基于生成式语义解耦的古籍视觉知识元重构与可信归档研究*

刘丽娜　蔡茜
（南京图书馆，南京 210018）

摘要：古籍数字化工作正向高质量、智能化方向转型，但现有以整页扫描件为基本管理单元的古籍数字化形态存在粒度错配、语义关联缺失、证据底本模糊等不足。生成式人工智能推动古籍图像处理从文字识别迈向语义理解与视觉重构，本研究据此提出视觉知识元概念并建立其形式化定义与分类体系。在视觉知识元理论基础上，进一步构建三层智能归档架构，使该架构与开放档案信息系统及中国行业标准进行对接，以实现证据不变、派生可控、过程可查的古籍数字化治理目标。本研究提出面向语义解耦的古籍视觉知识元重构工艺流程，并以南京图书馆馆藏为例开展场景应用与初步验证，为古籍资源从图像数字化走向语义数字化提供可落地的框架与路径。

关键词：视觉知识元；古籍数字化；生成式人工智能；语义解耦；OAIS；知识组织；数字人文
中图分类号：G250.7　DOI：10.3772/j.issn.1673–2286.2026.03.004
引文格式：刘丽娜，蔡茜. 基于生成式语义解耦的古籍视觉知识元重构与可信归档研究[J]. 数字图书馆论坛，2026，22（3）：31–39.

2022年中共中央办公厅、国务院办公厅印发《关于推进新时代古籍工作的意见》，明确提出推进古籍数字化，加强古籍数字化资源管理和开放共享[1]，对古籍知识服务能力提出了更高要求。然而，从知识组织视角看，当前主流数字古籍形态仍以整页扫描件为基本管理单元，粒度错配问题日益凸显。首先，古籍中印鉴的形制与叠压关系、题跋批注的作者身份与年代信息、版框界行等版式细节多依赖人工经验辨识判定，难以转化为可计算、可链接、可复用、可溯源的细粒度证据对象，制约了古籍知识服务的精准化与智能化发展；其次，整页模式容易将跨卷册检索封装在像素内部，使跨资源的逻辑链接缺少统一锚点，导致资源层面的语义关联链难以形成，从而不利于构建开放共享的古籍知识服务平台；最后，数字化服务输出的内容与证据底本之间缺少结构化隔离，两者存在混用现象，造成学术引用边界模糊与真实性风险。

1　研究综述

1.1　古籍数字化

中国古籍数字化起步于20世纪90年代，历经三个发展阶段，服务能力与覆盖范围不断提升。第一阶段以扫描保存为核心，上海图书馆早期开发《古籍影像光盘制作及检索系统》，国家图书馆整合多个古籍数据库构建起的中华古籍资源库，实现了珍贵古籍的数字化存储与基本检索，为公共文化服务奠定了资源基础[2]。第二阶段以检索服务为重心，中华书局下属古联（北京）

收稿日期：2026–02–04
*本研究得到2025年度江苏省图书馆学会一般课题“智慧图书馆的技术赋能与创新发展研究”（编号：25YB074）资助。

数字传媒科技有限公司负责建设和运营的籍合网[3]、中华经典古籍库[4]，字节跳动与北京大学合作研发的识典古籍[5]等平台相继上线，通过提供全文检索、在线阅览等服务，有效扩大了古籍资源的开放共享范围[6]。第三阶段以知识揭示为目标，王秋云等[7]从数字人文视角研究古籍资源深度揭示；陈涛等[8]构建了古籍文献通用知识组织模型，并以《永乐大典》“湖州府”相关卷册为例，构建关联数据和知识图谱，验证所构建古籍文献通用知识模型的科学性和可行性，推动古籍数字化从资源建设向知识服务转型。

然而，现有研究普遍存在粒度盲区。无论是影像数据库还是全文数据库，其基本组织单元仍是“页”或“书”，页内细粒度对象难以被独立标识、检索与关联，古籍知识服务的深度化挖掘仍待开发。虽然《全国智慧图书馆体系建设：古籍数字化和知识标引规范手册》[9]提出了知识标引的方向，但在技术实现层面仍缺乏系统化的对象模型与归档架构。

1.2 知识元理论及其视觉扩展

知识元理论试图对知识内容进行原子化切分与重组。索传军等[10]将知识元界定为文献中相对独立、最小、不可分割且表征一个知识点的元素，并强调其具有相对独立性、最小性、表征性。这一理论对古籍场景具有天然吸引力，因为古籍研究的知识对象往往不是整本书，甚至不是整段文字，而是被高频引用的细小证据点，如一枚藏书印、一处校勘批注、一个避讳缺笔等。高劲松等[11]进一步将知识元理论拓展至视觉模态，提出基于知识元的文物图像资源细粒度知识表示方法。值得注意的是，知识元理论在视觉模态中尚缺少落地标准，特别是古籍中大量高价值信息依附于版面结构与物理痕迹，在视觉形态上难以实现系统化建设。本研究以知识元为导引，结合潘云鹤[12-13]关于视觉知识理解的论述和Jiang等[14]关于多模态大模型视觉理解力的评估研究，本研究提出视觉知识元（Visual Knowledge Unit，VKU）这一概念，将它界定为古籍图像中具备空间边界可定位、语义属性可描述、生成过程可追溯3项特性的最小知识单元。

