arXiv:2305.04923v2 [cs.CV] 9 Jun 2024

学习评估人工智能生成图像的艺术性

Junyu Chen、Jie An、Hanjia Lyu、Christopher Kanan、*IEEE* 高级会员和 Jiebo Luo，*IEEE* 会士

摘要——**评估人工智能生成图像的艺术性，在图像生成领域仍是一项挑战。现有的大多数度量指标都无法用于实例级、无参考的艺术性评估。本文提出 ArtScore，这是一种旨在评估图像在多大程度上类似于艺术家的真实艺术作品（或相反，类似于照片）的度量指标，从而为艺术性评估提供了一种新方法。我们首先融合用于照片和艺术作品生成的预训练模型，得到一系列混合模型。随后，我们利用这些混合模型生成一组带有伪标注、艺术性程度各异的图像。每张写实图像都有一张对应的艺术风格图像，以及一系列从写实到艺术风格的插值图像。接着，我们使用该数据集训练神经网络，使其能够估计任意图像的量化艺术性水平。大量实验表明，与 Gram 损失和 ArtFID 等现有评估度量指标相比，ArtScore 预测的艺术性水平与人类艺术评价更加吻合。**

索引词—**艺术图像评估、神经风格迁移（NST）、生成对抗网络（GAN）、深度神经网络。**

I. 引言

IMAGE 生成在计算机视觉领域日益受到关注。神经风格迁移（NST） [1]–[4], 生成对抗网络（GAN） [5]–[8], 和扩散模型 [9]–[11] 是三种常见的图像生成框架。在提升图像生成质量方面已取得显著进展。然而，如何自动评估人工智能生成图像的艺术性，尚未得到充分探索。

度量指标，例如结构相似性指数（SSIM）[12], 、Inception Score（初始分数）（IS） [13], 、Fr´echet Inception Distance（FID） [14], 、内容损失（Content Loss） [1], 、ArtFID [15], 和 Gram 损失（Gram Loss） [1], 被图像生成算法广泛用于衡量生成结果的质量。然而，现有的度量指标**均无法**用于执行**实例级且无需参考图像的**艺术性评估。FID 和 IS 通常用于评估人工智能生成图像的整体质量，但它们并非专门为艺术性评估而设计，因此在艺术性测量方面不够准确。SSIM 和内容损失通常用于衡量照片到艺术作品的转换过程中内容语义的保留程度，但当内容的.....

稿件于2024年2月27日收到；2024年5月20日修订；2024年5月23日接受；当前版本日期为2024年6月1日。负责协调本稿件评审并批准其发表的副编辑为 Jianlong Fu。（通讯作者：*Jiebo Luo*。）作者来自罗切斯特大学计算机科学系，美国纽约州罗切斯特，邮编14627（电子邮箱：jchen175@ur.rochester.edu、jan6@cs.rochester.edu、hlyu5@ur.rochester.edu、ckanan@cs.rochester.edu、jluo@cs.rochester.edu）。项目页面：https://github.com/jchen175/ArtScore 数字对象标识符：10.1109/TMM.2024.3410672

两张图像有所不同。Gram 损失用于计算生成图像与真实图像之间的艺术距离，真实图像是在进行风格迁移时用作参考的图像；这限制了图像风格迁移的应用，因为进行艺术性评估需要目标真实艺术作品。与 FID [14], 类似，ArtFID 可以评估生成艺术作品与真实艺术作品分布之间的距离；当生成图像数量或用于比较的真实艺术作品数据集有限时，评估结果可能不准确，从而难以进行实例级艺术性评估。

为克服缺乏自动化艺术性评估指标的问题，现有研究广泛采用基于精心设计的用户研究进行主观评估 [4], [16]–[20]。然而，人工评估可能**耗费** 大量时间和精力，**难以** 复现；如果收集到的用户反馈数量不足，还可能**产生偏差** 。因此，我们需要一种客观指标，能够自动衡量算法生成图像的艺术性，并与主观人工评估高度一致 。恰当的评估指标对于图像生成算法至关重要。首先，它能够对不同算法进行公平、定性的比较，从而促进竞争并推动更优算法的发展。其次，这类指标可直接用于指导图像生成过程，以提升生成图像的质量和保真度。

为此，我们提出了一种新的指标来评估人工智能生成图像的艺术性水平，并将其命名为 ArtScore。ArtScore 以神经网络的形式实现，既支持实例级 评估，也支持无参考艺术性评估。为开发 ArtScore 模型，我们首先构建一个插值艺术性数据集，然后在新构建的数据集上训练神经网络，以回归伪标注的艺术性水平。整个过程可分为三个步骤：StyleGAN 适配、插值图像生成和 ArtScore 训练（图 1）。由人工标注图像的艺术性评分**成本高昂** ，且需要**丰富的专业知识**。前两个步骤旨在生成伪标注数据集，以避免成本高昂且耗时的人工标注 [21], ；我们收集的数据集中的图像呈现出不同的艺术性水平，每张照片级真实感图像都有一张对应的艺术图像，以及一系列从照片级真实感到艺术风格的插值图像。

从一个预训练的 StyleGAN2 [7] 模型（具有照片级真实感风格）出发，我们首先通过迁移学习训练一个具有美术绘画风格的模型，并采用 FreezeG [22] 和少样本适应 [23] 这两种迁移学习方案。FreezeG 会冻结生成器的前几层，并调整

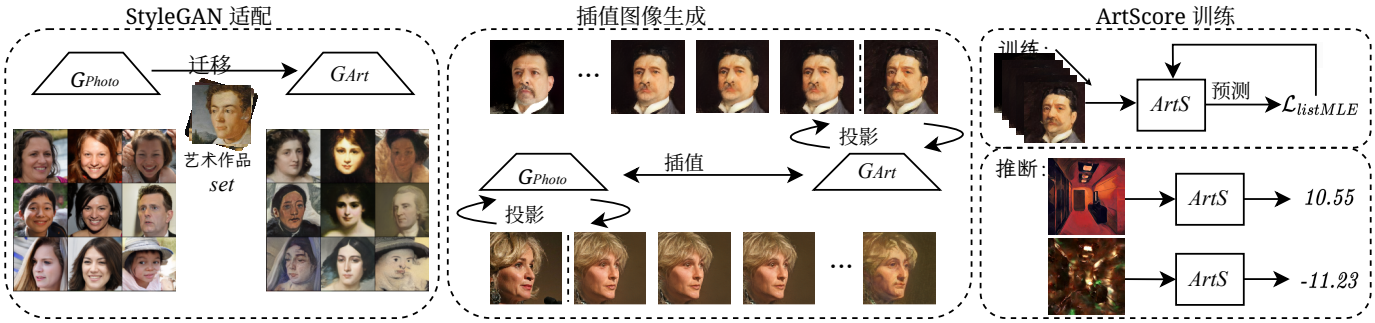


图 1. 所提出的框架由三个步骤组成：1）StyleGAN迁移；2）生成插值图像；3）训练 ArtScore。前两个步骤生成一个大型数据集，其中包含艺术性水平各异的图像。第三个步骤则利用构建的数据集学习区分不同的艺术性水平。

后续卷积模块，这种做法经验证能在相应领域取得良好效果。少样本方法可以生成具有不同艺术风格的多个 StyleGAN 模型，从而带来更多样的艺术图像。对于采用少样本设置的迁移学习，用于引导模型迁移的示例是来自不同艺术风格的代表性绘画作品，这些作品通过艺术风格聚类获得。

接下来，我们生成一个艺术性水平各异的伪标注数据集。基于具有照片写实风格的预训练 StyleGAN 模型，以及经过调整、具有艺术风格的模型，我们对 [24] 照片写实模型和艺术模型进行线性插值，得到能够生成艺术性水平各异图像的 StyleGAN 模型。通过调整基础模型与调整后模型之间的插值权重，我们可以控制混合模型的艺术性水平，从而使用相同的潜在输入生成结构相同但艺术性水平不同的图像。我们将来自 FFHQ [25], LSUN 马 [26], 和 LHQ [27] 数据集的照片写实图像投影到预训练 StyleGAN 模型的潜在空间 [7]中，并使用从照片风格向艺术风格插值的 StyleGAN 模型生成艺术性水平各异的图像。

