

Learning to Paint With Model-based Deep Reinforcement Learning

Zhewei Huang^{1,2}

WenHeng¹

ShuchangZhou¹

摘要

本文提出了基于深度强化学习的绘画智能体，展示了如何让机器像人类画家一样作画，他们可以使用少量的笔画来创作出美妙的画作。通过在基于模型的深度强化学习 (DRL) 中使用神经渲染器，我们的智能体学习确定每个笔划的位置和颜色，并制定长期计划，将纹理丰富的图像分解为笔划。实验表明，使用数百个笔画可以获得出色的视觉效果。训练过程不需要人类画家的经验或笔划跟踪数据。

关键词：深度强化学习；神经渲染器

1 引言

绘画是视觉艺术的一种重要形式，象征着人类的智慧和创造力。近几个世纪来，艺术家们使用了各种各样的工具来创作他们的杰作。但如果不花大量时间进行适当的训练，人们很难掌握这项技能。因此，教机器绘画是一项具有挑战性的任务，有助于揭示绘画的奥秘。此外，本课题的研究可以帮助我们构建绘画辅助工具。

我们训练了一个人工智能绘画代理，它可以在画布上按顺序绘制笔划，以生成类似于给定图像的绘画。神经网络用于生成控制笔划位置、形状、颜色和透明度的参数。以前的作品研究过教学机器，以学习绘画相关的技能，如素描^[1-3]、涂鸦^[4]和书写文字^[5]。相比之下，我们的目标是教机器处理更复杂的任务，例如绘制人类肖像和现实世界中的自然场景，这些场景具有丰富的纹理和复杂的结构组成。

我们解决了训练智能体绘制真实世界图像的三个挑战。首先，像人类一样绘画需要智能体有能力将给定的目标图像分解成有序的笔画序列。智能体需要直观地解析目标图像，了解画布的当前状态，并对未来的笔划有预见性的规划。为了实现这一规划，一种方法是给出每一步笔画分解的监督损失，如^[1]所示。然而，这种方法需要笔划分解的真实标签，通常比较难以定义。此外，纹理丰富的图像绘制通常需要数百个笔画才能生成与目标图像相似的绘画，这比涂鸦、素描或人物书写所需的数量高出数十倍，并增加了规划的难度。为了处理问题的不确定性和长期规划挑战，我们建议使用强化学习 (RL) 来训练智能体，因为 RL 可以最大化整个绘画过程的累积回报，而不是最小化每个步骤的监督损失。实验表明，RL 智能体可以用数百个步骤构建笔划分解计划。此外，我们应用对抗性训练策略^[6]来提高生成图像的像素级质量，因为该策略在其他图像生成任务中被证明是有效的^[7]。

其次，我们设计了连续的笔划参数空间，包括笔划位置、颜色和透明度，以提高绘画质量。以前的作品^[1,4,8]设计了离散的笔划参数空间，每个参数只有有限数量的选择，这对于纹理丰富的绘画来说是不够的。相反，我们采用了深度确定性策略梯度 (DDPG)^[9]，它很好地处理了智能体的连续动作空间。

第三，我们构建了一个高效的微神经渲染器，可以模拟画布上数百个笔划的绘制。大多数以前的作品^[1,4,8]通过与不可区分的绘画模拟环境交互来绘画，这些环境作为渲染效果很好，但无法提供关于生成图像的详细反馈。相反，我们训练一个神经网络，将笔划参数直接映射到笔划绘画。通过改变生成模式，渲染器还可以适应不同的笔划设计，如三角形和圆形。此外，差分渲染器可以与 DDPG 组合成单个基于模型的 DRL，该 DRL 可以以端到端的方式进行训练，这显著提高了绘画质量和收敛速度。

总之，我们的贡献有三个方面：

我们使用基于模型的 DRL 算法处理绘画任务，并构建智能体，将目标图像按顺序分解为数百个笔划，从而可以在画布上重新创建绘画。

我们构建了微神经渲染器，用于高效绘制和灵活支持不同的笔划设计，例如贝塞尔曲线、三角形和圆形。神经渲染器允许以端到端的方式训练基于模型的 DRL 智能体，从而提高绘画质量。