1.3 生成式人工智能在文化遗产领域的应用

生成式人工智能（Generative Artificial Intelligence，GenAI）为文化遗产的保护开发提供了新的技术路径。在文档分析领域，版面分析模型展现出强大的版面理解与结构化抽取能力，能够处理复杂的历史文献版面[15]。在古籍光学字符识别（Optical Character Recognition，OCR）领域，识典古籍项目利用深度学习技术实现了较高的古籍文字识别准确率，且吸纳2 600人的志愿者团队参与校对，形成了“技术驱动+社会协作”的服务模式[16]。在文化遗产修复领域，腾讯联合敦煌研究院开发了人工智能（Artificial Intelligence，AI）损伤识别技术，诊断敦煌千年壁画的损伤状况[17]。在技术层面，Qwen2.5-VL等视觉语言模型能输出结构化视觉数据并支持基于边界框的对象定位[18]；Qwen-Image-Layered等分层生成模型能够将单张图像分解为多个语义解耦的RGBA（Red-Green-Blue-Alpha）图层，在R（Red）、G（Green）、B（Blue）颜色模型基础上通过增加Alpha通道来控制颜色的透明度，为古籍中印章叠加、透背分离等问题提供新思路[19]。但现有研究较少关注如何为生成式输出建立可追溯的治理机制。

1.4 数字保存行业标准与AI治理

在国内，国家数字科技文献长期保存体系（National Digital Preservation Program，NDPP）借鉴开放档案信息系统（Open Archival Information System，OAIS），为公共文化机构的数字资源管理提供了参考范例[20]。同时，我国在古籍数字化与AI内容治理领域已建立较为完整的标准体系。《古籍元数据规范》[21]以都柏林核心元素集（Dublin Core Element Set，DC）为基础扩展古籍专有元素集；《汉文古籍版式描述规范》[22]为版式结构描述提供了标准化字段；《人工智能生成合成内容标识方法》[23]则为AI生成内容的标识提供了技术规范。然而，吴欣雨等[24]提出现有研究尚未充分回应OAIS国际框架与中国行业标准如何对接的关键问题，具体的架构设计与标准参照方案仍有待深入探索。

2 视觉知识元的概念界定与理论建构

2.1 概念界定：从平面像素到语义图层

古籍不同于现代一般类型的文献，其知识对象囊括内容型（文字、批注）、证据型（印章、题跋）和结构

型（版框、界行）等多种形态。这些对象具有两类共同属性：一是空间锚定性，它们必须能够回溯到图像中的具体位置，否则无法支撑版本学、目录学、校勘学的证据链；二是语义不完全显性，对象的意义常由位置关系、叠压关系与物理特征共同决定，单纯的文字转写难以完整表达。由于古籍页面呈现语义耦合特征，即多种对象叠合成一个像素，且无法通过简单裁切加以分离，因此提出语义解耦概念，将耦合的页面分解为语义自足的独立图层，再从各图层中切分出最小知识单元。语义解耦是从整页影像到VKU重构的必经步骤和前提。

基于上述分析，VKU应具备三重身份。第一，VKU必须具备明确边界与位置锚点，可以是字、词、符号、印章、插图、批注、缺损区域，也可以是版式结构组件，共同特点是能被边界化并被引用，从而支撑精准化的知识检索。第二，VKU作为证据优先的中介层，并不等同于最终的知识图谱节点，因此要先保证证据对象的可控封装，再开展知识元抽取与关系构建，以确保输出的可信度。第三，作为可治理的数字对象，VKU在GenAI介入后可能存在概率性偏差，因而VKU的定义必须包含溯源与置信表达。

进一步而言，VKU的语义完整性不仅指对象被标注了类型标签，而是要求每个VKU承载足以在知识网络中独立使用的语义信息。第一层为类型语义，明确对象归属的知识类别；第二层为内容语义，记录对象本身所携带的信息；第三层为关系语义，刻画对象与其他VKU之间的空间关系、叠压关系与归属关系。VKU的切分原则是保留语义自足性。由此，空间锚定性、语义完整性与溯源可追性共同构成VKU作为最小知识单元的三位一体约束。

2.2　语义透明度：Alpha通道的知识组织诠释

分层生成模型的核心输出是带有阿尔法（Alpha）通道的RGBA图层。在计算机图形学中，Alpha通道表示像素的不透明度，取值范围为0~1，数值为0时代表完全透明，数值为1时代表完全不透明。本研究提出将Alpha通道重新诠释为“语义透明度”，即知识对象独立性的数字化度量。这一概念强调古籍中的知识并非平铺在单一平面上，而是以叠压形态存在。分层生成模型通过Alpha通道，为每个语义层赋予独立的“存在权重”，从而实现知识的结构性分解。

