
  &= d + e
\end{align}
```

$$a = b + c \tag{7-2}$$

$$= d + e \tag{7-3}$$

`align` 环境会给每行公式都编号。

如果不需要按等号对齐，只需罗列数个公式，可用 `gather` 环境：

```
\begin{gather}
  a = b + c \notag \\
  f = d + e
\end{gather}
```

$$a = b + c$$

$$f = d + e \tag{7-4}$$

`gather` 环境同样会给每行公式都编号，如果某行不需要编号可在行末用 `\notag` 仅去掉某行的编号。

`align` 和 `gather` 有对应的不带编号的版本 `align*` 和 `gather*`。

另一个常见的需求是将多个公式组在一起公用一个编号，编号位于公式的居中位置。为此，`amsmath` 宏包提供了诸如 `aligned`、`gathered` 等环境，与 `equation` 环境套用。以 `-ed` 结尾的环境用法与前一节不以 `-ed` 结尾的环境用法一一对应。我们仅以 `aligned` 举例：

```
\begin{equation}
  \begin{aligned}
    a &= b + c \\
    d &= e + f + g \\
    h + i &= j + k
  \end{aligned}
\end{equation}
```

```
l + m &= n
\end{aligned}
\end{equation}
```

$$a = b + c$$

$$d = e + f + g$$

$$h + i = j + k$$

$$l + m = n$$

(7-5)

`split` 环境和 `aligned` 环境用法类似，也用于和 `equation` 环境套用，区别是 `split` 只能将每行的一个公式分两栏，`aligned` 允许每行多个公式多栏。

分段函数通常用 `amsmath` 宏包提供的 `cases` 环境，可参考文献 [59]

`amsmath` 宏包还直接提供了多种排版矩阵的环境，包括不带定界符的 `matrix`，以及带各种定界符的矩阵 `pmatrix`、`bmatrix`、`Bmatrix`、`vmatrix`、`Vmatrix`。其中中括号版的 `bmatrix` 最常用。这些矩阵环境需要在公式中使用，比如 `gather` 环境。

```
\begin{gather}
A= \begin{bmatrix}
x_{11} & x_{12} & \ldots & x_{1n} \\
x_{21} & x_{22} & \ldots & x_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
x_{n1} & x_{n2} & \ldots & x_{nn}
\end{bmatrix}
\end{gather}
```

$$A = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{bmatrix} \quad (7-6)$$

其中矩阵/向量加粗使用 `\bm{}` 命令。另外还可以使用 `array` 环境排版矩阵，类似 `tabular` 环境，用 `\\` 和 `&` 用来分隔行和列，这里不再赘述。

```
\begin{array}{l}[外部对齐tcb]{列对齐lcr}
行列内容
\end{array}
```

另外注意排版分式时，有两种方法：`\frac` 或者 `\dfrac`，效果分别为 $\frac{1}{2}$ 和 $\frac{1}{2}$ 。以上介绍的数学环境中，空格可参考文献 [59]，例如常用 `\quad`。

