

基于StyleGAN的草图到服装图像的生成方法

胡新荣, 陈雷, 罗瑞奇*, 刘军平, 梁金星, 彭涛, 李丽, 罗航

武汉纺织大学计算机与人工智能学院, 湖北 武汉

收稿日期: 2022年9月26日; 录用日期: 2022年10月24日; 发布日期: 2022年10月31日

摘要

草图生成服装图像被广泛应用于服装设计过程中,可直接展示服装的设计效果,节约大量成本。但在当前,普遍利用条件变量实现服装草图的图形化渲染或通过学习图像与草图的公共属性向量去映射图像等方法,生成的服装图案受草图约束严重,生成的图像深度信息缺失。因此本文针对上述问题提出了一种基于StyleGAN的从草图到服装图像的图像生成方法。首先,利用VGG网络,对多种被变换的草图图像,进行内容特征提取。有针对性的提取目标特征,加深其内容特征,降低其他特征图信息对生成图像的约束;其次,利用特征金字塔的多尺度的特征融合方法,通过小型映射网络生成中间风格向量,利用StyleGAN生成器,丰富图像深度信息。实验结果表明我们的方法在草图生成服装图像方面具有一定的优越性,可以更完整地保留草图的细节,生成图像的质量更佳。

关键词

草图, 服装, 图像生成, 特征融合, 风格向量

A Sketch-to-Clothing Image Generation Method Based on StyleGAN

Xinrong Hu, Lei Chen, Ruiqi Luo*, Junping Luo, Jinxing Liang, Tao Peng, Li Li, Hang Luo

School of Computer and Artificial Intelligence, Wuhan Textile University, Wuhan Hubei

Received: Sep. 26th, 2022; accepted: Oct. 24th, 2022; published: Oct. 31st, 2022

Abstract

Sketch-generated garment images are widely used in the garment design process, which can directly show the design effect of garments and save a lot of costs. At present, it is common to use conditional variables to achieve graphical rendering of garment sketches or to map images by

*通讯作者。

learning the common attribute vectors of images and sketches, etc. The generated garment patterns are severely constrained by sketches and the depth information of the generated images is missing. Therefore, this paper proposes a StyleGAN-based image generation method from sketches to garment images to address the above problems. First, using the VGG network, a variety of sketch images has been transformed, and the content features are extracted. Targeted extraction of target features deepens their content features and reduces the constraint of other feature map information on the generated images. Second, a multi-scale feature fusion method using a feature pyramid generates intermediate style vectors through a small mapping network and enriches the image depth information using a StyleGAN generator. The experimental results show the superiority of our method in the sketch generation of garment images, which can retain the details of sketches more completely and generate images with better quality.

Keywords

Sketching, Clothing, Image Generation, Feature Fusion, Style Vectors

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着服装行业的进步发展，人们对于服装的需求越来越多，于是服装企业对于简化服装生产流程的需求也因此增多。而在服装的生产过程中，需要先设计出服装设计草图，然后制作出服装样本，从而导致成本的提升以及资源的浪费，因此在设计阶段可以清楚的知道设计的服装效果，就可以为服装企业节约大量时间成本和金钱成本。于是如何精细化服装设计流程，便成了一个重要难题。为了解决这个问题，本文从可视化设计草图的角度，将设计师设计的草图图纸直接生成可视化的服装图像，简化生产流程，不需要制作服装样品，降低服装制作成本。

近年来，服装设计行业提出了很多在服装设计阶段的服装展示方法。大部分方法通过生成虚拟的服装，向用户展示服装的设计效果。这类方法在没有制作出样品的情况下，利用草图图片生成类似漫画效果的服装图片。但是这种方法存在一定的局限性，使得生成抽象的图片，无法适用于服装制造行业。与之不同的是，本文通过充分的提取原有草图图像的内容信息，采用一种风格映射的方式生成图片，解决了原有的方法过于参照草图边缘信息的问题，同时提高了图片色彩阴影对比，使生成的服装图片具有更高的真实度。

2018年，James Hays 等人提出了 SketchyGAN [1]，在生成器和判别器使用屏蔽剩余单元(MRU)块来代替卷积层，是一种基于 GAN 的端到端的多模态合成方法。同年，ContextualGAN [2]的作者将草图作为上下文弱约束对图像补全，使用联合图像来学习草图与其对应图像的联合分布。而 MUNIT [3]采用无监督的方法，将图像数据分为内容与风格，从不同的数据空间采样进行重构。到 2020 年，为从草图生成场景级图像，EdgeGAN [4]被用来生成前景图像，通过学习图像和边缘图的公共属性向量表示，最后映射到对应的图像。2021 年，pSp [5]为利用 StyleGAN [6]的强大图像生成能力，通过特征金字塔结构，将草图作为输入，提取 18 个中间风格向量作为 StyleGAN 生成器的输入，但该方法生成的图像受草图风格信息约束，中间风格向量存在大量草图风格因素，导致生成图像色彩单一。

