第1章

绪　论

1.1　引言

我国是一个统一的多民族国家，各民族的语言文字承载了丰富多彩的历史文化，多元一体是历史留给我们的一笔重要财富，也是我国的重要优势。我国各族人民共同缔造了中华人民共和国，都为中华民族的形成和发展做出了卓越贡献。56个民族都是中华民族大家庭的平等一员，共同构成了你中有我、我中有你、谁也离不开谁的中华民族命运共同体。实现中华民族伟大复兴是中国梦，是各民族的梦。

我国始终重视各少数民族的历史和文化传统，在各类政策、资金上都对少数民族文化研究有所倾斜，大力扶持了一批少数民族传统文化研究项目，其中少数民族古籍文献的保护研究是少数民族文化研究的重要组成部分[1]。一方面，古籍的文字内容承载了本民族在历史长河中的文化积淀，是宝贵的财富和遗产；另一方面，古籍本身的材质、制作工艺、传承渊源等，都是民族文化传承的载体。

藏族是我国多民族大家庭中的重要一员，其历史悠久文化璀璨，传承下来大量的古籍文献，生动地刻画了藏族先民在古代生活的各个方面，具有极高的研究和现实意义。笔者长期在甘肃工作生活，其省会“金城”兰州市是甘青川藏区文化交流的一个重要枢纽，2021年4月，第四届全国藏文古籍文献整理与研究高层论坛在兰州市召开，会议期间来自全国各地数十家科研机构、古籍整理单位、图书馆、高等院校的130多名专家围绕藏文古籍文献研究与弘扬中华民族优秀传统文化、建设青藏高原中华民族特色文化重要保护地、促进各民族交往交流交融等议题展开热烈研讨[2]。在会议的诸多议题中，藏文古籍的数字化保护以及相关资源的建设占据了一个重要的位置。

有一些藏文古籍由于年代久远，存在字迹模糊、纸张疏松、断裂和破损的情况，有些甚至不能翻阅，亟待抢救、整理、数字化存储和处理。这项研究工作属于图像处理、模式识别、文档图像分析与识别以及人工智能的范畴，其目标就是为藏文古

籍的数字化保护提供技术支持。采用传统方法，在对古籍样本图像深入分析和处理从而提高图像质量的基础上，提出将古籍页面图像分割为文本、图像、图形或表格等不同属性区域的版面分析方法；根据古籍手写和木刻的复杂文本特征，构建具有较高准确率的文本区域行、字切分算法模型；通过乌金体古籍文本字符图像的部件采样，创建基于部件图像的乌金体藏文、梵音藏文大字符集样本库构建方法，解决字符样本的多样性和对藏文古籍文字识别的适应性；研究多特征融合的大字符集藏文特征表示方法，训练具有较强鉴别能力和鲁棒性的分类模型。同时，采用深度学习方法构建和训练藏文古籍版面分析与识别模型，以克服传统方法的不足。设计和实现藏文古籍识别系统，将藏文古籍自动转换为相应的电子格式并进行有效的检索和再利用。

1.2 藏文古籍文档分析与识别研究的背景及意义

1.2.1 研究背景

中华文明源远流长，历经数千年绵延至今。海内外的数十万种古籍文献作为文明的传承载体，具有不可估量的历史文化价值。

以纸张作为载体的书籍本身就较为脆弱，水、火、战争甚至老鼠昆虫都会对其造成严重的侵蚀，历经沧桑能够侥幸存世的古籍大多纸张发黄变脆、色泽暗淡、字迹模糊，经不起反复翻阅和利用[3]。组织文物修复专家对破损古籍进行修复，并对古籍存放场所的温度、湿度以及光照条件进行严格控制，可以做到对古籍物理形态的抢救和保护[4-5]。但这种保护导致古籍只能束之高阁，仅保留了其作为历史文物的纪念价值而忽略了古籍内容所承载的文化价值，因此为了达到对古籍的保护、研究以及利用的多重目的，对其进行数字化保护势在必行。古籍整理靠人力完成，高度依赖专家经验，成本高、效率低，进展缓慢。光学字符识别（optical character recognition，OCR）是对文本图像中所包含的文字图像自动识别并输出为内码字符的计算机技术。因为古籍文献年代久远，所包含的文字量庞大，我们熟悉的汉文古籍异体字众多、字形字体多变、版式多样、页面模糊，这使得古籍文档图像分析与识别相较常规的 OCR 任务更具有挑战性。近年来，由于模式识别与深度学习融合得以快速发展，使得这一领域的应用显著提升了古籍 OCR 研究的效果，也极大地推动了古籍数字化的进程。