我们训练了一个神经网络（ArtScore），用于预测图像与真实艺术作品的接近程度，而这种判断可能因主观解读和标准而异。然而，在模型训练中，很难定义固定的绝对艺术性真实标签。因此，我们采用**学习排序**目标 [28], 而不是回归插值图像的绝对艺术性水平；模型的训练目标是对给定图像序列中图像的**相对艺术性**进行排序。学习排序损失项无需固定的绝对艺术性真实标签，同时仍能学到有效的艺术性评分。为进一步提高模型在推理阶段的鲁棒性，我们提出了一种新颖的数据增强策略。训练过程中，序列中的一张图像会被随机替换为另一序列中相对艺术性排名相同、但内容和风格不同的图像。这种做法有助于模型比较关联度较低的图像的艺术性。消融研究表明，采用学习排序损失项比回归固定的艺术性水平效果更好，而跨序列随机交换增强则进一步提升了模型性能。

我们在实例级和算法级评估了所提出度量指标的有效性，结果表明，ArtScore 测得的艺术性水平与**人工评估**的**一致性更高**。我们的实验还表明，将 ArtScore 与其他图像质量评估度量指标结合使用，可以获得更准确的图像艺术性排名，而**不使用 ArtScore 时，其他度量指标无法实现这一点**。此外，我们还进行了消融研究，以评估框架设计的有效性和必要性。

总而言之，我们的贡献主要有三点：

- 我们构建了一个大型数据集，可促进所提出的艺术性度量指标的有效学习，且无需依赖成本高昂、需要专业知识的人工标注。
- 我们提出了 ArtScore，可定量评估生成图像与美术绘画的相似程度。它是一种采用有效的学习排序目标进行训练的神经网络。据我们所知，ArtScore 是首个能够在无需参考图像的情况下进行实例级艺术性评估的方法。
- 我们通过实验证明，相较于现有度量指标，ArtScore 与人工评估的一致性更高；与其他现有度量指标结合使用时，它还能更好地定量评估神经风格迁移算法。

II. 相关工作

A. 图像生成框架

近年来，图像生成受到了广泛关注。神经风格迁移（NST）是一种重要的方法，可将风格图像中的艺术风格迁移到内容图像中。[1] 提出了一种迭代算法，用于优化 Gram 损失和内容损失。 [29]和 [30]提出了用于特定风格近似的前馈网络。后续技术实现了单个模型支持多种风格 [31], [32],以及通用风格迁移方法 [2], [3], [33], [34]。生成对抗网络（GAN）是另一种流行的方法，它涉及两个相互竞争的神经网络——生成器和判别器。各种先进且特定领域的 GAN 变体已被用于图像生成，例如 CAN [6],、StyleGAN [35], 和 CycleGAN [8], ，这些方法在生成逼真的伪造艺术作品图像方面取得了令人瞩目的成果。扩散模型 [9] 是一种近期提出且具有显著潜力的框架。它通过创建一系列噪声程度逐渐增加的原始图像版本，并通过迭代学习对其进行去噪。DALL-E 2 [10], 、稳定扩散 [11], 、MultiDiffusion [36],和 RAPHAEL [37]都是基于扩散模型的文本生成图像方法。为了训练 ArtScore 模型，我们构建了一个数据集。

使用 StyleGAN2 模型。在我们的实验中，为了评估训练好的模型，我们使用各种神经风格迁移方法和 StyleGAN 生成图像以供评估。

B. 图像生成评估

已有多种度量指标被提出，用于从特定维度评估生成的图像。Inception Score（IS） [13] 和 Fr´echet Inception Distance（FID） [14] 基于图像集的特征分布。IS 将条件标签分布与生成图像的边缘分布进行比较，以反映图像质量和多样性。FID 衡量生成图像集与真实图像集的深层特征分布之间的距离。结构相似性指数（SSIM） [12] 和学习感知图像块相似度（LPIPS） [38]基于对两幅图像的比较。SSIM 从像素层面衡量生成图像与真实图像之间的相似度，同时考虑亮度、对比度和结构。LPIPS 计算图像块级特征表示之间的平均 l2 距离。内容损失 [1] 和 Gram 损失 [1] 是神经风格迁移研究 [2], [17], [39], [40]中常用的度量指标。内容损失基于特征图的均方误差（MSE），衡量迁移图像与内容参考图像之间的内容保留程度。Gram 损失基于特征图 Gram 矩阵的均方误差，衡量风格化图像与风格参考图像之间的风格匹配程度。欺骗率 [41] 和 ArtFID [15] 利用在艺术绘画数据集上训练的预训练风格分类器，以改进风格表征。欺骗率衡量艺术家分类网络将多少比例的风格化图像判定为其艺术作品被用于风格化的艺术家所创作。ArtFID 同时纳入内容和风格：通过计算内容参考图像与风格化图像之间的 LPIPS 距离来衡量内容保留程度，并通过计算风格参考图像集与风格化图像集之间的 FID 来衡量风格匹配程度（ArtFID 风格）。最终的评估度量指标计算为$ArtFID = (1 + LPIPS) \cdot (1 + ArtFID\ \ Style)$。为评估我们提出的 ArtScore 的有效性，我们将其与艺术性评估中常用的度量指标进行比较。

C. StyleGAN 反演与迁移

潜在空间投影是 StyleGAN [35] 中一种广泛使用的技术，用于将给定图像归因于生成器。它涉及使用生成器为给定图像寻找匹配的潜在编码；因此。使得将潜在编码输入回生成器后，生成器能够生成与给定目标图像高度匹配的图像。这项技术有助于控制图像编辑中生成图像的属性 [42]。 [7]探索了一种基于优化的潜在空间投影方法：固定生成器，同时优化潜在编码，以尽量降低生成图像与给定目标图像之间的 LPIPS。

迁移学习已被用于将预训练的 StyleGAN 模型适配到新的风格 [23], [43], [44], ，从而仅凭有限样本训练出高质量的生成器。迁移后的生成器与源生成器密切相关：通过对二者的权重进行线性插值，网络插值可以近似实现两种所学习风格之间的插值 [24]。 [45] 表明，通过融合不同层，可以控制融合后生成器的结构性和风格化视觉效果，使其呈现出类似源模型或迁移后模型的效果。

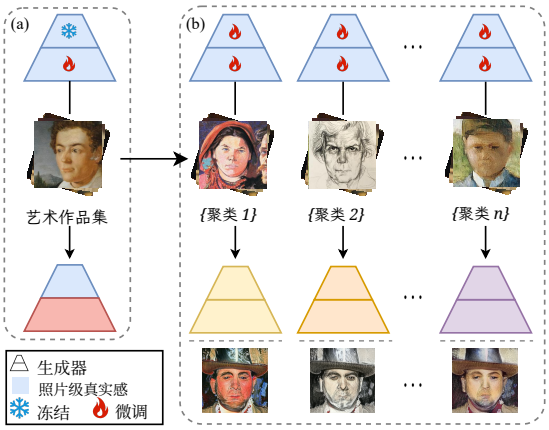


图 2. 我们采用 (a) FreezeG 和 (b) 一种少样本适应方法，将照片级真实感 StyleGAN 适配到艺术风格。在 FreezeG 设置 (a) 中，我们使用完整的艺术绘画数据集对写实模型进行微调，同时冻结较早的卷积层。在少样本适应设置 (b) 中，我们按风格对艺术绘画进行聚类，并抽取具有代表性的绘画作品，仅使用少数代表性样本对整个写实模型进行微调。

我们使用 StyleGAN迁移和插值来生成具有不同艺术性水平的生成器，并利用 StyleGAN 反演获取真实图像的潜在编码，将其作为这些生成器的输入，以合成训练数据集。

D. 学习排序

学习排序（LTR） [46] 旨在基于评分模型对对象列表进行排序。它通常用于各种应用，例如推荐系统和信息检索。在训练过程中，评分模型通过最小化排序损失函数，学习为输入对象列表预测分数。LTR 方法在如何将对象纳入损失函数方面有所不同。与点式方法和成对方法相比，后两者分别将排名迁移为对单个对象的回归或对对象对的分类，而 ListMLE 等列表式方法将整个列表作为输入，更好地考虑了相对位置信息 [28], [47]，并取得了更好的实验结果。所提出的 ArtScore 基于列表式 LTR 目标的优化进行训练。

