注意：本手稿由 UT-Battelle, LLC 根据与美国能源部签订的合同编号 DEAC0500OR22725 撰写。美国政府保留相关权利；出版商接受本文发表即表示承认，美国政府拥有非排他性、已付费、不可撤销、全球范围的许可，可出版本手稿的已发表形式或复制该形式，或允许他人为美国政府目的进行上述操作。美国能源部将根据美国能源部公共获取计划（http://energy.gov/downloads/doe-public-access-plan）向公众提供这些联邦资助研究成果。

从简约表示中提取物理与化学：通过不变变分自编码器进行图像分析

Mani Valleti，[1,a] Yongtao Liu，[2] 以及 Sergei V. Kalinin，[3,b]

[1] 田纳西大学布雷德森跨学科研究中心，美国田纳西州诺克斯维尔，邮编 37916 [2] 橡树岭国家实验室纳米相材料科学中心，美国田纳西州橡树岭，邮编 37831 [3] 田纳西大学材料科学与工程系，美国田纳西州诺克斯维尔，邮编 37916

电子显微成像、光学显微术和扫描探针显微镜方法正在产生越来越大量的图像数据，其中包含有关原子和介观结构及其功能特性的信息。这使得人们有必要开发机器学习方法，从数据中发现物理和化学现象，例如电子显微图像和扫描隧道显微图像中对称性破缺的表现，以及纳米颗粒的变异性。变分自编码器（VAE）正成为无监督数据分析的一种强大范式，能够解耦变化因素并发现最优的简约表示。在此，我们总结变分自编码器的最新进展，涵盖其背后的基本原理和直观理解。我们介绍不变变分自编码器，作为适应成像数据中尺度不变性和平移不变性，并将已知变化因素与待发现因素区分开来的方法。我们还将介绍通过控制变分自编码器架构所带来的机遇，包括条件变分自编码器、半监督变分自编码器和联合变分自编码器。本文讨论了若干变分自编码器应用案例，涉及玩具模型以及扫描透射电子显微镜实验数据集，并着重强调变分自编码器与基本物理原理之间的深刻联系。本文使用的所有代码均可在https://github.com/saimani5/VAE-tutorials获取，且本文可用作将这些方法应用于自有数据集时的应用指南。

[a] svalleti@vols.utk.edu[b] sergei2@utk.edu

1. 引言

自 Levenhoek 发明光学显微镜以来，[1] 显微成像在科学领域发挥的作用日益重要。现代光学显微术是生物学、[2] 材料科学、[3-6] 纳米技术、[7] 领域的主流表征工具，涵盖从桌面设备到定制共聚焦和光片成像系统的各种工具。随后，扫描探针显微镜真正打开了通往纳米世界的大门，[8] 使人们不仅能够观察结构，还能观察其电子、[9-11]磁学、[12,13] 化学、[14] 及机电[15] 功能特性。电子显微镜已成为材料科学、[16] 凝聚态物理、[17] 分子生物学、化学和催化领域的主要工具。[18]

电子显微成像、光学显微成像和探针显微成像的所有领域都普遍经历着从定性成像向定量成像的渐进式转变：在定性成像中，人们仅对图像的显著特征进行定性探究。在光学成像中，这一转变体现为多种生物成像模态、金相成像，以及包括显微拉曼和光致发光测量在内的各种高光谱成像模式。在电子显微成像和扫描探针显微镜中，相关实例包括绘制序参量场[19] 、了解晶界的内部结构[20] ，以及解码二维材料的复杂结构和缺陷化学。对于扫描探针显微镜而言，这还包括分子解折叠光谱学[21,22] ，该技术能够定量揭示单分子反应的机制；以及压电响应力显微镜[23] ，该技术能够量化铁电极化动力学。

过去十年间，现代成像工具采集的数据量呈指数级增长，如今数据形式包括大型图像、多通道多模态图像、高光谱数据集，以及视频等动态图像。向定量成像的转变，很大程度上得益于计算系统的引入；这些系统能够在单个工具、设施以及设施间网络的层面存储和分析数据。如今，几乎所有显微镜系统都配备了本地数据存储和分析功能。尽管这一转变如今已基本完成，但值得注意的是，这一趋势相对较新，主要发生在过去20年间。过去5年里，Google Colabs 和 Jupyter 笔记本的兴起，使个别研究人员能够轻松共享代码和分析工作流程。最后，Amazon AWS、Google Cloud 或 Microsoft Azure 等云服务能力的迅速增长，如今为

实现多个设施之间的数据流整合，以及成像数据的云存储和分析开辟了道路；分析结果则仅在个人设备上显示。

这些能力的可用性，以及包含局部结构、功能性和动力学信息的数据集，意味着有必要开发数据分析工具，以提取系统行为的显著特征。传统上，此类分析基于经典的计算机视觉算法，包括局部和半局部滤波器、傅里叶变换、拉东变换和霍夫变换等[24]，需要大量手动开发，且往往形成优雅但高度个性化的工作流程。图像领域监督式深度学习方法的兴起[25]提供了另一种途径，并使语义分割、实例分割等许多传统上复杂的问题变得简单。这只需将计算机成像和医学成像中的工作流程迁移至显微成像领域即可实现。集成方法的进步[26-30]还有助于应对分布外漂移效应，并开辟了将这些方法纳入显微镜实时运行的途径。[29,30]

然而，顾名思义，监督式方法仅适用于已知感兴趣现象的物理图景的情况。实际上，在许多应用中，科学家希望发现新的物理现象，例如电子显微图像和扫描隧道显微图像中对称性破缺的表现、纳米颗粒的变异性等。这些问题属于无监督物理发现问题，因此需要开发相应的算法。这些算法还应满足成像数据分析中的实际需求，即应对成像系统畸变、处理图像平面中的旋转不变性、支持完全无监督学习和半监督学习任务，并支持聚类与方差因素阐明的结合。

本文综述了变分自编码器（VAEs）的原理及应用，这是一种用于高维数据分析的强大方法。我们使用简单的玩具数据集讨论 VAE 背后的基本原理和直觉，并介绍其在有监督、半监督和无监督设置中的多个真实世界示例。我们特别强调 VAEs 与系统物理特性之间的联系。我们还进一步讨论不变自编码器，既用于发现材料结构的显著特征，也用于适应真实成像系统中不可避免的失真。最后，我们讨论包括条件变分自编码器、联合变分自编码器和半监督变分自编码器在内的变分自编码器架构，并讨论它们可应用的图像分析任务类型。所有

这些方法都作为 “VAE 教程” GitHub 仓库的一部分，并已公开发布，可通过以下地址访问：https://github.com/saimani5/VAE-tutorials。

2. 变分自编码器

变分自编码器由 Kingma[31,32] 和 Rezende[33] 几乎同时于2014年提出，是一类特殊的生成式概率模型。它们具有多种独特属性，使其特别适用于图像分析，尤其是基于物理的分析。下文将讨论变分自编码器的基本原理，并将其与经典自编码器进行比较；简要介绍变分自编码器之外的数学形式体系；并通过简单示例展示其行为。在本文中，我们希望强调，与许多机器学习（ML）工具一样，将变分自编码器应用于现实问题，需要培养通用直觉和领域特定直觉（它们会如何处理 YOUR 图像？）。因此，我们鼓励读者尝试相关笔记本，并将其用于自己的数据。

2.a. 自编码器简介

在讨论变分自编码器之前，我们先简要介绍简单自编码器（自编码器）的原理及其应用。一般而言，自编码器属于由编码器和解码器两部分组成的一类神经网络。编码器接收输入物体，例如图像或图像块，并将其依次通过一组卷积层或全连接层，这些层的维度逐步降低，直至瓶颈层或潜在层。在这一过程中，数据的维度从原始大小缩减为通常非常少量的潜变量。解码器接收潜在向量，并将其依次通过一组维度逐步增加的层，以重构原始数据。编码器和解码器的架构通常相对于瓶颈层呈对称结构，但并非必须如此。自编码器的训练过程旨在最小化输入数据与输出数据之间的差异，从而帮助自编码器有效地压缩并重构原始数据集。自编码器中的损失函数可表示为公式 1，其中 X 为输入图像，decoder(encoder(X)) 为重构图像，LF 为损失函数。

$$Loss = LF\left(X, decoder\left(encoder(X)\right)\right) \quad (1)$$

原则上，自编码器可以使用许多损失函数进行训练，包括但不限于交叉熵损失、平均绝对误差和均方误差。编码器和解码器的结构可以遵循任何常用的机器学习范式，例如全连接多层感知机、[34] 卷积网络、[35] 注意力层、[36] 以及图网络。[37] 架构的选择取决于输入数据的类型（例如光谱、图像或图表示）以及特定领域的要求。同样，编码器和解码器的结构与架构也可以不同。然而，输入数据和输出数据的维度保持不变；或者，我们也可以采用编码器-解码器架构，而不是自编码器，此时输出与输入不同（例如从图像到光谱[38]）。这些架构还广泛应用于建立结构-性质关系、定量结构-活性关系等方面，但不在本综述的讨论范围内。

乍看之下，自编码器似乎是一种非常违背直觉的应用——毕竟，它们试图重建数据本身。然而，自编码器真正的强大之处，在于能够将高维数据编码为低维向量，并从潜在向量解码还原数据。换句话说，它建立了（输入）→ (l_1 , …， l_n)和(l_1 ,…， l_n) → (输出)之间的关系，其中输入和输出理想情况下应尽可能接近。

这种转换带来了多项有用的性质和应用。首先，通过潜在向量对数据进行编码和解码，可以选择数据中的相关特征，并排除无关元素。为了对此进行演示，我们选用了 MNIST 数据集作为玩具数据集。[39] MNIST 数据集由介于 [0, 9] 之间的手写数字图像组成，通常是测试计算机视觉算法的首选数据集。在 MNIST 数据集上应用自编码器（AE）的结果如图 1 所示。原始输入在低维潜在空间中的编码可以通过两种不同的方式进行可视化。第一种方式是使用每个输入图像对应的潜在向量，在低维数据空间中以散点的形式绘制该图像。通过这种方法，可以根据输入图像在潜在空间中的相应位置，将所有输入图像表示在一张散点图中。此外，还可以使用类别标签为各个点着色，从而观察不同类别如何在潜在空间中形成聚类。下文将此类图称为潜在空间分布。

对于 MNIST 数据集，潜在空间维度设为 2，其对应的潜在空间分布如图1a所示，其中散点根据其类别 [0-9]进行着色。潜在空间分布表明，即使没有任何关于类别标签的先验信息，属于同一类别的图像也会在潜在空间中形成聚类。此外，共享某些共同特征的类别会彼此重叠。

在图 1b 中展示了可视化潜在空间构建过程的第二种方法。在这种方法中，训练阶段结束后会丢弃输入数据集。随后对训练好的潜在空间进行均匀采样，并将采样结果解码回输入空间。然后将这些解码后的图像绘制在潜在空间中各自对应的位置上。这种可视化潜在空间的方法称为解码潜在空间，或潜在表示；图 1b 展示了 MNIST 数据集的示例。解码潜在空间有助于我们在不使用输入数据集的情况下，直观地了解潜在空间的特征。

对于 MNIST 数据集，解码后的潜在空间展示了图像类别在潜在空间中的分布，以及不同图像类别特征的重叠情况。在自编码器中，潜在空间未经过正则化，因此，用于在每个潜在维度中均匀采样潜在空间的上下限，是根据潜在分布获得的。对于 MNIST 情况，解码潜在空间的上下限取自图 1a 中的潜在分布，并将两个潜在维度的范围均设为 [-100, 100] 。为使图示清晰易读，解码潜在空间中的坐标轴标签在某些位置被省略。这些解码后的潜在空间旨在展示输入在潜在空间中的变化。此外，由于潜在空间采用均匀采样，潜变量的绝对值不会带来额外的理解。

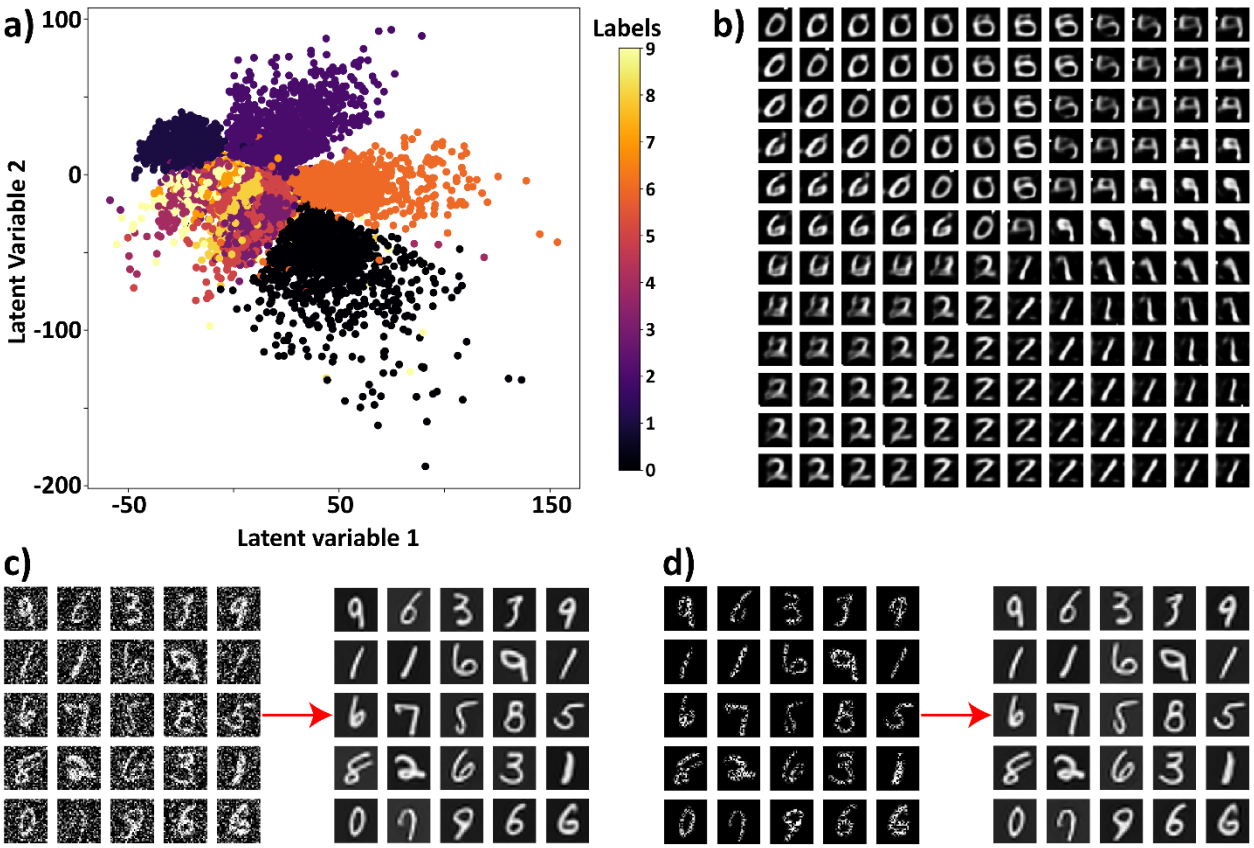


图 1。（a）在二维潜在空间中表示的 MNIST 数据集，并根据类别标签着色，（b）均匀采样并解码的潜在空间；（c）使用自编码器对 MNIST 数据集进行去噪的表示，其中展示了输入自编码器的带噪输入及其对应的清晰图像；（d）使用自编码器在 MNIST 数据集上填补缺失像素的应用表示，其中展示了含缺失像素的输入及自编码器生成的重构输出。

