Modeling and Simulation 建模与仿真, 2023, 12(6), 5608-5620
Published Online November 2023 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2023.126509

Hans汉斯

基于生成对抗网络的服装属性控制方法

晏金鋆，向　忠，钱　淼

浙江理工大学机械工程学院，浙江　杭州

收稿日期：2023年7月11日；录用日期：2023年11月17日；发布日期：2023年11月24日

摘　要

在服装设计领域中，图像生成技术帮助服装设计师能够根据市场需求快速调整款式，针对目前基于生成对抗网络的图像属性控制模型中图像语义信息难以被充分利用的问题，本文提出了一种新的基于生成对抗网络的服装图像属性控制模型，实现对服装不同属性的控制。采用了基于Unet++结构实现图像的编码与解码，相比普通的编解码器，能有效减少图像语义信息在编解码过程中的丢失，提高生成图像的质量；同时，使用CSAM注意力模块加强判别网络对输入特征通道域与空间域上的关注度，提高判别器网络对服装图像大范围属性的鉴别能力。对比了本文与其他生成模型在服装属性控制上的效果，本文所提出模型与StarGAN、AttGAN等主流图像属性控制模型相比，能够对服装图像属性进行更好地控制，并生成高质量的图像。

关键词

生成对抗网络，图像处理，图像属性控制，注意力机制

A Generative Adversarial Network-Based Approach to Garment Attribute Control

Jinyun Yan, Zhong Xiang, Miao Qian

School of Mechanical Engineering, Zhejiang Sci-Tech University, Hangzhou Zhejiang

Received: Jul. 11th, 2023; accepted: Nov. 17th, 2023; published: Nov. 24th, 2023

Abstract

In the field of clothing design, the image generation technology to help clothing designers can quickly adjust style according to market demand, in view of the current based on image semantic information which is difficult to make full use of network control model, this paper puts forward a new clothing based on the generated against network image attribute control model to realize the

文章引用: 晏金鋆, 向忠, 钱淼. 基于生成对抗网络的服装属性控制方法[J]. 建模与仿真, 2023, 12(6): 5608-5620.
DOI: 10.12677/mos.2023.126509

control of different attributes of clothing. Unet++ structure is adopted to realize image coding and decoding, compared with the ordinary codec, it can effectively reduce the loss of image semantic information and improve the quality of the generated image. At the same time, the CSAM attention module is used to strengthen the discriminator network's attention to the input feature channel domain and spatial domain, and improve the discriminator network's ability to identify a wide range of attributes of garment images. Compared with the effect of this paper and other generative models in the clothing attribute control, compared with the mainstream image attribute control models such as StarGAN and AttGAN, the proposed model can better control the clothing image attributes and generate high-quality images.

Keywords

Generative Adversarial Network, Image Processing, Image Attribute Control, Attention Mechanism

1. 引言

随着快时尚行业的蓬勃发展，生成对抗网络等人工智能技术开始逐渐推向时装行业，如虚拟试衣、款式搭配、服装设计等[1]。服装属性控制的任务目的是缩短快时尚背景下服装的设计周期，减少服装设计师工作量，它旨在将输入服装图像进行局部特定属性改变的同时保留服装其他细节，缩短服装改款时间。

服装款式图是服装设计师表达自己设计理念和向用户展示成衣效果的媒介，是服装从想法走向产品过程中的基础，好的服装设计师需要实时捕捉潮流走向，适应市场需求，不断对已经设计好的服装款式图进行修改调整。因此，本文通过研究生成对抗网络技术来实现对服装进行自动改款。

传统的服装改款方式主要是采用工作室专家利用电子手绘画板在事先画好的服装款式图的基础上进行微调和重新上色，这类通过人工的方法成本较高且效率较低。针对快时尚行业对成本控制以及快速设计的要求，服装数字化设计开始进入快时尚设计领域，邓欣[2]和谢雪勇[3]等人提出建立服装款式部件的数据库，然后用数据库中款式部件对应替换所需的部件，从而完成对服装的改款操作。但此类方法需要预先绘制出服装款式部件，并进行人工筛选和分类，其过程繁杂且效率低下。

近年来，随着人工智能技术的快速发展，生成对抗网络在服装生成领域中开始崭露头角，其通过将一个目标属性向量和噪声向量拼接后输入进一个生成器中从而生成带有目标属性信息的图像，并通过一个判别器来指导生成器进行更新迭代[4]。但这种方法虽然改变了图像的某种属性，也改变了图像其他的属性和原有的风格。CHOI Y 等人[5]提出的用于多域图像到图像翻译的统一生成对抗网络(StarGAN)开始逐渐运用到图像属性控制的任务中，这种方法将含有某属性(如长袖、短袖、无袖等)的图像集合定义为域，该网络通过将原图像与目标域标签输入进一个编解码器结构的生成器中，通过学习图像域之间的映射关系，从而达到在不改变图像其他信息的同时对图像的属性进行改变。但由于图像在编解码的过程中丢失了关键语义信息，导致生成的图像出现不正确的属性和伪影。

本文提出了一种基于生成对抗网络的服装属性控制方法，该方法利用 Unet++结构[6]，采用密集连接增强了不同尺度的特征在生成器中的传递，减少了图像在编解码过程中语义信息的丢失；同时，考虑到服装图像中改变的属性区域广泛性，引入注意力机制，加强了判别器网络对服装图像中大范围属性区域