实验表明，所提出的绘画智能体可以很好地处理多种类型的目标图像，包括手写数字、街景房屋编号、人像和自然场景图像等。

2 相关工作

基于笔划的渲染 (SBR) 是一种非真实感图像的方法，它通过在画布上放置离散的绘图元素 (如绘画笔划或点画^[10]) 来重建图像。大多数 SBR 算法通过贪心搜索在每一步或需要用户交互来解决笔划分解问题。Haeberli 等人^[11]提出了一种半自动方法，要求用户设置参数来控制笔画的形状，并为每个笔画选择位置。Litwinowicz 等人^[12]提出了一种单层画家式渲染，将画笔笔划放置在图像平面中的网格上，具有随机扰动的位置。一些工作还研究了使用不同笔划设计的效果^[13]以及从视频生成动画的相关问题^[14]。

最近的工作使用 RL 来改进图像的笔划分解。SPIRAL^[8]是一种经过对抗训练的 DRL 智能体，它学习图像中的结构，但无法恢复人像的细节。StrokeNet^[5]结合了微渲染器和循环神经网络 (RNN) 来训练智能体进行绘制，但未能对彩色图像进行概括。涂鸦 SDQ^[4]训练智能体使用 DQN 模拟人类涂鸦。早些时候，Sketch RNN^[1]使用顺序数据集在草图中获得良好的结果。Artist Agent^[15]探索使用 RL 自动生成单个画笔笔划。

3 本文方法

3.1 本文方法概述

绘画智能体的目标是将给定的目标图像分解为可以在画布上重新创建图像的笔画。为了模仿人类的绘画过程，智能体根据观察画布和目标图像的当前状态来预测下一笔。但是，每一步的笔画都需要与之前的笔画和未来的笔画很好地兼容，以减少完成绘画的笔画的数量。我们假设智能体应该在完成给定的比笔画数后最大化累积奖励，而不是当前笔画的奖励。为了实现这种延迟奖励设计，我们使用了一个 DRL 框架，图 1 中显示了整体架构的图表。

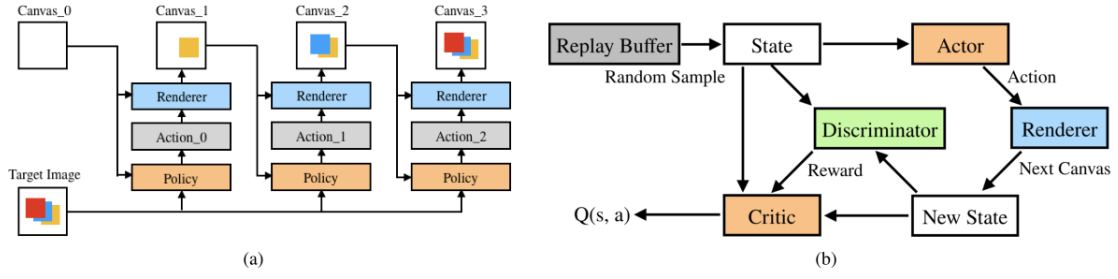


图 1: 总体架构

(a) 在推理阶段，actor 每一步根据画布状态和目标图像输出一组描边参数。然后，渲染器相应地在画布上渲染笔画。(b) 在训练阶段，actor 由 discriminator 和 critic 协助训练。奖励由 discriminator 在每一步给出，训练样本从 replay buffer 中随机抽取。

3.2 模型

给定一个目标图像 I 与一张空画布 C_0 ，智能体的目标是找到一个笔画序列 $(a_0, a_1, \dots, a_{n-1})$ ，其中，在 C_t 上渲染 a_t 能够得到画布 C_{t+1} 。当笔画按照序列渲染完成后，就得到了最终的画面 C_n ，本文的目标是让这个 C_n 尽可能与 I 相似。为此，本文将这个任务建模为一个马尔可夫决策过程，它的状态空间为 \mathcal{S} 、动作空间为 \mathcal{A} ，转移函数为 $trans(s_t, a_t)$ 、奖励函数为 $r(s_t, a_t)$ 。

