可在线访问：www.sciencedirect.com

ScienceDirect

Procedia CIRP 00 (2023) 000–000

www.elsevier.com/locate/procedia

第33届CIRP设计大会

通过潜在空间操控优化CAD模型

Jannes Elstner *,a，Raoul Sch¨onhof *,**,a，Steffen Tauber[b]，Marco F. Huber[c]

[a]弗劳恩霍夫制造工程与自动化研究所 *IPA*，*Nobelstraße 12*，*70569* 斯图加特，德国[b]*evia solution GmbH*，*Am Wallgraben 100*，*70565* 斯图加特，德国[c]工业制造与管理研究所 *IFF*，斯图加特大学，*Allmandring 35*，*70569* 斯图加特，德国

* 共同贡献；** 通讯作者。电子邮箱：raoul.schoenhof@ipa-extern.fraunhofer.de

arXiv:2303.12739v1 [cs.CV] 9 Mar 2023

摘要

在自动化领域的 CAD 模型优化中，神经网络目前仅发挥着较小的作用。优化自动化能力等抽象特征颇具挑战，因为这些特征很难模拟，对基于规则的系统而言过于复杂，而且可供机器学习方法使用的数据也很少，甚至完全没有。另一方面，StyleCLIP 等能够操控图像中抽象特征的图像操控方法已取得显著成果。它们依赖于预训练生成对抗网络的潜在空间，因此也能够利用大量未标注CAD数据。本文展示了这类方法同样适用于优化 CAD 零件的抽象自动化相关特征。为此，我们扩展了 StyleCLIP，使其能够处理以体素模型形式表示的 CAD模型，其中包括使用 3D StyleGAN 和定制分类器。最后，我们通过优化各种 CAD 模型的可抓取性，展示了系统优化自动化相关特征的能力。

©2023 作者。由 Elsevier B.V. 出版。本文为开放获取文章，采用知识共享署名-非商业性使用-禁止演绎许可协议（http://creativecommons.org/licenses/by-nc-nd/4.0/）同行评审由第33届CIRP设计大会科学委员会负责。

关键词：StyleCLIP；StyleGAN；体素；CAD优化；NeuroCAD

1. 引言

在制造领域，装配过程是影响成本和价值创造的核心环节 [1]。因此，持续降低装配成本是一项重要目标。降低成本的一种方法是提高装配过程的自动化水平。随着劳动力成本不断上升、熟练工人日益稀缺，这一点变得愈发重要。遗憾的是，产品能否实现自动化装配取决于许多参数，例如材料、表面属性、包装、尺寸或零件形状。然而，零件形状的影响往往对装配过程的诸多方面尤为关键。例如，零件可能需要从其他零件中分离出来（散装货物），或由机器人抓取；还需要经过设计，以便轻松地与其他零件连接。

要实现可行的装配过程，就必须考虑这些方面。这要求在产品设计中考虑自动化装配过程的特征和局限性。然而，设计产品并针对适于自动化装配进行优化，需要充分了解生产方面的知识，还需要一种能够优化CAD模型、提升自动化能力的系统。目前尚不存在这样的系统。

借助 NeuroCAD流水线，我们已经证明，学习、评估、分析和定位 CAD零件中自动化相关特征等多个方面都是可行的 [2, 3, 4]。

在本研究中，我们提出了一种初步方法，旨在针对抽象的自动化相关特征优化CAD零件。我们的方法基于图像到图像转换领域的研究成果，尤其是利用生成对抗网络潜在空间的图像操作方法，例如 StyleCLIP。本文第 2 节简要介绍 NeuroCAD流水线、StyleGAN 和 StyleCLIP 方法的最新进展[2, 3, 4, 5, 6, 7]。第 3 节介绍了用于优化抽象特征的系统架构。最后，在第 4 节中，我们通过优化组件的可抓取性，展示了系统的能力。

2212-8271 © 2023 作者。由 Elsevier B.V. 出版。本文为开放获取文章，采用知识共享署名-非商业性使用-禁止演绎许可协议（http://creativecommons.org/licenses/by-nc-nd/4.0/）。本文由第33届CIRP设计会议科学委员会负责同行评审。