针对 pSp 的不足，本文引入 VGG [7]网络模型，减少风格特征的提取，降低输入草图的风格对生成

的服装图像的影响。由于草图表现出更多的形状不确定性，因此需要从错乱的草图线条中模拟准确的内容。首先，使用第三方图像处理库处理本文提供的服装图像，生成对应的上衣服装草图。其次，随机遮蔽草图的小块区域，进一步变换每个草图。然后，利用 VGG [7]网络模型提取草图内容特征，降低对生成图像的约束。最后，利用 StyleGAN 生成器，得到更加丰富的图像深度信息。

2. 相关工作

在图像的编译领域中，生成对抗网络(GAN)是一种十分重要的网络模型。被提出的草图生成图像方法大多基于生成对抗网络(GAN)模型，作为一个生成模型，GAN 模型通过对抗学习，来避免了传统的生成模型在实际应用中的一些困难。CGAN [8]、CycleGAN [9]、FashionGAN [10]都能够学习制作服装的一些基本特征，比如配色方案、褶皱、阴影和折痕。CycleGAN 生成的照片更加逼真，但其无法继承织物图案的颜色。CGAN 在颜色分布更加多样化方面都优于 CycleGAN，但 CGAN 无法正确映射形状和颜色。FashionGAN 在 cGAN 的基础上，在织物和潜在向量之间建立了双射关系，从而可以用织物信息解释潜在向量。但 FashionGAN 所生成的服装图案受草图约束严重，生成的服装图像缺乏真实感，边界被固定，因此当草图提供的信息较低时，生成图像就会存在失真等现象。

最近 EdgeGAN [4]、SSS2I [11]、pSp [5]等方法在草图生成图像方面有着进一步的工作。Chengying Gao 等人提出了 SketchyCOCO [4]，利用手绘草图来生成整个场景的图像。它将草图分为前景和背景两部分，前景图像是其数据集中的鹿、斑马等各类动物。针对前景草图的抽象性和差异性，映射空间巨大，作者提出了新的神经网络算法 EdgeGAN，通过跨域数据来对一个表示对象进行通用的表示，来训练前景图像。但此方法生成的图像缺乏真实度，生成的前景图像分辨率较低。而 SSS2I [11]是一种以风格范例为导向的带有手绘草图到图像生成的方法，它提出的自监督的自编码器(AutoEncoder，简称 AE) [12]是通过将草图和 RGB 图像的内容和风格特征分离，然后分别提取特征，送入生成器，从而将草图转换为图片。而 pSp [5]是基于 StyleGAN [6]生成器的编码网络，利用 ω +潜空间的映射，用特征金字塔扩展了编码器主干，生成了三个级别的特征映射，使用简单中间网络 map2style [5]从中提取样式。用层次表示对齐的样式，随后根据其比例输入生成器生成输出图像，从而通过中间样式表示完成从草图到图像的转换。

为保存更具有输入图片的信息，本文使用基于 StyleGAN [6]网络结构的模型，利用 pSp [5]编码器编码草图到隐空间，不用训练判别器，无需额外的优化。基于此设计了草图生成服装图像框架，将草图的多姿态进行内容编码提取的特征图融合，丰富输入图像的内容信息。

3. 基于草图的服装图像生成方法

3.1. 模型的整体结构

本文在 pSp [6]编码器的基础上进行了进一步的研究工作。使用草图来约束生成图像的形状，并采用风格映射的方式来约束生成服装的颜色与纹理。

草图进行多次简单的变换后，将变换后的图像作为输入，经过内容编码器编码得到多重内容特征向量，并多层加权融合，在编码过程中保留更多的草图信息，减少其他信息，使生成的图像具有更高的草图含义。使用服装草图与之对应的服装图像作为成对图像来训练网络，通过混合正则化的方法，在训练过程中使用两个随机潜码，在生成图像时，随机选取一个交叉点，把一个潜码切换到另一个潜码，从而自然的生成具有更清晰纹理的服装图像。

模型整体结构如图 1 所示，该模型主要由一个生成器与一个编码器组成，编码器的目的是学习相对于平均样式向量的潜在代码。首先，对输入草图 I 遮蔽随机的小区域做为随机的变换，得到变换后的图像记为 I' 。 I' 经过内容编码器 $E_{content}$ 得到多层内容特征层 $f_{content}$ ， $f_{content}$ 具有草图的内容信息，而不具有其