需要局部更改字号时，可以使用 `tutorial` 文件夹 `lshort-zh-cn.pdf` 的 5.1 节进行更改，

如加\small 使得字号小一号。

7.4 定理

在 scutthesis.cls 文件的最后, 已经用\newtheorem 命令定义了几种定理环境, 包括: 定义、假设、定理、结论、引理、公理、推论、性质等等, 统称定理环境, 关于\newtheorem 的用法, 可参考^[59,61]或自行百度。要下面提供几个例子, 在横线之间的深色区域是代码, 效果在相应下方表示:

```
\begin{assumption}
  加权矩阵 $\{\bm{W}\}_1$ 和  $\{\bm{W}\}_2$  是对称矩阵, 且 $\{\bm{W}\}_2$ 非奇异。
  \label{assum_dca1}
\end{assumption}
```

假设 7.1: 加权矩阵 \mathbf{W}_1 和 \mathbf{W}_2 是对称矩阵, 且 \mathbf{W}_2 非奇异。

定理用法和假设类似:

```
\begin{theorem}
  如果假设\ref{assum_dca1}成立,  $\mathbf{F}$ 满足式\eqref{eq_F}的定义, 且 $\{\bm{W}\}_1$ 非
  奇异, 则有 $0 \leq e \left( \mathbf{F} \right) < 1$ , 其中 $e \left( \mathbf{F} \right)$ 是  $\mathbf{F}$ 
  的特征值。 \label{the_dca2}
\end{theorem}
```

定理 7.1: 如果假设7.1成立, \mathbf{F} 满足上式的定义, 且 \mathbf{W}_1 非奇异, 则有 $0 \leq e(\mathbf{F}) < 1$, 其中 $e(\mathbf{F})$ 是 \mathbf{F} 的特征值。

定理环境的编号可自定义, 但通常不需要再进行设置, 因为模板文件 scutthesis.cls 文件已经定义好。

2022 年 5 月更新:

根据最新的博士论文送审结果, 定理等环境统一把原来的斜体改成正体。在此引用一下参考文献^[61]的内容:

amsthm 提供了 \theoremstyle 命令支持定理格式的切换, 在用 \newtheorem 命令定义定理环境之前使用。amsthm 预定义了三种格式用于 \theoremstyle:plain 和 LATEX 原始的格式一致; definition 使用粗体标签、正体内容; remark 使用斜体标签、正体内容。

以上部分在 scutthesis.cls 文件最后一部分设置。

amsthm 还提供了一个 proof 环境用于排版定理的证明过程。proof 环境末尾自动加上一个证毕符号:

证明：显然有

$$E = mc^2$$

证毕

□

proof 的大更多用法见参考文献^[61]。scutthesis.cls 文件的最后，跟所有定理环境一样，只是把英文”Proof“改成中文“证明”。

7.5 参考文献

再次强调，使用其他参考文献管理软件的用户以及不使用任何软件的“裸奔”的用户不需要关注任何关于 zetero 的东西。

关于参考文献这块，很多同学有疑问。只有记住一点：不管用什么参考文献管理工具，最终目的是生成一个bib文件，bib文件里是特定格式的文献信息。bib文件当作文本打开，里面就是文献的元数据。

通常学位论文参考文献是基于 BibTeX 进行的，本模板最大的改进就是引入 BibLaTeX。关于这部分知识可参考文献^[59,61]的第六章，6.1 节参考文献和 BIBTEX 工具。

