

2026.07.27

AI VIRTUAL TRY-ON · 技术月报

AI虚拟换装 三周巅峰论文

基础模型 · 原生2K · 实时推理
arXiv 7月重磅论文全景解读

AI虚拟换装三周巅峰论文

基础模型 · 原生2K · 实时推理 | arXiv 7月重磅论文全景解读

日期：2026-07-27 阅读：6 分钟 标签：AI虚拟换装 · 虚拟试衣 · arXiv · 扩散模型 · 自回归模型

2026年7月，AI虚拟换装（Virtual Try-On）领域迎来三篇重磅论文集中爆发：Oxygen-TryOn 开启基础模型时代，WearWow 突破2K多服装合成天花板，STAR-VTON 以自回归架构实现4倍速度提升。三个技术路线从不同维度推动虚拟换装从"能用"走向"好用"。

本文三大核心看点

- 基础模型：Oxygen-TryOn 以 fashion-native 理念，用 RL 强化学习统一多品类换装，超越 GPT-Image-2 等商业系统
- 2K 超高清：WearWow 解决 $O(N^2)$ 内存爆炸，原生2K多服装合成首次达到工业级标准
- 实时推理：STAR-VTON 引入自回归范式，4倍加速为毫秒级实时交互铺路

01

01 · 基础模型时代开启：Oxygen-TryOn

7月23日提交的论文 **Oxygen-TryOn** (arXiv:2607.21694)，代表了虚拟换装领域从"专用模型"到"通用基础模型"的范式转变。与以往针对单一服装品类设计的换装系统不同，Oxygen-TryOn 的核心创新在于它是**专为时尚产业原生打造**（fashion-native），而非将通用图像编辑模型改造而来。

技术路线：三阶段训练 + 强化学习微调

Oxygen-TryOn 构建了大规模专用换装数据 pipeline，完成收集、制造、标注、过滤全流程，并设计了独特的三阶段训练配方：



关键创新：无 Mask 生成

传统虚拟换装方法依赖 **姿态假人（Human Parsing）+ Mask 修复** 范式，需要先分割出人体和服装区域，再基于 Mask 进行图像修复。Oxygen-TryOn 将换装重新定义为"多参考、理解驱动的生成任务"，彻底摆脱 Mask 依赖。

强化学习阶段采用混合奖励机制：内部换装奖励模型评估生成质量， 联合专有通用模型打分，共同监督细粒度一致性和指令级质量。此外，系统还能在同一次生成中处理姿态变化等通用编辑指令。

能力特点与评测结果

Oxygen-TryOn 支持：单品和多品换装、正面/侧面全身照、任意数量参考图片、 自由多品组合，同时忠实保留人物身份和服装外观。

评测维度	Oxygen-TryOn	对比系统
单品换装一致性	SOTA	FLUX.2
多品换装质量	领先	GPT-Image-2, Seedream5 Lite
商业系统超越	超越	Nano Banana Pro

在公开基准测试和自有 Oxygen-TryOn Bench 上，Oxygen-TryOn 在单品换装的一致性和真实感 均达到 SOTA，多品换装质量领先，超越 GPT-Image-2 等商业系统。

02

02 · 突破2K天花板：WearWow 原生多服装合成

7月22日提交的 **WearWow** (arXiv:2607.19923) 将虚拟换装的分辨率天花板 推向原生 **2K**，这是数字时尚领域的重要里程碑。 论文指出了该方向的两大根本性瓶颈。

核心挑战：O(N²) 内存爆炸

合成原生2K分辨率的多服装换装，面临的首要问题是**内存爆炸**。 当需要处理多件服装时，2K图像的条件信息以 O(N²) 规模增长， 导致传统方法在消费级硬件上几乎无法运行。

核心挑战：扩散模型的纹理退化

第二个瓶颈来自扩散模型本身的光谱偏差—— 在压缩的潜在空间中进行多步去噪时，**高频织物细节**（褶皱、纹理、光泽）会被过度平滑， 生成的服装看起来"模糊"或"塑料感"十足， 难以满足工业级视觉保真度要求。