语义透明度解决了视觉识别中的边界模糊困境。传统图像分割依赖“硬边界”，像素要么属于对象A，要么属于对象B。但古籍中大量对象存在自然过渡，表现为印章边缘的晕染、墨迹的洇散、霉斑的渐变。Alpha通道的连续取值恰好表达了这种“软边界”特性，使VKU的边界定义从二值判断转向概率表达。

由此可见，语义透明度解决的是VKU边界的连续性表达问题，而前文所述的语义完整性解决的是VKU内部信息的自足性问题，两者分别从“外部边界”和“内部内容”两个维度保障VKU作为最小知识单元的有效性。

2.3　形式化定义与分类体系

为支撑概念的可操作性，本研究将VKU定义为由9项分量构成的结构化对象，依次为：全局唯一标识、父页面标识、图层标识、对象类型、空间边界、视觉嵌入向量、语义属性集、溯源信息和置信度。全局唯一标识用于区分每个VKU的身份；父页面标识建立VKU与原始证据页面的映射；图层标识标记VKU所属的分层形态；对象类型标明所属知识类别；空间边界采用多边形坐标或Alpha遮罩表示对象在页面中的位置与范围；视觉嵌入向量用于相似性检索；语义属性集根据对象类型动态加载；溯源信息包含生成方式、模型版本、时间戳、随机种子等；置信度取值范围为0~1，衡量识别结果的可靠程度。

有效的VKU需满足3项约束：空间边界和溯源信息不得为空、语义属性集须属于该对象类型所对应的合法属性空间、语义属性集须涵盖该类型所要求的最小必备属性子集。例如，文字类VKU至少包含文本内容与识别置信度，印章类VKU至少包含印文内容与印文篆法。该约束确保每个VKU不仅具有合法的类型标签，而且承载了该类型所要求的最低限度语义信息，避免形式上有效但语义空洞的VKU进入知识服务体系。

为支撑古籍知识服务的系统化建设，本研究将VKU分为5类：文字类、图像类、符号类、结构类和缺损类，详情见表1。各类VKU的语义完整性体现为最小必备语义属性。在元数据映射方面参照《古籍元数据规范》《汉文古籍版式描述规范》《古籍特藏破损定级标准》[25]，从而实现VKU与现有行业标准体系的有效衔接。

表1　5类VKU详情

类别	内容	特征	语义属性	最小必备语义属性	参照的行业标准	典型应用场景
文字类	正文、注文、夹批、眉批、题跋中的文字片段等	具备明确的横轴X、纵轴Y组成的二维平面坐标边界，通常以多边形或矩形框形式定位在版框内的特定位置	文本内容、字体、墨色、书写年代、作者信息等	文本内容、识别置信度、所属段落上下文	《古籍元数据规范》	生僻字/异体字的识别与专家确证、去背透清洁版阅读
图像类	版画插图、装饰花纹、印鉴图样等	具有独立的像素图层边界，常表现为不规则形状，且可能与其他图层存在物理叠加	图像内容描述、色彩、构图、印文含义、印文篆法等	图像语义类别（印鉴/插图/装饰）、内容描述、关联文字VKU引用	《古籍元数据规范》	叠压印章的无损揭示、跨卷册印章关联检索
符号类	圈点、校勘符号、避讳缺笔标记等	点状或线状分布，通常依附于文字类VKU的侧边或内部空间	符号类型、标号意义、校勘意图等	符号类型、标注语义、所依附的文字VKU引用	《古籍元数据规范》	古籍自动化校勘、版本比对分析
结构类	版框、界行、书口、鱼尾、版心等	呈现为规则的线性或框架几何结构，定义了整个页面的空间骨架和布局	版式结构类型、尺寸参数、排列方式等	版式类型、空间参数、所包含的子VKU引用列表	《汉文古籍版式描述规范》	自动化版面分析、数字化古籍复原、版本学特征提取
缺损类	虫蛀孔洞、霉变污渍、水渍、撕裂等	呈现为随机分布的块状或点状负空间，具有渐变的Alpha透明度边界特征	缺损类型、几何尺寸、严重程度、修复优先级等	缺损类型、严重程度、影响范围内的文字VKU引用	《古籍特藏破损定级标准》	古籍劣化监测、修复优先级自动排序、非破坏性语义修复

3　系统架构设计

OAIS是国际公认的数字保存参考模型，其信息模型包括提交信息包（Submission Information Package，SIP）、存档信息包（Archival Information Package，AIP）和分发信息包（Dissemination Information Package，DIP），为数字资源的长期保存提供了通用语义框架[26]。在VKU重构语境下，AI驱动的古籍细粒度语义重构需要构建OAIS与中国行业标准双轨对接的智能归档架构。

3.1　架构总体设计与逻辑层次

古籍数字化治理通常呈现自下而上的分层特性，底层数据强调物理实体的真实性锚定，中间层强调语义信息的解耦与重组，顶层强调知识服务的交互与流通。基于视觉知识元与OAIS的古籍智能归档系统架构（见图1），由数据层、处理层、管理层、服务层4个逻辑层次构成：数据层是整个架构的基础，负责存储原始扫描件、分层图像、VKU结构化数据及各类元数据，不可变更；处理层承载VKU抽取、图层分解、语义解析等

