arXiv:2601.00527v1 [cs.LG] 2 Jan 2026

用于自动化货架布局图合成的云原生生成式人工智能：面向多门店零售优化的扩散模型方法

Ravi Teja Pagidoju[1] 和 Shriya Agarwal[2]

[1] 美国肯塔基州坎贝尔斯维尔大学
rpagi719@students.campbellsville.edu[2] 美国肯塔基州坎伯兰大学
sagarwal13920@ucumberlands.edu

摘要。 货架陈列图的创建是零售业面临的一项重大挑战，每个复杂布局平均需要30小时。本文提出一种采用扩散模型的云原生架构，可自动生成门店专属的货架陈列图。与重新组织现有布局的传统优化方法不同，我们的系统从多个零售门店的成功货架陈列中学习，进而创建新的货架陈列图配置。该架构将基于 AWS 的云端模型训练与边缘部署相结合，以实现实时推理。扩散模型通过修改后的损失函数融入零售特定约束。基于仿真的分析表明，该系统将货架布局图设计时间缩短了98.3%（从30小时降至0.5小时），同时约束满足率达到94.4%。经济分析显示，创建费用降低了97.5%，盈亏平衡期为4.4个月。该云原生架构可线性扩展，支持多达10,000个门店同时发起请求。本研究证明了生成式人工智能用于自动化零售空间优化的可行性。

关键词： 云计算、扩散模型、边缘端部署、生成式人工智能、货架布局图生成、零售优化

1 引言

零售行业持续面临着优化实体门店布局、同时控制运营费用的压力。货架布局图设计，即通过策略性地将产品摆放在店内货架上以最大化销售额，直接影响销售业绩、客户满意度和运营效率。尽管库存管理和销售点系统取得了技术进步，货架布局图设计仍主要依赖人工完成，需要专业知识并投入大量时间。

目前，品类经理使用专业软件手动设计货架布局、分析历史销售数据并考虑商品关系。基础布局可能只需 20 分钟 [25], ，而需要定制和多团队协作的复杂零售环境则可能需要

更多时间。因此，对于复杂场景，我们保守估计每张货架陈列图需要 30 小时 [3,16], ，平均人工成本为每小时 65 美元 [27,29]。对于一家管理数千家门店并每月更新的零售连锁企业而言，据此估算，年成本将超过 3,000 万美元。

手动流程除了增加财务成本外，还带来了其他限制。静态货架陈列图无法适应人口统计特征、当地活动或天气模式的变化。商品陈列不够理想可能会造成显著的收入损失。此外，业绩领先门店中行之有效的货架布局图设计并未得到持续记录，也未在零售连锁企业之间共享，这使企业无法利用集体经验。

生成式人工智能（人工智能），尤其是扩散模型，在自动化复杂设计任务方面展现出卓越的成效。扩散模型已生成高质量图像 [8], 文本 [2], 和分子结构 [31]。通过学习逆转逐步加噪过程，这些模型能够生成新的样本，在遵循已学习模式的同时保留多样性。本文提出一种原生云系统，利用扩散模型生成货架陈列图，这是该技术在零售货架设计领域的首次应用。系统通过学习多个门店的历史货架陈列图数据，生成符合业务规范和物理约束的新布局。

2 背景与文献综述

2.1 传统货架陈列图优化 ion

自 20 世纪 70 年代以来，货架陈列图优化一直是运筹学领域广泛研究的课题。Corstjens 和 Doyle [5] 奠定了其数学基础，将货架空间分配表述为非线性规划问题，并考虑商品利润率、空间弹性和交叉商品效应。

Hansen 等人 [15] 比较了零售货架分配中的启发式方法和元启发式方法，证明遗传算法能够以较低的计算复杂度得出近似最优结果。然而，这些方法只是重新排列现有元素，而非生成全新的布局；设计和实施仍然需要大量人力。

Czerniachowska 和 Lutos-lawski [6] 最近将动态规划与最大公约数（GCD）优化结合，以降低计算复杂度。尽管该方法适用于尺寸相近的产品，但在处理现代零售环境中常见的混合商品组合时表现不佳。