自编码器可广泛用于去噪和数据重建。对于这些应用以及许多其他应用，关键在于选择输入和输出。例如，在去噪任务中，将含噪数据集作为输入，并在重建数据集（输出）与干净数据集之间计算公式 1 中的损失函数。如果我们事先能够获得无噪声的（干净）数据集，该方法便能奏效。克服这一限制的一种方法是利用领域专业知识创建合成数据集，使其能够尽可能准确地代表当前数据集。随后，可以向该数据集添加噪声，并在合成数据集上训练自编码器。训练好的网络随后可用于清理原始数据集。我们通过向 MNIST 数据集添加噪声，并使用自编码器架构清理含噪图像，展示了这一方法。该过程的输入（含噪）和输出（干净）如图 1c 所示。

自编码器的另一项应用是图像重建，即修补缺失像素或恢复颜色。在缺失像素的情况下，自编码器的输入是不完整数据集，输出则是经过清理且完整的数据集。MNIST 数据集的缺失像素输入及其相应的填充输出如图 1d 所示。构成图 1 的分析及其代码可在随手稿附带的 Jupyter 笔记本“Intro to AE and VAE.ipynb”中找到。每个部分都附有一个这样的笔记本，用户可以将分析中使用的数据集下载到笔记本中。读者可以复现与手稿中讨论的结果相对应的完整分析。对于架构中耗时或占用内存较多的权重，手稿中会直接加载这些权重，以加快处理过程。

还要特别指出，这一过程也隐含地揭示了自编码器的局限性——缺失的信息并非凭空而来，而是源自原始数据集中各项特征之间的相关性。如果训练数据集中形状像汽车的物体都是红色的，那么重建数据集中的汽车即使经过重新着色，也仍会是红色。然而，如果数据集是在亨利·福特时代采集的（“你可以买任何颜色的汽车，只要它是黑色的”），那么重建结果就会产生误导。这是将神经网络用于重建和超分辨率问题时一个非常显著的局限性；本文还会反复指出这一局限性及其他局限性。

2.b. 变分自编码器简介

在上述对自编码器的阐释基础上，我们现在讨论本文综述的主题——变分自编码器。变分自编码器的编码器和解码器部分与自编码器相似，显著区别在于潜在层的结构。在这种情况下，变分自编码器并非让数据严格通过潜在瓶颈，而是对潜在空间进行采样。Kingma 最近发表的文章对变分自编码器的原理和特性作了精彩介绍。[32]

变分自编码器与自编码器具有相同的属性，但还能够解耦数据的潜在表示。尽管对这一过程的正式定义仍存在争议，但通常认为，它涉及将数据中的不同变化因素与不同潜变量相对应。变分自编码器因其生成能力而广受认可，能够生成与训练数据分布相似的新数据样本，这些样本此前未曾出现

，因此成为无监督学习领域中的强大工具。通过对位于两个输入数据样本的潜在表示之间的潜在点进行采样，变分自编码器可以生成包含两个输入数据点特征、但不属于输入数据集的新数据样本。这些新生成的样本可以看作是在潜在空间中对输入数据样本进行平滑插值的结果，并能为理解输入数据分布的结构提供有价值的见解。

用于阐释自编码器的 MNIST 数据集在此也用于说明传统变分自编码器构造的能力与不足。我们使用解码潜在空间和潜在分布作为分析变分自编码器结果的可视化工具。使用 MNIST 数据集训练的变分自编码器的解码潜在空间和潜在分布分别如图 2a 和图 2e 所示。在未另行说明的情况下，本文其余部分对变分自编码器及其变体的解码潜在空间进行绘图时，范围将设为 [-1.5、1.5] 。由于潜变量采用了先验分布，这些范围是可行的。先验分布通常被视为具有单位方差的独立的零均值正态分布，形式化部分将进一步讨论这一点。变分自编码器的解码潜在空间展示了数据集的重要特征如何编码到潜在空间中。例如，数据集中涉及的所有直线线段（“1”和“7”）都位于数据集的左侧，并逐渐向数据集右侧的圆形（“0”、“6”和“8”）转变。图 2e 所示的潜在分布展示了潜在空间是如何正则化的，即不同数据类别在潜在空间中紧密聚集，而对应不同数字的区域之间“非物理”结构的边界非常狭窄。这为变分自编码器的生成能力铺平了道路。

通过向数据集中引入额外的变异性（旋转 MNIST），可以揭示变分自编码器传统构建方式的不足。通常，我们选择这一特殊情形，是因为在成像中（与人类手写体不同），物体在图像平面内的方向事先未知。在旋转 MNIST 数据集中，每幅图像都旋转一个从范围 [-90°, 90°] 内随机采样的角度。所得数据集如图2c所示，并作为另一个变分自编码器网络的输入。在该旋转数据集上训练的变分自编码器，其解码潜在空间和潜在分布分别如图2c和图2f所示。解码潜在空间（图2c）不仅编码了未旋转数据集情形中讨论的特定特征，还编码了不同角度下的特征。图2f中的潜在分布按类别

标签进行着色，这些标签来自该数据集。由于某些类别存在旋转对称性，潜在空间中的类别标签现在变得模糊不清。潜在分布表明，变分自编码器对各类别的分离效果并不理想，因为看不到各类别形成清晰的簇。在此分析中增加潜变量的数量或许会有所帮助，但无法保证潜在空间能够解耦出生成图像时所涉及的角度。

为应对这一问题，我们使用旋转不变变分自编码器（rVAE）对数据集进行分析。这类变分自编码器确保其中一个潜变量是图像中物体的旋转角度。这种角度不变性（以及平移不变性）将在介绍 rVAE 的章节中进一步讨论。rVAE 的解码潜在空间和潜在分布分别如图 2d 和图 2g 所示。这里展示了 rVAE 在旋转 MNIST 数据集上的结果，因为在本节之后，我们将使用其他数据集来解释不同类别的变分自编码器的工作原理。rVAE 的解码潜在空间（图 2d）显示，所有数字都相对于水平轴以单一方向进行编码。潜在空间中的水平方向之所以如此，是因为在 rVAE 中，编码角度的零点是任意的。潜在分布（图 2g）显示，潜在空间分布中不同类别形成了多个簇，这表明与原始变分自编码器相比，rVAE 在分离不同类别方面表现更好。构成图 2 的分析及其代码可在随手稿附带的 Jupyter 笔记本“Intro to AE and VAE.ipynb”中找到。

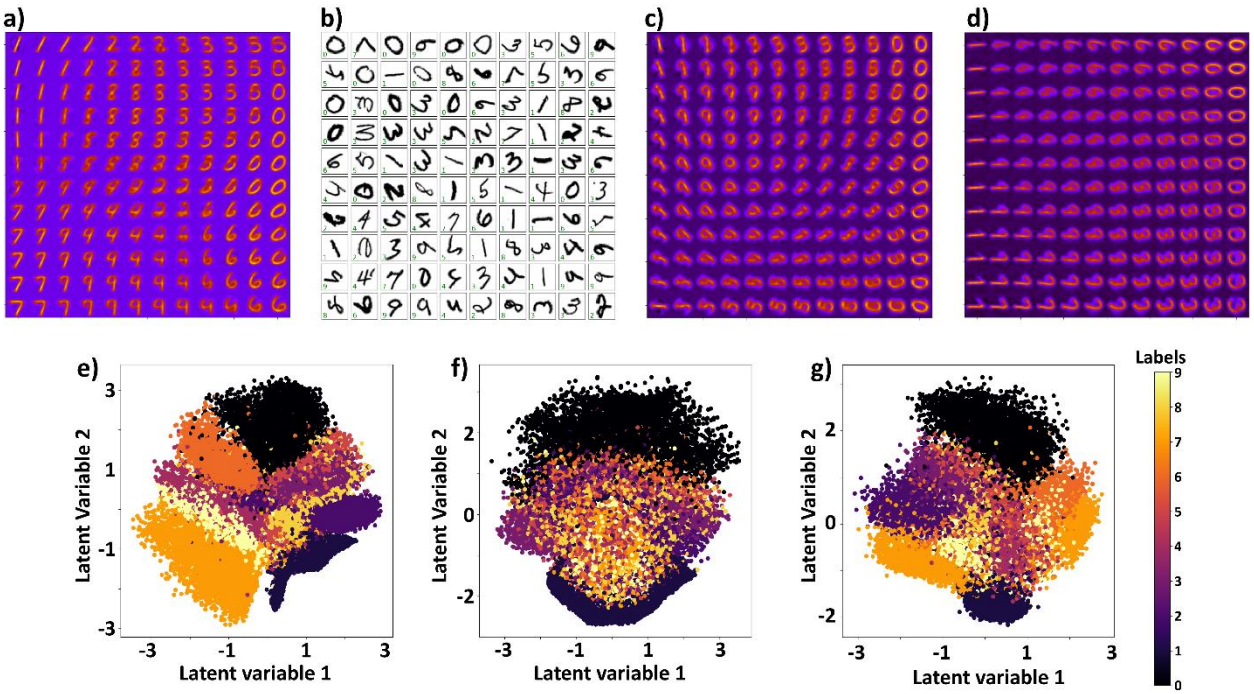


图2。（a）在 MNIST 数据集上训练的变分自编码器的解码潜在空间，除非另有说明，所有变分自编码器解码后的潜在空间中每个潜在维度的范围均为 [-1.5、1.5]（b）旋转 MNIST 数据集中的示例，其中每幅图像的旋转角度均从区间 [-90°、90°], 中随机采样；（c）在旋转 MNIST 数据集上训练时变分自编码器的解码潜在空间；（d）在旋转 MNIST 数据集上训练时 rVAE 的解码潜在空间。对于（e）在 MNIST 数据集上训练的变分自编码器、（f）在旋转 MNIST 数据集上训练的变分自编码器，以及（g）在旋转 MNIST 数据集上训练的 rVAE，使用图像的类别标签对潜在分布进行着色。（e）、（f）和（g）所用的色条相同，并显示于（g）中。

2.c. VAE的数学形式化

变分自编码器（VAE）是一种以无监督学习方式学习高维数据潜在空间的生成模型。VAE 由两部分组成：编码器 $q_\tau(z|x)$，也称为推断网络，其参数由 τ 确定；以及解码器 $p_\phi(x|z)$，也称为生成器网络，其参数由 ϕ 确定。后验分布 $p(z|x)$ 通过一种可处理分布进行近似，这种分布通常称为变分分布。在本文的分析中，变分推断采用高斯分布，其参数由编码器 $q_\tau(x|z)$ 预测。当向解码器提供一个潜在空间向量（z）时，它会输出

似然分布 $p(x|z)$ 的参数。VAE 的目标是学习一个紧凑且平滑的潜在空间，以便用于生成新的数据样本。

为实现这一目标，VAE 在潜在空间上引入先验分布 $p(z)$，通常为均值为零、方差为单位的标准高斯分布 $\mathcal{N}(0,I)$。该先验分布充当正则化器，确保潜在空间平滑且连续。编码器网络将输入数据映射为潜在空间中变分高斯分布的均值和标准差，以近似潜在空间。利用均值和标准差，可以从潜在空间中的高斯分布进行采样；VAE 因而能够从该分布采样，并将样本解码回原始数据空间，从而生成新的样本。

训练 VAE 时，需要最大化证据下界（ELBO）。VAE 的证据下界由两项组成：重构损失，以及学习到的潜在分布与先验分布之间的 KL散度。重构损失衡量输入数据与重构输出之间的差异，而 KL散度则促使学习到的潜在分布接近先验分布，从而确保潜在空间平滑且连续。VAE 通过最大化公式 2 所示的证据下界，学习最优的编码器网络和解码器网络。

	$ELBO_{vae} = -Reconstruction\ error - D_{KL}(q(z\|x)\|\|\mathcal{N}(0,I))$	(2)

总而言之，VAE 是一种生成模型，它通过在潜在空间上引入先验分布并最大化证据下界，学习平滑且连续的潜在空间。编码器将输入数据映射为潜在空间中高斯分布的均值和标准差，而解码器则将潜在空间映射为输出数据。VAE 通过最小化重构损失以及学习到的潜在分布与先验分布之间的 KL散度，或通过最大化证据下界进行训练。

2.d. 使用卡片的 VAE

在讨论了数学基础之后，我们现在开始测试变分自编码器在一个玩具数据集上的能力。从现在起，我们将使用该数据集评估变分自编码器的不同版本。为此，我们选择了扑克牌数据集，其中包含单色

图像，涵盖四种花色，即梅花、黑桃、方块和红桃。尽管扑克牌数据集十分简单，却构成了一组颇有意思的符号集合。例如，黑桃和梅花（不带尾部）具有三重对称性和镜像对称性。此外，水平翻转后的红桃和黑桃仅有细微差别，这使无监督算法难以区分它们。另一方面，方块的形状与其他三种花色截然不同。将方块花色旋转 90^0 相当于进行单轴压缩和缩放。这一现象使仿射变换应用于该数据集时的结果出现退化。扑克牌数据集在对称性方面具有这些复杂特征，因此对机器学习算法而言颇具挑战性；但与此同时，相关行为可以在二维潜在空间中直观地呈现，从而便于清晰地解释。

扑克牌数据集中的图像变异性，是通过对每张图像施加不同的仿射变换引入的。每张图像都会旋转一个角度；对于低旋转数据集，该角度从 [-12°，12°] 的范围内随机采样；对于高旋转数据集，则从 [-120°，120°] 的范围内随机采样。此外，图像会在 x 和 y 方向上以相同幅度进行剪切：低剪切情况下，剪切值从 [-1°，1°] 的范围内采样；高剪切情况下，则从 [-20°，20°] 的范围内采样。总共形成四个由这两种变换组合而成的数据集，分别为：i）低旋转、低剪切；ii）低旋转、高剪切；iii）高旋转、低剪切；iv）高旋转、低剪切。接下来的章节将使用这些数据集来分析不同类型的变分自编码器，为简洁起见，将它们称为扑克牌数据集（i-iv）。

