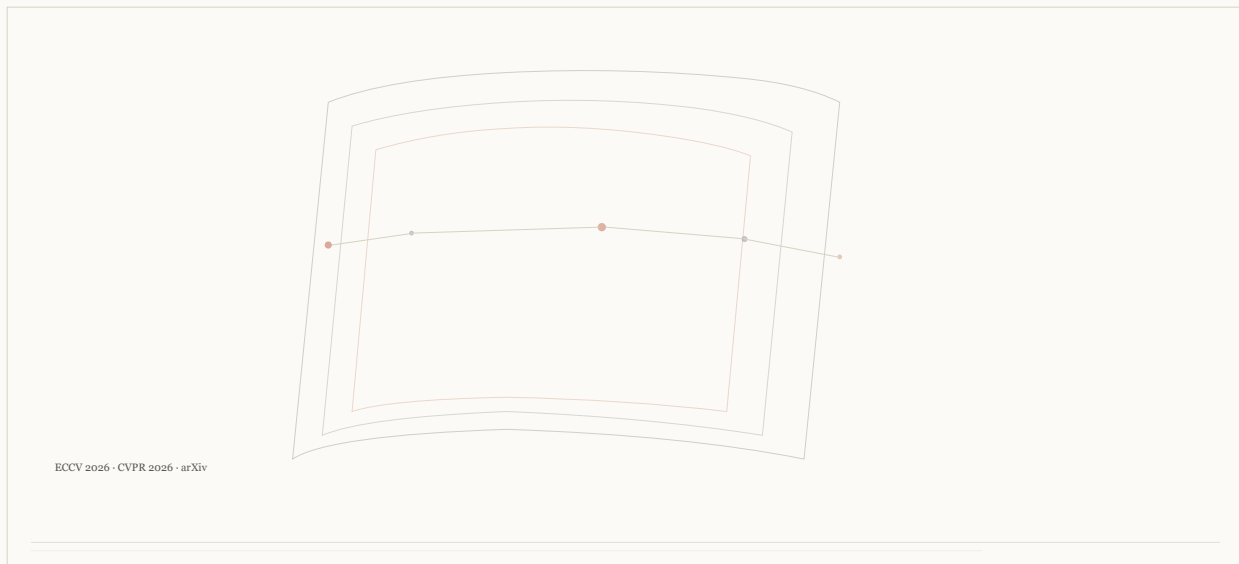




从单层到多层：VTON的第二次范式跃迁

2023—2025年，扩散模型驱动的AI虚拟换装（如OOTDiffusion、IDM-VTON、StableVITON）解决了 给定一件衣服+一个人、合成穿着效果的问题。这是一维对应——一件上衣换另一件上衣，一条裤子换另一条裤子。

但真实世界的穿衣逻辑远更复杂：衬衫外面套夹克，夹克外面再披大衣——或者试穿时先脱掉外套再换上新品。现有方法在多层叠加场景下集体失灵，因为它们依赖的cloth-agnostic representation（衣物无关表示）会丢弃关于已有衣物的关键上下文信息。

2026年7月的这一波进展表明，行业共识正在形成：

维度	2023—2025（第一代）	2026（第二代）
替换模式	单层替换	多层叠加+穿脱
物品支持	单一品类	任意品类
分辨率	512—1024	原生2K
模型路径	扩散模型微调	基础模型（CPT/SFT/RL）
参考图数量	1张	多张（商品图+实穿图）
指令跟随	无	支持姿态/编辑指令

Oxygen-TryOn：时尚原生的基础模型信仰

核心论文： *Oxygen-TryOn: Fashion-Native Foundation Model for Any-item Virtual Try-On*
(arXiv:2607.21694, 2026.7.23)

Oxygen-TryOn的身份宣言非常明确：不是在改造通用图像编辑器来做试衣，而是从零构建**时尚原生**（fashion-native）的基础模型。这不是修辞，而是体现在三个工程选择中：