的感知能力，以提升判别器对生成器生成图像和真实图像的鉴别能力，使生成器得到判别器的有效指导，从而改善图像属性控制的精准度。

2. 方法

2.1. 模型总体结构

本文模型总体结构如图 1 所示，模型由生成器 G 和判别器 D 组成。首先将服装图像 a 与目标属性标签向量 l_{target} 进行融合输入到生成器 G 中，得到带有 l_{target} 属性的图像 b，然后通过将图像 b 与服装图像 a 的原始标签向量 $l_{original}$ 再次通过生成器得到重构图像 a'。

此外，生成器要尽可能生成足够真实的图像去欺骗判别器。而判别器 D 用于判别图像是真实图像还是生成图像，判别器的判别过程如图 1 虚线左侧所示。判别器接受真实图像与生成图像作为输入，并输出图像的属性标签向量与图像的真实性得分。

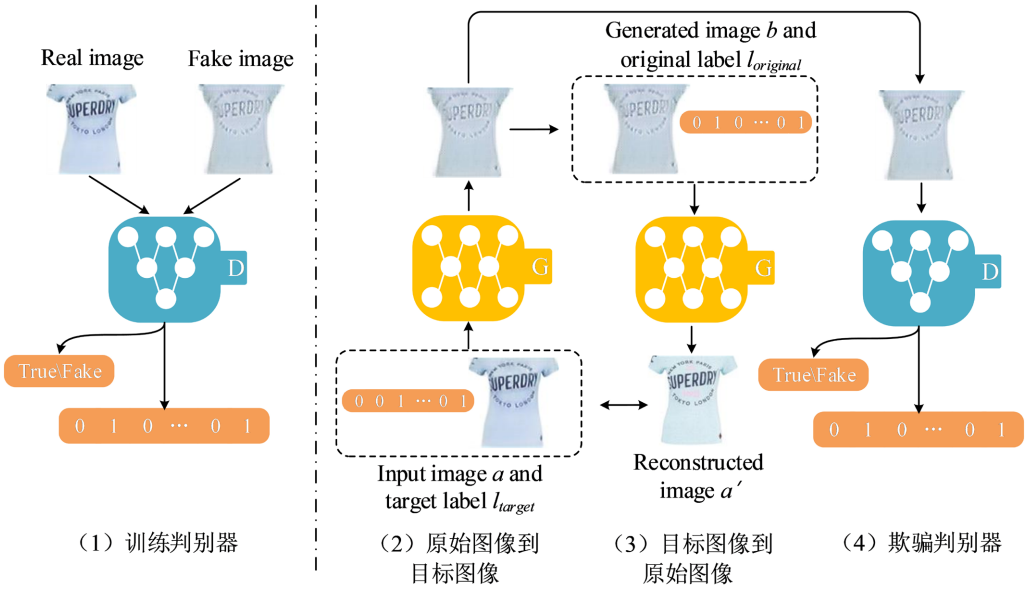


Figure 1. Overall structure diagram of the network
图 1. 网络总体结构图

2.2. 基于 Unet++结构的生成器搭建

不同尺度下的特征图包含着服装图像的颜色、纹理，形状等底层语义信息，因此在编解码器对图像进行上下采样过程中，存在大量的相似信息可以共享[7]。而由于编码器通过卷积层对图像下采样的过程中会导致图像特征信息的损失，令生成图像的质量与属性表达能力的下降，因此本文在 Unet++网络结构的基础上，对生成器进行改进，利用 Unet++中的密集跳跃连接结构，将编码器与解码器连接起来，增强不同尺度下特征图间信息传输能力。密集跳跃连接结构通过建立多个中转节点的方式，使编码器中的语义信息能够通过不同节点有选择性的融入到解码器特征图中，使得具有相同语义的特征能够融合起来，缓解生成器网络结构中的语义缺陷问题。

Unet++中对图像下采样过程中，由于运用大量池化操作对图像进行降维会导致图像特征信息丢失，本文采用了卷积层替代池化层，可以更好地保留原始服装图像的纹理和细节信息，使生成的图像更加自然和真实，生成器的具体结构如图 2 所示。

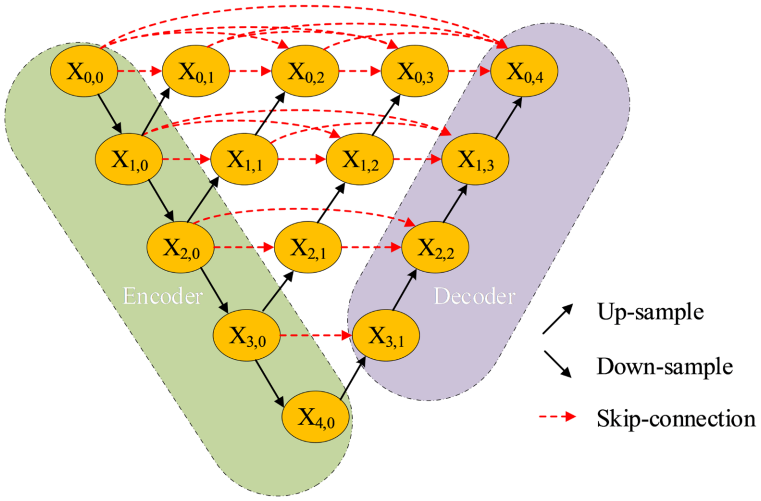


Figure 2. The generator structure
图 2. 生成器结构

令 $X_{i,j}\left(i \le m, j \le n\right)$ 表示各节点的输出，其中 i 表示编码器下采样层的索引，j 为跳跃连接节点处的索引，$X_{i,j}$ 由表示的特征输出计算为：

$$X_{i,j}=\begin{cases} D_{4\times4}\left(X_{i-1,j}\right) & i\neq 0\ ,\ j=0 \\ C_{1\times1}\left(\left[\left[X_{i,k}\right]_{k=0}^{j-1}, U_{4\times4}\left(X_{i+1,k-1}\right)\right]\right) & i\neq 0\ ,\ j>0 \end{cases} \tag{1}$$

