arXiv:2304.09384v1 [cs.CV] 19 Apr 2023

SP-BATIKGAN：一种用于对称图案生成的高效生成对抗网络

*†‡*Chrystian*，‡*Wahyono*

*三星印度尼西亚研究院，雅加达
†马克斯·普朗克生物控制论研究所，马克斯·普朗克学会，图宾根
‡加查马达大学计算机科学与电子学系，日惹

摘要

鉴于人工智能艺术引发的争议，我们的研究聚焦于让所有人，尤其是艺术家，都能利用人工智能，在有限的数据和条件下创作人工智能艺术。我们关注几何对称图案生成，这类图案常见于许多艺术作品中，例如葡萄牙瓷砖、摩洛哥瓷砖，以及东南亚文化遗产蜡染。对称图案生成是一项复杂的任务，先前的研究所创建的模型过于专用，只适用于特定图案。我们公开提供了[1]，首个专为此任务收集的高质量对称图案数据集，共包含 1,216 个图案，均直接取自设计文件。随后，我们提出对称图案约束（SPE）损失，以利用当前图像分布中已有的底层对称结构。SPE 可改进并加速任意生成对抗网络配置的训练。结合高效注意力，与有限条件下的先进生成对抗网络 FastGAN 相比，SP-BATIKGAN 将 FID 分数从 **110.11**降至 **90.76**，降幅达 **18%** ；模型多样性召回分数则从 **0.047** 提升至**0.204**，增幅达 **334%** 。

索引词— 生成对抗网络、应用、无监督图像生成、蜡染生成、对称图案生成

1. 引言

蜡染是东南亚的文化象征，也是联合国教科文组织认定的“人类口述和非物质遗产代表作”，是一种创作精美纺织品的古老艺术。近年来，许多研究 [1, 2, 3, 4, 5, 6] 开始利用人工智能创作具有独特复杂图案风格的艺术作品，例如蜡染。然而，这些早期初步研究面临诸多困难：生成对抗网络训练不稳定、需要大量数据样本、难以获取适用于无监督图像生成的高质量数据，而且每件作品的图案结构都十分复杂且差异很大。

在本研究期间，作者获得了印度尼西亚海洋奖学金以及马克斯·普朗克基金会汉里德卓越基金会的资助。[1]采用CC-BY-NC-SA许可证的ITB-mBatik数据集可在此获取：
https://data.mendeley.com/datasets/7hzr5539ws

首先，尽管大型数据集更受青睐且十分重要，图像质量也同样关键。其次，即使拥有数据集，将不同的图案类型或风格混合在一起，也不一定能保证取得良好效果；也就是说，即便有数千张图像，生成结果的质量仍可能不尽如人意[1, 4]。近期研究 [3] 在通用性与质量之间进行权衡，因为其方法是通过固定损失函数，将给定风格限制在单一特定图案类型上。

为解决这些问题，我们首先收集并预处理高质量的蜡染数据。随后，我们着重于推广对称图案生成任务。我们提出了一个框架，使生成模型能够在对称图案任务上高效训练生成器；这也是蜡染及其他文化艺术品的关键特征之一。基于该框架，我们进一步直观地推导出**对称图案约束（SPE）**损失。借助SPE，我们能够提升FastGAN[7]在对称图案生成任务上的图像质量和细节表现，达到最先进水平。

简而言之，我们的贡献如下：

- 我们将对称图案生成任务推广到其他不同的对称图案类型（第 3.2 节）
- 我们通过利用数据集中现有的对称结构推导出 SPE 损失，构建了一种直观的学习方法（第 3.2.1 节）
- SPE 损失提升了最先进水平的 FastGAN；与高效注意力结合后，SPE 损失可加速生成对抗网络训练、降低 FID，并大幅提高召回率（第 4.2 节）

2. 相关工作

先前相关的蜡染生成研究。近年来，人们对与 AI 相关的蜡染生成课题产生了浓厚兴趣，例如 [6], [5], 和 [2] ，它们分别采用文本到图像、条件生成对抗网络和神经风格迁移。对于无监督生成任务，早期研究（如 [4], ）采用简单的 DC-GAN 来进行蜡染图案合成。其他研究 [1] 则使用更多样化的数据集继续这项研究。然而，对于