每个数据集都训练了一个网络，图3展示了全部四个数据集对应的结果。图中分为4个部分，即（p-s），分别对应扑克牌数据集（i-iv）。图3a展示了从每个数据集中随机选取的25个示例。图3b展示了全部四个数据集的解码潜在空间。图3c展示了四个数据集的潜在分布，其中散点根据各数据点的原始类别标签着色。为简洁起见，潜在分布图省略了 x 轴和 y 轴的范围，因为这些图仅用于可视化潜在空间中的聚类。从潜在分布可以看出，变分自编码器能够出色地将图像编码到潜在空间中，且旋转幅度较小（图3cp和图3cq）。数据集中四个不同类别的簇及其间的边界清晰可见。从这两个数据集的解码潜在空间（图3bp和图3bq）也可以得出相同结论。

此外，由于变分自编码器的潜在空间具有平滑性，预计每个簇在潜在空间中的角度和剪切会平滑变化。在解码潜在空间中，每幅图像都明确属于四个类别之一，这也有助于我们得出结论：对于前两个数据集，变分自编码器能够在潜在空间中为所有类别生成最优编码。

后两个数据集（卡片 – iii 和 iv）则并非如此，这两个数据集的结果分别展示在图3r和图3s中。潜在空间分布（（图3cr和图3cs）显示，只有方块类别形成了清晰的聚类，而其他三个类别之间存在明显重叠。这种重叠源于高旋转情况下输入数据集中存在的不同对称性。就结构而言，方块与其他三个类别没有显著重叠。在解码后的潜在空间（图3br和图3bs）中也可以观察到这一点：除方块外，我们可以看到一些同时具有两个或更多类别特征的图像，而方块在两种情况下都占据了最上面的几行。这项分析揭示了变分自编码器在包含旋转的数据集上的不足，尤其是在存在旋转对称性的情况下。完整的分析可以通过随本文附带的“intro to AE and VAE.ipnyb”Jupyter 笔记本进行可视化和复现。

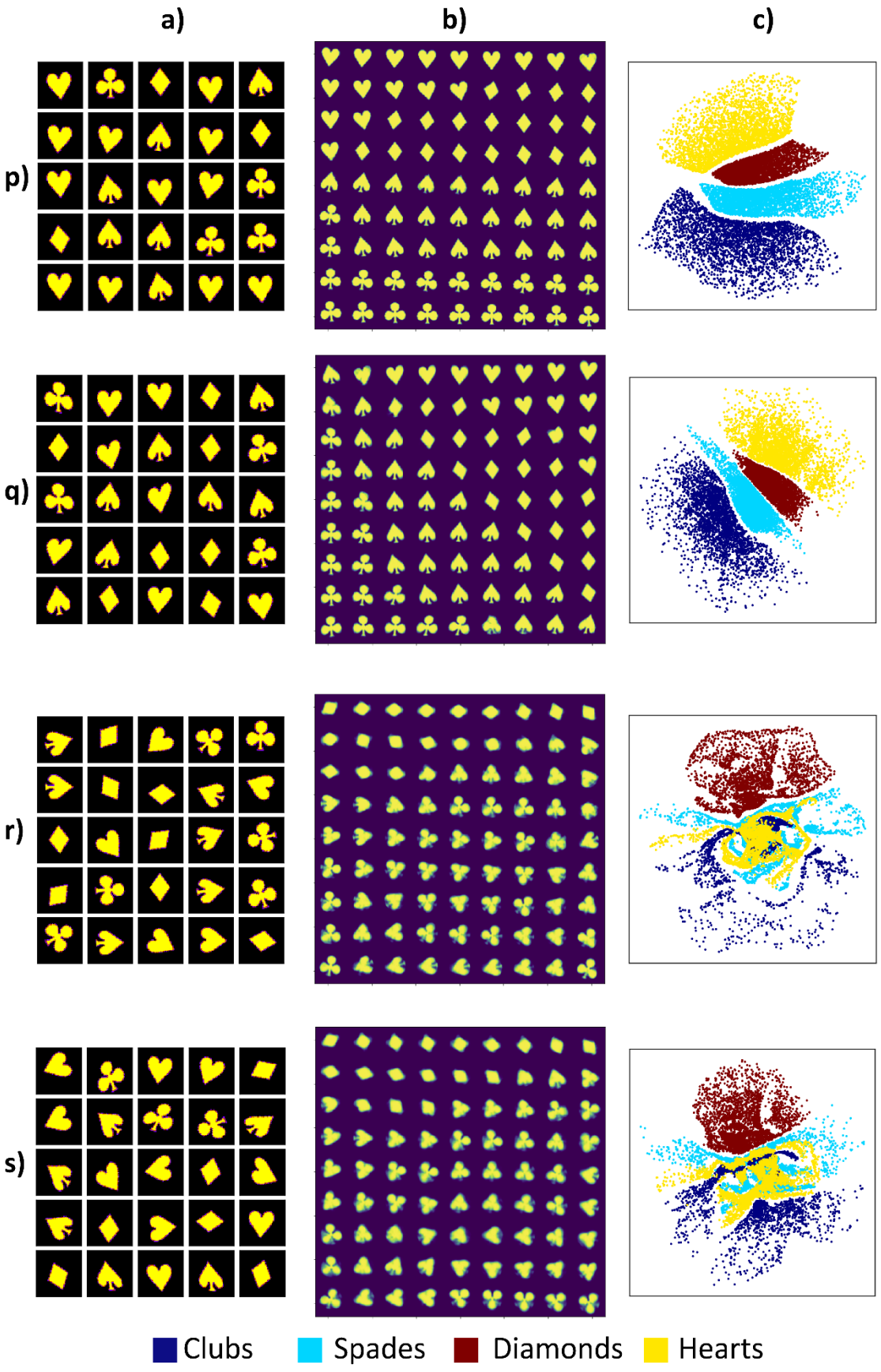

a)
b)
c)
p)
q)
r)
s)
Clubs
Spades
Diamonds
Hearts

图3。图中的各行，即（p-s），分别对应于通过对卡片图像应用变换而形成的不同扑克牌数据集（i-iv）。（a）从每个数据集中随机选取的25幅图像，用于展示各数据集的细节；（b）每个数据集的解码潜在空间；以及（c）每个数据集在潜在空间中的潜在分布，其中散点根据原始类别标签着色。

3. 选择描述符

将变分自编码器用于成像数据的关键在于描述符的选择。对于光谱成像数据 R(x, y, t)，其中 x 和 y 是坐标。自然的对象是在位置 [x, y], 处的光谱数据集 R(t)，其中 t 是一个描述符，可以是从光谱中提取的物理参数。对于成像数据，选择描述符有多种策略。其中一种是基于领域专业知识，从数据集中人工挑选物理定量描述符。为阐明这一过程，我们考察图 4a-d 中两个不同时间 t_0 和 t_1 的压电响应力显微镜（PFM）图像数据，该数据展示了 $PbTiO_3$ (PTO) 的铁电畴。图 4a 和 4c 分别展示了这两个时刻的 PFM 振幅响应。利用深度卷积神经网络对相应的畴壁进行语义分割，结果分别如图 4b 和 4d 所示。在此例中，物理描述符可以是：（1）某个区域内的极化状态，例如图 4a 和 4c 中亮区 P_1 与暗区 P_2 的比例；（2）畴壁数量，例如图 4b 和 4d 中分别有两条畴壁 W_1 和 W_2；（3）两条最近邻畴壁之间的距离，例如图 4b 和 4d 中的 d_1 和 d_2。从图像数据集中选取定量信息的另一种方式，是利用原子分辨率扫描透射电子显微镜图像中原子的邻域。不过，这种方法的效率取决于能否以机器学习技术可用的格式呈现邻域的全部信息。分子图、[40] SMILES、[41]SELFIES[42] 都是一些广为人知的技术，可将邻域信息编码为机器学习/深度学习技术能够处理的格式。人工挑选描述符或编码邻域这些方法往往无法涵盖数据集中的全部信息，和/或会造成非预期偏差。

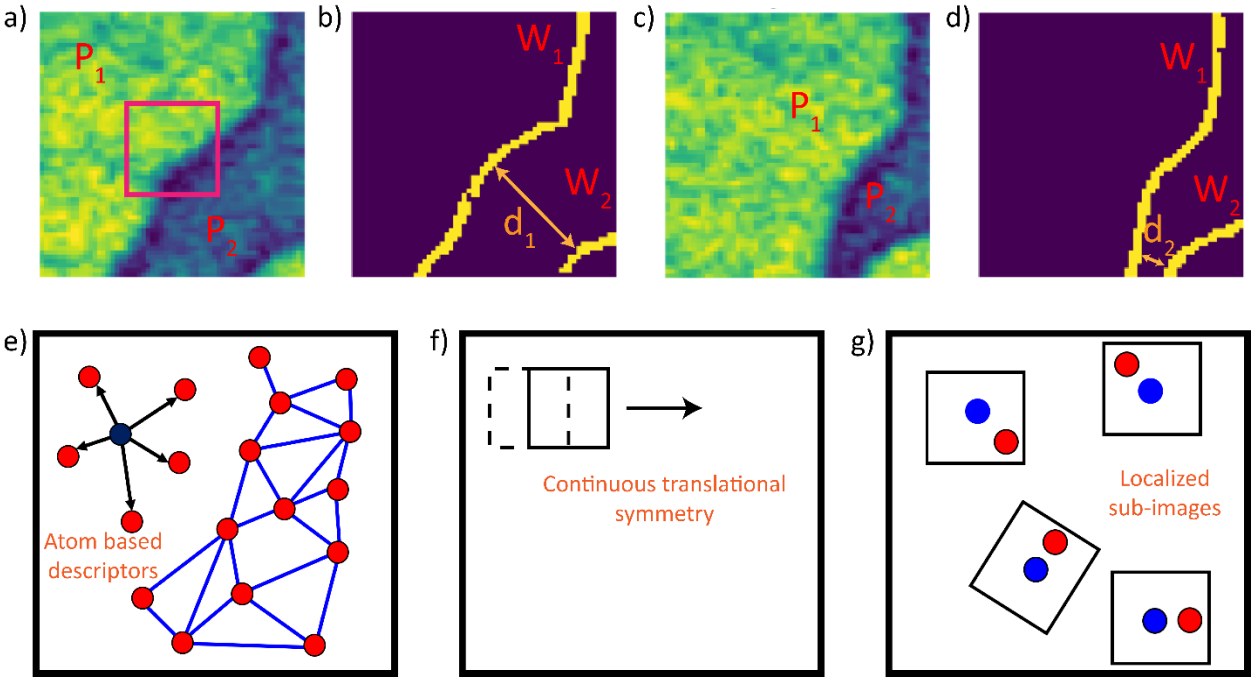


图4。（a-d）铁电PTO薄膜的压电响应力显微镜图像，改编自参考文献[43]。时刻（a）t_0和（c）t_1, 的PFM振幅图像，以及由深度卷积神经网络生成的、分别对应时刻（b）t_0和（d）t_1的畴壁图像。示意图展示了如何选择原子分辨图像中的原子邻域作为描述符。VAE图像分析中的特征工程包括图像块的选择。这些图像块可以基于矩形采样网格（f）（滑动窗口方法），也可以以单个特征（例如原子（g）或畴壁（a中的红色方框））为中心。对于动态数据，特征可以以延迟后的图像为中心。最后，可以根据物理问题和成像过程的物理规律，在有无旋转、平移或其他对称性的情况下对这些图像块进行分析。

一种更简单的系统描述符选择方式，是在原始图像的方格网上构建以网格点为中心的子图像。方格网的密度决定了从给定图像中采样的子图像数量。图 4f 展示了该方法的示意图。该方法能够避免在构建描述符时引入任何用户特定偏差。构建这些子图像的过程很简单。这种方法已用于此前的多项研究[44-46] 然而，对于包含感兴趣对象的图像，图块中心与对象之间的关系可能是任意的。因此，这些随机平移会成为数据中的主要变化因素。例如，图 4a-d 展示了铁电PTO薄膜的 PFM 数据；在 PFM测量期间，通过探针施加直流偏压以切换极化。极化状态变化和畴壁运动可通过连续PFM图像扫描记录。图 4a 展示了该数据在时间 t_0 时的一帧

连续 PFM 数据集。可以从该数据集中提取子图像，图 4a 展示了一个对应于红色方框标记区域的示例。由于在施加直流偏压时，极化状态和畴壁会发生变化，因此经过一段时间（例如，在时间 t_1），可以观察到极化状态变化（图 4c）和畴壁偏移（图 4d）。如下文所述，平移不变变分自编码器[47] 方法可以部分抵消这种影响。然而，当物体形状因与图像块边缘重叠等原因发生变化时，平移变分自编码器会受到限制。

构建子图像的第二种方法将图像块生成与基于物理的描述相结合。对于扫描透射电子显微镜数据，以单个原子为中心定义图像块是很自然的做法（图 4g）。在这种情况下，图像块集合包含图像中的完整信息，但通过这种构建方式分析的是单个原子的邻域。这进而需要引入旋转不变自编码器，因为否则物体在图像平面中可能具有任意方向。

最后，对于时间依赖的动态数据，可以构建包含时间信息的描述符。例如，在分析铁电畴壁运动时，可以将图像块的中心设在延时图像上。这样，描述符便会包含时间动力学信息。[43,48]

4. 构建不变变分自编码器

上述变分自编码器的示例阐释了变分自编码器的原理，以及在阐明感兴趣现象时描述符的选择。然而，将变分自编码器应用于图像分析还伴随着额外的挑战。如上文所述，我们最终旨在发现数据中的变化因素，或解耦数据表征。然而，在真实图像中，许多变化因素与感兴趣对象的位置、旋转或尺度的小幅变化有关。例如，在从显微图像中识别纳米颗粒形状时，我们通常假设纳米颗粒除图像平面内的旋转外应保持相同。同时，在选择描述符时确定锚点所存在的不确定性，会导致图像中出现平移偏移。这些考虑促成了不变变分自编码器的发展；不变变分自编码器通常指这样一种网络架构：将变化的物理因素（平移、旋转、尺度）分别分离到不同的潜变量中，并解耦剩余自由度中的变化因素。

4.a. 引入不变性

本节将探讨旋转不变自编码器的构造和工作原理。为简洁起见，以下将其称为 rVAE。在传统的变分自编码器设置中，潜变量包含输入数据集的基本特征，通常不受用户控制。例如，在 MNIST 数据集中，输入数据点的书写风格和类别（0-9）是潜在空间所学习到的主导特征。图像中物体的旋转和平移会被视为数据集的特征。潜在空间会将它们与其他特征一样作为输入变化来学习。例如，MNIST 数据集中的单个数字可以被随机旋转和平移，并可在此数据集上训练变分自编码器。在这种情况下，角度和平移预计会被编码到平滑变化的潜在空间中。在这一过程中，变分自编码器会将具有不同旋转和平移变换的同一物体视为不同实体，并为其生成不同的编码。rVAE 网络的设计使得 1-3 个潜在维度能够显式编码输入图像中存在的旋转和/或平移。通过在 rVAE 网络中同时引入旋转不变性和平移不变性，该网络能够在潜在空间中显式编码旋转信息和平移信息，从而为输入数据提供更稳健、更准确的表示。例如，具有相似物体但方向不同的图像，其编码在潜在空间中将保持相似，唯一区别在于显式编码角度的维度中的编码。

