

青稞AI

虚拟试衣

2D/3D

Sam (Daiheng Gao)

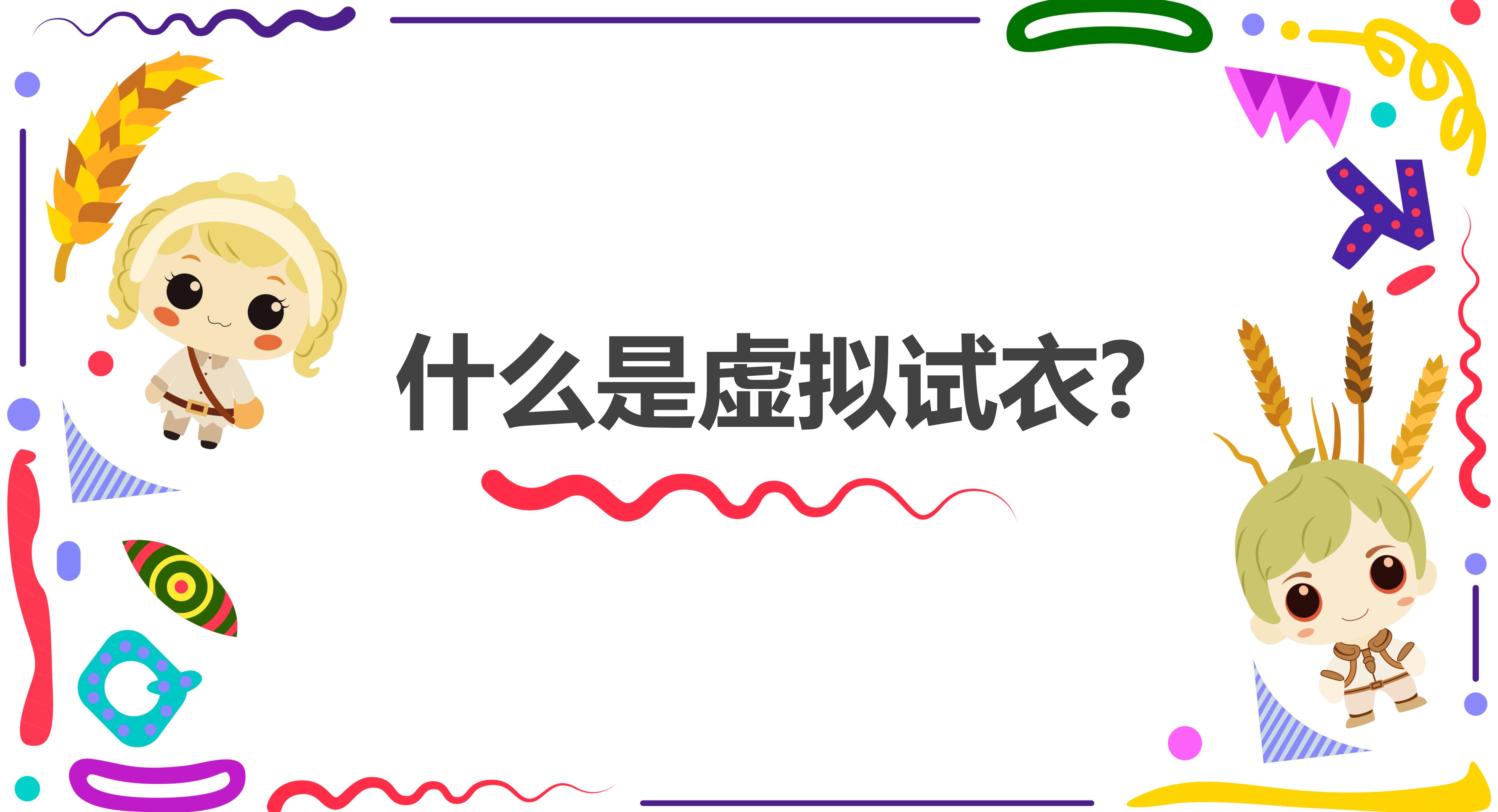
2024.11.08



- 01 虚拟试衣是什么
- 02 Diffusion
- 03 我的工作



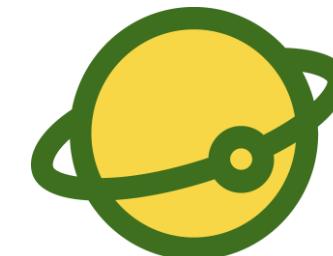
什么是虚拟试衣？



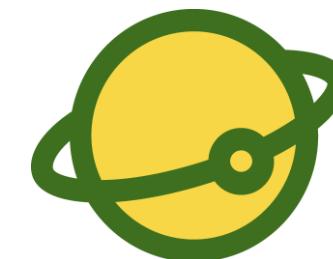


青稞AI

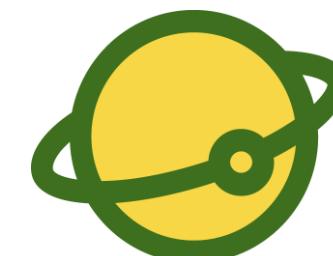
虚拟试衣：



(目的) 让模特穿上目标服饰



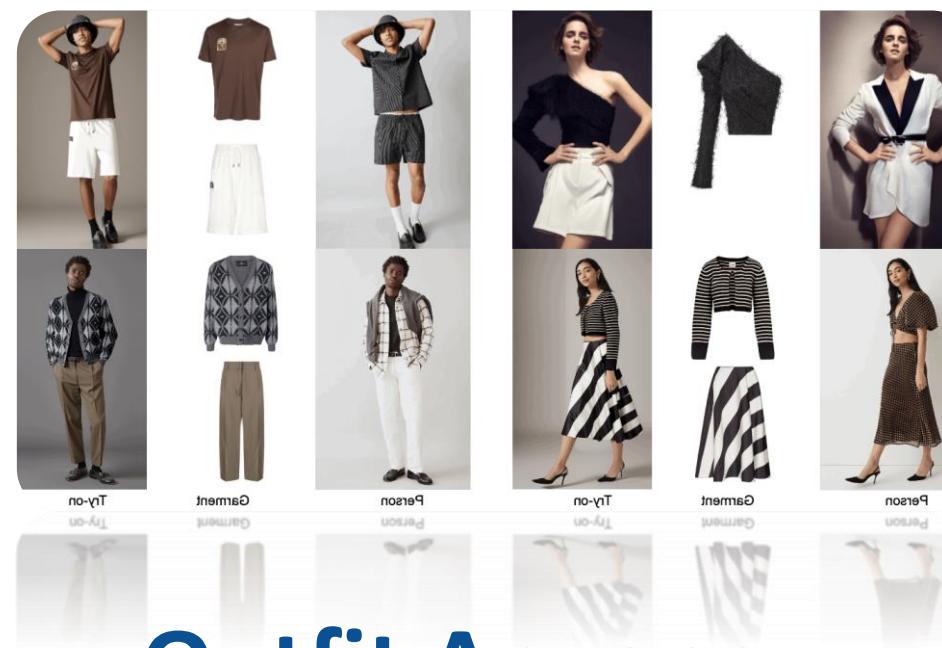
(类型) 2D、3D、视频



(难点) 服饰**多样性&一致性**
模特**身材**
布料解算
推理要求

SIFU

CVPR2024 Highlight



Outfit Anyone

Top #20 of 200,000+
Hugging Face spaces



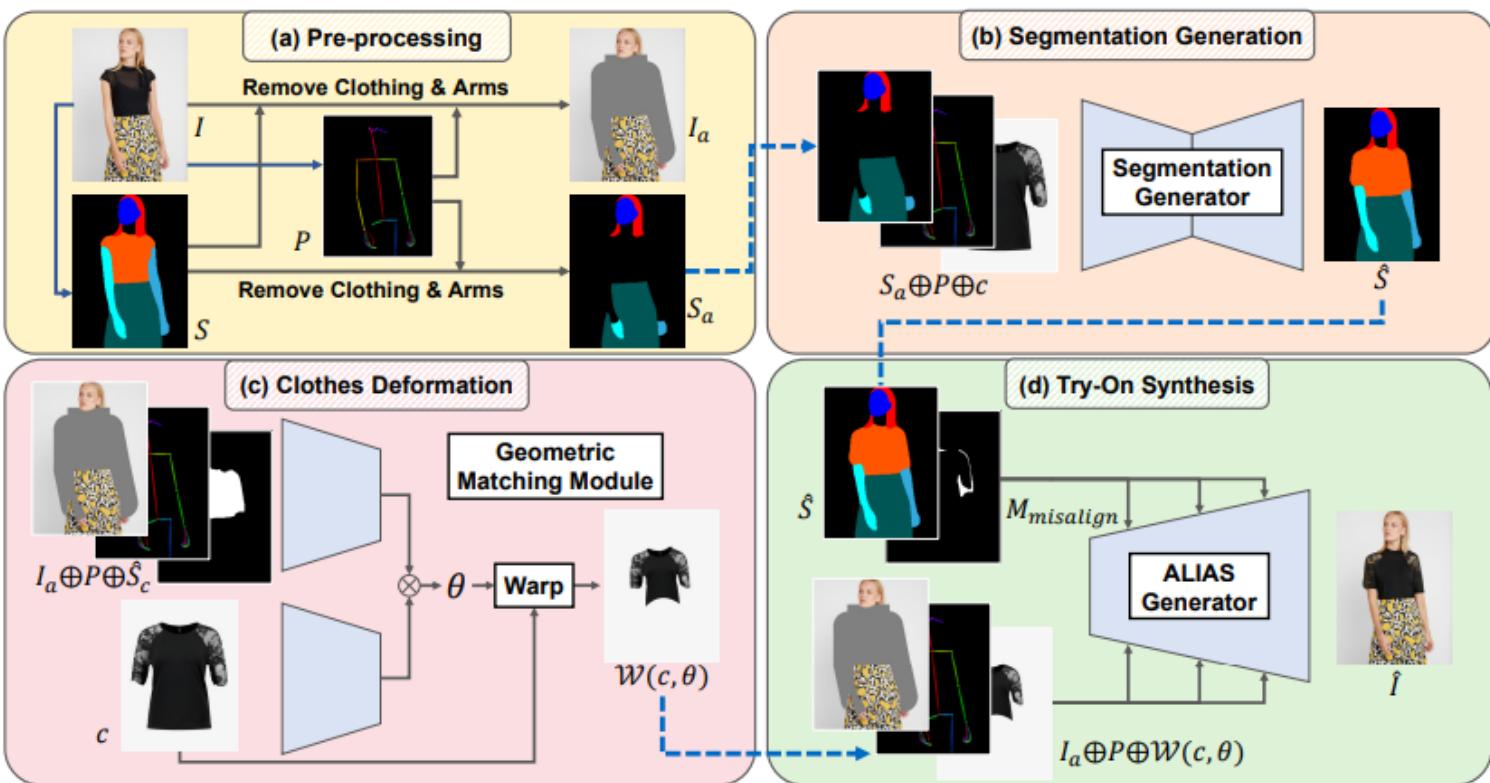
- [1] <https://river-zhang.github.io/SIFU-projectpage/>
- [2] <https://humanaigc.github.io/outfit-anyone/>
- [3] <https://ihe-kaii.github.io/DressCode/>



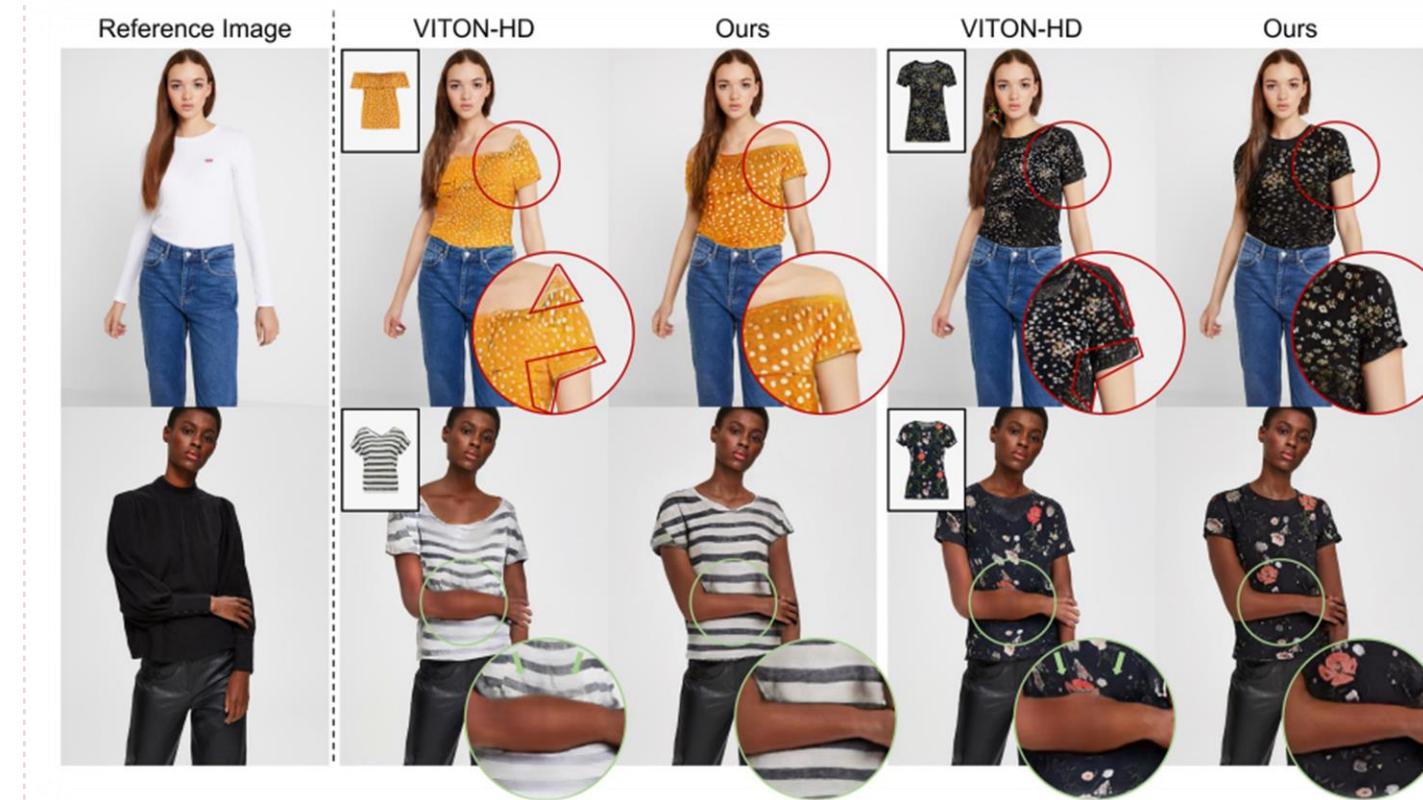
主流技术路线：



基于图像：从早期的GAN



VITON-HD: CVPR2021



HR-VITON: ECCV2022

主流技术路线：



基于图像：从早期的GAN到近期的Diffusion

TryOnDiffusion: A Tale of Two UNets

Luyang Zhu^{1,2*} Dawei Yang² Tyler Zhu² Fitsum Reda² William Chan²