图 1：ITB-mBatik数据集。最终提出的数据集中1,216组无缝且对称的图案对示例。

这两项研究都没有衡量任何指标，而且生成图像只呈现出模糊的蜡染特征，并带有过多噪声。针对无监督蜡染生成任务的最新研究是 BatikGAN [3], ，然而 BatikGAN 存在过多限制。即使采用了五种新的自定义损失函数，它也只考虑一种图案类型。生成图像的分辨率也很低，每个图像块仅为 32×32 像素。

有限计算资源和数据集条件下生成对抗网络的发展。从最初的生成对抗网络到撰写本文时的最先进水平（如 StyleGAN2 [8], ），用于训练生成对抗网络模型的数据量一直非常庞大。然而，近期的进展，如自适应判别器增强（ADA） [9]和面向高效数据生成对抗网络的可微增强 [10] ，可以将生成对抗网络对数据的需求降低到仅需数百张图像。这促使我们采用最先进的FastGAN [7] 作为基线；在有限计算资源和数据集环境下，其性能优于此前最先进水平的StyleGAN2。

3. 方法

为了获得良好的图像生成结果，输入图像的质量至关重要。对于我们的蜡染图案任务而言，传统服装图案很难获取，尤其是高分辨率图案，而这正是以往大多数研究的不足之处。以往研究使用了UI 蜡染数据集 [11] ，图像仅通过相机拍摄，质量明显低于标准。因此，为了开展后续研究，我们获取并创建了自己的、更现代的数字蜡染数据集。我们提出了一个更优质的新型高质量公开数据集，该数据集直接源自设计文件，称为万隆理工学院现代蜡染（ ITB-mBatik）数据集 [12], ，这是一个更适合无监督图像生成任务的数据集，如图1所示。

3.1. 完整图案生成的低效性与对称类型差异

不仅难以收集高质量图像，先前工作中的另一个显著问题 [1, 4] 是直接在完整图案上训练生成器的低效率。使用完整的重复图案拟合生成器会不必要地增加真实图像分布的复杂性，其中生成器-

器需要考虑许多不同的任意重复图案，从而导致结果不尽如人意。定义“图像块”为能够重构原始重复图案的最小图像子集，$\mathbf{p} \subset \mathbf{P}$ ，其中该图像块可以重构原始重复图案**图像块(p)** $= \mathbf{P}$。如果通过重复子集$\mathbf{p}$可以完整重构重复图案$\mathbf{P}$ ，那么与其冗余地训练生成器去学习整个重复图案，不如生成该子集图案，这是一种更高效的方法。

对于图案块生成，图案块图案本身具有很高的自由度和灵活性。除了生成图像块之外，为了限制图像块并使其便于处理，我们还规定，数据集中的每个图像块都必须具有一种共同的对称变换。使用具有不同对称性的生成对抗网络进行训练会导致不对称问题，并损害输出图像的美感。因此，对数据集施加约束，使其至少具有一种相同的对称变换，可以保持对称性并避免不对称。

3.2. 二维基本等距变换下的对称图案块图案生成任务

我们将任务定义在最常见的变换群中，即二维空间中的欧几里得群，或 E(2) 的等距变换（请注意，欧几里得群之外也存在对称性）。首先，将数据集中的图案块图像 $\mathbf{x}$ 定义为$\mathbf{x} \in \mathcal{X}$。若图案块经过变换后与自身相等，即$\mathbf{x} = T\mathbf{x}$，则称该图案块具有对称变换 T。我们希望得到的公共对称变换集 S，是一个由对称变换构成的集合，其中每种变换对数据集中的每个图像块都有效且成立。要构成对称图案生成任务，至少必须存在一种公共对称变换，即 $S \neq \varnothing$。最后，设 T_s 为 S 中的一种对称变换，使得以下性质成立：

$$\forall T_s \in S, \forall \mathbf{x} \in \mathcal{X}, (\mathbf{x} = T_s\mathbf{x}) \tag{1}$$