旋转和平移是我们日常生活中遇到的图像以及许多科学应用中的两个重要方面。在材料科学和凝聚态物理等领域，rVAE 网络同时捕捉旋转信息和平移信息的能力，对于分析显微图像尤其有用。例如，通过使用 rVAE 分析显微图像，研究人员可以更好地理解其中潜在的旋转模式和平移模式，以及它们对材料性质的影响。同样，在凝聚态物理中，旋转和平移对于理解原子和分子尺度上粒子的行为发挥着关键作用，而 rVAE 可用于分析和生成这些系统的多样化图像，从而深入了解其旋转动力学和平移动力学。总体而言，在 rVAE 网络中同时引入旋转不变性和平移不变性，为分析和生成各种科学应用中的多样化图像提供了强大工具，

使研究人员能够更好地理解其数据中的旋转模式和平移模式及其动力学。

rVAE 的编码器部分与传统 VAE 类似，其中全连接稠密神经网络（DNN）或卷积神经网络（CNN）会逐步降低数据集的维度，直至达到一个含有 k 个维度的瓶颈层。我们将包括 rVAE 在内的不同变体 VAE 的这些潜在维度称为 [z_1, z_2, ...、z_k]。这些潜在维度捕捉输入数据集中的变化，并将其平滑地编码到潜在空间中。根据用于训练的 rVAE 类型，rVAE 的瓶颈层还会串接 1–3 个额外特征，分别对应角度（θ）、x 轴平移（Δx）和 y 轴平移（Δy）。因此，rVAE 的潜在空间维度为 k+3 ，以实现旋转不变性和平移不变性，其中 k 是用户选定的潜变量数量。rVAE 使用一种特殊的解码器，称为空间生成器网络，该网络由 Bepler 等人于 2019年[49] 提出。空间生成器解码网络使用输入图像的坐标以及编码器的低维潜在表示来解码图像。首先，我们构建一个二维数组，其中包含输入图像所涉及像素的坐标。随后，利用与旋转和平移不变性对应的潜在特征（θ、 Δx、Δy）对二维坐标数组进行旋转和平移，这一变换由公式 3 给出，

	$x_t = x * \begin{bmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{bmatrix} + [\Delta x, \Delta y]$	(3)

其中，x 和 x_t 分别是原始坐标和变换后坐标， θ, Δx、 Δy 是 rVAE 的旋转和平移编码。

解码器的第一层将潜在空间编码 [z_1, z_2, z_3 ... z_k] 映射到一个更高维的空间。随后，使用另一个不同的神经网络将变换后坐标投影到这个更高维空间中。经过第一层后，将高维 z 向量复制 N 份，其中 N 是输入图像中的坐标或像素数量，并将每个高维坐标向量分别加到这些 z 向量副本上。完成此操作后，每个像素都对应一个高维编码。接着，我们像在 VAE 中那样，使用全连接 DNN 或 CNN 对这些编码进行解码，其中输出层的神经元/滤波器数量为一。通过将 z 向量复制 N 份、将高维坐标分别加到每个副本上，

再将它们解码回实空间，所得解码结果可以视为每个像素上的条件概率，其中每个解码器神经元均以变换后坐标为条件。图5a 展示了 rVAE 的示意图。通过最大化公式 4 给出的证据下界，可以学习 rVAE 编码器和解码器的参数。

	$ELBO_{rvae} = -Reconstruction\ error - D_{KL}(q(z\|x)\|\|\mathcal{N}(0,I))$ $- D_{KL}(q(\theta\|x)\|\|M(0,\kappa)) - D_{KL}(q(\Delta x\|x)\|\|\mathcal{N}(0,\sigma_s))$	(4)

其中，重构误差是重构图像与原始图像之间的误差，可以使用均方误差、交叉熵等损失函数进行计算；第二项表示潜在表示的后验分布与先验分布（$\mathcal{N}(0,1)$）之间的 KL散度；第三项表示角度的后验分布与参数为 κ 的 θ 上的冯·米塞斯先验之间的 KL散度；最后一项表示平移的后验分布与其标准差为σ_s的正态分布先验之间的 KL散度。这些潜变量先验分布中涉及的参数是网络的超参数，而关于输入数据集的领域知识有助于我们对其进行选择。整个架构的权重通过对证据下界$_{rvae,}$执行梯度上升来学习，正如 VAE 部分所述。

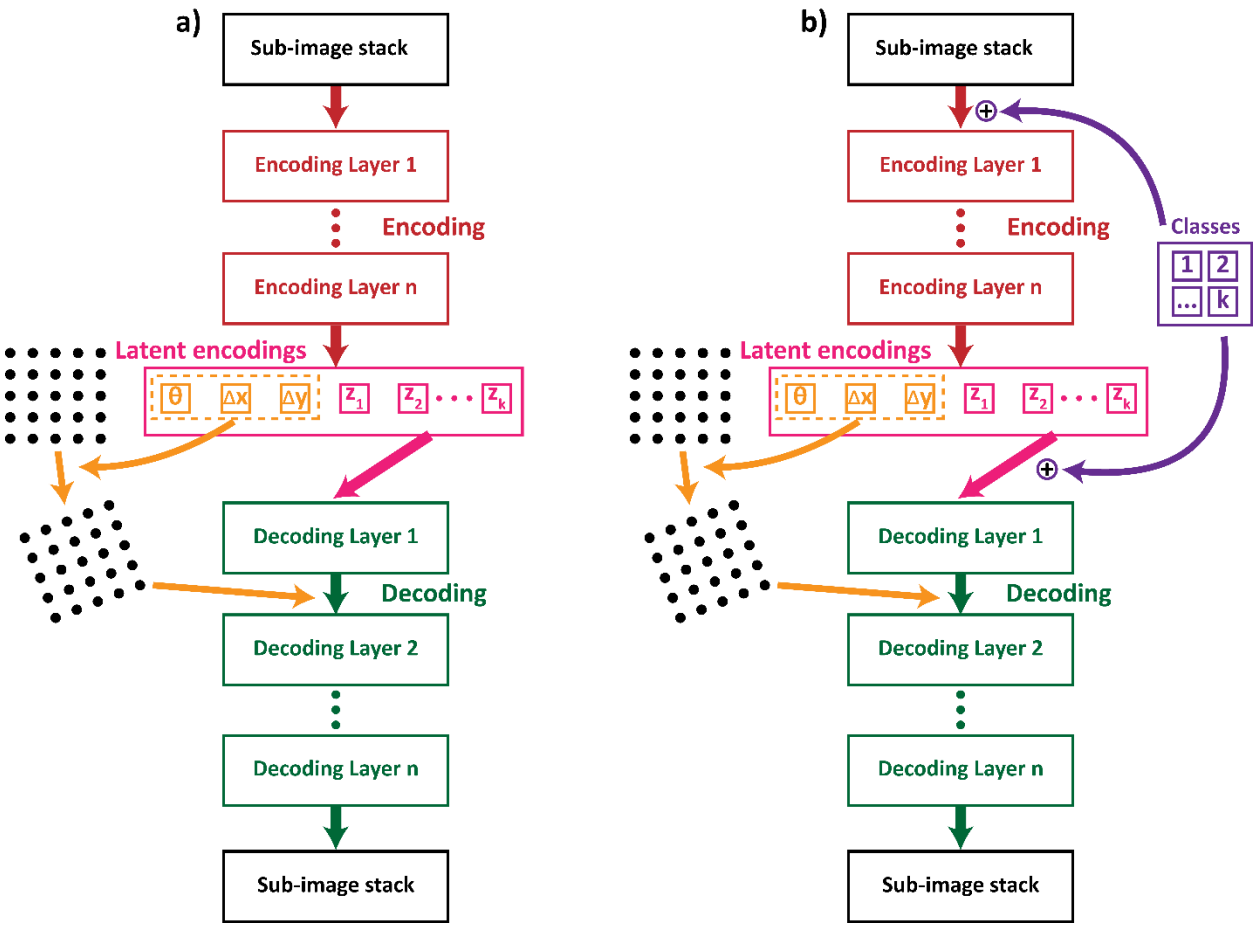


图5。（a）旋转不变变分自编码器（rVAE）和（b）条件旋转不变变分自编码器（crVAE）的工作示意图。

4.b. 卡片数据集上的 rVAE

为了探索 rVAE 的应用，我们使用前文 VAE 部分讨论的扑克牌数据集 i-iv。这些数据集由属于四种花色的原始四张卡片构成，即黑桃、方块、梅花和红桃，并通过对其进行随机旋转和剪切得到。旋转和剪切的随机值分别取自区间 [-α、 α] 和 [-s，其中 s], 。四个数据集的 α 和 s 的具体取值在 VAE 部分中进行了讨论。每个数据集分别训练一个旋转不变网络，编码器和解码器均包含 2 个隐藏层。低剪切情况下，即扑克牌数据集 i 和 iii，在隐藏层中使用 256 个神经元；高剪切情况下，即扑克牌数据集 ii 和 iv，在隐藏层中使用 512 个神经元。本文始终采用对称的编码器-解码器架构。预计 rVAE 能够显式解耦旋转特征，并形成一个平滑变化的潜在空间，用于编码其他变化

在数据集中，四个示例的剪切角均具有不同的范围。前两个旋转幅度较小的情况较为简单，因此手稿中未作讨论；相关内容可在随手稿附带的 Jupyter 笔记本中找到。此外，简单的变分自编码器已足以处理这两个数据集。潜在空间呈现出卡片的平滑变化，并且所有卡片都被分配到潜在空间的不同区域。后两个高剪切值案例的结果在图6中进行了讨论。

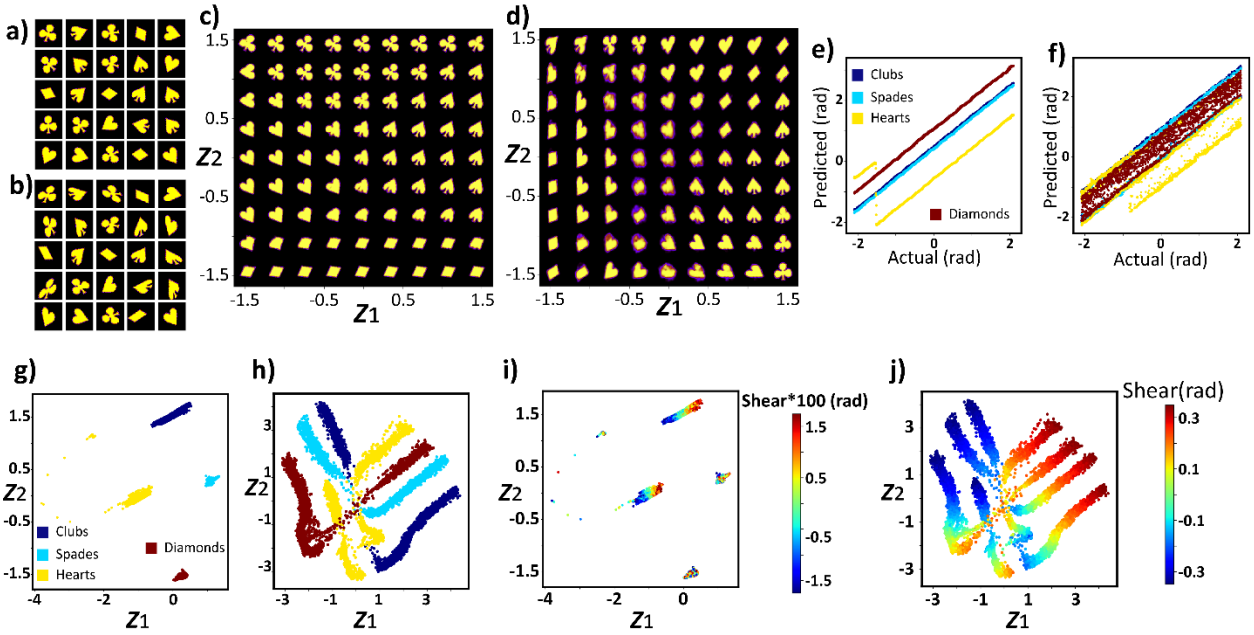


图6。随机采样的25幅图像构成了（a）卡片数据集 iii 和（b）卡片数据集 iv；解码潜在空间分别对应于（c）卡片数据集 iii 和（d）卡片数据集 iv；（e）卡片数据集 iii 和（f）卡片数据集 iv 的编码角度与真实角度，其中各点使用类别标签着色；（g）卡片数据集 iii 和（h）卡片数据集 iv 的潜在分布，其中各点使用类别标签着色；潜在空间中四个类别对应的真实剪切值变化分别如（i）卡片数据集 iii 和（j）卡片数据集 iv 所示。

图6a和图6b分别展示了应用于输入图像的两个高剪切数据集中的示例图像：低旋转（卡片数据集 iii）和高旋转（卡片数据集 iv）。图6c和图6d分别展示了训练得到的两个不同 rVAE 的解码潜在空间。对于与不变变分自编码器对应的解码后的潜在空间，将捕获角度和平移的潜变量设为零。这使我们能够了解来自每个

类别的图像在潜在空间中的编码方式，而不包含旋转和平移变换。解码角度的基准零点取决于输入图像的类别标签。

从解码后的潜在空间中也可以观察到这一点：每个类别在潜在空间中都有不同的方向。图6c和图6d分别绘出了数据集iii和iv中预测旋转角度与生成输入数据集时实际使用的角度。点的颜色对应于每张输入图像的类别标签。最后，图6g至6j展示了这两种情况下的潜在分布。在图6g和图6h中，潜在分布根据原始类别标签着色；在图6i和图6j中，则根据两个数据集各自的剪切值着色。潜在分布表明，rVAE能够区分低剪切数据集（卡片数据集 iii）中的四个不同类别，并在潜在空间中形成簇。

解码潜在空间中出现模糊图像，是因为给定的点位于输入数据集中的点之间。不同类别的数据点在潜在分布图中并未重叠。只有红桃花色被识别为两个不同的簇。其原因可以从解码潜在空间（图6c）左侧（$z_1 = -1.5$）看出：红桃花色被解码为两种不同的旋转方向。其中一种解码后的红桃花色与上方的梅花相似，另一种则与黑桃相似，可在解码潜在空间的中部找到。 从角度图（图6e）也可以推断出这一点，其中红桃花色对应两条斜率不同的线。从该图还可以看出，不同类别的编码角度基线零点各不相同。最后，图6i显示，对于每个类别，真实剪切角在潜在空间中平滑变化。卡片数据集 iii 中则偏离了这一规律。这可能是因为对于该数据集，施加剪切的范围相较于旋转角度较小。