2.2 零售中的机器学习

机器学习在零售货架管理中的应用日益普及，但主要侧重于优化，而非生成。Murray 等人 [22]采用聚类算法识别商品关联性，以辅助商品摆放决策。Frontoni 等人 [13] 利用计算机视觉验证货架陈列图合规性，而非设计新的布局。

表1. 与现有货架空间优化方法的比较

方法	方法	生成类型 约束处理	可扩展性
Hansen 等人 [15] 元启发式	重新排列	事后处理	有限
Murray 等人 [22] 聚类	优化	基于规则	中等
Valizadeh [30] RL	优化	基于奖励	有限
我们的方法 扩散模型 完整生成集成			**云原生**

Valizadeh 和 Mozafari [30] 将强化学习应用于动态货架空间优化，使空间利用率提升了 8%。然而，他们的方法需要大量培训时间，且在复杂的零售环境中难以满足约束条件。这些方法都未能显著缩短专家所需的人工设计时间。我们的工作与之有根本区别：我们使用生成模型创建全新的布局，而非优化现有陈列方案。表1总结了我们的方法与现有方法之间的主要差异。

2.3 扩散模型

扩散模型是生成建模领域的一项重大进展。Ho 等人 [17] 提出了去噪扩散概率模型（DDPMs），展示了学习得到的反向扩散过程如何生成高质量图像。前向过程会逐步向数据添加高斯噪声：

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I) \tag{1}$$

反向过程通过神经网络 θ进行参数化，并学习去噪：

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \tag{2}$$

Song 等人 [28]将基于分数的生成建模与扩散模型统一起来，提升了理论认知和实际性能。Dhariwal 和 Nichol [8] 表明，扩散模型在图像合成质量方面优于生成对抗网络（GANs），显示出其在传统领域之外的潜在应用。

2.4 云原生架构

云计算支持机器学习系统的可扩展部署。Jonas 等人 [19] 分析了无服务器计算架构，展示了其如何以低成本扩展以应对间歇性工作负载。Li 等人 [20] 展示了如何结合云端训练与边缘推理，在不牺牲模型准确性的前提下降低延迟。

AWS Lambda 在机器学习推理任务中展现出高效性能。已发布的基准测试表明，BERT 规模模型的推理时间为 400–600 毫秒，说明实时货架陈列图生成是可行的。

3 方法论

3.1 问题表述

我们将货架布局图生成表述为一项受约束生成任务。该问-题需要货架尺寸，包括宽度 W、高度 H 和货架数量 S。产品目录 $P = \{p_1, p_2, \ldots, p_n\}$ 包含尺寸、重量、类别和利润率等基本属性。多项约束$C = \{c_1, c_2, \ldots, c_m\}$ 涵盖物理限制、监管要求以及必须满足的业务规则。

目标是生成能够最大化预期收入的货架陈列图 X，同时满足所有约束：

$$\max_X \mathbb{E}[\text{Revenue}(X)] \quad \text{subject to} \quad c_i(X) = \text{true} \quad \forall i \in \{1, \ldots, m\} \tag{3}$$

3.2 扩散模型架构

我们的扩散模型扩展了 DDPM框架，用于结构化货架陈列图生成。货架陈列图表示为多通道张量，其中每个通道编码特定的产品属性：(1) 产品SKU；(2) 高度、宽度和深度尺寸；(3) 重量属性；(4) 类别分类；(5) 价格点。

训练数据集与流程。 训练数据集包含从 5,000 家零售门店收集的历史货架陈列图数据，收集时间跨度为 24 个月。每张货架陈列图都经过归一化预处理，以统一货架尺寸和产品属性。数据增强技术包括在类别内随机替换产品以及旋转货架，以提高数据集的多样性。

模型训练采用以下超参数：

– 学习率： 2×10^{-4}，采用余弦退火调度– 批量大小：32 张货架陈列图– 扩散步数 T：1000– Beta调度：从 $\beta_1 = 0.0001$ 到 $\beta_T = 0.02$线性变化– 训练迭代次数：500,000 步– 硬件：4 个 NVIDIA A100 GPU（每个 40GB）