数据引擎：氧气团队构建了一个数据引擎，系统性地采集、制造、标注和过滤大规模高质量试衣数据。这不是从LAION里顺手捡来的服装图片，而是专门为try-on设计的、覆盖全品类的训练语料。

三阶段训练：CPT（持续预训练）→ SFT（监督微调）→ RL（强化学习）。其中RL阶段尤为创新——使用混合奖励函数，联合一个内部试衣奖励模型（负责细粒度一致性）和一个专有引导通用模型（负责指令级质量），在单次pass中同时监督穿得对和听指令。

我们提出了Oxygen-TryOn，一个统一的多物品虚拟试衣基础模型。在单物品和多物品试穿基准上均达到SOTA，超越了Nano Banana Pro、GPT-Image-2、Seedream5 Lite和FLUX.2。

Oxygen-TryOn论文摘要

真正的任意物品：支持全/半身视图、可变参考图数量、自由多物品组合。在一个统一框架内，上衣、下装、全套甚至配饰都能处理。对比之前的系统只做一件衣服换另一件同一品类，这是本质上的能力拓展。

LVTON：叠穿变成可计算的

核心论文： *Layering Virtual Try-On* (arXiv:2607.22924, 2026.7.24, ECCV 2026)

如果说Oxygen-TryOn的方向是宽度——覆盖所有品类，那么LVTON的方向就是深度——理解衣物之间的**层次关系**。这是ECCV 2026接收的亮点工作，也是目前最系统地研究多层叠穿的论文。

核心洞察：LVTON团队发现，现有VTON范式从根本上就不适合叠穿。因为cloth-agnostic表示会丢弃已有衣物的信息，而单物品数据集根本不包含层次上下文。因此他们将问题解构为两个独立能力：

① General VTON Priors（通用试穿先验）

变形能力、身份保持——通过自动数据生成pipeline，从时尚视频中分割+修补来合成训练样本。

② Specific Layering Knowledge（特定叠穿知识）

叠加顺序、遮挡推理——在小型专用LVTON数据集上微调获得。

这种解构有其工程智慧：泛化能力来自大数据，层次逻辑来自精数据，两套能力互不干扰。

零样本泛化：LVTON方法在传统VTON基准上也能取得SOTA，甚至在零样本设定下也表现优异。这意味着学会叠穿并没有牺牲通用试穿能力，反而因为对空间关系有更深刻的理解而增强了。

我们的关键洞察是，LVTON挑战必须解构为两种不同能力：(1) 通用VTON先验（变形、身份保持）和 (2) 特定叠穿知识（叠穿顺序和遮挡推理）。

WearWow：清晰度战争的终局？

核心论文： *WearWow: Native 2K Multi-Garment Virtual Try-On via Adaptive Token Packing and Preference Alignment* (arXiv:2607.19923, 2026.7.22)

当Oxygen-TryOn和LVTON在能力边界上展开竞争时，WearWow选择在**清晰度**上建立壁垒。它是目前首篇实现**原生2K多衣物虚拟试衣**的端到端框架。

两个核心瓶颈及解法：

- 瓶颈一： $O(N^2)$ 内存爆炸
2K分辨率下，多参考图条件导致token序列爆炸式增长。解法：**Adaptive 2D Token Packing (ATP)**——利用衣物的空间稀疏性，将异构物品算法化打包到统一2D画布，并剪枝不携带信息的背景token。
- 瓶颈二： 高频织物纹理丢失
扩散模型在压缩潜空间中会平滑掉织物细节。解法：**Multi-dimensional Try-on Reward (MTR)**系统——包含Semantic Guidance Reward（驱动纹理恢复）和Cloth Distribution Reward（锚定物理分布）。