技术创新：自适应Token打包 (ATP)

WearWow 提出了自适应2D Token打包 (ATP) 机制：

ATP 三步流程：

- 1. 利用服装固有的稀疏性——服装在图像中并非全覆盖
- 2. 将异构物品算法打包到统一2D画布上
- 3. 剪枝非信息量背景token，最小化有效序列长度

结果：内存开销大幅降低，同时严格保留2D空间先验

技术创新：多维换装奖励系统 (MTR)

为解决纹理退化问题，WearWow 引入了**多维换装奖励（MTR）**系统：

MTR-S

语义引导奖励
Semantic Guidance Reward
显式驱动触觉恢复

MTR-C

服装分布奖励
Cloth Distribution Reward
隐式锚定物理分布

联合公式有效缓解了严重的奖励黑客问题。WearWow 还发布了**WearWow-2K** 数据集，提供物理正确的空间交互， 为极端质量合成提供训练基础。

评测结果显示，WearWow 在原生2K多服装合成领域建立新 SOTA， 超越了所有现有商业基准。

03

03 · 速度革命：STAR-VTON 自回归快速换装

7月13日提交的 **STAR-VTON** (arXiv:2607.11233) 引入了**视觉自回归（VAR）模型**，在虚拟换装领域实现了**4倍速度提升**， 为毫秒级实时交互铺平了道路。

问题：扩散模型的"质量-速度"困境

虚拟换装（VTON）是一个双条件图像生成问题，需要同时满足： 精确的人物身份保留、忠实的服装形变、细节合成。扩散模型虽能联合建模这些因素，但在压缩的潜在空间中进行多步去噪时，**高频细节丢失**问题难以避免。

更关键的是，扩散模型的**多步去噪**特性使其推理速度远慢于单步生成方法， 难以满足实时应用需求。

核心洞察：不同任务应使用不同表示空间

STAR-VTON 的核心洞察：

潜在空间

- 服装形变
- 人物布局
- 全局结构合成

像素空间

- 细粒度细节恢复
- 纹理/褶皱
- 高频视觉保真

≠

两阶段架构：潜在空间 + 像素空间

第一阶段（VAR-VTON）：在潜在空间中进行结构合成，整合服装条件和结构引导，实现高效潜在空间换装。VAR（视觉自回归）模型一次生成多个像素块，而非逐步去噪，大幅提升推理速度。

第二阶段（像素空间精炼）：引入**匹配感知精炼器**，在像素空间建立第一阶段生成结果与源服装之间的密集对应关系，直接映射细粒度像素级细节，恢复扩散模型丢失的高频信息。

