

→

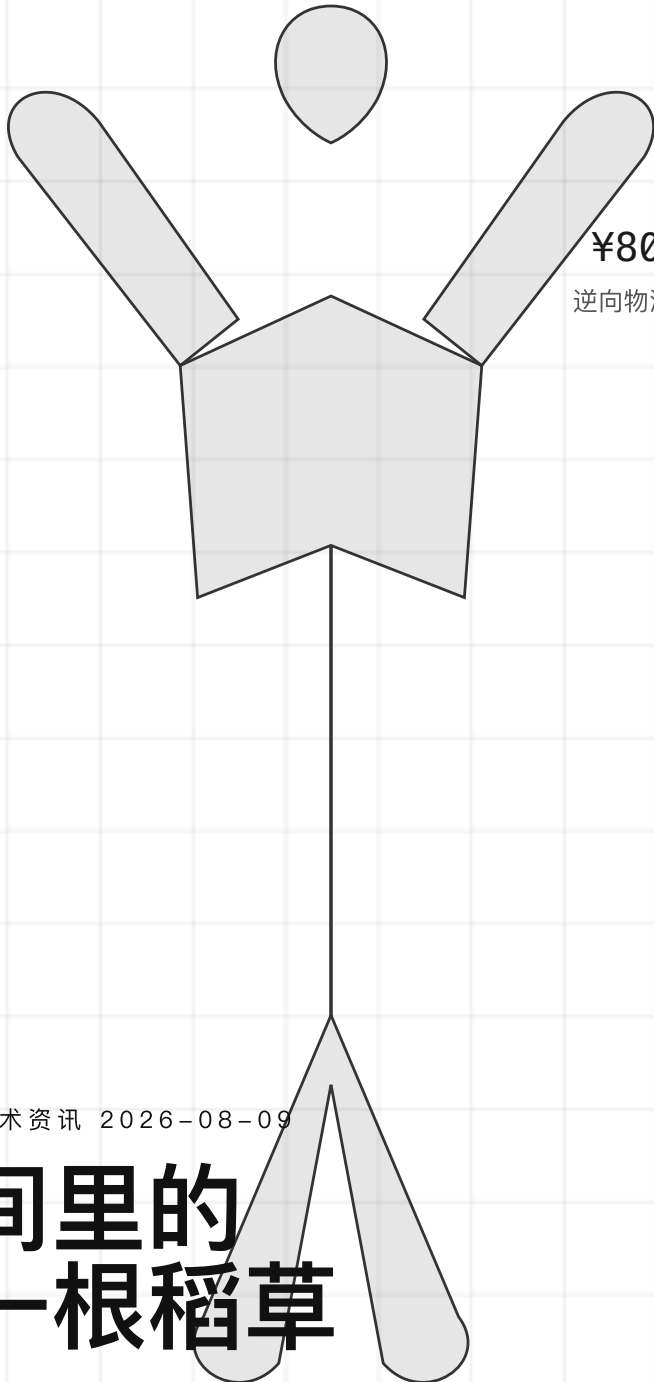
60%

女装退货率

←

¥800亿

逆向物流成本



AI 虚拟换装 · 技术资讯 2026-08-09

试衣间里的 最后一根稻草

从"玩具"到"降本工具": AI 换装如何对抗 60% 退货率, 学界又如何把"单层换装"推进到 "任意品类 + 可叠加"。

目录

1. 01被退货压垮的女装电商
2. 02为什么现在才算「能用」
3. 03学界范式一：Oxygen-TryOn，任意品类的
基础模型
4. 04学界范式二：挑战「层叠/脱层」
5. 05平台与产业的落地实证
6. 06趋势研判：多参考、可叠加、实时

试衣间里的最后一根稻草

AI 虚拟换装正在同时完成两场跃迁：学界从「单件替换」迈向「任意品类 + 层叠真实感」，产业从营销玩具变成对抗高退货率的降本工具。

日期：2026-08-09版式：longform · tufted配图：ai-generated · 封面 1 张

当女装电商被 60% 的退货率压垮，AI 试衣间成了最后一根稻草；而实验室里，换装技术正从「只换一件」走向「任意品类、层层叠加」。本期资讯把这两条线放在一起看：一端是数字，一端是通往数字的模型。