古籍数字化实质的目的就是保护和利用，这对古籍文献的再生性保护作用、文本深度挖掘、构建数据资源库以飨共享的知识服务平台等，使其在古籍保护及传播工作方面的能力无出其右。深度加工后的古籍内容，更以跨学科的“知识图谱”的形式辅助人们阅读与研究而产生二次价值，因此，被称为“高效率的知识内容”。古籍根据数字化加工及开发的程度，有存储、检索、交互、知识服务型数据库构建等形

式。就开发的层次，有学者概括为“表层数字化”和“深层数字化”，前者是图像或文本的简单存储，后者则是古籍内部知识元的标注，以及在知识元之间设计建立关联的原则等，是“内容和意义层面”的开发[6]。

就所谓“表层数字化”而言，基于不同的需求又可以分为不同的层次以及相应的技术手段。最基本的数字化是古籍文献保存的需求，这种保存除了上文提到的物理形态的保存，还包括其数字形态的保存，即对古籍文档进行拍照或者扫描，以数字图像的形式在网络空间上进行保存。进一步的数字化保护是对古籍文献内容获取的需求，更多的是希望它的内容能够得到更加广泛的传播并为人们所熟知。以图像形式保存的古籍不便于后期的检索和再学习，因此需要将文档图像转换为文字内码形式来保存。目前绝大部分可以进行全文检索的数字化藏文古籍都是靠人工录入的，这个过程的时间成本以及人力成本都过高。以自动化的方式，对藏文古籍图像进行分析识别，将其转换为内码保存，是目前古籍保护的发展趋势。

就“深层数字化”而言，其核心需求是对藏文古籍进行开发利用。所有的事物都只有在实践当中才会焕发新的生命，古籍也同样如此。例如，在对古籍文献进行电子出版时，不仅需要文字本身，还有文字在原始文档中的位置信息、文字书写时候的笔画和字形、文字附近的非文本内容等。这些信息都可以从自动化的分析和识别过程中得到。

藏文古籍文档图像分析与识别研究自然提到议事日程，与汉文古籍相比，藏文古籍有其自身的特点、难点和挑战性。数字化保护是本书研究的核心意义所在，研究贯穿了前文所述的不同层次的藏文古籍的保护需求，此外该项研究还在藏文的脱机输入接口、藏文古籍的知识库建设、藏文古籍的二次开发利用上有着广泛的应用前景。同时，研究的过程还是图像处理、模式识别等学科应用领域的拓展，研究中所产生的一些创新思路具有在其他相关领域进行泛化的潜力。

1.2.2 研究意义

1. 藏文古籍的数字化保护

藏族有着悠久的历史和灿烂的文化，在漫长的文明进程中，不仅创造了文字，还书写刻印了浩如烟海、博大精深的藏文文献，内容涉及佛教、哲学、历史、文学、艺术、藏医藏药与天文历算等多种学科。藏文典籍丰富浩繁仅次于汉文[7]，是中华民族的文化瑰宝，具有重要的社会和文化价值，也是珍贵的人类文化遗产。进行抢救式的数字化保护势在必行。最直接的方法是保护和抢救作为物质形态存在的古籍载体材料，但这并非长久之计。对文献进行拍照或扫描，以数字图像形式进行长久保存是首要任务，然而这种保存方式不便于后期的检索、学习及再利用。为了更好地进行藏文古籍的保护、信息处理与交换，通过对藏文古籍文档图像的分析与识别，可以在尽可能保持古籍原貌的情况下，将古籍图文重新排版印刷。同时使藏文典籍得以永久保存、永续利用，最终达到对古籍的保护、研究和利用的多重目的。