III. 方法

图 1 展示了计算 ArtScore 的完整流程。我们首先采用迁移学习流程，将预训练的照片级真实感 StyleGAN 模型适配到艺术风格。接着，我们融合用于生成照片和艺术作品的预训练 StyleGAN 模型，并使用融合后的模型生成一个艺术性水平各异的数据集，其中模型插值比例用作艺术性水平的伪标注。最后，我们基于学习排序目标训练一个神经网络，以衡量图像的艺术性。本节将介绍每个步骤的详细信息。

*A. StyleGAN*迁移

我们通过迁移学习，将预训练的 StyleGAN2 [7]模型从照片写实风格迁移到艺术风格。给定 StyleGAN 潜在空间中的相同潜在代码，原始预训练模型和迁移后的模型

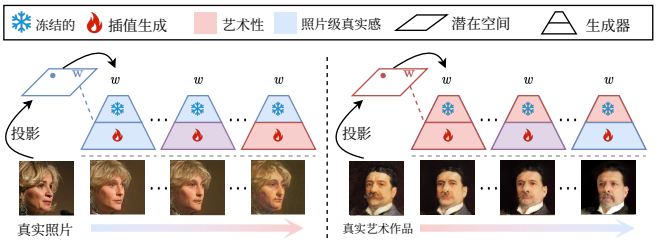


图 3. StyleGAN 模型插值与插值图像生成。我们对写实风格和艺术风格 StyleGAN 模型的低层卷积层进行线性插值。将真实照片和艺术作品投影到各自模型的潜在空间中，以提取潜在编码。这些编码用作插值模型的输入，从而生成一系列艺术性程度各异的图像。

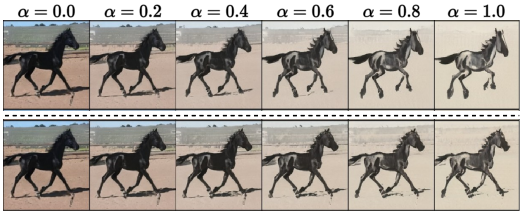


图 4. 融合写实模型与艺术模型的所有模块（上）或最后几个模块（下）的效果。请注意，仅融合低级层时，马的结构保留得更好。

分别生成照片写实图像和艺术图像，且两幅图像的内容相同。我们在基于 FFHQ [25]、LSUN 马 [26]，和 LHQ [27] 数据集预训练的 StyleGAN 模型上进行迁移学习，因为这些数据集对应的类别在艺术绘画中很受欢迎。我们从 WikiART [48]数据集中收集与这些领域对应的绘画，并用于迁移模型。我们探索了两种迁移学习方法：FreezeG [22] 和少样本适应 [23]（见图 2）。

FreezeG 冻结生成器的早期层，并调整生成器最后几个模块，从而在将 StyleGAN 迁移为艺术风格的同时，完整保留潜在编码结构。少样本方法只需少量示例，即可将照片级真实感生成器迁移为艺术风格；我们可以使用具有特定艺术风格的艺术作品作为示例，来控制迁移后的模型。通过切换示例图像，我们可以将同一个写实模型调整为不同的艺术模型，生成风格各异的图像，从而创作出更多样化的艺术图像。为此，我们通过聚类从不同艺术风格中选取有代表性的绘画作品。具体而言，我们采用 Gram 矩阵 [1]（堆叠成一个一维向量）作为风格表征，并使用 K-Means 算法进行聚类。选取最接近聚类中心的绘画作品作为该聚类的代表，以调整写实模型。我们对所有聚类进行少样本 StyleGAN 模型迁移，最终得到一个预训练的照片级真实感 StyleGAN 模型，以及与之对应的、具有不同艺术风格的艺术模型。

B. 伪标注数据集生成

本小节介绍了我们如何使用写实和艺术风格的 StyleGAN 模型合成各种图像，包括写实图像、艺术图像，以及如图 3 所示的两种风格之间具有不同艺术性水平的中间图像。我们首先对写实 StyleGAN 和艺术 StyleGAN 的网络参数进行插值

表 I 生成的伪标注艺术性数据集摘要。我们列出了每个领域中包含真实绘画作品或真实照片的序列数量。每个序列包含 12 张内容相同、风格不同的图像，其风格从照片级真实感逐渐过渡到艺术风格。

领域	绘画序列	照片序列	序列总数
Face	17,436	17,436	34,872
马	8,808	8,808	17,616
风景	45,627	45,627	91,254
All	71,871	71,871	143,742

模型 [24]，以得到能够生成真实风格与艺术风格之间插值图像的混合模型。插值模型的参数是写实模型和艺术模型的加权平均：

$$\theta_i^I = (1-\alpha)\cdot\theta_i^P + \alpha\cdot\theta_i^A, \tag{1}$$

其中，θ_i^I、θ_i^P 和 θ_i^A 分别是插值模型、写实模型和艺术模型的 i^{th} 卷积层参数。α 是用于控制所需艺术性水平的插值权重，同时也用作伪标注。

接下来，我们分别将一组真实绘画作品和照片投影到艺术领域和写实领域中对应的 StyleGAN 模型的潜在空间。我们使用 [49] 介绍的 StyleGAN 反演算法。对于通过少样本迁移学习方法生成的艺术风格 StyleGAN 模型，我们将用于 StyleGAN迁移示例的绘画作品投影到潜在空间。

最后，基于插值后的 StyleGAN 模型和检索到的潜在编码，我们使用艺术性水平各异的 StyleGAN 模型生成带有伪艺术性标签的图像。该过程从一幅真实绘画作品或一张照片开始，分别将其作为艺术性最高或最低的真实标签。接下来，我们以原始输入（即真实绘画作品或照片）为起点，生成一系列图像，同时平滑地过渡到另一种风格。给定相同的潜在编码，插值后的 StyleGAN 模型将生成语义内容相同、但艺术性水平各异的图像。如图 4 所示。这是通过仅融合最后 6 个卷积层，同时保持前面各层的参数不变来实现的。

C. 模型训练

本小节介绍如何在伪标注数据集上训练神经网络，以构建 ArtScore 模型。将伪艺术性标签作为回归目标，并直接以相应图像作为输入，无法取得令人满意的结果。原因在于，不同的真实艺术作品可能呈现出不同的艺术性水平。然而，在我们收集的伪标注数据集中，所有真实艺术作品的艺术性水平都相同，这可能会使网络在训练时产生混淆，导致准确率下降。为此，我们采用学习排序方案来训练神经网络，使其预测图像的相对排序，而非固定的绝对艺术性水平。给定插值图像序列 $\mathbf{x} = \{x_1, \cdots, x_n\}$ 及相应的艺术性水平 $\mathbf{y} = \{y_1, \cdots, y_n\}$，我们的目标是学习一个模型 f，使其输出的艺术性评分尽可能接近

真实排序。为此，我们使用 ListMLE 损失，其定义为

$$\mathcal{L}(f;\mathbf{x},\mathbf{y}) = -\log P(\mathbf{y}|\mathbf{x};f), \tag{2}$$

其中

$$P(\mathbf{y}|\mathbf{x};f) = \prod_{i=1}^{n} \frac{\exp(f(x_{\mathbf{y}^{-1}(i)}))}{\sum_{k=i}^{n}\exp(f(x_{\mathbf{y}^{-1}(k)}))} \tag{3}$$

这里，$\mathbf{y}^{-1}(i)$表示$\mathbf{y}$中位于第 i个位置的图像索引，因此，$f(x_{\mathbf{y}^{-1}(i)})$表示给定序列中排名第 i的图像对应的模型输出分数。为帮助模型比较内容和风格各异的图像的艺术性，我们在训练过程中引入了随机跨序列交换增强：序列中的一幅图像（即 由 StyleGAN 模型插值生成的图像序列）会被随机替换为另一序列中相对艺术性排名相同、但内容和风格不同的图像。实验发现，学习排序方法比直观的回归框架效果更好，而随机交换增强提升了模型在推理时的鲁棒性和性能。

IV. 实验

有效的评估指标应与人类判断 [15], [38], [50]高度一致。为验证我们框架的有效性，我们在一个经人工标注艺术性的数据集上，将 ArtScore 得出的艺术性水平与其他度量指标进行了比较。结果表明，就艺术性而言，所提出的 ArtScore 与人类测量结果的一致性更高。将 ArtScore 与其他图像质量评估指标结合使用，能够对 AI 生成图像的质量进行更准确、全面的评估。