本期速览

- 产业侧：60% 女装退货率 / 800 亿逆向物流成本，AI 试衣间有实证降本成效——京东 -25% 退货、Style3D 报告 -25% 退货 +30% 转化。
- 学界侧：Oxygen-TryOn 把试衣做成「任意品类」的 fashion-native 基础模型；LVTON (ECCV 2026) 攻克「叠层/脱层」。
- 平台侧：淘宝天猫 3D 虚拟试衣、抖音「虚拟上身」、京东 AI 穿搭助手 2025-2026 密集落地，且走向实时渲染数字分身。
- 趋势：学界与产业的交集 = 多参考、可叠加、实时——下一阶段的胜负手。

01

被退货压垮的女装电商

一个数字，压垮了无数女装商家：**60%**，这是中国女装电商的平均退货率。而在 抖音直播间，部分主播的退货率更是冲到 **80%**。卖 10 件退 6 件——剩下 4 件赚的钱，往往还不够覆盖 6 件的运费与折旧。

这笔账放大到全行业触目惊心：2025 年中国电商女装退货产生的 **逆向物流成本超过 800 亿元**，相当于一个中型省份的全年财政收入。退回的衣服要 重新消毒、熨烫、包装、二次上架，每一步都在烧钱；大量退货最终流向折扣店和尾货市场，品牌形象 跟着受损。换句话说，这 800 亿不只是物流账单，更是电商信任成本的一次性坦白。

60%

女装平均退货率

80%

抖音主播峰值

¥800亿

年逆向物流成本

退货危机的量级感知

痛点一句话

退货率高的根源就三个字：**不匹配**。买家秀和卖家秀差着十万八千里——模特穿是修身款，你穿是紧身款；屏幕上的雾霾蓝，到手是灰蓝色；标着 S 码，实际版型却偏大两个号。这是电商的固有缺陷：你没法在线上真正“试衣服”。

AI 试衣并不是新概念。2018 年京东就推出过 AR 试衣功能，但体验很差——模型僵化、衣服像贴上去的、色彩偏差大，用户试了一次就再也不想用。根本原因是当时的技术没到位：那代深度学习能做人脸识别，却搞不定布料的物理仿真——布料怎么下垂、褶皱怎么生成、不同面料的光泽怎么渲染，统统是难题。

转折发生在 2024—2025 年的生成式 AI 爆发期。虚拟试衣从此有了质的飞跃，而撑起这一跃迁的，正是下一节要展开的三项关键技术合力。

02

为什么现在才算「能用」

2018 年的 AR 试衣之所以失败，本质上是技术没到那一步。真正的转机来自 2024—2025 年的生成式 AI 爆发：它不是某一个模型的功劳，而是三项技术在工作流里形成合力，才让虚拟试衣从「能看」变成「能用」。

01	02	03
3D 人体重建	布料物理仿真	尺寸匹配算法
一张正面 + 一张侧面照，重建带体型数据的 3D 模型；肩宽/胸围/腰围/臀围误差 <2cm，手机摄像头即可。	最难的一环。用「神经渲染」让模型自学布料行为：纯棉怎么垂、真丝怎么飘、雪纺怎么透，不再靠手工调参。	结合品牌版型、用户历史购买、同身材用户上身效果，动态推算「这家店的这个款式该选什么码」。

支撑「能用」的三项技术

这三项能力恰好补齐了 2018 年那个不可能三角：人体重建解决了「合不合身」，布料物理仿真解决了「像不像真的」，尺寸匹配解决了「该买哪一号」。当三者能在手机端跑通，虚拟试衣才真正具备了走进 电商详情页的条件。

技术坐标

布料物理仿真是其中最难、也是决定观感上限的一环。它不是预设动画，而是让模型在神经渲染的驱动下自己「学会」不同面料的物理行为——这正是生成式 AI 相对传统图形学最本质的差异。

03

学界范式一：Oxygen-TryOn，任意品类的基础模型

产业端在忙着降本，学界则在重新定义试衣这件事本身。2026 年 7 月提交的 Oxygen-TryOn (arXiv:2607.21694) 提出了一个关键判断：与其改造一个通用的图像 编辑器来「勉强做试衣」，不如直接构建一个**时尚原生 (fashion-native)** 的试衣基础 模型，用专用的数据引擎和试衣专用训练把它做好。

它接受的输入十分灵活：一个或多个参考单品（干净的电商图，或自然场景里穿着着的照片），加上一张 目标人物图，就能合成这个人穿着这些单品的逼真图像，覆盖几乎所有时尚品类。这打破了过往系统的两个 局限——过往系统多在摄影棚设定下处理单一品类；近年多参考方法也仍以「单件服装」为