风格信息。进一步融合后，得到特征 f_{map} 。为对应 StyleGAN 生成器对于不同层级的输入需求，将 f_{map} 作为特征金字塔底层结构，生成三种不同等级的特征。map2style [5]结构是将卷积层变成 512 维的向量，利用 map2style 全连接层来生成 18 个编码，最后得到的隐代码的大小是 18×512 。再通过一个全连接层 A 输入到 StyleGAN 生成器的各个 style 中。

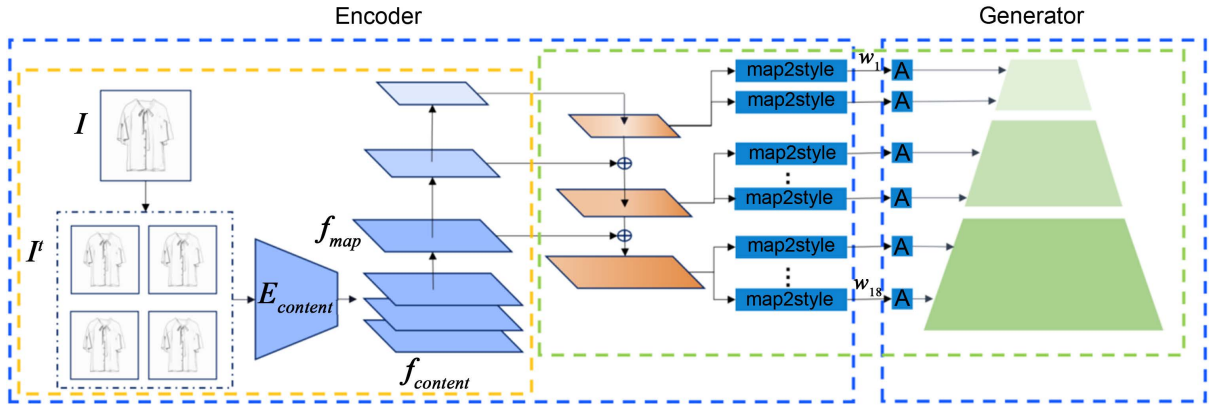


Figure 1. Model overall structure diagram
图 1. 模型整体结构图

同时，对于模型的输出，给定一个输入图像 b ，定义 a 为预先训练的生成器的平均样式向量。模型的输出被定义为：

$$p(b) := G(E(b) + a), \tag{1}$$

其中 $E(\cdot)$ 和 $G(\cdot)$ 分别表示编码器和 StyleGAN [6] 生成器。这个公式目的在于学习关于平均样式向量的潜在代码，这可以导致更好的初始化。

3.2. 基于内容编码器的多重内容特征提取

草图生成图像的过程中，输入的草图和预期生成的服装图像的边缘之间具有实质性的像素对齐属性。相反，真实的草图表现出更多的不确定性和变形，因此需要模型从错位的草图线幻觉适当的内容。以数据增强的自我监督的方式增强了内容的内容特征提取能力。我们通过掩蔽随机的小区域来一步变换每个草图，使不连贯。在图 1 中可以找到 I' 的一个示例集。

如图 2 所示，为提取更多的输入图像的内容信息，减少风格等其他信息的影响，本文在 pSp [5] 编码器的基础上，增加使用了预训练的 VGG [7] 网络 $E_{content}$ 。在输入时，对输入草图先进行简单的随机的四种变换，在训练时，将四种变换随机的应用到原输入草图 I 上，得到变换后的图像 I' 。将变换后的 I' 作为 $E_{content}$ 的输入得到多重内容特征 $f_{content}$ ，进一步特征融合后，得到金字塔结构底层 f_{map} 。

编码器主要是将输入的原图转换成隐代码，一共有 18 层。StyleGAN [6] 的作者提出了不同层级的编码实际上代表着不同层次的特征，可以将其分为三层次，分别为 coarse、medium 和 fine。为对应 StyleGAN 的 18 层网络结构，首先将 f_{map} 作为输入，使用 ResNet 主干上的标准特征金字塔 [13] 来提取特征图，输出三个不同尺寸的特征图，然后每个特征图都生成多个隐代码。其中因为不同尺寸和不同深度的特征图包含的语义信息不一样，小特征图对应 styles 的第 0~2 层，中特征图对应 styles 第 3~6 层，大特征图对应 styles 第 7~18 层。每一层隐代码再进行仿射变换，然后输送给预训练好的 StyleGAN [6]，最后输出图像。