2. 广泛的应用前景

计算机终端的文字输入方式有两种，一种是键盘输入，另一种是识别输入。藏文键盘输入基本满足实用要求，而藏文文字识别的研究还有待深入，其中既包含基础研究，还包含应用研究与技术开发。概括起来，藏文古籍文档分析与识别的应用范围至少有以下 4 个方面：①藏文文字识别是多文种智能计算机的重要接口之一，终端对各种文字的识别结合起来，形成多功能文本阅读器，自动生成电子数据库，加快藏文古籍数据库建设步伐，便于对藏文古籍进行深入研究，有利于知识的共享与交流。②古籍全文数字化主要有三种方式：古籍全文图像方式、古籍全文文本方式和古籍全文图文方式。而古籍全文图文则是将全文图像化和全文文本化相结合，以达到全文检索与提供原始文献的目的。③应用于印刷行业，将藏文古籍通过数码识别，原样排版印制或重新排版印制，满足藏学研究、文献保护与传承。④甘肃、青海、西藏等有大量壁画、唐卡绘画艺术品上有文字题记，人工手段抄录速度慢、工作量大，通过图文转换的识别技术，在大幅提高工作效率的同时，还可以保留壁画、唐卡等艺术品的原貌并赋予其新的延伸功能。

3. 跨境语言相关研究的国际学术话语权

藏语是跨境语言，属汉藏语系藏缅语族藏语支。藏语主要分布于中国的西藏、青海、四川、甘肃、云南等省区，以及印度、尼泊尔、不丹等国家，使用总人数六百多万。在我国的藏语有卫藏、康、安多三种方言，各种方言之间差别较大，但文字并无差异。藏文古籍图像分析与识别的人工智能研究，对构建中国周边跨境文档图像分析研究的国际学术话语权有重要意义，可以提升学术实力、学术地位和影响力，更有利于国际文化交流。

综上，藏文古籍文档分析与识别研究的意义在于：①由于年代久远，很多藏文古籍存在纸张疏松、页面污损、笔迹模糊或笔画断裂等情况，所获取的古籍图像质量较低。将图像处理、模式识别、机器学习、深度神经网络等相关理论与藏文古籍保护结合起来，解决该类低质文档图像的预处理、版面分析、文本行字切分、古籍藏文识别和版面复原等问题，达到藏文古籍数字化保护的目的。②构架古籍图像预处理、版面分析、行字切分的算法模型，创建训练各个模型的数据集；研究多特征融合、更具鉴别能力和鲁棒性的古籍藏文特征表示方法与分类识别算法，提高计算机对藏文的感知和理解能力，丰富和发展图像处理、模式识别等学科的应用研究范围。

1.3 乌金体藏文古籍文档分析与识别研究内容

从文档图像预处理、文档版面分析、行字切分与识别、版面复原等实证研究，到相关数据集的建设，本书均以木刻北京版《甘珠尔》的文档图像为具体的实例进行介绍。

藏文古籍形制别具一格，具有浓厚的民族特色和时代烙印。其装帧有多种形式，包括卷轴装、经折装、蝴蝶装、线装以及梵夹装等[8]，其中的梵夹装是藏文古籍文献的主流形式，也是本书主要的研究对象。这种装帧形式源于印度的贝叶经，文档的页面为不加装订的长条散页，每一册文档有多页，通常用木板夹起来，或用绳子捆住，或用丝绸包裹。因此藏文古籍的数目量词常为“帙”“包”“函”等。典型的藏文古籍如图 1-1 所示。

(a)　(b)

图 1-1　梵夹装的藏文古籍

(a) 丝绸包裹成册；(b) 夹板并捆绑成册

一些收集到的藏文古籍文档页面如图 1-2 所示，这些文档在形式上都各有不同，通常来说，其版面特征包含了界线、边栏、明目与版心等。界线是指在纸面左右两侧的竖线，其作用是为了每行文字的左右对齐；边栏通常是一个矩形框，将纸上的所有内容囊括其中，以便排版整齐美观，也有一些手写的文档没有边栏；明目则通常在正文侧面，主要是当前页在全书中的章节、页码等内容；版心是页面的正文，有些印制精美的古籍每页中还会配有插图。

藏文古籍的纸张一般是以藏区特产的“瑞香狼毒草”等植物原料制作，其本身的颜色不同，且页面也较为粗糙。在进行书写或刻版印刷时，所使用的颜料也各有不同。常规的文本除使用墨汁书写外，还会用到银粉或者金粉，彩色的文字部分则使用了不同的矿物原料。

