目录

AI虚拟换装技术日报

从图像到视频 · 从 UNet 到 DiT · 7月开源项目速览

日期： 2026年7月12日　　出品： 小沐 AI · 无人值守采集

本周 AI 虚拟换装领域持续升温。vivo 团队的 MagicTryOn 基于 Wan2.1 扩散 Transformer 实现视频换装突破，CatVTON 以不到 8G 显存的轻量方案领跑行业，ICCV 2025 的 Any2AnyTryon 提供多场景统一框架。从图像到视频、从 UNet 到 DiT， 虚拟换装技术正在经历范式级转变。

本周要点

1. CatVTON（ICLR 2025）持续活跃，总参数量 899M，推理仅需 8G 显存
2. vivo MagicTryOn 采用 Wan2.1 Diffusion Transformer，全自注意力建模时空一致性
3. Any2AnyTryon（ICCV 2025）提出自适应位置嵌入，支持多场景统一换装
4. StreetTryOn（WACV 2025）聚焦真实户外场景换装基准
5. 技术趋势：轻量化、视频化、DiT 架构取代传统 UNet

01 开篇：一周行业概览

2026年7月，AI 虚拟换装（Virtual Try-On）领域呈现出一派繁忙景象。 从 GitHub 开源社区的动态来看，多个重量级项目在过去一周内均有活跃更新， 反映出这一方向的研究热度不仅没有消退，反而在持续升温。

本周值得关注的核心事件有三项：首先是 vivo 相机研究团队开源的**MagicTryOn** 在7月11日仍有代码更新，该项目基于 Wan2.1 扩散 Transformer 实现视频换装，代表了当前技术前沿； 其次是 ICLR 2025 接收的 **CatVTON**，凭借仅 8G 显存即可推理 的轻量级特性，已成为行业参考基准； 第三是 ICCV 2025 的 **Any2AnyTryon**，提出了一种多场景统一的换装框架。

> 从图像到视频的跨域升级，是当前虚拟换装技术的核心挑战。
>
> — 技术社区共识

核心判断

这些项目的共同特征是：从传统的 UNet 架构转向更强大的 Diffusion Transformer (DiT) 架构， 参数量不断突破的同时也在追求极致的轻量化。

02 CatVTON 持续领跑：轻量级方案的极致优化

由郑州大学崇政团队开发的 **CatVTON** 是 ICLR 2025 接收的论文成果， 也是当前 GitHub 上虚拟换装方向星标最高的开源项目（1789 星）。 其核心创新在于将复杂的虚拟换装任务压缩到了一个极致的轻量化框架中。

CatVTON 的三大技术支柱使其在学术界和产业界均获得了广泛关注： 第一，轻量网络设计，总参数量控制在 899.06M，远低于同类模型； 第二，参数高效训练策略，仅需 49.57M 参数即可参与训练，大大降低了复现和二次开发的门槛； 第三，简化推理流程，在 1024×768 分辨率下仅需不到 8G 显存即可完成推理。

数据对比

传统 UNet 架构的虚拟换装模型通常需要 12G+ 显存才能处理 1024×768 分辨率， CatVTON 将这一门槛降低了约 33%，使得在消费级 GPU（如 RTX 3060 12G）上 运行高质量的虚拟换装成为可能。

值得注意的是，截至 2026 年 7 月 11 日，CatVTON 项目仍在积极更新维护， 社区贡献者持续提交优化建议和改进代码。这表明该项目不仅是一篇学术产出， 更是一个长期维护的工程基座。对于希望在其基础上进行二次开发的团队而言， CatVTON 提供了良好的起点。

总参数量	**899.06M**
可训练参数	**49.57M**（参数高效训练）
推理显存	**< 8G**（1024×768）
论文	ICLR 2025

✦ CatVTON 关键指标

03 MagicTryOn：vivo 团队的 Wan2.1 DiT 突破

vivo 相机研究团队在 2025 年 5 月发布了论文"MagicTryOn: Harnessing Diffusion Transformer for Garment-Preserving Video Virtual Try-on"， 并于 2025 年 6 月全面开源。截至 2026 年 7 月 11 日，该项目已积累 559 星标， 社区活跃度持续保持高位。