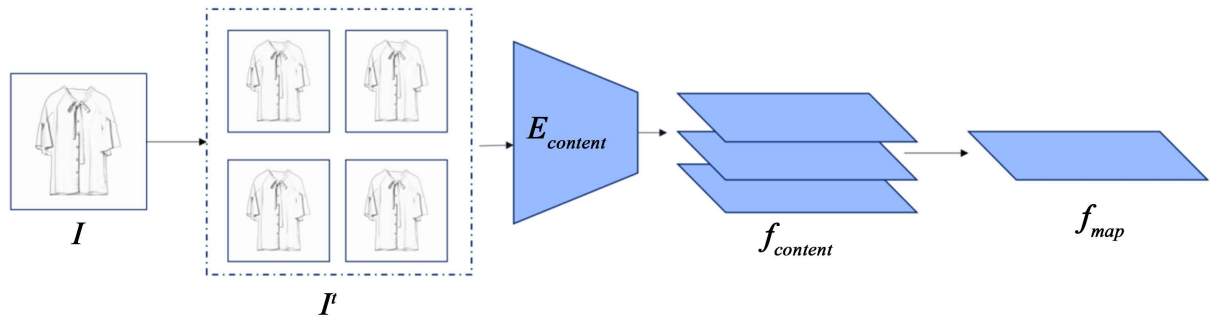


Figure 2. Content-based feature extraction module

图 2. 内容特征提取模块基线实验对比

3.3. 编码器的损失函数设计

编码器使用几个目标的加权组合进行训练。

在编码器的工作过程中，首先，为了进一步加强内容信息的提取，确保 $E_{content}$ 只能提取内容而不能提取风格，因此采用 triplet loss 来确保效果更好，损失公式为

$$\mathcal{L}_t = \max(d(f_t, f_{pos}) - d(f_t, f_{neg}) + \alpha, 0) \quad (2)$$

其中， $d(\cdot)$ 为均方距离， α 为边距， f_t 和 f_{pos} 是来自对应于同一 I 的草图的特征，而 f_{neg} 是来自一个随机不匹配的草图。这种自我监督过程使 $E_{content}$ 对草图上的变化更加稳健，并使其能够从扭曲和不完整线条的草图中推断出更准确和完整的内容。

其次，利用像素级的 \mathcal{L}_2 损失：

$$\mathcal{L}_2(b) = \|b - (b)\|_2 \quad (3)$$

此外，为了学习知觉相似性，利用了 LPIPS [14]损失，与更标准的知觉损失[15]相比，它能更好地保持图像质量[16]：

$$\mathcal{L}_{LPIPS}(b) = \|F(b) - F(p(b))\|_2 \quad (4)$$

其中， $F(\cdot)$ 表示感知特征提取器。

总之，总损失函数被定义为：

$$\mathcal{L}(b) = \alpha_1 \mathcal{L}_2(b) + \alpha_2 (\lambda_1 \mathcal{L}_{LPIPS}(b) + \lambda_2 \mathcal{L}_t) \quad (5)$$

其中 $\alpha_1, \alpha_2, \lambda_1, \lambda_2$ 是定义损失权重的常数。这组经过策划的损失函数允许更准确地编码到 StyleGAN [6]中。

4. 实验与分析

在这节中，我们将通过实验数据验证我们的方法的有效性和可行性。

4.1. 数据集

本文在选择和收集数据集的过程中，考虑服装具有不同类别标签的属性，且收集具有成对的草图与其对应的服装图像具有一定的挑战性。综合考虑后，我们建立了一个草图-服装的成对数据集。数据集共包括 31,735 张分辨率为 512×512 的服装图片。为了避免产生不符合要求的结果，本文采用其中 9687 张上衣服装，以白底图片为基准，用相关算法对服装预处理，使数据集没有其他的背景干扰信息的提取。使用 python 开源的第三方图像处理库 PIL (Python Image Library)生成对应的上衣服装草图，以便进行训练。部分数据集如图 3 所示。



Figure 3. Partial sketch-clothing dataset

图 3. 部分草图 - 服装数据集

4.2. 实验环境和超参数

本文实验环境配置为：Ubuntu16.04 64 位操作系统，处理器为 Intel(R) Xeon(R) Gold 6226R CPU @29.GHz，显卡为 A100-PCIE-40GB，采用 Python3.8 和 PyTorch 1.9.1 训练框架。将批处理大小设置为 16，共 500,000 个 epoch。使用 Ranger 优化器，一个校正的 Adam 和前瞻技术的组合，恒定的学习率为 0.001 训练。