参考文献引用和著录是基于 ZOTERO 这个软件进行的。视频教程见^[62]。此外，为了符合毕业论文撰写规范，需设置参数。按照视频教程安装完必要的插件（如 Better BibTeX）后，在编辑->首选项进行设置。图7-5到图7-15所示的是我的 zotero 软件设置。其中最重要的是7-14的设置要排除的选项，多余的显示会让审稿人反感，按照论文撰写规范进行即可。在毕业论文撰写时，在编辑->首选项->Better BibLaTeX->Fields 中，Fields to omit from export 填 month,abstract,note,extra,file,keywords,type,url,doi，就是在参考文献著录中排除这些多余的项，避免过于复杂。而在写本模板使用说明时，没有排除 url，因为很多参考资料是网页。

使用zotero，科学上网很重要，通常我们使用谷歌学术搜索文献并利用chrome的zotero插件直接捕获文献著录信息。但我使用蓝灯，代理服务器均遇到过被谷歌学术封锁的情况。只能不断换科学上网方法。这里我现在用的chrome插件：谷歌上网助手，它可以轻松捕获谷歌学术的著录信息，注册一个账号即可使用。谷歌上网助手有可能和某些代理冲突。这些都是科学上网的问题，已经超出了本项目的范围，听说百度一下 v2ray 可发现新大陆，可惜我试了Vultr的服务器依然被谷歌封。知网捕获中文参考文献著录信息的话不需要考虑这个问题，直接在知网首页搜索文献然后点击插件既可以选想捕获的著录了。

在 zotero 软件点击文件->导出文献库，如图7-16所示，再在导出对话框图7-17选择导出格式为 Better BibLaTeX，同时勾选 Keep updated 选项保持自动更新，再点击 ok，在弹出的对话框图7-18确定保存路径和文件名，例如我的是 MyLibrary.bib，这也是我整

个读书生涯的文献库 bib 文件。如果写小论文的话通常导出格式是 BibTeX 或者 Better BibTeX（这里按照期刊的要求来即可，文献管理软件的好处就是快速自动生成一个文件库）。关于 BibTeX 和 BibLaTeX 的区别这里不做展开。

得到文献库后，在 scutthesis.tex 文件第九行使用 \addbibresource 命令，添加文献库。引用某文献时秩序在 zotero 选中某文献条目，然后按 Ctrl+Shift+C，复制引用关键字（Citation Key）到剪切板（快捷键可自定义）。然后在 tex 文件编辑界面直接粘贴，默认的时上标形式，若需要非上标形式，可以改为 \parencite{xxx}，其中 xxx 是 Citation Key。这里的操作和认为设置的首选项参数有关，需要在编辑->首选项->导出界面的默认格式一栏选中相应的项，同时在编辑->首选项->高级->快捷键设置为默认值。

2020 年 12 月 2 日测试：下载最新 zotero，从知网和谷歌捕获文献（刚打开网页最好稍等一会再点击插件，谷歌可能需要现人机验证），对文献 [63]、[64] 进行引用。

2021 年 9 月 14 日测试：使用 endnote 的用户也可以利用导出的 bib 文件生成参考文献著录信息，导出选项是 bibTeX，貌似没有更多导出设置选项。导出设置没有 zotero 那么灵活丰富，得到 bib 文件后要引用某论文需要自行查找标签（label，也有软件叫引用关键字 Citation Key）{xxx} 然后手打 \cite{xxx}。欢迎熟悉 endnote 的同学来信告诉我更好的办法。

另外有同学反映，换了电脑后重新导出的 bib 文件 Citation Key 值不同，记得设置好 Better BibTeX 之后，在著录条目界面全选著录（或仅选想更新的著录）然后右键选 Better BibTeX 更新 refresh 一下。然后在 Automatic export 选项点击 Export now 立即更新 bib 文件（按理说勾选了自动更新选项他会自动更新，但为了确保万无一失还是点一下）。