我们使用自行构建的小型数据集进行优化。第 5 节和第 6 节总结了我们的工作，并展望了未来。

2. 相关工作

2.1. CAD 优化

借助神经网络优化三维拓扑结构时，针对各种各样的问题领域，已有许多方法。优化对象包括应力、变形、热传递或流体力学等机械特征。这些方法通常会取代进化算法等优化方法或搜索策略。特别是在有限元方法领域，已有显著改进，例如：[9, 10]不过，三维形状优化算法通常只处理那些可通过解析方程或数值模拟描述的任务。无法实际模拟的问题也无法优化。目前尚无基于学习到的特征对三维形状进行优化的方法。

2.2. 潜在空间操控

过去几年，许多研究探索了基于预训练生成对抗网络（GAN） [8, 7, 11, 12]潜在空间的图像操控。对于潜在空间操控方法，首先会将图像嵌入 GAN 的潜在空间，这一过程称为 GAN 反演。随后，在潜在空间中操控所获得的潜在编码，例如沿着与场景空间布局等属性相对应的方向进行遍历 [13]。

StyleGAN [5, 6] 是最常用于潜在空间操控的 GAN，这得益于其解耦的潜在空间、先进的图像质量，以及支持针对特定尺度进行操控的架构。

2.3. StyleCLIP

StyleCLIP [7] 是一种使用文本提示操控图像的方法。为此，StyleCLIP 依赖预训练的 StyleGAN 生成器和预训练的对比语言-图像预训练（CLIP） [14] 模型，通过评估文本提示与给定图像的匹配程度来引导操控。

StyleCLIP 使用三种不同的方法来操纵图像：潜在优化、使用潜在映射器进行优化，以及全局方向。我们的 CAD 优化基于前两种方法。全局方向方法似乎并不适用，因为它假设针对某个感兴趣属性的所有图像操纵在潜在空间中都采用相同的步长。这一假设对 CAD 优化的限制过多，因为优化可能针对某个特定组件具有很强的针对性。

2.3.1. 潜在优化

图像编辑被建模为一个优化问题，并通过直接优化潜在编码来求解。具体而言，给定来自 StyleGAN 扩展潜空间 $\mathcal{W}+$ 的源潜在编码 t 作为指令，图像编辑被建模为以下优化问题：

$$\arg\min_{w\in\mathcal{W}+} D_{CLIP}(G(w),t) + \lambda_{L2}\|w - w_s\|_2 + \lambda_{ID}\mathcal{L}_{ID}(w), \quad (1)$$

其中，D_{CLIP} 是一个函数，用于计算其两个参数的 CLIP 嵌入之间的余弦距离 [7]。潜在空间中的L_2 距离用于控制操纵后的图像与源图像之间的相似度。此外，在操纵人脸时，还会加入身份损失，以保留能够确立人物身份的特征。

2.3.2. 潜在映射器

潜在映射器是一种针对特定文本提示t 训练的映射网络；给定图像的源潜在编码，它可以推断出潜在空间中的操纵步长。给定图像源潜在编码的数据集后，潜在映射器 M_t 经过训练，以最小化以下损失：

$$\begin{aligned}\mathcal{L}(w) = {} & D_{CLIP}(G(w + M_t(w)), t) \\ & + \lambda_{L2}\|M_t(w)\|_2 \\ & + \lambda_{ID}\mathcal{L}_{ID}(w).\end{aligned} \quad (2)$$

此处，通过将推断出的操纵步长 $M_t(w)$ 加到潜在编码 w 上，得到优化后的潜在编码。操控后图像与源图像之间的相似度由操纵步长的 L_2 范数控制。

3. 方法

我们的 CAD 优化基于 StyleCLIP，同时采用其潜在优化方法和潜在映射器方法。StyleCLIP 依靠 StyleGAN 提供解耦潜在空间，并依靠 CLIP 提供视觉语言表征。要将 StyleCLIP 用于 CAD模型 优化，StyleGAN 必须能够对 CAD模型 进行编码，而 CLIP 则必须替换为能够评估优化目标的损失网络。对于后者，我们使用比较器网络，根据原始部件与优化后的部件的优化程度对二者进行比较。请注意，与 StyleCLIP 相比，我们的优化并非由文本驱动，系统中也不使用任何自然语言处理。