3.2.1. 对称图案约束

由于我们知道每个输出图像都必须具有对称性，根据公式 1，这启发我们从一开始就直接对生成器中的每个图案施加对称性约束。我们将这一损失称为**对称图案约束（SPE）**，记为 $\mathcal{L}_{\text{SPE}}$。我们将生成器输出图像与其几何对称对应图像之间的相似度进行汇总。图 2 展示了 $\mathcal{L}_{\text{SPE}}$ 仅使用生成器的原始输出，并在*DiffAug* T 对真实图像和生成图像进行增强之前的情况。最后，为了汇总生成图像及其对称对应图像，我们的SPE损失基于相似度损失 $\mathcal{L}_{\text{Sim}}$。因此，一般而言，对称图案约束损失的形式化定义如下：

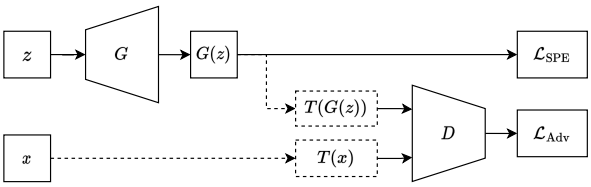

图 2：对称图案约束损失概览。设 z，G，G（z）分别为潜在输入、生成器和生成图像，x，D，以及虚线 T 分别表示真实图像、判别器和 DiffAug 操作。对于 $\mathcal{L}_{\text{SPE}}$，我们直接获取应用 DiffAug 之前的生成图像 G（z）。与 $\mathcal{L}_{\text{Adv}}$ 对抗损失不同，SPE损失仅应用于生成器。



图 3：对称图案约束操作示意图。我们在 ITB-mBatik数据集上使用的全部四种操作。左图显示生成器输出的原始图像及翻转轴线，右图显示翻转结果，用于比较两幅图像的相似度。

$$\mathcal{L}_{\text{SPE}}(G(z)) = \frac{1}{|S|}\sum_{T_s \in S} \mathcal{L}_{\text{Sim}}(G(z), T_s G(z)) \qquad (2)$$

其中，$G(z)$ 是直接从生成器取得的生成输出图像，$\frac{1}{|S|}$ 是一种在 S 基数较高时使用的归一化方法。

3.2.2. 面向 *ITB-mBatik* 的对称图案约束实现

然而，在我们的实现中，为了实用性，我们降低了方法的通用性。首先，Cartan-Dieudonn´e 定理 [13]指出，复杂的高维等距变换可以由多个基本的简单反射复合而成。因此，为了降低损失的复杂度，我们将推导一种损失函数，重点关注 ITB-mBatik数据集 中一组可观察且直观的反射对称性。

我们发现了 4 种简单且最直观的对称变换，满足公式 1：沿水平轴、垂直轴以及正斜率对角线的反射和沿负斜率对角线的反射，如图 3 所示。此外，在实际应用中，来自 R 的变换操作必须可微，并支持自动微分以进行反向传播。因此，设 R = {h-flip、v-flip、p-flip、n-flip}，然后我们通过构造 R 或 $\mathcal{P}$（R）的幂集，进一步优化并找出最佳子集操作。不过，为了减少计算量，我们之后在 SPE 损失中仅试验了四种子集。

接下来，我们以公式 2 为模板，将 SPE 实现为：

$$\mathcal{L}_{\text{SPE}}(G(z)) = \frac{1}{|R|}\sum_{T_r \in R} \mathcal{L}_{\text{Sim}}(G(z), T_r G(z)) \qquad (3)$$

其中，对于 $\mathcal{L}_{\text{Sim}}$，我们使用简单的相似度指标 L2，但$\mathcal{L}_{\text{Sim}}$ 也可以使用 L1 或 LPIPS[14] 等其他函数。

4. 实验

数据集。我们使用了自己的 ITB-mBatik数据集 [12]。我们从原始设计文件中预处理、裁剪并缩放重复区域的“图像块”，使其达到 1024×1024 像素。此外，我们通过多阶段采样处理平移对称性，并增加数据集的训练规模。