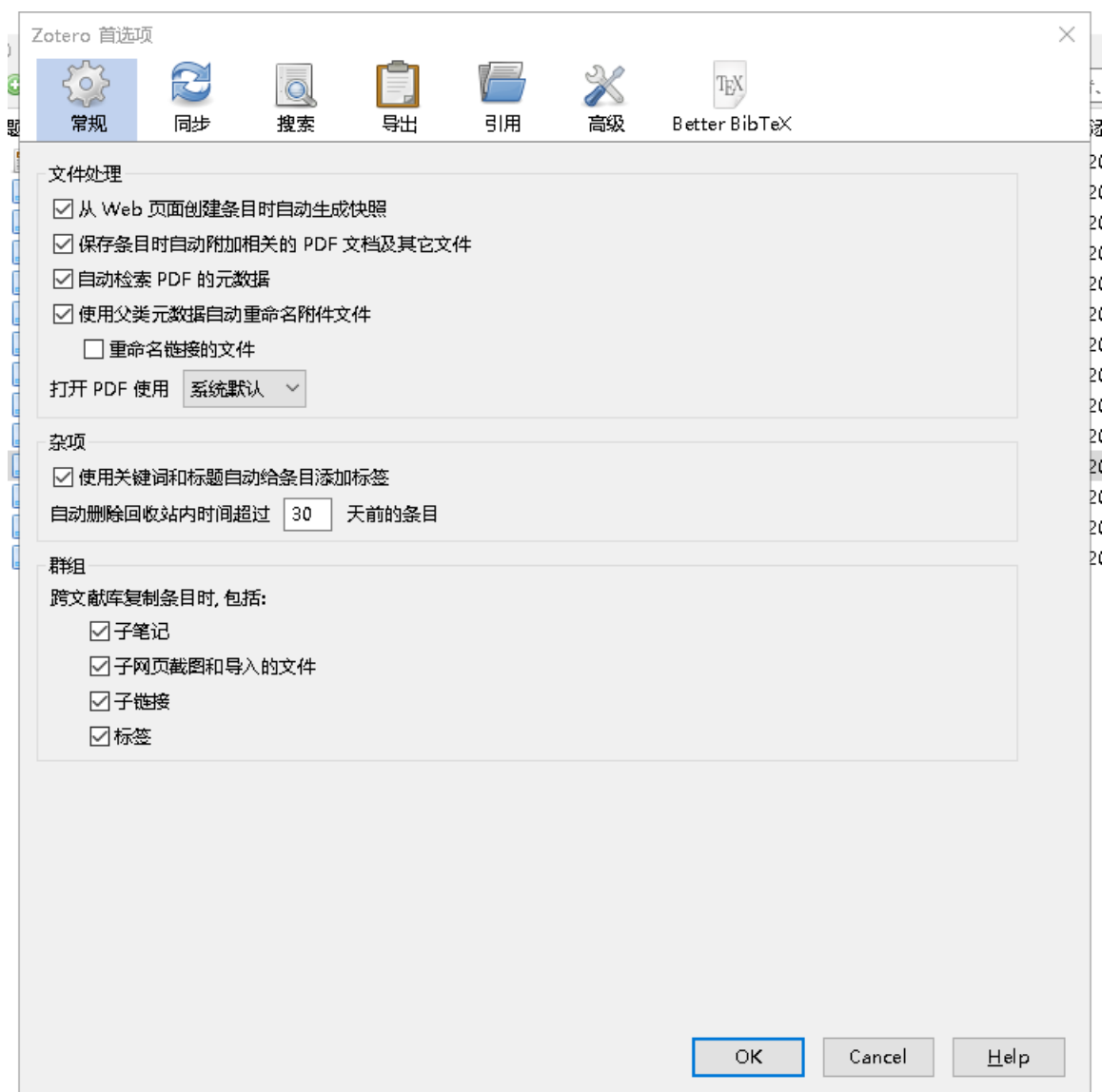


图 7-5 常规

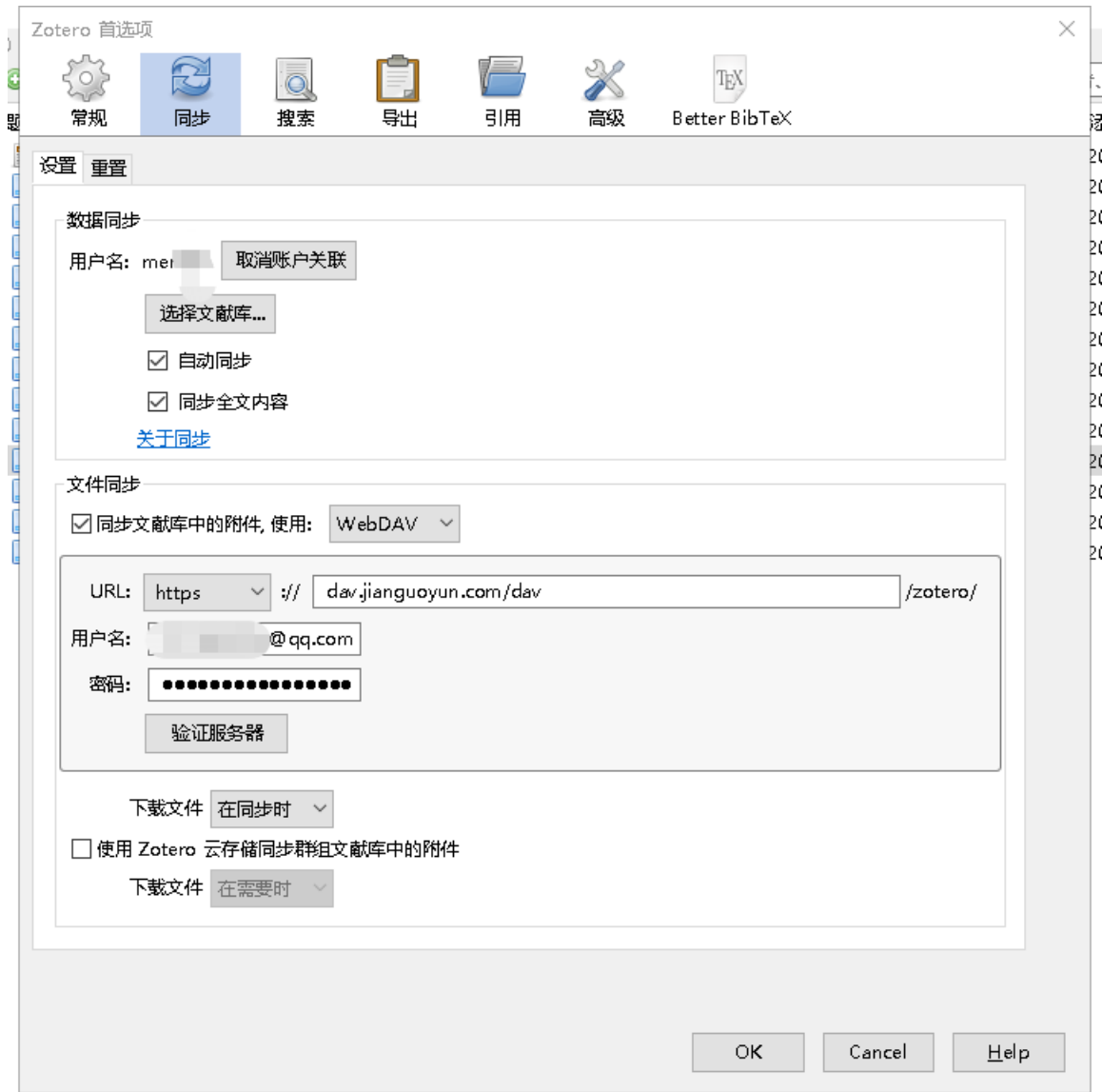


图 7-6 同步 1

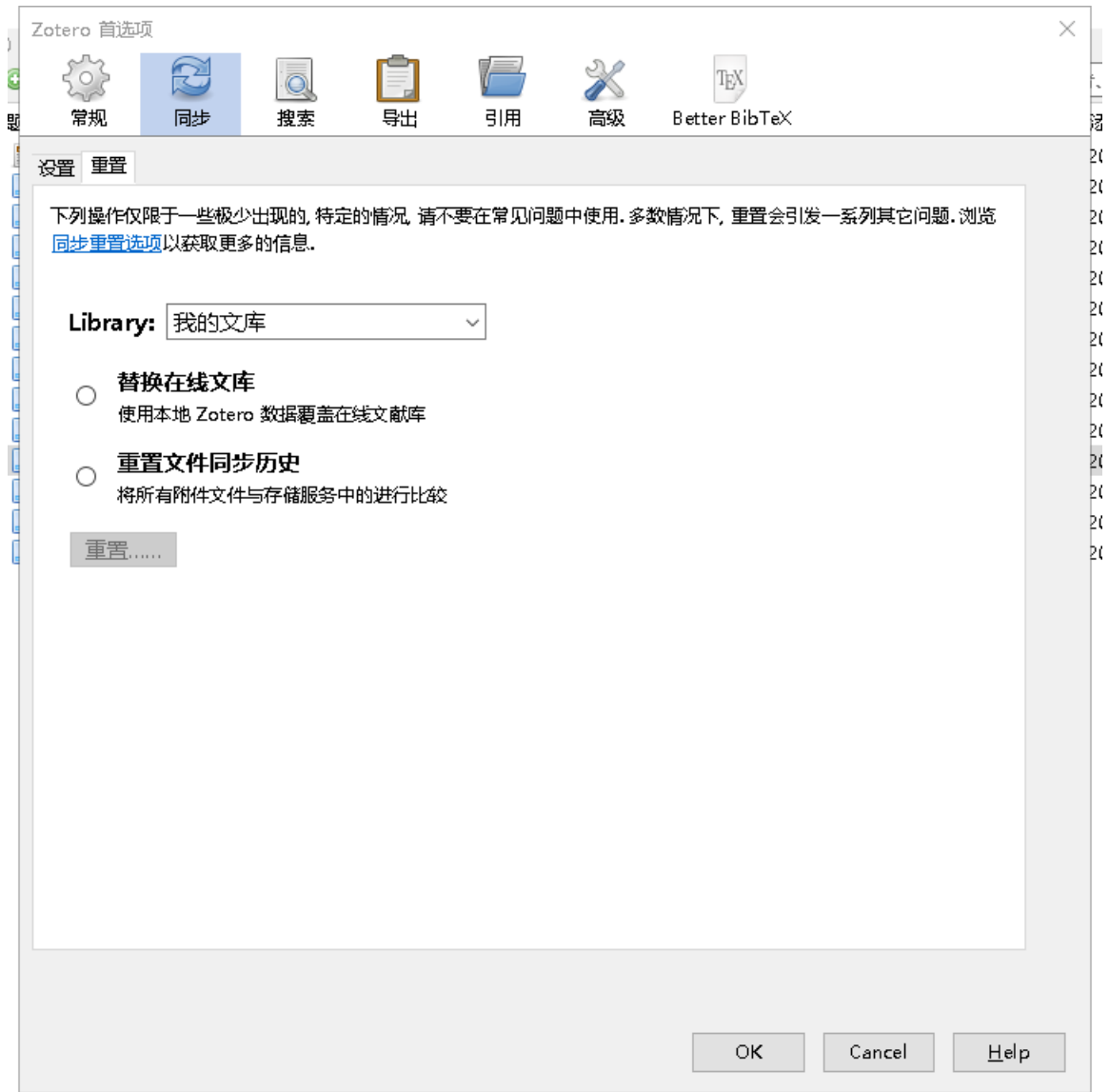


图 7-7 同步 2

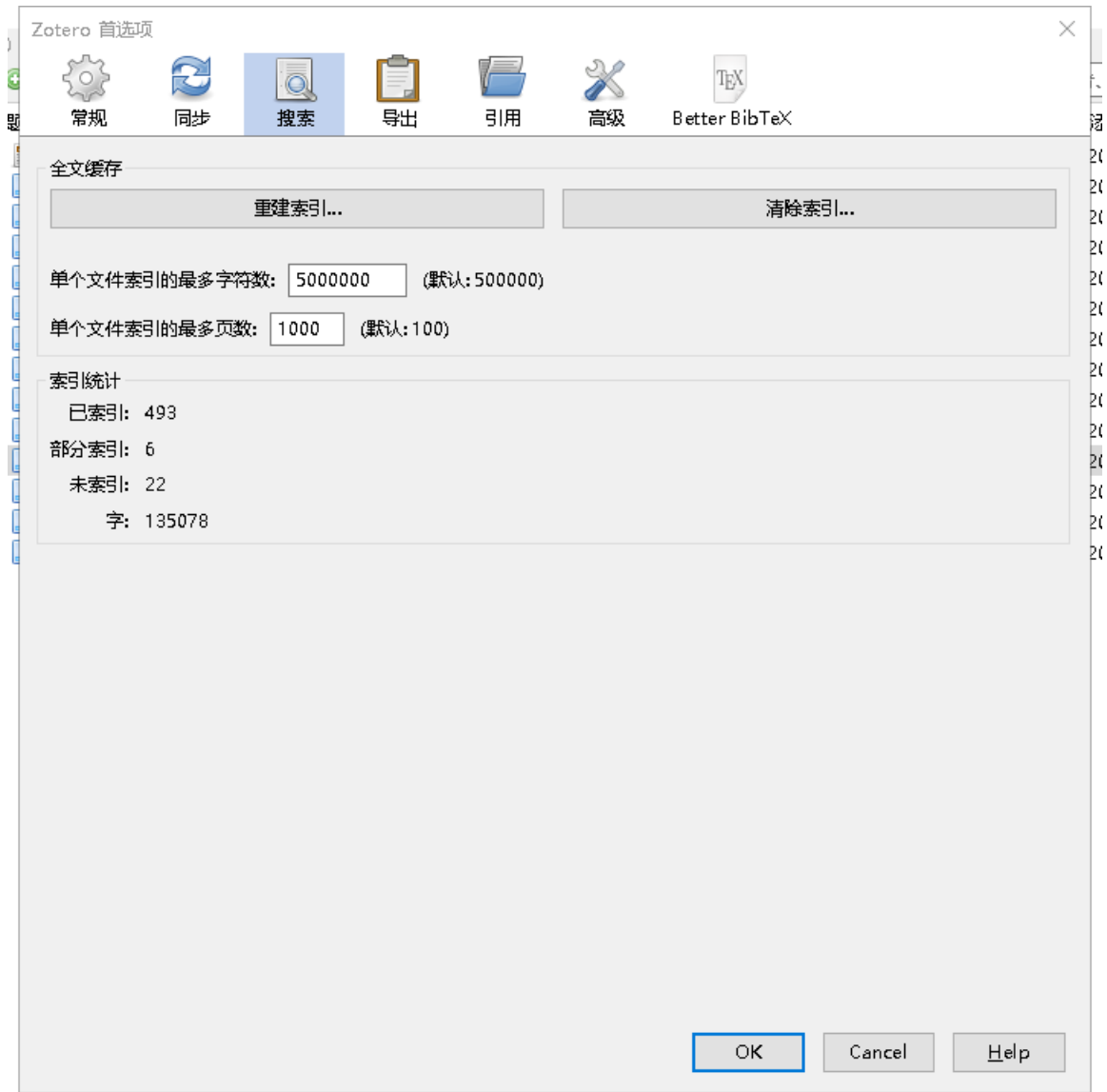


图 7-8 搜索

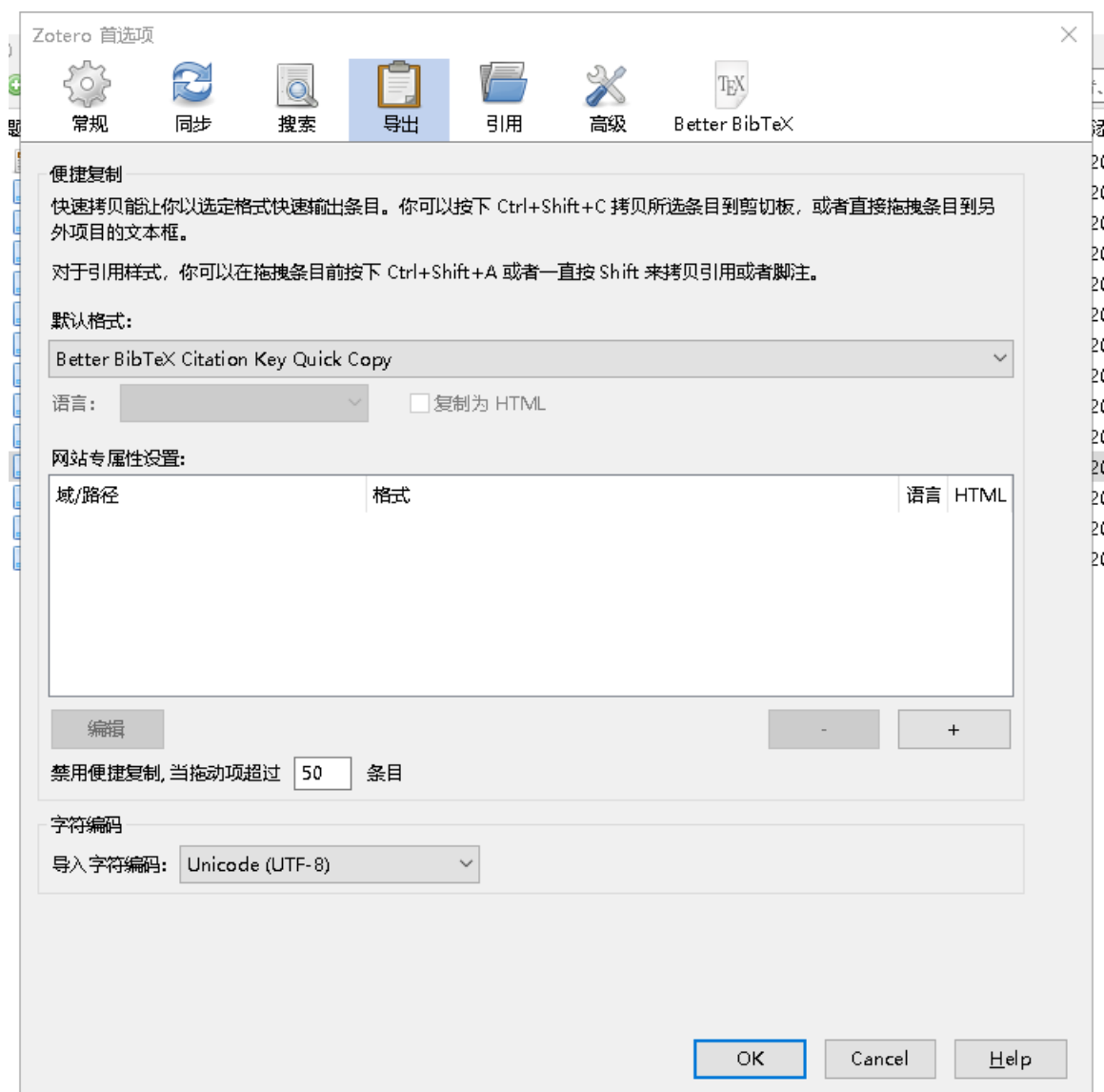


图 7-9 导出

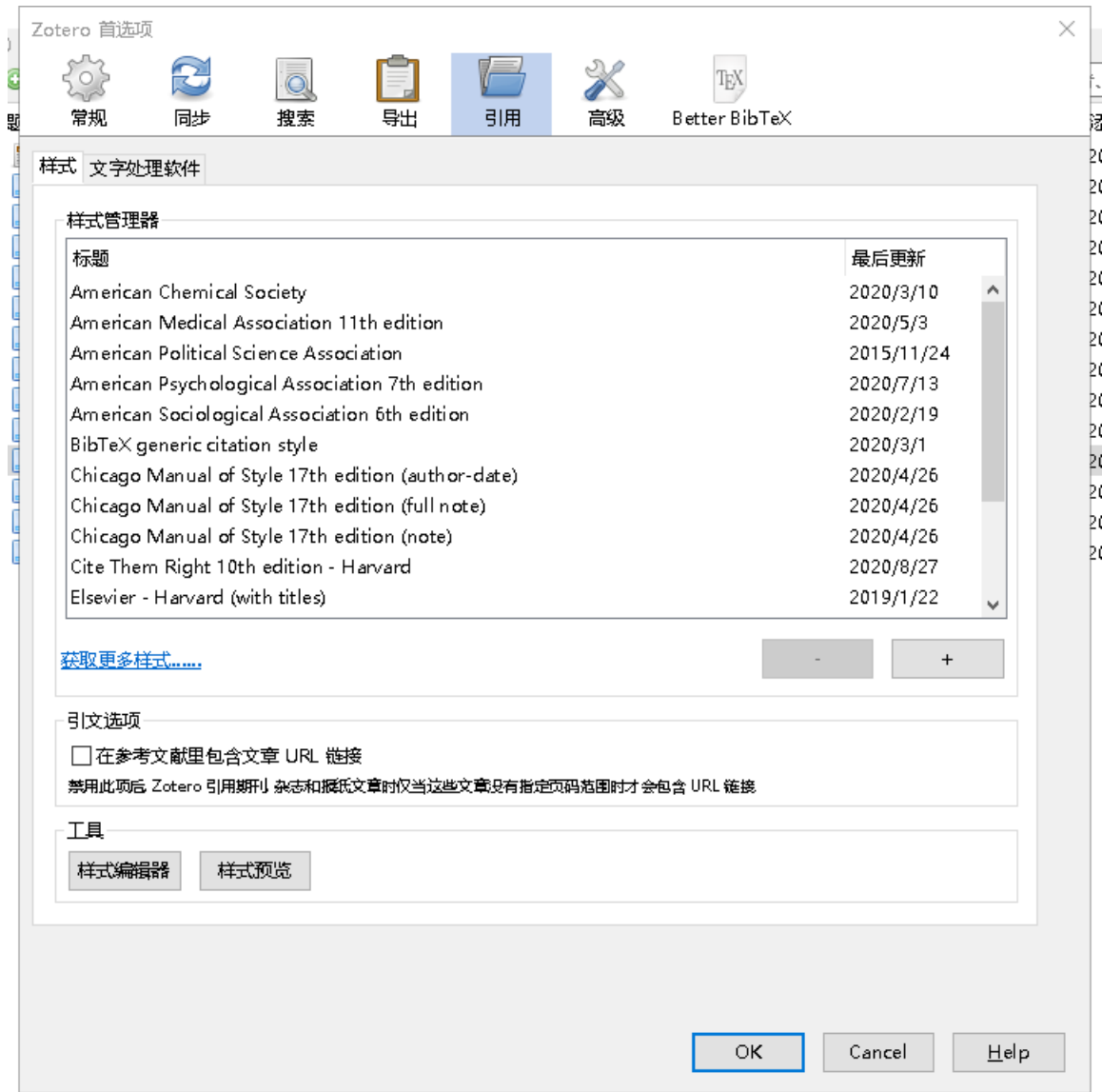


图 7-10 引用

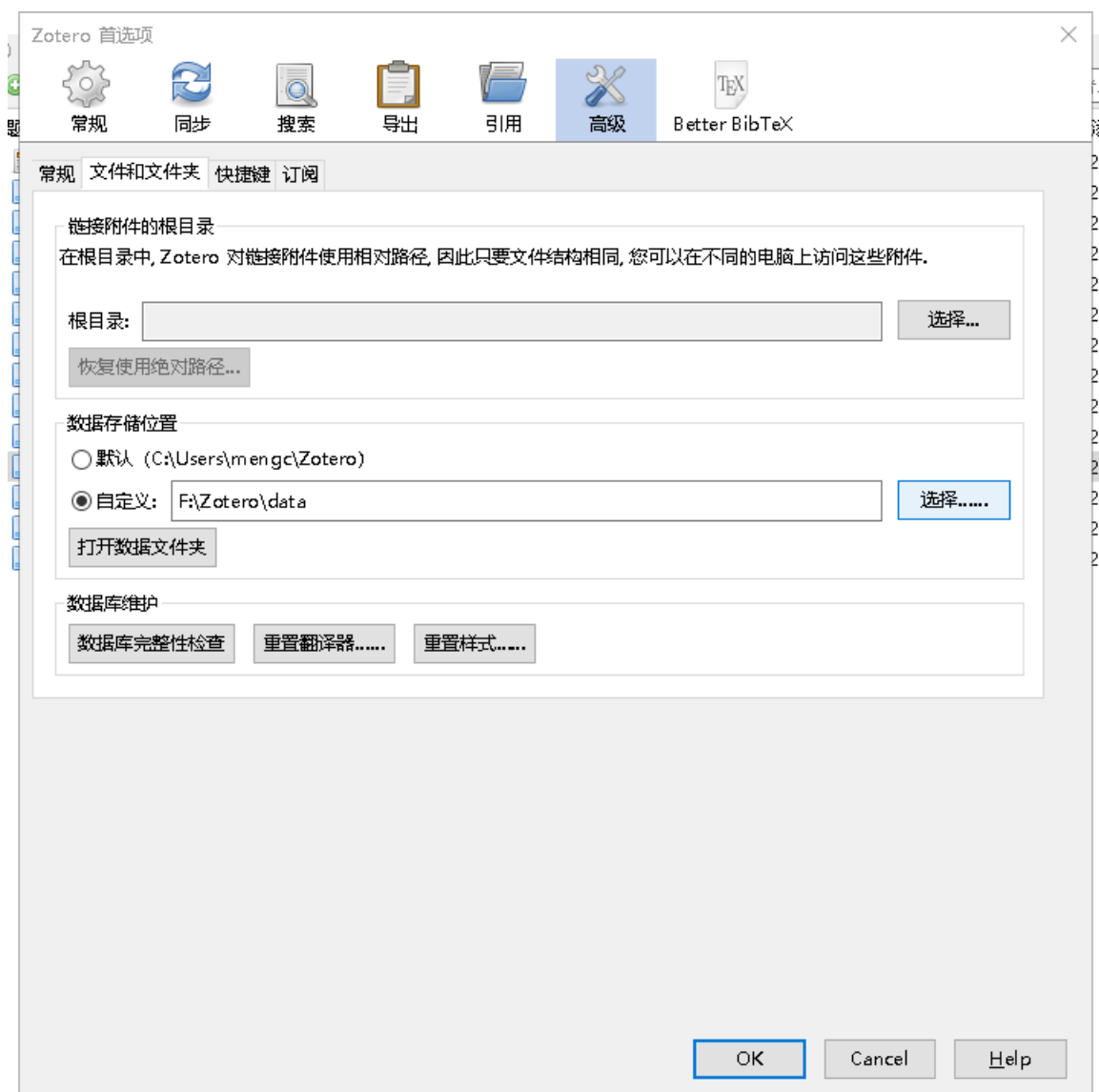


图 7-11 高级 1

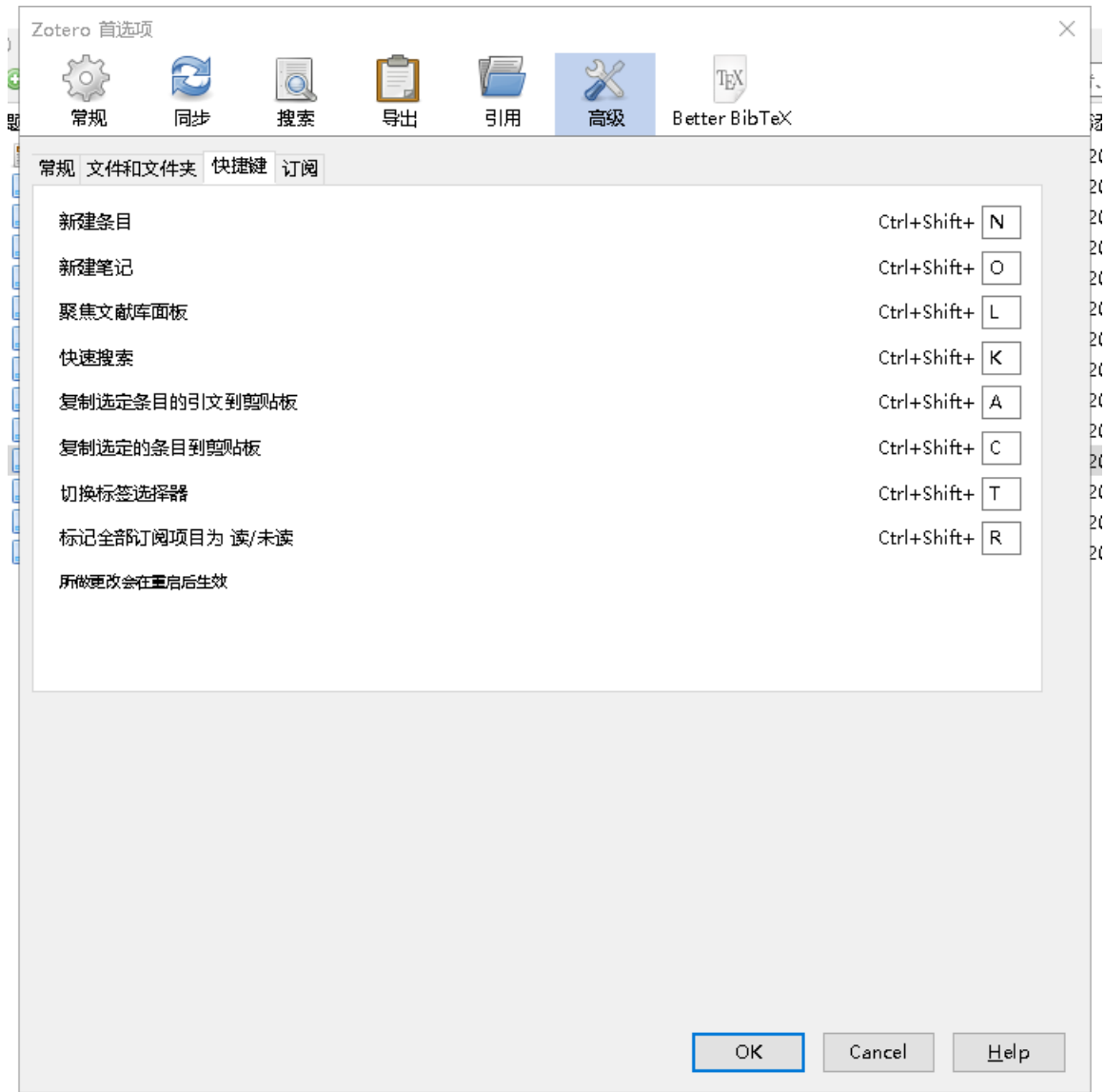


图 7-12 高级 2

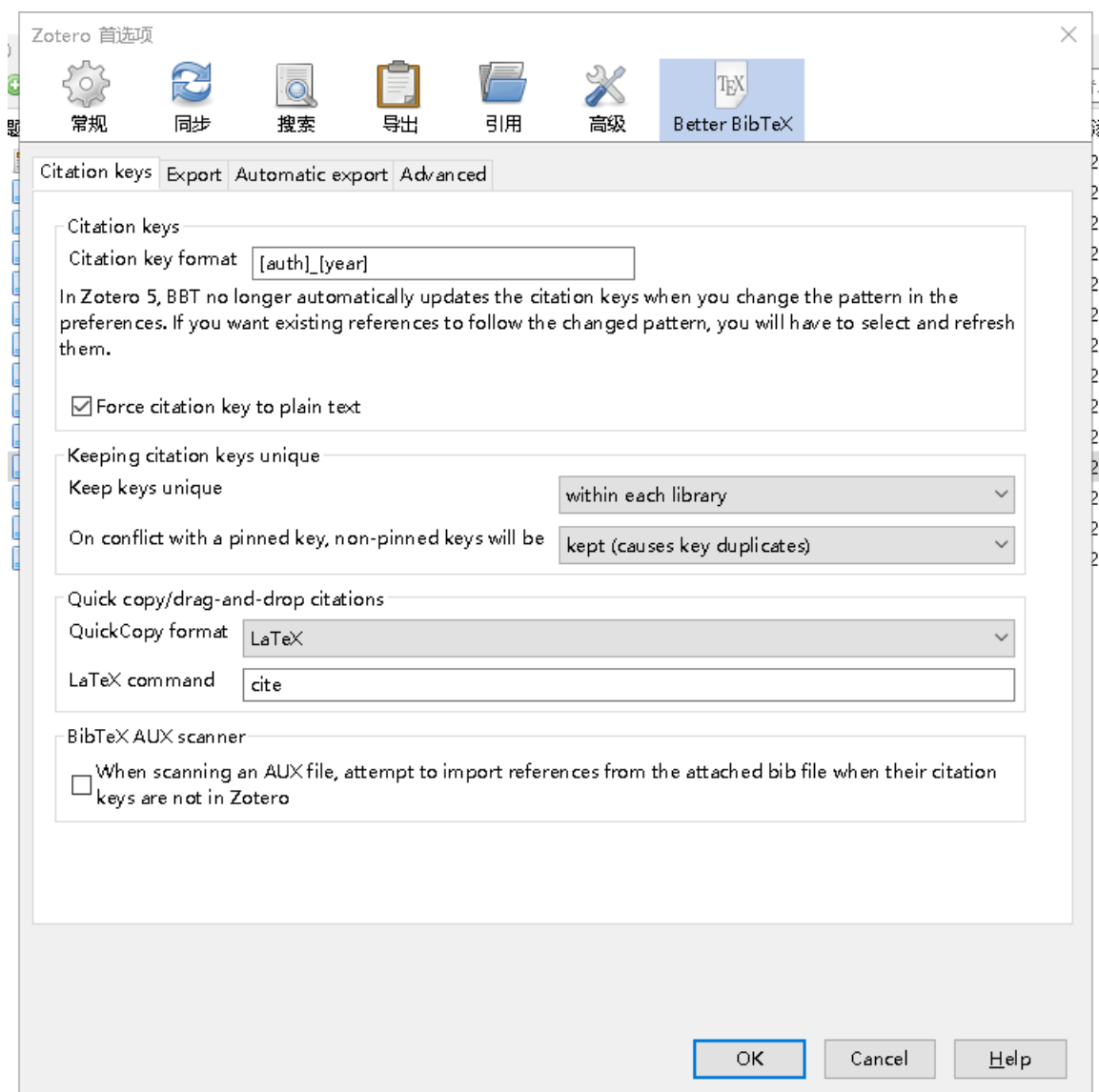


图 7-13 Better BibTeX1

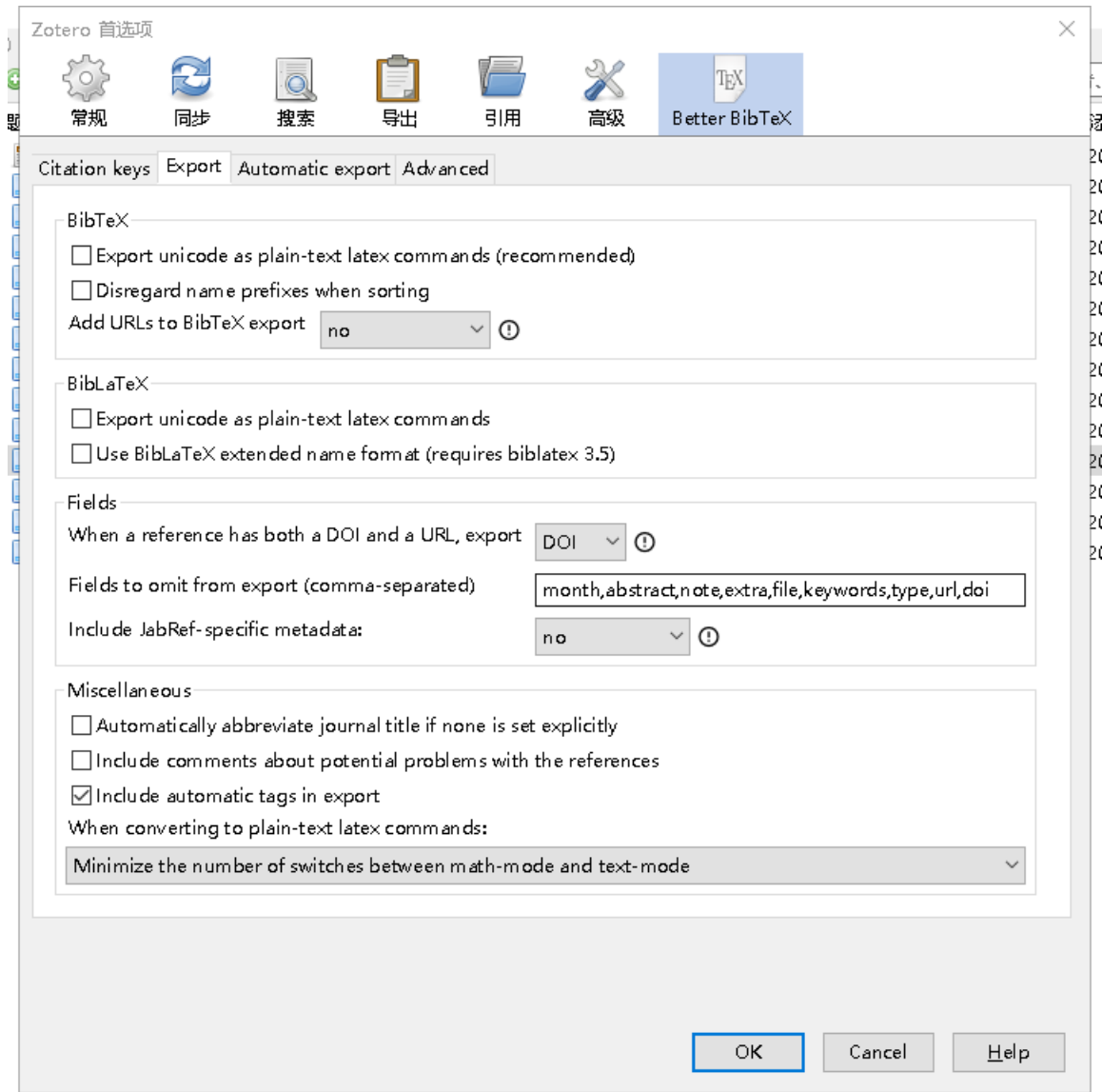


图 7-14 Better BibTeX2

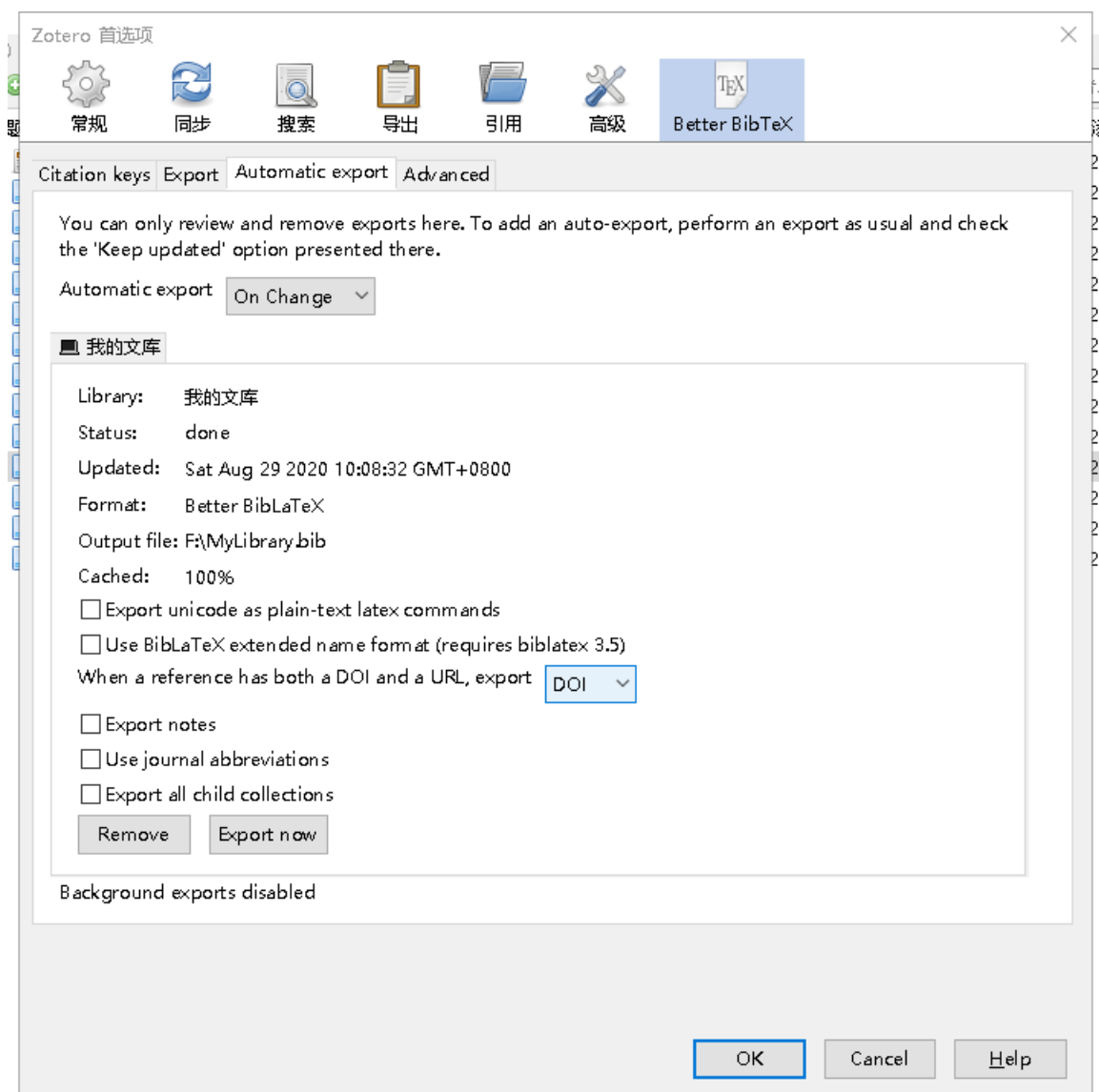


图 7-15 Better BibTeX3

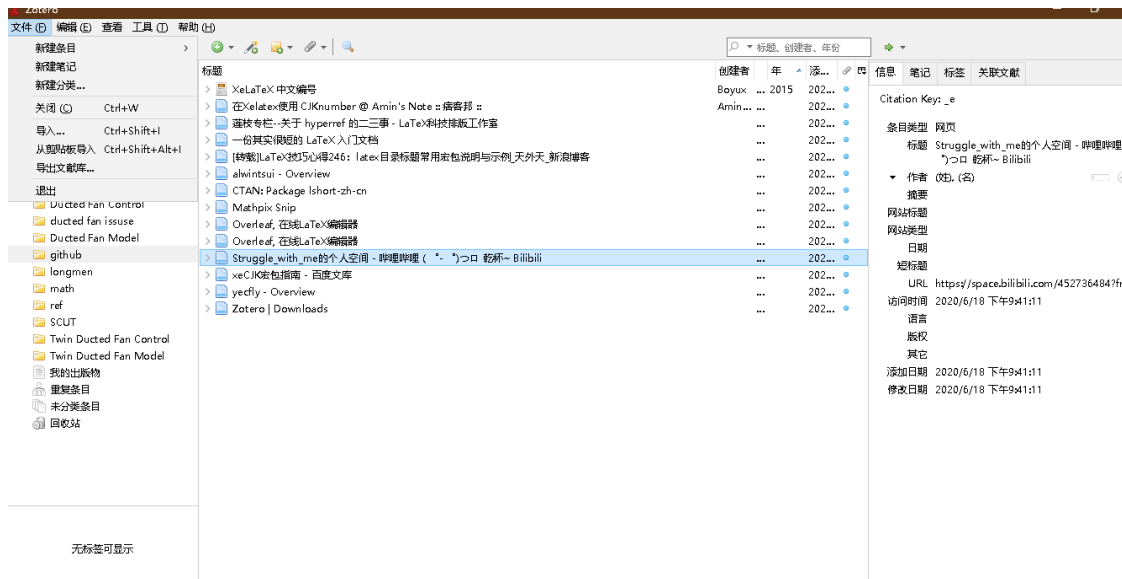


图 7-16 导出文献库

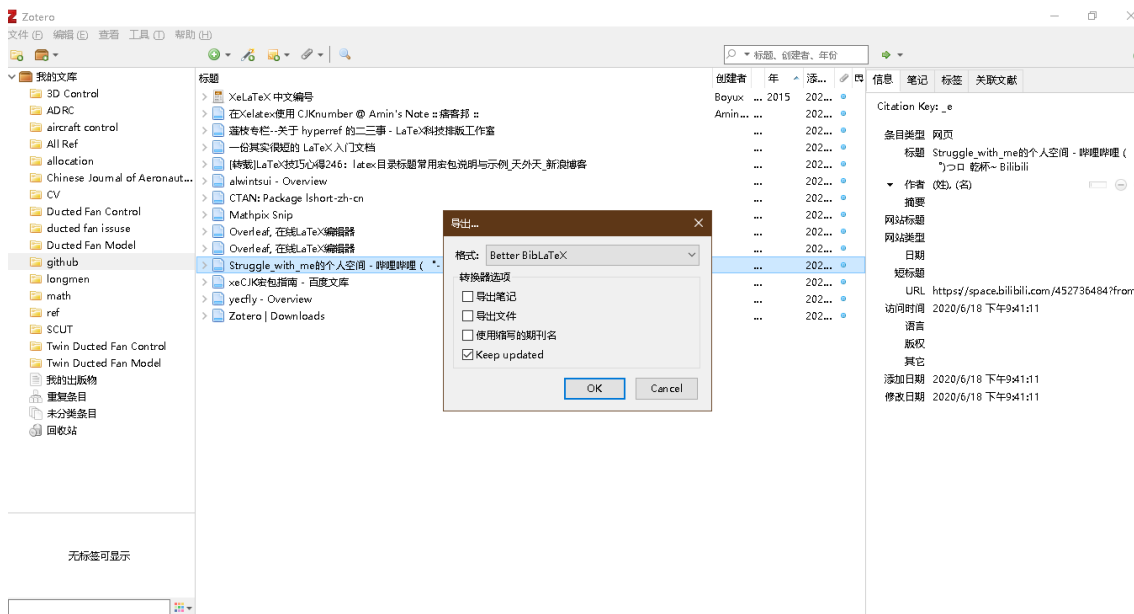


图 7-17 导出格式

图 7-18 导出文件名

第八章 列举环境

以下资料来自宏包说明和网络，翻译不一定正确：

在 LaTeX 中有三种基本的列举 (列表) 环境，即 `enumerate`(编号)、`itemize`(分条目) 和 `description`(描述) 环境。调整 latex 的列表环境时，使用 `enumitem` 宏包可以方便的调整间距（注意区分包名和环境名）和自定义编号样式。

8.1 调整间距

三种基本环境无论哪一种，间距的调整都是一样的。调整间距的参数命令包括两类：垂直间距和水平间距。各种距离的定义如图8-1所示。下图的来源一直找不到，可能是旧版本的宏包说明，新版已经删掉了下面的注释了。

现先总结出所推荐的间距设置，无编号的：

```
\begin{itemize}[topsep = 0 pt, itemsep= 0 pt, parsep=0pt, partopsep=0pt, leftmargin=36
pt, itemindent=0pt, labelsep=6pt, listparindent=24pt]
\item 第一项。内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容
内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容

\item 第二项。内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容
内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容

\end{itemize}
```

效果：

- 第一项。内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容