我们采用基于体素的表示来表示 CAD模型，因为体素网格是二维图像在三维空间中的对应形式，而大多数潜在空间操控方法都以二维图像为操作对象。具体而言，我们将数据栅格化为分辨率为 64x64x64 的二值网格，以尽可能捕捉更多细节，同时满足计算预算。如何选择合适的分辨率是体素模型面临的典型难题，我们已在先前的工作 [2] 中讨论过。

3.1. StyleGAN

为了在三维体素模型上训练 StyleGAN，我们采用了与 [15]，类似的设置：将所有二维操作替换为相应的三维操作，并在

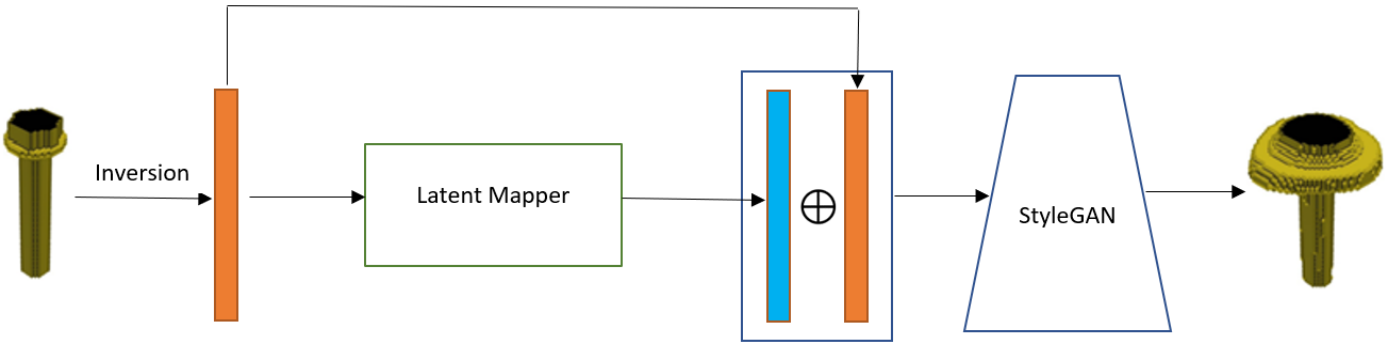


图 1. 使用潜在映射器进行可抓取性优化。原始部件被反演为源潜在编码（橙色），并输入潜在映射器。映射器生成一个向量（蓝色），将其添加到源潜在编码中，从而得到优化后的潜在编码。StyleGAN 根据优化后的潜在编码生成优化后的组件。在训练过程中，使用比较器将优化后的组件与原始部件进行比较。灵感来自 [7]。

生成器末尾添加 tanh 激活函数。此外，我们移除了生成器中的噪声输入，因为它们会模拟随机变化 [6]。，而 CAD 模型通常不存在这种变化。最后，为了提高训练的稳定性并促进收敛，我们采用了 [16]中所述的类别条件化。

鉴于三维深度学习对 GPU 显存的高需求，我们大幅减少了 StyleGAN 的可训练参数数量。原始 StyleGAN2 生成器和判别器各约有 3000 万个可训练参数，而我们将其分别减少至约 210 万个。

对于生成对抗网络的评估，Fr´echet Inception Distance (FID) [17] 是最常用的指标。FID 依赖于一个在二维自然图像上训练的 InceptionV3 [18] 网络，因此不能直接用于评估三维生成对抗网络。我们遵循 [15]的方法，改为在合成的三维体素模型的轴向面、冠状面和矢状平面的中间切片上计算 FID 分数。

在对 StyleGAN 进行训练和评估后，我们使用 [6] 提出的反演方法的简化版本，将组件嵌入生成器的潜在空间。

3.2. 比较器

需要将 CLIP 模型替换为能够评估优化目标的损失网络。通常，损失网络是分类器或回归器，但我们选择使用比较器网络 [19]。，根据一对 CAD模型的优化程度对其进行比较。

使用比较器的主要优势在于构建数据集要容易得多，尤其是在我们最感兴趣的用例中：优化自动化属性。与评估 CAD模型的自动化属性相比，判断给定的两个模型中哪一个得分更高要容易得多。另一方面，比较器只能回答给定的两个模型中哪一个更好，却不能说明好多少。此外，也无法指定具体的优化目标，例如像 [11],中那样设定期望的目标年龄。在优化问题中，这并不会造成问题，因为目标通常难以确定或论证。以优化生产成本为例，选择任何最优价格作为目标，本质上都是任意的。在实践中，优化的起点可以只是希望 CAD模型的成本更低。