 Chitwan Saharia² Mohammad Norouzi² Ira Kemelmacher-Shlizerman^{1,2}
¹University of Washington ²Google Research



Figure 1. TryOnDiffusion generates apparel try-on results with a significant body shape and pose modification, while preserving garment details at 1024×1024 resolution. Input images (target person and garment worn by another person) are shown in the corner of the results.

Try On Diffusion: CVPR2023

OutfitAnyone: Ultra-high Quality Virtual Try-On for Any Clothing and Any Person

Ke Sun^{1*} Jian Cao^{1*} Qi Wang¹ Linrui Tian¹ Xindi Zhang¹ Lian Zhuo¹
 Bang Zhang¹ Liefeng Bo¹ Wenbo Zhou² Weiming Zhang² Daiheng Gao^{2,3}
¹Intelligent Computing, Tongyi, Alibaba Group ²USTC ³Formation.ai

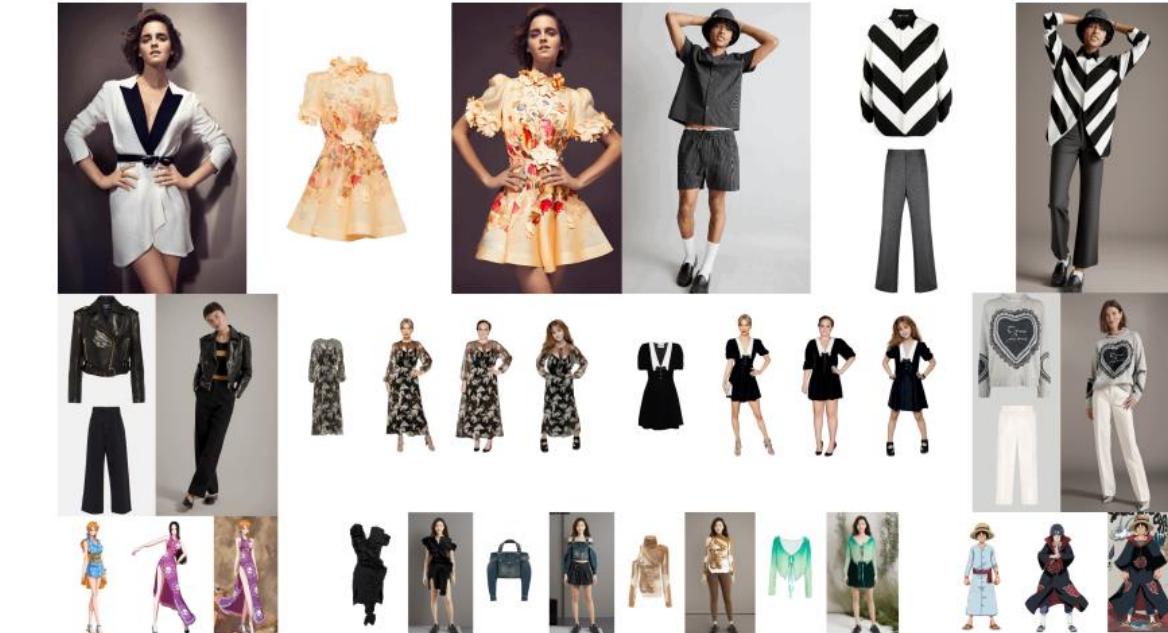


Figure 1. We introduce **OutfitAnyone**, a diffusion-based framework for 2D Virtual Try-On. By far, it has garnered over 5,000 stars on GitHub and ranked within the top 20 among all the Hugging Face spaces.

