arXiv:2304.01182v1 [cs.RO] 3 Apr 2023

学习读取盲文：利用扩散模型弥合触觉现实鸿沟

Carolina Higuera[1],Byron Boots[1],和 Mustafa Mukadam[21华盛顿大学， [2]Meta AI

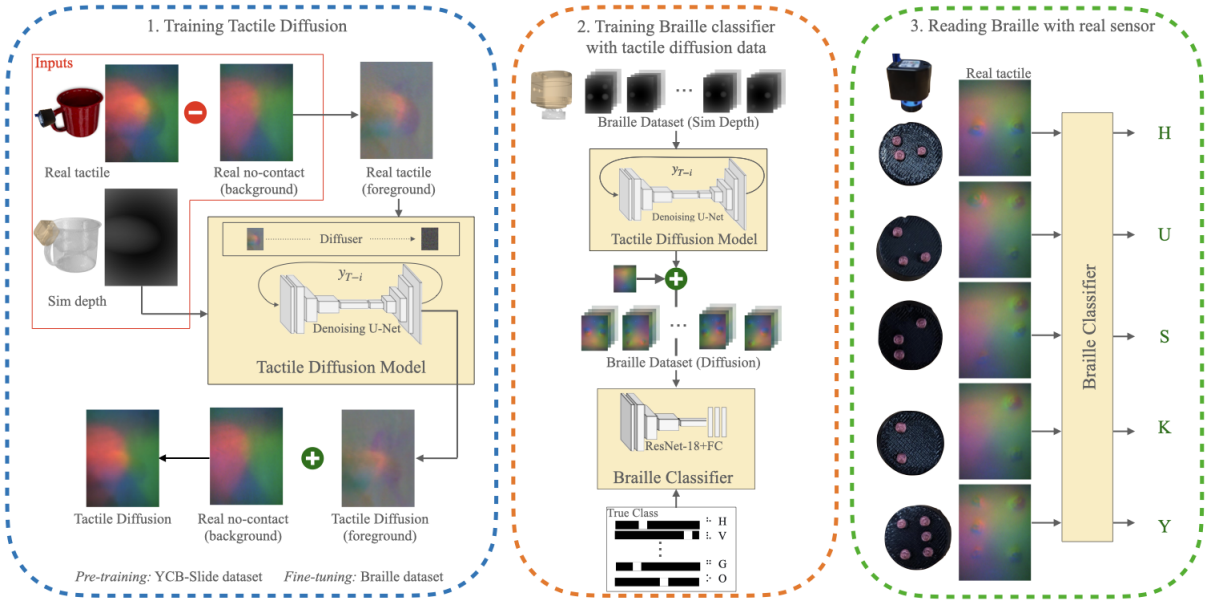

图1：（左）我们的触觉扩散模型在 YCB-Slide 上训练，并使用 20% 的真实盲文数据进行微调，能够根据模拟接触深度生成逼真的触觉图像。（中）我们使用触觉扩散生成的数据训练盲文分类器。（右）在使用 DIGIT 传感器读取真实盲文时，该分类器的表现优于使用仿真数据和其他方法训练的分类器。

摘要—**当大规模采集真实世界数据成本过高时，仿真基于视觉的触觉传感器有助于训练用于接触密集型任务的模型。然而，要准确建模凝胶变形的光学响应并纳入接触动力学，使得仿真到现实迁移颇具挑战。以往研究尝试通过数据增强、微调或训练生成模型来缩小仿真到现实差距。本文提出首个利用概率扩散模型捕捉凝胶变形所引起复杂光照变化的方法。我们的触觉扩散模型能够根据模拟接触深度生成逼真的触觉图像，从而弥合基于视觉的触觉感知中的现实鸿沟。在使用 DIGIT 传感器进行真实盲文阅读任务时，采用我们的扩散模型训练的分类器准确率达到 75.74%，优于使用仿真数据和其他方法训练的分类器。项目页面：**

h
t
t
ps://github.com/carolinahiguera/Tactile-Diffusion

I. 引言

我们提出了一种触觉扩散模型，用于为基于视觉的触觉传感器生成合成图像，逼真地呈现弹性体变形引起的光照变化。改进触觉传感器的渲染技术，将为机器人操作、机器人远程操作、触觉技术和假肢等多种应用提供更好的反馈。逼真地渲染触觉数据，使我们能够在仿真中训练模型，并将其部署到真实环境中，通过虚拟触觉向用户提供反馈，例如，盲文字符的粗糙度和纹理特征。

将基于视觉的触觉传感器与物理引擎结合进行仿真，能够快速生成大量数据，从而用于原型开发和调试操作策略及触觉模型。TACTO [1]和 Taxim [2]等模拟器让我们能够更接近于渲染物体与凝胶交互产生的触觉印痕和高度图。然而，迁移到真实系统并非易事。总体而言，与仿真中的触觉图像相比，真实触觉图像存在分布外差异。对凝胶变形的光学响应进行建模，并纳入接触动力学，使基于视觉的触觉传感器仿真面临巨大挑战。要克服这一困难，我们可以直接依赖真实数据 [3], [4], [5], [6], ，或使用数据增强和微调 [7], [8]。

基于学习的生成模型也在积极探索如何逼真地渲染接触状态下的压痕。这个问题被表述为图像到图像的转换，因此我们可以借鉴计算机视觉文献中的成功方案。使用这些基于学习的模型有若干优势，例如无需标定触觉传感器、对凝胶变形力学的领域特定知识需求较少，以及有望持续改进。当前主流的

文献中的技术是训练生成对抗网络（GAN）以缩小仿真到现实差距。例如，使用 CycleGAN 训练仿真到现实和现实到仿真生成器 [9],在模拟数据和背景上训练条件生成对抗网络 [10]。其他研究提出了反向转换方式，即使用生成对抗网络将真实数据转换为仿真数据，从而使在仿真中学习到的策略能够迁移到真实环境中 [11], [12]。