比较器架构基于 [19]提出的双通道网络。比较两个 CAD 模型时，会沿通道轴将它们拼接起来，并作为一幅双通道图像输入卷积网络。

给定体素模型 V_1 和 V_2，，训练比较器以预测两者中哪一个优化程度更高。这可以通过训练比较器来最小化以下损失实现：

$$\mathcal{L}(V_1, V_2, y_{true}) = H(y_{true}, C(V_1, V_2)), \tag{3}$$

其中，C 表示比较器，y_{true} 表示标签，H 用于计算其参数之间的二元交叉熵。因此，优化程度比较被视为一个二分类问题，包含两个类别：“V_1 优化程度更高”和“V_2 优化程度更高 / V_1 优化程度更低”。

3.3. 优化方法

我们的优化方法在 StyleGAN 的中间潜空间 $\mathcal{W}$ space 中运行，而不是在原始扩展潜空间 $\mathcal{W}+$ 中运行，因为使用 $\mathcal{W}$ 可以获得更高的感知质量和更一致的结果。这也意味着，对于潜在映射器，我们采用 StyleCLIP 的设置，仅使用一个映射网络，而不是针对不同特征层级使用三个映射网络。对于损失函数，我们将 CLIP 损失替换为优化损失，该损失使用比较器比较源组件和优化后的组件。优化后的组件与原始组件之间的相似度仍由潜空间中推断出的操纵步长的 L_2 范数控制，同时还加入了数据空间中的 L_2 损失，因为从潜空间到数据空间的映射并不总是平滑的。我们还移除了身份损失，因为 CAD模型不像人脸那样具有身份。

潜在映射器的最终损失如下：

$$\begin{aligned}\mathcal{L}(V, w) = {} & H(0, C(V, G(w + M(w)))) \\ & + \lambda_1 \|M(w)\|_2 \\ & + \lambda_2 \|G(w + M(w)) - V\|_2,\end{aligned} \tag{4}$$

其中G 为预训练生成器，C 为预训练比较器，V 为反演潜变量w的源组件。同样，潜在优化的优化问题重新定义如下：

$$\arg\min_{w\in\mathcal{W}} H(0, C(V, G(w))) + \lambda_1 \|w - w_s\|_2 + \lambda_2 \|G(w) - V\|_2, \quad (5)$$

其中w_s 是源组件的源潜在编码。

4. 结果与讨论

为了测试我们的优化方法，我们创建了一个围绕组件可抓取性的小型玩具数据集。组件的可抓取性通常描述工业机器人抓取组件的难易程度，这取决于组件的几何结构，尤其是平坦表面的面积。为简化问题，我们将可抓取性简化为工业夹爪抓取组件的难易程度，并将组件限定为螺钉和螺栓这两类。用于训练比较器的数据集包含 1000 对螺钉/螺栓，并标注了每对中可抓取性更好的组件。

4.1. StyleGAN 的训练与评估

StyleGAN 需要能够编码我们希望优化的组件，因此我们使用 MCB 数据集的一个子集 [20], 进行训练，该子集筛选出了视觉上与螺钉和螺栓相似的类别。最终得到 9 个类别，共包含 22 788 个组件。鉴于数据量较少，我们采用自适应伪增强 [21] 作为有限数据训练策略。

评估 StyleGAN 时，我们计算了训练数据集中全部 22 788 张图像与 45 576 张合成图像之间的逐切片 FID；每个训练数据集标签对应两张合成图像。由于 InceptionV3 网络无法直接处理 64x64x1 的中间切片，因此在输入网络之前，先通过上采样将其转换为 128x128x3，并复制通道维度。

我们对 StyleGAN 进行训练，直到逐切片 FID 收敛。这大约耗时两天，冠状面、轴向面和矢状面 FID 分别收敛至 90.77、76.12 和 61.89。图2 展示了合成组件的示例。

我们还尝试在 ABC 数据集[22], 上训练 StyleGAN。该数据集更加复杂，规模也大得多，包含超过一百万个 CAD模型。尽管尝试了各种训练参数和架构修改，训练仍未成功。目前尚不清楚是我们的设置不合适，还是 StyleGAN 不适用于在更通用、更复杂的 CAD 数据集上进行训练。