内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容
- 第二项。内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容
内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容

有编号的：

```
\begin{enumerate}[topsep = 0 pt, itemsep= 0 pt, parsep=0pt, partopsep=0pt, leftmargin
=44pt, itemindent=0pt, labelsep=6pt, label=(\arabic*)]
\item 第一项。内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容
内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容

\item 第二项。内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容
内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容

\end{enumerate}
```

效果：

- (1) 第一项。内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容

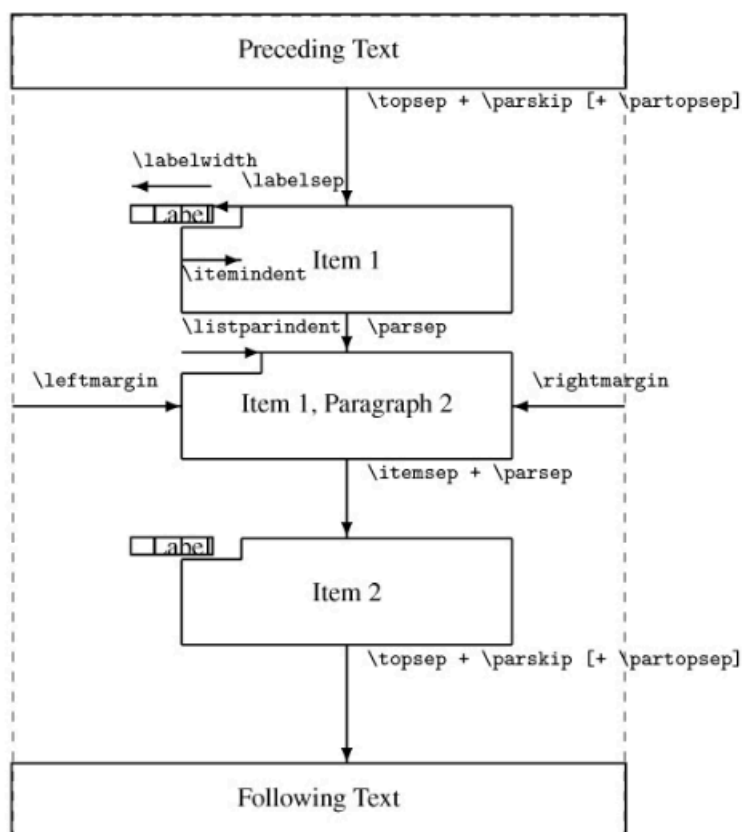


图 8-1 enumitem 包对各种间距的定义

内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容

(2) 第二项。内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容

内容

下面两节分别讨论参数设置规则。

8.1.1 垂直间距

摘抄宏包说明：

- `topsep` 控制列表环境与上文之间的距离。第一项和前一段之间的空间。
- `itemsep` 条目之间的距离
- `parsep` 条目里面段落之间的距离
- `partopsep` 条目与下面段落的距离。当环境开始一个新段落时，额外的空间被添加到 `\topsep`。

论文中希望上述距离都为 `0pt`，如：

```
\begin{itemize}[topsep = 0 pt, itemsep= 0 pt, parsep=0pt, partopsep=0pt]
  \item 第一项。
  \item 第二项
  \item 第三项。
\end{itemize}
```

效果为：

- 第一项。
- 第二项
- 第三项。

8.1.2 水平间距

水平间距调整比较复杂，对照宏包说明给出的图，下面内容参考了宏包原文和网络资料：

- 为页面的左边距）和该列表的左边距之间的空间。必须是非负数。它的值取决于表，则为页面的左边距）和该列表的左边距之间的空间。必须是非负数。它的值取决于列表级别。
- `rightmargin` 列表环境右边的空白长度。类似于 `\leftmargin` 但用于右边距。它的值通常是 `0pt`。
- `labelsep` 标号与列表第一项文本左侧的距离。标签框的末尾和第一项的文本之间的空间。它的默认值为 `0.5 em`。
- `itemindent` 条目的缩进距离。添加到项目第一行文本部分的水平缩进的额外缩进。通过减去 `labelsep` 和 `labelwidth` 的值，相对于该参考点计算标签的起始位置。它的值通常是 `0pt`。注：理解这个变量时，查看图8-1的顺序应该按照箭头从左到右，先 `leftmargin` 再 `itemindent`，然后再 `labelsep`，最后 `labelwidth`。即箭头的起始点是