其中 $C_{1\times1}\left(\cdot\right)$ 表示卷积核大小为 1×1 的卷积模块用以改变特征图的通道数，$\left[\cdot\right]$ 表示特征图拼接操作，$D_{4\times4}$ 卷积核大小为 4×4 的下采样操作，$U_{4\times4}$ 表示卷积核大小为 4×4 的上采样操作。网络第 $j=1$ 的节点只接受编码器前一层的输入；$j=1$ 的节点除了接受编码器前一层输入外，还接受编码器的下一层输入；$j>1$ 的节点接受编码器同一 i 层前 j 个节点的输出和 $i+1$ 层中第 j 个节点上采样后的输出，$X_{m,n}$ 即为编解码器最终的输出。

通过 Unet++中的密集跳跃连接思想将编码器多尺度的特征融合到解码器中，使得编码器下采样特征通过网络中的中转节点，有选择性地通过跳跃连接与解码器特征相连，解码器在密集跳跃连接的帮助下，能够获取到编码器特征中的有用的语义信息来弥补编码器和解码器特征映射之间的语义差距，基于 Unet++结构的生成器在保留原始的服装图像视觉信息的情况下能更好地对服装图像进行属性的修改转换。

2.3. 融合 CSAM 模块的判别器搭建

由于服装图像属性控制任务目标是在保留原始服装图像整体风格与细节对图像局部像素区域进行修改，通常情况下的 GAN 网络判别器输出一个概率值作为对图像真实性的得分，缺少对图像不同区域的重要性的考虑，导致模型缺乏对图像局部纹理与细节的感知，因此本文采用了 PatchGAN [8]中的方法，通过多个卷积层将输入进判别器中的图像映射为一个矩阵后输出，矩阵中的每个值对应着输入图像一小块区域的感受域，代表着原图像局部的真实性得分，通过这种方式，迫使模型能够更加关注图像的细节信息。此外，判别器通过一个额外的卷积操作输出图像的属性标签，PatchGAN 判别器的总体结构如图 3 所示，卷积核参数已在图中标注出，其中 K 代表卷积核大小，S 代表步长，P 代表特征图边缘填充数量。其中每一层网络由卷积与 LeakyReLU 激活共同构成，最终经过两个卷积操作后分别输出图像的每一类别的置信度与真实性得分。

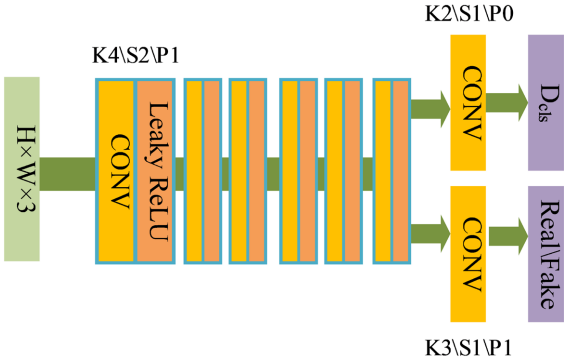


Figure 3. PatchGAN discriminator structure
图 3. PatchGAN 判别器结构

判别器的感受野受限于卷积核。由于服装图像属性区域的广泛性，在训练过程中难以从全局的角度捕捉服装图像的内容，导致属性区域无法准确地被判别器其感知，从而对生成器提供指导，生成出属性准确且清晰的服装图像。注意力机制可以扩大神经网络的感受野，在输入特征图中聚焦于当前任务更为关键的信息，降低对其他信息的关注度。因此为了扩大判别器的感受野，提升模型对服装图像属性区域的感知能力，本文设计了 CSAM (Channel and Spatial Attention Mechanism)模块，并将其融入进判别器中，使模型能够生成属性准确且清晰的服装图像，如图 4 所示。

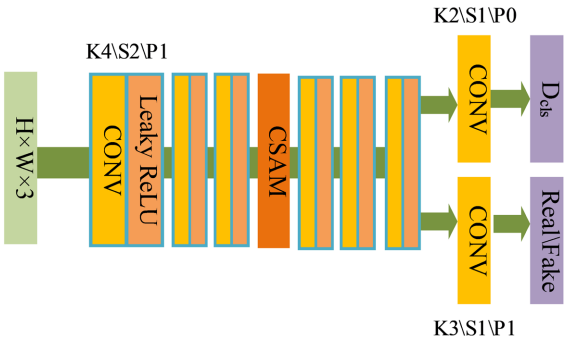


Figure 4. The improved discriminator structure
图 4. 改进后的判别器结构

CSAM (Channel and Spatial Attention Mechanism)模块的具体结构如图 5 所示。输入特征图 $F \in R^{C\times H\times W}$ 输入模块后分别通过通道注意力与空间注意力两步操作，得到通道注意力权重 $M_C(F) \in R^{C\times 1\times 1}$ 与空间注意力权重 $M_S(F_0) \in R^{1\times H\times W}$，从而学习到丰富的特征信息。