4.3. 评价指标

为了定量地评估模型效果，使用了常见的度量生成模型质量和多样性的方法 Fréchet Inception Distance [17] (FID) 计算真实图片和假图片在特征层面的距离，测算两组图片之间的分布相似性，并作为生成图像的质量和多样性以及草图与图像的匹配程度的度量。较低的 FID 表示生成数据的分布更接近真实样本的分布。用学习感知图像块相似度 [14] (Learned Perceptual Image Patch Similarity, LPIPS) 也称为“感知损失” (Perceptual Loss) 度量两张图像之间的差别。学习生成图像到真实图像的反向映射强制生成器学习从假图像中重构真实图像的反向映射，并优先处理它们之间的感知相似度。较低的 LPIPS 的值表示两张图像之间具有更高的相似度。

4.4. 实验结果与分析

为了方便与其他草图生成图像的方法进行比较，本文将制作的草图 - 服装的成对数据集投入其他方法中，并以此与本文所提出的方法进行对比。训练损失变化如图 4 所示。

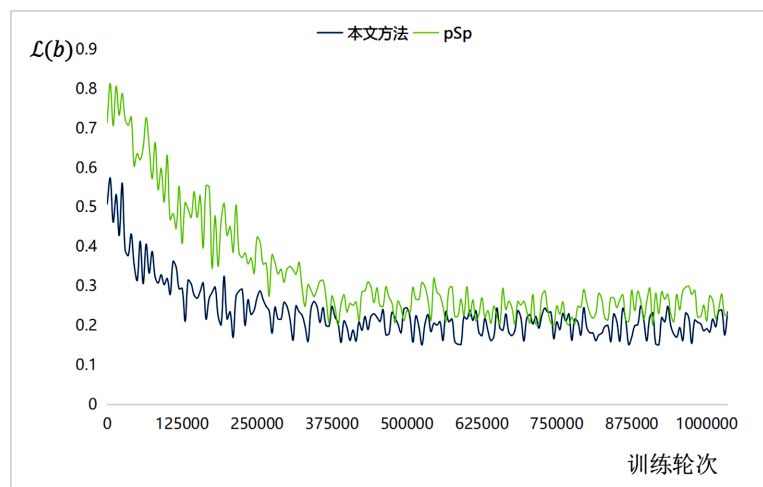


Figure 4. Contrast curve of training loss of improved method
图 4. 改进方法训练损失对比曲线基线实验对比

本文可视化了训练过程中如式(5)所示的总损失函数。如图所示,本文第一阶段的训练在 300000 次迭代便达到可用程度,而原始的 pSp 方法需要进行 375000 次迭代。后续实验则表明,本文提出的内容特征提取方法有效提高了最终的图像质量。

4.4.1. 定量分析

本文选用三个基线模型 EdgeGAN [4]、AutoEncoder [11]与 pSp [5]。EdgeGAN 用一个属性向量捕获用户的表达式意图,然后从中生成所需的图像。而 AutoEncoder 是将内容编码向量与风格编码向量作为解码器的输入,从而生成图像。pSp 建立在一个预先训练过的 StyleGAN [6]生成器和 W+潜在空间之上。通过扩展了一个具有特征金字塔的编码器主干,使用简单的中间网络提取样式,然后根据比例输入生成器,生成输出图像。

由于基线模型并没有服装相关的实验结果,因此本文选择从头开始训练这三个模型。本文只保留了 256×256 的输出图像,因此将本文所设计的草图-服装数据集作为三个基线模型的输入,分别生成 256×256 的输出图像,并进行 FID [17]和 LPIPS [14]评估。FID 和 LPIPS 评估结果如表 1。

Table 1. Baseline experimental results on the dataset

表 1. 数据集上的基线实验结果

	FID↓	LPIPS↓
pSp [5]	34.372	0.289
AutoEncoder [11]	30.649	0.276
EdgeGAN [4]	164.765	0.534
Ours	27.716	0.213

本文方法与其他方法的 FID 得分和 LPIPS 得分的对比结果如表 1 所列。从表 1 中可以看出相比于原始的 pSp [5]框架,FID [17]指标降低了 6.656,LPIPS [14]降低了 0.076。相比于 AutoEncoder [11]与 EdgeGAN [4]而言,FID 分别降低了 2.933, 137.065,LPIPS 分别降低了 0.063, 0.321。可以看到,本文方法早草图生成服装图像的实验中 FID 得分在所有方法中最低,说明生成的图像质量更高。另外,采用本文方法生成的服装图像的 LPIPS 得分在所有方法中最低,说明采用本文方法可以生成更加真实的服装图像。