基准点。若 `itemindent=0pt`，则 `leftmargin-labelsep`-编号长度的结果就是编号起始位置。

- `labelwidth` 包含标签的框的标称宽度。如果标签的自然宽度为 $< \text{labelwidth}$ ，则默认情况下，标签在宽度为 `(labelwidth)` 的框内右对齐排版。否则，使用自然宽度的框，这会导致该行上的文本缩进。可以通过为 `\makelabel` 命令提供定义来修改标签的排版方式。
- `listparindent` 条目下面段落的缩进距离。除了以 `litem` 开头的段落之外，列表的每个段落的开头都有额外的缩进。可以为负数，但通常为 `0pt`。

无编号的水平间距，给出两张方案

第一种：

- 第一项。内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容
- 第二项。内容

第二种：

- 第一项。内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容
内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容
- 第二项。内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容内
容内容内容内容内容内容内容内容内容内容内容内容内容内容内容内容

推荐第一种。

有编号的水平间距，下面给出三种方案：注：labelsep 是某一项文字和编号框的距离，一般就设为一个空格 6pt，要使编号左侧缩进两格，itemindent-labelsep 要等于编号长度。注意编号是右对齐，向左扩展的。

第一种方案是整体右移两格，文字距离编号一个空格，然后第二行文字和第一行对齐：

- [illegible]

第二种方案是和论文撰写规范的格式一样，注意不是论文撰写规范规定的格式，规

(1) 第一项。内容内容内容内容内容内容内容内容内容内容内容内容内容
内容内容内容内容内容内容内容内容内容内容内容内容内容

(2) 第二项。内容

内容

第三种方案是整体右移两格，文字距离编号一个空格，第二行文字不再右移：

[illegible][illegible]

推荐第一种。

除上述小括号数字的编号方法外，还有斜体字母等。在使用 `enumerate` 的时候，`label` 的问题就是使用计数的字符，是阿拉伯数字、罗马、中文、还是希腊字符的问题。

1) 第一项。

2) 第二项

3) 第三项。

a , 第一项。

b. 第二项

c. 第三项。

(I) 第一项。

(II) 第二项

(III) 第三项。

结 论

本文主要是展示如何使用修改“祖传模板”得到的新模板，在使用时直接替换成自己的论文内容即可。总结下来最最最麻烦的是科学上网，只有科学上网才能获取文献信息生成 bib 文件，后面就好办了。

本模板难免有不足之处，主要是我本人的论文涉及的格式有限，有些地方没探索到自然就没去设置。比如附录，附录的图文并茂等等，我本人是没有研究的，这里仅仅做了一些初步的工作，不过对很多同学来说本模板是够用的。希望有能帮助到华工的小伙伴们，有不足之处请多多理解，可以通过邮件联系我，上班之余我会尽量回复。

本模板会一直更新——2022-5-28

参考文献

- [1] Zaremba W, Sutskever I, Vinyals O. Recurrent neural network regularization[J]. ArXiv preprint arXiv:1409.2329, 2014.