评测结果：4倍加速，质量不降

4×

比扩散模型更快的推理速度，且不损失质量

更重要的是，像素空间精炼器作为**即插即用模块**，可以改进现有任何VTON方法——这意味着整个社区都能受益于这项提速成果。

04

04 · 三大技术路线对比与行业趋势

纵观三篇论文，可以清晰地看到2026年虚拟换装领域的三条并行技术路线，它们并非相互替代，而是面向不同场景的互补选择。

技术路线对比

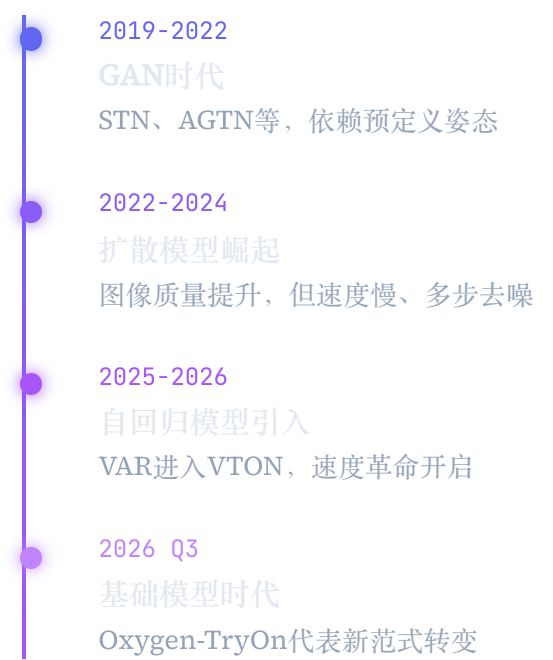
论文	技术路线	核心创新	最强场景	发布时间
Oxygen-TryOn	基础模型 + RL	fashion-native 三阶段训练 无Mask生成	多品类通用换装 商业系统对比	2026-07-23
WearWow	扩散模型优化	Adaptive Token Packing 多维奖励系统 MTR 2K原生合成	超高分辨率 多服装场景	2026-07-22
STAR-VTON	自回归模型	VAR引入VTON 两阶段架构 像素空间精炼	实时推理 移动端部署	2026-07-13

范式转变：从"专用"到"基础模型"

Oxygen-TryOn 的最大意义不在于在某个benchmark上超越了多少系统， 而在于它代表了一种**范式转变**—— 虚拟换装正在从"针对每个品类训练专用模型"走向 "一个基础模型处理任意品类换装"。

这与计算机视觉领域 CLIP、DINO、Stable Diffusion 等基础模型的出现路径类似： 通用性提升 → 应用边界扩大 → 行业渗透加速。 当换装模型的"通用性"问题被解决， 电商、AR试衣、AI设计工具的落地门槛将大幅降低。

技术演进时间线



05 · 行业落地：从技术突破到商业应用

三篇论文的技术突破并非停留在学术层面，虚拟换装的商业化路径已经相当清晰。从电商到 AR 试衣，从 AI 设计工具到游戏角色定制， 每一条路径都在等待"好用"这个临界点的到来。

电商场景：实时换装改变购物体验

传统电商的"试穿"依赖用户自拍上传，不仅体验割裂，还存在光线、角度、背景不统一的问题。当 STAR-VTON 的**4倍加速**成熟落地，实时虚拟换装将成为电商详情页的标准功能—— 用户点击即试，无需等待，用户体验从"异步"变为"同步"。

AR 试衣：移动端部署的突破口

AR 眼镜和手机AR试衣的核心瓶颈是**推理延迟**。扩散模型的多步去噪在移动端难以实时运行，而自回归模型（STAR-VTON） 和蒸馏技术的发展正在打开这扇门。当换装推理从秒级压缩到毫秒级， AR 试衣的沉浸感才能真正实现。

AI 设计工具：设计师的工作流重构

Oxygen-TryOn 的 fashion-native 基础模型思路，对于 AI 设计工具开发者尤为重要—— 通用基础模型 + 少量垂直数据微调，可能比从零训练专用模型更高效。 这意味着中小型创业公司也能基于开源基础模型快速构建垂直换装应用。

游戏与虚拟形象

游戏角色定制、虚拟主播形象、社交头像个性化—— 这些场景对换装的**速度和多样性**要求极高， 对图像质量要求相对宽松。WearWow 突破2K分辨率意味着 游戏内虚拟形象的换装效果首次达到发布级画质。

商业落地关键问题：

- 推理成本：基础模型参数量大，边缘部署仍需优化
- 隐私合规：人体图像处理需符合 GDPR、个保法要求
- 多品类泛化：模型在极端品类（异形服装、特殊材质）上的表现仍需验证
- 生态整合：API标准化、服务商整合仍是早期市场问题

06 · 技术挑战与未来方向

尽管2026年三篇论文代表了重要突破，虚拟换装领域仍有多项核心挑战待解。 这些问题将是未来1-2年学术和工业界的重点攻克方向。

挑战一：极端品类的泛化能力

现有系统在标准品类（T恤、衬衫、裤装）上表现优异， 但面对**异形服装**（如和服、婚纱、戏服）和**特殊材质**（如皮革、金属装饰、羽毛）时，生成质量下降明显。 这类品类的训练数据稀缺，且空间形变规律与日常服装差异巨大。