A. 实现细节

数据集。 WikiART [48] 是一个大型数据集，包含 8 万多幅独特画作，并提供有关艺术家、风格和流派的详细元数据。我们从 WikiART 数据集中收集了一组艺术画作，包括艺术人脸、艺术马画和风景画，用于将预训练的 Style-GAN2 [7] 模型从写实风格迁移到艺术风格。对于艺术人脸域，我们使用了一个现成的预处理数据集 [51]。为了收集艺术马画，我们从 WikiART 中检索了 2,000 幅画作，这些画作的 CLIP [52] 嵌入与提示词“这是一幅马画”的 CLIP 嵌入具有最高的余弦相似度；此外，还收集了配套元数据（即流派、名称、描述和标签）中包含关键词“horse”的所有画作。我们手动剔除了误检的马画。对于风景画域，我们使用 WikiART 中所有属于“风景”流派的画作。总计收集了 5,812 幅艺术人脸画作、2, 936 幅马画和 15,209 幅风景画。

StyleGAN迁移学习。 我们采用两种设置，将预训练的照片级真实感 StyleGAN2 模型迁移到艺术模型。在第一种设置中，我们使用 FreezeG 技术，固定生成器的早期层，并对生成器的最后几个模块进行微调。我们通过超参数调优来确定要融合的层，实验表明，仅调优最后 5 或 6 个模块，或对整个模型进行调优，都能产生视觉上令人愉悦的结果。对于这种方法，我们使用 0.0002 的学习率、16 的批量大小和总计 10,000 次训练迭代，并使用所有收集到的、属于相应领域的画作进行适应。在第二种设置中，我们使用少样本适应方法。首先，我们根据 Gram 矩阵所表示的艺术风格，使用 K-Means 对画作进行聚类。我们将聚类数量设为 20，这一数量大致对应于数据集中独特艺术风格的数量。对于每个聚类，我们选取最接近聚类中心的 10 幅画作作为代表，并使用它们来适应照片级真实感模型。我们在全部 20 个聚类上执行少样本 StyleGAN 模型迁移，使用 0.0002 的学习率、4 的批量大小和总计 5,000 次训练迭代。其他所有超参数均与官方实现保持一致。

插值图像生成。 我们使用插值参数 α 控制生成图像的艺术性程度，取值范围为 0.0 至 1.0，步长为 0.1。每个序列包含 12 张图像，其中包括一张真实照片或一幅绘画，以及 11 张艺术性程度各异的插值图像。表 I 展示了所构建数据集的统计信息。

ArtScore 训练。 我们采用在 ImageNet 上预训练的 ResNet50 作为所提出 ArtScore 的骨干网络，并将全连接层替换为一个三层 MLP，其隐藏层大小为 768，输出层大小为 1。我们使用 AdamW 优化器 [53] ，以 $1e{-}4$ 的学习率优化 ListMLE 损失 [28]。批量大小设为 16 个序列（即 192 张图像）。为防止过拟合，dropout 率设为 0.5。除随机交换增强外，我们还对训练图像应用常用的数据增强方法，包括随机翻转和旋转。数据集按 0.8、0.1 和 0.1 的比例随机划分为训练集、验证集和测试集。我们训练模型 10 个 epoch，并在模型于验证集上达到最低损失时保存其参数，作为最终评估器；该评估器在测试集上的 NDCG 分数为 0.965。

B. 与人工评估的定量一致性

[15] 通过一项大规模用户研究，系统比较了 13 种不同的神经风格迁移算法在照片转艺术风格迁移中的效果。我们采用其论文中给出的用户研究评分，作为衡量这些算法生成艺术作品质量的基准。随后，我们通过计算斯皮尔曼排名相关系数 [54]，评估我们提出的 ArtScore 及其他现有度量指标与人工评估的一致程度。按照 [15], 中的实验设置，我们使用相同的内容参考与风格参考配对，为文献中的 12 种风格迁移方法 [1], [2], [19], [33], [55]–[62] 生成 50,000 512×512 张风格迁移图像。由于某种方法 [63]的实现细节不足，我们将其排除。内容图像采样自 Places365 数据集 [64]和 COCO 数据集 [65]。风格图像采样自 WikiART [48]数据集。我们计算了三组不同的度量指标：(1) 内容保留度量指标，包括 LPIPS、内容损失、SSIM 和 L2 距离，使用风格化图像与内容参考进行计算；

表 II 根据式（4）给出的排序聚合方法，用户研究 [15] 所得排名与 ArtScore 及其他度量指标所得排名之间的相关性。对于 ArtFID 度量指标，我们报告 [15]中的结果（论文），以及使用我们实验中生成的风格化图像重新实现的结果（REP）。最高相关性以粗体标示，次高相关性加下划线。

方面	度量指标	斯皮尔曼 $\rho\uparrow$	p 值 $\downarrow$
艺术性	ArtScore	0.6364	0.0261
风格	Gram 损失	-0.3286	0.2969
	ArtFID_风格	0.6504	0.0220
	+ArtScore	0.7549 (+0.1045)	0.0045
内容	LPIPS	0.4126	0.1826
	+ArtScore	0.7138 (+0.3012)	0.0091
	内容损失	0.4755	0.1182
	+ArtScore	0.5835 (+0.1080)	0.0464
	SSIM	0.5315	0.0754
	+ArtScore	0.5965 (+0.0650)	0.0406
	L2	0.2837	0.3715
	+ArtScore	0.5298 (+0.2461)	0.0764
风格+内容	ArtFID∞ （论文）	0.9301*	0.0000*
	ArtFID∞	0.7063	0.0102
	+ArtScore	0.7790 (+0.0727)	0.0028
	ArtFID_风格+LPIPS	0.6714	0.0168
	+ArtScore	0.7552 (+0.0489)	0.0045
	ArtFID_风格+CL	0.7874	0.0024
	+ArtScore	**0.8196** (+0.0322)	**0.0011**
	ArtFID_风格+SSIM	0.8120	0.0013
	+ArtScore	0.7719 (-0.0401)	0.0033
	ArtFID_风格+L2	0.6843	0.0141
	+ArtScore	0.7356 (+0.0513)	0.0064

（2）风格匹配度量指标，包括 ArtFID 风格（即 ArtFID 度量指标中使用的风格度量）和 Gram 损失，这些指标基于风格化图像和风格参考进行计算；（3）基于所提出的 ArtScore 的艺术性度量指标，该指标不需要任何参考。为了获得某个算法的综合 ArtScore 度量指标，我们首先使用训练好的 ArtScore 计算每幅风格化图像的艺术性评分，然后将这些评分输入 sigmoid 函数并计算其平均值。因此，某个算法最终的 ArtScore 被归一化到 0 到 1 的范围内，分数越高表示该算法越能够生成类似美术绘画的图像。

为了衡量度量指标及其组合与人类评价分数之间的对齐程度，我们计算斯皮尔曼 ρ相关系数，结果如表 II 和图 5 所示。对于单个度量指标，我们可以直接计算相关性；而对于多个度量指标，我们探索了三种组合方法。

排名。 我们将 M_i 记为第 i 个度量指标，将 Alg_j 记为第 j 个算法。我们根据每个度量指标 M_i计算各算法 Alg_j 的排名，记为 R_{ij}。算法 Alg_j 的总体排名R_j 定义如下：

$$R_j = \sum_{i=1}^{n} R_{ij} \tag{4}$$

其中，n 是所使用度量指标的总数。随后使用 R_j 作为计算相关性的聚合分数。

相加。 这里，我们首先将度量指标值归一化到 0 到 1 之间，并转换为数值越小越好的形式，记为M_{norm}。我们将 M_{normij} 记为第 j 个算法在第 i 个归一化度量指标上的取值。然后我们使用

$$A_j = \sum_{i=1}^{n} M_{normij} \tag{5}$$

将标准化指标合并，得到一个综合得分。

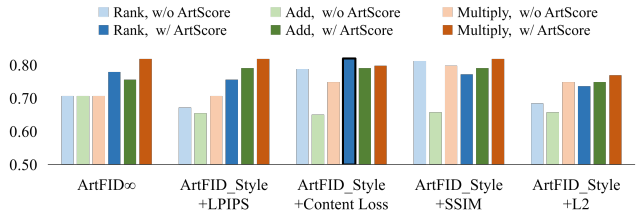


图 5. 来自 [15] 的用户研究排名与内容保留度量指标和风格匹配度量指标所产生排名之间的相关性，分别展示了加入和未加入我们提出的 ArtScore 度量指标的情况。我们测试了三种聚合方法，分别由公式 4、公式 5 和公式 6 表示。浅色表示未加入 ArtScore 的结果，深色表示加入 ArtScore 的结果。总体而言，在所有聚合方法中，加入 ArtScore 都能持续提高图像艺术性排名的准确率。