特征图中每一条通道包含了不同的特征信息，通道注意力集中关注于特征图中哪些通道是有意义的。通道注意力对特征图的每一通道施加不同的权重，让判别器能够侧重关注到特征图中有意义的服装图像特征，并将有意义的特征送入到空间注意力之中。通道注意力结构如图 5 所示，其中 $\otimes$ 表示矩阵的点乘操作，$\oplus$ 表示矩阵逐元素相加操作。首先由通道注意力将输入的特征图 $F \in R^{C\times H\times W}$ 经过两个并行的最大池化层 Maxpool(·)和平均池化层 Avgpool(·)，分别得到特征图 $F_M \in R^{C\times 1\times 1}$ 和 $F_A \in R^{C\times 1\times 1}$，然后经过同一个卷积核大小为 1 × 1 卷积模块 $\mathrm{Conv}_{1\times 1}(\cdot)$，在该模块中，它先将通道数压缩为原来的 $1/r$ (r：通道压缩倍率)倍，再扩张到原通道数后得到两个结果。将这两个输出结果进行逐元素相加，再通过一个 Sigmoid 激活函数 $\delta(\cdot)$ 得到通道注意力权重 $M_C(F) \in R^{C\times 1\times 1}$，再将这个输出结果乘原图得到最终输出 $F_0 \in R^{C\times H\times W}$，该过程可以用公式(2)~(5)进行表示。

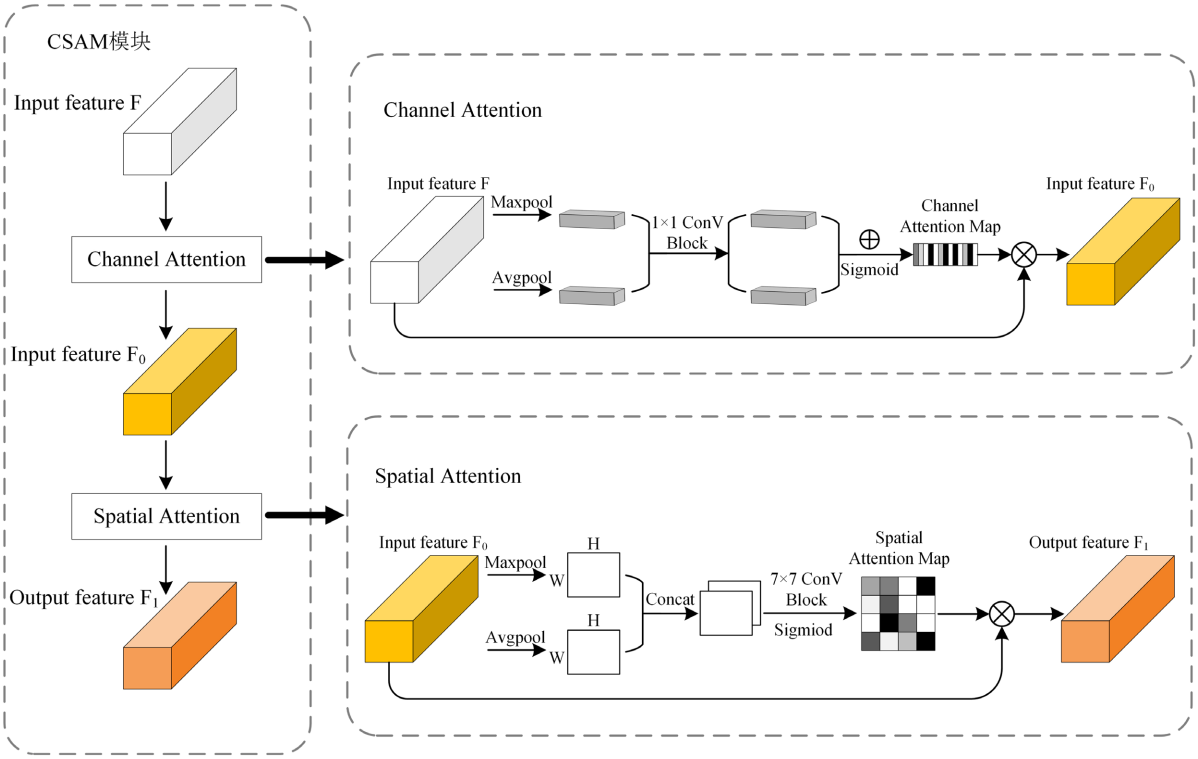


Figure 5. The CSAM module structure
图 5. CSAM 模块结构

$$F_M = \text{Maxpool}(F) \tag{2}$$

$$F_A = \text{Avgpool}(F) \tag{3}$$

$$M_C(F) = \delta\left(\text{Conv}_{1\times1}(F_M) + \text{Conv}_{1\times1}(F_A)\right) \tag{4}$$

$$F_0 = M_C(F) \otimes F \tag{5}$$

空间注意力模块的作用是增大网络对图像全局与局部的感知能力，并扩大网络的感受野，增强判别器对更大的服装属性区域的识别能力。而为避免过多池化层的加入导致空间特征信息的损失，本文采用大尺寸卷积核的卷积模块对输入特征进行压缩和解压，目的是为了进一步增大感受野。首先通过卷积核大小为 7 × 7 卷积模块 $\text{Conv}_{7\times7}(\cdot)$ 将输入特征图 $F_0 \in R^{C\times H\times W}$ 的通道进行压缩得到特征图 $F_0' \in R^{C\times H\times W}$，然后通过卷积核大小为 7 × 7 的卷积模块对 $F_0' \in R^{C\times H\times W}$ 通道数进行还原，并经过 Sigmoid 激活函数 $\delta(\cdot)$ 得到空间注意力权重，最后将权重 $M_S(F_0)$ 乘原图变回原始尺寸的特征图 $F_1 \in R^{C\times H\times W}$。该过程可以用公式(6)~(8)进行表示。

$$F_0' = \text{Conv}_{7\times7}(F_0) \tag{6}$$

$$M_S(F_0) = \delta\left(\text{Conv}_{7\times7}(F_0')\right) \tag{7}$$

$$F_1 = M_S(F_0) \otimes F_0 \tag{8}$$

本文使用了 PatchGAN 判别器结构，并在判别器中引入 CSAM 注意力模块，可以帮助判别器更好的感知服装图像属性控制的区域，加强判别器对服装图像鉴别能力，使得判别器能更好地指导生成器更新，在保证生成图像的质量同时提高了属性控制的准确度。

