Text2shape深度检索模型：生成基于案例的推理背景下机械零件再设计的初始案例

推理

Tianshuo Zang †、Maolin Yang †、Wentao Yong 和 Pingyu Jiang *

制造系统工程国家重点实验室，西安交通大学，中国西安 710049；ibubble@stu.xjtu.edu.cn (T.Z.)；maolin@xjtu.edu.cn (M.Y.)；yongwentao@stu.xjtu.edu.cn (W.Y.) * 通信作者：pjiang@mail.xjtu.edu.cn

† 这些作者对本研究的贡献相同。

摘要： 在基于案例的推理背景下，为新的设计需求从历史案例库中检索相似方案，是机械零件再设计的第一步。然而，当案例库规模较大时，人工检索方法效率较低。此外，简单推理算法（例如基于规则的推理、决策树）难以涵盖复杂设计方案中的所有特征。为此，我们建立了一种 text2shape 深度检索模型，以支持基于文本描述的机械零件形状检索，其中的文本用于描述目标机械零件的结构特征。具体而言，我们采用特征工程来识别目标机械零件的关键结构特征。基于识别出的关键结构特征，我们构建了一个包含 1000 个样本的训练集，每个样本由一段描述一组结构特征的文本以及相应的结构特征三维形状组成。我们定制了循环神经网络和三维卷积神经网络算法，以构建 text2shape 深度检索模型。我们采用正交实验进行车削建模。最终，模型的准确率最高达到 0.98，因此该模型可有效检索机械零件再设计的初始案例。

关键词： text2shape 检索；再设计；卷积神经网络；循环神经网络

1. 引言

基于案例的推理机制可通过检索与新设计需求最相似的历史设计方案，有效应用于产品再设计活动 [1]。随后，可进一步修改检索到的方案，使其符合新的设计需求 [2]。通常，可采用两种方法执行检索任务：人工检索和基于简单推理算法的检索。然而，这两种方法都有缺点。例如，当设计方案库规模较大时，人工检索方法效率低下；而简单推理算法无法涵盖复杂设计方案中的所有结构特征。例如，在皮带轮设计任务中使用决策树模型时 [3,4], , 首先需要建立一组产生式规则，再将其组合成决策树。在此过程中，每条产生式规则都会考虑皮带轮的一项功能或结构特征。因此，如果需要考虑大量设计特征，决策树就会变得庞大而复杂，进而导致推理效率下降。

随着深度学习技术的快速发展，基于深度学习的形状检索模型已成为解决上述问题的一种可行方案 [5–8]。根据输入数据的格式，检索模型可分为多维检索和单维检索，而单维检索又可进一步分为基于文本的检索 [9]，基于图像的检索 [10–15], 草图到形状 [16–18], 以及描述符到形状[19–22]。通常，原始产品设计需求以文本描述的形式呈现，因此，在基于案例的推理驱动的产品

重设计。

然而，在为机械零件再设计任务构建基于深度学习的文本到形状检索模型时，至少需要考虑两个问题。首先，原始设计需求通常以文本描述的形式表达，属于一维数据；而机械零件的设计方案则以三维形状的形式呈现。因此，要实现从低维数据到高维数据的准确检索，必须弥合数据维度差距 [23]。另一方面，用于机械零件检索的 text2shape 模型比用于简单形状检索任务（例如，通常研究的桌子、书桌和椅子检索模型）的模型更复杂，因为机械零件的结构特征比简单形状更复杂，因此其数据特征空间也会更加复杂 [24]。

为此，本文构建了一个用于三维形状检索的集成深度学习模型。该模型能够根据描述新设计需求的输入文本，检索机械零件的三维形状。由此，检索到的三维形状可作为基于案例的推理背景下再设计任务的初始设计方案。我们的检索模型具有鲜明的工程设计背景。为此，我们构建了一个包含 1000 个连杆样本的专用训练集，每个样本均由一段文本描述及其对应的机械零件三维形状模型组成。基于该训练集，我们构建了一个集成 CNN-RNN 与三维卷积神经网络的 text2shape 检索模型，并针对训练集的潜在空间对模型结构和参数进行了专门定制。还值得一提的是，样本的文本描述是根据特征工程流程编排的；该流程分析目标机械零件（即连杆）的关键结构特征，因此，与基于自由文本描述的检索相比，该模型具有更高的检索准确率。此外，我们还采用了基于正交试验的调优方法，以便快速确定深度学习模型的合适超参数。

本文其余部分安排如下。第2节回顾与构建我们的 text2shape 深度检索模型相关的工作。第3节详细介绍如何通过特征工程构建训练集，以及如何构建适用于该训练集的深度检索模型。第4节通过案例研究展示该模型在连杆检索任务中的可操作性。第5节讨论本研究的贡献，并简要总结全文。

2. 相关工作

如第1节所述，本研究的核心思路是利用文本到形状检索模型，识别与新设计需求最相似的历史设计方案的三维形状。设计需求以文本形式描述。因此，文本数据表示、三维形状数据表示和文本到形状检索是本研究的三项基础技术。在本小节中，我们回顾这三项技术，并讨论相关研究空白。

2.1. 文本数据表示与分析

Jiang等人[26] 提出了一种基于深度神经网络的文本表示模型，称为基于主题的连续词袋模型。该模型最重要的特点是能够实现全局与局部上下文融合。基于通过该模型获得的词分布表示，研究者提出了一种基于 TF-IWF 加权池化机制的短文本表示方法。Yan等人[27] 提出了一种原始表示学习方法，旨在扩展场景文本图像的内部表示。该方法的一个关键点是将特征图中的元素建模为无向图节点。该方法的框架主要由池化聚合器、加权聚合器和原始表示学习网络组成。聚合器的作用是将原始表示转换为高层视觉文本表示。该方法在场景文本识别任务中表现出很高的效率。Luo等人[28] 提出了一种将潜在狄利克雷分配（LDA）文本表示与卷积神经网络相结合的方法，用于文本数据的情感分析。该方法使用 LDA 在潜在空间中训练短文本的主题分布，并使用 GRU-CNN 作为特征向量的分类器，从而提高了文本分类的准确率。

Sinoara等人[29] 提出了一种根据文档集合的嵌入表示来表示文档集合的方法，该方法通过计算枢轴词与上下文词之间的距离，获得上下文中各词的权重。由此提供了一种有效的方法来提高嵌入向量的可解释性。Song等人[30] 提出了一种联合变分自编码器（VAE），用于表示案例文本的嵌入表示。该方法使用 VAE 将案例文本的统计特征和内容特征嵌入同一潜在空间，与采用单一特征嵌入机制的模型相比，取得了更好的性能。Shun等人[31] 提出了一种多义词感知向量表示模型（PAVRM）。该模型能够生成准确的词向量表示，还能通过上下文聚类算法识别语料库中的多义词。Hou等人[32] 提出了一种基于实体的概念知识感知模型（ECKA），该模型能够将语义信息融入短文本表示中。该模型基于卷积神经网络算法开发，能够从词语、实体、概念和知识层中提取语义特征。该模型可以解决因文本稀疏且篇幅短而导致分类效果不佳的问题。

由上述文献可知，DNN 和卷积神经网络均可有效用于表示文本数据的特征。然而，在我们的研究中，用于描述机械零件结构特征的文本相对较长，因此特征空间会更加复杂。因此，与使用单一模型相比，采用集成模型来表示描述机械零件结构特征的文本可能更为有效。