相乘。 这里我们同样使用标准化后的指标值，并采用

$$Mul_j = \prod_{i=1}^{n} (1 + M_{normij}) \tag{6}$$

用于合并归一化后的度量指标。这是 ArtFID 中用于合并 LPIPS 和 ArtFID 风格度量指标的公式，可确保比较公平。

如表 II 所示，当使用单一度量指标评估风格迁移算法时，ArtScore 和 ArtFID 风格与人类判断显著相关，且表现优于内容相关指标。这可能是因为这两种度量指标均基于在大规模艺术绘画数据集上训练的模型，并且对艺术风格拥有丰富的先验知识。如表 II 和图 5 所示，在与其他度量指标结合时，事实证明，在不同的聚合设置下，ArtScore 能有效提升与人类判断的相关性。具体而言，将 ArtScore 得出的排名与 LPIPS 结合，其相关系数高于强基线 ArtFID；ArtFID 由 LPIPS 和 ArtFID 风格组成。此外，将 ArtScore 与 ArtFID 结合可进一步提升相关性。这些结果表明，ArtScore 能够准确、定量地评估风格迁移方法生成的图像质量。

有趣的是，我们观察到 GramLoss 度量指标与人类评价分数呈负相关，而将其与其他度量指标结合使用会降低相关性。这表明 Gram 损失可能与人类艺术评价并不十分一致。以往的研究也显示了类似的模式 [17], [39], [40], ：在人工评估研究中，较低的 Gram 损失并不一定意味着更高的偏好或更好的风格保留。

C. 用户研究

我们开展了一项用户研究，以在实例级评估我们提出的 ArtScore 与现有度量指标和人类艺术评价之间的一致性。参与者看到图像对，并被要求选出其中哪张图像更接近美术绘画。比较包含三种不同情形：使用相同、不同或不使用风格-内容参考生成的图像对。我们使用与算法级实验相同的 12 种风格迁移方法，生成采用相同或不同参考的艺术图像，并利用预训练的 StyleGAN3 [66] 无条件合成艺术图像。

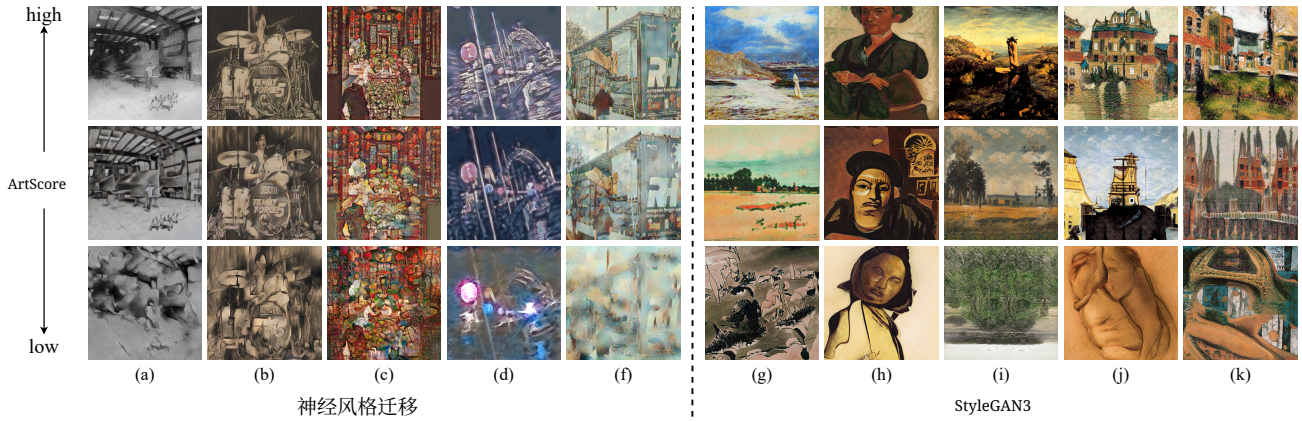


图 6. 对条件生成（使用神经风格迁移算法）和无条件生成（使用 StyleGAN3）的图像进行定性比较。ArtScore 值较低和较高的图像之间的一个主要区别是，高分图像往往更接近美术绘画。这体现在构图更有条理、笔触更细腻、伪影更少。

表 III 我们通过度量各指标的艺术性比较结果与人类判断相符的比率，评估每项指标与人类艺术判断的对齐情况。所提出的 ArtScore 表现最佳。我们使用麦克尼马尔检验，考察 ArtScore 与其他度量指标在艺术评价方面的差异。

比较	度量指标	准确率	χ^2	p-值
神经风格迁移（NST）与相同风格和内容参考	ArtScore	**0.6433**	-	-
	Gram	0.4503	21.1250	0.0000
	LPIPS	0.5848	2.5423	0.1108
	内容	0.5819	2.5478	0.1105
	SSIM	0.6053	1.0992	0.2944
	L2	0.5731	3.7786	0.0519
神经风格迁移（NST）与不同风格和内容参考	ArtScore	**0.5639**	-	-
	Gram	0.4511	5.9225	0.0149
	LPIPS	0.5301	0.5378	0.4633
	内容	0.4361	8.6429	0.0033
	SSIM	0.4699	4.8403	0.0278
	L2	0.5226	0.8547	0.3552
StyleGAN3	ArtScore	**0.6426**	-	-

为确保数据的可靠性，我们在用户研究中加入了一道注意力检测题。每位参与者总共需要回答 46 道题，包括三种比较类型（相同、不同和无风格-内容参考）各 15 对图像，以及注意力检测题。我们会剔除未通过注意力检测的回答。我们还会排除参与者对两张图像的偏好人数相同的图像对，以避免出现任何有争议的回答。

为了评估各项度量指标与人工艺术评价的一致性，我们以人工评估得出的艺术性比较结果作为参考标准。随后，我们根据各项度量指标所产生的比较结果，计算其准确率分数。我们还进行了麦克尼马尔检验 [67]，以确定 ArtScore 与其他度量指标在与人工艺术评价的一致性方面的差异是否具有统计显著性。在这项比较中，我们排除了 ArtFID，因为它基于生成图像与真实艺术作品的深层特征分布，且在仅提供有限图像时无法进行实例级比较。

表 III 所示的结果表明，在比较通过风格迁移方法生成的艺术作品时，与其他度量指标相比，ArtScore 与人类判断更加一致。值得注意的是，当被比较的图像基于不同的风格-内容参考时，ArtScore 的表现显著优于 Gram 损失、内容损失和 SSIM。这与我们在算法级实验中的发现一致。其他度量指标无法直接用于比较不同图像的艺术性，在艺术评价任务中也可能得出不准确的结果。内容损失、SSIM、LPIPS 和 L2 根据深层特征或像素值衡量内容保留情况。虽然通过与真实照片参考进行比较，它们在一定程度上能够衡量生成图像的构图结构如何，但却完全**忽略了风格信息**。因此，它们无法捕捉艺术评价中的这一重要方面，例如笔触与美术作品中的笔触有多相似。另一方面，Gram 损失只衡量风格，完全**忽略内容**。例如，它可能忽略扭曲内容的伪影，而这些伪影会削弱作品的美术感。此外，在比较无条件生成的图像时，其他所有度量指标都无法发挥作用，因为没有成对的参考图像就无法计算这些指标。另一方面，在这种情况下，ArtScore 与人类判断仍然保持一致。

对于每项度量指标，在比较具有不同风格-内容参考的图像时，其准确率分数都会下降。我们推测，这是因为当待比较的图像相似度和一致性较低时，即使对人类而言，艺术判断也会变得更加困难且更具主观性：当待比较图像的参考内容不同时，有争议的配对占所有配对的比例会从 8.47% 上升至 12.03%。