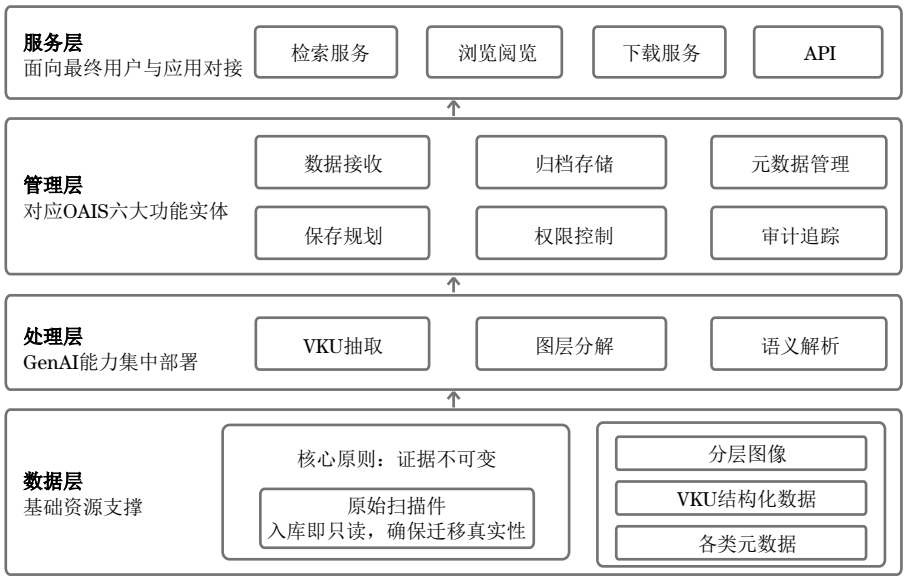

图1　基于视觉知识元与OAIS的古籍智能归档系统架构图

计算任务，是GenAI能力的集中部署位置；管理层对应OAIS的六大功能实体；服务层面向最终用户。

架构设计的核心理念是分层差异治理，遵循两条基本原则：一是存算分离原则，原始证据与计算派生在物理上隔离，确保即使模型发生幻觉或迭代，图书馆始终保有回滚至原始状态的能力；二是风险显性化原则，凡经AI生成或修补的区域，必须在元数据中强制标注置信度与生成参数，将隐性的算法风险转化为显性的治理对象。

3.2　三层信息包与保存描述信息扩展

为实现证据不变、派生可控、过程可查的治理目标，本研究在OAIS框架下将内容对象按证据等级划分为三层，即证据层、内容层和溯源层（见图2）。证据层是内容层的唯一输入源，即使生成式模型出现系统性错误或迭代升级，始终可以从证据层重新生成内容层；内容层存储经语义解耦后生成的VKU实例及其结构化元数据；溯源层记录从证据层到内容层的所有转换过程。这三层共同构成完整的AIP，从而实现GenAI输出向可归档、可审计、可回滚对象的有效转化。

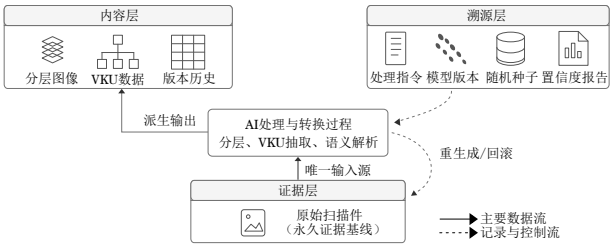

图2　基于OAIS的三层AIP设计

保存描述信息（Preservation Description Information，PDI）是AIP的组成部分，PDI由溯源信息、参考信息、固定性信息、语境信息与访问权限信息5个组成部分[27]。本研究针对GenAI场景对PDI进行扩展，在溯源信息中增设“生成参数记录”子类，包括模型标识、模型版本、推理时间戳、随机种子、提示词、推理步数、置信度、质量标签等字段。这一理论设计的核心逻辑：只有通过PDI完整记录了模型的生成轨迹，生成式重构才是可信的。此外，参照《人工智能生成合成内容标识方法》，所有AI生成或修补的VKU在元数据中必须明确标注“AI生成”标识，在用户界面中以水印、图标或悬浮等可视化方式呈现，确保用户知情权。

三层信息包（SIP、AIP、DIP）与PDI扩展构成了可信归档的数据架构基础，但“可信”的实现还需要具体的技术手段保障证据链的完整性与不可篡改性。当前可供选择的技术路径主要有3条：①内容来源认证，即内容来源与真实性联盟（Coalition for Content Provenance and Authenticity，C2PA）标准，通过在派生对象中嵌入可验证的来源声明清单，为每个VKU附加操作历史与模型指纹，该方案轻量成熟、实施成本低，适合单机构内部的可信保障；②哈希链与数字签名，对证据层的原始扫描件计算密码学哈希值形成不可篡改的证据基线，内容层的每次版本更新均附带数字签名与时间戳，适合多机构协作场景下的证据互认；③区块链存证，将关键哈希值（如证据层原始件摘要、内容层版本摘要）上链存证[28]。