指标。在评估过程中，我们将衡量五种无监督图像生成指标。为更深入地了解生成器的性能，我们将使用 Fr´echet Inception 距离 [15], 改进的精确率与召回率 [16], 以及密度和覆盖率 [17]。

GAN 训练概要。我们使用带有 DiffAug 的 FastGAN 作为基线。我们将输出图像尺寸设为 254×254；对于生成对抗网络的损失函数，我们认为不同损失函数的表现相近 [18], , 因此选用铰链损失 [19], 因为它简单且计算速度快。优化器方面，我们使用 Adam 优化器 [20]，学习率为 $2e-4$, β_1 =0.5, $\beta_2 = 0.999$。随后，我们训练生成对抗网络五次，并报告其中最佳的 FID 结果。每个生成对抗网络训练 200k 个小批次，批次大小为 8。最后，每训练 1k 个小批次，我们都会保存模型，以便进行评估并在必要时提前停止训练。

使用的环境。硬件方面，研究环境运行于 i5-3470 中央处理器、16GB 随机存取存储器，以及配备 12GB 显存的 RTX 3060 图形处理器上。软件方面，研究环境使用 PyTorch 构建生成对抗网络模型，其中包含生成器和判别器；并使用 TensorFlow 运行 InceptionV3 和 VGG-16 进行评估。

4.1. 对称图案约束分析

首先，我们对 R := {hv、np、hvnp、 [hv、np]} 进行实验，其中 h、v、n、p 分别表示水平翻转、垂直翻转、负对角线翻转和正斜向翻转；[hv、np]则表示 hv翻转和 np 翻转交替使用的操作。

表 1 展示了每组 R 操作组合的最佳结果。我们发现，hv翻转组合将 FID 从 110.106 降至 97.852，达到最低值，并使基线召回率提升一倍，从 0.047 增至 0.094。

表 1：SPE损失实验结果。**粗体** 数字表示最佳得分，下划线表示次佳得分。

SPE损失配置		FID↓	保真度 Pre↑	保真度 Den↑	多样性 Rec↑	多样性 Cov↑
基线		110.106	0.821	0.863	0.047	0.557
L2	***hv***	**97.852**	0.762	0.696	**0.094**	0.490
	np	103.782	**0.838**	0.866	0.068	0.564
	hvnp	102.624	0.835	**0.904**	0.067	**0.570**
	$[hv, np]$	111.291	0.819	0.762	0.048	0.463

表2：基于注意力的生成器结合SPE约束的实验结果。**粗体**数字表示最佳分数，下划线表示次优分数。注意力机制名称旁边的第二列表示应用注意力的特征分辨率。

模型		FID↓	保真度 Pre↑	保真度 Den↑	多样性 Rec↑	多样性 Cov↑
基线+SPE		97.852	0.762	0.696	0.094	0.490
+SAN	8^2	122.27	0.797	0.774	0.011	0.428
	16^2	139.78	**0.875**	0.773	0.007	0.298
+EAttn	8^2	99.62	0.772	0.775	0.087	0.461
	16^2	99.62	0.772	0.763	0.102	0.480
	32^2	**90.76**	0.737	0.661	**0.204**	**0.541**

但两项保真度指标略有下降。不过，有趣的是，如果使用更多对称操作，例如 $hvnp$，SPE 在所有指标上都优于基线。但在这种情况下，由于 FID 是主要优化指标，因此我们进一步进行了实验，将 SPE损失配置设为 hv翻转。

4.2. 使用基于注意力的生成器加速 SPE

注意力机制在许多领域都达到了最先进水平，例如基于注意力的生成对抗网络[21]，，其中将注意力机制自注意力网络（SAN） [22] 加入了StyleGAN2 [23]。然而，原始的基于注意力的生成器计算开销很大。因此，我们还针对有限的设置，尝试了计算上更合理的替代方案——高效注意力（EAttn） [24]，。