4.2. 可抓取性优化

StyleGAN 训练完成并随后训练潜在映射器后，我们使用不同参数运行了优化方法。在图 3 中，我们提供了几个

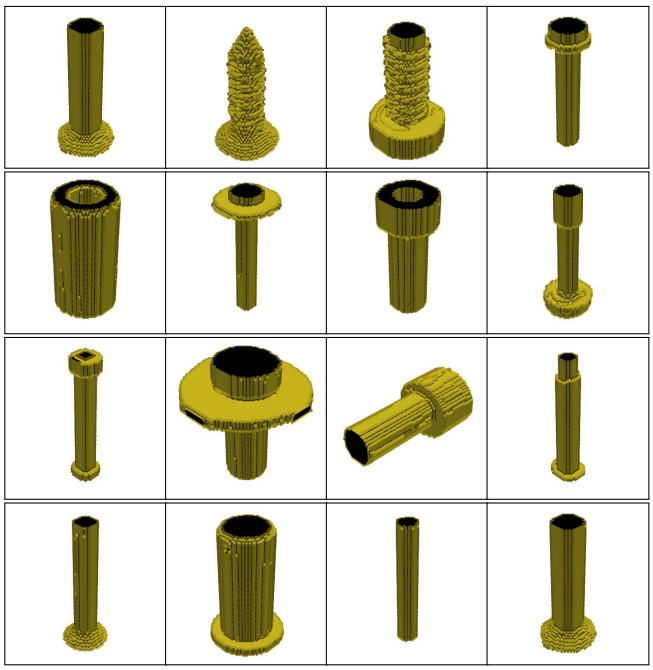


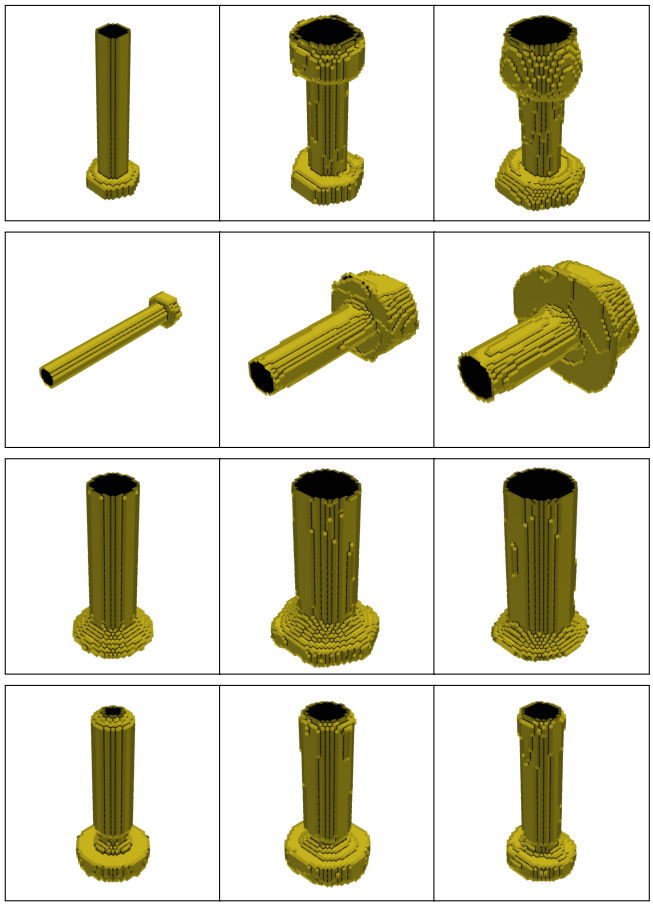
图 2. 在 MCB 数据集上完成训练后，从收敛的 StyleGAN 生成器合成的 16 个随机组件。

我们精心挑选了若干感知质量较高且合理的可抓取性优化示例。可抓取性主要根据可供抓握的平坦表面面积来建模，而优化后的组件主要增大了螺钉和螺栓头部的尺寸，这与我们的可抓取性模型相符。遗憾的是，许多优化结果更像图 4 中的示例，存在明显缺陷，例如外观不完整、变形，或带有明显的伪影，总体而言并不真实。