Outfit Anyone: Arxiv 2024

本质是一种图像编辑

主流技术路线:



基于3D：从早期的固定3D Mesh(SMPL)

CLOTH3D: Clothed 3D Humans

Hugo Bertiche^{1,2}, Meysam Madadi^{1,2}, and Sergio Escalera^{1,2}

¹ Universitat de Barcelona, Spain

² Computer Vision Center, Spain
hugo_bertiche@hotmail.com

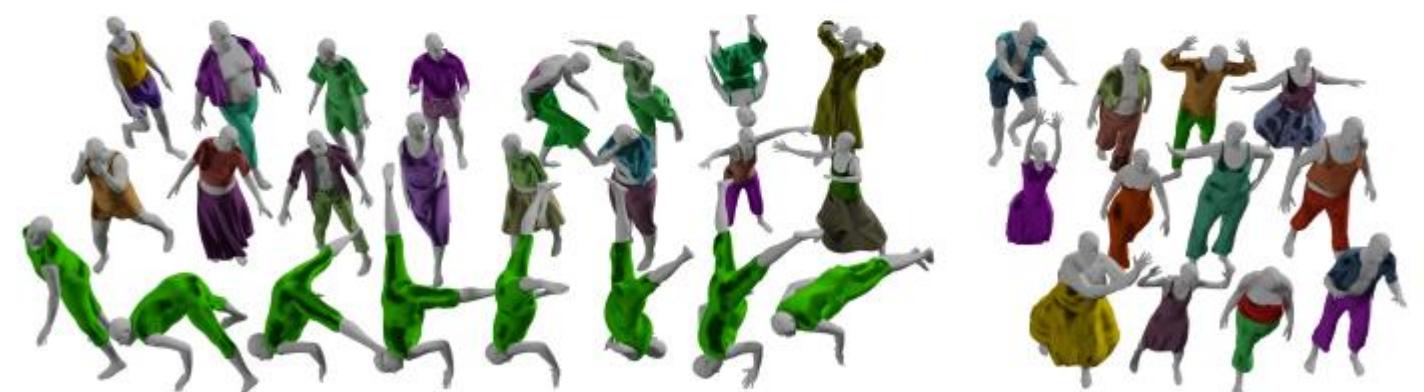
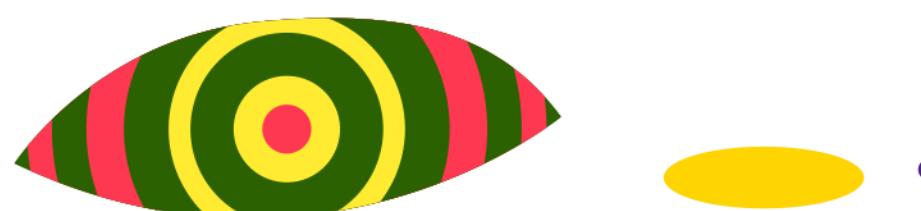
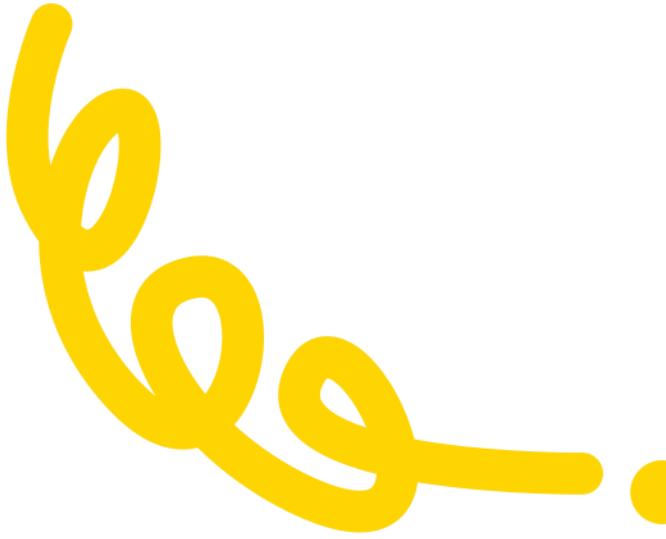


Fig. 1: Left: CLOTH3D³ is the first big scale dataset of animated clothed humans. It contains thousands of different outfits and subjects, high variability of poses and rich cloth dynamics. Right: generated 3D garments with proposed GCVAE.



Cloth3D: ECCV2020

Pix2Surf: CVPR2020



主流技术路线：



基于3D：从早期的固定3D Mesh(SMPL)，到近几年的NeRF，3D GS，DMTet的方案

HumanGaussian: Text-Driven 3D Human Generation with Gaussian Splatting

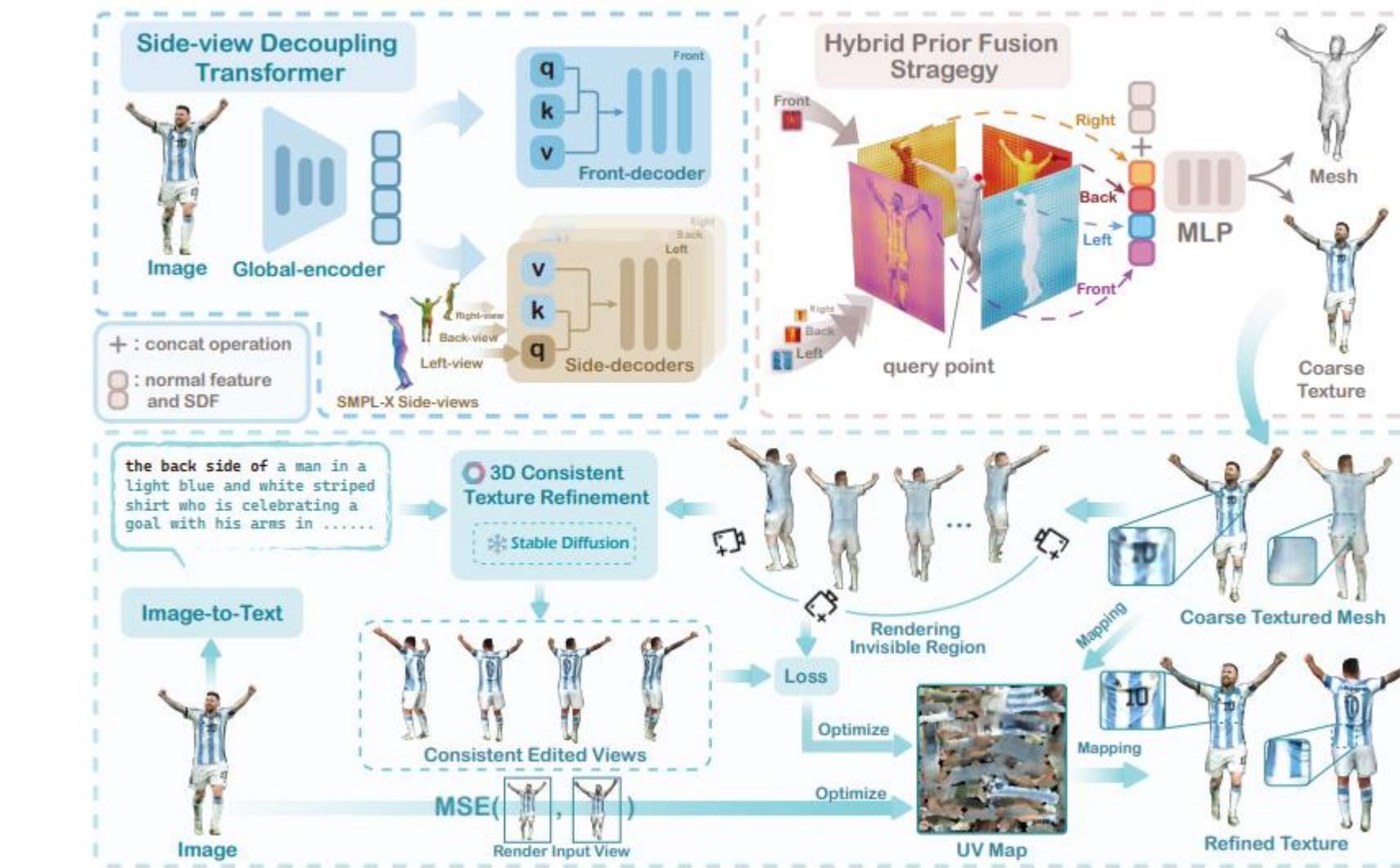
Xian Liu¹, Xiaohang Zhan², Jiaxiang Tang³, Ying Shan², Gang Zeng³, Dahua Lin¹, Xihui Liu⁴, Ziwei Liu⁵