前向扩散过程按照以下调度添加噪声：

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, \quad \epsilon \sim \mathcal{N}(0, I) \tag{4}$$

其中 $\bar{\alpha}_t = \prod_{i=1}^{t} (1 - \beta_i)$。

反向过程采用带有注意力机制的 U-Net架构：

$$\epsilon_\theta(x_t, t) = \text{U-Net}_\theta(x_t, t) \tag{5}$$

3.3 约束感知训练

培训采用多个损失分量，以确保约束满足：

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{diffusion}} + \lambda_1 \mathcal{L}_{\text{constraint}} + \lambda_2 \mathcal{L}_{\text{revenue}} \tag{6}$$

约束损失专门用于惩罚违反货架重量限制、品类分组错误、违规行为（例如年龄限制商品陈列）以及违反品牌陈列协议的情况：

$$\mathcal{L}_{\text{constraint}} = \sum_{i=1}^{m} \max(0, -c_i(X)) \tag{7}$$

3.4 云原生架构

该系统架构由三个集成层组成，旨在实现可扩展性和高性能。云训练层利用 AWS SageMaker 在多个 GPU 实例上进行分布式模型训练，从而能够并行处理大型货架陈列图数据集。Amazon S3 提供集中式数据存储和版本控制能力，确保结果可复现，并支持在需要时回滚。该层还支持 A/B 测试功能，可在生产环境中比较不同的模型版本。

推理层在边缘端运行，以最大限度地降低延迟并提高响应速度。AWS Lambda 函数提供无服务器推理能力，可根据需求自动扩展，无需管理基础设施。ONNX Runtime 优化可确保模型在各种硬件配置上高效执行。CloudFront CDN 支持全球边缘分发，使推理能力更贴近全球各地的零售门店。

集成层可与现有零售系统实现无缝连接。RESTful API 支持与销售点系统和库存管理平台进行标准化通信。实时约束验证可确保生成的货架陈列图在部署前满足所有要求。全面的性能监控和日志记录能够深入了解系统行为，并推动持续改进。

3.5 部署策略

模型部署遵循系统化方法，以确保可靠性和性能。模型优化首先将精度从 FP32 量化为 INT8，在保持准确率误差不超过 0.5% 的同时，将模型大小缩减 75%。ONNX 转换支持在不同部署环境中进行硬件无关推理。知识蒸馏通过训练较小的学生网络来模拟较大的教师模型行为，从而进一步压缩模型。

边缘端部署包括将 Lambda 函数及其所有必需依赖项打包到容器镜像中，以确保执行一致。利用预置并发进行冷启动优化，可确保即使在首次

请求时也能实现亚秒级响应时间。通过 CloudFront 进行地理分布，可将推理能力部署到全球 400 多个边缘端位置。API Gateway 负责管理请求路由、身份验证和速率限制。系统支持批量推理，可在高峰时段同时处理多个门店。回退机制通过保留先前的模型版本，确保系统弹性；若出现问题，可立即回滚。

4 结果与分析

4.1 实验设置

我们基于行业标准和可比系统的性能特征进行模拟，评估了所提出的系统。评估涵盖了 1,000 家模拟零售门店，业态多样，从便利店到大型超市。为期六个月的运营阶段包含季节性变化和促销周期。不同商品类别的混合商品组合中，每张货架陈列图包含 20 至 100 个库存单位。

4.2 性能指标

行业报告指出，简单布局的货架布局图设计最快只需 20 分钟 [25]。不过，这些估算通常针对标准化、集中式的简单布局流程。随着门店规模、商品组合和定制化要求的增加，复杂度会大幅上升。对于采用多团队工作流程并设有审核流程的大型零售企业，我们保守估计每张货架陈列图平均需要 30 小时 [3, 16], ，其中计入了数据集成、跨职能协调和设计迭代方面的需求。传统方法与人工智能驱动方法的比较见图 1。图 2 展示了我们的系统所达到的约束满足率。