在本研究中，我们提出利用扩散概率模型 [13] 从仿真数据生成逼真的触觉图像。鉴于这类模型在图像转换 [14], [15], [16], [17], 方面表现出色，我们利用它们捕捉凝胶变形造成的复杂光照变化，以缩小基于视觉的触觉传感中的仿真到现实差距。据我们所知，扩散模型虽已用于机器人学，以生成物体操作 [18] 和规划 [19], [20], [21],的目标状态图像，但尚未用于触觉感知的域适应。我们以物体与凝胶接触的仿真高度图为扩散模型提供条件，并运行去噪过程，生成能够逼真呈现凝胶变形颜色渐变的触觉图像。

我们按照图1所示，在`YCB-Slide` 数据集 [22], 上训练触觉扩散模型。该数据集提供了 DIGIT 传感器与 YCB 物体之间丰富多样的接触交互数据。随后，我们使用少量真实盲文数据对模型进行微调，以适应下游任务领域。我们展示了触觉扩散模型如何（i）生成逼真的触觉仿真图像，突出物体的纹理特征，从而（ii）借助 DIGIT 传感器，在真实盲文阅读任务中实现零样本迁移。使用我们的扩散模型训练的盲文分类器准确率达到 75.74%，优于使用仿真数据及其他方法训练的分类器。

II. 相关工作

A. 基于视觉的触觉仿真

基于视觉的触觉传感器仿真可以加快机器人学中富接触操作任务的策略学习。目前这是一个活跃的研究领域，面临两项主要挑战。首先，如何捕捉接触动力学；其次，如何在传感器变形的情况下对其光学和照明特性进行建模，以获得逼真的感知图像。TACTO [1] 旨在专门应对第二项挑战，同时将接触力的仿真留给底层物理引擎，例如 PyBullet [23]。该模拟器会同步渲染引擎中物体和传感器的位姿，并获取渲染后的触觉和深度印迹。针对 DIGIT [24] 和 OmniTact [25] 传感器的实验结果表明，可能需要采用数据增强或域适应等技术来缩小仿真到现实差距。

Taxim [2]使用二阶多项式查找表模拟 Gelsight传感器 [26]的光学响应。需要使用真实数据进行标定以拟合参数，而且必须针对每个传感器单独执行。根据接触区域内物体的形状构建高度图，并使用金字塔高斯核平滑边缘。提取高度图中每个点的法向值，再通过查找表将其映射为强度值，以渲染触觉图像。 [27] 探讨了如何使用 Pybullet 和 Taxim，为抓取稳定性任务采集接触力和触觉数据。

[28] 提出了一种不同的方法，着重解决捕捉物体与传感器之间接触动力学这一挑战。该方法不渲染触觉图像，而是模拟覆盖传感器接触表面的法向力和剪切力。具体而言，它将传感器垫表示为一组点，并使用基于惩罚的触觉模型来表征每个点上的力。实验表明，该方法能够近似模拟 GelSlim传感器 [29] 的弹性体，从而生成稠密触觉力场。

一些缩小仿真到现实差距的尝试采用基于物理的渲染 [30]或数值方法来模拟弹性体的变形。例如， [8] 采用双变量二维高斯滤波，随后使用Phong着色模型 [31] 渲染传感器的内部照明。研究报告称，GelSight传感器的结构相似性指数（SSIM）约为0.859；然而，反射函数和高斯核的参数必须手动设置。

B. 基于视觉的触觉合成图像生成

用于模拟逼真触觉图像的数据驱动方法所需的领域特定知识较少，并且可以通过在线方式持续改进。为克服模拟物体压痕导致膜变形时特有颜色渐变的难题，生成对抗网络（GAN） [32] 已成为触觉感知领域广泛采用的域适应技术。其核心思想是将网络用作转换器，使仿真图像呈现出真实数据的风格。 [9]提出使用以未配对的模拟数据和真实数据训练的CycleGAN，生成仿真到现实和现实到仿真的生成器。 [10]训练条件生成对抗网络，根据NeRF模型渲染的RGB-D图像以及触觉传感器的参考背景图像生成触觉数据。相对于DIGIT传感器采集的真实图像，其报告的结构相似性指数（SSIM）为 0.865。现实到仿真的方法也在持续探索中。例如， [11], [12]将真实图像转换为仿真图像，并将其用作强化学习策略的状态，以完成边缘与表面跟踪、物体滚动和推动等任务。他们在TacTip、DIGIT和DigiTac这三种传感器上评估了该方法，报告称边缘与表面轨迹的结构相似性指数（SSIM）均高于0.98。

C. 机器人学中的扩散模型

鉴于扩散概率模型近期在计算机视觉领域取得成功， [13], [33], [34], [35], [36], [37], 研究人员开始探索其在机器人学中的优势。[18] 提出将其用于操作任务，根据文本描述生成呈现类人式物体摆放的目标状态图像。 [38]使用扩散模型处理复杂的多步骤三维规划任务，以重新摆放物体。 [19] 提出了一种轨迹级扩散概率

模型，这种方法能够通过使用不同的引导函数，为多种任务规划轨迹。 [20] 专注于学习用于六自由度抓取的SE(3)扩散模型，并将抓取姿态分布表示为代价函数。 [21]提出了一种将扩散模型用于离线强化学习的进化规划方法，可借助奖励梯度的引导实现自我改进。据我们所知，这是首项将扩散模型用于触觉感知的工作。