2.2. 三维形状数据表示与分析

Feng 等人 [33] 提出了一种用于三维形状表示的网格自然网络，命名为 MeshNet。该模型引入了人脸形状单元和特征分割机制，从而构建了一个用于人脸三维形状表示的通用模型。在该模型中，建立了空间块和结构描述符块来表示初始特征，并建立卷积块来聚合相邻特征。Gao 等人[34] 提出了一种用于三维网格表示学习的局部结构感知各向异性卷积操作（LSA Conv）。LSA 卷积操作能够学习每个节点的自适应权重矩阵，且无需进行预处理和后处理。Rinon 等人 [35] 提出了一种基于无监督学习 MRGAN 算法的三维形状表示方法，用于部件分离。其网络架构采用树结构图，根节点的每个分支代表一个形状不同的部件。该方法通过采用根混合训练策略和一组损失函数，能够促进分支分离，并避免分支退化或过度生长。Jiang 等人 [36] 提出了一种用于三维形状表示的局部隐式网格（LIG）方法，并在规则网格的每个区域中采用隐式特征向量编码机制，以解决可扩展性和通用性问题。Zheng 等人 [37] 提出了一种用于三维形状表示的无监督深度隐式模板方法。其核心思想是将深度隐式函数（DIFS）表示为模板隐式函数的条件变形，并提出了空间扭曲长短期记忆网络，将条件空间变换分解为多个逐点变换，以确保泛化能力。Zhang 等人 [38] 开发了一种多分辨率深度隐式函数（MDIF），用于表征三维模型的形状特征。该模型的显著特点是采用特殊的层次网络结构，能够表示不同水平的三维形状细节，并通过引入多分辨率机制和解码器优化来提升形状生成性能。Tristan 等人 [39] 提出了一种基于隐式架构的新型形状表示方法，可实现快速可微渲染，从而解决隐式形状表示中渲染效率低下的问题。Wei 等人 [40] 提出使用规范视图来表示三维模型的形状特征。规范视图可以将任意视图的原始特征转换为固定数量的视图特征。此外，研究还提出了归一化视图的特征分离约束，以确保归一化视图表示中的视图特征能够嵌入欧几里得空间中的散点，并保证特征的可辨别性。Edgar 等人 [41] 提出了一种新型的基于中层面片的表面表示方法。该方法将深度图的部分点云补全、形状插值和对象生成模型作为新的应用。模型中的面片能够表示任意拓扑形状，同时支持任意分辨率提取。Hao

等人 [42] 提出了一种名为几何特征统计直方图（GFSH）的三维特征描述子。该描述子首先通过构建加权协方差矩阵，求得稳定可靠的局部参考框架（LRF），然后对多种几何分布特征进行统计，从而实现对三维特征的描述。该方法能够缓解三维特征描述子在有效性、鲁棒性、紧凑性和效率之间难以平衡的问题。

从上述文献可以看出，使用复杂的深度学习网络直接表示三维形状的数据结构或三维形状的二维图像，其计算过程可能十分复杂且耗时。因此，如何同时提高三维形状表示的效率和准确率，仍是其工程应用中有待填补的研究空白。

2.3. 从文本描述到三维形状的检索

Han 等人 [43] 提出了一种名为 Y2Seq2Seq 的基于视图的模型。该模型基于视图序列和词序列的联合重建来学习跨模态匹配关系。Y2Seq2Seq 结构通过两个耦合的“Y”型序列到序列（Seq2Seq）结构（即模型的主体结构），能够有效连接跨模态语义。Yue 等人 [44] 提出了一种用于文本到三维形状检索的三模态训练方案，涵盖体素、图像和文本。通过应用多视图三维模型渲染和端到端训练方法，该模型的表现优于双模态模型。Wu 等人 [45] 提出了一种多模态映射的两阶段学习方法。第一阶段结合低层特征和语义信息，以表示跨模态语义特征向量；第二阶段则采用编码器—解码器范式来学习匹配关系。该方法通过将多模态数据投影到低维嵌入空间，能够保留特征信息和语义信息。Chuan 等人 [46] 提出了一种三维点云与文本数据的联合嵌入方法，以支持文本与三维形状之间的匹配，并采用三元组排序损失计算三维形状与文本数据之间的距离。Qiao 等人 [14] 建立了一种面向装配结构的两级检索模型。在第一级检索模型中，采用语义检索从模型库中识别匹配度最高的结构。在第二级检索模型中，采用一种基于属性邻接图的几何结构检索机制来输出最终结果。Li 等人 [47] 建立了一种基于语义树的三维形状检索模型。该语义树由 WordNet 和普林斯顿形状基准本体库构建；在其支持下，该模型在并行实验中表现更佳。

从上述文献可以看出，多模态检索确实比单模态检索具有更好的持久性。然而，要求提供更多类型的搜索条件，也会降低检索所需历史设计方案的便利性。另一方面，大多数关于文本到形状检索的研究都在计算机科学领域开展，只有少数研究致力于机械零件形状检索任务。因此，应针对我们的特定应用场景构建专用数据集及其相应的深度学习模型。

3. Text2shape深度检索模型

如第1节所述，text2shape检索模型能够识别出与新设计需求最相似的历史设计方案，而识别出的设计方案可在基于案例的推理过程中用作产品再设计的初始方案。本节首先采用特征工程识别目标机械零件（即连杆）的关键结构特征，并为text2shape检索模型构建一个包含1000 个样本的训练集；随后建立由循环神经网络和三维卷积神经网络组成的集成模型，以拟合训练集的潜在空间。该模型的应用场景和构建步骤如图1所示。借助检索模型，输入新设计需求中结构特征的文本描述后，即可从历史设计方案库中检索出一系列与这些文本描述最相似的三维形状，从而为设计师提供初始设计方案，以供后续再设计流程使用。