MagicTryOn 的技术路线具有鲜明的前瞻性。与当时主流的 UNet 架构不同， 它直接采用了 Wan2.1 扩散 Transformer（Diffusion Transformer, DiT）作为骨干网络。 DiT 架构在图像和视频生成领域已被证明优于传统 UNet， MagicTryOn 是首批将 DiT 应用于视频虚拟换装方向的研究之一。

该框架的三大核心创新点：

- **全自注意力时空建模**：采用全自注意力机制（full self-attention）来捕捉视频帧间 的时空一致性，避免了传统方法中因帧间不一致导致的闪烁问题。
- **由粗到细的服装保留策略**：引入 coarse-to-fine 的服装保留机制， 确保换装后的服装纹理、图案和颜色与参考服装高度一致。
- **掩码感知损失**：针对服装区域的掩码感知损失函数（mask-aware loss）， 专门增强服装区域的保真度，减少背景污染和服装形变。

> MagicTryOn 的核心贡献在于证明了 DiT 架构在视频换装任务上的潜力， 为后续研究开辟了新方向。
>
> — 技术社区分析

项目提供了三个版本模型：14B 完整版（最高质量）、1.3B 轻量版（中等速度）、 以及 Turbo 加速版（2026 年 4 月更新，速度大幅优化）。 社区当前最关注的话题是 Wan2.1-I2V-14B 的加速方法， 以及是否会将加速技术整合到新版模型中。

安装提示

MagicTryOn 推荐使用 Python 3.12.9 + CUDA 12.3 + PyTorch 2.2 环境。 Flash Attention 需手动下载安装包（特定版本）， 国内用户可通过 HF_ENDPOINT 环境变量配置镜像加速权重下载。

04 Any2AnyTryon：ICCV 2025 的多场景统一框架

Any2AnyTryon 是 ICCV 2025 接收的论文成果，其核心贡献在于提出了一种**面向多种虚拟换装任务的统一框架**。 与传统方案针对不同任务（人→人、人→动漫、动漫→人等）分别训练不同模型不同，Any2AnyTryon 通过自适应位置嵌入（Adaptive Position Embeddings）技术， 实现了"一个模型解决所有换装场景"的目标。

该技术的关键创新在于位置嵌入机制的自适应设计。传统扩散模型使用固定的位置编码， 无法灵活适应不同体型、不同风格之间的空间映射关系。 Any2AnyTryon 的位置嵌入模块能够根据输入模型图和服装图动态调整空间注意力分布， 从而实现跨风格、跨体型的精确服装迁移。

- **基础架构**：基于 FLUX.1-dev 扩散模型
- **核心机制**：自适应位置嵌入（Adaptive Position Embedding）
- **数据资源**：配套 LAION-Garment 服装数据集
- **微调支持**：提供 LoRA 微调方案
- **论文**：ICCV 2025

该项目在 2026 年 6 月 24 日仍有更新。社区讨论焦点集中在微调步数的最佳实践 和服装重构（garment reconstruction）的质量优化上， 反映出研究者正在从"能不能换"向"换得更好"的方向深入。

> **技术提示**
>
> 测试时推荐使用 AutoMasker 生成分割掩码用于重绘（repaint）评估， 支持配队和未配对数据的混合评估策略。

---

05 StreetTryOn：迈向真实户外场景的换装基准

大多数虚拟换装研究依赖于经过精心布光和拍摄的室内数据集（如 VITON-HD、DressCode）。 然而，真实应用场景（如电商导购、社交媒体的实时换装滤镜）中的图像远非如此理想—— 背景复杂、光照变化、服装褶皱、姿态多样。

针对这一差距，StreetTryOn（WACV 2025）提出了首个面向真实户外场景的虚拟换装基准。 其核心贡献包括两个方面：

- **In-the-Wild 数据集**：收集真实拍摄场景下的换装测试数据， 包含复杂背景、自然光照、日常姿态等真实挑战。
- **跨域换装（Cross-Domain）评测**：定义并系统评估模型在域间迁移能力—— 即在理想数据集上训练的模型，能否在真实场景图像上仍保持换装质量。