III. 触觉扩散模型

本文聚焦于缩小基于视觉的触觉传感器图像渲染中的仿真到现实差距这一问题。TACTO [1] 和 Taxim [2],等模拟器能够加速模型学习，使其从接触中提取有用信息，并学习用于富接触操作任务的策略。然而，由于真实图像与仿真图像之间存在光学差异，这些模型的仿真到现实迁移可能颇具挑战性，因为这些模拟器无法完全复现类Gelsight传感器弹性体上复杂且非理想的光传输特性。我们的目标是利用扩散模型，根据仿真中的深度图像生成逼真的DIGIT传感器 [24]合成图像。我们探索扩散模型捕捉传感器内部照明以及物体与弹性体接触时因压痕引起的膜变形的能力。

A. 训练触觉扩散模型

我们将生成逼真合成触觉图像的问题视为图像到图像转换。给定对齐的仿真触觉图像与真实触觉图像对，我们的目标是将触觉模拟器输入的灰度深度图像转换为一幅合理的彩色图像，使其呈现接触状态下的压痕周围的颜色渐变。为此，我们采用了 [14] 中提出的图像着色方法。

条件扩散模型使我们能够将去噪过程表述为 $p(y|x)$，其中 x 是用于条件输入或查询模型的图像，而y 是生成的彩色图像。我们的目标是学习逆向过程，以恢复真实触觉图像 y_0，其中包含有关接触状态下压痕的信息；模型以触觉模拟器生成的相应深度图像 x作为条件。

在正向过程中，原始图像 y_0 会在 $t \in T$ 个时间步内通过添加噪声 ϵ_t 逐渐扩散。噪声取自标准正态分布，混入图像的噪声量由超参数 β_t控制：

$$\begin{aligned} \hat{y}_1 &= \sqrt{1-\beta_1} y_0 + \sqrt{\beta_1}\epsilon_1 \\ \hat{y}_t &= \sqrt{1-\beta_t}\hat{y}_{t-1} + \sqrt{\beta_t}\epsilon_t \quad \forall t \in 2 \cdots T \end{aligned} \tag{1}$$

然而，也可以通过迭代代入上述公式，直接采样得到 $\hat{y}_t$ ，如下所示：

$$\hat{y}_t = \sqrt{\alpha_t} y_0 + \sqrt{1-\alpha_t}\epsilon, \quad \text{where } \alpha_t = \prod_{s=1}^{t} 1-\beta_s \tag{2}$$

解码器执行逆向过程，在以 x 为条件的情况下，从 $\hat{y}_t$ 映射回 $\hat{y}_{t-1}$ ，直至恢复原始图像 y_0。为此，我们训练一个解码器 $f_\theta(x, \hat{y}_t, \alpha_t)$，其参数为 θ ，用于恢复添加到图像 $\hat{y}_t$ 上的噪声向量 ϵ_t 。为了训练解码器，我们优化一个类似于去噪分数匹配 [13] 的目标函数。我们的损失函数对应于以下目标：

$$L(\theta) = \mathbb{E}_{(x,y_0)}\, \mathbb{E}_{\epsilon,t}\left[\|\epsilon - f_\theta(x, \hat{y}_t, \alpha_t)\|^2\right] \tag{3}$$

B. 将仿真深度作为条件图像进行推理

编码器的正向过程经过构造，使得最终潜变量 $\hat{y}_T$ 的条件分布 $p(\hat{y}_T|y_0, x)$ 和边缘分布$p(\hat{y}_T|x)$均成为标准正态分布。因此，我们可以从纯高斯噪声开始推理过程，然后经过 T个迭代步骤，映射回凝胶变形 y_0。

将网络重新参数化为预测添加的噪声后，我们可以通过重新排列 y_0 中的项公式 (2)来近似初始图像：

$$\begin{aligned} y_0 &\approx \frac{1}{\sqrt{\alpha_T}}\left[y_T - \sqrt{1-\alpha_T}\hat{\epsilon}\right] \\ &= \frac{1}{\sqrt{\alpha_T}}\left[y_T - \sqrt{1-\alpha_T} f_\theta(x, \hat{y}_T, \gamma)\right] \end{aligned} \tag{4}$$

根据贝叶斯定理，我们可以得到条件扩散分布 $q(\hat{y}_{t-1}|\hat{y}_t, y_0)$，以恢复添加噪声前的图像，该分布对应于正态分布：

$$\mathcal{N}\left[\frac{(1-\alpha_{t-1})}{1-\alpha_t}\sqrt{1-\beta_t}\hat{y}_t + \frac{\sqrt{\alpha_{t-1}}\beta_t}{1-\alpha_t} y_0, \frac{\beta_t(1-\alpha_{t-1})}{1-\alpha_t}\mathbf{I}\right] \tag{5}$$

将 y_0 根据公式 (4)得到的近似值代入公式 (5)中的条件扩散分布后，反向过程的每次迭代都可以表示为：

$$\hat{y}_{t-1} = \frac{1}{\sqrt{1-\beta_t}}\hat{y}_t - \frac{\beta_t}{\sqrt{1-\alpha_t}\sqrt{1-\beta_t}}\hat{\epsilon} + \sigma\epsilon \tag{6}$$

其中， $\sigma\epsilon$ 表示条件扩散分布的方差，并且它取自标准正态分布 [39]。

IV. 实验

在本节中，我们评估触觉扩散在盲文阅读任务中的表现，如图1所示。首先，我们使用 YCB-Slide 数据集中的丰富多样的数据集，对基于视觉的触觉传感器与标准 YCB 物体之间的滑动交互进行预训练 [22]。随后，我们使用少量真实的盲文触觉图像样本对模型进行微调。我们利用触觉扩散从仿真数据生成逼真的触觉图像，以训练盲文分类器。我们认为，阅读盲文是一项有趣且自然的触觉任务，它阐明了在触觉仿真中渲染逼真的物体纹理对于构建用于下游任务的触觉模型的重要性。我们展示了触觉扩散使分类器能够实现零样本迁移，从而使用 DIGIT 传感器读取真实盲文文本。