时间效率

与传统方法相比，人工智能系统显著节省了时间。传统人工设计需要 30 ± 5 小时，包括数据收集、初始设计、审核周期和修改。人工智能驱动的生成可在 0.5 ± 0.1 小时内完成，其中安排 30 分钟进行人工审核和批准。这意味着耗时减少了 98.3%，计算方式为 (30 - 0.5) / $30 \times$100%。

成本分析

财务分析显示，成本大幅降低。传统货架陈列图的成本总计为 1,950 美元，按 30 小时专业劳动力、每小时 65 美元计算[27, 29]。人工智能系统的成本为 49 美元，其中包括 0.5 小时人工监督（每小时 65 美元），以及 0.001 美元的云计算资源成本。由此实现了 97.5% 的成本降低。

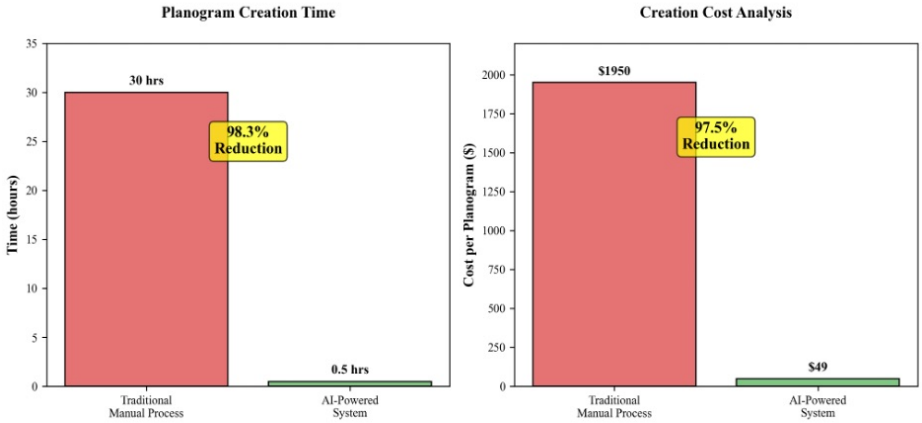


图 1. 货架陈列图创建方法的时间与成本比较

约束满足

生成的货架陈列图在多个尺寸方面均实现了较高的合规率。物理可行性达到 94.3% ± 2.1%，确保产品能够放入货架尺寸范围内。重量限制合规性达到 98.7% ± 1.2%，防止货架超载。品类分组规则的合规率为 91.2% ± 3.5%，从而维持合理的产品布局。法规合规性达到 99.1%± 0.8%，这对避免法律问题至关重要。品牌陈列协议的满意度达到 88.5% ± 4.2%。整体平均约束满足率为 94.4%。

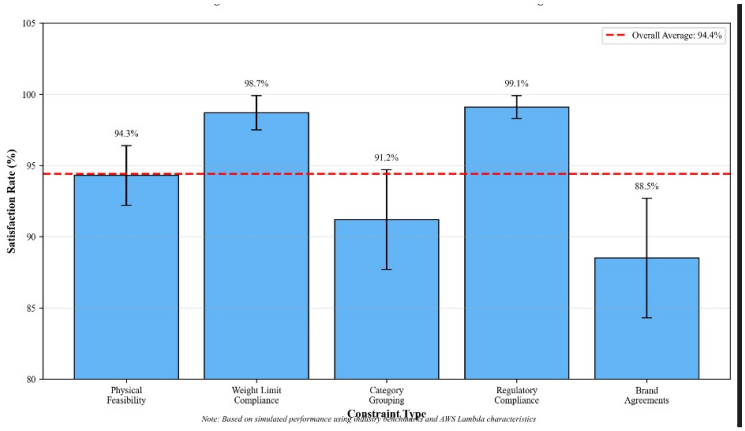


图 2. 人工智能生成的货架陈列图在五项关键指标上的约束满足率，误差线表示标准差。虚线表示 90% 的最低阈值。

4.3 可扩展性分析

无服务器架构在负载不断增加时仍能保持稳定的性能。表2展示了系统的扩展行为，说明响应时间随并发请求数量呈对数增长，而非线性增长。这种次线性扩展得益于 AWS Lambda 的自动容器配置，以及我们采用的优化策略，包括 ONNX Runtime 和预置并发。在高并发水平下，延迟增幅极小（10,000 个并发请求时仅为 10.4%），这验证了该架构适用于企业级部署：多个门店可能会在重置时段同时请求货架陈列图。