有趣的是，我们注意到 Gram 损失的表现并不优于随机猜测，这再次表明 Gram 损失可能与人工艺术评价并不十分一致。

D. 定性分析

我们进行了定性分析，以深入了解 ArtScore 高低不同的生成图像之间的差异。结果见图 6。我们采用定量实验中生成的图像，并选取 ArtScore 各异的示例。在 NST 生成的示例中，我们观察到，高 ArtScore 图像往往结构更清晰、更自然（例如，(a) 中的屋顶、(b) 和 (c) 中的人物，以及 (f) 中的卡车），引入的伪影也更少（如 (d) 和 (f) 所示）。在 StyleGAN3 生成的示例中，我们观察到高分图像往往

表 IV 实例级评估的消融实验。我们展示了艺术性评估器在以下情况下的性能：1）训练时仅使用单一图像域；2）训练时将学习排序目标替换为均方误差损失；以及 3）训练时移除随机交换增强。最佳性能以粗体突出显示，次佳性能加下划线。

度量指标	准确率		
	NST 相同参考	NST 不同参考	StyleGAN3 无参考图
ArtScore	**0.6433**	**0.5639**	**0.6426**
仅人脸	0.5205	0.5188	0.4729
仅马	0.5468	0.5150	0.5054
仅风景	0.5877	0.5564	0.5884
均方误差损失	0.6404	0.5627	0.5812
无随机交换	0.5556	0.5414	0.6245

表 V 算法层面评估的消融实验。我们展示了单独使用 ArtScore（上方）以及与 ArtFID 风格和内容损失指标搭配使用时（下方）的性能。最佳性能以粗体标出，次佳性能以下划线标出。

–

度量指标		斯皮尔曼 $\rho\uparrow$	p 值 $\downarrow$
ArtScore		**0.6364**	**0.0261**
仅人脸		0.4336	0.1591
仅马		0.5455	0.0666
仅风景		0.5594	0.0586
均方误差损失		0.5594	0.0586
不使用随机交换		0.5734	0.0513
ArtScore		**0.8196**	**0.0011**
仅人脸		0.7860	0.0024
仅马	+ArtFID 风格	0.7951	0.0020
仅风景	+内容损失	0.8000	0.0018
均方误差损失		0.8000	0.0018
无随机交换		0.8163	0.0012

细节更多（(i) 和 (j)）、结构更和谐（(h) 和 (j)），笔触也更细腻（(g) 和 (k)）。因此，高分图像更接近美术绘画。这些观察结果表明，所提出的 ArtScore 能有效捕捉人工智能生成图像中与美术相关的特质，例如清晰度、自然度和和谐感。

E. 消融实验

消融实验结果报告于表 IV、V 和 VI 中。在评估将 ArtScore 与其他度量指标相结合的有效性时，我们采用 ArtFID 风格和内容损失作为额外的度量指标，并采用基于排序的聚合，因为如表 II 所示，这种组合产生了最高的相关性。

仅使用单一域。我们仅使用属于人脸域、马域或风景域的图像序列训练 ArtScore，以评估其性能。与使用完整数据集相比，这种设置会导致性能显著下降，凸显了在多个艺术作品域上训练艺术性评估器的重要性。这样，模型便能从内容和风格各异的多样化数据集中学习。有趣的是，我们发现人脸域的结果最差，其次是马域，而风景域的表现最佳。我们推测，这可能是因为各个域的对齐程度不同。人脸域的对齐程度很高，因为源生成器和迁移后的生成器都是使用经过预处理和对齐的数据训练的。马域则处于中等——

表 VI
算法层级评估：考察序列长度和数据集大小的影响。我们展示了采用两种不同设置训练的 ArtScore 的性能：1）缩短序列长度，以表示更粗粒度的层级。

艺术与真实风格之间的插值（上），以及 2）所构建数据集的子集，表示训练期间使用更少的序列（下）。最右列表示我们在主要结果中采用的设置。

序列长度	4	6	8	10	12
均方误差损失	0.4056	0.4406	0.5245	0.5385	0.5594
ListMLE损失	0.5175	0.5734	0.6084	0.6364	0.6364
训练集比例	**20%**	**40%**	**60%**	**80%**	**100%**
均方误差损失	0.3916	0.5035	0.5175	0.5385	0.5594
ListMLE损失	0.4266	0.5804	0.6084	0.6084	0.6364

得到了适度的对齐，因为只要图像中至少有一匹马，训练数据就不需要对齐。相比之下，风景域的对齐程度最低，不像马域受物体约束，也不像人脸域受位置约束。因此，我们假设，在真实绘画作品的情境中，对对齐程度的要求有所不同会更加自然，因为画家创作作品时未必总会设置这类约束。总体而言，结合这些域能取得最佳结果。

采用更粗粒度的插值。在主要结果中，我们以 0.1 为步长，将插值参数 α 从 0.0 变化到 1.0，得到每个序列 12 张图像，其中包括一张真实图像和 11 张具有不同程度艺术性的插值图像。我们通过改变 α，探索每个序列图像数量更少的伪标注数据集变体； α表示艺术风格与真实风格之间更粗略的差异。如表 VI 所示，较长的序列长度通常有助于提升相关性表现。这与我们的动机一致：训练 ArtScore 区分所构建数据集中的细粒度差异，从而增强其在推理时的稳健性。

使用更少的序列。我们仅使用所构建数据集的一个子集来训练 ArtScore。如表 VI 所示，使用更多序列进行训练能让模型受益。更全面、更多样化的训练集能够增强模型的泛化能力，并提升其在推理时对未见图像的表现。

替换学习排序目标。我们用均方误差损失替换 ListMLE损失来训练 ArtScore。随后，模型经过优化，以便为输入的图像序列输出准确的伪排名值。我们的评估结果（表 IV、V）表明，在不同条件下（如序列长度和训练集大小，见表 VI），ListMLE 在实例层面和算法层面评估中的表现均优于均方误差损失。使用均方误差损失时，即使不同图像的真实艺术性可能有所不同，模型也会被促使为各自序列中排名相同的图像给出相同分数。这一额外约束可能会妨碍学习过程。

移除随机交换论证。训练艺术性评估器时，我们移除不同序列之间的随机交换数据增强。同样，我们观察到性能下降。这项实验表明随机交换增强是有效的，它能提供这样一种场景：ArtScore 需要比较内容和风格不同的图像的艺术性，使其】【。

在推理过程中更加稳健。

V. 结论与未来工作

本文提出了一种名为 ArtScore 的新度量指标，用于根据人工智能生成图像与美术绘画的相似度评估其质量。我们的框架包括 StyleGAN 适配、插值图像生成和 ArtScore 训练，可自动标注艺术性水平，从而实现不同算法之间公平且定量的比较。实验结果表明，与所有现有度量指标相比，所提出的 ArtScore 与人类艺术判断更加一致。此外，将 ArtScore 与选定的现有度量指标相结合，可以高度准确地评估生成图像的整体质量。无论单独使用还是与其他指标结合使用，ArtScore 都为图像生成领域的研究人员和艺术家提供了宝贵的工具。我们希望本研究能够激励更多关于开发可靠艺术评价度量指标的研究，从而推动该领域的发展。

未来，我们计划进一步发展 ArtScore，将其作为损失项纳入图像生成模型的训练或生成过程。例如，ArtScore 可用作扩散模型采样的引导函数 [68]，并用于正则化扩散模型的训练 [69],，使生成结果更接近美术绘画。虽然我们的重点主要面向人工智能研究人员，但我们也认识到，艺术家和艺术研究者对于“艺术性”这一概念可能有更细致的见解。目前还没有数据集提供此类标注，而从专家处收集数据也面临诸多挑战，包括成本、多样性以及艺术判断本身所固有的主观性。因此，我们遵循相关研究中的常见做法，采用自动化方法和用户研究来评估艺术性。未来，我们计划通过案例研究整合专家见解，以加深对艺术性的理解。