4　技术实现路径

VKU的生成依赖“理解”与“重构”两类能力的协同。理解力是指视觉语言模型负责解析页面的语义结构，识别对象类型并确定空间边界；重构力是指分层生成模型负责将纠缠的视觉层解耦为独立图层。本节从工艺流程角度详细阐述VKU的生成机制，而非聚焦于算法细节。

4.1　模型选型依据

在视觉语言模型选型方面，本研究选择Qwen2.5-VL作为语义解析阶段的核心模型。首先，由于古籍页面尺寸和长宽比多变，Qwen2.5-VL的原生动态分辨率机制恰恰能处理任意长宽比的高分辨率图像，无需缩放裁剪，从而避免印章、批注等细小对象因预处理而丢失细节；其次，该模型的视觉定位能力可输出包含边界框坐标的结构化数据，直接对接VKU形式化定义中的空间边界字段，减少了定位信息的二次转换损耗；最后，Qwen2.5-VL经过大规模中文语料训练，对古籍中的繁体字、异体字具备基本识别能力。

分层生成模型Qwen-Image-Layered可以解决传统分割模型硬性归类的缺陷。Alpha通道提供连续的透明度值，自然表达软边界；生成式架构能够推断被遮挡区域的内容，实现智能补全；RGBA图层天然具备层次结构，保留叠压顺序。而且Qwen-Image-Layered模型采用的RGBA变分自编码器架构，使模型能够根据图像复杂度动态调整输出层数，实现深度的语义解耦。

Qwen-Image-Layered与Qwen2.5-VL同属Qwen系列，二者在特征表示空间上具有天然的亲和性，有利于“理解—重构”两个阶段的无缝衔接。后续研究将在更多古籍样本上引入书生·万象（InternVL）等开源模型，进行系统化的对比评估，以进一步验证选型的合理性。

4.2 面向语义解耦的工艺流程

（1）语义解析。该环节的任务是从原始页面影像中提取元素清单，由视觉语言模型承担核心解析工作，工艺要点在于提示词设计。图书馆员需定义提取目标，模型据此输出结构化的元素清单，每个元素包含类型判定、边界检测、初步置信度以及可选的语义解释4项信息。系统同时生成元素间的空间关系图，其中空间关系边集记录包含、覆盖、相邻、分离等拓扑关系。

在实际操作中，提示词设计采用分阶段策略。第一阶段为全局版面分析，输入整页图像并要求模型输出版面整体布局，包括版框范围、文字区域、非文字区域的大致分布；第二阶段为局部对象检测，针对第一阶段识别出的各区域逐一发送精细化提示，要求模型检测区域内的具体对象，例如“识别此区域中的所有印章并输出边界框坐标”。面向古籍场景的提示词设计需要指定古籍特有的对象类别，像鱼尾、版心、避讳缺笔这类在通用文档中罕见的元素，同时附加古籍特有指令，例如“区分正面文字与透背墨迹”“注意识别叠压区域中被遮挡的文字”。

（2）分层重构。该环节旨在将耦合的页面分解为语义独立的图层。古籍场景的基线配置包含文字层（墨迹笔画）、印章层（朱印）、病害层（透背、霉斑、虫蛀等）、底纹层（纸张纹理）4个层次。分层重构的输入为原始页面图像与提取的元素清单，输出为多个RGBA图层。具体流程为：①根据元素清单确定分层方案，将各对象按语义类别分配至对应图层，叠压区域的层序依据物理规律确定，印章在文字之上，底纹在所有内容之下；②为每个目标图层构建文本提示引导，进行分层生成，模型再据此输出各图层的RGBA图像；③对Alpha通道进行等边界精化和重叠区域一致性校验后处理，去除图层间的溢出像素。在透背场景中还需要执行方向性判断，区分正面渗出与背面渗入的墨迹走向，以确保透背层的分离方向正确。

（3）封装与质量分流控制。该环节遍历元素清单为每个元素生成对应的VKU实例。封装流程按6个步骤依次执行：①标识分配与层级绑定，为每个VKU生成全局唯一标识，并绑定其父页面标识与图层标识；②边界计算，从分层图像的Alpha遮罩中提取对象边界多边形，写入空间边界；③语义属性填充，根据对象类型调用对应的语义解析规则，将识别结果写入语义属性集，并校验其是否满足该类型的最小必备语义属性要求；④嵌入向量计算，对VKU区域图像执行视觉编码，生成嵌入向量供后续相似性检索使用；⑤溯源信息组装，将生成参数写入溯源信息；⑥置信度评估，综合模型输出概率与Alpha遮罩质量计算最终置信度。

封装过程中提及的置信度是VKU治理的核心指标。本研究引入基于置信度的三级分流机制，参考文档分析与OCR领域的工业级置信度分级实践，设定具体分流阈值[29–31]。对于置信度≥0.85的VKU将自动入库并标记为“已发布”；当0.70≤置信度<0.85时，该区间的字符识别存在形近字混淆等不确定性[29–30]，故保留机器处理结果，但需标记为“待审核”，在服务前端向用户展示不确定性提示；对于置信度<0.70的识别结果，模型输出的可靠性不足[31]，此类对象将被自动拦截并强制进入“专家确证队列”，由人工进行最终裁决。封装完成后，VKU集合进入质量验证与分流处理环节。