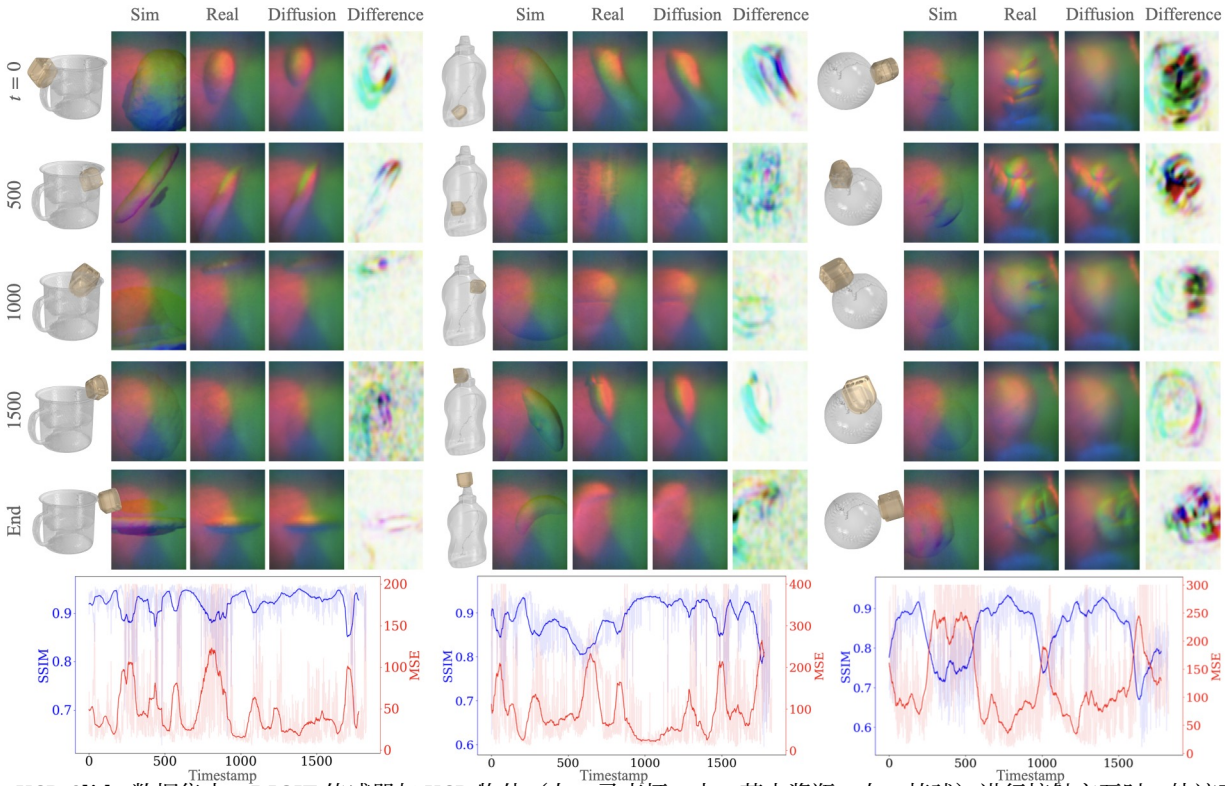

图 2：YCB-Slide 数据集中，DIGIT 传感器与 YCB 物体（左：马克杯，中：芥末酱瓶，右：棒球）进行接触交互时，轨迹不同时间戳对应的仿真图像、真实图像、触觉扩散生成图像，以及生成图像与真实图像之间的差异。触觉扩散生成的图像与真实图像结构相似（结构相似性指数通常高于 0.80）。差异主要来自接触区域的轻微错位，或未能呈现传感器相机捕捉到的特定纹理信息，例如芥末酱瓶标签上的文字。

A. 触觉扩散预训练

我们使用 YCB-Slide 数据集 [22], , 其中包含 DIGIT 传感器与 10 个 YCB 物体之间的滑动接触交互。该数据集包含真实触觉图像、传感器位姿以及真实网格模型；每个物体有 5 条轨迹，每条持续 60 秒。利用传感器位姿和物体网格模型，我们使用 TACTO 仿真器重现每条轨迹，并收集每个时间步的仿真深度图。用于训练扩散模型的数据集包含 18 万对对齐的仿真深度图与真实触觉图像，其中 80% 用于训练，20% 用于测试。

触觉扩散模型的解码器采用条件 U-Net 主干网络，每个下采样和上采样阶段分别包含 2 个 ResNet 块。按照 [14], [40]，我们通过拼接来引入仿真深度图作为条件。我们使用线性噪声调度器（$1e^{-4}$，0。02），训练和推理时的时间步数为 $T = 500$ 。我们使用 Adam [41] 优化器，学习率为 $1e^{-4}$。

为了训练我们的触觉扩散模型，我们遵循图1 （左）所示的流程。我们通过进行前景提取来预处理真实数据。具体而言，我们从所有图像中减去传感器采集的一张真实无接触图像。不同的 DIGIT 传感器由于制造工艺的差异，背景可能略有不同，因此这样可以使扩散模型不受背景影响，而只关注发生在凝胶变形周围的颜色渐变。扩散模型训练和推理的流程示意图如图1所示。在推理时，我们以纯高斯噪声作为去噪过程的起点，并将仿真深度图作为输入来对模型进行条件控制。其直观含义是询问模型：使用真实传感器时，仿真深度图会呈现出什么样子。推理完成后，我们通过将背景加回生成的前景来进行后处理。作为背景，我们可以添加将用于下游任务的任意 DIGIT 传感器的无接触图像。在训练和推理过程中，我们使用一块 RTX- 3080 GPU，对一批包含 30 张图像的数据进行推理所需的时间为 23 秒。

