# 虚拟换衣系统科技报告

**系统总体架构**

随着互联网的普及,传统的线下购物方式不再是购物的主体,线上购物平台不断发展。近几年来，电子商务的发展势头十分迅猛，服装购物作为电商平台提供的主要服务之一，用户对虚拟试衣的需求越来越迫切。在这个背景下,虚拟换衣技术应运而生,人们足不出户就可以在虚拟试衣平台上体验到不同服装的试穿效果。

现阶段传统的虚拟试衣平台通常存在几个问题： 第一，3D建模成本大，计算复杂，模型推理速度慢。第二，需要用户自己手动输入身高体重等数据，得到的三维人体模型不具有针对性，很难让用户有真实代入感。第三，平台上提供的衣服品类有限，一般不支持用户自己提供的衣服照片。

针对目前传统的虚拟换衣平台所面临的问题和不足，本研究旨在深度学习前沿技术设计和实现一套用户可以独自操作并获得较好的换衣结果的虚拟换衣系统。通过该系统用户可以上传自己想要试穿的衣服照片，利用摄像头捕获人物图像后上传到系统后台进行处理，进行虚拟换衣相关算法操作后返回结果到显示屏幕上，让人们更加便捷的实现虚拟试穿。

**需求分析**：

1. 上传想要试穿的衣服。用户可以上传想要试穿的衣服到系统上，或者选择系统提供的参考样例衣服。在用户选择好试穿衣服之后，该衣服会显示在屏幕上。
2. 拍取人像。用户通过本地摄像头进行拍照，将拍取的人像照片上传到系统上，该人像会显示在屏幕上。
3. 虚拟换衣。在获得上述人像和试穿衣服后，输到后台进行算法处理，得到的换衣结果中，人像照片中的人穿着所选择的要试穿的衣服，保留未选择的其余部分，不参与换衣过程。该换衣结果返回到屏幕上。

**解决方案**

本研究的虚拟换衣系统示意图如图1所示，采集设备主要用于采集人像，显示设备主要用于显示输入和输出结果，客户端用于将采集到的人像上传到系统，以及用户和后台的交互。服务器用于处理模型，包含对获取到的人像基于Openpose [1] 2D 姿态点检测算法进行关键点提取; 对人像利用人体解析算法PGN[2] 进行语义分割,得到的分割结果包含手臂、腿、人脸、头发、衣服、裤子等20个分割类别; 根据提取出来的关键点和分割结果对所选择试衣图像进行TPS薄板样条[3]扭曲和人物图像融合。人物图像穿着试穿衣服的结果会被返回于显示屏幕上。

[1] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, “Open- pose: Realtime multi-person 2d pose estimation using part affin- ity fields,” *TPAMI*, 2019.

[2] Gong K, Liang X, Li Y, et al. Instance-level human parsing via part grouping network[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 770-785.

[3] Bookstein F L. Principal warps: Thin-plate splines and the decomposition of deformations[J]. IEEE Transactions on pattern analysis and machine intelligence, 1989, 11(6): 567-585.