藏文自吞弥桑布札创制后经过千余年的实践创新，字体发展到几十种。从大的方面讲，可归纳为乌金体与乌梅体两大类。乌金体即有冠体，因其有一个显著特点，每个字母最上一笔是横直的，字母排列时上端必须在一条直线上，形似平顶帽，故此得名。其整体书写效果整齐划一。吐蕃时期王室发布的文告、执照、碑文以及钟铭，特别是佛经写卷几乎都是用乌金体。依据藏传佛教后弘期藏文书籍“软字精校精刻”的刊印要求，佛经写卷大都采用乌金体，该类字体可以认为是藏文的官方字体风格[9]。乌梅体即无冠体，这种字体的上端没有横直的一笔，似去掉帽子，故此得名。乌金体清晰易认，乌梅体飘逸洒脱，类似草体，二者共同构建了藏文书法艺术。两种字体的形态差异很大。乌金体藏文古籍无论木刻版或手写本，均可归结为手写藏文的范畴。藏文字符不等高、不等宽；除常用的藏文外，还有大量低频

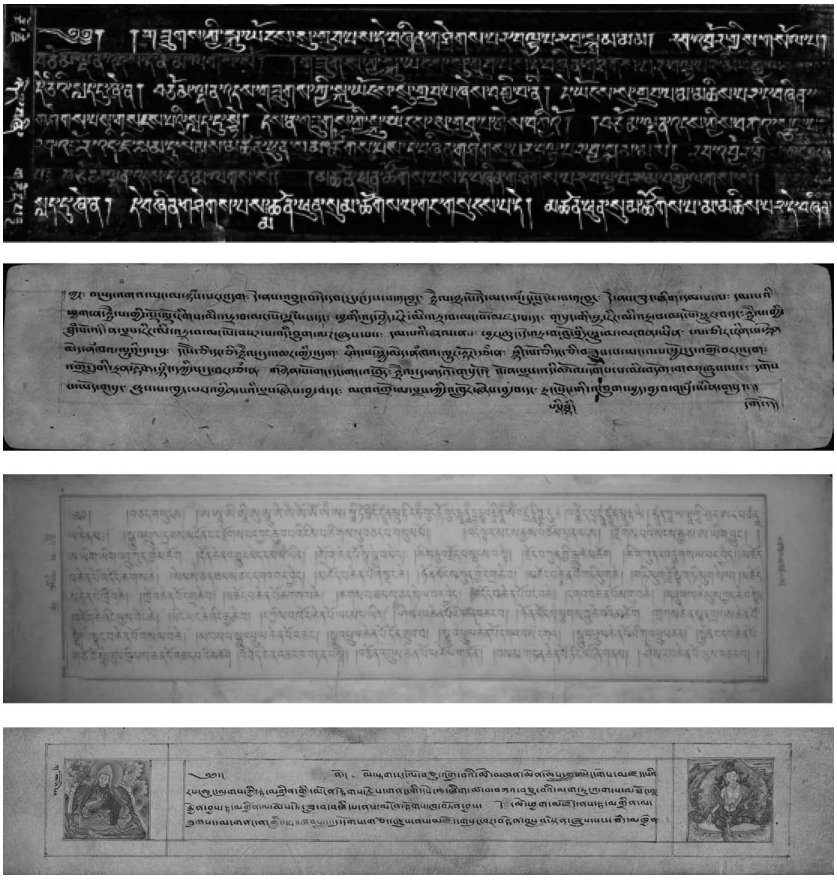


图 1-2　不同类型的藏文古籍页面

的梵音藏文；版面有污渍、字迹模糊，以及行之间的笔画粘连、行内字符之间的粘连等复杂现象。

在图 1-2 中，前三幅图像都是乌金体藏文，从中可以看出，尽管在这些文档中，文字的颜色大小各不相同，但是它们中的每一行文字都是沿着基线对齐展开的。图 1-2 的最后一幅图所展示的文字为乌梅体。本书的文档分析与识别研究仅限在乌金体范围。

藏文古籍有手写和木刻两种形式，其中藏文雕版印刷术自蒙元时期起，在西夏、元大都刊行至传入藏区之后，藏区各地陆续建起了印经院，并不断地进行技术更新，大规模地刊刻印刷，为藏文文献的保存和藏族文明的发展提供了更加有效的手段。相比手写的藏文古籍，木刻版的书籍发行量大、成本低廉，对文字普及和藏族文化的传播，都起着至关重要的作用。