- [2] Shi X, Chen Z, Wang H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting[J]. Advances in neural information processing systems, 2015, 28.
- [3] Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. ArXiv preprint arXiv:1406.1078, 2014.
- [4] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[J]. ArXiv preprint arXiv:1609.02907, 2016.
- [5] Fragkiadaki K, Levine S, Felsen P, et al. Recurrent network models for human dynamics[C] //Proceedings of the IEEE International Conference on Computer Vision. 2015:4346-4354.
- [6] Jain A, Zamir A R, Savarese S, et al. Structural-rnn: Deep learning on spatio-temporal graphs[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 5308-5317.
- [7] Ghosh P, Song J, Aksan E, et al. Learning human motion models for long-term predictions[C]//2017 International Conference on 3D Vision (3DV). 2017: 458-466.
- [8] Martinez J, Black M J, Romero J. On human motion prediction using recurrent neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2891-2900.
- [9] Gui L Y, Wang Y X, Liang X, et al. Adversarial geometry-aware human motion prediction[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 786-803.
- [10] Tang Y, Ma L, Liu W, et al. Long-term human motion prediction by modeling motion context and enhancing motion dynamic[J]. ArXiv preprint arXiv:1805.02513, 2018.

- [11] Gui L Y, Wang Y X, Ramanan D, et al. Few-shot human motion prediction via meta-learning[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 432-450.
- [12] Guo X, Choi J. Human motion prediction via learning local structure representations and temporal dependencies[C]//Proceedings of the AAAI Conference on Artificial Intelligence: vol. 33: 01. 2019: 2580-2587.
- [13] Liu Z, Wu S, Jin S, et al. Towards natural and accurate future motion prediction of humans and animals[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 10004-10012.
- [14] Chiu H k, Adeli E, Wang B, et al. Action-agnostic human pose forecasting[C]//2019 IEEE Winter Conference on Applications of Computer Vision (WACV). 2019: 1423-1432.
- [15] Gopalakrishnan A, Mali A, Kifer D, et al. A neural temporal model for human motion prediction[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 12116-12125.
- [16] Sang H F, Chen Z Z, He D K. Human Motion prediction based on attention mechanism[J]. Multimedia Tools and Applications, 2020, 79(9): 5529-5544.
- [17] Corona E, Pumarola A, Alenya G, et al. Context-aware human motion prediction[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 6992-7001.
- [18] Pavllo D, Feichtenhofer C, Auli M, et al. Modeling human motion with quaternion-based neural networks[J]. International Journal of Computer Vision, 2020, 128(4): 855-872.
- [19] Aksan E, Kaufmann M, Hilliges O. Structured prediction helps 3d human motion modelling[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 7144-7153.
- [20] Mao W, Liu M, Salzmann M, et al. Learning trajectory dependencies for human motion prediction[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 9489-9497.
- [21] Mao W, Liu M, Salzmann M. History repeats itself: Human motion prediction via motion attention[C]//European Conference on Computer Vision. 2020: 474-489.

-
- [22] Cui Q, Sun H, Yang F. Learning dynamic relationships for 3d human motion prediction[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 6519-6527.
 - [23] Li M, Chen S, Zhao Y, et al. Dynamic multiscale graph neural networks for 3d skeleton based human motion prediction[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 214-223.
 - [24] Li M, Chen S, Chen X, et al. Symbiotic graph neural networks for 3d skeleton-based human action recognition and motion prediction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.
 - [25] Li B, Tian J, Zhang Z, et al. Multitask Non-Autoregressive Model for Human Motion Prediction[J]. IEEE Transactions on Image Processing, 2020, 30: 2562-2574.
 - [26] Liu J, Yin J. Multi-grained Trajectory Graph Convolutional Networks for Habit-unrelated Human Motion Prediction[J]. ArXiv preprint arXiv:2012.12558, 2020.
 - [27] Lebailly T, Kiciroglu S, Salzmann M, et al. Motion Prediction Using Temporal Inception Module[C]//Proceedings of the Asian Conference on Computer Vision. 2020.
 - [28] Dang L, Nie Y, Long C, et al. MSR-GCN: Multi-Scale Residual Graph Convolution Networks for Human Motion Prediction[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 11467-11476.
 - [29] Cui Q, Sun H. Towards Accurate 3D Human Motion Prediction From Incomplete Observations[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 4801-4810.
 - [30] Shi L, Wang L, Long C, et al. Social Interpretable Tree for Pedestrian Trajectory Prediction[C]//AAAI Conference on Artificial Intelligence. 2022.
 - [31] Shi L, Wang L, Long C, et al. SGCN: Sparse Graph Convolution for Pedestrian Trajectory Prediction[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2021.
 - [32] Duan J, Wang L, Long C, et al. Complementary Attention Gated Network for Pedestrian Trajectory Prediction[C]//AAAI Conference on Artificial Intelligence. 2022.

- [33] Butepage J, Black M J, Kragic D, et al. Deep representation learning for human motion prediction and classification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 6158-6166.
- [34] Li C, Zhang Z, Lee W S, et al. Convolutional sequence to sequence model for human dynamics[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 5226-5234.
- [35] Liu X, Yin J, Liu J, et al. TrajectoryCNN: a new spatio-temporal feature learning network for human motion prediction[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020.
- [36] Barsoum E, Kender J, Liu Z. Hp-gan: Probabilistic 3d human motion prediction via gan[C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2018: 1418-1427.
- [37] Kundu J N, Gor M, Babu R V. Bihmp-gan: Bidirectional 3d human motion prediction gan[C]//Proceedings of the AAAI conference on artificial intelligence: vol. 33: 01. 2019: 8553-8560.
- [38] Hernandez A, Gall J, Moreno-Noguer F. Human motion prediction via spatio-temporal inpainting[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 7134-7143.
- [39] Jain D K, Zareapoor M, Jain R, et al. GAN-Poser: an improvised bidirectional GAN model for human motion prediction[J]. Neural Computing and Applications, 2020, 32(18): 14579-14591.
- [40] Liu Z, Lyu K, Wu S, et al. Aggregated multi-gans for controlled 3d human motion prediction[C]//Proceedings of the AAAI conference on artificial intelligence: vol. 35: 3. 2021: 2225-2232.
- [41] Cui Q, Sun H, Kong Y, et al. Efficient human motion prediction using temporal convolutional generative adversarial network[J]. Information Sciences, 2021, 545: 427-447.
- [42] Chao X, Bin Y, Chu W, et al. Adversarial refinement network for human motion prediction[C]//Proceedings of the Asian Conference on Computer Vision. 2020.

-
- [43] Lyu K, Liu Z, Wu S, et al. Learning human motion prediction via stochastic differential equations[C]//Proceedings of the 29th ACM International Conference on Multimedia. 2021: 4976-4984.
- [44] Liu Z, Wu S, Jin S, et al. Investigating pose representations and motion contexts modeling for 3D motion prediction[J]. IEEE transactions on pattern analysis and machine intelligence, 2022, 45(1): 681-697.
- [45] Oord A v d, Dieleman S, Zen H, et al. Wavenet: A generative model for raw audio[J]. ArXiv preprint arXiv:1609.03499, 2016.
- [46] Yu F, Koltun V, Funkhouser T. Dilated residual networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 472-480.
- [47] Pavllo D, Grangier D, Auli M. Quaternet: A quaternion-based recurrent model for human motion[J]. ArXiv preprint arXiv:1805.06485, 2018.
