 DOI: https://doi.org/10.12349/iser.v6i4.5321

Design and research of virtual try-on system based on diffusion model

Quansheng Wang Meina Zhang*

Liaoning University of Science and Technology, Anshan, Liaoning, 114051, China

Abstract

This paper presents an intelligent virtual try-on system based on diffusion model, which aims to break through the limitations of traditional virtual try-on technology in the aspects of reality, diversity and user experience. In the process of research, we plan and design a new style of dual-path diffusion architecture, which focuses on half-body and full-body fitting scenes to make optimization processing respectively. This system integrates the OpenPose human pose estimation module with the fine human analysis module, and achieves the natural alignment state of clothing and human body by generating accurate clothing region masks. Experimental results show that, compared with some existing methods, our system has the ability to generate higher quality and more realistic fitting results while maintaining the identity characteristics of the model, especially in dealing with complex textures and wrinkles.

Keywords

Diffusion model, human pose, garment alignment, Generative AI

基于扩散模型的虚拟试衣系统的设计与研究

王泉盛 张美娜 *

辽宁科技大学，中国 · 辽宁 鞍山 114051

摘 要

本文拿出了一种依据扩散模型打造的智能虚拟试衣系统，其目的在于突破传统虚拟试衣技术在真实感、多样感以及用户体验等层面存在的局限之处。在研究进程当中，我们规划设计出了一种全新样式的双路径扩散架构，该架构分别着眼于半身和全身试衣场景来做出优化处理。此系统将OpenPose人体姿态估计模块与精细人体解析模块整合到了一起，借助生成精准的服装区域掩码这一方式，达成了服装和人体的自然对齐状态。实验结果显示，跟现有的一些方法相互对比来看，本系统在维持模特身份特征的同时，有能力生成质量更高且更加逼真的试衣成效，尤其在应对处理复杂纹理以及褶皱方面，其表现格外突出。

关键词

扩散模型；人体姿态；服装对齐；生成式AI

1 引言

随着电子商务不断地快速发展起来，虚拟试衣技术也渐渐变成了能够提升在线购物体验的一项颇为重要的工具[1]。不过，传统的虚拟试衣技术其实是面临着不少挑战的，这里面主要存在着像是服装和人体的匹配精度不高、没办法很准确地去模拟服装所具有的物理特性，还有就是个性化方面的需求也难以很好地满足等一系列的问题。传统的那些方法主要是依靠图像合成以及三维建模来开展相关工作的，可往往就是没办法真实地把试穿的效果给呈现出来，如此一来也就使得用户在体验方面表现得不是很好。

近些年来，扩散模型于图像生成这一领域当中获得了颇为重要的突破。与传统的生成对抗网络相比，扩散模型所生成出来的图像，其质量明显更高，而且图像的细节之处也处理得更为细致。该模型采取的方式是逐步往里面添加噪声，然后再逆向将清晰的图像给恢复出来，通过这样的操作过程，其展现出了极为强大的图像生成方面的能力。在虚拟试衣这个领域里面，扩散模型给生成那种精准的人体和服装相互匹配的效果带来了全新的可能性，它能够以更好的方式去模拟服装所具有的那种自然贴合的感觉以及动态呈现出来的效果，如此一来，便能够对传统技术所存在的不足之处起到弥补的作用。

【基金项目】2025 年辽宁科技大学大学生创新创业训练计划项目经费支持。

【作者简介】王泉盛（2005–），男，中国山东潍坊人，在读本科生，从事计算机视觉研究。

【通讯作者】张美娜（1981–），女，中国辽宁鞍山人，硕士，副教授，从事自然语言处理，软件工程研究。

2 深度学习在虚拟试衣中的应用

在深度学习不断发展的进程中，虚拟试衣技术慢慢开始引入那些较为先进的神经网络模型 [3]，如此一来，试穿效果在准确性以及真实感方面都有了明显的提升。

GAN 系列模型已经在虚拟试衣领域得到了颇为广泛的运用。这些模型能够生成出来的图像和真实试穿效果是比较相近的，如此一来便使得服装和用户体型之间的匹配程度得到了提升。GAN 可以依据用户图像以及服装相关信息去生成那种具有较高真实性的虚拟试穿图像，通过这样的方式也就提升了用户在体验方面的感受。