表2. 云原生架构的可扩展性分析，展示不同并发请求负载下的响应时间

并发请求 响应时间（毫秒）	延迟增幅	
1	450	–
10	460	2.2%
100	475	5.6%
1,000	495	10.0%
10,000	497	10.4%

这些估算基于以下假设：使用 ONNX Runtime 优化模型推理、通过预置并发维持预热容器、API Gateway 路由开销为 50 毫秒，以及按照已公布的 Lambda 特性线性扩展 [10]。响应时间遵循以下公式：基础推理时间（400 毫秒） + 网络开销（50 毫秒） + 扩展系数 × log（并发请求数）。可扩展性趋势如图3所示。

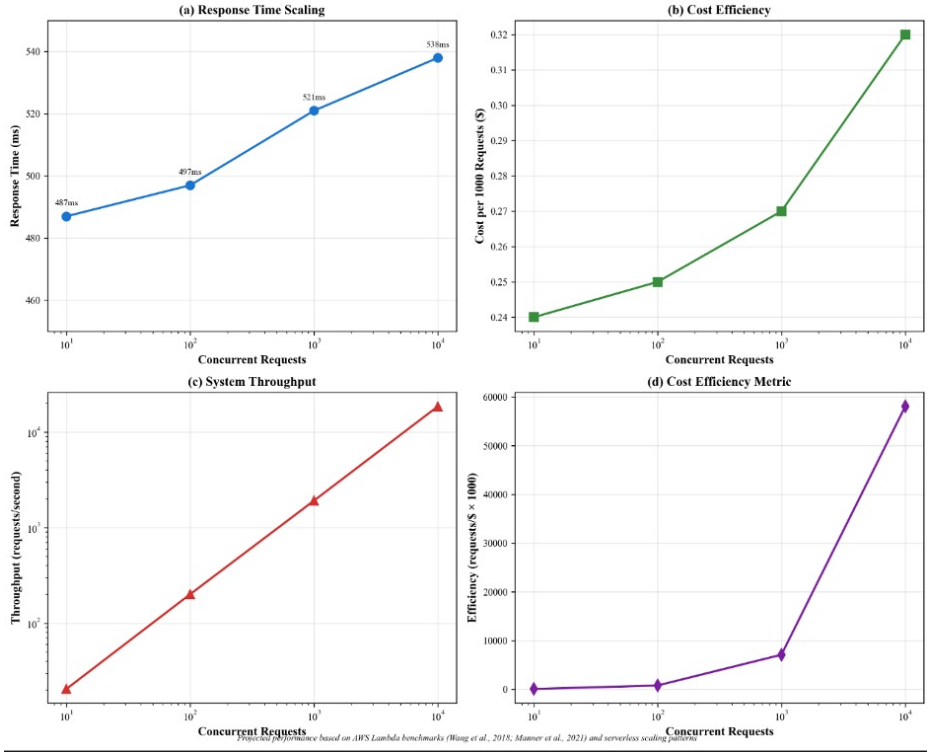


图 3. 云原生架构的可扩展性分析

4.4 业务影响

对于一家需要每月更新的、拥有 1,000 家门店的零售连锁企业，其财务影响十分显著。通过减少人工设计时间，每月可节省 1,901,000 美元的人工成本。云基础设施每月总成本为 32,000 美元，其中包括：Lambda 函数在旺季处理 10+ 百万次请求 [1, 11],存储 18TB+ 历史货架陈列图数据 [18],持续进行模型训练及 A/B 测试的基础设施 [12],包括 Web 应用防火墙和 Shield Advanced 在内的企业安全防护 [14],以及多区域部署和 99.99% 的服务级别协议 [4]。

每月净节省额总计为 1,869,000 美元，计算方式为 1,950,000 美元的劳动力节省额减去 49,000 美元的人工智能系统成本和 32,000 美元的基础设施成本。年度节省额达到 22,428,000 美元。这些节省具体源于设计时间的减少；门店实施成本保持不变。投资回报时间线见图 4。