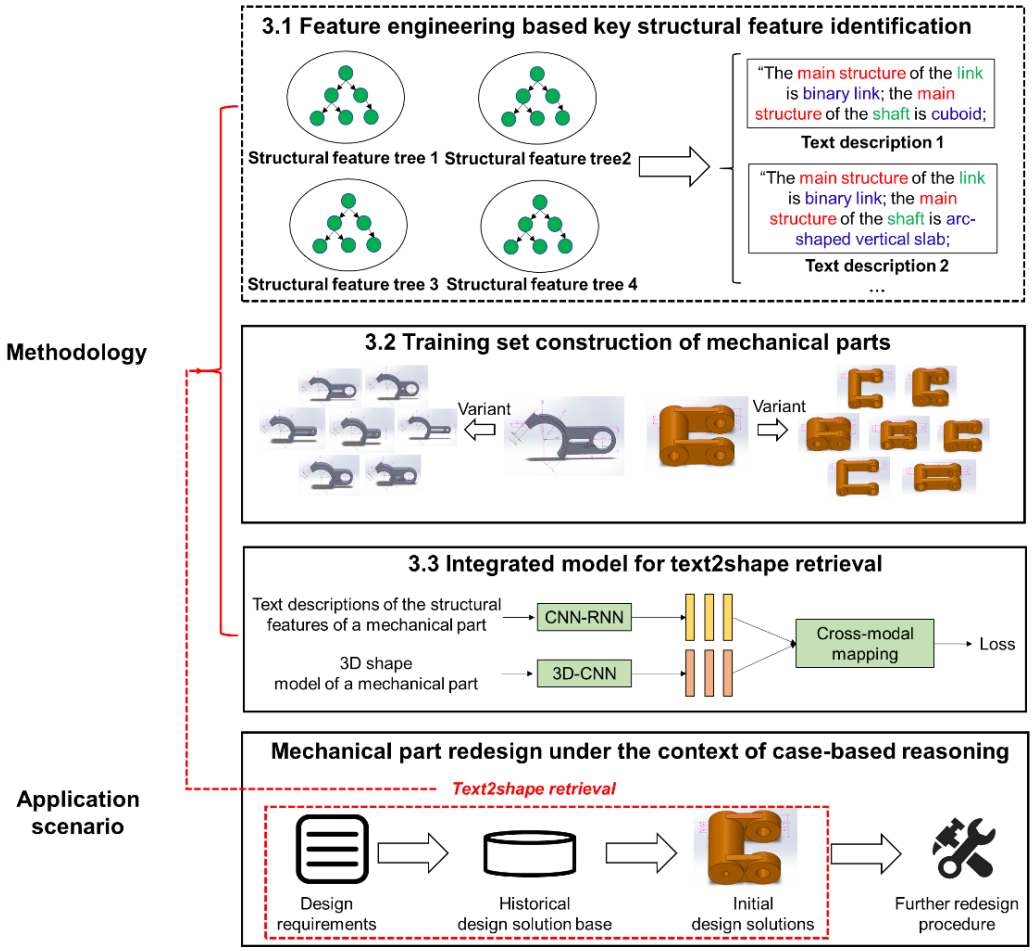


图1。 文本到形状检索模型的应用场景与构建步骤。

3.1 基于特征工程的关键结构特征识别

特征工程大致可以看作一系列从原始数据集中提取更明确特征的方法，从而提升深度学习模型的性能[48,49]。本小节建立了树结构分布，以揭示连杆的关键结构特征，并据此构建训练集（见下一小节），用于训练text2shape检索模型。基于特征工程的关键结构特征识别主要步骤如图2所示。值得一提的是，本文选用连杆作为目标机械零件，因为连杆广泛应用于不同类型的机械装配体中，且包含丰富多样的结构特征。

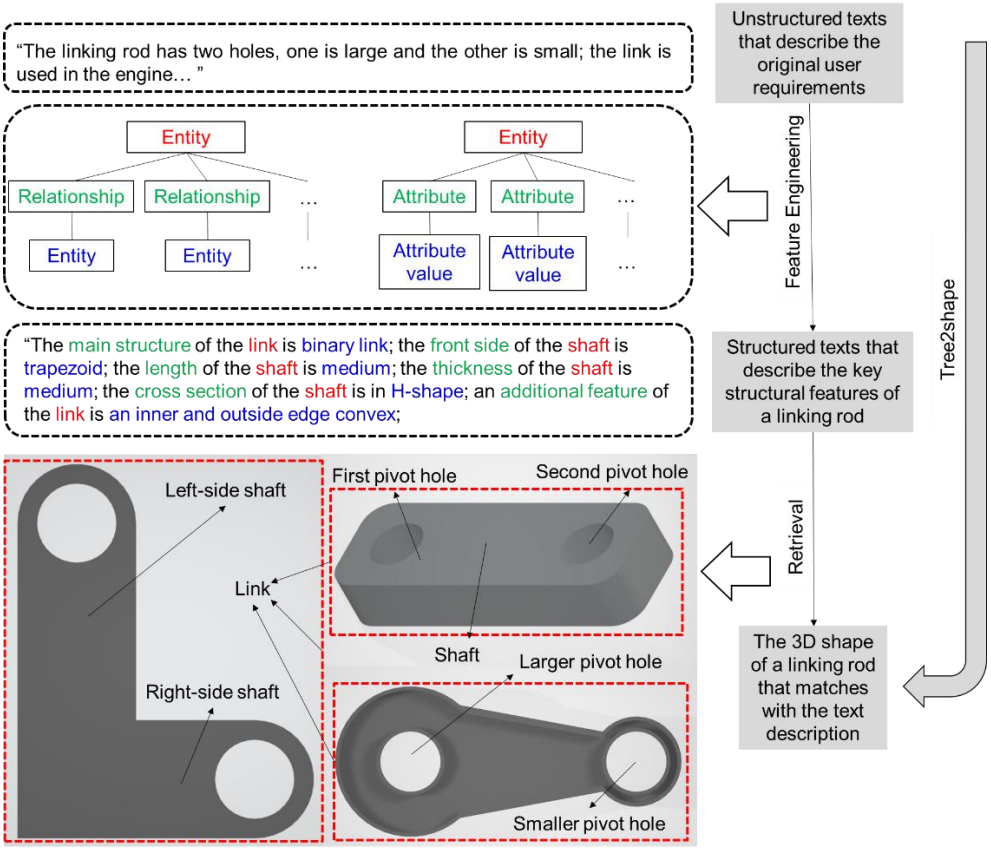


图2. 基于特征工程的关键结构特征识别的主要步骤。

步骤1：关键结构特征识别与术语定义。客户提出的原始设计需求通常是非结构化的，并且可能用不同术语描述相同的结构特征。因此，深度学习模型要表示描述原始设计需求的文本数据，难度会大幅增加。为此，应首先依据专业设计师的领域知识识别目标机械零件的关键结构特征，并确保这些结构特征的组合能够描述目标机械零件系列中的任何变型。此外，还应为每种结构特征指定专用术语，以避免歧义。为此，本文归纳出一组用于定义连杆的关键结构特征（例如，二元连杆、较大的枢轴孔、内径），并为每种特征统一命名，如图3所示。

步骤2. 基于三元组的特征关系表达。 在此步骤中，将每对结构特征之间的关系描述为"实体–关系–实体"，并将每个结构特征及其可能的特征值描述为"实体–属性–属性值"的形式。例如，第一个枢轴孔是连杆的结构相关特征；因此，可以生成一个三元组："连杆–结构特征–第一个枢轴孔。"类型是第一个枢轴孔的一个属性，而分离类型可以是类型的一个可能特征值；因此，这个结构特征可以表示为"第一个枢轴孔–类型–分离类型。"需要注意的是，特征的属性大致可分为两类：结构相关属性和尺寸相关属性。例如，分离类型是第一个枢轴孔的结构相关属性，而内径是第一个枢轴孔的尺寸相关属性。还值得一提的是，结构相关属性的值也可以拥有自己的属性和属性值。例如，在图3中，附加特征是一种结构相关属性，内边缘凸起是其属性值之一；而厚度是内边缘凸起的尺寸相关属性，厚度也有其自身的属性值（即小、中、大）。

步骤3：树状结构特征描述。将关键结构特征表示为三元组后，可以按树状结构描述的形式将它们连接起来，如图3所示。总共识别出了八个特征实体，

包括（即第一个枢轴孔、大型枢轴孔、轴、连杆等），每个实体都可以根据其结构相关属性和尺寸相关属性展开为树状结构。值得一提的是，这些实体中的每一个也都可以通过“结构特征”关系与连杆实体相连。

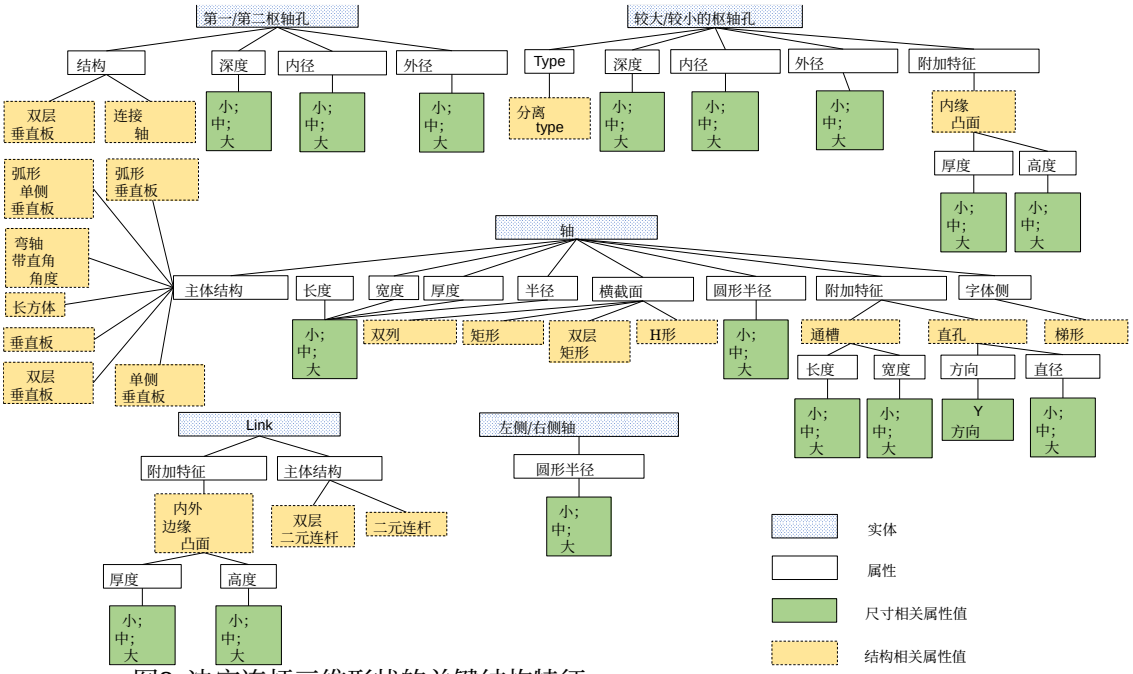


图3. 决定连杆三维形状的关键结构特征。

3.2. 训练集构建

本小节为我们的 text2shape 深度检索模型构建了一个包含 1000 个样本的训练集。每个样本都包含一个连杆的三维形状模型，以及一段描述该连杆关键结构特征的文本。