4.4.2. 定性分析

为了从人类视觉角度直观的比较生成图像的质量,采用感知研究[18][19]的方式进一步分析。本文将实验结果与其他基线结果分别随机抽取 10 张生成的服装图像,并随机选取 30 名与实验无关人员对每张图像打分,分别从整体、形状、色彩、阴影、细节等 5 个部分评分,评分表如下表 2。

Table 2. Scoring indicators
表 2. 评分指标

评分指标	分数(总分 100 分)
整体观感	30
形状观感	15
色彩观感	20
阴影观感	20
细节纹理	15

整体观感的评分需评分者从生成图片和原始图片之间的相似度与生成图片的图片质量等两方面综合考虑;形状观感的评分需评分者考虑生成图像的形状大小与原始图片的形状大小的相似度;色彩观感部分需考虑生成服装的色彩分布是否均匀且自然;阴影观感部分需对关注生成服装的真实度综合评分;细节纹理的评分在于生成的服装是否能充分映射出服装的花纹等细节信息。

图像每张图像 100 分,每人对该 10 张图像打分后,汇总取均分后的各项平均分数代表其对样本数据的该项感知平均分。对全部评分者的每项评分指标汇总求平均分,汇总平均分情况如下表 3。

Table 3. The comprehensive mean score results
表 3. 综合平均得分结果

	Ours	pSp	AutoEncoder	EdgeGAN
整体观感平均分(30)	27.1	25.9	26.0	8.9
形状观感平均分(15)	14.4	12.8	10.9	3.3
色彩观感平均分(20)	13.6	12.1	14.2	7.5
阴影观感平均分(20)	17.6	16.3	16.9	4.7
细节纹理平均分(15)	13.8	12.5	12.7	2.7
平均总分	86.5	79.6	80.7	27.1

可以看到,除色彩观感上较 AE 方法不足之外,本文的方法在整体在其他人员的感官看法上,都存在着一定的优势,较为优于其他方法。

为进一步对本文的实验进行定性分析,本文对原 pSp [5]框架进行消融实验。

在原 pSp 编码器的实验中,原作者提出的是一个通用的图像到图像的转换框架,其保留更多输入图像的除内容信息之外的其他信息。因此在草图到图像的训练效果较差。考虑到输入的草图不仅具有其内容特征,同时也含有风格特征,因此加入了内容特征提取模块。