注意力机制于服装迁移所起的作用在于：注意力机制应用于深度学习之时，对提升模型处理复杂输入数据的能力是有帮助的。在虚拟试衣这一场景下，注意力机制能够助力模型更为妥善地聚焦于图像里服装以及人体的关键区域，从而达成更为精准的服装迁移与适配操作。

3 扩散模型研究进展

扩散模型的基本原理在于，它属于一种生成模型，其运作方式是逐步往图像里添加噪声，之后再凭借逆向的过程来把图像恢复到清晰的状态。相较于传统的生成对抗网络，也就是 GAN 而言，扩散模型在生成图像时，于图像的细节呈现以及质量把控这两方面，都展现出了更为突出的能力。经过长时间不断地迭代操作，扩散模型所生成出来的图像能够达到更加清晰的程度，看上去也更为自然，并且还具备着很高的细节度。

4 人体姿态估计与人体解析技术

人体姿态估计技术的运作方式是通过对图像里关节点位置加以分析，进而实现对人体具体姿态以及动作的识别。此技术能够助力虚拟试衣系统去捕捉用户处于不同姿势之时的体型方面的变化，以此来达成对服装穿着效果的精确模拟。

人体解析技术会进一步对人体的各个部位加以分析，像头部、躯干、四肢等部位，从而提取出更为精细的体型特征。在虚拟试衣中，上述这些信息对生成契合的服装效果来讲是极为重要的，其能够助力服装和用户的体型达成更为精准的匹配状态。

5 系统的设计与方法

5.1 系统的整体架构

该虚拟试衣系统在整体架构方面选用了模块化设计方式，其借助多个功能模块彼此协同运作的模式，来给予精准且高效的虚拟试衣感受。此系统的整体架构包含了如下几个较为关键的部分：

人体姿态估计模块（OpenPose）这一模型，可用来提取模特图像里面的人体关键点相关信息。而这些所提取出来的关键点，会充当后续开展服装试穿流程之时的一种参考依据，其目的在于保障服装和人体之间能够达到较好的贴合程度。

人体解析这一模块的主要职责在于对人体图像展开分析，从中提取出关于用户体型的具体详尽信息。在结合姿态估计所得到的数据之后，系统便能够较为精准地对用户的体型特征予以识别，进而实现对服装匹配效果的优化提升。

服装区域掩码生成这一模块，主要是运用深度学习方面的相关算法来生成服装区域的掩码。通过该操作，能够把服装图像里的一些相关区域和背景分离开来，如此一来，便为后续的合成操作提供了便利条件。

双路径扩散模型 [2] 构成了该系统极为关键的核心图像生成模块。在这一系统当中，借助半身试衣模型以及全身试衣模型，系统会依据用户上传的模特图像以及服装图像，进而生成虚拟试穿图像。

多样性采样及可控生成机制的作用在于它能够让用户依照自身的具体需求，对试穿效果所涉及的样式、颜色以及贴合度等各类参数加以调节，进而生成多种多样且各不相同的虚拟试穿效果。