致谢

本研究部分得到了罗切斯特大学戈根数据科学研究所的支持。

参考文献

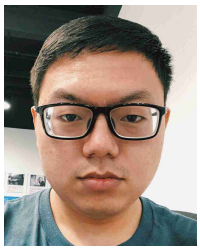
[1] L. A. Gatys、A. S. Ecker 和 M. Bethge，“艺术风格的神经算法”，*arXiv* 预印本 *arXiv:1508.06576*，2015年。[2] X. Huang 和 S. Belongie，“通过自适应实例归一化实现实时任意风格迁移”，发表于*ICCV*，2017年。[3] J. An、H. Xiong、J. Huan 和 J. Luo，“通过神经架构搜索实现超快速照片级真实感风格迁移”，发表于*AAAI*，2020年。[4] S. Huang、J. An、D. Wei、J. Luo 和 H. Pfister，“Quantart：量化图像风格迁移，以实现高视觉保真度”，*arXiv* 预印本 *arXiv:2212.10431*，2022年。[5] I. Goodfellow、J. Pouget-Abadie、M. Mirza、B. Xu、D. Warde-Farley、S. Ozair、A. Courville 和 Y. Bengio，“生成对抗网络”，发表于*NeurIPS*，2014年。[6] A. Elgammal、B. Liu、M. Elhoseiny 和 M. Mazzone，“CAN：创意对抗网络，通过学习风格并偏离风格规范来生成‘艺术’”，*arXiv* 预印本 *arXiv:1706.07068*，2017年。[7] T. Karras、S. Laine、M. Aittala、J. Hellsten、J. Lehtinen 和 T. Aila，“分析并改进 StyleGAN 的图像质量”，发表于*CVPR*，2020年。[8] J.-Y. Zhu、T. Park、P. Isola 和 A. A. Efros，“使用循环一致对抗网络进行非配对图像到图像的迁移”，发表于*ICCV*，2017年。[9] J. Ho、A. Jain 和 P. Abbeel，“去噪扩散概率模型”，发表于*NeurIPS*，2020年。[10] A. Ramesh、P. Dhariwal、A. Nichol、C. Chu 和 M. Chen，“使用 CLIP 潜变量进行分层文本条件图像生成”，*arXiv* 预印本 *arXiv:2204.06125*，2022年。[11] R. Rombach、A. Blattmann、D. Lorenz、P. Esser 和 B. Ommer，“使用潜在扩散模型进行高分辨率图像合成”，载于 CVPR，2022年。[12] Z. Wang、A. C. Bovik、H. R. Sheikh 和 E. P. Simoncelli，“图像质量评估：从误差可见性到结构相似性”，TIP，第13卷，第4期，第600–612页，2004年。[13] T. Salimans、I. Goodfellow、W. Zaremba、V. Cheung、A. Radford 和 X. Chen，“改进的 GAN 训练技术”，载于 NeurIPS，2016年。[14] M. Heusel、H. Ramsauer、T. Unterthiner、B. Nessler 和 S. Hochreiter，“通过双时间尺度更新规则训练的 GAN 收敛到局部纳什均衡”，载于 NeurIPS，2017年。[15] M. Wright 和 B. Ommer，“ArtFID：神经风格迁移的定量评估”，载于 PR，2022年。[16] P. Chandran、G. Zoss、P. Gotardo、M. Gross 和 D. Bradley，“用于结构感知风格迁移的自适应卷积”，载于 CVPR，2021年。[17] J. An、S. Huang、Y. Song、D. Dou、W. Liu 和 J. Luo，“Artflow：通过可逆神经流实现无偏图像风格迁移”，载于 CVPR，2021年。[18] N. Kolkin、J. Salavon 和 G. Shakhnarovich，“通过松弛最优传输和自相似性进行风格迁移”，载于 CVPR，2019年。[19] S. Liu、T. Lin、D. He、F. Li、M. Wang、X. Li、Z. Sun、Q. Li 和 E. Ding，“Adaattn：重新审视任意神经风格迁移中的注意力机制”，载于 ICCV，2021年。[20] W. Xu、C. Long、R. Wang 和 G. Wang，“Drb-gan：用于艺术风格迁移的动态 resblock 生成对抗网络”，载于 ICCV，2021年。[21] Y. Hu、H. Hua、Z. Yang、W. Shi、N. A. Smith 和 J. Luo，“Promptcap：利用 GPT-3 进行视觉问答的提示引导图像描述生成”，载于 ICCV，2023年。[22] Bryandlee。（2021）冻结生成器以实现伪图像迁移。[在线]。可访问：https://github.com/bryandlee/FreezeG [23] U. Ojha、Y. Li、J. Lu、A. A. Efros、Y. J. Lee、E. Shechtman 和 R. Zhang，“通过跨域对应关系实现少样本图像生成”，载于 CVPR，2021年。[24] X. Wang、K. Yu、S. Wu、J. Gu、Y. Liu、C. Dong、Y. Qiao 和 C. Change Loy，“Esrgan：增强型超分辨率生成对抗网络”，载于 ECCV workshops，2018年。[25] T. Karras、S. Laine 和 T. Aila，“用于生成对抗网络的基于风格的生成器架构”，载于 CVPR，2019年。[26] F. Yu、A. Seff、Y. Zhang、S. Song、T. Funkhouser 和 J. Xiao，“Lsun：利用深度学习并结合人工参与构建大规模图像数据集”，arXiv 预印本 arXiv:1506.03365，2015年。[27] I. Skorokhodov、G. Sotnikov 和 M. Elhoseiny，“对齐潜在空间与图像空间，以连接原本无法连接的对象”，载于 ICCV，2021年。[28] F. Xia、T.-Y. Liu、J. Wang、W. Zhang 和 H. Li，“学习排序的列表式方法：理论与算法”，载于 ICML，2008年。[29] J. Johnson、A. Alahi 和 L. Fei-Fei，“用于实时风格迁移和超分辨率的感知损失”，载于 ECCV，2016年。[30] D. Ulyanov、V. Lebedev、A. Vedaldi 和 V. S. Lempitsky，“纹理网络：纹理与风格化图像的前馈合成”，载于 ICML，2016年。[31] H. Huang、H. Wang、W. Luo、L. Ma、W. Jiang、X. Zhu、Z. Li 和 W. Liu，“视频实时神经风格迁移”，载于 CVPR，2017年。[32] X. Gong、H. Huang、L. Ma、F. Shen、W. Liu 和 T. Zhang，“神经立体图像风格迁移”，载于 ECCV，2018年。[33] L. Sheng、Z. Lin、J. Shao 和 X. Wang，“Avatar-net：通过特征装饰实现多尺度零样本风格迁移”，载于 CVPR，2018年。[34] J. An、T. Li、H. Huang、L. Shen、X. Wang、Y. Tang、J. Ma、W. Liu 和 J. Luo，“通过零通道剪枝在高分辨率图像上实现实时通用风格迁移”，arXiv 预印本 arXiv:2006.09029，2020年。[35] T. Karras、S. Laine 和 T. Aila，“用于生成对抗网络的基于风格的生成器架构”，载于 CVPR，2019年。[36] O. Bar-Tal、L. Yariv、Y. Lipman 和 T. Dekel，“MultiDiffusion：融合扩散路径以实现可控图像生成”，载于 ICML，2023年。[37] Z. Xue、G. Song、Q. Guo、B. Liu、Z. Zong、Y. Liu 和 P. Luo，“RAPHAEL：通过扩散路径的大型混合模型实现文本到图像生成”，载于 NeurIPS，2024年。[38] R. Zhang、P. Isola、A. A. Efros、E. Shechtman 和 O. Wang，“深度特征作为感知度量的惊人效力”，载于 CVPR，2018年。[39] Y. Deng、F. Tang、W. Dong、C. Ma、X. Pan、L. Wang 和 C. Xu，“Stytr2：基于 Transformer 的图像风格迁移”，载于 CVPR，2022年。[40] K. Hong、S. Jeon、H. Yang、J. Fu 和 H. Byun，“领域感知的通用风格迁移”，载于 ICCV，2021年。[41] A. Sanakoyeu、D. Kotovenko、S. Lang 和 B. Ommer，“用于实时高清风格迁移的风格感知内容损失”，载于 ECCV，2018年。[42] M. Huh、R. Zhang、J.-Y. Zhu、S. Paris 和 A. Hertzmann，“将图像变换并投影到类别条件生成网络中”，载于 ECCV，2020年。