为了更清晰地表明本文在原模型上提出的改进,且对生成的图像质量有着一定的提升,进行内容特

征提取模块的消融实验。本文分析了内容特征提取模块的功效，并将部分实验结果放在了消融实验结果中。消融实验结果如图 5 所示。

其中，Sketch 项为作为输入的草图原始图像，Image 项为作为生成目标的原始服装图像，Ours 项为本文方法所生成图像，pSp 项是其方法所生成图像。

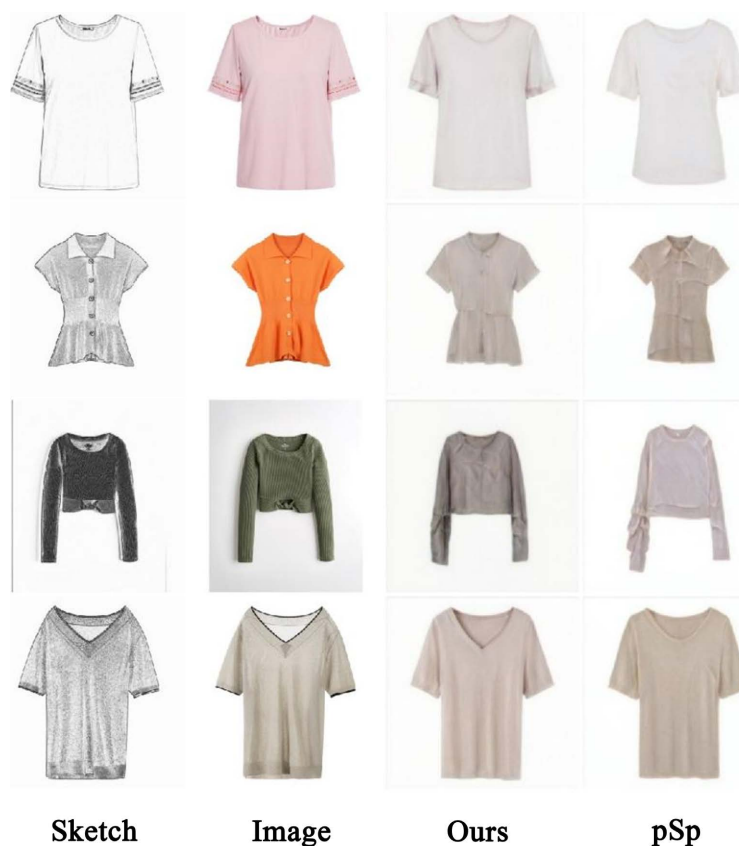


Figure 5. Comparison of the baseline experiments

图 5. 基线实验对比

本文选择从头开始训练 pSp 模型，使用本文提供的草图 - 服装数据集，并使用生成的测试集图像数据与本文的结果进行比较。从图 6 中可以看出，本文的方法对于原方法的改进是多方面的，经过第一行样本图像对比可知，原始图像在服装袖口处存在部分花纹，原方法无法有效生成相应纹理，而本文的方法可以生成对应纹理；从第二行与第三行样本图像对比可知，对复杂草图图像生成时，原方法生成的服装图像会生成不存在的褶皱，而本文的方法生成的图像更加真实且正常；第四行样本图像可以明显看出，服装样式为 V 领服装，而原方法却生成了圆领服装图像。

综上，本文着重提取草图内容信息的方法能够生成与草图一致的服装图像，并且在袖口、衣领以及服装样式等处，具有更高的细节描绘，生成的服装图像效果更好。

此外，为进一步表述本文的实验结果相较于其他方法质量更高，为方便与其他方法进行对比，本文使用上述草图 - 服装数据集分别作为基线模型方法的输入，从头开始训练模型。对比效果图如下图 6。前两列分别为草图原图与服装原图，后四列分别为在本文提供的数据集下所对应生成的服装图像。

图 6 中反映了不同的模型在草图 - 服装数据集上训练测试样本产生图像的典型情况。本文引入内容特征提取模块降低了草图风格对生成图像的影响，同时明显增强了生成服装图像的质量。

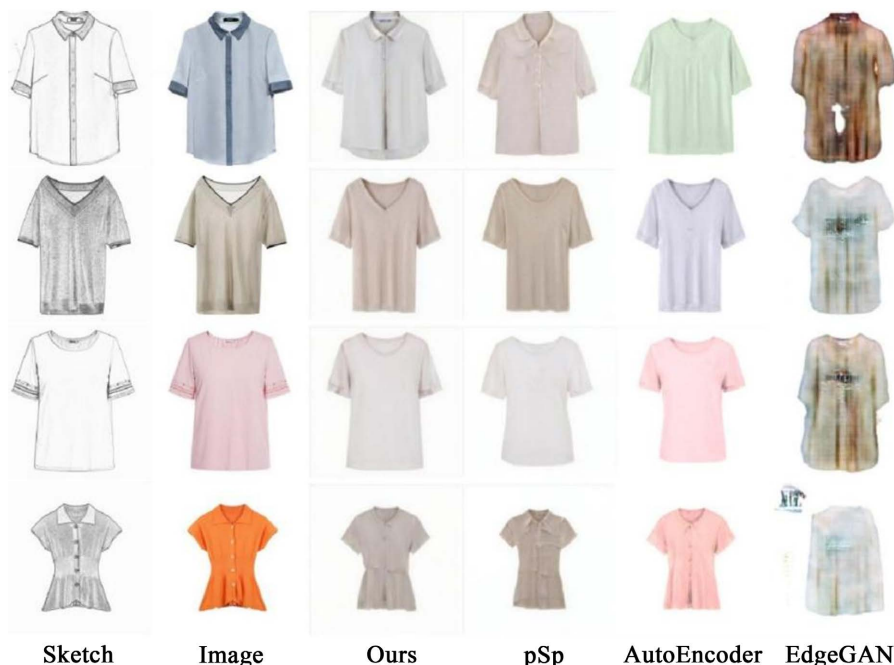


Figure 6. Comparison of the baseline experiments
图 6. 基线实验对比

5. 结论

我们将草图生成图像引入服装生产设计之中，有效改善了服装工业生产过程中，服装设计师可以根据设计草图，直观的了解生成样本的服装样式问题。通过利用内容特征提取模块，极大改进了原网络架构生成图像仍然存在草图风格特征的问题。实验结果表明，本文的改进模型具有良好的质量和效果，对于服装工业制造有着很高的实用意义。

尽管本文的工作得到了比较满意的结果，但是仍然存在着一一些问题需要去加强和改进。本文的服装数据集通过 python 第三方库生成服装对应的草图，致使部分草图语义信息过于杂乱。同时，服装设计师对于生成服装的色彩需要是十分丰富多彩的，因此需要调控生成服装图像的色彩。