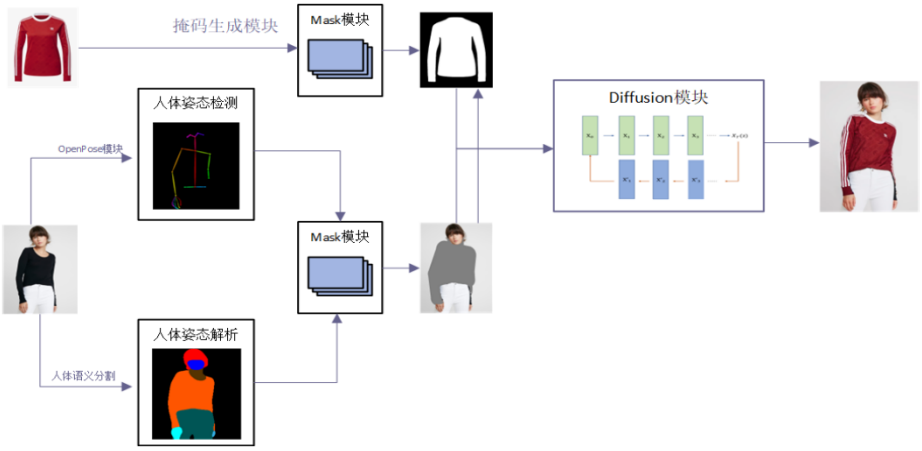

图 1 基于扩散模型的虚拟试衣方法架构

5.2 人体姿态估计模块

OpenPose 的工作流程大体上能够划分成若干主要步骤，这其中涵盖了输入预处理环节、人体姿态估计环节、热图处理环节、峰值检测环节以及姿态连接环节。而各个环节其具体的操作过程如下所示。

5.2.1 输入预处理

用户所上传的图像，第一步是要经过预处理操作 [7]，预处理涵盖了对图像做缩放处理以及对其颜色格式加以转换等内容。具体来讲，要对图像的分辨率做出调整，比如说调整成 384x512 这样的规格，与此同时，还要对图像的颜色进行标准化的处理。

5.2.2 人体姿态估计

在经过预处理的图像之上，OpenPose 借助卷积神经网络来生成一组热图，这里的每一个热图都是用来呈现图像里某个关节其概率的具体分布情况的。在这些热图当中，每一个像素所具有的值其实就意味着该位置出现特定关节的可能性大小。经过好几个阶段所开展的卷积以及池化相关操作之后，该模型最终成功生成能够涵盖所有关节的多个热图。

5.2.3 热图处理与峰值检测

在生成出来的热图中，运用高斯滤波器来对它加以平滑方面的处理，其目的在于能够减少其中存在的噪声，同时还可以提升关键点所具有的精度。高斯滤波器具体的公式如下所示：

$$G(x,y)=\frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \quad (1)$$

$G(x,y)$ 所表示的是高斯滤波器具体的值，σ 指的是标准差。

经过这样一番处理之后，系统便能够精准地提取出关键点所在的精确位置。在此之后，系统会运用峰值检测算法，在每一个关节所对应的热图当中去寻觅局部的最大值。

$$P(x,y)=\text{True} \quad \text{if} \quad \forall\left(x',y'\right)\in N(x,y), H(x,y)>H\left(x',y'\right) \quad (2)$$

在这之中，N(x,y) 所代表的是处于点 (x,y) 邻域范围之内的其他一些点，H(x,y) 是热图在这个相应点上所具有的值。

5.2.4 骨骼连接

当关键点被检测出来以后，系统就会依照骨骼结构所具有的逻辑关系，把这些检测到的点一一连接起来，进而形成一个完整的姿态。而这一连接的过程是能够凭借下面所给出的公式来加以描述的，假设存在两点，分别记为 A(x1,y1) 以及 B(x2,y2)，那么连接这两点的直线段便可以用如下的形式来表示：

$$\text{Line}=\{(x,y)|x=x_1+t(x_2-x_1), y=y_1+t(y_2-y_1), 0\le t\le 1\} \quad (3)$$

这里面，t 属于一个参数，其作用在于表明处在连线的各个点的相应位置。

5.3 人体解析模块

人体解析模块着重于对用户体型详细特征予以分析并提取出来。它运用了一个依据 ResNet 架构搭建而成的深度神经网络 [3]，借助将多个技术，像 PSP 模块、ASPP 模块以及边缘学习模块等相互结合的方式，来解析人体图像。