中心。Oxygen-TryOn 则支持全身/半身视角、可变的参考数量、以及自由的多单品组合，同时保持人物身份与 服装外观的一致。

范式转变

它不再用 mask 做图像修补（把衣服抠掉再填新的），而是把试衣重构为**多参考、理解驱动的生成 任务**——「给 AI 看同一人穿不同衣服的照片 + 给你想要的新衣，让它直接生成换装后的效果」。

Oxygen-TryOn 三阶段训练配方

阶段	全称	作用
CPT	连续预训练	在自建试衣数据上打底，建立试衣领域的基础能力。
SFT	监督微调	用高质量标注样本对齐试衣任务的目标输出。
RL	强化学习	用混合奖励联合监督细粒度一致性与指令级质量。

arXiv:2607.21694

其中 RL 阶段采用了**混合奖励**：自研的试衣奖励模型 + 一个专有、rubric 引导的通用 模型，两者联合监督「细粒度一致性」与「指令级质量」。有意思的是，它在同一轮生成里也能跟随通用编辑 指令——比如改变姿态——而不只是呆板地换衣服。在公开基准与自研的 Oxygen-TryOn Bench 上，它的 单件试衣一致性与真实感已达到 SOTA，多件试衣也领先。

任意品类

全身/半身视角

多参考输入

自由多单品组合

跟随编辑指令

数据引擎驱动

Oxygen-TryOn 的能力边界示意

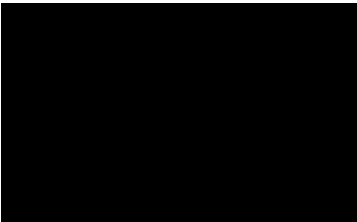
04

学界范式二：挑战「层叠/脱层」

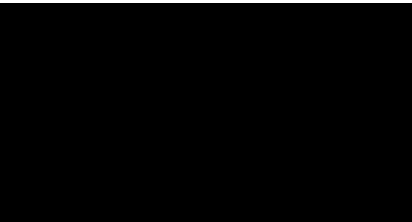
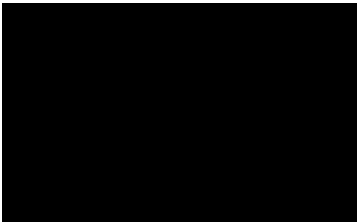
Oxygen-TryOn 解决的是「品类广度」，而 2026 年 7 月、被 ECCV 2026 接收的 **Layering Virtual Try-On (LVTON, arXiv:2607.22924)** 解决了另一个更贴近日常的问题——**层叠**。

现实世界的时尚本质是层叠：衬衫外面加一件夹克，或依次穿上、再脱下一层。可现有的大多数虚拟试穿（VTON）方法恰恰擅长的是**单层替换**，并不设计来「叠层/脱层」一套既有穿搭。作者认为，现有范式的 cloth-agnostic（与布料类别无关的）表征，加上只含单件的数据集，把层叠所需的上下文——比如遮挡顺序——给丢弃了。

单件替换（现有）



层叠/脱层（LVTON）



夹克 + 衬衫，顺序 + 遮挡

层叠 vs 单层替换

应对之道是把 LVTON 挑战拆成两种可分离的能力：一是**通用 VTON 先验**（形变、身份保持），二是**特定的层叠知识**（层叠顺序、遮挡推理）。前者来自一个自动数据生成管线——从时尚视频 经分割 + 修补合成训练样本；后者则在一个小型、专用的 LVTON 数据集上微调学习。这种拆解让模型既保留了试衣的通用水准，又学会了层叠的逻辑。

ECCV 2026

在这套设计下，LVTON 在其基准上达到 SOTA，在传统 VTON 基准上泛化更佳——微调后刷新 SOTA，且具备 zero-shot 能力。它和 Oxygen-TryOn 一起，把「试衣」从「换一件衣服」推向了「理解一套真实的穿着」。

05

平台与产业的落地实证

如果说学界在定义「技术能走到哪」，那么平台和品牌正在定义「技术能不能赚钱、值不值得铺」。2025-2026 年，多个玩家已经跑出了可量化的真实效果——而且方向惊人一致：指向降本与转化。

主要玩家与已跑出的数据

玩家 / 功能	形态	效果数据
京东 AI 穿搭助手	大模型 · 换装	使用用户退货率比未用低 25%
抖音「虚拟上身」	直播实时渲染	2026 Q1 内测，跟随主播角度同步
淘宝天猫 3D 虚拟试衣	元宇宙魔盒 · 数字分身	10 秒生成分身，免费接入
Style3D 虚拟试衣	品牌 B 端方案	报告 -25% 退货 / +30% 转化