3.2.1. 构建三维形状模型

首先，从不同的工程应用场景中选取 15 种不同类型的连杆，它们各自具有不同的结构相关特征属性，彼此之间各不相同。使用*Solidworks*2021以 STL 文件的形式绘制连杆的原始三维模型。在此过程中，三维模型的尺寸被限制在 64 mm3 的立方空间内。随后，通过改变各模型的尺寸相关特征属性，将每个三维模型修改为 48 至 64 个变体模型。例如，绘制轴长不同但类型相同的连杆，即可获得该连杆的不同变体模型。然而，这项工作是首次尝试在尺寸敏感型输入条件下检索三维形状的研究之一；因此，为提升深度检索模型的性能，尺寸相关特征属性仅采用三个取值（即小、中、大）。

其次，为了便于深度检索模型表征连杆的三维模型，随后将 STL 文件形式的三维模型转换为三维体素模型，其中包含表面体素信息和内部体素信息，并在转换过程中采用了 AABB包围盒法。值得一提的是

确保所有三维体素模型都应在规则立方空间中构建，例如 4^3、16^3 或 64^3（在本研究中，采用 16^3 以在建模精度和训练效率之间取得平衡）。这样可以确保三维体素模型在深度检索模型中解析并重新渲染后不会发生变形。

最终，以三维体素模型的形式构建了原始15个连杆的共1000个变体模型。图4展示了其中几个模型示例。



图4. 训练集中的三维形状模型示例。

3.2.2. 准备三维模型的文本描述

获得连杆的三维体素模型后，为每个模型配上了一段文本，用于描述其关键结构特征。每段文本描述都介绍了结构相关特征属性（例如连杆类型、轴的主要结构、轴的横截面）和尺寸相关特征属性（例如第一／第二枢轴孔内径、轴长）。还值得一提的是，该段落中的句子均采用“实体–属性–属性值”的形式表达。例如，一段文本描述可以是：“连杆的主体结构是二元连杆；轴的主要结构是长方体；轴长为大；轴宽为中；轴厚为中”。基于第3.1节中确定的关键结构特征（如图3所示），可以准确地定义1000个样本范围内的任意连杆。图5展示了训练集中几段文本描述的示例。

AAJ002	The main structure of the link is binary link; the main structure of the shaft is arc-shaped vertical slab; the radius of the shaft is large; the length of the shaft is medium; the thickness of the shaft is large; the cross section of the shaft is rectangle; the inner diameter of the first pivot hole is large; the outer diameter of the first piovt hole is large; the deepth of the first pivot hole is large; the inner diameter of the second pivot hole is large; the outer diameter of the second piovt hole is large; the deepth of the second pivot hole is large.
AAJ003	The main structure of the link is binary link; the main structure of the shaft is arc-shaped vertical slab; the radius of the shaft is medium; the length of the shaft is large; the thickness of the shaft is large; the cross section of the shaft is rectangle; the inner diameter of the first pivot hole is large; the outer diameter of the first piovt hole is large; the deepth of the first pivot hole is large; the inner diameter of the second pivot hole is large; the outer diameter of the second piovt hole is large; the deepth of the second pivot hole is large.
AAJ004	The main structure of the link is binary link; the main structure of the shaft is arc-shaped vertical slab; the radius of the shaft is medium; the length of the shaft is medium; the thickness of the shaft is large; the cross section of the shaft is rectangle; the inner diameter of the first pivot hole is large; the outer diameter of the first piovt hole is large; the deepth of the first pivot hole is large; the inner diameter of the second pivot hole is large; the outer diameter of the second piovt hole is large; the deepth of the second pivot hole is large.
AAJ005	The main structure of the link is binary link; the main structure of the shaft is arc-shaped vertical slab; the radius of the shaft is large; the length of the shaft is large; the thickness of the shaft is large; the cross section of the shaft is rectangle; the inner diameter of the first pivot hole is large; the outer diameter of the first piovt hole is large; the deepth of the first pivot hole is medium; the inner diameter of the second pivot hole is large; the outer diameter of the second piovt hole is large; the deepth of the second pivot hole is medium.
AAJ006	The main structure of the link is binary link; the main structure of the shaft is arc-shaped vertical slab; the radius of the shaft is large; the length of the shaft is medium; the thickness of the shaft is large; the cross section of the shaft is rectangle; the inner diameter of the first pivot hole is large; the outer diameter of the first piovt hole is large; the deepth of the first pivot hole is medium; the inner diameter of the second pivot hole is large; the outer diameter of the second piovt hole is large; the deepth of the second pivot hole is medium.

图5. 描述连杆结构特征的文本示例，这些特征可以表示为“实体–属性–属性值”的形式。

3.2.3. 制定由多名参与者构建训练集的指导规范

众所周知，高质量的训练集是成功训练深度学习模型的前提，但手动构建大型训练集既繁琐又耗时。因此，本文中的三维形状模型及其对应的文本描述由多名参与者共同准备，并在此过程中采用了一些规则，以避免出现可能损害利用该训练集训练的深度检索模型性能的错误。

（a）术语一致：在文本描述中使用不同术语表达相同特征，会增加训练所需的训练样本数量。因此，所有参与者都应使用相同的术语来表达这些特征。

（b）为三维模型使用相同的布局方向：提供文本描述时，连杆的三维模型应位于相同的坐标系中，并采用相同的布局方向。例如，为了准确描述轴中孔结构的方向（即X、Y或Z方向），所有连杆的枢轴孔都应朝向相同的方向。

(c) 枢轴孔的统一区分：对于枢轴孔内径相同的连杆，使用“左侧/右侧枢轴孔”作为统一术语，进一步描述两个枢轴孔的其他结构特征。对于枢轴孔内径不同的连杆（此类连杆通常用于内燃机），使用“较大/较小的枢轴孔”作为统一术语。

(d) 尺寸相关属性值的统一标准：如前所述，为了降低深度检索模型的训练难度，尺寸相关属性值简化为小/中/大，并以样本的平均值作为基准来确定这些值。

(e) 最终统一检查：构建训练集后，应由一名设计师检查多个参与者提供的三维模型及其对应的文本描述。

3.3 用于文本到形状检索的卷积神经网络与循环神经网络集成模型

在本小节中，建立了一个由循环神经网络和卷积神经网络构成的集成模型，用于学习三维体素模型与其对应文本描述之间的匹配关系。整个模型可归纳为三个部分，即基于卷积神经网络和循环神经网络的文本数据表示（由 [50,51]中的工作扩展而来）、基于三维卷积神经网络的三维体素数据表示，以及基于三元组损失的相似度计算。该模型的整体结构如图6所示。