WearWow还配套构建了WearWow-2K数据集——原生2K三元组，具有物理正确的空间交互关系，让模型自然学会无掩码的精细生成。

三条赛道，一个目标

三篇论文在技术和侧重点上各有取舍，但在根本上指向同一个方向：**AI虚拟换装正在从技术demo转变为生产级工具**。

维度	Oxygen-TryOn	LVTON	WearWow
核心能力	多品类任意试穿	多层叠穿推理	原生2K超高清
训练策略	CPT+SFT+RL	自动数据pipeline+微调	2K数据集+奖励系统
输入类型	多参考图+人物图	人物图+衣物层	多衣物图+人物图
分辨率	1024	1024	原生2048
特别优势	RL混合奖励	零样本+叠穿逻辑	2K+微纹理保持
发布时间	2026.7.23	2026.7.24 (ECCV)	2026.7.22

三者之间并不完全是竞争关系。Oxygen-TryOn适合做通用入口，LVTON的技术可以补充叠穿能力，WearWow的ATP+MTR方案也适用于前两者的架构。三种思路的融合，很可能就是2026下半年VTON最重要的一条主线。

从实验室到电商：商用化路径

这些技术对电商行业的冲击是直接的。以下是三个最有可能的落地方向：

方向一：品牌独立虚拟试衣间

Oxygen-TryOn类的基础模型可以部署为品牌官网的试衣工具。消费者上传人像照片，即可试穿全品类商品，同时试穿上衣+下装+外套的组合。不再需要等待退货换货。

方向二：UGC穿搭社交

LVTON让用户在APP中自由叠穿搭配：先选打底衫，再叠穿外套、围巾，甚至做穿脱演示。这种体验在时尚社区中有极强的种草效应。

方向三：超高清版型打样

WearWow的2K分辨率意味着设计师在打样前就能看到衣服在模特身上的精细效果——包括面料纹理、褶皱、光泽，可以替代部分实物打样流程。

当前局限：

- 实时推理仍需要高端GPU，边缘设备部署有待优化
- 多衣物场景下的物理模拟（布料褶皱、重力拖曳）不够真实
- 长序列穿脱（穿→脱→再穿）时累积误差

挑战与展望

站在2026年7月底回看，VTON领域已经走过了一条清晰的进化路径：

1. 第一代：GAN-based，低分辨率，单品类
2. 第二代：Diffusion-based，512-1024px，单层替换
3. 第三代：Foundation model-based，多品类+多尺度，正在上演

尚未完全解决的问题包括：

多衣片物理模拟：目前模型在视觉上看起来合理，但衣物的物理交互（叠穿紧致度、衣领堆叠、褶皱传递）远未达到物理模拟器的精度。融合神经渲染与物理模拟可能是一条路径。

实时性：WearWow的2K生成需要数十秒到数分钟。电商场景下用户能接受的等待时间是秒级。剪枝量化、蒸馏和并行解码是工程化关键。

隐私与个性化：用户上传全身照片在隐私法规下尤其敏感。联邦学习、本地推理和匿名化处理是部署的必要组件。

数据质量瓶颈：所有三篇论文都提到数据是最大瓶颈。高品质的A→B试衣三元组的标注成本极高。Oxygen-TryOn的数据引擎和LVTON的自动数据pipeline提供了两种不同解决思路，但距离大规模生产级数据还有距离。

结语

过去五年，AI虚拟换装一直在换掉一件衣服的框架内打转。2026年7月， 我们终于看到了三个不同方向的同时突破：氧气团队的**宽度**、LVTON的**深度**、WearWow的**精度**。

在技术演进的时间线上，这很可能是一个**密度极高的转折点**。当这三条技术路线开始融合——支持任意品类、理解多层关系、输出超高清效果的统一平台——那将是AI在时尚领域真正的iPhone moment。

对于国内从业者来说，拥有数据生态和场景优势的电商平台（淘宝、抖音、唯品会）可能更快落地。而对于技术团队，现在正是选择路线的最佳时机：做通用入口、深度叠穿、还是视觉保真——或者说，如何融合三者。

本文发布于2026年7月30日。所引论文均可在arXiv公开获取：2607.21694, 2607.22924, 2607.19923。