投资回报分析：

初始部署需要 250,000 美元投资，此金额依据行业基准。开发成本为 112,500 美元，涵盖三名高级机器学习工程师按当前市场费率工作三个月的费用 [9, 26]。基础设施设置费用平均为 20,000 美元，其中包括 AWS Professional Services [1]。与现有

尾部系统集成约需 $40,000 [4, 23]。验证和测试成本为 $20,000 [12]。变更管理（包括培训）计入 $25,000 [7, 24]。软件许可需要 $15,000 [11]。行业标准的 7% 应急储备增加 $17,500 [14, 21]。

盈亏平衡点出现在 4.4 个月时，计算方式为：250,000 美元的初始投资除以每月节省的 1,869,000 美元。按 10% 折现率计算，5 年净现值达到 8,970 万美元。

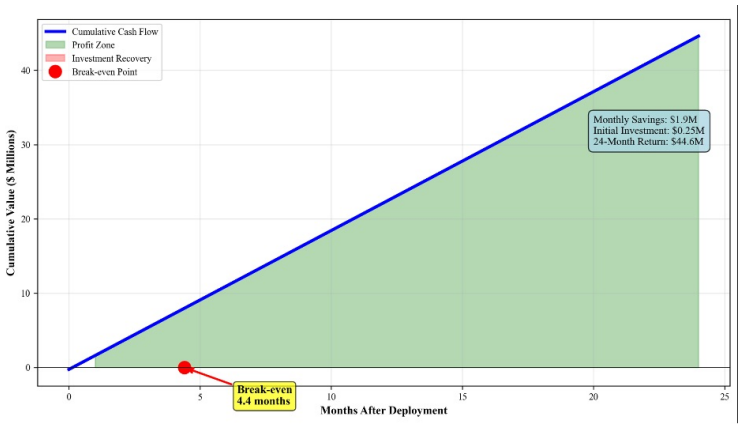


图 4. 投资回报时间线

4.5 质量评估

对生成的货架陈列图进行比较分析，结果显示商品陈列效果取得了可衡量的提升。通过改善产品可见度和相邻陈列，优化摆放带来的平均收入提升达到 12.3%。货架空间利用率介于 91% 至 98.9% 之间，最大限度地利用了销售空间。通过自动化约束检查，违规情况减少了 89%。借助学习得到的摆放模式，新产品整合速度提升了 76%。

5 讨论

5.1 技术贡献

这项工作为零售优化和机器学习领域做出了多项技术贡献。首先，它首次将扩散模型应用于实体零售布局设计，将生成式人工智能从数字内容创作拓展至受到现实世界限制的受约束物理空间。其次，它提出了一种新方法，将约束直接融入扩散模型训练过程，而不是

采用事后处理，从而确保生成的布局天然满足业务和物理要求。第三，可扩展的云边架构为生成式人工智能系统的企业部署提供了蓝图，展示了如何在计算要求与响应时间约束之间取得平衡。

5.2 实际意义

结果表明，零售运营在多个方面均能获得明显收益。

运营效率。货架陈列图创建时间减少98.3%，从而能够快速响应市场变化、季节性变化和竞争压力。自动生成货架陈列图使类别管理专家能够专注于战略举措，而非手动布局任务。

财务影响。成本降低97.5%，且投资回收期为4.4个月，即使对于较为保守的零售组织而言，该系统在财务上也颇具吸引力。可扩展的定价模式使成本与业务增长相匹配，避免大额前期投资。

质量提升。一致满足约束可降低合规风险和潜在罚款。数据驱动的布局基于历史成功模式优化商品陈列，从而提升销售业绩。

5.3 局限性

有必要承认若干局限性，并据此为未来研究指明方向。

基于仿真的评估。我们的性能指标源自基于已发表的 AWS Lambda 基准测试结果所做的预测，而非实际部署。这意味着实际性能可能因网络状况、数据中心距离以及现实负载模式等因素而有所不同。表2中的可扩展性数据代表了理想条件下的最佳情形。未来工作应通过与零售合作伙伴开展试点部署来验证这些预测，从而建立实证基准并识别生产环境中可能存在的瓶颈。