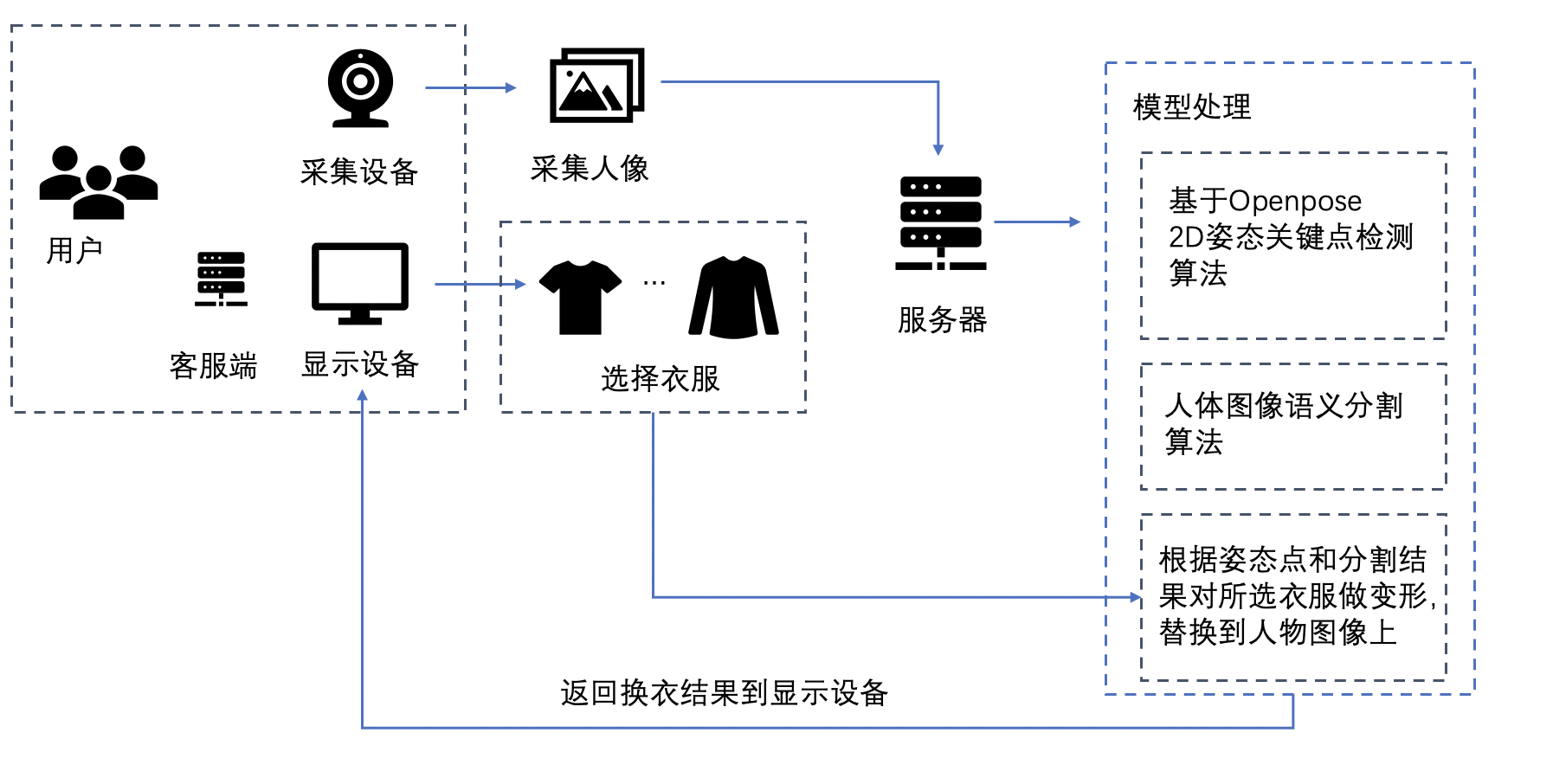


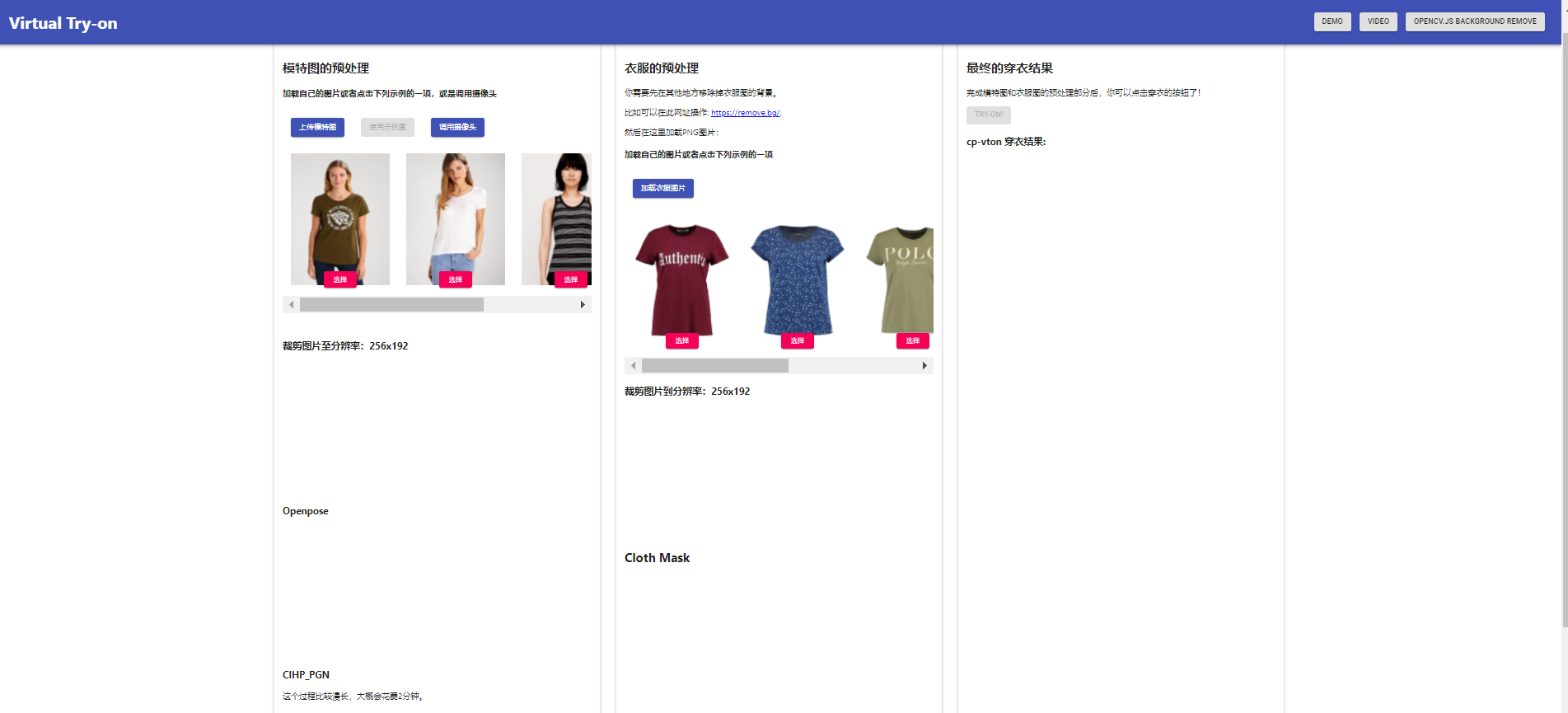
图1

**系统详细设计与实现**

**2.1 web前端**

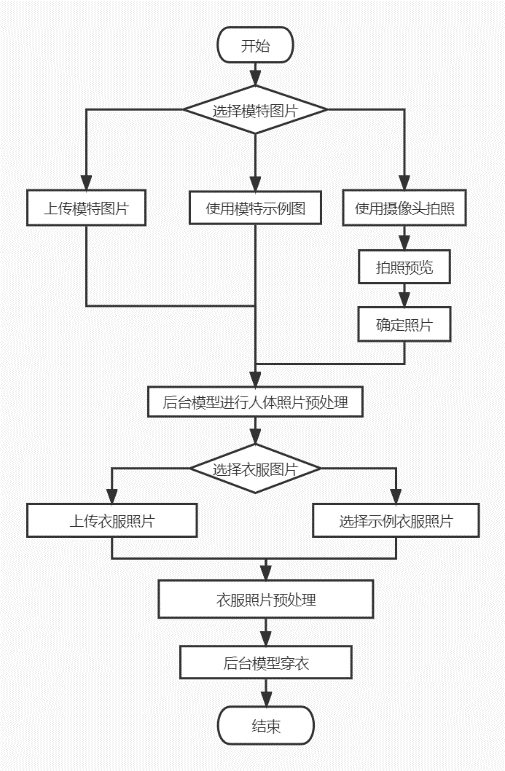
**2.1.1** 前端实现框架

前端主要使用了react开发框架，同时选用多种工具搭配构建项目，如Vue Router、Vuex和Axios等，主要通过HTML、JavaScript和CSS语言进行开发。前端的组件使用Material-UI组件库。该组件库基于React实现，实现了谷歌Material Design设计规范。Material-UI组件是相互独立的，自支持的，工作时仅注入当前组件所需要的样式。前端页面的跳转与局部刷新完全由路由控制，可以按需加载。多个组件的共享数据由状态管理工具进行同步。前端通过Web API与后端交互，并使用WebSocket实现动态数据交互。