¹CUHK ²Tencent AI Lab ³PKU ⁴HKU ⁵NTU

Project Page: <https://alvinliu0.github.io/projects/HumanGaussian>

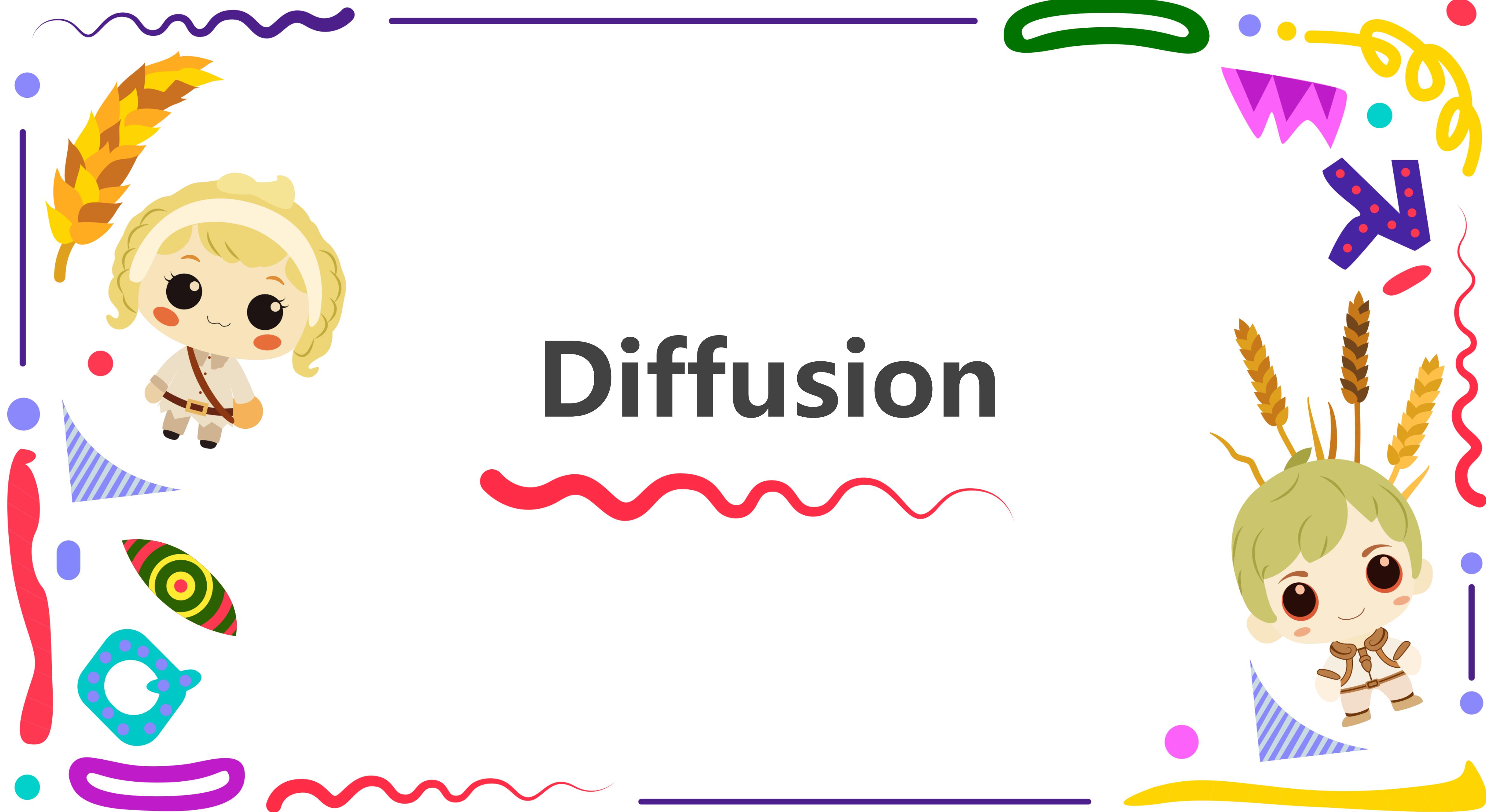


Human Gaussian: CVPR2024

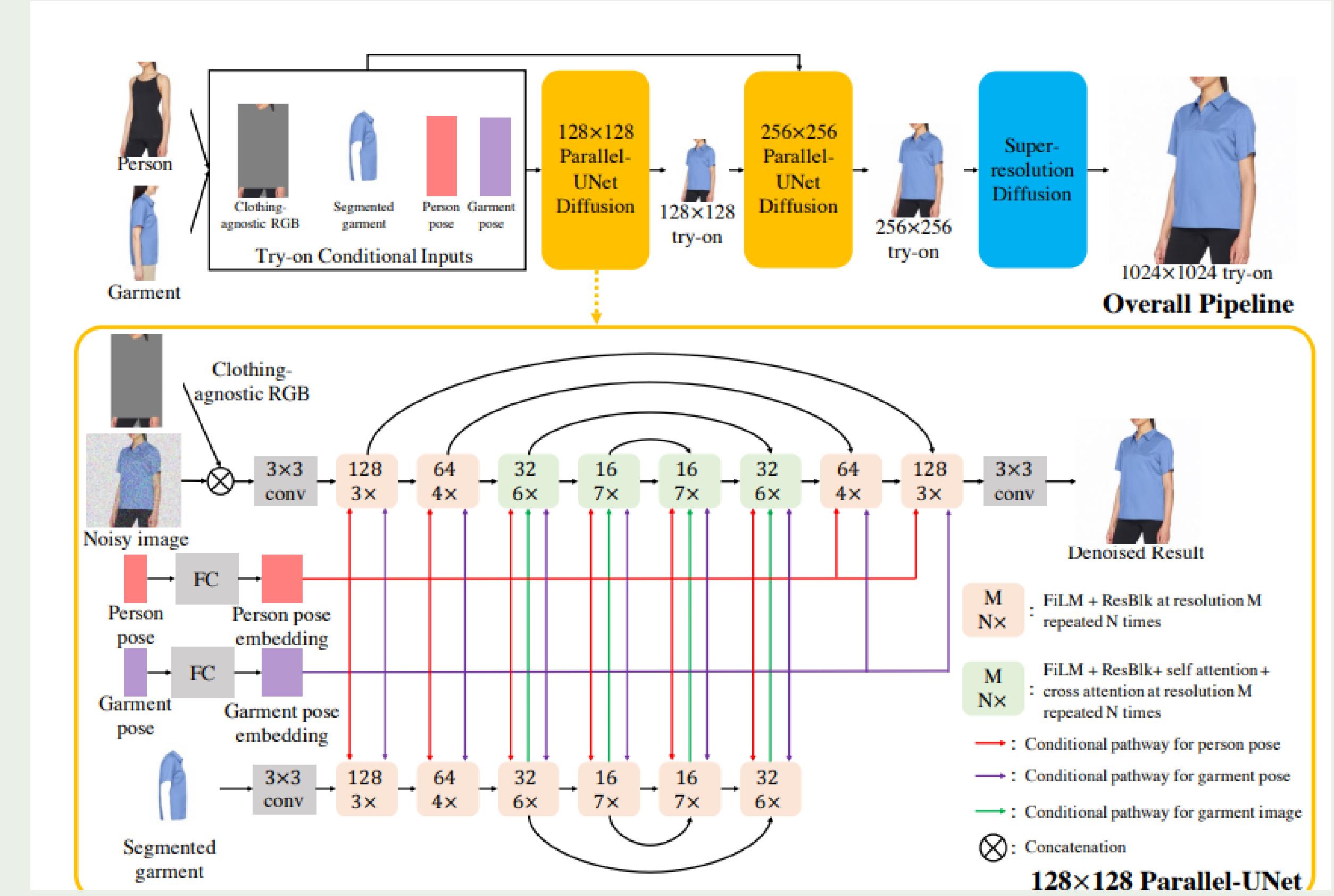


SIFU: CVPR2024

Diffusion



TryOnDiffusion (CVPR2023)