在之前的实验中，基于注意力的生成器由于参数量增加以及训练数据规模较小，难以训练和收敛。然而，采用SPELoss后，图4显示每个模型的性能都得到了显著提升。现在，来自 $8^2, 16^2, 32^2$的所有EAttn生成器的FID都低于100，其中最小的EAttn 32^2 从127.06提升至创纪录的90.76。

然而，这一创纪录的低 FID 是以牺牲保真度和多样性之间的平衡为代价的。每一种生成器配置

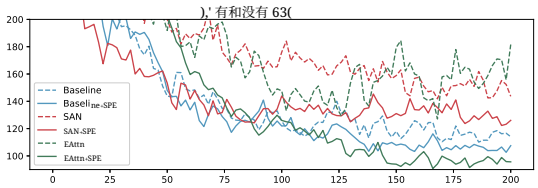

图4：采用SPE的基于注意力的生成器的FID结果。每个模型均训练了200k次迭代。为清晰起见，我们省略了不同分辨率下其他表现较差的注意力模型，仅列出 8^2 上表现最佳的SAN和 32^2上表现最佳的高效注意力（更多详情请参见表2）。



图5：采用高效注意力时，基线与所提方法SPE之间的生成器结果差异。生成器结果取自每个FID最佳的快照，其中(a)基线为186k，(b)基线+SPE为200k，(c)基线+SPE+EAttn为166k。

采用 SPE 后，保真度指标都会略有降低。不过，这一点是意料之中的，因为 FastGAN 的设计初衷就是实现高保真度。作为回报，随着多样性提高，我们可以在图 5 中看到，与基线和带有 SPE+EAttn 的基线相比，SPE+EAttn 的细节明显更加精细，对称性也更好。我们还发现，与之前的方法相比，EAttn 能够覆盖更多不同的风格，并呈现更丰富的细节。

5. 结论

随着生成对抗网络取得巨大进展，尤其是在有限计算资源和数据条件下的进步，其主流应用也将随之发展。因此，本文探讨了蜡染生成任务，这一主题在东南亚广受欢迎。在研究过程中，我们发现，对称图案任务的成果可推广至葡萄牙瓷砖或摩洛哥瓷砖等类似任务。通过构建更优质的高质量数据集、探索在有限计算资源下将注意力机制应用于生成对抗网络的低成本方法，并采用SPE损失实现训练加速，我们在图像质量（FID）和细节（召回率）方面，显著提升了对称图案生成效果，超越了当时最先进的Fast GAN。我们认为，这项研究将成为未来人工智能艺术及面向艺术家的对称图案生成应用的起点和基础。

6. 参考文献

[1] Agus Eko Minarno、Moch. Chamdani Mustaqim、Yufis Azhar、Wahyu Andhyka Kusuma和Yuda Munarko，《深度卷积生成对抗网络在蜡染图案生成器中的应用》，载于*2021*年第*9*届国际信息与通信技术会议（*ICoICT*），2021年，第54–59页。

[2] Aditya Firman Ihsan，《使用神经网络进行蜡染风格迁移的研究》，载于*2021*年第九届信息与通信技术国际会议（*ICoICT*），2021年，第313–319页。[3] Wei-Ta Chu和Lin-Yu Ko，《BatikGAN：用于蜡染创作的生成对抗网络》，载于*2020*年多媒体艺术作品分析与吸引力计算联合研讨会论文集，美国纽约州纽约市，2020年，MMArt-ACM '20，第13–18页，计算机协会。[4] Miqdad Abdurrahman、Nabila Husna Shabrina和Dareen K. Halim，《用于蜡染图案合成的生成对抗网络实现》，载于*2019*年第五届新媒体研究国际会议（*CONMEDIA*），2019年，第63–67页。[5] Syenne Ecstexela和Sung-Ho Bae，《一种利用深度神经网络进行创意摄影的新型蜡染图案生成方法》，载于韩国信息科学学会论文集，2019年6月。[6] Aifa Nur Amalia、Arief Fatchul Huda、Diena Rauda Ramdania和Mohamad Irfan，《构建用于文本到图像合成的生成对抗网络蜡染数据集》，载于*2019*年*IEEE*第五届无线与远程信息处理国际会议（*ICWT*），2019年，第1–7页。[7] Bingchen Liu、Yizhe Zhu、Kunpeng Song和Ahmed Elgammal，《面向高保真少样本图像合成的更快速、稳定的生成对抗网络训练》，2021年。[8] Tero Karras、Samuli Laine、Miika Aittala、Janne Hellsten、Jaakko Lehtinen和Timo Aila，《分析并提升StyleGAN的图像质量》，载于*CVPR*论文集，2020年。[9]Tero Karras、Miika Aittala、Janne Hellsten、Samuli Laine、Jaakko Lehtinen和Timo Aila，《使用有限数据训练生成对抗网络》，载于*NeurIPS*论文集，2020年。