3. 结果与讨论

3.1. 数据集与实验设置

本文使用带有属性向量的 VITON [9]服装图像数据集进行实验，该数据集标记了 22 种不同的属性，共 14,221 张图像，每张图像的分辨率为 192 × 256，80%的图像被划分为训练集，20%划分为测试集。由于袖长、颜色这两种是服装图像中最显著和重要的视觉特征，因此本文选取数据集中的袖长与颜色这两种属性进行研究。其中服装属性通过独热编码进行表示，独热编码通常可用来表示没有大小关系的类别特征，在独热编码中，属性向量 C_k 表示为：$C_k=(c_1 \quad c_2 \quad \cdots \quad c_n)$。其中 k 代表每一种属性种类，n 代表属性的个数，c_n 的值为 0 或 1，1 代表该属性值存在，0 代表该属性值不存在。服装图像袖长标签 C_1 与颜色标签 C_2 连接后输入到网络中，图 6 展示了实验所用到数据集的分布情况。

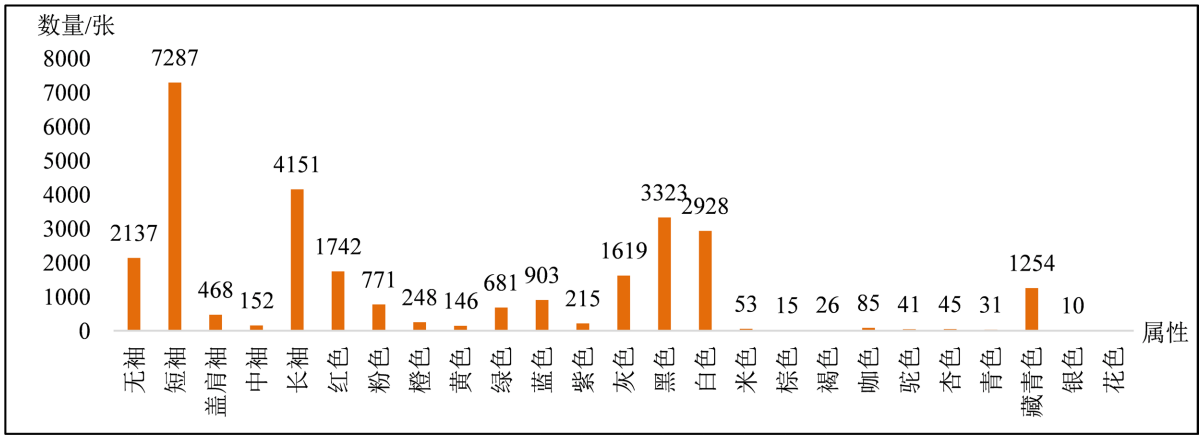


Figure 6. Datasets statistics
图 6. 数据集统计

实验环境设置如下：服务器操作系统为 Ubuntu20.04，Python 版本为 3.8.0，使用的深度学习框架为 PyTorch1.9.0，显卡为 NVIDIA RTX A5000，实验选用 Adam 算法作为模型参数优化器，BatchSize 的大小设置为 32，迭代次数为 20 万次。在训练之前对图像进行预处理，对服装图像进行中心裁剪与缩放，变成 128 × 128 × 3 的三通道 RGB 图像，再对三通道 RGB 图像进行归一化，将三通道 RGB 图像的像素从 0~255 之间的整数映射为 0~1 之间的浮点数，最后再输入模型进行训练和测试。

3.2. 评价指标

针对生成对抗网络性能的定量评估，本文采用分类准确率、Fréchet Inception 距离(FID)作为服装属性控制效果的评价指标。

1) 分类准确率

分类准确率作为在图像属性控制任务中用于衡量生成图像属性是否精准的指标得到了广泛应用[7]。其想法是，如果图像属性得到正确的转换，那么在真实图像上训练的分类器也能对属性转换后的图像进行正确的分类。因此本文采用在图像分类任务中广泛使用的 VGG16 网络[10]作为分类器，并在真实的服装图像数据集上进行训练，准确率达到了 94.3%。然后，本文将生成图像得到分类准确率对生成对抗网络模型进行评分，得分越高说明生成图像的属性转换越成功，该模型性能越好。

$$\text{Acc}=\frac{\text{分类正确样本}}{\text{总样本数}}\times 100\% \tag{9}$$

2) FID (Fréchet Inception Distance)

FID 是一个广泛使用的图像质量的评价指标[11]，其原理是使用 Inception 网络[12]对真实图像与生成图像进行特征提取，得到真实图像与生成图像分别服从的分布，计算分布之间的距离，距离越小说明真实图像与生成图像相似度越高。本文对通过计算真实图像分布与生成图像的 FID 分数，以衡量模型的图像生成能力，较低的 FID 意味着两个分布之间更接近，说明模型生成图像的效果越好。FID 的计算公式如公式(10)所示。

$$S_{\mathrm{FID}}(x,g)=\left\|\mu_x-\mu_g\right\|_2^2+Tr\left(\Sigma_x+\Sigma_g-2\left(\Sigma_x\Sigma_g\right)^{\frac{1}{2}}\right) \tag{10}$$