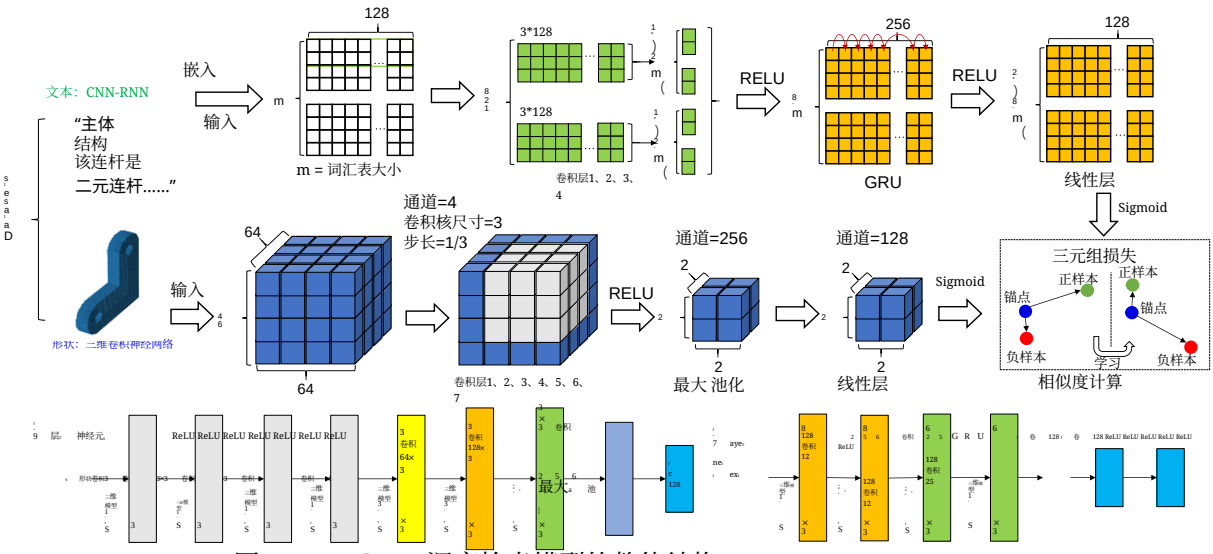


图6. text2shape深度检索模型的整体结构。

3.3.1. 使用卷积神经网络和循环网络集成模型表示文本描述的数据特征 N

采用卷积神经网络和循环神经网络集成网络，将描述连杆结构特征的文本数据表示为矩阵。集成网络的结构如图6右下方所示。可以看出，该网络包含四个卷积层，每个卷积层之后均接有一个 ReLU 激活函数，以避免过拟合。前三个卷积层的卷积核尺寸为 3 × 128，步长为 1，填充为 1。第四个卷积层的卷积核尺寸为 3 × 256，其步长和填充与前三个卷积层相同。为兼顾循环神经网络和卷积神经网络的优势，在第四个卷积层与全连接层之间添加了一个门控循环单元（GRU）层。经过第四个卷积层后，数据维度增至 256；GRU 层能够将数据转换为序列，使数据之间具有时间依赖关系。最后，在 GRU 层末端连接两个全连接层，最终输出的通道数为 128。

3.3.2. 使用基于卷积神经网络的模型表示三维体素模型的数据特征

采用三维卷积神经网络将三维体素数据表示为矩阵，其主体结构如图6左下方所示。该三维卷积神经网络包含七个卷积层，每个卷积层后均接一个 ReLU 激活函数，以避免过拟合。前四个卷积层的卷积核尺寸为 3 × 3 × 3，步长均设为 1，通道数设为 4。第五、第六和第七卷积层的卷积核尺寸为 3 × 3 × 3，步长设为 3，通道数分别为 64、128 和 256。

在第七卷积层之后接入一个卷积核为 2 × 2 × 2 的最大池化层进行下采样，其步长设为 1。最后，在池化层之后接入一个全连接层，最终输出的通道数为 128。

3.3.3. 计算　使用基于三元组损失的方法衡量文本描述与三维体素模型之间的相似度　od

文本到形状检索是一项从一维数据到三维数据的跨模态匹配任务。从图5中的示例可以看出，对于每条训练数据，文本描述了多个不同的结构特征。在此过程中，检索模型首先将文本描述表示为一个向量，并将三维形状表示为另一个向量。随后，在计算过程中，如果三维形状不包含文本所描述的结构特征，检索模型就会增大两个向量之间的距离，反之亦然。

为了提高text2shape检索模型的准确率，需要计算并整合两种距离，即表示文本的向量与表示三维形状模型的向量之间的距离，以及表示三维形状模型的向量与表示文本的向量之间的距离。对此，检索模型中所采用的损失函数如下所示。

$$Loss = Loss_{t2s} + \mu Loss_{s2t} \tag{1}$$

分别从第3.3.1节和第3.3.2节获得文本描述和三维形状模型的输出向量后，可应用三元组损失 [52] 来评估输入文本与三维模型之间的距离。三元组损失可根据锚点、正样本和负样本的值计算。计算过程会缩短锚点与正样本之间的距离，并增大正样本与负样本之间的距离。三元组损失可通过下式计算。

$$L = max\left(d(Anchor, Positive) - d(Anchor, Negative) + margin, 0\right) \tag{2}$$

三元组损失函数有三种计算方法，分别是简单三元组、困难三元组和半困难三元组。简单三元组无需进一步优化计算，当 L = 0 时应采用，且

$$d(anchor, positive) + margin < d(anchor, negative) \tag{3}$$

当锚点与正样本之间的距离较大时，应采用困难三元组，且

$$d(anchor, negative) < d(anchor, positive) \tag{4}$$

当满足公式（5）时，应采用半困难三元组。在这种情况下，锚点和负样本的向量彼此较为接近，但它们之间仍有一定间隔。

$$d\left(anchor, positive\right) < d(anchor, negative) < d\left(anchor, positive\right) + margin \tag{5}$$

最终，检索模型采用半困难三元组 。在训练过程中，计算表示文本和三维形状的各向量之间的距离，并将其与其他距离进行比较。在此基础上，损失值越小，表示三维形状模型与文本越相似，反之亦然。

4. 案例研究

为了验证第3节中工作的有效性，我们开发了一个基于 Python 的深度检索模型，并开展了一系列检索实验。在此过程中，我们采用正交实验对深度检索模型的超参数和网络结构进行了调优。

4.1. 运行环境

实验平台搭建在一台配备 Intel Xeon Silver 4216 CPU 和 GeForce RTX 3090 GPU 的深度学习工作站上。该工作站的运行环境如表1所示。基于该平台，我们使用第3.2节构建的数据集，对第3.3节构建的模型进行了训练。

表1. 运行环境。

项目	项目内容
操作系统	Microsoft Windows 10
Python 环境	Python 3.9
虚拟环境	Anaconda3-5.2.0
编程软件	PyCharm 2022.1.3

4.2. 训练集准备

为构建训练集而收集的原始样本包括一组文本描述及其对应的三维形状模型。原始文本描述的长度各不相同，因为这些描述由设计师根据自身对相应三维形状所包含结构特征的理解手动编写；由*Solidworks*生成的原始三维形状采用*STL* 文件格式。

基于原始样本，在使用样本训练检索模型之前，先进行了预处理流程。在此过程中，使用自主开发的代码将 STL 格式的三维形状转换为*NRRD* 格式的三维体素模型。采用*NRRD* 格式，是因为该格式能够存储维度、大小、空间方向、空间原点等信息，并且可以通过 Python 插件包方便地处理。每个三维体素模型均构建在一个大小为 16^3 的规则立方空间中。对于文本描述，则使用自主开发的代码将每个样本中的文本长度缩减至最多 256 个词。此外，选取出现次数超过两次的词作为训练文本表示模型的词汇表，并将所有文本描述以*CSV*格式存储。