> 在理想数据集上达到 SOTA 不代表能在真实场景中使用。 StreetTryOn 的提出填补了研究与应用之间的重要空白。
>
> — 虚拟换装社区分析

该项目在 2026 年 7 月 8 日仍有更新，表明社区正在持续完善这一基准的评测协议和数据覆盖。 对于关注虚拟换装落地应用的研究团队而言，StreetTryOn 的评测结果具有更高的参考价值。

---

06 技术架构演进：从 UNet 到 DiT 的范式转变

回顾虚拟换装技术的发展轨迹，可以清晰地看到一条架构演进的脉络。 早期的虚拟换装方案（如 CVVTON、DressCode）主要基于 CNN 编码器-解码器架构， 核心思路是将人体图和服装图通过卷积网络提取特征后拼接合成。

扩散模型时代（2023-2025 年初），VITON-HD、IDM-VTON、CatVTON 等项目将扩散过程引入换装任务， 使用 UNet 作为去噪网络的主干。这一阶段的关键进步在于：

- 扩散模型提供了更高质量的生成能力
- CLIP/DeepLab 等预训练模型提供了更好的语义理解
- 参数高效微调（PEFT）降低了训练成本

而 2025 下半年开始的最新趋势，则是扩散 Transformer（DiT）的全面崛起。 MagicTryOn 是最早将 DiT 应用于视频换装的代表项目。DiT 架构的核心优势在于：

- **自注意力机制**：能够捕捉全局上下文关系，避免 UNet 感受窗口的局限性
- **时序一致性**：通过自注意力自然建模帧间关系，减少视频闪烁
- **可扩展性**：模型规模可按需扩展（1.3B → 14B → 更大）

架构对比

UNet 类（CatVTON）：推理快、显存低，适合图像场景
DiT 类（MagicTryOn）：质量高、一致性佳，适合视频场景
混合方案（Any2AnyTryon + FLUX）：灵活性强，支持多场景统一

07 社区生态与未来趋势

2026 年 7 月的 AI 虚拟换装社区呈现出三个明显的趋势特征：

第一，轻量化与高性能的持续博弈。CatVTON 以不到 8G 显存的推理门槛开创了轻量级换装的新标准。 而 MagicTryOn 则提供了从 1.3B 到 14B 的完整模型谱系。 未来一年，我们可能会看到更多在 4-6G 显存消费级 GPU 上运行的 高质量换装模型出现。

第二，从图像到视频的自然延伸。MagicTryOn 是最早探索视频换装的项目之一， 其采用的全自注意力时空建模方法为后续研究提供了参考方向。 随着 Wan2.1-I2V-14B 等视频生成模型的成熟， 视频换装的质量有望迎来质的飞跃。

第三，生态系统的逐步完善。Any2AnyTryon 提供了 LAION-Garment 数据集和 LoRA 微调工具，StreetTryOn 构建了真实场景评测基准， 这些基础设施的累积为整个社区的长期发展奠定了基础。

- **电商场景**：虚拟试衣将成为电商平台的标准配置， 降低退货率、提升转化率
- **社交娱乐**：基于轻量模型的实时换装滤镜， 在移动端实现即时换装体验
- **设计工具**：设计师使用 AI 换装快速验证设计方案
- **教育科研**：开放基准和基准数据集持续推动学术研究

值得关注

MagicTryOn 社区正在讨论将加速技术应用到新版模型， Wan2.1-I2V-14B 的百倍加速方法已有初步探索， 可能成为下一阶段的关键突破点。

08 结语

从 ICLR 2025 的 CatVTON 到 WACV 2025 的 StreetTryOn， 从 ICCV 2025 的 Any2AnyTryon 到 vivo 团队的 MagicTryOn， 2026 年上半年虚拟换装领域的学术产出和开源贡献均达到了新的高度。

技术层面，扩散 Transformer 正在取代 UNet 成为新标准； 应用层面，从图像到视频、从理想数据集到真实场景的跨越正在加速； 生态层面，开源项目、数据集和评测基准的累积为长期发展提供了坚实基础。

本期日报基于 GitHub 开源项目数据实时采集生成， 数据来源包括 GitHub API、项目 README 及社区 Issue 动态。 本文覆盖了五个最具代表性的项目，按 Star 数与活跃度排序： CatVTON、MagicTryOn、Any2AnyTryon、StreetTryOn、TryOnDiffusion。 后续将持续跟踪虚拟换装技术演进，欢迎关注。