其中，x 代表真实图像，g 代表模型生成的图像。Tr 表示矩阵的迹，μ 表示为数据概率分布的均值，Σ 表示数据概率分布的协方差。

3.3. 基于 Unet++的生成网络效果评估

为验证 Unet++结构中密集跳跃连接的必要性，本文在 StarGAN 结构的基础上使用不同结构的生成器进行对比实验，分别为生成器采用不添加任何跳跃连接的编解码器结构的 StarGAN-w、生成器使用一个跳跃连接的 StarGAN-1s、生成器仅使用了两个跳跃连接 StarGAN-2s、和在生成器所有层间使用跳跃连接的 StarGAN-Unet 和本文提出的基于 Unet++进行改进的生成网络 StarGAN-Unet++。实验数据集采用 VITON 数据集，对于每种结构的生成器，在相同实验环境下进行测试，最终求出每种方法对应的 FID 与分类准确率，结果如表 1 所示。

Table 1. StarGAN variant model quantitative index assessment
表 1. StarGAN 变体模型量化指标评估

方法	StarGAN-w	StarGAN-1s	StarGAN-2s	StarGAN-Unet	StarGAN-Unet++
分类准确率	71.5%	70.6%	68.4%	68.1%	**73.3%**
FID	59.14	57.16	54.84	51.35	**45.22**

可以看出，在生成器中加入跳跃连接后，模型的各项指标都有较明显的改善，对比表 1 中 StarGAN-1s、StarGAN-2s、StarGAN-Unet 三种方法可以看到分类准确率随着跳跃连接个数的增加有所降低，而 FID 随着跳跃连接个数的增加而逐渐降低，这是因为编码器特征与解码器特征之间存在有可以共享信息，随着在编码器和解码器加入跳跃连接越来越多，编码器和解码器中越来越多的信息得到了共享，但编码器特征与解码器同尺度的特征之间存在语义上的不完全相同，编解码器中同尺度的特征图中含有不同的语义信息，通过直接与解码器特征拼接相连，造成了解码器中特征语义信息的混乱，从而导致属性控制能力有所降低。而采用 Unet++生成器的 StarGAN-Unet++，相比于不采用任何跳跃连接的生成模型 StarGAN-w 在 FID 上降低了 13.92，在分类准确率上提高了 5.2%。

本文方法得到了最优的分类准确率和最优的 FID 值，说明本文方法在属性控制的准确性上要优于其他四种方法并且通过本文方法生成的服装图像具有最接近真实图像的分布。这是由于通过 Unet++中的密集跳跃连接改善了生成网络的信息流动，让不同尺度的特征在编码器中的特征图通过中间节点有选择性地与解码器中的特征图进行融合，缓解了采用一般的跳跃连接结构所产生的语义缺陷问题。实验结果表明，采用 Unet++的生成模型降低了图像特征信息在跳跃连接中传递的损失，从而提高了模型对属性控制的准确度，增强了生成图像的质量。

3.4. CSAM 模块效果评估

为了验证本文提出的 CSAM 注意力模块的有效性，本文采用不使用任何注意力机制的基础模型、SE 注意力模块[13]、Self-Attention 注意力模块[14]、CBAM 注意力模块[15]作为对比模型进行实验。实验中保持损失函数、生成器结构不变且学习率不变，在判别器的相同位置插入不同注意力模块。实验结果如表 2 所示。

Table 2. Comparison of quantitative indicators for different attention modules
表 2. 不同注意力模块的量化指标对比

模型	W/O	SE	Self-Attention	CBAM	CSAM
分类准确率	73.3%	68.3%	69.2%	73.9%	**75.2%**
FID	45.22	64.14	52.13	42.96	**41.47**

从表 2 可以看出，相比不使用任何注意力机制的模型，SE 模块和 Self-Attention 模块在插入进判别网络中 FID 指标有所上升，分类准确率相比原始模型有所降低，这是因为 SE 模块仅在通道尺度上对特征施加注意力权重，使判别器忽略了空间尺度上的信息，而 Self-Attention 模块仅在空间尺度上对特征施加注意力权重，两种单一尺度的注意力机制都使网络丢失了重要的特征信息，从而导致判别器无法将正确的梯度信息传给生成器。CBAM 采用了混合注意力机制，同时在通道域与空间域上对特征施加注意力权重，使得分类准确率相比原始模型有所提高，FID 指标也有所下降，但 CBAM 模块采用多次的池化操作导致特征信息有所损失，使判别器难以得到有效训练。

本文所提出的 CSAM 注意力模块在分类准确率和 FID 指标上与其他四种模型对比均有着显著优势，CSAM 模块在通道注意力部分让判别器能够侧重关注到特征图中有意义的服装图像特征，并将有意义的特征送入到空间注意力进一步处理；同时在空间注意力部分增大了判别器的感受野，更好的感知服装图像属性控制的区域，加强判别器对服装图像鉴别能力，使得判别器能更好地指导生成器对服装属性的控制，从而提高了图像的属性控制的准确度。

3.5. 服装属性控制效果评估

1) 效果分析：为了评估本文所提出的方法，本文将 Unet++GAN 与目前主流方法 StarGAN、Fashion-Attgan、AttGAN 的生成的图像进行了比较，图 7 展示了四种方法生成结果的图像细节对比。

从图 7 中可以看到 AttGAN、Fashion-AttGAN 和 StarGAN 生成的图像丢失了图像的部分信息，尤其是带有服装上的文字图案无法得到很好的保留，如第一张与第三张测试图像中的文本信息都有着不同程度的缺失。而本文所提出的方法生成的图像更加清晰，较大程度保留了原服装的纹理和花纹图案细节，生成的服装图像更为真实。