为评估触觉扩散模型的性能，我们在测试数据上逐像素计算结构相似性指数（SSIM）和均方误差（MSE）。该数据集对应于 `YCB-Slide` 数据集中未用于训练扩散模型的轨迹。在图 2中，我们展示了 DIGIT 传感器的模拟、真实和生成触觉图像的若干快照，以及它们之间的差异。对于马克杯和芥末酱瓶，生成图像在整个轨迹上的结构相似性指数均高于 0.80。这意味着，与真实传感器采集的图像相比，扩散模型能够生成在结构、凝胶变形和亮度信息方面相似的触觉图像。马克杯图像的差异主要是接触区域轻微错位；对于芥末酱瓶，除错位外，模型还

无法准确呈现传感器相机捕捉到的非常具体的纹理信息（例如$t = 500$中的标签文本）。然而，这类纹理并不提供任何有用的接触信息，尽管它会对图像相似度指标产生负面影响，尤其是均方误差。对于棒球，扩散模型更难以根据仿真数据生成精细细节。比较真实触觉图像与生成图像时，可以发现真实传感器捕捉到了大量纹理信息，而扩散模型将其概括为一条接触路径。尽管从结构相似性指数和均方误差来看，图像相似度较低，但扩散模型准确捕捉到了接触区域。

B. 触觉扩散微调

我们使用盲文接触交互的（仿真深度图、真实）图像对，对触觉扩散模型进行微调。用于微调和测试的数据集由真实触觉数据组成，这些数据是在 DIGIT 传感器与 27 个 3D 打印盲文字符（字母 A-Z 和 #）接触时采集的。每个物体的尺寸为 $2 \times 2 \times 1$cm。数据采集装置由一台 Franka Panda 机械臂和一个定制的 3D 打印法兰组成，DIGIT 传感器安装在法兰上，使传感器面与末端执行器垂直。以物体中心为基准，接触点分布在一个 3D 网格上，网格尺寸为 $3 \times 3 \times 4$，XY 轴方向的步长为 ± 5mm，Z 轴方向的步长为 -2mm，并进行 $0°$、$\pm 45°$ 和 $\pm 90°$ 的偏航旋转。每个物体由此得到 180 张分辨率为 320×240 的图像。此外，我们还使用 PyBullet 和 TACTO 复现了相同的数据采集过程，以获得每次交互中传感器凝胶与物体接触时的深度穿透图像。我们的数据集包含 1080 对微调图像和 837 对测试图像。

我们使用微调数据集中的 20% 对先前在 YCB-Slide 上训练的触觉扩散模型进行微调。在 图 3中，我们展示了不同字母对应的模拟、真实和由触觉扩散生成的触觉图像样本，并以条件生成对抗网络作为基线。尽管文献中常用条件生成对抗网络生成触觉图像，但相关模型并未开源，因此我们使用自行实现的条件生成对抗网络，并以相同的仿真深度图像作为条件。

从定性角度看，触觉扩散能够细致地呈现物体凹陷周围的颜色变化，从而表示相应的凝胶变形。这些凹陷对应于构成每个盲文字符特征的凸点。对于部分字符，条件生成对抗网络也能做到这一点，但在呈现凝胶变形程度时会出现错误，或完全未能呈现。这与 [42], 中的发现一致：作者解释说，生成对抗网络能够以牺牲多样性为代价换取保真度，生成高质量样本，但无法覆盖整个分布。总体而言，与扩散模型生成的图像相比，我们发现条件生成对抗网络更难进行条件控制，且纹理细节较少。我们的触觉扩散模型生成的图像与真实图像也存在一些差异。不过，这些差异大多是凸点压痕位置未对齐所致。

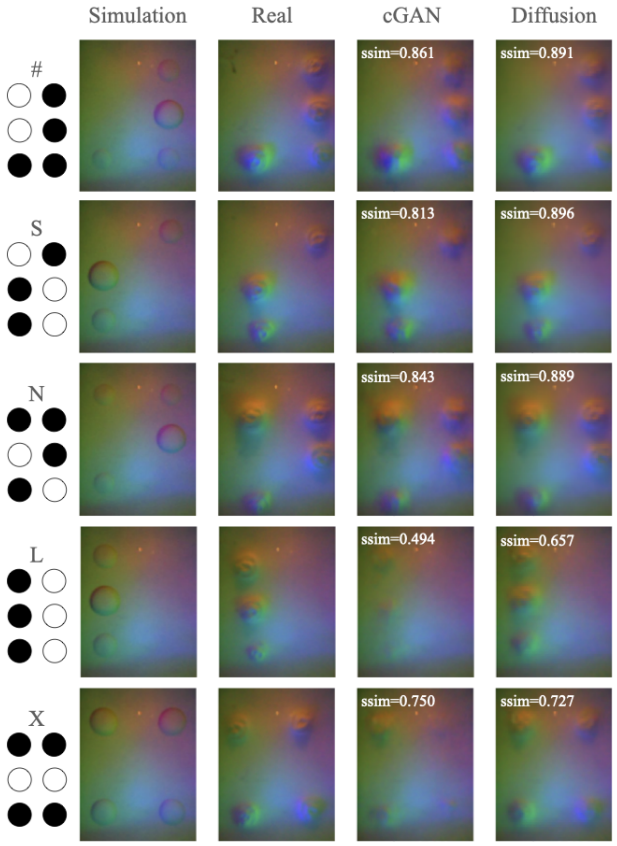

图 3：盲文测试数据集中模拟、真实、条件生成对抗网络和触觉扩散图像的示例对比。与真实传感器图像相比，触觉扩散生成的图像始终具有更高的结构相似性指数。请注意，触觉扩散不会遗漏盲文字符中任何圆点的表示。