对于高旋转和高剪切数据集，rVAE 仍优于原始变分自编码器，因为它在潜在空间中形成了簇（图6h）。尽管潜在空间中形成的簇在不同类别之间存在重叠，但这种重叠程度远小于原始变分自编码器的情况。在 rVAE 中，红桃花色与其他三种花色均有重叠。在解码潜在空间的中心区域，四种花色似乎都发生了重叠。请注意，不同类别对应的流形不能相互重叠（如果物体不同），因此会形成特征性的“交叉”图案。

在如此高的旋转和剪切下，类别之间的结构差异非常小。解码潜在空间（图6d）也显示，四种花色之间存在显著重叠

在此数据集中，导致该图中出现多个模糊图像。它还展示了来自一个类别的图像如何平滑地转变为来自另一个不同类别的图像。解码角度 与 真实角度的对比图（图6f）显示了预测角度时存在歧义。数据集中的内在对称性以及高剪切使图像无法辨识。 真实剪切图显示，即使在高剪切变换下，每个类别的图像属性仍然平滑变化。尽管该数据集的编码优于原始变分自编码器，但由于发生重叠，潜在空间中的花色分类仍然存在歧义。这是因为所应用的变换完全使图像变形，即使是 rVAE 也无法仅使用未标注数据对输入图像进行分类。所有四种情况的完整分析和结果图均可在仓库中名为“rvae_cards.ipnyb”的 Jupyter 笔记本中找到，该笔记本与手稿一同提供。

4.d. 实验数据集上的 rVAE

为了解 rVAE 在实验数据集上的应用，我们选择了电子束激活操控的石墨烯中硅原子的扫描透射电子显微镜图像。[50-54] 该数据集由电子束操控过程中石墨烯系统的 50 张快照组成。石墨烯系统既包含晶态区域，也包含非晶区域，且各个原子周围存在若干点缺陷。图 7a 显示了数据集第零张快照中的碳原子和硅原子（亮点）。为初步获取用于 rVAE 分析的特征向量，我们将整个数据集输入预训练的深度卷积神经网络（DCNN），以获得原子坐标。有关该 DCNN 的详情及工作原理，详见以下参考文献[43,55]。该 DCNN 是一种语义分割网络，可将图像中的每个像素分类为以下三种类别之一即碳原子、硅原子和背景。图 7b 展示了该图像经过 DCNN 处理后的输出，其中三种类别分别以不同颜色表示。随后，通过计算构成每个原子的像素质心，得到原子坐标。图 7c 显示了这些原子坐标， 其中碳原子以蓝点表示，硅原子以亮红点表示。这些原子坐标叠加绘制在原始扫描透射电子显微镜图像上。如前所述，通过裁剪这些原子周围一定窗口大小的图像，构建已识别原子的特征向量数据集。窗口大小的选取以能够容纳中心原子周围三个六边形石墨烯环为准。需要注意的是，后续分析对用于生成这些特征向量的窗口大小

敏感。为所有快照中已识别的原子构建特征向量，并将其作为 rVAE 网络的输入。rVAE 网络的编码器网络和解码器网络各包含两个隐藏层，每层有 512 个神经元。我们使用具有两个用户指定潜在维度（z_1 和 z_2）的旋转和平移不变 VAE 版本。该版本的 rVAE 总共有 5 个潜在维度：即，一个用于编码角度的维度（θ）、两个用于编码 x 和 y 方向平移的维度（Δx 和 Δy），以及两个用户指定的潜在维度（z_1 和 z_2），用于编码数据集中每个原子的邻域。

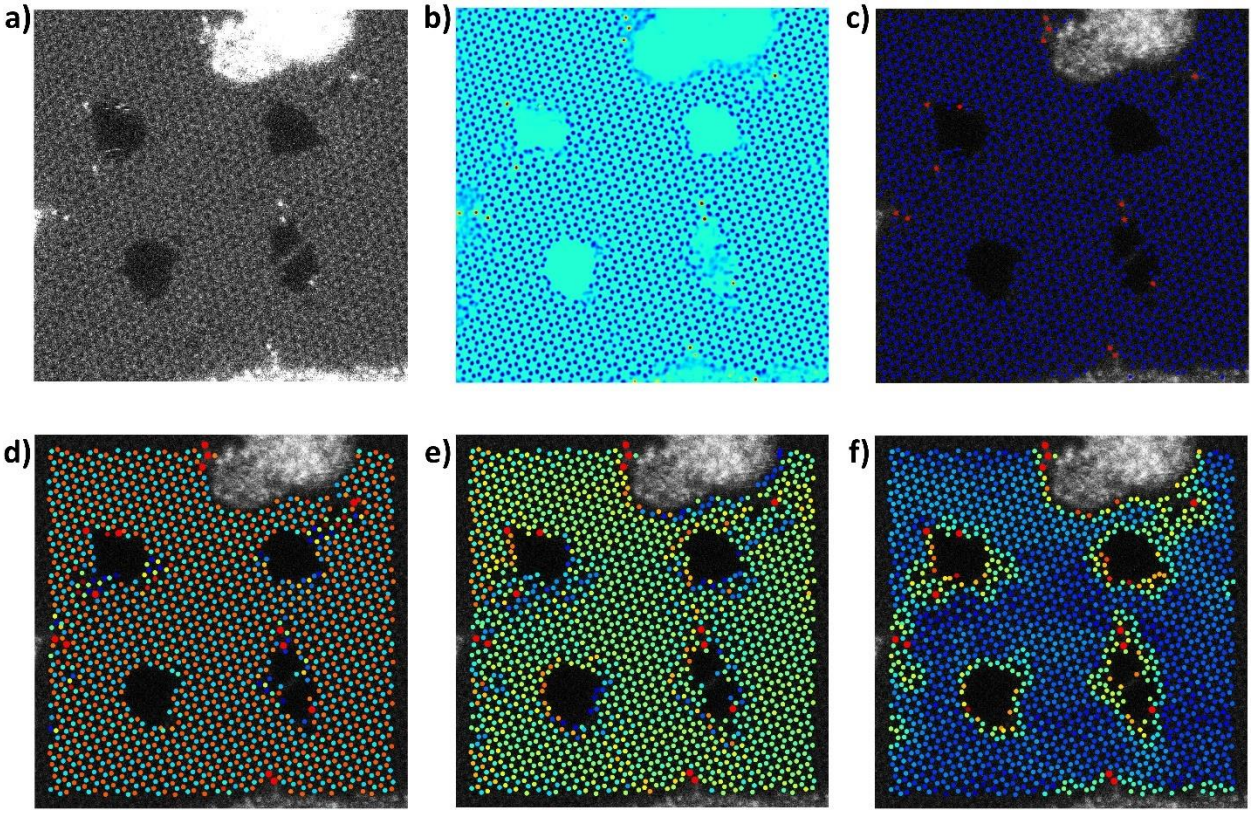


图 7. (a) 石墨烯数据集第零帧的原始扫描透射电子显微镜图像；(b) 深度卷积神经网络的语义分割输出，清晰地区分了两类原子与背景类别；(c) 根据深度卷积神经网络输出中原子像素簇的质心获得的原子坐标。第零帧中的每个原子均根据 rVAE 网络的以下编码着色：(d) 角度编码 (θ)、(e) 第一潜变量 (z_1) 和 (f) 第二潜变量 (z_2)。

在 rVAE 网络的训练阶段之后，图 7d-f 重新绘制了第零帧，此时原子坐标使用特征向量的编码进行着色。编码角度、第一潜变量（z_1）和第二潜变量（z_2）分别用于绘制图 7d、7e 和 7f 中原子坐标的颜色。对于理想石墨烯层，中心原子周围有两个相差 60° 的对称晶胞。rVAE 能够识别这两个结构的最佳旋转编码，如图 7d 中对应两个旋转变体的棋盘格图案所示。图 7f 中的第二潜变量（z_2）编码了其接近非晶区域的程度。当原子的裁剪图像中存在非晶区域时，该潜变量的值会发生变化；从非晶区域附近该潜变量颜色的骤变可以推断出这一点。第一潜变量（z_1）似乎对特征向量中图像的其他变异性形式进行编码，例如缺陷和应变。编码的绝对值无助于理解 rVAE 的工作原理，因此为简洁起见，省略了图 7d-f 的色标。

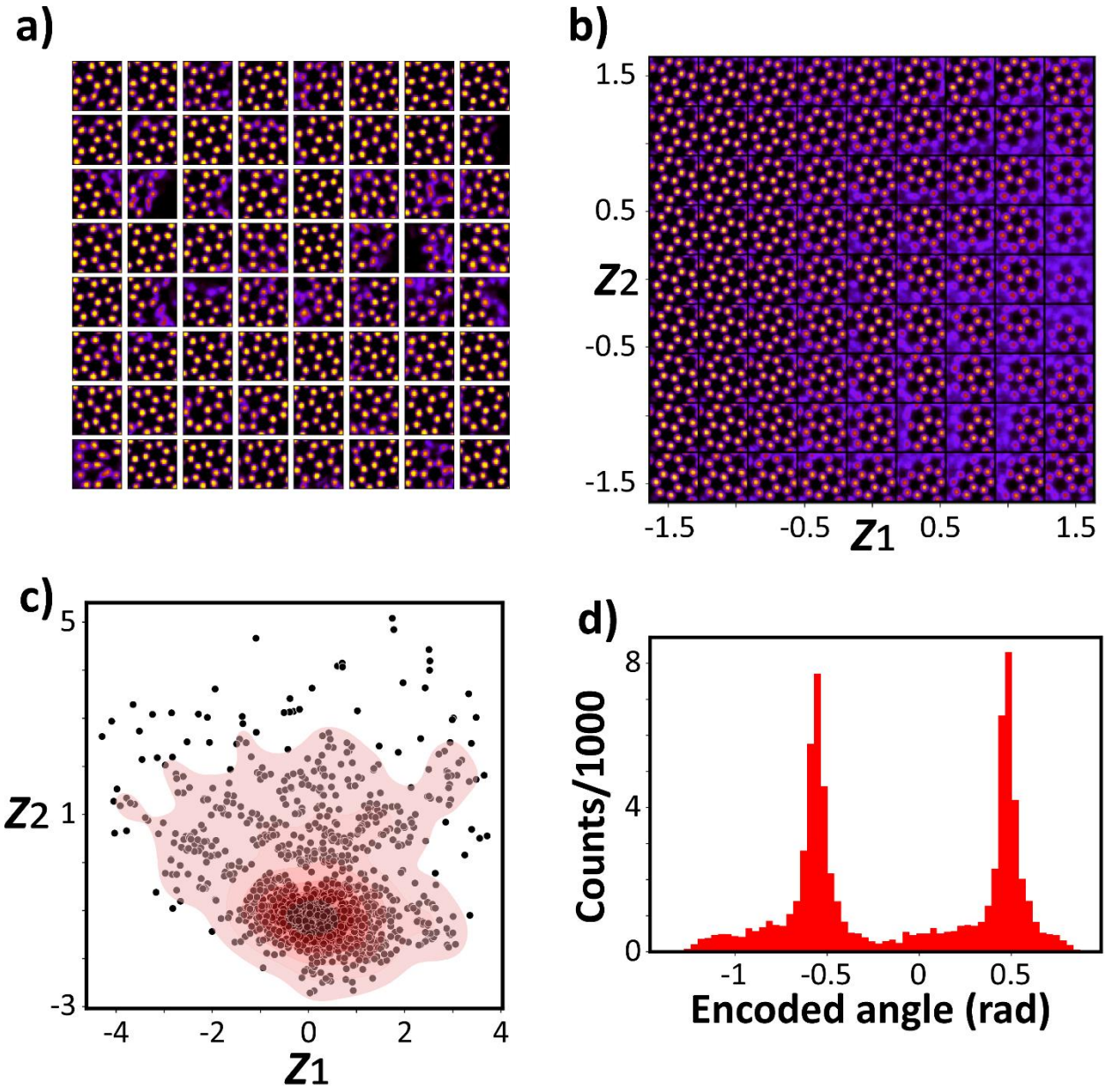


图 8。（a）从 rVAE 网络的输入中随机选取的 64 幅图像；（b）在石墨烯数据上训练的 rVAE 网络的解码潜在空间；（c）潜在分布的核密度图；（d）角度编码（θ）的直方图，单位为弧度，分为 60 个区间。

图 8 进一步讨论了 rVAE 编码。图 8a 展示了从输入数据集中随机选取的 64 个图像（特征向量），为我们提供了关于训练数据集的见解。图 8b 展示了该网络的解码潜在空间，阐明了编码随潜在空间变化的情况。从解码

潜在空间中，某些区域会编码理想石墨烯结构，并将平滑变化映射到含有非晶区域的图像中。尽管肉眼无法察觉，图像中另一种变化因素——即应变——也预计会在潜在空间中得到平滑编码。要证明这一点还需要进一步分析，但这超出了本文的讨论范围。

图 8c 显示了第零帧中原子坐标的潜在分布（z_1-z_2）核密度图。图像中心附近的大型聚类对应于理想石墨烯结构，因为大多数输入特征向量都对应于这些结构。远离该聚类的点则是带有缺陷和/或接近非晶区域的特征向量。最后，图 8c 将五十帧中所有原子的编码角度绘制成一个包含 60 个分箱的直方图。图中两个尖锐的峰证实，rVAE 确实找到了石墨烯层中的两种旋转对称变体。这些位于 5 个潜在维度中的编码构成了恰当的特征向量，能够针对每个原子分别编码角度分量和邻域。为简洁起见，本文省略了与平移分量对应的潜在维度相关讨论。整套分析可通过随本文提供的 Jupyter 笔记本“rVAE_graphene.ipnyb”进行可视化和复现。

5. 条件变分自编码器

本节将讨论条件旋转不变变分自编码器，以下简称 crVAE，并在手稿的其余部分中均采用这一称谓。在条件变分自编码器的构造中，我们假设输入数据集被划分为离散类别，且每个数据点的类别标签均已预先获知。或者，条件也可以是连续向量函数。通常，条件表示数据中已知的变异因素。

首先，我们考虑离散条件 rVAE。解码器网络，也称为生成器网络，随后会学习在数据点的类别标签条件下，对编码潜在空间进行解码。在生成器网络中，rVAE 编码得到的 z 向量会与该数据点的独热编码类别标签拼接，然后通过空间解码器网络进行解码，从而构成 crVAE。为使编码器-解码器网络保持对称，我们也将类别标签与展平输入图像拼接。由于解码器部分以类别标签为条件，crVAE 会为每个类别独立学习一个潜在空间。由于 rVAE 的构造本身已分别以旋转和平移编码为条件，因此对于 crVAE，类别

标签可以采样为对应于某个潜在空间的类别；随后，分别为旋转潜在向量和平移潜在向量采样一个向量，并在潜在空间中随机选取一个点，使用解码器网络生成图像。crVAE 构造的一个显著特点是，潜在表示可以以离散类别标签或生成该数据集的连续真实值为条件（对于扑克牌数据集，该值为剪切）。用于优化 crVAE 的证据下界与 rVAE 相同，见公式 4；crVAE 工作原理的示意图见图5b。