本研究以C2PA标准方案构建VKU封装环节的来源声明清单。对于每一个AI生成的派生图层，系统均构建对应的来源声明清单，该清单由三部分组成：一是动作断言，用于记录图像的编辑历史，以保证处理过程的可追溯性；二是软件代理信息，用于记录所使用模型的指纹信息；三是溯源掩码的可视化缩略图，作为视觉证据的直观索引。

5 实证研究

为论证VKU与归档架构的适用性逻辑，本研究以南京图书馆建设的江苏省古籍数字资源集成平台[32]为选样库，从5 000余部全文图像中挑选符合场景的样本，从4个典型场景进行实证研究。

5.1 场景一：叠压印章的无损揭示

对于印章盖于正文文字之上的“朱墨叠压”现象，传统处理方式存在以下问题：若优先识别正文文字，印文区域将成为干扰噪声；若优先保留印章，被覆盖的正

文文字将丢失。使用分层生成模型可将页面分解为用于独立阅读的纯净文字层、可用于鉴赏的独立印章层，并通过遮挡关系推导出时序关系层，用于版本考证（见图3）。文字层通过智能化推断补全了被印章遮挡的部分，印章层则保留完整的印鉴图样与晕染细节。Alpha通道为每枚印章生成包含印文、印文篆法、印泥颜色、叠压关系的VKU实例，使跨卷册的印章关联检索成为可能。

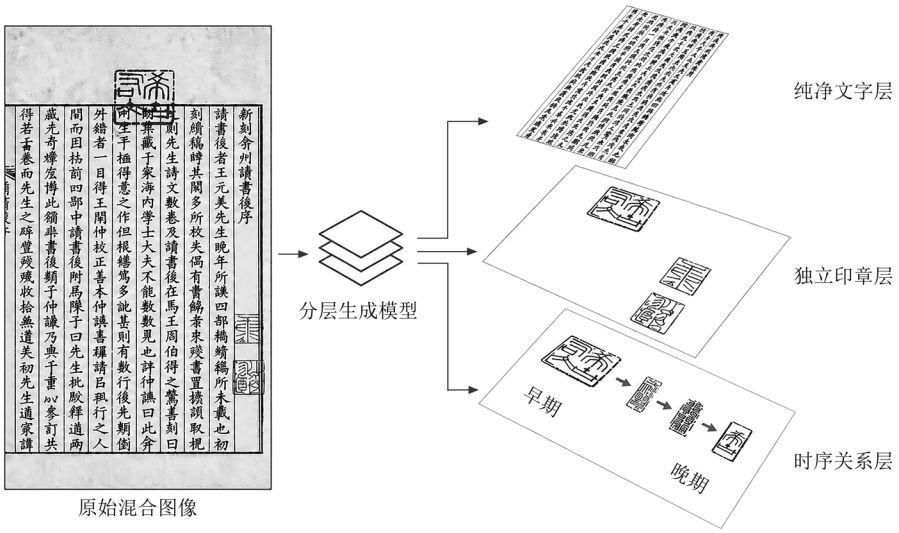

图3　基于GenAI的古籍印章分层解耦图

5.2　场景二：透背古籍的语义修复

透背现象在清薄纸刻本和民国石印本中尤为突出。在对此类古籍进行数字化采集时，常出现背面墨迹透过纸张干扰正面的现象，采用传统二值化处理极易造成笔画断裂或墨迹残留。本研究将透背视为一个独立的语义图层加以处理，通过分层生成模型将页面分解为正面文字层、透背层和底纹层3个组件进行独立封装，从理论上消除了背景噪声对OCR的干扰；底纹层则保留纸张纹理信息，可支撑后续的纸张鉴定研究。这种处理方式实现了非破坏性修复的治理目标。

5.3　场景三：缺损识别与修复优先级排序

传统修复工作需逐页目验评估缺损状况，工作量大且难以形成系统化排序。经过VKU重构后，视觉语言模型将自动识别虫蛀孔洞、霉变污渍、水渍、撕裂等缺损区域，生成包含缺损类型、几何尺寸、严重程度的缺损类VKU，基于VKU的空间边界信息，可进一步计算缺损区域与文字区域的空间重叠比例，判断缺损情况对该区域文字可读性的影响程度。然后根据古籍的物理受损类型和受损位置，划分出修复优先等级，高优先级修复区涉及关键正文的缺损；中优先级修复区涉及题跋、印章等缺损；低优先级修复区涉及指向页边空白处的污渍或缺损。