**2.1.2 用户操作流程**

web前端系统的操作流程如图所示，进入页面之后，用户需要选择一种方式提供模特图片，可选的方式包括上传本地模特图片、使用前端系统已有模特图片和使用摄像头拍照。选用摄像头拍照时，需要拍照后确认图片无误后再上传。获得人体照片后，信息会被传递到后台，后台模型进行照片的预处理。第二步中，用户需要选择衣服图片，方式包括上传本地衣服照片和选择系统示例衣服照片。选择完毕后，前端系统可以完成衣服图片的预处理。第三步中，用户点击穿衣后，由后台系统结合前两步的结果，进行虚拟穿衣。前端会随之展示虚拟穿衣结果。



2.1.3 前端重要功能细节

前端中较为重要的功能包括本地文件上传和调用本地摄像头采集照片。

在本地文件上传中，处理流程为客户端添加事件，创建fileReader实例，用户选择本地图片后，函数获取本地图片路径，根据路径读取文件，读取成功后用axios，在封装的接口中将本地文件上传到后端。

在调用本地摄像头采集照片功能时，首先使用WebRTC API识别设备可用摄像头。WebRTC代表Web Real Time Communication，即网页即时通信的API，这个API能帮助Web应用开发人员通过简单的JavaScript编程就能实现功能丰富的实时多媒体应用。获取用户相机权限后，调用函数进行拍照，采集到拍照结果后，调用canvas标签对应Context的API drawImage将显示摄像头内容绘制到canvas标签页上，并且调用上传接口，将图片上传到后端。

3、接口设计

本系统的前端通过Web API与后端进行交互，从而实现对后端功能的调用。系统和后台交互的重要接口如下表：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 功能 | 路径 | 参数 | 返回值 | 方法 |
| 上传人体模特图 | /openpose-api | file | rendered\_img,keypoints | post |
| 上传衣服图片 | /cpvton | modelFile  modelParse  cloth  clothMask  pose | warpCloth  warpMask  tryon | post |

**2.2 后台模型处理**

2.2.1. Openpose提出关键姿态点

为了提取出能代表人体动作的关键骨骼点，本系统利用OpenPose对传输进来的人像进行处理。OpenPose 是由美国卡耐基梅隆大学( CMU) 基于卷积神经网络和监督学习算法开发的一个用于检测 2D 图片中人体关键点( 面部，四肢等) 的开源实时系统。OpenPose 区别于传统识别算法的是，研究时使用了自下而上的人体关键点信息识别，即寻找出人体的各个部位，再通过部分亲和域( Part Affinity Fields) 连接各部分。这使得 OpenPose 具有高鲁棒性、可实时输出 的优点。所以本系统选择使用 OpenPose 识别人体关键骨骼点信息。