[43] R. Gal、O. Patashnik、H. Maron、A. H. Bermano、G. Chechik 和 D. Cohen-Or，《Stylegan-nada：基于 CLIP 引导的图像生成器领域自适应》，ACM TOG，2022年。[44] J. Xiao、L. Li、C. Wang、Z.-J. Zha 和 Q. Huang，《通过宽松的空间结构对齐实现少样本生成模型自适应》，载于 CVPR，2022年。[45] J. N. Pinkney 和 D. Adler，《用于实现领域间可控图像合成的分辨率相关 GAN 插值》，arXiv 预印本 arXiv:2010.05334，2020年。[46] Z. Cao、T. Qin、T.-Y. Liu、M.-F. Tsai 和 H. Li，《学习排序：从成对方法到列表式方法》，载于 ICML，2007 年。[47] T. Qin 、 T.-Y. Liu 、 J. Xu 和 H. Li，《 LETOR：用于信息检索中学习排序研究的基准集合》，IR，2010年。[48] K. Nichol。（2016年）《按数字绘画》，WikiArt。 [在线]。可访问：https://www.kaggle.com/c/painter-by-numbers [49] R. Abdal、Y. Qin 和 P. Wonka，《Image2StyleGAN：如何将图像嵌入 StyleGAN 潜在空间？》，载于 ICCV，2019 年。[50] J. Lin 、 H. Hua 、 M. Chen 、 Y. Li 、 J. Hsiao 、 C. Ho 和 J. Luo，《VideoXum：视频的跨模态视觉与纹理摘要》，TMM，2023年。[51] A. Ushio。（2022年）WikiArt 人脸/WikiArt 通用图像集。 [在线]。可访问：https://github.com/asahi417/wikiart-image-dataset [52] A. Radford、J. W. Kim、C. Hallacy、A. Ramesh、G. Goh、S. Agarwal、G. Sastry、A. Askell、P. Mishkin、J. Clark 等，《从自然语言监督中学习可迁移的视觉模型》，载于 ICML，2021年。[53] I. Loshchilov 和 F. Hutter，《解耦权重衰减正则化》，arXiv 预印本 arXiv:1711.05101，2017年。[54] C. Spearman，《两个事物之间关联的证明与测量》，《美国心理学杂志》，1961年。[55] X. Luo、Z. Han、L. Yang 和 L. Zhang，《一致的风格迁移》，arXiv 预印本 arXiv:2201.02233，2022年。[56] D. Chen、J. Liao、L. Yuan、N. Yu 和 G. Hua，《连贯的在线视频风格迁移》，载于 ICCV，2017年。[57] J. Svoboda、A. Anoosheh、C. Osendorfer 和 J. Masci，《用于任意图像风格迁移的两阶段同伴正则化特征重组》，载于 CVPR，2020年。[58] Y. Deng、F. Tang、W. Dong、W. Sun、F. Huang 和 C. Xu，《通过多自适应网络实现任意风格迁移》，载于 ACM MM，2020年。[59] D. Y. Park 和 K. H. Lee，《利用风格注意力网络实现任意风格迁移》，载于 CVPR，2019年。[60] J. Huo、S. Jin、W. Li、J. Wu、Y.-K. Lai、Y. Shi 和 Y. Gao，《用于语义对齐风格迁移的流形对齐》，载于 ICCV，2021年。[61] X. Li、S. Liu、J. Kautz 和 M.-H. Yang，《学习用于快速图像与视频风格迁移的线性变换》，载于 CVPR，2019年。[62] Y. Li、C. Fang、J. Yang、Z. Wang、X. Lu 和 M.-H. Yang，《通过特征变换实现通用风格迁移》，载于 NeurIPS，2017年。[63] Z. Chen、W. Wang、E. Xie、T. Lu 和 P. Luo，《迈向超高分辨率神经风格迁移：通过缩略图实例归一化》，载于 AAAI，2022年。[64] B. Zhou、A. Lapedriza、A. Khosla、A. Oliva 和 A. Torralba，《Places：一个包含一千万张图像的场景识别数据库》，TPAMI，2017年。[65] T.-Y. Lin、M. Maire、S. Belongie、J. Hays、P. Perona、D. Ramanan、P. Doll´ar 和 C. L. Zitnick，《Microsoft COCO：上下文中的常见物体》，载于 ECCV，2014年。[66] J. Pinkney。（2021年11月）在 WikiArt 数据集上训练的 StyleGAN3 模型。[67] [ ] 3 在线。可访问：https://lambdalabs.com/blog/stylegan- Q. McNemar，《关于相关比例或百分比之差的抽样误差的注记》，Psychometrika，1947年。[68] A. Bansal、H.-M. Chu、A. Schwarzschild、S. Sengupta、M. Goldblum、J. Geiping 和 T. Goldstein，《扩散模型的通用引导》，载于 CVPR，2023年。[69] K. Black、M. Janner、Y. Du、I. Kostrikov 和 S. Levine，《通过强化学习训练扩散模型》，载于 ICLR，2023年。



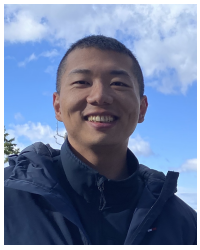
Junyu Chen 于2020年获得中国北京北京大学心理学理学学士学位和经济学文学学士学位，并于2022年获得美国纽约州罗切斯特罗切斯特大学数据科学理学硕士学位。目前，他正在计算机科学系攻读博士学位。

他的主要研究兴趣包括计算机视觉、持续学习和深度生成模型。



Jie An 于2016年和2019年分别获得中国北京大学信息与计算科学学士学位和应用数学硕士学位。他目前正在美国纽约州罗切斯特大学计算机科学系攻读博士学位。

他的主要研究兴趣包括计算机视觉、深度生成模型和 AI+ 艺术。



Hanjia Lyu 于2017年获得中国上海复旦大学经济学文学学士学位，并于2020年获得美国纽约州罗切斯特罗切斯特大学数据科学理学硕士学位。目前，他正在计算机科学系攻读博士学位。

他的主要研究兴趣包括数据挖掘、网络科学、计算社会科学、机器学习和健康信息学。



Christopher Kanan （IEEE 高级会员）于2004年获得美国俄克拉荷马州斯蒂尔沃特俄克拉荷马州立大学的计算机科学与哲学学士学位，于2006年获得美国加利福尼亚州洛杉矶南加州大学的计算机科学硕士学位，并于2013年获得加利福尼亚州圣迭戈加利福尼亚大学圣迭戈分校的博士学位，其研究方向包括机器学习、认知科学和计算机视觉。获得博士学位后，他曾在加利福尼亚州帕萨迪纳的加州理工学院担任博士后学者，随后 мәжбур 他曾在美国国家航空航天局喷气推进实验室担任研究技术人员，直至2015年。目前，他是美国纽约州罗切斯特罗切斯特大学计算机科学副教授。他的实验室主要专注于深度学习基础研究，尤其是终身（持续）机器学习、抗偏差人工智能、医学计算机视觉和语言引导的场景理解。



Jiebo Luo （IEEE 会士）于2011年在柯达研究实验室（罗切斯特）度过了15余年的丰硕职业生涯后，加入美国纽约州罗切斯特的罗切斯特大学计算机科学系。他撰写了500余篇技术论文，并持有90余项美国专利。他的研究兴趣包括计算机视觉、自然语言处理、机器学习、数据挖掘、计算社会科学和数字健康。

罗教授还是计算机协会（ACM）的会士，也是协会的会士。人工智能促进协会（AAAI）、国际光学与光子学学会（SPIE）和国际模式识别协会（IAPR）。他在 2020–2022年 担任 IEEE 多媒体汇刊主编。他曾担任 2010年 ACM 多媒体会议、2012年 IEEE 计算机视觉与模式识别会议（CVPR）、2016年 ACM 多媒体检索国际会议（ICMR）和 2017年 IEEE 图像处理国际会议（ICIP）的程序委员会共同主席，以及 2018年 ACM 多媒体会议的大会共同主席。他曾担任以下期刊的编委会成员：IEEE 模式分析与机器智能汇刊、IEEE 多媒体汇刊、IEEE 电路与视频技术汇刊、IEEE 大数据汇刊、《模式识别》、《机器视觉与应用》和*ACM* 智能系统与技术汇刊。