5.a. 扑克牌数据集上的 crVAE

为了探究 crVAE 的工作机制，我们使用了 VAE 一节中讨论过的同一组四个扑克牌数据集。对于架构，在所有四种情况下，我们都为编码器和解码器的两个隐藏层各使用了 256 个神经元。这里不讨论前三种情况的结果，因为其解释较为直观，且与第四个数据集的结果趋势相似。我们选择了卡片数据集 iv，因为上一节中的 rVAE 得到的编码并非最优。

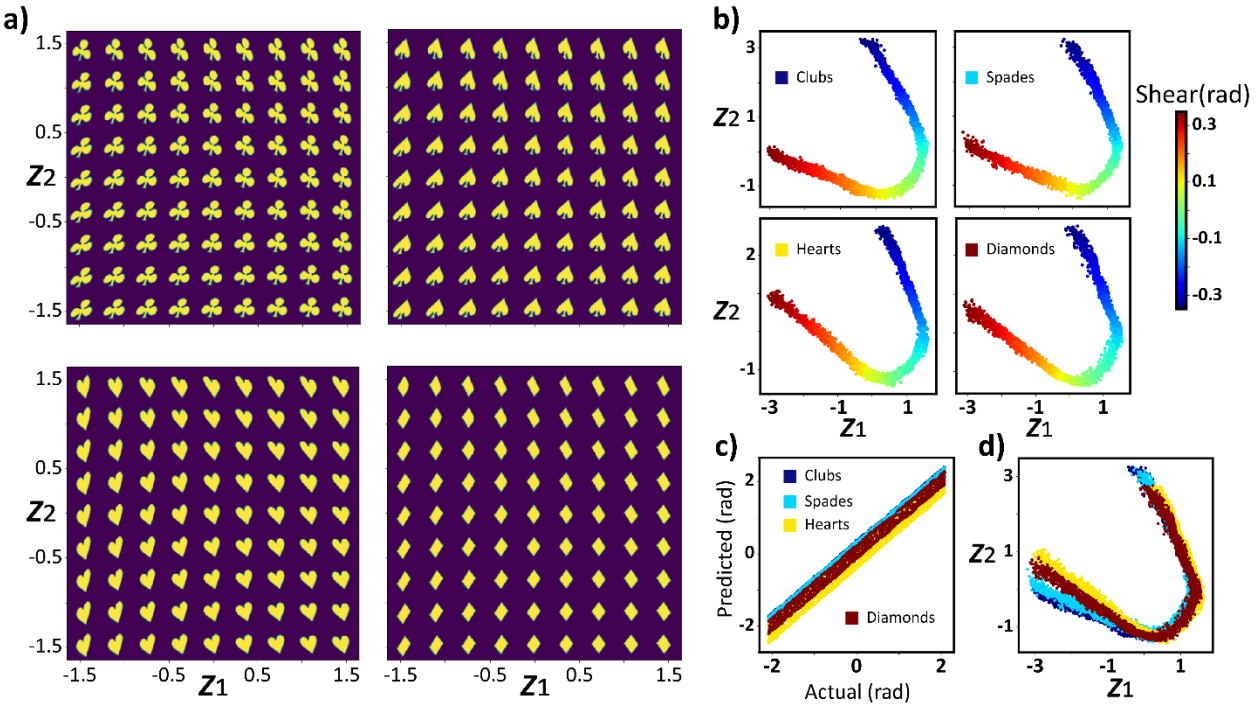


图9。（a）输入 cards 数据集-iv 中四个类别形成的四个独立潜空间，（b）以真实剪切值着色的四个类别的潜在分布，（c）用于旋转输入图像的编码角度与真实角度的对比，（d）四个潜在分布

将这些空间显示在同一张图中，并根据每个数据点的真实标签着色。(d) 的色条与 (c) 中所示的相同。

图9讨论了扑克牌数据集 iv 上的 crVAE 分析。图9a展示了梅花、黑桃、红桃和方块各类别的解码潜在空间。由于类别标签不再是变异性的特征，并且旋转分量已由旋转潜在维度 θ 编码，因此数据集中剩下的唯一变异因素就是施加在图像上的剪切值。该剪切值应在二维潜在空间中平滑变化。在图9b中，将潜在分布按真实剪切值着色后，可以观察到四个类别中的这一现象。图9c展示了编码角度与真实角度的关系。编码角度似乎是关于真实角度的单个线性函数，但存在一定方差，这源于数据集中固有的对称性。最后，图9d展示了按类别标签着色的潜在分布。尽管不同类别的这些编码似乎彼此重叠，但每个类别的编码都位于由类别标签决定的独立潜在空间中。因此，crVAE 是对类别数据集进行编码的更优方法。不过，这一方法的前提是事先知道类别标签。关于全部四个扑克牌数据集（i-iv）的 crVAE 完整分析，可通过随手稿附带的 ‘crvae_cards.ipnyb’ Jupyter 笔记本访问。

5.b. crVAE 在实验数据集上的应用

为了阐明 crVAE 在实验数据集上的工作原理，我们使用了 rVAE 部分中讨论的石墨烯数据集。由于 crVAE 要求将数据集划分类别，我们首先根据中心原子周围环的数量和类型，将裁剪图像划分类别。石墨烯层中的每个碳原子都是三个六边形结构的一部分，但分类时只使用涉及中心原子的这三个环。对于这个庞大的实验数据集，分类任务耗时较长，因此提出了一种用于分类各个子图像的自动化工作流。首先，根据 DCNN网络 识别出的原子构建原子图。如果两个原子之间的距离小于预定义阈值，则在这两个原子之间建立一条边。设置该阈值，使每个原子仅与其三个最近邻相连。为适应石墨烯层中的缺陷，将该距离设定为大于理想的碳-碳键长，同时限制每个原子最多可以拥有

三条边。每个快照的图一次性构建完成，然后通过在给定图中寻找环路，在整幅图像中识别五至七元环。随后创建一个虚拟通道，将每个已识别环的中心像素赋予等于环大小的值。在裁剪子图像以形成输入数据集之前，将该虚拟通道添加到原始扫描透射电子显微镜图像中。然后利用虚拟通道中的信息对每个裁剪子图像进行分类。尽管仅根据中心原子的邻域可以构建大量类别，但为进行分析，仅考虑与其关联数据点数量最多的六个类别。需要注意的是，尽管这种图像分类方法耗时较少，但在构建图中的边时，或由于扫描透射电子显微镜图像中原子识别不当，仍可能出现错误。

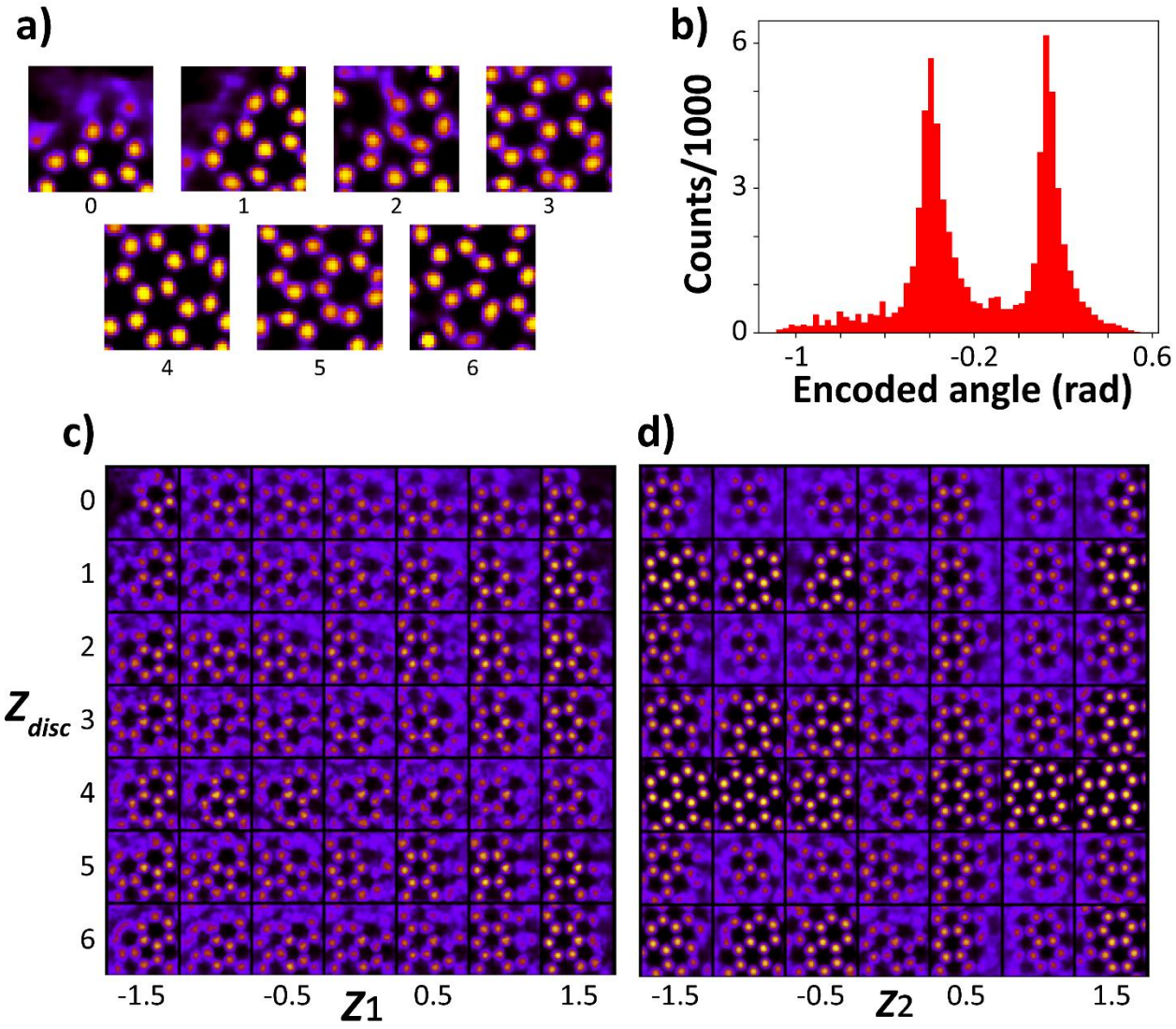


图10。（a）用于石墨烯数据集 crVAE 网络的全部七个类别的代表性数据点；（b）将角度编码（θ）划分为六十个区间后的直方图；（c 和 d）遍历流形，其中改变 x 轴上的一个潜在维度和 y 轴上的类别编码，同时将其他潜变量固定为零。（c）中改变第一潜变量（z_1），（d）中改变第二潜变量（z_2）。

用于分析的七个类别如图 10a 所示，分别为：0：含有少于 2 个环的图像；1：含有 2 [6, 6] 个环的图像；2：含有 2 [5, 6] 个环的图像；3：含有3 [6, 6, 7] 个环的图像；4：含有 3 [6, 6, 6] 个环的图像；5：含有三个 [5, 6, 7] 环的图像；6：含有 3 [5, 6, 6] 个环的图像。需要注意的是，分类时只考虑与中心原子相连的三个六边形结构，而不考虑环的取向。对于

例如，在被归为类别 6 的图像中，五元环可以出现在任意位置。首先，图 10b 展示了本次分析所用子图像的编码角度直方图。与 rVAE 的情况类似，该直方图呈现出典型的双峰，这表明数据集中的所有类别都识别出了石墨烯子层的两种旋转变体。由于这个包含七个类别的数据集会产生七个独立潜空间，为简洁起见，这里不予展示。所有七个流形都可以在本节随附的 Jupyter 笔记本中查看。

为简洁地展示潜在空间，我们采用一种名为遍历流形的新方法，来可视化具有离散类别的数据集的潜在空间。在此方法中，除一个连续潜变量外，其余变量均固定，并沿该维度展示每个类别的一张图像。第一个遍历流形如图 10c 所示，其中 z_1 发生变化；第二个遍历流形则展示 z_2 发生变化的情形。这两个流形中的角度和两个平移潜在维度均设为零。被固定的潜变量始终设为零。此方法类似于 rVAE 的解码潜在空间，其中旋转和平移编码均设为零。此可视化方法的缺点是，当数据集中没有或只有少量图像编码在固定维度的零值处时，遍历流形中会出现不属于训练集的图像。例如，图 10d 中的类别 4（第 5 行）呈现理想的石墨烯晶格；但在图 10c 中将 z_2 固定为零时，我们会看到大量噪声。这表明类别 4 的 $z_2 = 0$ 中并未存储太多结构信息。不过，这些遍历流形为我们提供了一种简洁的表示，展示了不同类别如何编码到潜在维度中。

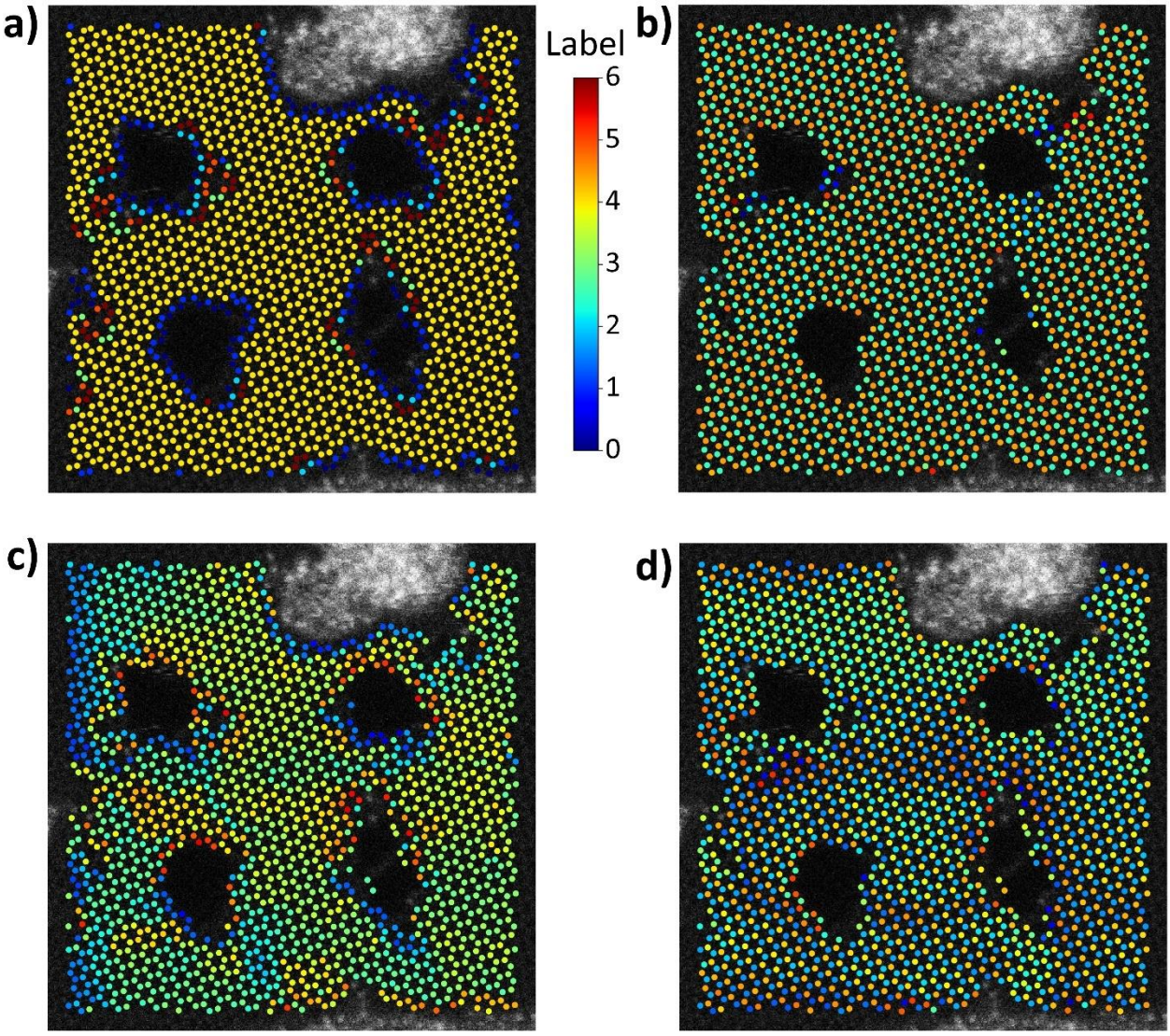


图 11. 语义分割网络在第零帧中识别的原子，分别根据（a）真实类别标签、（b）角度编码（θ）、（c）第一潜变量的值（z_1）以及（d）在石墨烯数据集上训练的 crVAE 网络的第二潜变量值（z_2）着色。