1.4 乌金体藏文及其结构特征

1.4.1 藏文文字特点

文字的诞生并非必然，而是社会生产力以及文明发展到一定程度后的产物，藏族先民自古生活在青藏高原以及周边地区，形成了自身的独特文化，积累了大量需要记载和传承的内容。在内部需求的刺激以及外部文化的影响下，藏族先民主动地开始了文字创造工作。据1322年成书的《布顿教佛史》记载，由于吐蕃长期只有语言没有文字，各类知识要么口口相传，要么依赖其他周边民族的文字进行记载，给本民族文化的发展带来了很大困难，因此在松赞干布时期（公元7世纪），特派吞弥桑布札赴印度学习文字[10]。吞弥桑布札在印度学成后，以当时在印度使用最为广泛的天城体梵文为基础，结合藏语的发音特点和当时已有的不完善的文字符号体系，创制了藏文的基础字母和基本的构字组词法，自此藏族人民有了自己的文字，藏族文化的发展开启了一个新的纪元。

从语言学的角度讲，藏文本质上是拼音文字，这类文字在长期的使用过程中，必然会出现“字随音变”的情况，即由于不同地域方言的分化，相应语言所对应的拼音文字符号也会自然发生变化，长此以往这种变化所导致的文字差异有悖于文字创制的初衷。此外藏文创制完成后，一个最为重要的应用方向就是佛教经书的翻译。在翻译过程中，一方面由于原文的版本不同，既有梵文的原版经书直接翻译的，也有将汉译本经书进行二次翻译的；另一方面，经书中存在着大量的没有现实事物直接对应的抽象概念，这些概念的意义受翻译者自身对原始经文的理解以及宗教学派的不同而有所差异。这就造成相同的内容由不同译者翻译出来后大相径庭，为后续标准的制定和进一步的学习带来较大的困难，因此对藏文语法和词法的修订也是势在必行的。

目前学术界普遍推测藏文一共经历了3次重大的修订过程，在藏文语言学中将其称为“厘定”。我们现在所使用的藏文，基本上是公元11世纪第3次文字厘定后的成果。在这3次厘定过程中，主要的工作包括如下3个方面。第一，对大量藏传佛教中专有名词的翻译进行了统一；第二，对一些常见词的缩写语法进行了规范；第三，淘汰了一批文字草创初期产生的异体字，对很多字的拼写方法进行了简化。藏文经过3次厘定后，在各个方面都有了长足的改进，已经逐渐发展成为一门较为成熟的文字体系，其应用领域也有了极大的拓展。

吞弥桑布札在创制藏文基本字母时，最终得到了现在使用的藏文基本字母表，其中包含30个辅音字母和4个元音字母。早期我们在藏文的拉丁文转写键盘输入研究中，按照藏文字母与拉丁字母转写的对应关系，设计和开发了藏文智能输入系统[11-15]，并于2022年推出了新的输入系统，使其更加快捷、方便[16]。

除基本的字母符号以外，藏文中还有40个标点符号、24个图形符号以及20个藏文数字，这些数字中除了10个整数还有10个较为特殊的半数值，例如符号𼫏表示阿拉伯数字0.5，这也是藏文独有的特点。

表1-1所示为30个藏文字母及其拉丁转写、表1-2为反字字母及梵文元音的拉丁转写（为了输入法的键位安排，个别字母和传统的拉丁转写不完全一致）。

表1-1　藏文字母与对应的拉丁转写

ཀ	ཁ	ག	ང	ཅ	ཆ	ཇ	ཉ	ཏ
ka	kha	ga	nga	ca	cha	ja	nya	ta
ཐ	ད	ན	པ	ཕ	བ	མ	ཙ	ཚ
tha	da	na	pa	pha	ba	ma	tsa	tsha
ཛ	ཝ	ཞ	ཟ	འ	ཡ	ར	ལ	ཤ
dza	wa	zha	za	va	ya	ra	la	sha
ས	ཧ	ཨ	ི	ུ	ེ	ོ		
sa	ha	a	i	u	e	o		