状态转移函数 状态空间由智能体在环境中观察到的所有可能信息构成。我们将一个状态分为三个部分：画布的状态、目标图像和步数。形式上， $s_t = (C_t, I, t)$ 。 C_t 和 I 是位图，步数 t 作为附加信息构造智能体剩余的步数。转移函数 $s_{t+1} = trans(s_t, a_t)$ 给出了状态之间的转换过程，这是通过在当前画布上绘制一个笔画来实现的。

动作 绘画智能体的动作 a_t 是一组参数，这些参数控制在步骤 t 时将要绘制的笔画的位置、形状、颜色和透明度。我们将智能体的动作定义为将状态映射到确定性动作的策略函数 π ，即 $\pi : \mathcal{S} \rightarrow \mathcal{A}$ 。在步骤 t 时，智能体在预测下一个笔画的参数之前观察状态 s_t 。状态的演化基于转换函数 $s_{t+1} = trans(s_t, a_t)$ ，该函数执行 n 步。

3.3 奖励函数定义

选择一个合适的度量来测量当前画布和目标图像之间的差异，对于训练绘画智能体是至关重要的。奖励的设计如下，

$$r(s_t, a_t) = L_t - L_{t+1} \quad (1)$$

$r(s_t, a_t)$ 是 t 时刻的奖励， L_t 是 I 和 C_t 之间的度量损失， L_{t+1} 是 I 和 C_{t+1} 之间的度量损失。

为了使最终的画布与目标图像相似，智能体应该在整个回合中最大化累积奖励。在每一步中，智能体的目标都是最大化未来折扣奖励的总和 $R_t = \sum_{i=t}^T \gamma^{(i-t)} r(s_i, a_i)$ ，其中折扣因子 $\gamma \in [0, 1]$ 。

4 复现细节

4.1 与已有开源代码对比

本文的开源代码地址：<https://github.com/megvii-research/ICCV2019-LearningToPaint>。本次复现对原有工作进行改进：

动作束 对于许多 RL 任务来说，跳帧是一个强大的技巧，通过限制智能体只观察环境，每 k 帧

而不是一帧做一次动作。这一技巧使智能体有更好的能力学习时间较远的状态和动作之间的联系。智能体预测一个动作并在下一个 $k-1$ 帧重用它，从而以更少的计算成本获得更好的性能。

受到这个技巧的启发，本次改进使用 **Action Bundle**，智能体在每一步预测 k 个笔画，渲染器按顺序渲染这些笔画。这个实践鼓励了对动作空间和动作组合的探索。渲染器可以同时渲染 k 个笔画，大大加快了绘画过程。

通过实验发现，设置 $k = 5$ 是一个很好的选择，可以显著提高性能和学习速度，将奖励折扣因子从 γ 修改为 γ^k 以保持一致性。

WGAN 奖励 GAN 作为一种特殊的损失函数被广泛应用于迁移学习、文本模型和图像恢复，因为 GAN 具有很强的测量生成数据与目标数据之间分布距离的能力。Wasserstein GAN (WGAN) 是原始 GAN 的改进版本，使用 Wasserstein-1 距离，也称为 earth-mover 距离。

如需伪代码，采用如下的写作方式进行描述

Procedure 1 The discriminator in WGAN.