为定量评估这两个模型，我们在真实触觉图像测试数据集上，逐像素计算真实图像与生成图像之间的结构相似性指数和均方误差。我们的触觉扩散模型的结构相似性指数为 0.908，逐像素均方误差为 36.02（每个通道的最大像素值为 255）；而条件生成对抗网络的结构相似性指数为 0.879，均方误差为 47.99。这些指标证实，触觉扩散模型生成的合成图像优于条件生成对抗网络生成的图像，而且在结构上与使用真实 DIGIT 传感器采集的该盲文数据集图像相似。

C. 用于阅读盲文的触觉扩散

为了评估使用触觉扩散模型处理仿真数据是否有助于仿真到现实迁移，我们训练了一个分类器，用于检测通过基于视觉的触觉传感器读取盲文文本时对应的字母。我们使用 TACTO 收集了一个单独的数据集，其中包含 DIGIT 传感器与 27 种盲文字符网格模型接触交互产生的深度图像。与用于微调的数据不同，在此数据集中，传感器始终竖直放置以读取盲文，但我们也考虑了传感器方向的微小变化。接触点位于一个 3D 网格上；对于偏航旋转，XY 轴上的步长为 ± 3mm，Z 轴上的步长为 -2mm，旋转范围从$-25°$ 到 $+25°$，步长为 $5°$。

表I：盲文分类任务的指标。

训练数据来源	% 真实数据微调	准确率 %	精确率	召回率
Sim	-	30.23	0.34	0.30
	20	64.99	0.71	0.65
	80	73.11	0.80	0.73
	100	73.95	**0.81**	0.74
模拟 + 数据增强	-	43.48	0.61	0.43
	100	73.23	0.76	0.73
cGAN	-	31.18	0.40	0.31
触觉扩散	-	**75.74**	0.79	**0.76**
Real	-	100.0	1.00	1.00

使用 100% 真实数据训练条件生成对抗网络，使用 YCB-Slide + 20% 真实数据训练触觉扩散

我们使用当前所有常见的基于视觉的触觉传感器仿真到现实问题解决方法来训练该分类器。这些方法包括使用仿真数据训练模型 + 数据增强（光照随机化）、仿真+ 使用真实数据进行微调、仿真 + 数据增强 + 使用真实数据进行微调、条件生成对抗网络和触觉扩散。对于最后两种方法，我们按照图1（中间）表I中的流程，对仿真深度图像进行了数据适配。表I展示了所有这些模型在我们的盲文测试数据集（共 837 张图像）上的测试表现（准确率、精确率和召回率）。

使用真实数据训练下游任务是理想的，因为这样可以确保训练数据和测试数据遵循相同的分布。直接使用真实数据训练的盲文分类器，可以根据盲文字符在传感器凝胶上的压痕完美地区分字母。然而，采集真实数据成本高、耗时长，且难以规模化。实际应用中，我们希望仅在仿真中训练下游任务，并在部署到现实环境时取得良好性能。直接迁移基于原始仿真数据训练的模型，凸显了使用基于视觉的触觉数据时存在的仿真到现实差距。数据增强提升了模型的泛化能力，但仍不足以使训练数据集呈现真实数据的分布。使用真实数据对模型进行微调无疑有助于提升其性能。我们研究如何以尽可能少的真实微调数据获得尽可能高的性能。触觉扩散在由 YCB-Slide 接触数据构成的通用数据集上训练，并仅使用 20% 的任务相关盲文数据进行微调。在这些条件下，该模型在真实触觉数据上的零样本准确率达到 75.74%。若要通过微调仿真模型达到相近的性能，则需要采集多 4 倍的真实触觉数据。

对比不同的数据适配技术，触觉扩散的表现显著优于条件生成对抗网络。这是预料之中的，因为条件生成对抗网络有时会漏掉某些盲文字符凸点的表示。触觉图像中的这些异常可能会被识别为另一个字符，从而导致误分类，影响下游任务。与图像合成中的其他问题一样，[42],从纯噪声一步生成逼真的图像，可能是条件生成对抗网络产生噪声行为的原因。总体而言，这些结果展现了触觉扩散在缩小基于视觉的触觉传感器仿真到现实差距方面的良好前景。

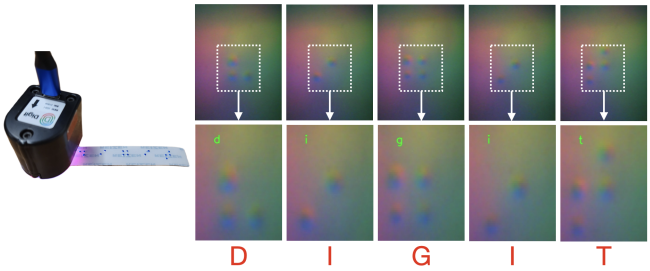

图4：我们使用盲文标签机在乙烯基胶带上打出 *DIGIT* 一词。随后使用 DIGIT 传感器读取胶带，并采用基于触觉扩散训练的分类器进行识别。对于真实尺寸字符的盲文压痕出现的细微变化，该分类器也能很好地泛化。

我们测试了使用触觉扩散数据训练的分类器，借助 DIGIT 传感器读取实际尺寸的盲文文本。我们用盲文标签机在乙烯基胶带上打出单词 *DIGIT*。我们对凝胶上的盲文凹痕进行放大，以近似分类器训练所用的完整盲文图案的尺寸。在图4 中，我们展示了使用触觉扩散训练的分类器对实际尺寸盲文进行零样本读取的快照。