表1-2　反字字母及梵文元音的拉丁转写

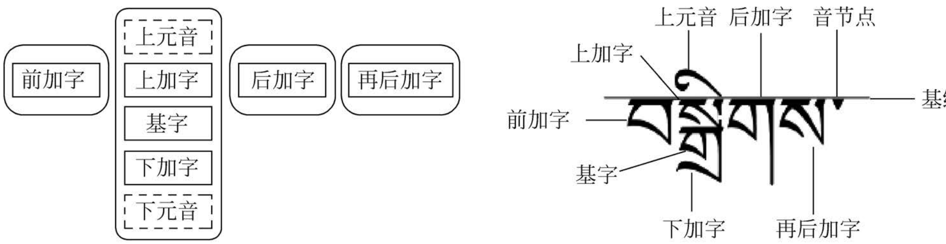
ྡ	ྣ	ྵ	ྷ	ྟ	ཻ	ཽ	ཾ
Da	Na	SHa	THa	Ta	E	I	M
ཿ	ྃ	ྈ	ྒ	ྲ	ྐ	ྂ	
O	H	A	G	R	X	B	

1.4.2　藏文音节

藏文句子由音节组成，音节具有“字”的意义，音节之间以音节点也叫隔音符隔开。音节有严格的“左右拼写、上下叠加”的规则，一个音节最多包含前加字、基字、上加字、下加字、上元音（或下元音）、后加字和再后加字七个字母，且最多出现一个元音（上元音或下元音）。一个音节的其他成分都可以省略，但至少有基字字母。一个音节中的各个纵向单位称为字符或字丁。基字字母所在字丁的叠加最多四层。

如图1-3所示是藏文音节结构及示例。藏文总是沿着一条基准线（下文简称基线）书写。基线是藏文的重要的位置信息。将基线上方所有字母统称基线上方笔画，对藏文来说基线以上只有三个元音字母，基线下方字母或所有字母的叠加组合统称基线下方笔画。

图1-3(b)的音节中有四个字丁：ད、ཐྲེ、ཀྱ、ས。藏文中的字丁有500多个。根据



(a) (b)

图 1-3 藏文音节结构及示例

(a) 藏文音节结构；(b) 藏文音节示例

音节拼写规则，音节可分为单字丁音节、双字丁音节、三字丁音节和四字丁音节，如ཁྲི、སྐྱིག、བཀད、མགྲིགས分别为单字丁、双字丁、三字丁和四字丁音节。理论上可拼写的藏文音节有 14000 多个[17]，实际使用的音节有 8000 多个[18]。藏文的文字识别可以以音节为单位，也可以以字丁或部件为单位。乌金体藏文古籍文档字丁和部件识别是本书的主要研究内容。

1.4.3 梵音藏文

藏文源于梵文，但又和梵文不完全等同，在字形、构字规则以及拼写规则上都存在较大差异。在使用藏文对梵文经书进行翻译时，很多专有词汇为了能够尽可能贴近原文，都是直接以梵文转写形式翻译而非意译，因此在藏文特别是在古籍中存在大量的梵音转写，也称为梵音藏文。这些文字和藏文文字的构成方式有较大不同。

梵音藏文是梵文的藏文转写形式，藏文和梵音藏文字符集包括：ISO/IEC 10646-1：Tibetan Character Collection[19]即藏文基本集、《信息技术 藏文编码字符集(扩充集 A)》[20]以及《信息技术 藏文编码字符集(扩充集 B)》[21]。

整理字符集时发现，扩充集 A 与扩充集 B 有一些重复的字丁，它们是：①“ཪྵ”扩充集 B 第 5 个字丁(F0004，第 1 行第 5 列)与扩充集 A 第 945 个字丁(F60B，第 60 行第 1 列)重复；②“ཀྵ”扩充集 B 第 404 个字丁(F0193，第 26 行第 4 列)与扩充集 A 第 822 个字丁(F653，第 52 行第 6 列)重复；③“ཐྱ”扩充集 B 第 4724 个字丁(F1273，第 296 行第 4 列)与扩充集 A 第 1304 个字丁(F871，第 82 行第 8 列)重复。扩充集 B 内部有两个重复字丁，如图 1-4 所示，F1144 和 F1145 完全一样。删除重复的字丁，最后确定字符集包括基本集的 42 个、扩充集 A 的 1536 个和扩充集 B 的 5662 个，共计 7240 个字丁。除基本集、扩充集 A 和扩充集 B 这 3 个标准内所收录的字符外，还有大量未被收录的字符也有其相应的 Unicode 编码。为了研究的规范化，本书所研究的内容限定为上述 7240 个字丁。