预处理后，共制备了 1000 个样本，可用于训练过程。

4.3. 训练细节

4.3.1. 使用基于正交试验的方法调整深度检索模型

为了高效确定超参数和网络结构设置的最佳配置，包括批量大小、卷积层数、学习率和训练轮次，我们采用了基于正交试验的调优方法。

首先，进行了包含四个因素、每个因素三个水平的正交试验，如表 2 所示。结果表明，准确率主要与批量大小、卷积层数和学习率有关，因为改变这些因素的设置会显著影响准确率。此外，结果还表明，较大的批量大小、较高的卷积层数和较低的学习率能够带来更高的准确率。第一次正交试验中的最高准确率为 71.09%。

表 2。 第一次正交试验。

编号	批量大小 (A)	学习率 (B)	训练轮次 (C)	卷积层数 (D)	准确率 (E) /%
1	16	0.001	50	3	3.77
2	16	0.0001	200	4	64.77
3	16	0.00001	100	5	46.45
4	32	0.001	200	5	30.00
5	32	0.0001	100	3	55.62
6	32	0.00001	50	4	69.38
7	64	0.001	100	4	48.44
8	64	0.0001	50	5	71.09
9	64	0.00001	200	3	57.81
T_{1j}	114.99	82.21	144.24	117.2	
T_{2j}	155	191.48	150.51	182.59	447.33
T_{3j}	177.34	173.64	152.58	147.54	
R_j	62.35	109.27	8.34	65.39	
顺序	B > D > A > C				
最佳级别	A3	B2	C3	D2	
最佳组合	A3B2C3D2				

根据第一次正交试验的结果，设计了第二次实验，以扩大各因素的取值范围。第二次实验采用四个因素和四个水平，取值范围比第一次更大，如表3所示。共进行了16组实验；当批量大小为4、卷积层数 为7且学习率为0.000001时，最高准确率提高到了95%。

表3。 第二次正交试验。

编号	学习率 (A)	批量大小 (B)	卷积层数 (C)	训练轮次 (D)	准确率 (E) /%
1	0.001	4	4	50	64.44
2	0.001	16	5	100	48.86
3	0.001	32	6	200	61.25
4	0.001	64	7	500	42.19
5	0.0001	4	5	142	53.89
6	0.0001	16	4	500	42.61
7	0.0001	32	7	50	48.12
8	0.0001	64	6	100	58.59
9	0.00001	4	6	142	90.00
10	0.00001	16	7	200	58.52
11	0.00001	32	4	100	82.50
12	0.00001	64	5	50	60.16
13	0.000001	4	7	100	95.00
14	0.000001	16	6	50	85.23
15	0.000001	32	5	500	74.37
16	0.000001	64	4	200	69.53
T_{1j}	216.74	303.33	259.08	257.95	
T_{2i}	203.21	235.22	237.28	284.95	1035.26
T_{3j}	291.18	266.24	295.07	243.19	
T_{4j}	324.13	230.47	243.83	249.17	
R_j	120.92	72.86	57.79	41.76	
顺序	A > B > C > D				
最佳级别	A4	B1	C3	D2	

编号	学习率 **(A)**	批量大小 **(B)**	卷积层数 **(C)**	训练轮次 **(D)**	准确率 **(E)** /%
Best 组合	A4B1C3D2				

基于第二次实验的结果，开展了第三次正交实验，以进一步探索超参数和网络结构设置的最优配置。本次实验采用了四个因素和两个水平，如表4所示，旨在确保涵盖前几次实验中被忽略的因素配置。从结果可以看出，当批量大小为4、卷积层数为7、学习率为0.00001时，最高准确率达到98%。

表4。 第三次正交实验。

编号	学习率 **(A)**	批量大小 **(B)**	卷积层数 **(C)**	准确率 **(D)** /%
1	0.00001	4	6	91.67
2	0.00001	4	7	77.78
3	0.00001	16	6	98.30
4	0.00001	16	7	59.66
5	0.000001	4	6	96.67
6	0.000001	4	7	97.78
7	0.000001	16	6	94.89
8	0.000001	16	7	86.93
T_{1j}	327.41	363.9	381.53	703.68
T_{2j}	376.27	339.78	322.15	
R_j	48.86	24.12	59.38	
顺序	C > A > B			
最佳级别	A2	B1	C1	
Best 组合	A2B1C1			

4.3.2. 使用极差分析和方差分析验证实验结果

在本小节中，对正交实验的极差值和方差值进行分析，以确保实验结果未受到实验误差的影响 [53]。这里以第三次实验的结果进行示范分析。

在分析极差值时，可以从*MiniLab*软件中获得极值。由此可以看出，网络层这一因素的*Delta*值最大（如表5所示），这与表3中的结果一致，并且三个因素的顺序也相同。这里，*Delta*是一个表示响应水平平均值的参数，*Delta*值越大，表明其影响越强。

表5。响应水平平均值。

水平	**Lea**学习率 **(A)**	批量大小 **(B)**	**Con**卷积层数 **(C)**
1	81.85	90.97	95.38
2	94.07	84.94	80.54
Delta	12.22	6.03	14.84
顺序	2	3	1

借助图7，可以轻松分析各因素的平均值。坐标的横坐标表示各因素的水平，纵坐标表示平均值的平均值。从图中可以看出，当学习率处于第二水平、批量大小处于第一水平，且卷积层数处于第一水平时，平均值的平均值最大。

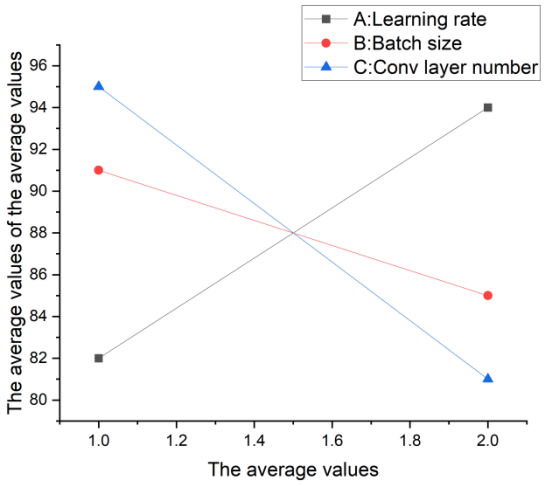


图7。 平均值的主效应图。

可以看出，根据上述分析结果，*A2B1C1*配置获得了最高的平均值。然而，仍需进行方差分析，以确保这三个因素对实验结果具有统计显著性。借助 *MiniLab*软件，可以获得自由度、调整平方和，、调整均方和F值（如表6所示）。根据F分布图可以看出，当显著性水平为0.25时，上述数值表明因素A和C会产生显著影响。

表6。 方差分析。

因素	自由度	调整平方和	调整均方	F值	p值
A	1	298.41	298.41	2.75	0.173
B	1	72.72	72.72	0.67	0.459
C	1	440.75	440.75	4.06	0.114
误差	4	434.70	108.68		
总计	7	1246.58			

最终，根据对极差值和方差值的分析，可以看出正交实验的结果是正确的。

4.4. 检索案例研究

本小节所述的案例研究测试了上述训练的检索模型能否识别出与输入文本描述最相似的三维形状模型。训练后的模型应能够识别出与输入文本描述损失值最小的八个三维形状模型。研究共进行了十多次检索实验，结果均令人满意。其中一项实验的输入和输出如图8所示。