5.3.1 上下文信息融合

PSP 模块以及 ASPP 模块借助多尺度处理的方式，把不同尺寸的各类信息融合起来，以此强化模型针对复杂场景予以理解的能力，特别是在对人体细节部分展开处理的时候，其效果更为明显。

(1)PSP 模块中的金字塔池化

将输入特征图按照不同的尺寸来开展池化操作，这里分别运用尺寸为 1x1、2x2、3x3 以及 6x6 的池化窗口。倘若输入特征图其尺寸是 H × W 的话，那么在完成池化之后，所输出的特征图尺寸便会变为 H′ × W′，相应公式如下：

$$H'=\frac{H}{S},\ W'=\frac{W}{S} \quad (4)$$

(2)ASPP 模块中的膨胀卷积

膨胀卷积在运作过程中，会对卷积核的感受野予以扩展，如此一来，它便能够抓取到更多的上下文方面的信息。假定卷积核其大小设定为 k × k，而相应的膨胀率确定为 d 的话，那么经过扩展之后所形成的感受野则是：

$$\text{dilated size}=k+(k-1)\times(d-1) \quad (5)$$

借助这样的方式，ASPP 模块便能够对多尺度的信息予以捕捉，进而在复杂的图像当中提取出更多具备实用价值的特征。

5.3.2 边缘特征提取

边缘学习模块对图像里人体边缘的提取能力有了更进一步的强化，以此来保证服装和人体轮廓能够达成精准的匹配。

边缘卷积这一操作，其主要方式是针对图像里的每一个像素去计算它的梯度，以此来把图像当中的边缘信息提取出来。而关于梯度操作可以通过特定方式来进行表示的，具体可表述为：

$$G(x,y)=\sqrt{\left(\frac{\partial I}{\partial x}\right)^2+\left(\frac{\partial I}{\partial y}\right)^2} \quad (6)$$

在这里面，G(x,y) 所代表的是图像处于位置 (x,y) 之时的边缘强度情况，图像在 x 方向以及 y 方向上的梯度。

5.4 双路径扩散模块

本系统精心创设了一个双路径扩散模型 [2]，这个模型一方面可应用于半身试衣场景，另一方面能用于全身试衣情境。该模型会把生成的服装图像和用户的体型相互融合起来，以此达成高度贴近真实状况的虚拟试衣成效。

5.4.1 试衣模型的设计

半身试衣模型运用了条件扩散网络，把用户的上身图

像以及服装图像输入进去之后，便能生成那种虚拟试穿效果，且该效果呈现出自然贴合的状态。其具体操作步骤如下：

起初，用户所上传的上半身图像以及服装图像，会依照自定义的一套处理流程来展开预处理操作，这里面涉及到对图像尺寸的调整以及使其达到标准化的状态。图像在输入的时候，会借助 AutoProcessor 与 CLIPVisionModel 来完成编码工作，进而从中提取出图像所具有的嵌入特征。随后依靠 UNetGarm2DConditionModel 模型，服装图像的相关特征会被提取出来，并且会和人体图像的特征相互融合到一起，这样便能够生成出初步的虚拟试穿图像。关于扩散过程，其数学表达式如下：

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \Delta_t(\mathbf{y}) \quad (7)$$

其中，$\mathbf{x}_t$是扩散的当前图像所处的状态，$\mathbf{x}_{t-1}$则是前一个步骤呈现出来的状态，表示的是在每一个具体的时间步骤 t 上所发生的噪声添加或者移除这样的过程。

5.4.2 多样性采样与可控生成机制

为了让生成结果具备更多样性，同时提升用户与之交互的程度，本系统特别设计出了多样性采样机制以及可控生成机制。用户能够依照自身的实际需求，对生成图像的风格、色彩、款式等诸多参数加以调整，如此一来，便可以获取到多种多样的虚拟试穿效果。在多样性采样机制方面，系统会依据如下公式来对生成过程里的样本采样做出相应调整：

$$P(\mathbf{z}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\mathbf{z}-\mu)^2}{2\sigma^2}\right) \quad (8)$$