新浪科技 / Style3D / 公开报道汇总（2026-07）

京东 2025 年上线的 **AI 穿搭助手**基于大模型：用户上传全身照，AI 识别体型、推荐合适 版型的衣服，还能「换装」看效果。官方数据是使用它的用户退货率比未使用用户低 25%——如果全线推广，每年有望省下数十亿的退货成本。

抖音电商 2026 年第一季度开始内测「**虚拟上身**」：主播直播时，观众可一键生成自己的 3D 模型，看到衣服穿在自己身上的效果。关键区别在于——这不是截图式贴图，而是**实时渲染**，跟随主播的展示角度同步变化。淘宝天猫则通过「元宇宙魔盒」计划，让参与计划的服饰商家免费接入 3D 虚拟试

衣：消费者在商品详情页点「AI 试衣」，10 秒内生成专属数字分身，实时查看不同尺码、颜色、搭配的上身 效果。

海外视角同样印证了量级。Style3D 在 2026 年 3 月的博客里援引数据：全球时尚产业每年因退货损失高达 **5500 亿美元**；超过 70% 的全球时尚消费者期待互动式数字购物体验；采用先进虚拟试衣算法 的电商品牌报告退货降低 25%、客单价与购物篮规模提升。

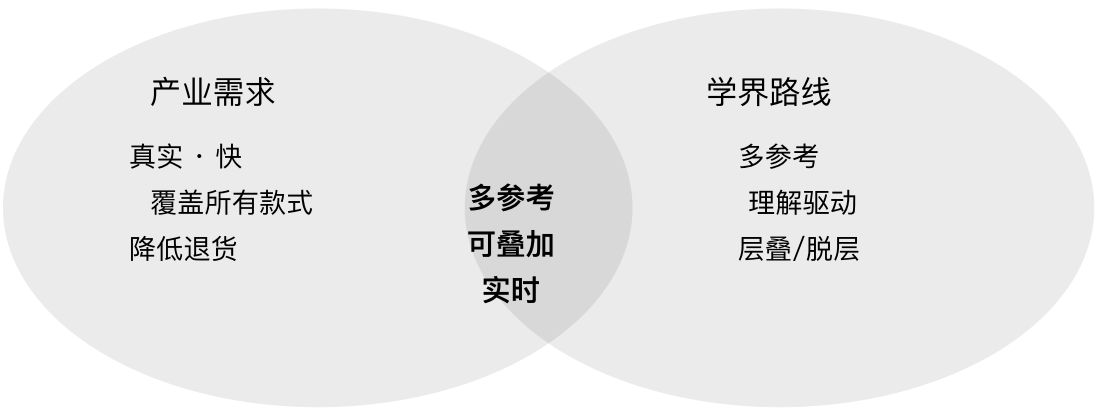
仍需冷静

需要说明的是，多数「降低 25% 退货」是平台在特定人群/试点下的统计口径，并非全行业恒定结论；且紧身、修身、西装等对版型敏感的品类，AI 仍可能出现肩线歪斜、腰线不准的问题。它解决了一大部分「不匹配」， 但还没有解决全部。

06

趋势研判：多参考、可叠加、实时

把学界与产业两条线放在一起，能看到一个清晰的收敛方向。产业端的需求是「真实 + 快 + 覆盖所有款式」， 学界的技术路线是「多参考、理解驱动、支持叠加」。两者的交集，正是虚拟试衣下一阶段的三个关键词：**多参考、可叠加、实时**。



产业需求与学界路线的交集

多参考让试衣可以「一次看多件、多种搭配」，对应电商的搭配推荐；**可叠加**让「夹克加衬衫」这类日常穿搭也成立，对应真实的穿法而非理想化的单件；**实时**则把虚拟试衣 从「上传照片等几秒」推进到直播中跟随角度渲染——这正是抖音「虚拟上身」已经在做的事。

值得注意的还有一条被低估的线：**理解驱动的生成**。无论是 Oxygen-TryOn 的 fashion-native 重构，还是 LVTON 对层叠逻辑的显式建模，它们都在摆脱「像素级修补」的老思路，转向「理解穿搭结构再生成」。这条路一旦与实时渲染结合，体验上限会更高，但算力和端侧部署的挑战也随之而来。

一句话结论

虚拟试衣的胜负手，已经从「能不能换装」变成了「换得多不多、叠得真不真、来得快不快」。学界和产业 正在同一套语言下，把「试衣」变成一套能读懂穿搭、再实时生成真实效果的完整能力。