[10] Shengyu Zhao、Zhijian Liu、Ji Lin、Jun-Yan Zhu 和 Song Han，《用于数据高效生成对抗网络训练的可微增强》，2020。

[11] Yohanes Gultom、Aniati Murni Arymurthy 和 Rian Jo- sua Masikome，《使用深度卷积网络迁移学习进行蜡染分类》，《计算机与信息科学期刊》，第 11 卷，第 2 期，2018 年。[12] Chrystian Chrystian，《ITB-mBatik 数据集》，2023 年。[13] Jean Gallier，卡当-迪厄多内 ´定理，第 197–247 页，Springer New York，纽约州纽约市，2001 年。[14]Richard Zhang、Phillip Isola、Alexei A Efros、Eli Shechtman 和 Oliver Wang，《深度特征作为感知度量的惊人有效性》，发表于*CVPR*，2018 年。[15] Martin Heusel、Hubert Ramsauer、Thomas Unterthiner、Bernhard Nessler 和 Sepp Hochreiter，《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》，发表于神经信息处理系统进展，I. Guyon、U. V. Luxburg、S. Bengio、H. Wallach、R. Fergus、S. Vishwanathan 和 R. Garnett 编，2017 年，第 30 卷，Curran Associates, Inc。[16]Tuomas Kynk¨a¨anniemi、Tero Karras、Samuli Laine、Jaakko Lehtinen 和 Timo Aila，《用于评估生成模型的改进精确率与召回率指标》，*CoRR*，第 abs/1904.06991 卷，2019 年。[17] Muhammad Ferjad Naeem、Seong Joon Oh、Youngjung Uh、Yunjey Choi 和 Jaejun Yoo，《生成模型的可靠保真度与多样性度量》，2020 年。[18] Mario Lucic、Karol Kurach、Marcin Michalski、Sylvain Gelly 和 Olivier Bousquet，《生成对抗网络是否生而平等？一项大规模研究》，2018 年。[19] Jae Hyun Lim 和 Jong Chul Ye，《几何生成对抗网络》，2017 年。[20] Diederik P. Kingma 和 Jimmy Ba，《Adam：一种随机优化方法》，2017 年。[21] Ning Yu、Guilin Liu、Aysegul Dundar、Andrew Tao、Bryan Catanzaro、Larry Davis 和 Mario Fritz，《生成对抗网络的双重对比损失与注意力》，发表于 *IEEE* 国际计算机视觉会议*(ICCV)*，2021 年。[22] Hengshuang Zhao、Jiaya Jia 和 Vladlen Koltun，《探索用于图像识别的自注意力》，发表于 *IEEE/CVF*计算机视觉与模式识别会议论文集*(CVPR)*，2020年6月。[23] Tero Karras、Samuli Laine、Miika Aittala、Janne Hellsten、Jaakko Lehtinen 和 Timo Aila，《分析并提升 StyleGAN 的图像质量》，*CoRR*，第 abs/1912.0 卷，2019 年。[24] Zhuoran Shen、Mingyuan Zhang、Haiyu Zhao、Shuai Yi 和 Hongsheng Li，《高效注意力：具有线性复杂度的注意力》，发表于冬季计算机视觉应用会议，2021 年。