不同方法的实验对比结果图如图 8 所示。AttGAN 模型的图像生成效果和其他三种模型相比效果较差，该模型几乎不能正确改变服装的任何属性，其原因在于 AttGAN 在训练过程中，分类学习的任务与重构学习任务同时进行，但 AttGAN 在损失函数设计上存在缺陷，分类损失和重构损失之间存在冲突，再加上服装属性较人脸属性需要改变的图像区域更广。Fashion-AttGAN 在服装生成的过程中存在着明显的图像缺失和伪影的问题，这是因为 Fashion-AttGAN 在损失函数上进行改进，赋予模型在图像重构中更多的自由度的同时，削弱了模型对图像属性的感知能力。StarGAN 模型生成的服装图像存在着颜色失真的情况，图像在改变袖长的同时也改变了其颜色，在图案细节与纹理还原上的表现有所欠缺，StarGAN 缺少了对服装图像整体与局部关系的理解。

注：GT 表示真实图像，AG 表示 AttGAN，FG 表示 Fashion-AttGAN，SG 表示 StarGAN，UG 表示 Unet++GAN。

Figure 7. Image reconstruction to generate detail comparison
图 7. 图像重构生成细节对比

注：AG 表示 AttGAN，FG 表示 Fashion-AttGAN，SG 表示 StarGAN，UG 表示 Unet++GAN。

Figure 8. Image rendering diagram generated by different models
图 8. 不同模型生成图像效果图

本文所提出的 Unet++GAN，通过嵌入 CSAM 注意力模块，增大了网络在空间上的感受野，在保证生成图像的质量同时提高了属性控制的准确度。此外，使用 Unet++的生成网络有效提升了网络对图像的细节和纹理的重建能力与图像相关区域的编辑能力，从而使模型能够生成更加自然、真实的服装图像。

2) 对比验证：为了进一步评估 Unet++GAN 模型的效果，本文通过分类准确率、FID 两个客观指标对不同模型进行了对比验证，实验结果如表 3 所示。

Table 3. Experimental results of different models
表 3. 不同模型的实验结果

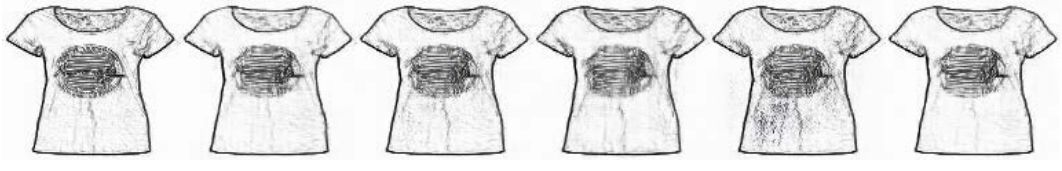
模型	AttGAN	Fashion-AttGAN	StarGAN	Unet++GAN (本文方法)
分类准确率	46.9%	68.5%	72.3%	75.2%
FID	88.18	65.24	46.73	41.47

从表 3 可以看出，本文方法与 AttGAN 方法相比，FID 下降了 49.71，分类准确率提升了 28.3%；本文方法与 Fashion-AttGAN 方法相比，FID 下降了 23.77，分类准确率提升了 6.7%；本文方法与 StarGAN 方法相比，FID 下降了 5.26，分类准确率提升了 2.9%。表明了本文方法在图像的属性控制上更加精确，生成的图像更加接近于真实图像分布。Unet++GAN 通过判别器中加入 CSAM 混合注意力机制，首先在通道上施加注意力权重，让判别器能够侧重关注到特征图中有意义的服装图像特征，并将有意义的特征送入到空间注意力之中，然后利用空间注意力增大判别器对图像的感受野，增强判别器对范围较大的服装属性区域的判别能力。

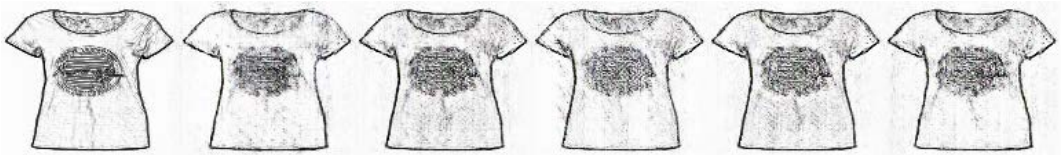
Unet++GAN 通过 Unet++结构的生成器加强编解码器不同尺度特征之间的融合与传递，克服了 Unet 结构中直接将编解码器不同语义特征相连导致模型对服装属性控制能力的减弱的问题，Unet++使得解码器能够自适应地获取到编码器特征中的有用的语义信息来弥补编码器和解码器特征映射之间的语义差距，在基于 Unet++生成器与 CASM 混合注意力机制判别器的共同作用下，本文方法所生成的图像得以在保留原始服装图像纹理与细节的情况下，改变由属性标签所控制的图像区域。

3.6. 服装线稿属性控制效果评估

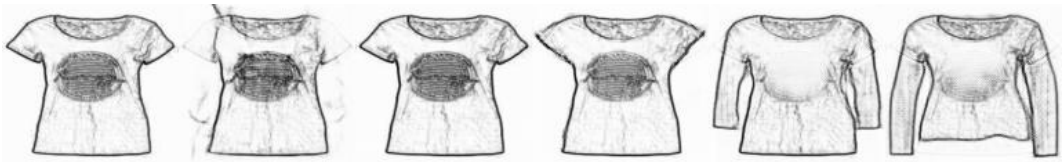
因为服装线稿图与服装款式图都属于服装设计师表达其设计理念的一种载体，并且对服装线稿图像进行属性控制的任务与对服装款式图进行属性控制的任务相类似，所以本文也在服装线稿数据集上进行了实验，以更好地评价 Unet++GAN 模型的性能，结果如图 9 所示。