在这当中，μ 所代表的是生成分布的均值，而 σ 代表的则是其标准差。

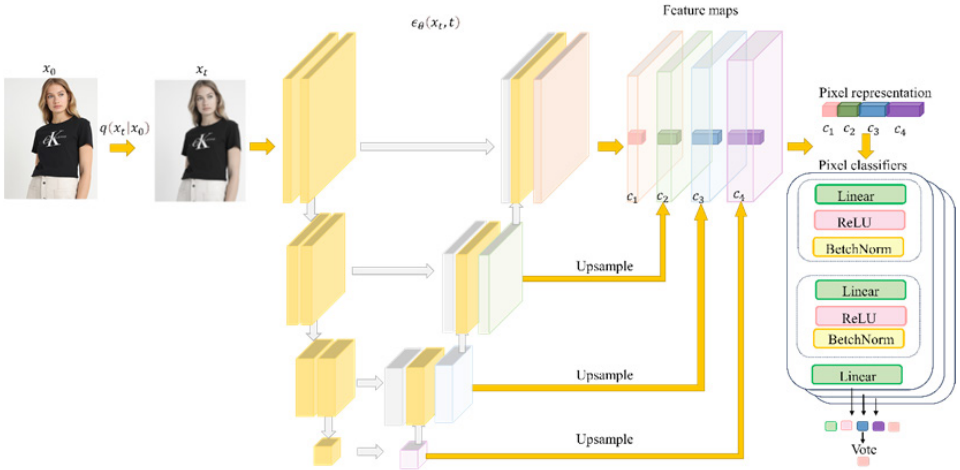

图 2 扩散模型结构示意图

6 实验与结果分析

就生成图像的质量来讲，我们运用 FID 以及 SSIM 对双路径扩散模型所生成的服装图像展开评估。实验得出的结果表明，在诸多评估指标的衡量之下，该模型展现出了出色的表现。从性能层面来看，双路径扩散模型有着比较高的生成速度，同时其计算成本也相对较低。

从表格呈现的评估指标能够观察到，本方法（Ours）在生成图像的质量，结构一致性方面以及感知自然度等方面，全都有着出色的表现。相较于其他的一些方法而言，本方法在生成图像时所呈现出的真实感以及细节的保真度上，是具备着明显优势的，图像的结构以及纹理都维持了更高程度的一致性。

表 1 试衣效果定量指标评价结果

评价指标	HR-VITON	GP-VTON	VITON-HD	Ours
KID ↓	16.77	15.52	18.63	10.36
FID ↓	12.58	10.86	13.90	9.52
SSIM ↑	0.855	0.887	0.864	0.885
LPIPS ↓	0.091	0.071	0.098	0.064

7 结语

此研究推出一种依托扩散模型构建的智能虚拟试衣系统，鉴于传统虚拟试衣技术在真实感层面、多样性表现以及用户体验环节存在诸多问题，特意设计了双路径扩散模型，且对其做了相应优化。实验得出的结果显示，跟现有的一些方法相互对比来看，该系统于生成质量方面、细节保真程度方面以及感知自然的程度方面，均能够获取颇为显著的提升，尤其在处置那些有着复杂纹理以及服装细节的情况时，展现出较为强劲的优势。该系统一方面给时尚零售领域引入了技术层面的创新之举，另一方面也为日后虚拟试衣系统的发展给予了全新的思路以及技术层面的路径选择。

参考文献

[1] 崔馨心,朱琳.元宇宙虚拟试衣的应用价值与前景分析[J].中国服饰,2024,(09):56-58.

[2] 陈相宜.基于外观流和扩散模型的虚拟试衣[J].现代信息科技,2025,9(05):17-24.

[3] 杨浩哲,郭楠.基于图像的虚拟试衣综述——从深度学习到扩散模型[J/OL].计算机工程与应用,1-21[2025-04-03].http://kns.cnki.net/kcms/detail/11.2127.TP.20241209.1400.008.html.