Input: image X
Output: score S

```
for  $i$  in target frame indices do
     $x = \text{Conv2d}(x)$      $x = \text{ReLU}(x)$      $x = \text{Conv2d}(x)$      $x = \text{ReLU}(x)$      $x = \text{Conv2d}(x)$ 
     $s = \text{AVG}(x)$ 
end
```

4.2 实验环境搭建

- 服务器版本：Ubuntu 16.04.1 LTS
- GPU：Tesla P100
- PyTorch: 1.7.0
- tensorboardX
- opencv-python

4.3 界面分析与使用说明

使用以下命令进行测试：

```
pythontest.py --max_step = 100 --actor = actor.pkl --renderer = renderer.pkl --img = image/test.png --divide = 4
ffmpeg -r 10 -f image2 -i output/generated%d.png -s 512x512 -c:v libx264 -pix_fmt yuv420p video.mp4 -q:v 0 -q:a 0
```

5 实验结果分析

本文使用了四个数据集，包括 MNIST^[16], SVHN^[17], CelebA^[18] and ImageNet^[19]。我们表明智能体在绘制各种类型的现实世界图像方面具有出色的表现。

在本文中，我们的目标是训练一个能够绘制任何图像的智能体，而不仅仅是训练集中的图像。因此，我们额外分割出测试集来测试训练智能体的泛化能力。对于 MNIST，我们使用官方定义的测试集。对于其他数据集，我们随机抽取 2000 张图像作为测试集。