V. 讨论

总结。 本文介绍了扩散概率模型的应用：以仿真数据为条件，为基于视觉的触觉传感器生成逼真的触觉图像。模型以仿真中的深度图像为输入，该图像表示物体压痕造成的凝胶变形高度图。在此数据的条件下，模型会生成前景图像，逼真地呈现凝胶变形的颜色渐变。随后，作为后处理步骤，我们可以添加来自真实传感器的无接触背景图像。这使得该方法能够抵御同类型传感器之间因弹性体制造差异而产生的细微变化。我们的实验表明，在模拟 DIGIT 传感器沿 YCB 物体表面运动的轨迹时，生成的触觉图像与真实图像具有相似的结构。此外，我们还展示了触觉扩散模型能够仅使用仿真数据训练模型，并将其迁移至真实场景，用于盲文分类等下游任务。我们认为，扩散模型是缩小基于视觉的触觉感知仿真到现实差距的一种有前景的工具。

局限性。 与我们的盲文分类器类似，触觉扩散模型可用于对仿真数据集进行后处理，以离线方式训练下游任务模型。目前将扩散模型集成到仿真循环中的主要局限，是执行去噪过程所需的计算时间。为了加快推理速度，可以探索其他方案，例如在潜在空间而非像素空间中运行扩散过程 [36]，以及使用生成对抗网络进行去噪扩散 [43]。

致谢

作者感谢 Sudharshan Suresh、Mike Lambeta 和 Roberto Calandra 在 TACTO 仿真器和 DIGIT 传感器方面提供的帮助。

参考文献

[1] S. Wang、M. Lambeta、P.-W. Chou 和 R. Calandra，《Tacto：一种快速、灵活的开源高分辨率基于视觉的触觉传感器仿真器》，IEEE机器人与自动化快报，第 7 卷，第 2 期，第 3930–3937 页，2022年。[2] Z. Si 和 W. Yuan，《Taxim：一种基于示例的 GelSight 触觉传感器仿真模型》，IEEE机器人与自动化快报，第 7 卷，第 2 期，第 2361–2368 页，2022年。[3] W. Yuan、Y. Mo、S. Wang 和 E. H. Adelson，《利用触觉感知和深度学习进行主动服装材料感知》，载于2018年IEEE国际机器人与自动化会议（ICRA），IEEE，2018年，第 4842–4849 页。[4] S. Dong、D. K. Jha、D. Romeres、S. Kim、D. Nikovski 和 A. Rodriguez，《用于插入的 Tactile-RL：泛化至几何形状未知的物体》，载于2021年IEEE国际机器人与自动化会议（ICRA），IEEE，2021年，第 6437–6443 页。[5] Y. She、S. Wang、S. Dong、N. Sunil、A. Rodriguez 和 E. Adelson，《使用触觉响应式夹爪进行线缆操作》，《国际机器人研究杂志》，第 40 卷，1401，2021 12 14，1385 第 - 期，第 – 页。[6] D. Ma、S. Dong 和 A. Rodriguez，《利用分布式触觉测量进行相对运动跟踪的外部接触感知》，载于2021年IEEE国际机器人与自动化会议（ICRA），IEEE，2021年，第 11 262–11 268 页。[7] M. B. Villalonga、A. Rodriguez、B. Lim、E. Valls 和 T. Sechopoulos，《通过几何接触渲染进行首次触碰时的触觉物体位姿估计》，载于机器人学习会议，PMLR，2021年，第 1015–1029 页。[8] D. F. Gomes、P. Paoletti 和 S. Luo，《为仿真到现实学习生成 GelSight 触觉图像》，IEEE机器人与自动化快报，第 6 卷，第 2 期，第 4177–4184 页，2021年。[9] W. Chen、Y. Xu、Z. Chen、P. Zeng、R. Dang、R. Chen 和 J. Xu，《通过 CycleGAN 实现 GelSight 触觉传感器的双向仿真到现实迁移》，IEEE机器人与自动化快报，第 7 卷，第 3 期，第 6187–6194 页，2022年。[10] S. Zhong、A. Albini、O. P. Jones、P. Maiolino 和 I. Posner，《触摸 NeRF：利用神经辐射场生成触觉感知数据》，载于第六届机器人学习年会，2022年。[11] A. Church、J. Lloyd、N. F. Lepora 等，《通过现实到仿真图像转换实现触觉策略仿真到现实迁移》，载于机器人学习会议，PMLR，2022年，第 1645–1654 页。[12] Y. Lin、J. Lloyd、A. Church 和 N. F. Lepora，《Tactile Gym 2.0：用于比较低成本高分辨率机器人触觉的仿真到现实深度强化学习》，IEEE机器人与自动化快报，第 7 卷，第 4 期，第 10 754–10 761 页，2022年。[13] J. Ho、A. Jain 和 P. Abbeel，《去噪扩散概率模型》，神经信息处理系统进展，第 33 卷，第 6840–6851 页，2020年。[14] C. Saharia、W. Chan、H. Chang、C. Lee、J. Ho、T. Salimans、D. Fleet 和 M. Norouzi，《Palette：图像到图像扩散模型》，载于2022年ACM SIGGRAPH会议论文集，2022年，第 1–10 页。[15] J. Choi、S. Kim、Y. Jeong、Y. Gwon 和 S. Yoon，《ILVR：去噪扩散概率模型的条件控制方法》，arXiv预印本 arXiv:2108.02938，2021年。[16] S.-I. Cheng、Y.-J. Chen、W.-C. Chiu、H.-Y. Tseng 和 H.-Y. Lee，《利用扩散模型从笔画和草图生成自适应逼真的图像》，载于IEEE/CVF冬季计算机视觉应用会议论文集，2023年，第 4054–4062 页。[17] T. Wang、T. Zhang、B. Zhang、H. Ouyang、D. Chen、Q. Chen 和 F. Wen，《图像到图像转换只需预训练》，arXiv预印本 arXiv:2205.12952，2022年。[18] I. Kapelyukh、V. Vosylius 和 E. Johns，《DALL-E-Bot：将大规模扩散模型引入机器人学》，arXiv预印本 arXiv:2210 02438，2022年。[19]M. Janner、Y. Du、J. B. Tenenbaum 和 S. Levine，《利用扩散进行规划以实现灵活的行为合成》，arXiv预印本 arXiv:2205.09991，2022年。[20] J. Urain、N. Funk、G. Chalvatzaki 和 J. Peters，《SE(3)-DiffusionFields：通过扩散学习用于联合抓取与运动优化的代价函数》，arXiv预印本 arXiv:2209 03855，2022年。[21] Z. Liang、Y. Mu、M. Ding、F. Ni、M. Tomizuka 和 P. Luo，《Adapt-Diffuser：作为自适应自演化规划器的扩散模型》，arXiv预印本 arXiv:2302.01877