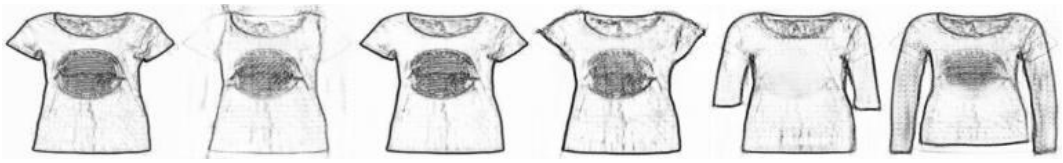
(a) AttGAN



(b) Fashion-AttGAN



(c) StarGAN



(d) Unet++GAN

注：图像从左至右第 1 列为原图像，第 2~6 列服装属性分别为无袖、短袖、盖肩袖、中袖、长袖。

Figure 9. Control effect of image attributes
图 9. 线稿图像属性控制效果

本文所使用的线稿数据集由真实的服装图像经过线稿提取算法逆向生成得到。由于线稿图像是单通道的灰度图，没有颜色属性，因此本文主要针对服装的袖长属性进行研究，并在线稿数据上进行了实验。本文训练了 AttGAN 模型、Fashion-AttGAN 模型和 Unet++GAN 模型，并对他们的定性和定量结果进行比较。从图 9 可以看出，AttGAN 与 Fashion-AttGAN 无法对线稿图像的袖长属性进行控制，而本文提出的 Unet++GAN 模型的属性转换较 StarGAN 更准确，线稿边缘更加清晰自然。定量结果如表 4 所示，Unet++GAN 模型与 StarGAN 模型相比分类准确率提高了 4.6%，FID 得分降低了 5.9%，说明了改进后的 Unet++GAN 生成图像的质量更高，属性控制更加准确。因此，本文所提出的方法在服装线稿的属性控制任务中较其他三种方法效果更佳。

Table 4. Comparison of the quantitative indicators of the sketch image generation model
表 4. 线稿图像生成模型量化指标对比

模型	AttGAN	Fashion-AttGAN	StarGAN	Unet++GAN (本文方法)
分类准确率	39.5%	37.2%	66.3%	**69.4%**
FID	218.22	207.61	146.87	**138.63**

4. 结论

本文提出一种基于生成对抗网络的服装图像属性控制模型，用于控制服装图像的属性。通过引入用 Unet++改进过后的编解码器改善了生成网络的梯度流，使编码器特征能够更好地与解码器特征进行融合，结合了 CSAM 注意力模块有效提升了网络对图像属性区域的感知能力，提升了判别器对服装属性相关区域的判别能力。实验表明，相对于其他属性控制模型，本文模型能够根据属性信息控制改变服装图像属性，生成高质量的服装图像。在后续研究中考虑扩充数据集，并改进注意力模块和网络模型的结合方式，以进一步改善模型的生成效果。

参考文献

[1] 施倩, 罗戎蕾. 基于生成对抗网络的服装图像生成研究进展[J]. 现代纺织技术, 2023, 31(2): 36-46.

[2] 邓欣. 服装款式部件的数据库构建及应用[J]. 天津纺织科技, 2016(4): 32-33.

[3] 谢雪勇, 张辉. XML 在服装部件信息数据存储上的优势[J]. 纺织科技进展, 2012(2): 84-86.

[4] Mirza, M. and Osindero, S. (2014) Conditional Generative Adversarial Nets.

[5] Choi, Y., Choi, M., Kim, M., *et al*. (2018) Stargan: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 8789-8797. https://doi.org/10.1109/CVPR.2018.00916

[6] Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., *et al*. (2019) Unet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Transactions on Medical Imaging*, **39**, 1856-1867. https://doi.org/10.1109/TMI.2019.2959609

[7] Isola, P., Zhu, J.-Y., Zhou, T., *et al*. (2017) Image-to-Image Translation with Conditional Adversarial Networks. *The Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 5967-5976. https://doi.org/10.1109/CVPR.2017.632

[8] Demir, U. and Unal, G. (2018) Patch-Based Image Inpainting with Generative Adversarial Networks.

[9] Han, X., Wu, Z., Wu, Z., *et al*. (2018) Viton: An Image-Based Virtual Try-On Network. *The Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7543-7552. https://doi.org/10.1109/CVPR.2018.00787

[10] Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition.

[11] Heusel, M., Ramsauer, H., Unterthiner, T., *et al*. (2017) GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. *Proceedings of the* 31*st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6629-6640.

[12] Szegedy, C., Vanhoucke, V., Ioffe, S., *et al*. (2016) Rethinking the Inception Architecture for Computer Vision. *The Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 2818-2826. https://doi.org/10.1109/CVPR.2016.308

[13] Hu, J., Shen, L. and Sun, G. (2018) Squeeze-and-Excitation Networks. *The Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7132-7141. https://doi.org/10.1109/CVPR.2018.00745

[14] Zhang, H., Goodfellow, I., Metaxas, D., *et al*. (2019) Self-Attention Generative Adversarial Networks. *Proceedings of the International Conference on Machine Learning*, 2019, 7354-7363.

[15] Woo, S., Park, J., Lee, J.-Y., *et al*. (2018) CBAM: Convolutional Block Attention Module. *The Proceedings of the European Conference on Computer Vision* (*ECCV*), Munich, 8-14 September 2018, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1