- [48] Li Y, Wang Z, Yang X, et al. Efficient convolutional hierarchical autoencoder for human motion prediction[J]. The Visual Computer, 2019, 35: 1143-1156.
- [49] Yan S, Xiong Y, Lin D. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]//Thirty-second AAAI conference on artificial intelligence. 2018.
- [50] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. 2015: 234-241.
- [51] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
- [52] Aksan E, Kaufmann M, Cao P, et al. A spatio-temporal transformer for 3d human motion prediction[C]//2021 International Conference on 3D Vision (3DV). 2021: 565-574.
- [53] Cai Y, Huang L, Wang Y, et al. Learning progressive joint propagation for human motion prediction[C]//European Conference on Computer Vision. 2020: 226-242.
- [54] Yu J, Lin Z, Yang J, et al. Generative image inpainting with contextual attention[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 5505-5514.

- [55] Zamir S W, Arora A, Khan S, et al. Multi-stage progressive image restoration[C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 14821-14831.
- [56] Sofianos T, Sampieri A, Franco L, et al. Space-Time-Separable Graph Convolutional Network for Pose Forecasting[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 11209-11218.
- [57] Alwintsui - Overview[EB/OL]. GitHub. <https://github.com/alwintsui>.
- [58] Yecfly - Overview[EB/OL]. GitHub. <https://github.com/yecfly>.
- [59] 一份其实很短的 LaTeX 入门文档[EB/OL]. 始终. <https://liam.page/2014/09/08/latex-introduction/index.html>.
- [60] 莲枝专栏—关于 Hyperref 的二三事 - LaTeX 科技排版工作室[EB/OL]. <https://www.latexstudio.net/archives/4800.html>.
- [61] CTAN: Package Lshort-Zh-Cn[EB/OL]. <https://ctan.org/pkg/lshort-zh-cn>.
- [62] Struggle_with_me 的个人空间 - 哔哩哔哩 (゜-゜)つロ 乾杯 ~ Bilibili[EB/OL]. <https://space.bilibili.com/452736484?from=search&seid=12208069428001748893>.
- [63] Renduchintala A, Jahan F, Khanna R, et al. A Comprehensive Micro Unmanned Aerial Vehicle (UAV/Drone) Forensic Framework[J]. Digital Investigation, 30: 52-72.
- [64] 蒙超恒, 裴海龙, 程子欢. 涵道风扇式无人机的优先级控制分配[J/OL]. 航空学报, 41(10): 327-338. <https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2020&filename=HKXB202010026&v=H33nFWoKPiMVe8lDuZG26q9EEFHrc40qK0CS6t2FdhaWKR8ppb9it6SMeovM2l8e>.

附录 1

在论文撰写规范中，下面两段话让人费解：

1. 对需要收录于学位论文中但又不适合书写于正文中的附加数据、方案、资料、详细公式推导、计算机程序、统计表、注释等有特色的内容,可做为附录排写,序号采用“附录 1”、“附录 2”等。
2. 公式序号按章编排,如第一章第一个公式序号为“(1-1)”,附录 2 中的第一个公式为“(2-1)”等。

论文撰写规范要求的附录和通常书籍上使用附录 A、附录 B 等编号的不一样，容易和正文混淆。特殊的要求和代码的耦合，使我不得不使用比较笨的方法来设计附录部分的模板。

1.1 测试一级标题 section

1.1.1 测试二级标题 subsection

1.1.1.1 测试三级标题 subsection

[illegible]

$$\begin{cases} \dot{v}_1(t) = v_2(t) \\ \dot{v}_2(t) = R^2 \left(-\zeta_1 [v_1(t) - v_c(t)]^\alpha - \zeta_2 \left[\frac{v_2(t)}{R} \right]^\beta \right) \end{cases} \quad (1-1)$$

$$\begin{cases} \dot{v}_1(t) = v_2(t) \\ \dot{v}_2(t) = R^2 \left(-\zeta_1 [v_1(t) - v_c(t)]^\alpha - \zeta_2 \left[\frac{v_2(t)}{R} \right]^\beta \right) \end{cases} \quad (1-2)$$



图 1-1 测试测试测试



图 1-2 测试测试测试

1.2 测试测试测试

1.2.1 测试测试测试

[illegible]

参数符号	数值	参数符号	数值	参数符号	数值
A_x, A_y, A_z	0.04082 m ²	ρ	1.225 kg/m ³	I_b	0.000029
k_{ϖ}	1.13342×10^{-6}	d_{ϖ}	1.13342×10^{-7}	k_{δ}	0.01495
$C_{D,x}, C_{D,y}$	0.43213	$C_{D,z}$	0.13421	q_a	1.49
l_a	-0.1121 m	d_{ds}	0.01495	d_{af}	0.01495
R	0.11 m	b	2	S	0.04082 m ²
$C_{l_{\alpha}}$	2.212 /rad	$C_{l_{\max}}$	1.05	$C_{l_{\min}}$	-1.05
l_2	0.06647 m	l_1	0.17078 m	m	1.53 kg
$C_{d,o}$	0.9	$C_{d,g}$	0.9	C_{duct}	0.78497
I_x	0.02548	I_y	0.02550	I_z	0.00562

参数符号	数值	参数符号	数值	参数符号	数值
I_x	054593	I_y	0.017045	I_z	0.049226
l_1	0.0808 m	l_2	0.175 m	l_3	0.06647 m
l_4	0.2415 m	l_5	0.1085 m	m	3.7 kg

附录 2

在论文撰写规范中，下面两段话让人费解：

1. 对需要收录于学位论文中但又不适合书写于正文中的附加数据、方案、资料、详细公式推导、计算机程序、统计表、注释等有特色的内容，可做为附录排写，序号采用“附录 1”、“附录 2”等。
2. 公式序号按章编排，如第一章第一个公式序号为“(1-1)”，附录 2 中的第一个公式为“(2-1)”等。

论文撰写规范要求的附录和通常书籍上使用附录 A、附录 B 等编号的不一样，上述要求最终的效果是这些编号容易和正文的混淆。特殊的要求和代码的耦合，使我不得不使用比较笨的方法来设计附录部分的模板。

2.1 测试测试测试

2.1.1 测试测试测试

测试测试测试测试测试测试测试测试测试测试测试测试测试测试测试测试

$$\begin{cases} \dot{v}_1(t) = v_2(t) \\ \dot{v}_2(t) = R^2 \left(-\zeta_1 [v_1(t) - v_c(t)]^\alpha - \zeta_2 \left[\frac{v_2(t)}{R} \right]^\beta \right) \end{cases} \quad (2-1)$$

$$\begin{cases} \dot{v}_1(t) = v_2(t) \\ \dot{v}_2(t) = R^2 \left(-\zeta_1 [v_1(t) - v_c(t)]^\alpha - \zeta_2 \left[\frac{v_2(t)}{R} \right]^\beta \right) \end{cases} \quad (2-2)$$



图 2-1 测试测试测试



图 2-2 测试测试测试

表 2-1 测试测试测试

参数符号	数值	参数符号	数值	参数符号	数值
A_x, A_y, A_z	0.04082 m^2	ρ	1.225 kg/m^3	I_b	0.000029
k_{ϖ}	1.13342×10^{-6}	d_{ϖ}	1.13342×10^{-7}	k_{δ}	0.01495
$C_{D,x}, C_{D,y}$	0.43213	$C_{D,z}$	0.13421	q_a	1.49
l_a	-0.1121 m	d_{ds}	0.01495	d_{af}	0.01495
R	0.11 m	b	2	S	0.04082 m^2
$C_{l_{\alpha}}$	$2.212 / \text{rad}$	$C_{l_{\max}}$	1.05	$C_{l_{\min}}$	-1.05
l_2	0.06647 m	l_1	0.17078 m	m	1.53 kg
$C_{d,o}$	0.9	$C_{d,g}$	0.9	C_{duct}	0.78497
I_x	0.02548	I_y	0.02550	I_z	0.00562

表 2-2 测试测试测试

参数符号	数值	参数符号	数值	参数符号	数值
I_x	0.054593	I_y	0.017045	I_z	0.049226
l_1	0.0808 m	l_2	0.175 m	l_3	0.06647 m
l_4	0.2415 m	l_5	0.1085 m	m	3.7 kg

攻读博士/硕士学位期间取得的研究成果

一、已发表（包括已接受待发表）的论文，以及已投稿、或已成文打算投稿、或拟成文投稿的论文情况(只填写与学位论文内容相关的部分):

序号	作者（全体作者，按顺序排列）	题目	发表或投稿刊物名称、级别	发表的卷期、年月、页码	与学位论文哪一部分（章、节）相关	被索引收录情况
1	蒙超恒、裴海龙、程子欢	涵道风扇式无人机的优先级控制分配	航空学报	已录用，2020年5月	2.1、2.2、3.4、4.1、4.2、5.1和5.3节	EI
2	蒙超恒、裴海龙、程子欢	Dynamic Control Allocation for A Twin Ducted Fan UAV	2020 International Conference on Guidance, Navigation and Control	已录用，2020年8月	2.3、4.3和5.2节	EI

注：在“发表的卷期、年月、页码”栏：

1. 如果论文已发表，请填写发表的卷期、年月、页码；
2. 如果论文已被接受，填写将要发表的卷期、年月；
3. 以上都不是，请据实填写“已投稿”，“拟投稿”。

不够请另加页。

二、与学位内容相关的其它成果（包括专利、著作、获奖项目等）

致 谢

这次你离开了没有像以前那样说再见, 再见也他妈的只是再见
我们之间从来没有想象的那么接近, 只是两棵树的距离
你是否还记得山阴路我八楼的房间, 房间里唱歌的日日夜夜
那么热的夏天你看着外面, 看着你在消逝的容颜
我多么想念你走在我身边的样子, 想起来我的爱就不能停止
南京的雨不停地下不停地下, 就像你沉默的委屈
一转眼, 我们的城市又到了夏天, 对面走来的人都眯着眼
人们不敢说话不敢停下脚步, 因为心动常常带来危险
我多么想念你走在我身边的样子, 想起来我的爱就不能停止
南京的雨不停地下不停地下, 有些人却注定要相遇
你是一片光荣的叶子, 落在我卑贱的心
像往常一样我为自己生气并且歌唱
那么乏力, 爱也吹不动的叶子

蒙超恒

2020 年 7 月 10 日

于华南理工大学