如图2所示，2（a）为摄像头拍摄下的人物图像，2（b）为经过openpose处理后得到的骨骼节点可视化结果。

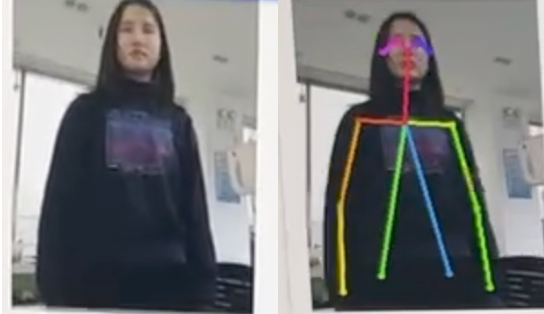


图 2 (a)人物图像 (b) 关键骨骼点

2．2.2 PGN获得人体语义分割

为了提取出人物图像中不同的区域，本系统采用PGN对传输进来的人像进行分割处理。PGN是中山大学和商汤基于实例级人体解析任务提出的部分分组网络，利用一个统一的网络对两个连续的分割部分进行分组，包括部分级像素分组（part-level pixel-grouping）和实例级部分分组（instance-level part-grouping）。首先，部分级像素分组可以借助以单个像素为一个部分标签的语义部分分割任务得到解决，这个方法学习了类别特征。之后，给定一组独立的语义部分，实例级部分分组可以根据预测的实例感知边缘（instance-aware edge）确定所有部分到底属于哪个实例，其中由实例边缘分隔的部分将被分给不同的人物实例。这种联合优化语义部分分割和实例感知边缘检测的无检测统一网络被称为部分分组网络。由于两个模块通过共享一致的分组目标而保持了相互之间的高度相关性，PGN得以进一步整合了一个细化分支，利用互补的语境信息使两个目标相互受益。所以本系统选择使用 PGN获得人体语义分割结果。如图3所示，3（a）为摄像头拍摄下的人物图像，3（b）为经过PGN处理后得到的人体语义分割可视化结果。



图3 (a) 人物图像 (b) 语义分割结果

2.2.3虚拟换衣算法

2.2.3.1对试穿衣服进行tps扭曲

在得到人物图像中的关键姿态点后，本系统通过将该姿态点输入到网络对姿态进行处理，从而将试穿衣服根据该姿态进行扭曲和变形。本系统采用TPS薄板样条插值算法对试穿衣服进行变换。薄板样条是一种常见的插值方法。因为它一般都是基于2D插值，所以经常用在在图像配准中。在两张图像中找出N个匹配点，应用TPS可以将这N个点形变到对应位置，同时给出整个空间的形变(插值)。薄板样条插值具有数值稳定性好和收敛性好的特性，且能近似曲线拟合复杂的形状，变形精确度和变形效果由特征点选择和特征点的数量决定。所以本系统选择使用 TPS来变换试穿衣服的形态，从而使得试穿衣服和人物图像中人的姿态保持一致，并尽可能的在变换过程中不丢失衣服的纹理色彩信息。如图4所示， 4（a）为用户所选择的试穿衣服，4（b）为经过TPS插值处理后得到的扭曲衣服的可视化结果。



4 （a） 试穿衣服 （b）扭曲的试穿衣服

2.2.3.2融合图像

通过前面获得的人体语义分割结果，可以去除掉人物图像中原来的衣服区域，保留剩下的外观区域，将扭曲的试穿衣服和人物图像中剩下的区域一起输入到一个U-net网络结构中，分别通过两个分支得到掩膜和换衣的生成图像。此时的生成图像保留了人物图像中不需要参与变换的大部分外观，但是在需要变换的衣服区域得到的纹理会较为粗糙，因此需要生成掩膜进行感兴趣区域提取，根据感兴趣区域替换，将TPS的扭曲试穿衣服与生成图像进行第二次融合，得到最终的结果。如图5所示，5(a)为人物图像， 5(b)为用户所选择的试穿衣服，5(c)是经过换衣算法生成的最终换衣结果。



图5 (a) 人物图像(b)试穿衣服 (c) 换衣结果

**结论**

本系统利用深度学习技术设计和实现了一套用户可以独自操作并获得较好的换衣结果的虚拟换衣系统，能够满足人们足不出户就可以选择自己想要试穿的衣服进行试穿，得到图像的结果。本系统主要工作与贡献如下：

1. 设计虚拟换衣系统各个功能模块，在此基础上设计系统的总体架构。
2. 设计并实现了web前端。使用WebRTC API识别设备可用摄像头，获取用户相机权限后，调用函数进行拍照，采集到拍照结果后，调用canvas标签对应Context的API drawImage将显示摄像头内容绘制到canvas标签页上。
3. 设计并实现了后台模型。利用Openpose提取出人物图像的关键骨骼点来表示姿态；利用PGN对人物图像进行语义分割来提取出要替换的衣服区域和非替换区域；利用tps 薄板样条插值来扭曲试穿衣服使得变形为和人物一样的姿态；通过图像融合和感兴趣区域提取将扭曲衣服无缝拟合到人物图像身上

本系统仍然存在可以改进和完善的地方：

（1）本系统目前只支持模型处理无背景的模版衣服，这需要用户自行搜索，而网络上更为可见的是穿着衣服的人物图像。

（2）本系统提供的换衣模型对人物图像的尺度和姿势有较为严格的约束，目前无法处理太大或太小的尺度以及十分复杂的人体姿势。