挑战二：时序一致性

当前换装研究主要聚焦于**单帧静态**生成， 视频换装（连续帧间服装一致性）是更困难的问题。 当人物走动、转身、跳跃时，服装需要保持物理连续性—— 这需要将换装与视频生成技术（如DiT）深度结合。

挑战三：可控性与编辑灵活性

Oxygen-TryOn 支持姿态编辑，但整体可控性仍有限。 用户可能希望指定"只换上衣不改裤子"、"换装同时保留特定配饰"、 "调整服装松紧度"——这类**细粒度条件控制**需要更完善的条件机制设计。

未来方向展望



视频换装

连续帧一致性 + 时序服装形变



细粒度控制

区域指定、属性调节、编辑撤回



端侧部署

手机/AR眼镜上的实时推理



多模态融合

文本+图片+3D联合驱动的换装

07 · 关键结论与行动建议

本期三篇论文揭示了虚拟换装领域正在经历的关键转变。 以下是核心结论和面向不同读者的行动建议。

三大核心结论

- 01

基础模型范式已至

Oxygen-TryOn 代表了换装从"专用模型"到"通用基础模型"的转变，类比CLIP在视觉领域的意义——通用性提升将大幅降低行业应用门槛。
- 02

速度与质量不再对立

STAR-VTON 证明自回归模型可以在4倍加速的同时保持质量——这打破了"扩散模型=高质量=慢"的旧有认知，为实时应用打开大门。
- 03

2K多服装成为新战场

WearWow 的突破意味着多服装换装首次达到工业级分辨率标准——电商场景的全套搭配推荐将不再受制于分辨率限制。

面向不同读者的行动建议

技术创业者

关注 STAR-VTON 的即插即用精炼器，可用其改进现有产品换装质量；同时跟进 Oxygen-TryOn 的开源动态，评估基础模型接入成本

投资人

虚拟换装的基础模型路线已验证，商业化路径清晰；重点关注有扩散模型优化经验 + 时尚行业数据的团队

AI/ML从业者

三篇论文代码若开源，是良好的学习样本；关注自回归模型在视觉生成领域的更多应用（VAR + VTON 结合点值得关注）

时尚行业从业者

2K多服装合成成熟将改变电商内容生产流程；建议开始评估AI换装在自身业务场景的集成可行性

08 · 参考文献与延伸阅读

本期内容基于以下三篇 arXiv 论文，更多技术细节请参考原文。

Oxygen-TryOn: Fashion-Native Foundation Model for Any-item Virtual Try-On

Liu, Y., Fu, X., Xu, Z., et al. (14 authors)

[arXiv:2607.21694](#) · Submitted 23 Jul 2026

<https://arxiv.org/abs/2607.21694>

WearWow: Native 2K Multi-Garment Virtual Try-On via Adaptive Token Packing and Preference Alignment

Zhang, X., Du, R., Chang, S., et al. (10 authors)

[arXiv:2607.19923](#) · Submitted 22 Jul 2026

<https://arxiv.org/abs/2607.19923>

STAR-VTON: Structure-Detail Decoupled Autoregressive Generation for Fast and High-Fidelity Virtual Try-On

Yang, L., Hu, X., Li, Y., et al. (6 authors)

[arXiv:2607.11233](#) · Submitted 13 Jul 2026

<https://arxiv.org/abs/2607.11233>

延伸阅读

更多 arXiv 虚拟换装论文，可访问[arXiv Virtual Try-On 搜索页面](#)。当前共有 **535+** 篇相关论文，2026年7月新提交约20篇。

本期编辑说明：

本文由 AI 自动采集 arXiv 最新论文、生成编辑方案并排版发布。原始素材来自论文摘要和公开信息，数据以论文原文为准。如有技术细节疑问，建议阅读原论文。