参考文献

- [1] Chen, W.L. and Hays, J. (2018) Sketchygan: Towards Diverse and Realistic Sketch to Image Synthesis. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 18-23 June 2018, 9416-9425. <https://doi.org/10.1109/CVPR.2018.00981>
- [2] Lu, Y.Y., Wu, S.Z., Tai, Y., *et al.* (2018) Image Generation from Sketch Constraint Using Contextual Gan. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Proceedings of the 2018 European Conference on Computer Vision*, Springer Press, Berlin, 205-220. https://doi.org/10.1007/978-3-030-01270-0_13
- [3] Huang, X., Liu, M.Y., Belongie, S. and Kautz, J. (2018) Multimodal Unsupervised Image-to-Image Translation. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Computer Vision—ECCV 2018*, Springer, Cham. https://doi.org/10.1007/978-3-030-01219-9_11
- [4] Gao, C., Liu, Q., Xu, Q., Wang, L., Liu, J. and Zou, C. (2020) SketchyCOCO: Image Generation from Freehand Scene Sketches. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 13-19 June 2020, 5173-5182. <https://doi.org/10.1109/CVPR42600.2020.00522>
- [5] Richardson, E., *et al.* (2021) Encoding in Style: A StyleGAN Encoder for Image-to-Image Translation. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 20-25 June 2021, 2287-2296. <https://doi.org/10.1109/CVPR46437.2021.00232>
- [6] Karras, T., Samuli, L., Aittala, M., Hellsten, J., Lehtinen, J. and Timo, A. (2020) Analyzing and Improving the Image

- Quality of StyleGAN. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 13-19 June 2020, 8107-8116. <https://doi.org/10.1109/CVPR42600.2020.00813>
- [7] Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. <https://arxiv.org/abs/1409.1556>
- [8] Mirza, M. and Osindero, S. (2014) Conditional Generative Adversarial Nets. <https://arxiv.org/abs/1411.1784>
- [9] Zhu, J.Y., Park, T., Isola, P., *et al.* (2017) Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. 2017 *IEEE International Conference on Computer Vision*, Venice, Italy, 22-29 October 2017, 2242-2251. <https://doi.org/10.1109/ICCV.2017.244>
- [10] Cui, Y.R., Liu, Q., Gao, C. and Su, Z. (2018). FashionGAN: Display Your Fashion Design Using Conditional Generative Adversarial Nets. *Computer Graphics Forum*, **37**, 109-119. <https://doi.org/10.1111/cgf.13552>
- [11] Liu, B., Zhu, Y., Song, K. and Elgammal, A. (2021) Self-Supervised Sketch-to-Image Synthesis. *Proceedings of the AAAI Conference on Artificial Intelligence*, **35**, 2073-2081. <https://doi.org/10.1609/aaai.v35i3.16304>
- [12] Kingma, D.P. and Welling, M. (2014) Auto-Encoding Variational Bayes. <https://arxiv.org/abs/1312.6114>
- [13] Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B. and Belongie, S. (2017) Feature Pyramid Networks for Object Detection. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 21-26 July 2017, 936-944. <https://doi.org/10.1109/CVPR.2017.106>
- [14] Zhang, R., Isola, P., Efros, A., *et al.* (2018) The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. 2018 *IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 18-23 June 2018, 586-595. <https://doi.org/10.1109/CVPR.2018.00068>
- [15] Johnson, J., Alahi, A. and Li, F.-F. (2016) Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In Leibe, B., Matas, J., Sebe, N. and Welling, M., Eds., 2016 *European Conference on Computer Vision*, Springer, Cham, 694-711. https://doi.org/10.1007/978-3-319-46475-6_43
- [16] Guan, S.Y., Tai, Y., Ni, B.B., Zhu, F.D., Huang, F.Y. and Yang, X.K. (2007) Collaborative Learning for Faster StyleGAN Embedding. <https://arxiv.org/abs/2007.01758>
- [17] Heusel, M., Ramsauer, H., Unterthiner, T., *et al.* (2017) GANs Trained by a Two Timescale Update Rule Converge to a Local Nash Equilibrium. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 4-9 December 2017, 6626-6637. <https://dl.acm.org/doi/10.5555/3295222.3295408>
- [18] 王建欣, 史英杰, 刘昊, 黄海峤, 杜方. 基于 GAN 的手绘草图图像翻译研究综述[J/OL]. 计算机应用研究: 1-10. <https://kns.cnki.net/kcms/detail/detail.aspx?doi=10.19734/j.issn.1001-3695.2022.01.0027>, 2022-06-27.
- [19] Chen, S.-Y., Su, W.C., Gao, L., Xia, S.H. and Fu, H.B. (2020) DeepFaceDrawing: Deep Generation of Face Images from Sketches. *ACM Transactions on Graphics*, **39**, Article No. 72. <https://doi.org/10.1145/3386569.3392386>