我们在图 2 中显示了在不同数据集上训练的代理的测试损失曲线。

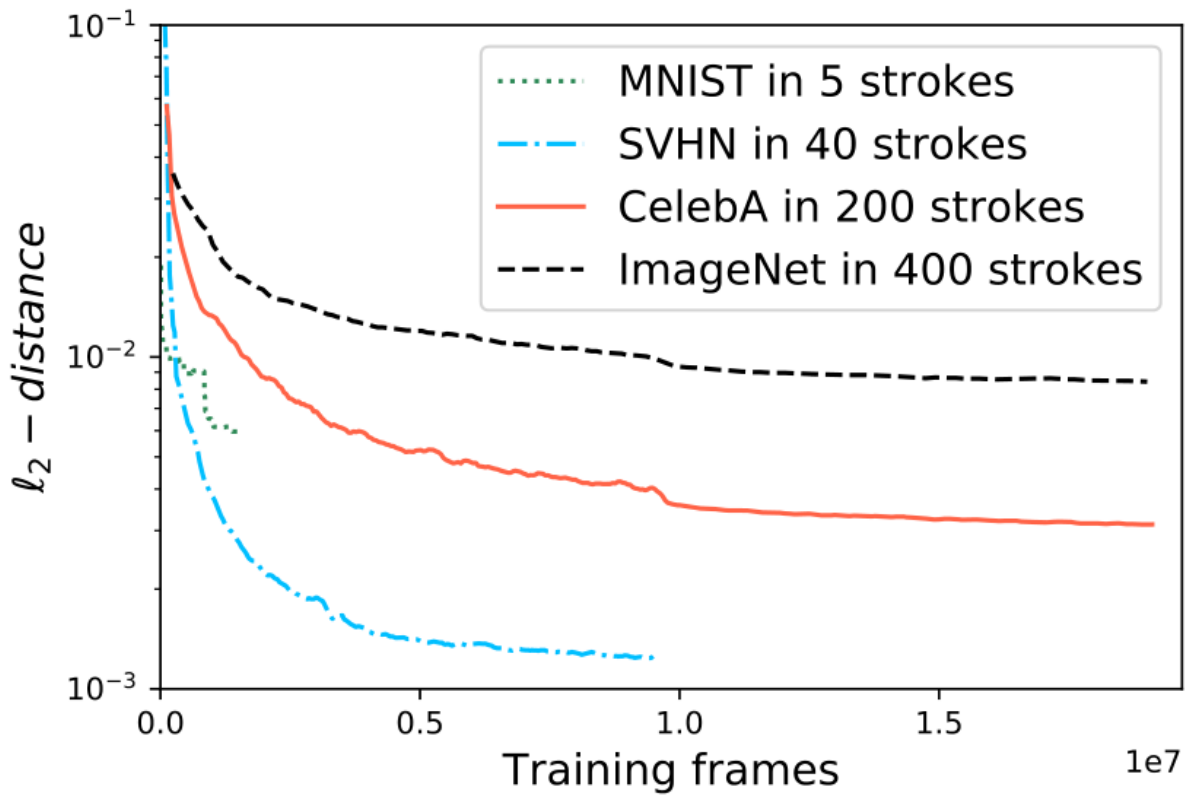


图 2: 测试集的 l_2 损失

图 3是我们训练的智能体在四个不同的数据集上的绘画示例。



图 3: 在多个数据集上绘画结果

6 总结与展望

在本文中，我们训练智能体将目标图像分解为一个有序的笔画序列，以模仿人类在画布上的绘画过程。该训练基于深度强化学习框架，该框架鼓励智能体为基于连续笔画的绘画制定长期计划。此外，我们构建了一个可微神经渲染器来渲染笔画，这允许使用基于模型的 DRL 算法来进一步提高重建图像的质量。训练的智能体可以预测数百甚至数千笔来生成一幅生动的画。实验结果表明，该模型可以处理多种类型的目标图像，并在绘制人体肖像和纹理丰富的自然场景等真实图像时取得良好的表现。

参考文献

- [1] HA D, ECK D. A neural representation of sketch drawings[J]. arXiv preprint arXiv:1704.03477, 2017.
- [2] CHEN Y, TU S, YI Y, et al. Sketch-pix2seq: a model to generate sketches of multiple categories[J]. arXiv preprint arXiv:1709.04121, 2017.
- [3] SONG J, PANG K, SONG Y Z, et al. Learning to sketch with shortcut cycle consistency[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 801-810.
- [4] ZHOU T, FANG C, WANG Z, et al. Learning to doodle with deep q networks and demonstrated strokes [C]//British Machine Vision Conference: vol. 1: 2. 2018: 4.
- [5] ZHENG N, JIANG Y, HUANG D. Strokenet: A neural painting environment[C]// International Conference on Learning Representations. 2018.
- [6] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
- [7] LEDIG C, THEIS L, HUSZÁR F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4681-4690.
- [8] GANIN Y, KULKARNI T, BABUSCHKIN I, et al. Synthesizing programs for images using reinforced adversarial learning[C]// International Conference on Machine Learning. 2018: 1666-1675.
- [9] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning [J]. arXiv preprint arXiv:1509.02971, 2015.
- [10] HERTZMANN A. A survey of stroke-based rendering[C]// . 2003.
- [11] HAEBERLI P. Paint by numbers: Abstract image representations[C]// Proceedings of the 17th annual conference on Computer graphics and interactive techniques. 1990: 207-214.
- [12] LITWINOWICZ P. Processing images and video for an impressionist effect[C]// Proceedings of the 24th annual conference on Computer graphics and interactive techniques. 1997: 407-414.
- [13] HERTZMANN A. Painterly rendering with curved brush strokes of multiple sizes[C]// Proceedings of the 25th annual conference on Computer graphics and interactive techniques. 1998: 453-460.
- [14] LIN L, ZENG K, LV H, et al. Painterly animation using video semantics and feature correspondence[C]// Proceedings of the 8th International Symposium on Non-Photorealistic Animation and Rendering. 2010: 73-80.
- [15] XIE N, HACHIYA H, SUGIYAMA M. Artist agent: A reinforcement learning approach to automatic stroke generation in oriental ink painting[J]. IEICE TRANSACTIONS on Information and Systems, 2013, 96(5): 1134-1144.
- [16] LECUN Y. The MNIST database of handwritten digits[J]. <http://yann.lecun.com/exdb/mnist/>, 1998.

- [17] NETZER Y, WANG T, COATES A, et al. Reading digits in natural images with unsupervised feature learning[J]., 2011.
- [18] LIU Z, LUO P, WANG X, et al. Deep learning face attributes in the wild[C]//Proceedings of the IEEE international conference on computer vision. 2015: 3730-3738.
- [19] RUSSAKOVSKY O, DENG J, SU H, et al. Imagenet large scale visual recognition challenge[J]. International journal of computer vision, 2015, 115(3): 211-252.