藏文和梵音藏文统称大字符集，在模式识别中就是 7240 个字丁。一般把藏文

称为现代藏文。在使用藏文字母进行构字时，其规则性较强，在语法上对基字和位于不同位置加字的搭配组合有严格的限定，但是在进行梵文的藏文转写时没有相关的规则，可以任意前后搭配组合。同时梵音藏文的书写在进行上下叠加时没有层数的限制。图 1-4 所示为梵音藏文。在整理的字符集中叠加的最多层数为 7 层。相比藏文和常用的几十个梵音藏文外，古籍中出现的其他梵音藏文，由于其叠加的灵活性和不确定性给藏文古籍文档图像的识别带来较大的挑战。以字丁为识别单位的藏文古籍识别中，字丁类别多、相似字丁多给识别带来了难度。

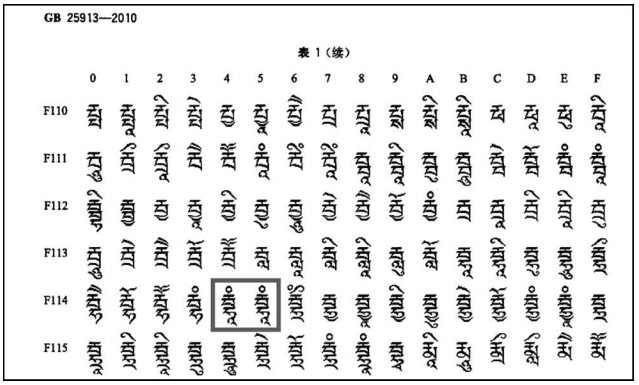


图 1-4　扩充集 B 部分字丁及两个重复梵音藏文

1.4.4　藏文-梵音藏文的部件

为了提高藏文-梵音藏文字丁样本质量和生成效率，降低采样成本，提出了藏文-梵音藏文基于部件组合的样本生成方法，确定遵循 3 个原则：①部件集越小越好。藏文-梵音藏文部件就是基本集中的单个字母、可以进行上下叠加的组合的编码单元。如下所示：ཀ་ཁ་ག་ང་ཅ་ཆ་ཇ་ཉ། ཏ་ཐ་ད་ན་པ་ཕ་བ་མ། ཙ་ཚ་ཛ་ཝ་ཞ་ཟ་འ་ཡ། ར་ལ་ཤ་ས། ཧ་ཨ། ཊ་ཋ་ཌ་ཎ་ཥ། ཱ་ཪ་ཀ། ི་ུ་ེ་ཻོཽཾྀྃ ྭྱྲྐྑྒྔྕྖྗྙ ྟྠྡྣྤྥྦྨྩྪྫ ྚྛྜྞྐྵྺྻྼྶྷྸ。②部件最小化。字丁中笔画不相连的基本集字符可以是部件，如ཨོཾ是基本集中的字丁，它可由 3 个基本集部件ཾ、ོ和ཨ上下叠加组合而成。③书写习惯最大化。书写字丁时，笔画相连的基本集字符本着最大化原则，把相连的两个基本集字符作为部件，并将其作为新部件加入部件集。如字丁ཀྱུ，其国际标准 Unicode 码为 3 个，即由基本集中的ཀ、ྱ和ུ构成，如果按照原则②，应该是 3 个部件。但是由于第 2 层和第 3 层印刷和书写时相连的特点，将其拆分为ཀ和ྱུ 2 个部件。与ུ相连的共有 27 个部件：ཀུཁུགུཐུཔུཕུམུཛུཚུཇུཇུཟུཏུནུཡུཧུཤུལུཙུསུངུཙྪུཉུཥུཊུཊུཏུ།，同样与ྲ相连的共有 29 个部件：ཀྲཁྲགྲཐྲཔྲཕྲམྲཛྲཚྲཇྲཊྲཟྲཏྲནྲཡྲཧྲཤྲལྲརྲསྲངྲཉྲཥྲཊྲཌྲཌྲཟྲཞྲདྲ།，与ྭ书写相连的有 30 个部件：ཀྭགྭཁྭབྭཔྭམྭཛྭཚྭཁྭཇྭ ཟྭཏྭནྭཡྭཧྭཤྭལྭརྭསྭངྭཙྭཉྭཥྭཊྭཌྭཌྭཟྭཞྭདྭའྭ།。3 个字符འ、ཞ、ཟ与ོ书写相连，形成 3 个部件：འོ、ཞོ、ཟོ，部件集由 81 个基本集字符和 89 个相连部件组成，如表 1-3 所示共 170 个部件。有了这 170 个部件，一方面，我们可以利用