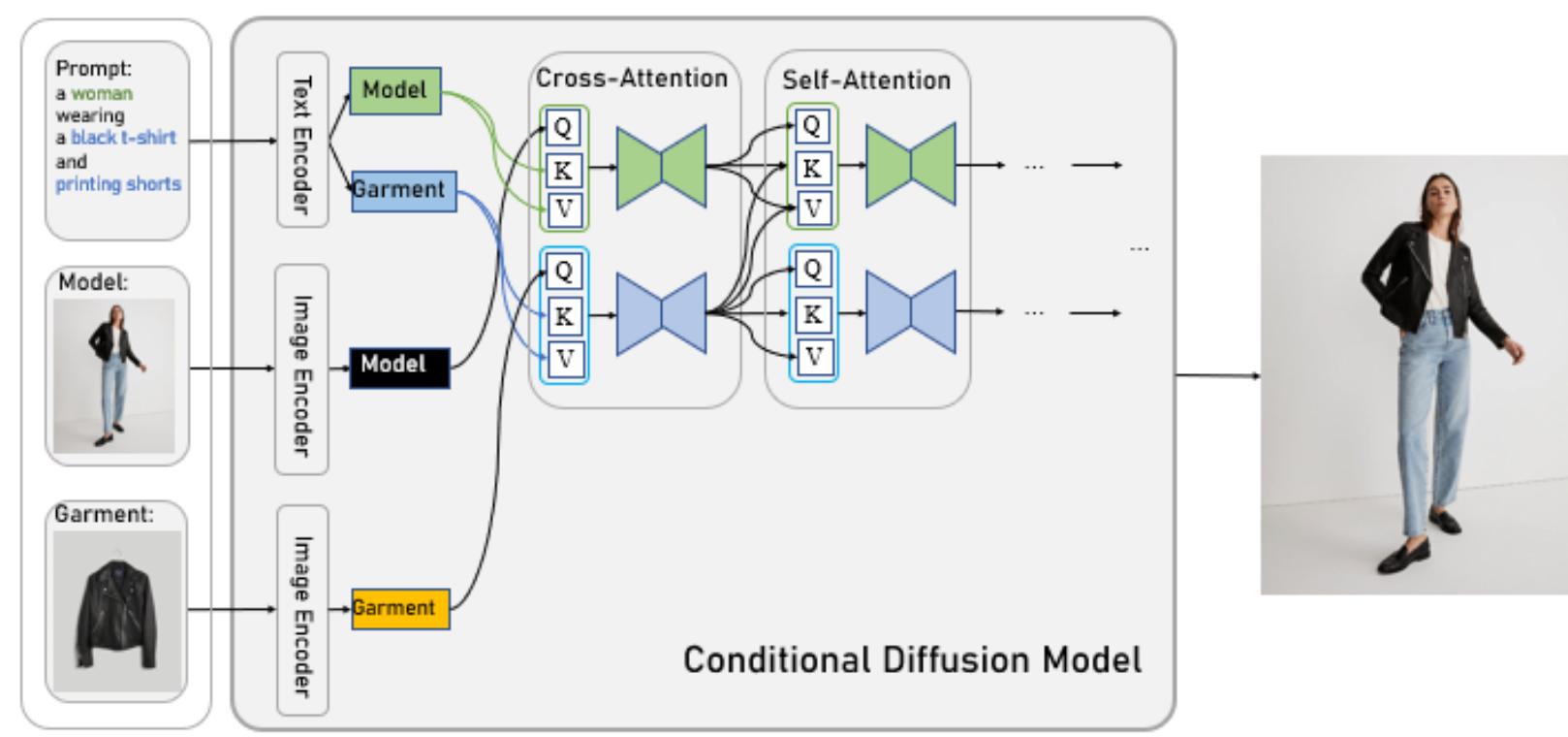


Google的TryOnDiffusion开创了用扩散模型做虚拟试衣这个领域

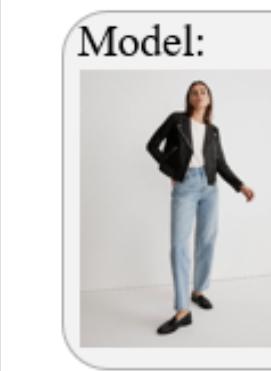
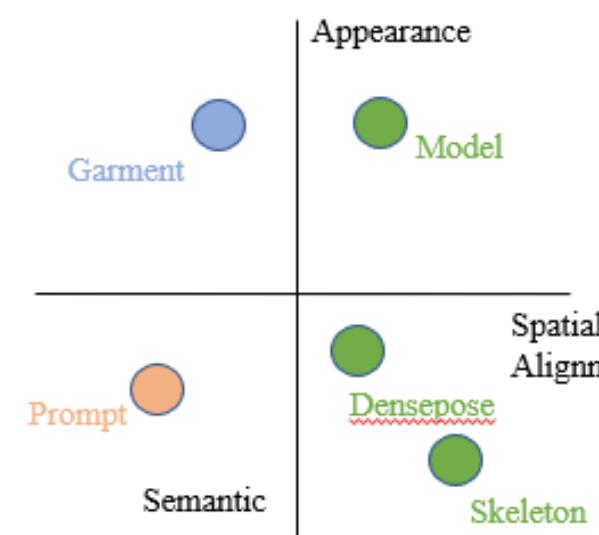
TryOnDiffusion->Outfit Anyone

在扩散模型基础上, 针对服装和模特设计双流U-Net结构, 并在self attention层对特征进行混合, 即可实现绝佳的试衣效果。

Dual-Path Conditional Diffusion Model



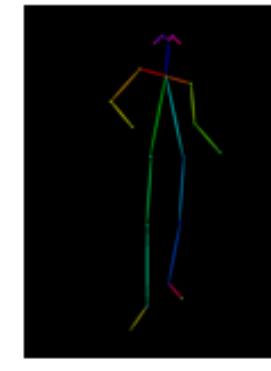
- 输入信号的区分: **高效整合拆分, 降低模型复杂度**
 - 模特信号:** 相互之间存在空间对齐关系
 - 服饰信号:** 与模特空间不对齐, 需要保持纹理细节
 - 文本信号:** 与模特空间不对齐, 语义层次信息
- 输入信号 -> Embedding的转化: 因材适配, 提高表征能力
 - 模特信号:** VAE + SD
 - 服饰信号:** VAE + SD, 保持纹理细节, 与模特Embed特征空间一致
 - 文本信号:** Text-Encoder
- Embedding间的融合方式: 灵活配置, 减轻模型改动
 - 模特信号:** Channel变换
 - 模特 & 服饰:** Self-Attention
 - 模特 & 文本:** Cross-Attention
- 优势:
 - 双流SD模型, 既保证服饰特征的高质量保留, 也保证服饰特征的高质量还原
 - 几乎无非训练参数, 且模特&服饰特征空间一致 -> 快速收敛 (6k可初见试衣效果)



Model:



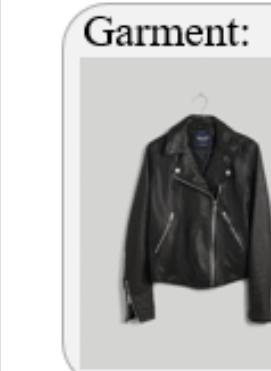
保持模特ID:
通过mask, 区分
需要保留的部分
和需要生成的部分



保持模特姿态:
通过skeloton, 保
持模特的姿态



保持模特体型:
通过densepose,
smpl等, 保持模
特的身材和姿态



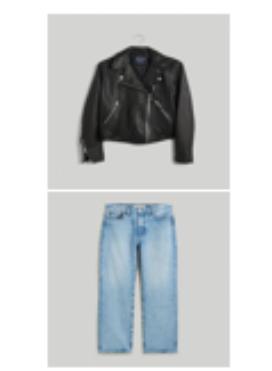
Garment:



单件上衣: 结合
保持mask, 可做
到任意上衣和固
定下衣搭配试穿

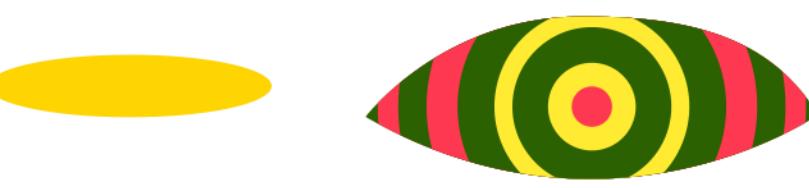


单件下衣: 结合
保持mask, 可做
到任意下衣和固
定上衣搭配试穿

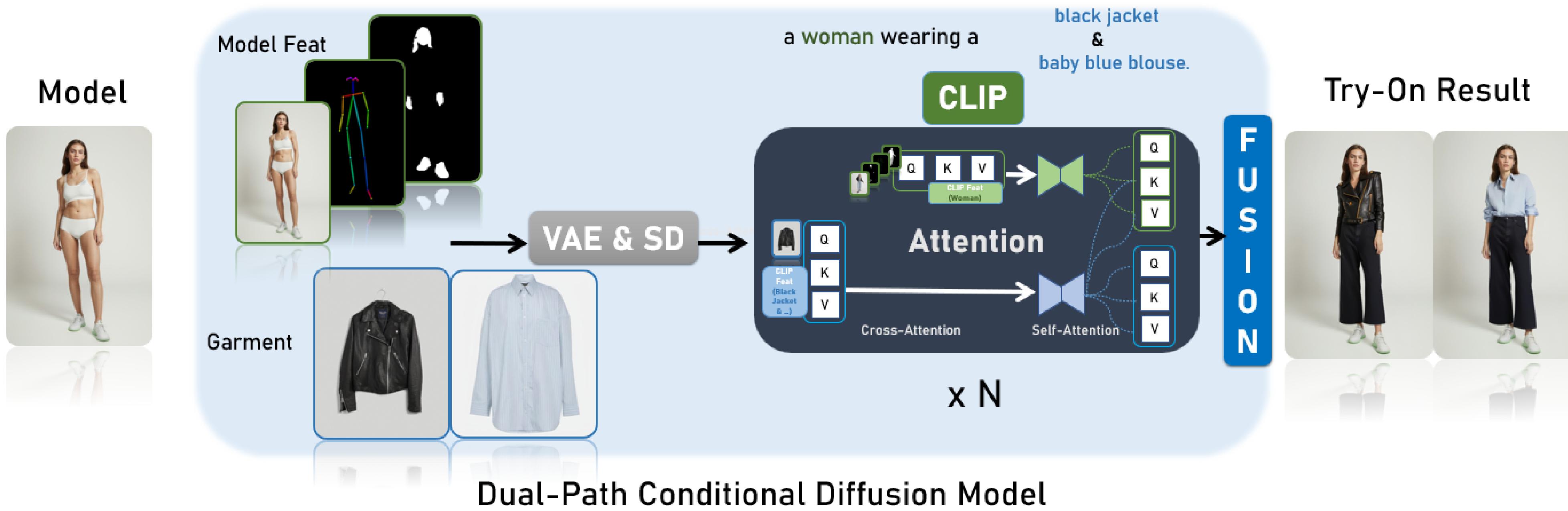


套装: 支持套装的
一次性试穿

OutfitAnyone



精准的数据预处理和高质量数据是成功的关键



总结

 **能力:** 扩散模型对数据分布的梯度进行建模, 生成效果显著优于传统的生成模型, 因此, 图像/视频领域, **完全转向扩散模型的趋势是不可阻挡的。**

 **市场:** 从TryOn Diffusion, 到Outfit Anyone和快手的KColors虚拟试衣, 目前的虚拟试衣技术已经进入商业验证环节, **实时性和效果开始达到要求。**

 **未来:** 虚拟试衣作为重要的原子化技术, 可以极大的赋能传统B端市场, 预计2025年, **有超过1亿元美金的市场规模。**



我的工作



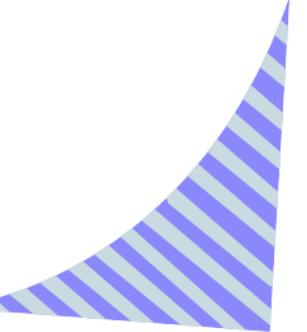
Cloth2Tex (3DV 2024)



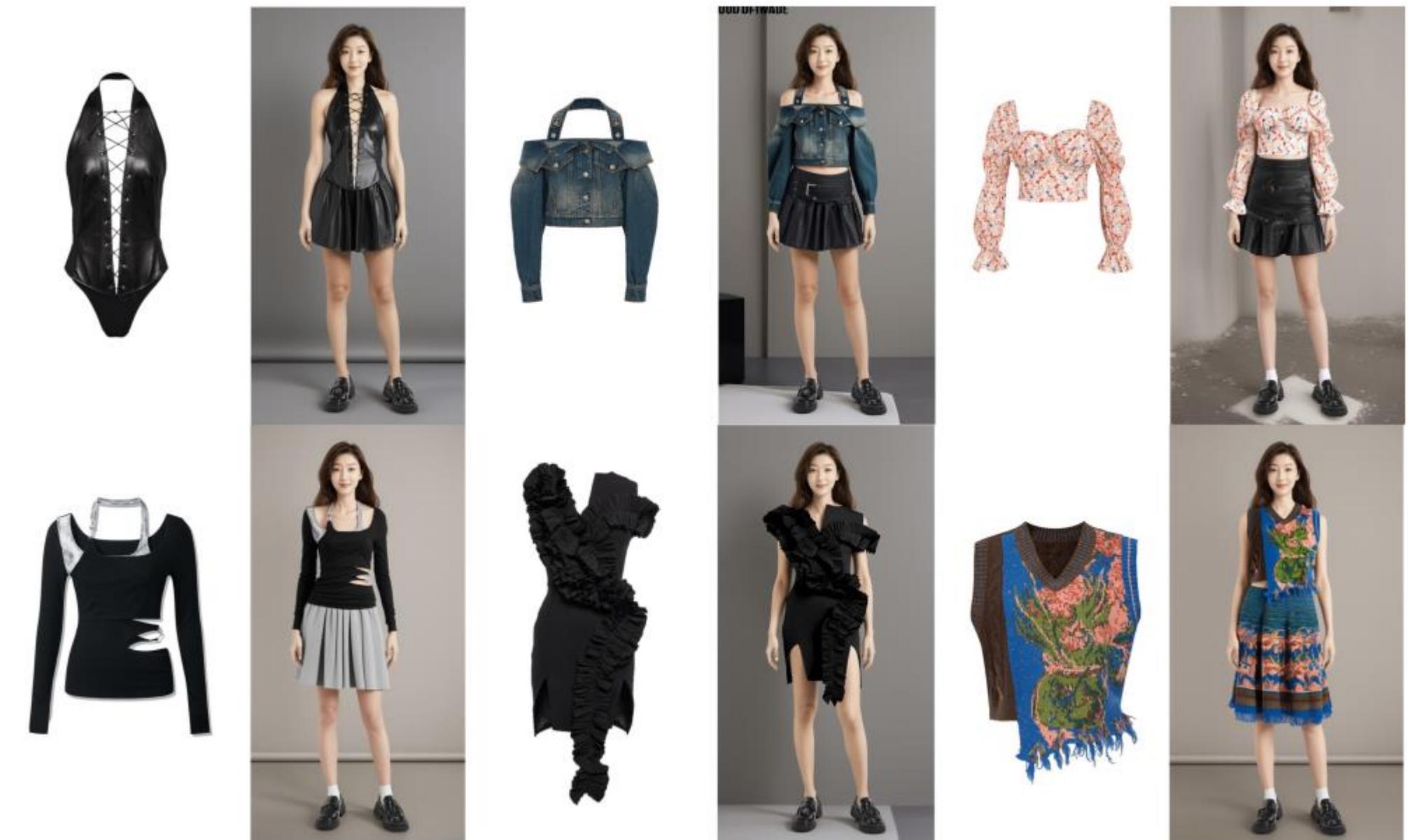
Outfit Anyone (Arxiv 2024)