不切实际的优化使实际应用变得困难，也会对潜在映射器的训练产生负面影响；在训练过程中，这些优化可能会使比较器感到困惑，因为比较器仅使用逼真的图像进行训练。当前的潜在空间操控方法并不以生成逼真结果为目标，而是依赖潜在空间的特性。具体而言，这些方法依赖于源潜在编码附近区域中的潜在编码，因此生成的图像大多也较为逼真。对于 CAD 模型的操控而言，这还不够，因为 CAD 模型通常遵循更严格的形状规则，而潜在空间中的操控不太可能遵循这些规则。

因此，问题源于数据表示方式，并非体素模型独有，也涉及其他常见表示方式，例如点云或基于视图的表示。未来工作可以探索使用自然语言表示的生成模型，其中 CAD 模型通过创建它们时使用的操作来定义[23]。与我们的 3D StyleGAN 相比，使用自然语言表示 CAD 模型的 GAN 模型已被证明能够处理复杂得多的数据集，同时大幅降低训练成本。

最后，人们通常希望保护



(a) 原始部件　(b) 潜在映射器　(c) 潜在优化

图 3. 使用潜在映射器和潜在优化进行的可抓取性优化。

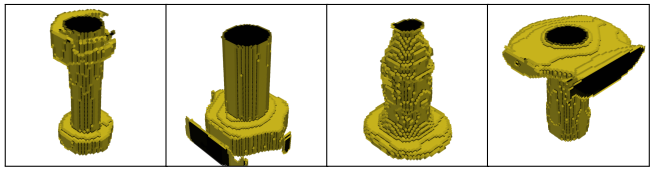


图 4. 不切实际或失真的优化示例

组件中的特定结构，例如它们承担核心功能时。可以扩展潜在优化方法，通过添加一个额外的 L_2 损失项来惩罚对受保护结构的改动，或者在 G(w) 中撤销对受保护结构的改动，从而实现这一点。扩展潜在映射器则更具挑战性，因为它需要一种机制，在训练期间选择合理的受保护结构。

我们实验所使用的所有代码都可以在我们的代码仓库中找到 [24]。

5. 结论

在本研究中，我们展示了如何调整 StyleCLIP 等图像操控方法，以优化 CAD模型。我们还通过优化螺钉和螺栓的可抓取性，展示了系统的能力。许多优化结果具有较高的感知质量，按照我们对可抓取性的定义也合乎情理，但结果仍存在明显缺陷，例如不切实际或扭曲的优化结果。考虑到我们的数据集仅包含 1000 个样本，而且从体素表示转向自然语言表示有望修复系统已发现的缺陷，我们仍然认为，利用潜在空间操控来优化抽象CAD特征具有巨大潜力。我们认为，像我们这样的系统可以为部件设计提供一种替代方案，减少对领域专家知识与经验的依赖，从而推动制造自动化。

6. 未来工作

这两种优化方法都有若干局限。优化模型在对称性和功能性方面的质量并不稳定。有时，优化的特征不正确，或者优化结果失真。此外，训练 3D-CNN 的计算开销很大。从可用性的角度来看，还需要识别并标记某些受保护的功能结构。要实现可用于工业的优化算法，就必须应对这些挑战。

致谢

研究人员谨向德国联邦教育与研究部（BMBF）以及卡尔斯鲁厄项目管理机构（PTKA）致谢，感谢其通过“KiMont”项目提供资助（资助编号 02K20K512）。

参考文献

[1] Spur, Guenter, Helwig, H.-J.，《装配技术导论》，载于：G. Spur, T. Stoeferle，《制造技术手册》，第5卷，《连接、搬运与装配》，卡尔·汉瑟出版社，慕尼黑，1986年。[2] Schoenhof, R., Fechter, M.，《借助深度学习实现自动化能力评估》，Procedia CIRP Design，第433-438页，2020年。[3] Schoenhof, R., Werner, A., Elstner, J., Zopcsak, B., Awad, R., Huber, M.，《借助可解释人工智能方法实现自动化设计评估中的特征可视化》，Procedia CIRP Design，第331-336页，2021年。[4] Schoenhof, R., Elstner, J., Manea, R., Tauber, S., Awad, R., Huber, M.，《借助深度残差自编码器简化CAD特征学习》，Procedia CIRP Design，第84-88页，2022年。[5] Karras, T., Laine, S., Aila, T.，《生成对抗网络的基于风格的生成器架构》。载于 IEEE/CVF 计算机视觉与模式识别会议论文集，第4401-4410页，2019年。[6] Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.，《分析并改进 StyleGAN 的图像质量》。载于