，2023年。

[22] S. Suresh、Z. Si、S. Anderson、M. Kaess 和 M. Mukadam，“Midas-touch：基于滑动触觉的分布蒙特卡洛推断”，arXiv预印本 arXiv:2210.14210，2022年。[23] E. Coumans 和 Y. Bai，“PyBullet：用于游戏、机器人学和机器学习物理仿真的 Python 模块”，2016年。[24] M. Lambeta、P.-W. Chou、S. Tian、B. Yang、B. Maloon、V. R. Most、D. Stroud、R. Santos、A. Byagowi、G. Kammerer 等，“DIGIT：一种用于手内操作的低成本、紧凑型高分辨率触觉传感器新设计”，IEEE机器人与自动化快报，第5卷，第3期，第3838–3845页，2020年。[25] A. Padmanabha、F. Ebert、S. Tian、R. Calandra、C. Finn 和 S. Levine，“OmniTact：一种多方向高分辨率触觉传感器”，载于2020年IEEE国际机器人与自动化会议（ICRA）。IEEE，2020年，第618–624页。[26] W. Yuan 、 S. Dong 和 E. H. Adelson，“Gelsight：用于估计几何形状和力的高分辨率机器人触觉传感器”，Sensors，第17卷，第12期，第2762页，2017年。[27] Z. Si、Z. Zhu、A. Agarwal、S. Anderson 和 W. Yuan，“基于触觉感知和仿真到现实迁移的抓取稳定性预测”，载于2022年IEEE/RSJ智能机器人与系统国际会议（IROS）。IEEE，2022年，第7809–7816页。[28] J. Xu、S. Kim、T. Chen、A. R. Garcia、P. Agrawal、W. Matusik 和 S. Sueda，“用于机器人操作的高效可微触觉仿真”，载于第六届机器人学习年会。[29] E. Donlon、S. Dong、M. Liu、J. Li、E. Adelson 和 A. Rodriguez，“GelSlim：一种高分辨率、紧凑、稳健且经过校准的触觉感知指套”，载于2018年IEEE/RSJ智能机器人与系统国际会议（IROS）。IEEE，2018年，第1927–1934页。[30] A. Agarwal、T. Man 和 W. Yuan，“基于物理渲染的视觉触觉传感器仿真”，载于2021年IEEE国际机器人与自动化会议（ICRA）。IEEE，2021年，第1–7页。[31] B. T. Phong，“计算机生成图像的照明”，Communications of the ACM，第18卷，第6期，第311–317页，1975年。[32] 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”，Communications of the ACM，第63卷，第144页，2020年11月，第139页。[33] A. Ramesh、P. Dhariwal、A. Nichol、C. Chu 和 M. Chen，“基于CLIP潜变量的层次化文本条件图像生成”，arXiv预印本 arXiv:2204.06125，2022年。[34] A. Nichol、P. Dhariwal、A. Ramesh、P. Shyam、P. Mishkin、B. McGrew、I. Sutskever 和 M. Chen，“GLIDE：利用文本引导扩散模型实现逼真图像生成与编辑”，arXiv预印本 arXiv:2112.10741，2021年。[35] G. Kim、T. Kwon 和 J. C. Ye，“DiffusionCLIP：用于稳健图像操控的文本引导扩散模型”，载于IEEE/CVF计算机视觉与模式识别会议论文集（CVPR），2022年6月，第2426–2435页。[36] R. Rombach、A. Blattmann、D. Lorenz、P. Esser 和 B. Ommer，“基于潜扩散模型的高分辨率图像合成”，载于IEEE/CVF计算机视觉与模式识别会议论文集，2022年，第10 684–10 695页。[37] C. Saharia、W. Chan、S. Saxena、L. Li、J. Whang、E. Denton、S. K. S. Ghasemipour、B. K. Ayan、S. S. Mahdavi、R. G. Lopes 等，“通过深度语言理解实现逼真文本到图像的扩散模型”，arXiv预印本 arXiv:2205.11487，2022年。[38] W. Liu、T. Hermans、S. Chernova 和 C. Paxton，“StructDiffusion：用于新颖对象语义重排的以对象为中心的扩散模型”，arXiv预印本 arXiv:2211.04604，2022年。[39] S. J. Prince，《理解深度学习》。MIT出版社，2023年。[在线]。可访问：https://udlbook.github.io/udlbook/ [40] C. Saharia、J. Ho、W. Chan、T. Salimans、D. J. Fleet 和 M. Norouzi，“通过迭代细化实现图像超分辨率”，IEEE模式分析与机器智能汇刊，2022年。[41] D. P. Kingma 和 J. Ba，“Adam：一种随机优化方法”，arXiv预印本 arXiv:1412.6980，2014年。[42] P. Dhariwal 和 A. Nichol，“扩散模型在图像合成方面优于生成对抗网络”，神经信息处理系统进展，第34卷，第8780–8794页，2021年。[43] Z. Xiao、K. Kreis 和 A. Vahdat，“利用去噪扩散生成对抗网络应对生成式学习三难困境”，载于国际学习表征会议（ICLR），2022年。