在修复资源有限的情况下，可优先处理缺损严重且影响文字可读性的区域。同时，通过缺损VKU的时序记录可支撑劣化监测功能，对比不同时间点的变化以评估保存环境有效性。

5.4　场景四：生僻字的专家确证机制

古籍中的生僻字、异体字是GenAI识别与推断的薄弱环节，模型可能将残损笔画误判为其他字形，或对缺失部分作出错误推断，这是当前古籍数字化应用AI技术的主要风险所在。参照前文质量控制中提及的置信度理论，当置信度分流机制识别出置信度＜0.70的文字类VKU时，这类VKU会被界定为低置信度VKU，将自动进入专家确证队列。专家界面左右对照显示原图与AI解耦的文字图层、置信度值及模型推断补全的文字，存疑部分高亮提示，支持“确认”“修正”“标注不确定”3种操作。专家所有修订仅记录于内容层，不触及证据层原始扫描图像，系统同步记录修订人、时间及具体修改内容，生成版本号并写入溯源层。

这一机制的设计逻辑是承认AI能力边界，通过“机器筛选+专家确证”的协作模式提升整体效率。专

家只需处理低置信度VKU，效率可望显著提升。

5.5 小规模测试与概念验证

为了从概念推演迈向实证探索，本研究在南京图书馆馆藏古籍中选取10页具有代表性的样本开展原型测试。兼顾叠压印章（4页）、透背干扰（3页）和虫蛀缺损（3页）等典型问题，以检验VKU在不同场景下的基本适用性。

测试按“解析—重构—封装”工艺流程执行。语义解析环节使用Qwen2.5-VL对每页图像进行对象识别与空间定位。分层重构环节使用Qwen-Image-Layered按文字层、印章层、病害层、底纹层的基线配置进行图层分解。每页样本生成的VKU实例数在12~15个，测试结果由2位具有10年以上古籍工作经验的馆员进行定性审核。在印章分离任务中，生成的文字层经馆员审核认为基本可用于阅读与检索，但印章边缘区域存在少量残留色素。在透背修复任务中，馆员反馈分离后的正面文字层在视觉清晰度上有明显改善。在缺损识别任务中，模型对受损类型识别和缺损严重程度的评估结果经馆员审核获得了较为积极的反馈。

需要说明的是，受测试条件限制本次原型测试为小规模探索性验证，样本量有限且以定性评估为主，目的在于检验VKU从概念到实践的基本可行性，而非对框架效能进行全面定量评价。后续研究将在更大规模的样本集上设计更规范的实验方案。

6 结论

本研究针对当前古籍数字化中页级单元与细粒度知识服务需求之间的粒度错配问题，构建了从概念界定到架构设计再到工艺流程的系统化解决方案。以VKU作为古籍资源组织的新型中介层，使古籍资源组织从文献单元管理过渡到要素单元管理；将国际长期保存语义框架与本土规范系统化融合，构建了OAIS与中国行业标准双轨对接的智能归档架构；进一步提出分层AIP与PDI过程描述的组合策略，使VKU的生成过程也成为可保存对象的一部分，为古籍生成式处理提供了可落地的解释路径，为古籍资源从图像数字化走向语义数字化提供理论基础与实践框架。进一步在知识服务层面设计了基于置信度的封装策略，以证据层永久保真保存为底线，将分割、去噪、补全等派生结果约束在不同层级中，用户获取的不再是古籍的图片集合，而是可被检索、对照、解释的语义资产。

本研究主要的贡献在于理论与框架设计，在实践层面未开展大规模系统实验和精度评估，仅以南京图书馆古籍数字化难题为切入点，进行了小规模定性测试，初步给出了从物理扫描到语义资产转化的治理方案。在后续研究中，将重点关注多类型古籍场景与VKU之间的互操作映射及其稳定性，在内容标识与来源说明机制上进一步与国家强制性要求对齐，以期为推动古籍数字化服务体系的可持续发展提供可参照的方案。

参考文献

[1] 中共中央办公厅 国务院办公厅印发《关于推进新时代古籍工作的意见》[EB/OL]. [2025-12-20]. https://www.gov.cn/zhengce/2022-04/11/content_5684555.htm.

[2] 韩小凡. 我国古籍数字化平台知识服务研究[D]. 南京：南京师范大学，2022.

[3] 籍合网[EB/OL]. [2026-02-20]. https://www.ancientbooks.cn/.

[4] 中华经典古籍库[EB/OL]. [2026-02-20]. https://jingdian.ancientbooks.cn/6.

[5] 识典古籍[EB/OL]. [2026-02-20]. https://www.shidianguji.com.

[6] 张晓冰，张佩. 古籍数字化出版的挑战与发展路径研究：以“识典古籍”为例[J]. 北京印刷学院学报，2024，32（9）：1-6.

[7] 王秋云，龙欢. 我国古籍领域开展数字人文研究与实践现状分析[J]. 图书馆界，2023（4）：19-25.

[8] 陈涛，赵晓飞，杨鑫，等. 古籍文献通用知识模型研究与设计[J]. 信息资源管理学报，2025，15（1）：139-153.

[9] 国家图书馆. 全国智慧图书馆体系建设：古籍数字化和知识标引规范手册[M]. 北京：国家图书馆出版社，2023.

[10] 索传军，戎军涛. 知识元理论研究述评[J]. 图书情报工作，2021，65（11）：133-142.

[11] 高劲松，付家炜. 基于知识元的文物图像资源细粒度知识表示方法研究[J]. 情报科学，2022，40（1）：16-24.