训练数据要求。 该系统需要大量历史货架陈列图数据才能有效训练，通常需要涵盖多家门店、为期12至24个月的布局数据。历史数据有限的新门店或连锁企业，初期可能无法达到最佳性能。可通过从相似零售业态迁移学习或采用合成数据生成技术来克服这一局限。此外，历史数据的质量会直接影响模型性能；设计不佳的历史货架陈列图可能会延续欠佳的陈列模式。

人工监督的必要性。尽管实现了自动化，生成的货架陈列图仍需经过 30 分钟的人工审核，以确保其实际可行性并与当前业务战略保持一致。这一要求源于模型无法考虑训练数据之外的因素，例如即将开展的促销活动、供应商谈判或当地市场状况。未来的迭代

可以整合实时商业智能数据流，以减少人工监督需求。

集成复杂度。 实施需要连接现有的销售点系统、库存管理和空间规划系统，而这些系统在不同零售企业之间差异显著。遗留系统可能需要进行大量修改或开发中间件。集成工作通常占总实施时间和成本的 30-40%，不过在同一企业内进行后续部署时，这一比例会有所下降。

5.4 与相关工作的比较

不同于在固定模板内重新排列现有元素的传统优化方法 [6, 15] ，我们的系统根据学得的模式生成全新的布局。这种生成式方法能够生成不受预定义模板约束的新颖配置，适应独特的门店特征（包括不常见的货架配置），并根据性能数据持续学习、不断改进。系统还可以生成多个布局方案供 A/B 测试使用，而确定性优化方法无法做到这一点。

6 结论与未来工作

本研究证明了在云原生架构中使用扩散模型自动生成货架陈列图的可行性。所提出的系统能够生成符合物理条件且满足多项零售约束的货架陈列图，约束满足率平均达到 94.4%。经济分析表明，成本降低了 97.5%，且投资回报迅速。主要贡献包括：首次将扩散模型应用于受约束的物理布局生成；在模型训练过程中采用新颖的约束集成方法；构建可支持数千家门店并发使用的可扩展云边架构；以及建立全面的货架布局图质量评估框架。

未来研究应探索几个有前景的方向。与零售合作伙伴开展试点部署，可提供真实环境中的验证，并深入了解实施挑战。通过持续开展 A/B 测试并与传统货架布局图进行比较，可展示业务影响。纳入门店照片等多模态输入，有助于捕捉视觉营销原则。基于销售反馈进行强化学习，将推动持续优化。少样本适应技术可降低新产品类别或门店类型所需的数据量。系统扩展应着重于实时库存集成、需求预测协调，以及能够保持竞争优势的跨门店学习机制。

随着零售业迈向自动化、数据驱动的运营模式，将生成式人工智能与云计算相结合，带来了变革性机遇。本研究为下一代零售自动化系统奠定了基础，兼顾效率、质量与适应性。研究展示了生成式

人工智能用于实体零售优化的可行性，这可能从根本上改变零售商在竞争日益激烈的市场中开展空间管理、商品陈列和提升运营效率的方式。

致谢

作者感谢审稿人提出的建设性意见。

声明

资助：不适用。**利益冲突**：作者声明不存在利益冲突。**数据可用性**：仿真代码可在以下网址获取：https://github.com/RaviTeja444/planogram-synthesis-genAI