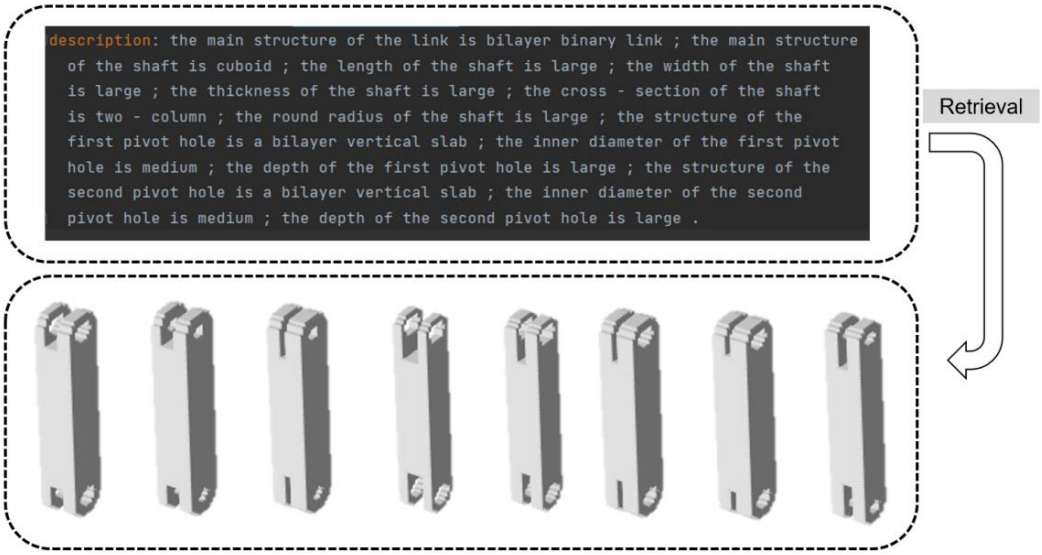


图8. 检索实验示例。

5. 讨论与结论

本文按照基于特征工程的关键结构特征识别、训练集构建和深度检索模型构建这几个步骤，开发了一个text2shape深度检索模型。本研究的贡献可从两个方面总结。

从方法论角度来看，text2shape深度检索模型能够从历史设计方案库中获取与新设计需求最相似的设计方案，从而为基于案例的推理情境下的再设计任务提供初始设计方案。此外，与常用的人工检索方法和基于简单推理算法的方法相比，基于text2shape深度检索模型的方法具有更高的检索效率和准确率。

从技术角度来看，本研究以连杆为检索目标，建立了一个特定的text2shape匹配训练集，并构建了一个由循环神经网络和卷积神经网络组成的集成模型，以拟合训练集的潜在空间。为避免过拟合问题，我们对检索模型的网络结构进行了修改。在三维卷积神经网络模型中增加了若干卷积层，并在每个卷积层后接入一个RELU函数。与计算机科学领域的研究不同，本文所使用的训练集是针对这一特定应用场景构建的，并对深度检索模型的结构和参数进行了修改，使其适配特定训练集中数据特征所蕴含的潜在空间。此外，我们采用了基于正交试验的方法进行模型调优，从而有助于快速确定适用于深度检索模型和数据集的超参数与网络设置的最佳配置。

值得一提的是，本文的工作实际上是我们尝试利用深度学习模型支持产品设计活动的第一步。之所以将结构特征表示为三元组数据（如第3.1节所示），主要是为了开展未来工作的下一步：届时将以 graph2shape 生成模型取代 text2shape检索模型，因为知识图谱能够直观地表示目标产品的整体结构，而生成模型（而非检索模型）则能够根据设计需求的输入图描述生成新的设计方案。模型生成的新设计方案也将不同于历史案例库中的现有设计方案。此外，我们还将在未来工作中，通过构建新的训练集以及与这些训练集相匹配的深度学习模型，考虑功能需求与三维形状之间的匹配关系。在新的训练集中，输入将是功能需求的文本或图描述，输出则是相应的三维形状。由此即可实现从原始客户需求到产品三维形状的检索/生成。此外，需要指出的是，仍需通过制作不同类型连杆的变体来获取更多训练样本

以提升检索模型的泛化能力。

参考文献

1. Das, R.; Godbole, A.; Naik, A.; Tower, E.; Jia, R.; Zaheer, M.; Hajishirzi, H.; McCallum, A. 基于子图的案例推理知识库问答。arXiv 2022, arXiv:2202.10610. 2. Li, S.; Chen, L. 面向快速重新设计的基于模式的推理：一种主动方法。Res. Eng. Des. 2010, 21, 25–42. 3. Brooks, B. 滑轮模型：风险决策的描述性模型。Saf. Sci. Monit. 2007, 11, 1–14. 4. Wang, Z.; Long, C.; Cong, G.; Ju, C. 利用深度表示学习高效检索体育比赛。载于KDD ‘19论文集：第25届ACM SIGKDD国际知识发现与数据挖掘会议论文集，美国阿拉斯加州安克雷奇，2019年8月4–8日。5. Malkiel, I.; Nagler, A.; Mrejen, M.; Arieli, U.; Wolf, L.; Suchowski, H. 用于纳米光子结构设计与检索的深度学习。arXiv 2017, arXiv:1702.07949. 6. Tangelder, J.W.H.; Veltkamp, R.C. 基于内容的三维形状检索方法综述。Multide. Tools Appl. 2008, 39, 441–471. 7. Yang, Y.B.; Lin, H.; Zhu, Q. 基于内容的三维模型检索：综述。Chin. J. Comput. 2004, 27, 1297–1310. 8. Yuan, J.F.; Hameed, A.R.; Li, B.; Lu, Y.; Schreck, T.; Bai, S.; Bai, X.; Bui, N.-M.; Do, M.N.; Do, T.-L.; 等. 三维场景形状检索方法比较。Comput. Vis. Image Underst. 2020, 201, 103070. 9. Chen, K.; Choy, C.B.; Savva, M.; Chang, A.X.; Funkhouser, T.; Savarese, S. Text2Shape：通过学习联合嵌入，根据自然语言生成形状。载于Computer Vision-ACCV 2018论文集，澳大利亚珀斯，2018年12月2–6日。10. Hu, N.; Zhou, H.; Liu, A.A.; Huang, X.; Zhang, S.; Jin, G.; Guo, J.; Li, X. 基于二维图像的三维形状检索的协同分布对齐。《视觉学杂志》Commun. Image R. 2022, 83, 103426. 11. Angrish, A.; Bharadwaj, A.; Starly, B. MVCNN++：利用多视图卷积神经网络进行计算机辅助设计模型形状分类与检索。J. Comput. Inf. Sci. Eng. 2021, 21, 011001. 12. Kim, E.Y.; Shin, S.Y.; Lee, S.; Lee, K.J.; Lee, K.H.; Lee, K.M. 采用共享二维卷积核的三平面卷积，用于三维分类与形状检索。Comput. Vis. Image Underst. 2020, 193, 102901. 13. Song, D.; Li, T.B.; Li, W.H.; Nie, W.Z.; Liu, W.; Liu, A.A. 通用跨域三维模型检索。《IEEE Multimed.》2021, 23, 2721–2731. 14. Qiao, H.; Wu, Q.Y.; Yu, S.L.; Du, J.; Xiang, Y. 一种基于装配信息的三维装配模型检索方法。Assem. Autom. 2019, 39, 556–565. 15. Sun, X.Y.; Wu, J.J.; Zhang, X.M.; Zhang, X.M.; Zhang, Z.T.; Zhang, C.K.; Xue, T.F.; Tenenbaum, J.B.; Freeman, W.T. Pix3D：用于单图像三维形状建模的数据集与方法。载于2018年IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，美国犹他州盐湖城，2018年6月18–23日。16. Qin, F.W.; Qiu, S.; Gao, S.M.; Bai, J. 基于草图和无监督变分自编码器的三维CAD模型检索。Adv. Eng. Inform. 2022, 51, 101427. 17. Manda, B.; Dhayarkar, S.; Mitheran, S.; Viekash, V.K.; Muthuganapathy, R. “CADSketchNet”——用于借助深度神经网络进行三维CAD模型检索的标注草图数据集。Comput. Graph. 2021, 99, 100–113. 18. Kuo, W.; Angelova, A.; Lin, T.Y.; Dai, A. Mask2CAD：通过学习分割与检索实现三维形状预测。载于Computer Vision-ECCV论文集，英国格拉斯哥，2020年8月23–28日。19. Li, Z.; Xu, J.Y.; Zhao, Y.; Li, W.H.; Nie, W.Z. MPAN：面向基于点云的三维形状检索的多部件注意力网络。《IEEE Access》2020, 8, 157322–157332. 20. Kim, H.; Yeo, C.; Lee, I.D.; Mun, D. 基于深度学习的管道组件目录检索，用于工厂三维CAD模型重建。Comput. Ind. 2020, 123, 103320. 21. Wang, Z.; Lin, H. 基于拉普拉斯算子和联合贝叶斯模型的三维形状检索。Vis. Inform. 2020, 4, 69–76.