IEEE/CVF 计算机视觉与模式识别会议论文集，第8110-8119页，2020年。[7] Patashnik, O., Wu, Z., Shechtman, E., Cohen-Or, D., Lischinski, D.，《StyleCLIP：文本驱动的 StyleGAN 图像操控》。载于 IEEE/CVF 国际计算机视觉会议论文集，第2085-2094页，2021年。[8] Shen, Y., Gu, J., Tang, X., Zhou, B.，《解读生成对抗网络的潜在空间以实现语义人脸编辑》。载于 IEEE/CVF 计算机视觉与模式识别会议论文集，第9243-9252页，2020年。[9] Spathopoulos, Stefanos C. and Stavroulakis, Georgios E.《基于有限元分析和人工神经网络方法的钣金成形回弹预测》。Applied Mechanics，第1卷，110，2020年，97页。[10]Sumudu Herath, Udith Haputhanthri，《利用条件生成对抗网络进行拓扑优化设计与失效预测》，International Journal for Numerical Methods in Engineering，10.1002/nme.6814，2021年。[11] Alaluf, Y., Patashnik, O., Cohen-Or, D.，《风格决定一切：使用基于风格的回归模型进行年龄转换》。ACM图形学汇刊（TOG），40(4)，第1-12页，2021年。[12] Richardson, E., Alaluf, Y., Patashnik, O., Nitzan, Y., Azar, Y., Shapiro, S., Cohen-Or, D.，《以风格编码：用于图像到图像转换的 StyleGAN 编码器》。载于 IEEE/CVF 计算机视觉与模式识别会议论文集，第2287-2296页，2021年。[13] Yang, C., Shen, Y., Zhou, B.，《场景合成的深度生成表示中涌现的语义层次结构》。International J ournal of Computer Vision，129(5)，第1451-1466页，2021年。[14] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sutskever, I.，《从自然语言监督中学习可迁移的视觉模型》。载于国际机器学习大会论文集，第8748-8763页，2021年。[15] Hong, S., Marinescu, R., Dalca, A. V., Bonkhoff, A. K., Bretzner, M., Rost, N. S., Golland, P.，《3D-StyleGAN：用于三维医学图像生成建模的基于风格的生成对抗网络》。

深度生成模型，以及数据增强、标注与瑕疵，第 24-34 页，2021年。[16] Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., Aila, T., 使用有限数据训练生成对抗网络。神经信息处理系统进展，33，第 12104-12114 页，2020年。[17] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S., 采用双时间尺度更新规则训练的生成对抗网络收敛至局部纳什均衡。神经信息处理系统进展，30，2017年。[18] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 重新思考用于计算机视觉的 Inception 架构。载于 IEEE 计算机视觉与模式识别会议论文集，第 2818-2826 页，2016年。[19] Zagoruyko, S., Komodakis, N., 通过卷积神经网络学习比较图像块。载于 IEEE 计算机视觉与模式识别会议论文集，第 4353-4361 页，2015年。[20] Kim, S., Chi, H. G., Hu, X., Huang, Q., Ramani, K., 面向分类与检索任务的大规模标注机械零部件基准数据集，基于深度神经网络。载于欧洲计算机视觉会议，第 175-191 页，2020年。[21] Jiang, L., Dai, B., Wu, W., Loy, C. C., Deceive D：用于有限数据 GAN 训练的自适应伪增强。神经信息处理系统进展，34，2021年。[22] Koch, S., Matveev, A., Jiang, Z., Williams, F., Artemov, A., Burnaev, E., Panozzo, D., ABC：用于几何深度学习的大型 CAD 模型数据集。载于 IEEE/CVF 计算机视觉与模式识别会议论文集，第 9601-9611 页，2019年。[23] Wu, R., Xiao, C., Zheng, C., DeepCAD：用于计算机辅助设计模型的深度生成网络。载于 IEEE/CVF 国际计算机视觉会议论文集，第 6772-6782 页，2021年。[24] https://github.com/RaoulusHPC/KiMont_StyleGAN3D。