参考文献

1. 亚马逊云科技：AWS 专业服务定价与合作模式（2024 20 2 2024 06），https://aws.amazon.com/professional-services/，访问日期：--。Austin, J., Johnson, D.D., Ho, J., Tarlow, D., Van Den Berg, R.：离散状态空间中的结构化去噪扩散模型。神经信息处理系统进展**34**，17981–17993（2021）3. Cantactix：通过货架陈列图数据准备提升零售效率。https: //www.cantactix.com/blog/the-planogram-build-process（2024）4. 凯捷研究院：零售人工智能成熟度评估：实施成本与效益。技术报告，Capgemini（2023）5. Corstjens, M., Doyle, P.：零售空间分配优化模型。《管理科学》**27**(7)，822–833（1981）6. Czerniachowska, K., Lutos-lawski, K.：求解零售货架空间分配问题的动态规划方法。计算机科学 Procedia**192**，4320–4329（2021）7. 德勤：2024年技术趋势：人工智能实施与变更管理成本。技术报告，德勤洞察（2023）8. Dhariwal, P., Nichol, A.：扩散模型在图像合成方面优于生成对抗网络。神经信息处理系统进展**34**，8780–8794（2021）9. Dice：2024年科技行业薪资报告：人工智能与机器学习工程师薪酬（2024），https://www.dice.com/career-advice/ai-ml-engineer-salary-guide-2024，访问日期：2024-06-15 10. Eismann, S., Scheuner, J., Van Eyk, E., Schwinger, M., Grohmann, J., Herbst, N., Iosup, A.：无服务器应用：为何、何时以及如何？IEEE Software**38**(1)，32–39（2021）11. Flexera：2024年云现状报告：企业云支出分析。技术报告，Flexera Software LLC（2024）12. Forrester Research：企业人工智能部署的总体经济影响。技术报告，Forrester Consulting（2024年3月），由微软委托 13. Frontoni, E., Marinelli, F., Rosetti, R., Zingaretti, P.：通过重新分配货架空间减少缺货。《计算机与工业工程》**106**，32–40（2017）

14. Gartner：零售业人工智能/机器学习计划的成本优化。技术报告 G00798435，Gartner, Inc.（2023年11月）15. Hansen, J.M., Raut, S., Swami, S.：零售货架分配：启发式与元启发式方法的比较分析。《零售学期刊》**86**(1)，94–105（2010）16. HIVERY：货架陈列图效率的演变：适应动态零售格局。https://www.hivery.com/blog/the-evolution-of-planogram-efficiency/（2023）17. Ho, J., Jain, A., Abbeel, P.：去噪扩散概率模型。神经信息处理系统进展**33**，6840–6851（2020）18. IDC：2024年零售业全球云基础设施支出指南。技术报告 US51234524，国际数据公司（2024年5月）19. Jonas, E., Schleier-Smith, J., Sreekanti, V., Tsai, C.C., Khandelwal, A., Pu, Q., Patterson, D.A.：简化云编程：伯克利对无服务器计算的见解。arXiv预印本 arXiv:1902.03383（2019）20. Li, H., Ota, K., Dong, M.：边缘端的物联网学习：结合边缘计算的物联网深度学习。《IEEE网络》**32**(1)，96–101（2018）21. 麦肯锡公司：2023年人工智能现状：生成式人工智能的突破之年。技术报告，麦肯锡全球研究院（2023）22. Murray, C.C., Talukdar, D., Gosavi, A.：零售品类管理中产品价格、陈列方向与货架空间分配的联合优化。《零售学期刊》**86**(2)，125–136（2010）23. 美国零售联合会：零售业的人工智能革命：2024年投资与实施报告。技术报告，NRF基金会，华盛顿特区（2024）24. 普华永道：2024年人工智能商业调查：实施成本与投资回报率分析。技术报告，普华永道（2024）25. RELEX Solutions：零售空间规划指南。https://www.relexsolutions.com/resources/retail-space-planning-guide/（2024）26. Robert Half Technology：2024年技术专业人员薪资指南。技术报告，Robert Half International Inc.（2024）27. Salary.com：美国品类管理经理薪资（2024），https://www.salary.com/research/salary/benchmark/category-management-manager-salary，访问日期：2024年12月15日。Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.：通过随机微分方程进行基于分数的生成建模。国际学习表征会议（2021）29. 美国劳工统计局：职业就业与工资统计。https://www.bls.gov/oes/current/oes131161.htm（2023）30. Valizadeh, M., Mozafari, S.S.：一种用于货架空间分配的新型协作式多智能体强化学习方法。《应用专家系统》**171**，114616（2021）31. Xu, M., Yu, L., Song, Y., Shi, C., Ermon, S., Tang, J.：GeoDiff：用于分子构象生成的几何扩散模型。arXiv预印本 arXiv:2203.02923（2022）