效果



效果



动态效果

Outfit Anyone + Animate Anyone



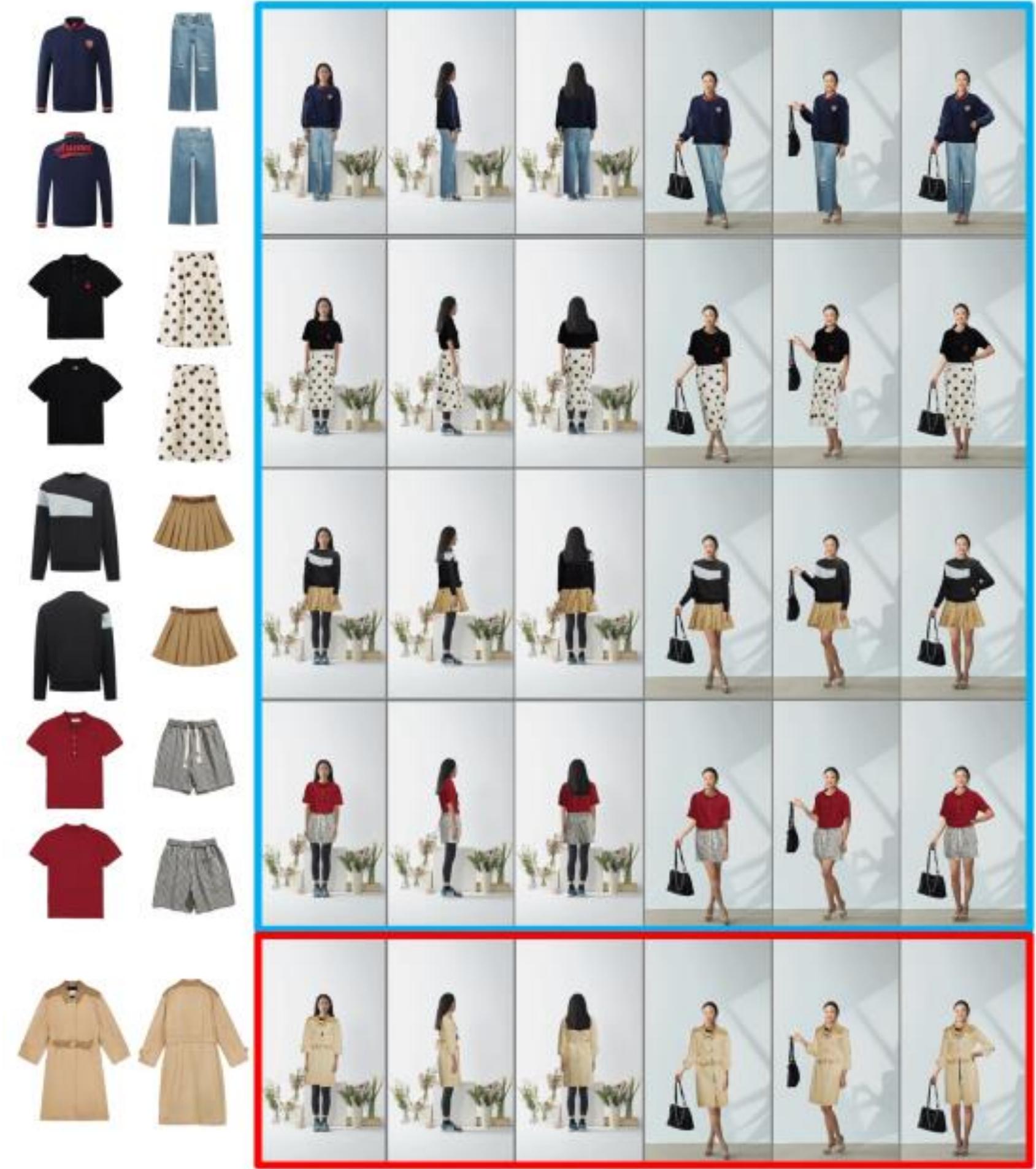
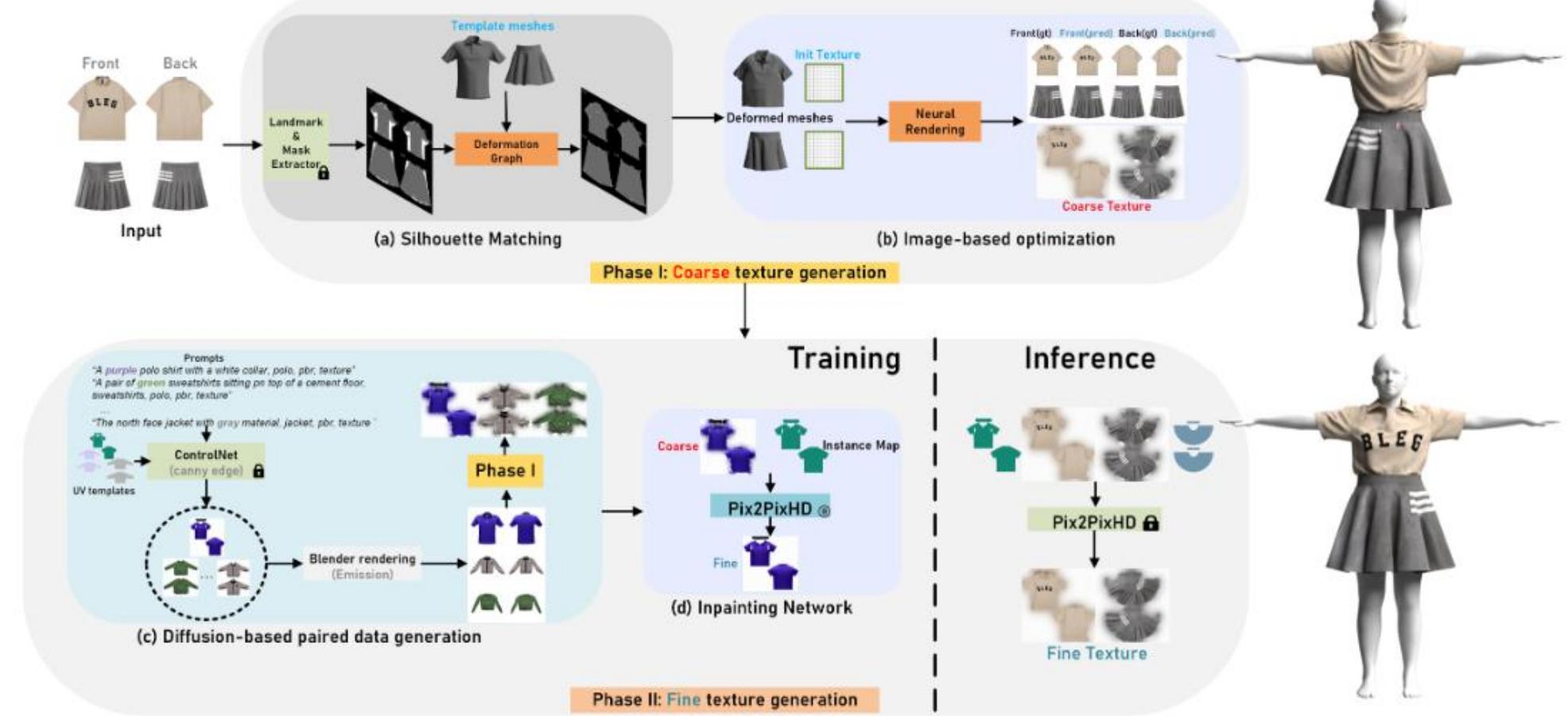
Person

Garment

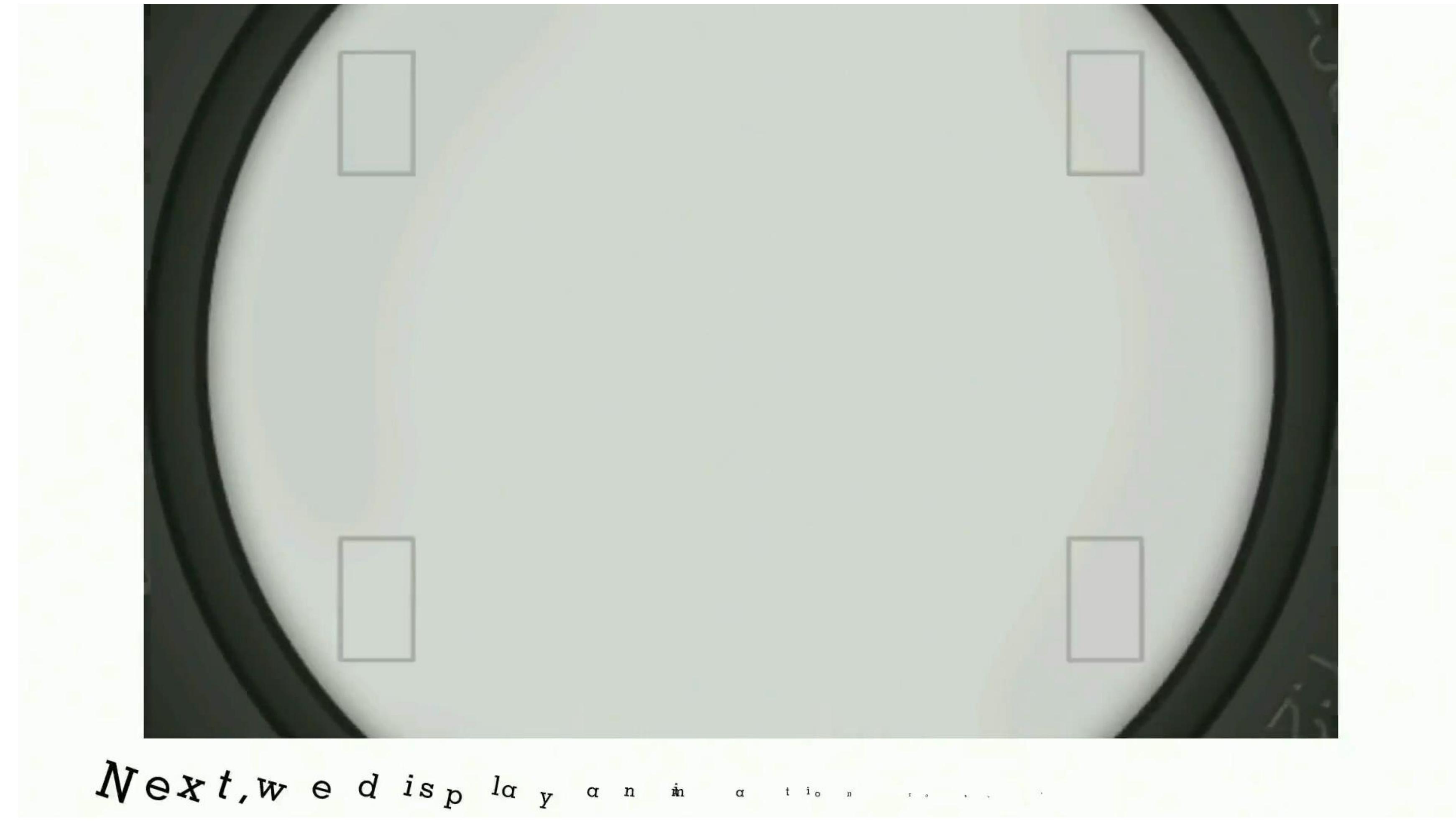
Try-on

Animation

效果



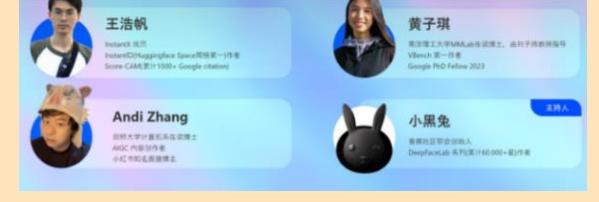
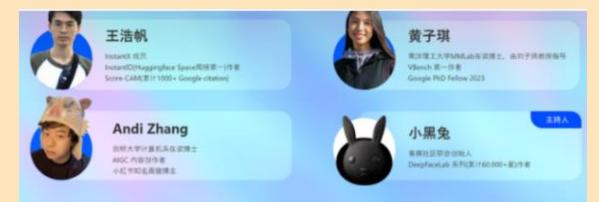
动态效果