图 11 展示了石墨烯数据集的 crVAE 分析在实空间中的更多结果。图 11a 中，用与原子对应的类别为数据集第零帧中纳入分析的原子着色。这些类别标签作为 crVAE 的输入。图 11b 中，原子坐标按编码角度着色，呈现出 rVAE 中观察到的棋盘格图案线。最后，使用潜变量的值为原子着色

见图 11c（z_1）和图11d（z_2）。如果不了解数据集中各类别的变异性，就无法从这些编码中得出太多结论。不过，图11d（z_2）还显示出棋盘格模式，这表明角度中的部分信息渗入了第二个由用户选定的潜变量 z_2。当 z_1, z_2 潜变量对数据集中的平移/旋转变化进行编码时，就可能出现这种情况。例如，在该数据集中，非应变旋转和平移分量由（θ、 Δx 和 Δy）编码，而缺陷被划分为不同类别，因此图像中仅剩下结构中的应变。如果 z_1 对此进行编码，那么留给另一个潜变量编码的信息就不多了。在这种情况下，减少潜在维度的数量有助于更好地解释潜变量。不过，与 rVAE 的相应编码相比，crVAE 的编码能够更好地作为特征向量，在任何机器学习分析中对每个原子的邻域进行编码。关于石墨烯数据集上 crVAE 的完整分析可参见随手稿附带的 Jupyter 笔记本“crvae_graphene.ipnyb”。

6. 半监督变分自编码器

变分自编码器属于无监督学习机器学习算法这一类。我们很少会遇到这样的训练数据集：其中所有数据点都被划分为不同类别，并且所有数据点的类别标签都已先验知晓。当这些信息已知时，crVAE 能够出色地将数据集投影到低维潜在空间中，同时明确 disentangle 训练数据中涉及的旋转和/或平移。实际遇到的情况是，数据通常以已发表论文、目录及其他形式的小规模人工标注数据集呈现，而我们的目标是将这些数据作为部分已知标签加以利用。另一种情况是，可以由人工对一小部分数据进行标注，然后我们希望将该标签传播到具有强旋转或平移无序的系统中。ssrVAE（半监督旋转不变自编码器）能够从小规模的已标注子集（有监督数据）推广到无监督数据；这些无监督数据来自不同的变化分布，并且伴随强烈的旋转/平移无序。为实现这一点，ssrVAE 利用了 rVAE 和 crVAE 的重要特征。

ssrVAE 的架构包含两个编码器，即 encoder_z 和 encoder_y。输入带有标签的有监督数据集时，encoder_y 充当分类器网络，它是一个带有 softmax激活函数的多层感知机，用于预测该

数据点的类别标签，并使用交叉熵损失训练 encoder_y 网络。encoder_z 网络与我们在 crVAE 中见到的类似：它将图像投影到低维潜在空间，同时明确使用 1–3 个潜变量，将数据集中的旋转和平移解耦。输入无监督训练数据时，encoder_y 会尝试预测每个数据点的类别标签，并将该类别标签与潜在空间中的 z 向量拼接起来，解码过程则以类别标签为条件。无监督设置与有监督设置之间唯一的区别在于：有监督设置使用类别标签训练 encoder_y，而无监督设置使用 encoder_y 预测的类别标签。由于 ssrVAE 尝试使用 encoder_y 网络预测的类别标签重建数据集，因此它既利用了 crVAE 中的条件机制，也明确地将旋转和/或平移潜在向量解耦。

6.a. 卡片上的 ssrVAE

为了阐明 ssrVAE 的工作原理，我们使用扑克牌数据集-i 的一小部分（旋转 = 12o，剪切 = 1o）来尝试预测扑克牌数据集-iii 的标签（旋转 =120o，剪切 = 1o）以及扑克牌数据集-iv 的标签（旋转 = 120o，剪切 = 20o）。通过这种方式为 ssrVAE 的有监督和无监督设置选择不同的数据集分布，我们得以了解 ssrVAE 如何将其从有监督数据中学习到的知识外推到无监督数据集。有监督训练集通过为每张卡片选择 200 张图像构成，并应用从扑克牌数据集 i 中随机采样的旋转值和剪切值进行变换。无监督数据集包含每张卡片 3000 张图像，其旋转值和剪切值来自扑克牌数据集 ii 和 iv。对于这些无监督数据集，我们训练了两个网络，同时保持有监督数据集不变。在高旋转、低剪切的情况下，encoder_z 和解码器层使用的神经元数量为 128；而在高旋转、高剪切的情况下，2 个隐藏层中的每一层均使用 256 个神经元。最后，为了观察每个网络的表现，我们根据构成无监督数据集的分布，构建了两个验证数据集（每种情况一个），每种扑克牌花色包含 1000 张图像。 为简洁起见，下面仅在本手稿的图 12 中讨论高旋转和高剪切（扑克牌数据集-iv）的结果。

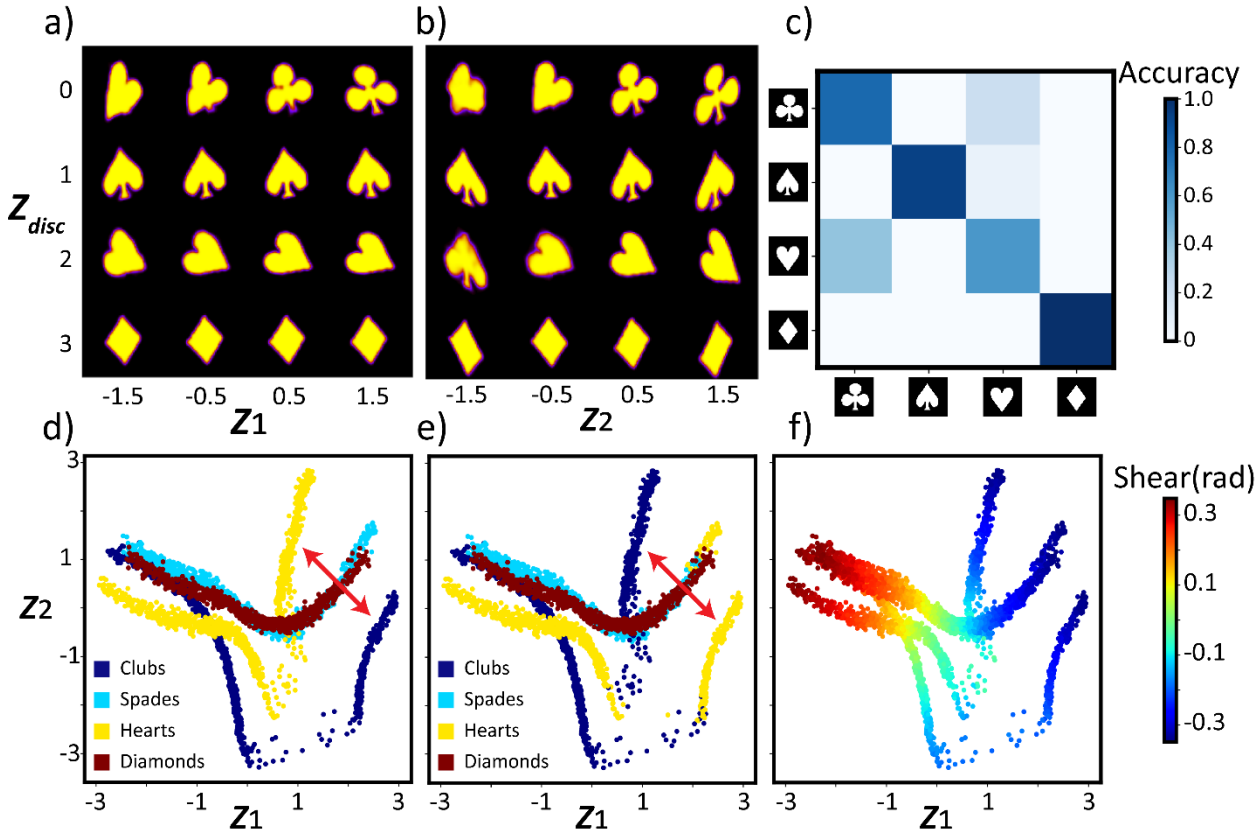


图 12. 在扑克牌数据集上训练的 ssrVAE 的结果，其中无监督数据集取自 cards 数据集-iv，有监督数据集取自 cards 数据集-i。遍历流形，其中 (a) 改变第一潜变量 (z_1)，(b) 改变第二潜变量 (z_2)。(c) 验证数据集上的混淆矩阵。验证数据集的潜在分布分别以 (d) 真实标签、(e) ssrVAE网络预测的标签，以及 (f) 数据的真实剪切值着色。

图 12a 和 12b 使用遍历流形来可视化解码后的潜在空间，具体讨论见 crVAE 部分。在图 12a 中，z_2 保持为零，同时沿 x 轴改变另一个潜变量 z_1；而在图 12b 中则相反。在两个遍历流形中，离散类别标签均沿 y 轴变化。为了生成这些流形，训练阶段结束后会丢弃 encoder_y 和 encoder_z，并将每个类别的类别标签与采样得到的 z 向量拼接起来，以解码潜在向量。可以观察到，类别 0（梅花，第一行）和类别 2（红桃，第三行）存在一些误分类。这一点也可以从图 12c 所示的归一化混淆矩阵中推断出来。梅花和黑桃的准确率约为 50%，图像会被相互识别。潜在分布的

验证数据集绘制于图12d和图12e中，并分别使用图12d中的真实类别标签和图12e中的预测类别标签进行着色。

图12d和图12e中使用红色箭头标出了梅花和黑桃之间的错误分类，其中潜在分布中的两个分支似乎交换了颜色。这是由于有监督数据集与无监督数据集之间的分布差异较大，以及高度剪切的红桃类别与梅花类别之间具有结构相似性。最后，图12f中的验证数据集潜在分布使用真实剪切值进行着色，显示出潜在空间中的剪切平滑变化。尽管不同类别的剪切值潜在分布似乎存在重叠，但正如在 crVAE 的情况下所讨论的，每个类别都存在于独立的潜在空间中。这些编码各自所在的潜在空间取决于 encoder_y 预测的类别标签，该标签由图12e中的颜色表示。需要注意的是，在有监督训练数据集中，剪切的最大值为 1o，而 ssrVAE 将其泛化到了无监督数据集，其中剪切的最大值为 20o。训练后的 ssrVAE 通过将在有监督数据集中学习到的特征进行泛化，在无监督数据集上达到了 82% 的准确率。 总之，当输入 800 张低旋转（ 12o ）和低剪切（ 1o ）的图像时，ssrVAE 能够以 82% 的准确率进行分类，显式解耦旋转，并在包含 12,000 张图像的高旋转（120o）和高剪切（20o）数据集中形成其他变化的平滑变化潜在空间。两个数据集的完整分析可通过随手稿提供的‘ssrvae_cards.ipnyb’ Jupyter 笔记本进行访问和可视化。

6.b. ssrVAE 在石墨烯数据集上的应用

crVAE 部分讨论的石墨烯数据集设置同样用于说明 ssrVAE 的工作原理。输入图像中属于 crVAE 部分所讨论的七个类别之一的图像，仅有 15% 被用作有监督数据集。数据集中的一小部分（4%）被留作验证数据集。其余数据集，包括不属于 crVAE 部分所讨论的七个类别的图像，被用作无监督数据集。遍历流形方式的可视化结果如图13a和图13b所示。在图13a中改变第一潜变量（z_1），同时将 z_2 保持为零；在图13b中则反之。这些流形有助于了解不同类别如何编码在潜在维度 z_1–z_2 中。验证数据集上的混淆矩阵如图13c所示。该网络能够预测

类别 4，即准确率最高的理想[6, 6, 6] 。验证数据集上的最终准确率为 74.6%

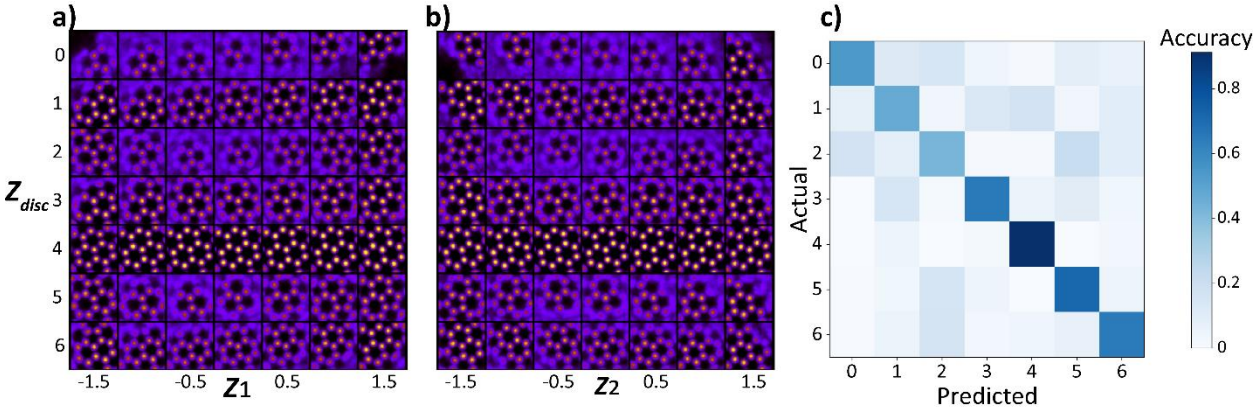


图 13。在 (a) 第一潜变量 (z_1) 变化和 (b) 第二潜变量 (z_2) 变化时遍历流形。(c) ssrVAE 设置下石墨烯验证数据集的混淆矩阵。

潜在编码的进一步结果如图 14 所示。为获得这些编码，将整个数据集用作输入，并在图 14a 中使用编码角度为第零帧中的每个原子着色。前几节中观察到的棋盘格图案在此情形下同样普遍存在。潜在维度 z_1 和 z_2 分别用于为图 14b 和图 14c 着色。由于 ssrVAE 以 encoder_y 预测的类别为条件，这些编码值位于不同的潜在空间中。这意味着邻域中的结构变异性已经由 encoder_y 编码。正如 crVAE 部分所讨论的，潜在维度编码了应变和微观变异性，这些变异性可以从图 13a 和图 13b 中的遍历流形观察到。最后，在图 14d 中，根据 encoder_y 预测的类别为原子着色。理想石墨烯晶格类别 4 占据了帧的大部分区域。