22. Son, H.; Lee, S.H. 使用深度神经网络通过对象部件分割创建局部本体，在单类别几何中进行三维模型检索。《机械科学与技术杂志》2021, 35, 5071–5079. 23. Udaiyar, P. 使用深度神经网络进行跨模态数据检索与生成。硕士论文，RIT学术成果库，罗切斯特理工学院，美国纽约州纽约市，2020. 24. Huang, L. 前注意形状特征空间。《视觉学杂志》2020, 20, 10. 25. Wang, Y.; Zhang, H.; Shi, G.; Liu, Z.; Zhou, Q. 一种结合互注意力的文本增强知识图谱表示学习模型。《IEEE Access》2020, 8, 52895–52905. 26. Jiang, Z.; Gao, S.; Chen, L. 基于深度学习和主题信息的文本表示方法研究。《Computing》2020, 102, 623–642. 27. Yan, R.; Peng, L.; Xiao, S.; Yao, G. 用于场景文本识别的原始表示学习。载于：2021年IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，美国田纳西州纳什维尔，2021年6月20–25日。28. Luo, L.X. 结合LDA文本表示与GRU-CNN的网络文本情感分析方法。《个人与普适计算》2019, 23, 405–412. 29. Sinoara, R.A.; Jose, C.C.; Rossi, R.G.; Navigli, R.; Rezende, S.O. 用于文本分类的知识增强文档嵌入。《Knowledge-Based Systems》2019, 163, 955–971. 30. Song, R.; Gao, S.; Yu, Z.; Zhang, Y.; Zhou, G. Case2vec：用于案例文本嵌入表示的联合变分自编码器。《国际机器学习与控制论杂志》2021, 12, 2517–2528. 31. Guo, S.; Yao, N. 用于文本分类的多义词感知向量表示。《IEEE Access》2020, 8, 135686–135699. 32. Hou, W.; Liu, Q.; Cao, L. 基于命名实体、概念和知识的认知方面短文本表示。《应用科学》2020, 10, 4893. 33. Feng, Y.; Feng, Y.; You, H.; Zhao, X.; Gao, Y. MeshNet：用于三维形状表示的网格神经网络。载于：AAAI'19：AAAI人工智能会议论文集，美国夏威夷州檀香山，2019年1月27日–2月1日。34. Gao, Z.; Yan, J.; Zhai, G.; Zhang, J.; Yang, Y.; Yang, X. 学习局部邻域结构以实现鲁棒三维形状表示。载于：AAAI人工智能会议论文集，美国加利福尼亚州帕洛阿尔托，2022年2月22日–3月1日。35. Gal, R.; Bermano, A.; Zhang, H.; Daniel, C.O. MRGAN：通过无监督部件解耦进行多根三维形状表示学习。载于：2021年IEEE/CVF国际计算机视觉会议研讨会（ICCVW 2021）论文集，加拿大不列颠哥伦比亚省蒙特利尔，2021年10月11–17日。36. Jiang, C.; Sud, A.; Makadia, A.; Huang, J.; Nießner, M.; Funkhouser, T. 用于三维场景的局部隐式网格表示。载于：2020年IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，美国华盛顿州西雅图，2020年6月13–19日。37. Zheng, Z.; Yu, T.; Dai, Q.; Liu, Y. 用于三维形状表示的深度隐式模板。载于：2021年IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，美国田纳西州纳什维尔，2021年6月20–25日。38. Chen, Z.; Zhang, Y.; Genova, K.; Fanello, S.; Bouaziz, S.; Häne, C.; Du, R.; Keskin, C.; Funkhouser, T.; Tang, D. 用于三维形状表示的多分辨率深度隐式函数。载于：2021年IEEE/CVF国际计算机视觉会议（ICCV 2021）论文集，加拿大魁北克省蒙特利尔，2021年10月10–17日。39. Tristan, A.A.; Stavros, T.; Sven, D.; Allan, J. 使用概率有向距离场表示三维形状。arXiv 2021, arXiv:2112.05300. 40. Wei, X.; Gong, Y.; Wang, F.; Sun, X.; Sun, J. 学习规范视图表示，以利用任意视图进行三维形状识别。载于：2021年IEEE/CVF国际计算机视觉会议（ICCV 2021）论文集，加拿大魁北克省蒙特利尔，2021年10月10–17日。

41. Tretschk, E.; Tewari, A.; Golynanik, V.; Zollhöfer, M.; Stoll, C.; Theobalt, C. PatchNets：基于补丁的通用深度隐式三维形状表示。载于：计算机视觉-ECCV 2020论文集，英国格拉斯哥，2020年8月23–28日。42. Hao, L.; Wang, H. 用于三维局部形状实值和二值特征表示的几何特征统计直方图。《图像与视觉计算》2022, 117, 104339. 43. Han, Z.; Shang, M.; Wang, X.; Liu, Y.; Zwicker, M. Y2Seq2Seq：通过联合重建和预测视图与词序列，实现三维形状与文本的跨模态表示学习。载于：2019年AAAI人工智能会议论文集，美国夏威夷州檀香山，2019年1月27日–2月1日。44. Ruan, Y.; Lee, H.; Zhang, K.; Chang, A.X. TriCoLo：用于细粒度文本到形状检索的三模态对比损失。arXiv 2022, arXiv:2201.07366. 45. Wu, Y.; Wang, S.; Huang, Q. 用于跨模态检索的多模态语义自编码器。《神经计算》2019, 331, 165–175. 46. Tang, C.; Yang, X.; Wu, B.; Han, Z.; Chang, Y. Part2Word：通过匹配部件与词语，学习点云和文本的联合嵌入。arXiv 2021, arXiv:2107.01872. 47. Li, D.; Wang, Y.; Zhang, Y.; Mu, X. 基于语义树的三维模型快速检索方法。载于：2018年虚拟现实与智能系统国际会议（ICVRIS）论文集，中国长沙，2018年8月10–11日。48. França, T.; Braga, A.M.B.; Ayala, H.V.H. 通过特征工程应对稀疏辨识中的噪声数据。《专家系统与应用》2022, 188, 115995. 49. Kanjilal, R.; Uysal, I. 人类活动识别的未来：深度学习还是特征工程?《神经处理快报》2021, 53, 561–579. 50. Reed, S.; Akata, Z.; Yan, X.; Logeswaran, L.; Schiele, B.; Lee, H. 生成对抗式文本到图像合成。载于：第33届国际机器学习会议论文集，美国纽约州纽约市，2016年6月19–24日。51. Maxim0815 Text2shape [源代码], 2020。在线获取：https://github.com/maxim0815/text2shape（访问日期：2022年3月17日）。52. Schroff, F.; Kalenichenko, D.; Philbin, J. FaceNet：用于人脸识别和聚类的统一嵌入。载于：2015年IEEE计算机视觉与模式识别会议（CVPR）论文集，美国马萨诸塞州波士顿，2015年6月7–12日。53. Wang, C.; Wang, C.; Pan, H.; Yue, Y. 基于正交试验法的结构参数对燃烧器低氮性能的影响。《热工程案例研究》2022, 39, 102404.