青稞是谁？



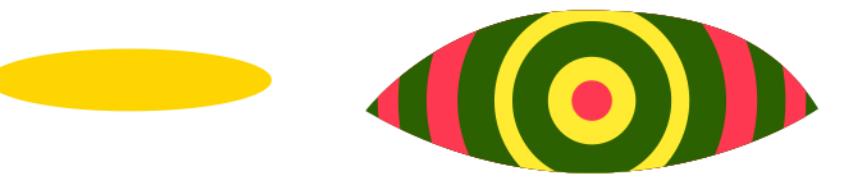
a) 模型成果

模型	类型	备注	公开课
Outfit Anyone (Sota试衣模型)	Virtual Try-on, Diffusion, 图生图 https://huggingface.co/spaces/HumanAIGC/OutfitAnyone	1,000,000+ API calls. #14 of all huggingface spaces (200,000+). 5,500+ github stars.	
Instant-ID (人脸风格化)	Personalized Image Synthesis, Diffusion, 图生图 https://huggingface.co/spaces/InstantX/InstantID	#9 of all huggingface spaces (200,000+). 10,000+ github stars.	
LLaMA-Factory (高效LLM微调框架)	Fine tuning, LLM, WebUI版大模型微调 (100+大模型支持) https://github.com/hiuga/LLaMA-Factory	29,300+ github stars. ACL 2024	
Intern-VL (GPT-4O开源平替)	GPT4O, LLM, 最强多模态 开源大模型 https://github.com/OpenGVLab/InternVL	5,100+ github stars. CVPR 2024	

超牛技术成果&模型开发能力

以2024年为例, 青稞的开发者们开发了很多知名项目, 在 Github&Huggingface 都获得了很大的影响:

如Huggingface space 排名前20的 **OutfitAnyone** 和 **Instant-ID**, 对标GPT-4O的书生万象,最好用的 webUI 版 LLM 微调框架 **LLaMA-Factory** 等.



b) 公开课直播

分享前沿技术&吸引精准流量

2024年5月开始密集启动:

通过邀请清华、北大、Stanford、UCLA、MIT、NTU 等学界专家，以及阿里、智谱等业内专家分享最新 AI 技术研究成果，保持和最先进生产力的密切关系，为社区的保持活力，并借此促成一些咨询和技术合作项目。

The grid displays 18 AI research talks, each with a thumbnail, title, subtitle, and direct link. The talks are categorized into three rows:

- Row 1:**
 - mPLUG-Ow13: 探索长序列模型架构的通用多模态大模型
 - SGLang v0.2: 面向LLM和VLM的快速、高效通用服务引擎
 - S-LoRA: 实现多LoRA大模型的高效并行化推理
 - AWQ: 激活值感知的LLM低效权重量化
 - InternVL 2.0: 通过渐进式策略扩展开源多模态大模型的性能边界
 - SEED-Story: 生成长篇图文故事的多模态大型语言模型
- Row 2:**
 - YOLO-World: 基于视觉语言模型的实时开放词汇物体检测
 - 经典大模型公开课第二季
 - MiniCPM-V: 端侧可用的GPT-4V级多模态大模型
 - Mobile-Agent: 基于多模态Agent架构的手机智能体
 - LLaMA Factory: 从预训练到RLHF, 大模型高效训练框架
 - VillagerAgent: 减少幻觉、提高任务分解效率的多智能协作体框架
- Row 3:**
 - LLaMA Pro: 扩展Transformer块优化的大型语言模型继续预训练
 - PiSSA: 收敛快、误差小的大模型参数高效微调方法
 - 具身多模态大模型的视觉表征预训练研究
 - VideoBooth: 文本和图像提示共同驱动的视频生成
 - MixEval: 混合评测数据集来拟合大语言模型的人类评估
 - 实时渲染3DGS中的反走样及逆渲染应用

Each card includes a small profile picture of the speaker, their name, title, and the date of the live broadcast.

公开课 <https://qingkelab.github.io/talks/>

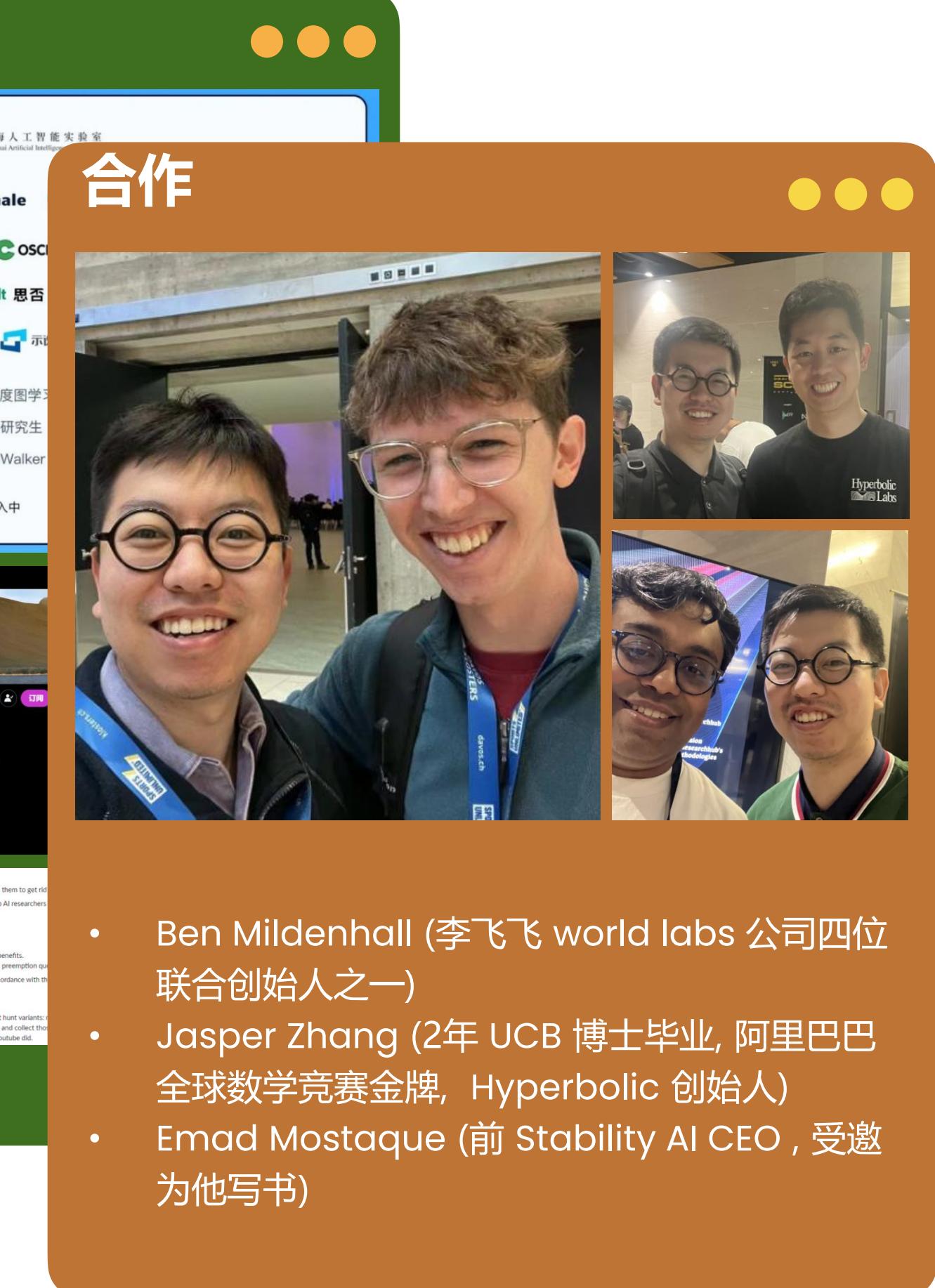


青稞AI



c) 活动





- Ben Mildenhall (李飞飞 world labs 公司四位联合创始人之一)
- Jasper Zhang (2年 UCB 博士毕业, 阿里巴巴全球数学竞赛金牌, Hyperbolic 创始人)
- Emad Mostaque (前 Stability AI CEO , 受邀为他写书)

分享AI进展&吸引广泛流量

(国内)通过和 OpenBMB (面壁智能), 智谱, 商汤等机构进行合作和共同宣发, 推广品牌.

(AI 社区)2024年8月开始, 配合 AK 和王铁震, 提案关于 Huggingface space/model/dataset 等板块对顶级开发者的激励计划.

(国外)和海外顶级 AI 公司进行合作, 奠定良好声誉



青稞AI

THANK
YOU

青稞, AI开发者社区



samuel.gao023@gmail.com



314913739