[12] 潘云鹤. 从视觉知识到视觉理解[J]. 中国工业和信息化，2023（Z1）：51-58.

[13] PAN Y H. On visual understanding[J]. Frontiers of Information Technology & Electronic Engineering，2022，23（9）：1287-1291.

[14] JIANG T，XIA S，XU Y，et al. VKnowU: Evaluating Visual Knowledge Understanding in Multimodal LLMs[EB/OL].

[2026−02−21]. https://arxiv.org/abs/2511.20272v1.

[15] DONG Q，ISMAYIL M，HAMDULLA A，et al. The YOLO model that still excels in document layout analysis [J]. Signal，Image and Video Processing，2023，18（2）：1539−1548.

[16] 何蕊. 唤醒古籍[N]. 北京日报，2024−12−26（16）.

[17] 王丹阳. AI陪你逛敦煌千年数字图书馆[N]. 羊城晚报，2025−06−04（A01）.

[18] BAI S，CHEN K，LIU X，et al. Qwen2.5-VL Technical Report [EB/OL]. [2025−12−25]. https://arxiv.org/abs/2502.13923v1.

[19] YIN S，ZHANG Z，TANG Z，et al. Qwen-Image-Layered: Towards Inherent Editability via Layer Decomposition[EB/OL]. [2025−12−25]. https://arxiv.org/abs/2512.15603v1.

[20] 张晓林，吴振新，赵艳，等. 国家数字科技文献资源长期保存体系的战略与实践[J]. 图书馆杂志，2017，36（12）：14−19.

[21] 中华人民共和国文化部. 古籍元数据规范：WH/T 66—2014[S]. 北京：国家图书馆出版社，2015.

[22] 中华人民共和国文化和旅游部. 汉文古籍版式描述规范：WH/T 100—2023[S]. 北京：中华人民共和国文化和旅游部科技教育司，2023.

[23] 国家市场监督管理总局，国家标准化管理委员会. 网络安全技术人工智能生成合成内容标识方法：GB 45438—2025[S]. 北京：中国标准出版社，2025：13−15.

[24] 吴欣雨，吴振新，曹晓丽，等. 国际数字资源长期保存领域技术体系及相关工作进展分析[J]. 图书情报工作，2025，69（23）：137−148.

[25] 中华人民共和国文化部. 古籍特藏破损定级标准：WH/T 22—2006[S]. 北京：中华人民共和国文化部，2006.

[26] Space Data System Practices-reference model for an open archival information system (OAIS): ISO 14721:2025[S/OL]. [2026−01−16]. https://www.iso.org/standard/87471.html.

[27] 王紫丞. 基于OAIS的电子档案单套管理模型构建研究[D]. 长沙：中南大学，2024.

[28] 刘越男，吴云鹏. 区块链技术在档案管理中应用路径研究[J]. 档案学通讯，2020（6）：57−64.

[29] Understanding Confidence Scores: API V2 Documentation [EB/OL]. [2025−12−28]. https://v2.docs.werk24.io/concepts/confidence/.

[30] LEVEREGE. Computer Vision Basics: Confidence & Accuracy[EB/OL]. [2026−01−08]. https://www.leverege.com/blogpost/computer-vision-basics.

[31] MICROSOFT AZURE. Confidence score in Custom question answering[EB/OL]. [2026−01−08]. https://learn.microsoft.com/en-us/azure/ai-services/language-service/question-answering/concepts/confidence-score.

[32] 江苏省古籍数字资源集成平台[EB/OL]. [2025−12−20]. http://zggj.jslib.org.cn/#/home.

作者简介

刘丽娜，女，硕士，副研究馆员，研究方向：数字资源开发，E-mail：396068237@qq.com。
蔡茜，女，学士，馆员，研究方向：智慧图书馆。

Research on the Reconstruction and Trustworthy Archiving of Ancient Book Visual Knowledge Elements Based on Generative Semantic Decoupling

LIU LiNa　CAI Qian
(Nanjing Library, Nanjing 210018, P. R. China)

Abstract: The digitization of ancient books is moving toward higher quality and intelligence. However, the current whole-page scan as the basic management unit suffers from granularity mismatch, weak semantic association, and ambiguous evidence provenance. Generative Artificial Intelligence is shifting ancient book image processing from text recognition toward semantic understanding and visual reconstruction. In response, this study introduces the concept of visual knowledge unit with its formal definition and classification. Based on visual knowledge unit theory, a three-layer intelligent archiving architecture is constructed and integrated with the Open Archival Information System and relevant Chinese industry standards, aiming to achieve immutable evidence, controllable derivation, and verifiable processes. This study further presents a reconstruction workflow for semantic decoupling, followed by application scenarios and preliminary validation using ancient books from the Nanjing Library collection. The work provides a practical framework and pathway for moving ancient book resources from image-based digitization to semantic digitization.

Keywords: Visual Knowledge Unit; Ancient Book Digitization; GenAI; Semantic Decoupling; OAIS; Knowledge Organization; Digital Humanities

（责任编辑：常春）