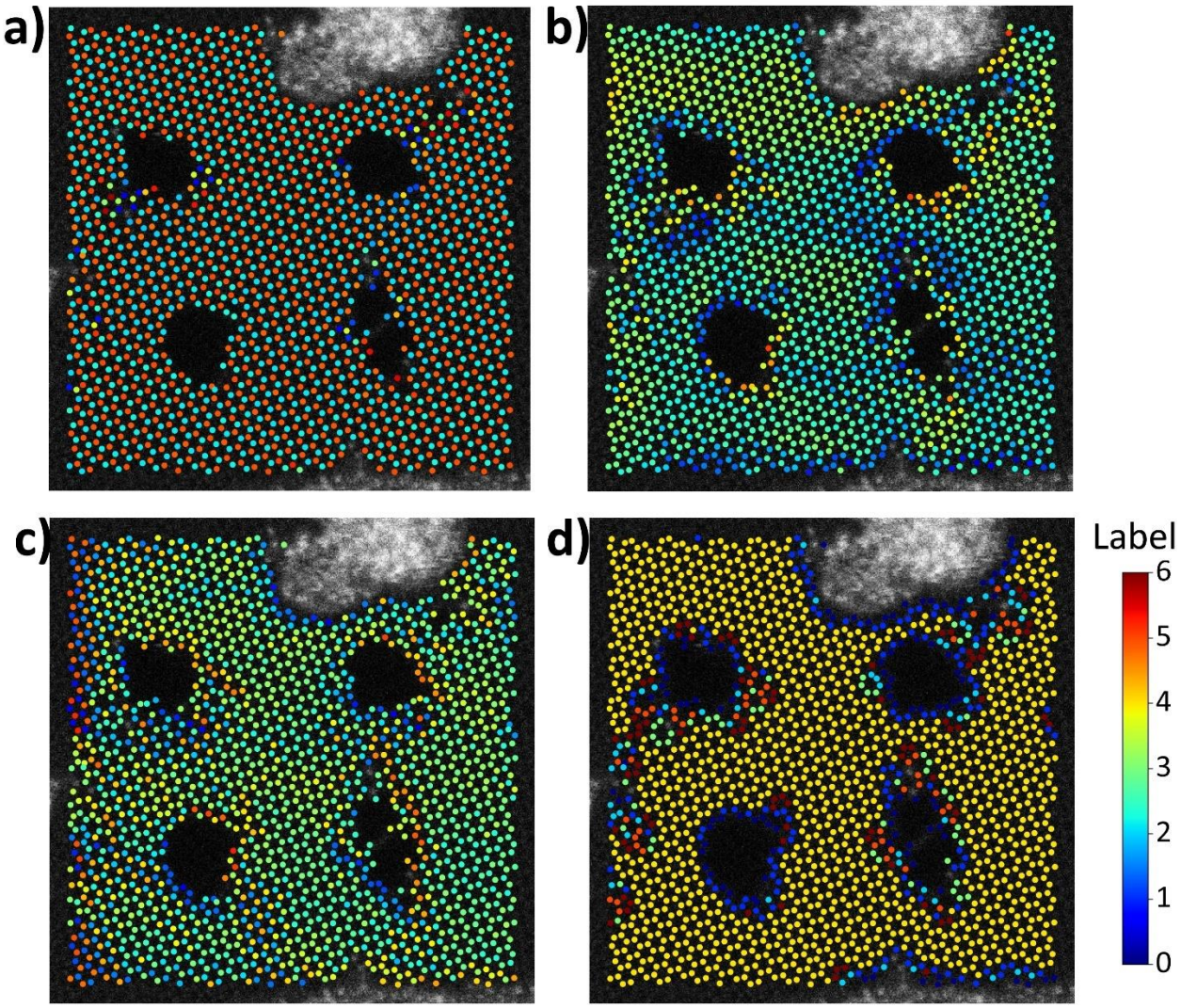


图 14。语义分割网络在第零帧中识别出的原子分别使用（a）角度编码（θ）、（b）第一潜变量的值（z_1）、（c）第二潜变量的值（z_2）以及（d）在石墨烯数据集上训练的 ssrVAE 网络预测的类别标签进行着色。

7. 联合变分自编码器

通常，成像数据包含来自多个类别的物体，同一类别内的图像也存在变异性。在监督式机器学习和半监督式机器学习设置中，数据集至少有一部分带有标签。随后，算法会将从已知类别中学习到的信息传播到无标签数据集。然而，数据集中的类别并不明显，尤其是在难以获得大量带标签数据集的学科领域。无监督学习

聚类技术会根据输入数据集中的隐藏模式或数据分组，将数据集划分为离散类别。在将无标签数据集划分为离散类别的同时学习低维流形，这一问题并不适合采用传统的机器算法。

本节将讨论一类适用于解决这一问题的不变变分自编码器，称为联合旋转不变变分自编码器（jrVAE）。名称中的“联合”一词描述了 jrVAE 潜在空间的一个特点：其潜在空间由离散潜在维度和连续潜在维度共同构成。这类似于 ssrVAE 在分类器网络未经训练时的无监督设置。jrVAE 的离散潜在维度负责将数据集解耦为不同类别，而连续维度则学习类内变化。这些特点也可以与旋转和平移不变性分别结合，其中 1–3 个潜在维度用于捕捉图像中的旋转和平移。

jrVAE 的证据下界由下面的公式 5 描述，其中重构误差是指原始图像与重构图像之间的误差。第二项包含三个不同的组成部分，分别对应于连续潜变量与其相应先验分布之间的 Kullback-Leibler（KL）散度。第一部分 $D_{KL}(q(z|x)||\mathcal{N}(0,I))$ 是潜变量（z）与标准正态先验分布之间的 KL 散度。随后是 $D_{KL}(q(\theta|x)||M(0,\kappa))$，它表示角度维度（θ）与参数为 0 和 κ 的冯·米塞斯先验之间的 KL 散度。角度（θ）维度上的先验分布是包裹正态分布，在区间 [-π,π) 上模拟均匀分布 θ 。第二项的最后一项 $D_{KL}(q(\Delta x|x)||\mathcal{N}(0,\sigma_s))$ 指的是平移维度与其标准差为 σ_s 的正态先验之间的 KL 散度。最后一项是对离散维度的惩罚，表示为离散潜变量与其先验 $p(y) = Cat(y|\lambda)$ 之间的 KL 散度，该先验是参数为 λ 的离散分类分布。

	$ELBO_{jrvae} = -Reconstruction\ error$ $-\gamma(D_{KL}(q(z\|x)\|\|\mathcal{N}(0,I)) + D_{KL}(q(\theta\|x)\|\|M(0,\kappa))$ $- D_{KL}(q(\Delta x\|x)\|\|\mathcal{N}(0,\sigma_s)) - C_z) - \gamma(D_{KL}(q(y\|x)\|\|p(y)) - C_y)$	5

针对给定数据集，需要优化的证据下界超参数包括 κ：冯·米塞斯先验中关于 θ 的参数； σ_s：Δ 的正态先验的标准差x； λ：离散潜变量的分类先验分布的参数。除此之外，jrVAE 网络还高度依赖另外三个超参数，即 γ、C_z 和 C_y。连续通道容量和离散通道容量分别为 C_z 和 C_y。当在权重系数 γ 保持不变的情况下将它们设为较高值时，KL 项的影响会减弱，从而更加重视重构误差。这三个项的作用与 β-VAEs 中参数 β 的作用相同。在训练初期，会使用较低的通道容量，以便首先更好地实现数据的解纠缠，随后降低该容量，以改善重构效果。需要特别注意的是，尽管其他变体的变分自编码器（主要是 ssrVAE）对 β 超参数的依赖较弱，但根据我们的经验，jrVAE 的有效性在很大程度上取决于对这些参数的优化。

7.a. jrVAE 在卡片数据集上

在简要介绍 jrVAE 后，本节使用扑克牌数据集来展示其能力。针对四个数据集分别训练了一个 jrVAE 网络，其中隐藏层使用了 1024 个神经元。编码器和解码器网络均使用四个隐藏层。两个用户选定的潜变量（z_1, z_2）、三个不变潜变量（θ、 Δx 和 Δy）以及一个具有 4 个维度的离散潜变量（y）共同构成潜在空间。jrVAE 预测的类别标签是离散潜变量中取值最高的维度。本节讨论扑克牌数据集 i 和 iii 上 jrVAE 的结果。不过，全部四个数据集的结果都可以在随本节附带的 Jupyter 笔记本中找到。扑克牌数据集 i 的遍历流形如图 15a 和 15b 所示，其中图 15a 中改变了 z_1，图 15b 中改变了 z_2，而其余连续变量均固定为零。遍历流形表明，对于扑克牌数据集 i，jrVAE 能够以最佳方式

识别出四个类别，而无需预先了解类别。四个潜在空间、潜在分布和角度图都可以在 Jupyter 笔记本中可视化。扑克牌数据集 iii 的遍历流形如图 15c（改变 z_1）和 15d（改变 z_2）所示。在此情况下，jrVAE 对方块的表现十分理想，因为对应方块（类别 3）的类别由同一方向上的所有方块花色填充。

该网络还将所有梅花识别为同一个类别（类别 0），但其方向不止一种。我们观察到，当数据集中的旋转较多时，其他变体的变分自编码器在角度潜在维度上也会出现多个基态零点。最后，由于黑桃和红桃之间存在结构相似性，这两个数据集都落在类别 1 和类别 3 中。我们还可以看到，红桃被识别为具有两种不同的基态方向。尽管 jrVAE 预测的类别是任意的，但会根据真实类别占多数的原则为四个类别分别赋予类别标签。根据这些预测类别标签，jrVAE 在扑克牌数据集 iii 上的准确率为 82.11%。由于我们事先知道真实类别标签，因此可以计算准确率。所有四个数据集的完整分析、所有类别的潜在空间、潜在分布、角度图，以及潜在空间图中的剪切变化，均见随手稿提供的“jrVAE_card.ipnyb”Jupyter 笔记本。

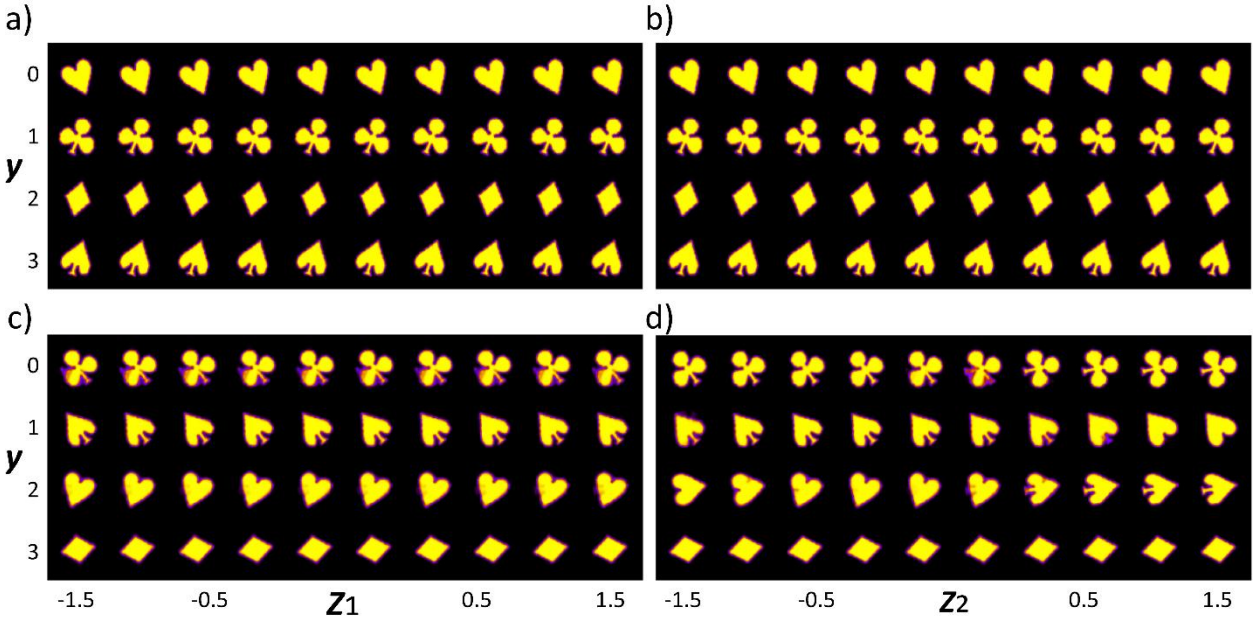


图 15。在扑克牌数据集 I 上应用 jrVAE 时的流形遍历：(a) 改变 z_1，(b) 改变 z_2；在卡片数据集 iii 上：(c) 改变 z_1，(d) 改变 z_2。

如前所述，jrVAE 对超参数 γ、Cz 和 Cy 高度敏感。应针对每个数据集优化这些超参数，以获得理想结果。在本综述中，我们使用了 AtomAI 软件包提供的默认参数，并利用该软件包创建 jrVAE 网络。优化这些超参数超出了本综述的讨论范围；文献 ref.[56] 提出了一种优雅的方法，可针对每个数据集优化这些参数。默认参数通常会产生不够理想的结果，因此只能作为理解数据集的初步步骤。

7.b. 石墨烯上的 jrVAE

图 16 讨论了 jrVAE 应用于石墨烯数据集的结果。用于此次应用的网络架构与扑克牌数据集所使用的架构相似，唯一的变化是离散潜在维度（y）的维数；在本例中，该维数被设为 2。图 16a-c 分别展示了角度维度中的编码 z_1，和 z_2，并将其叠加在第零帧检测到的原子坐标上。jrVAE 中同样观察到了角度潜在维度中的棋盘格图案，这表明 jrVAE 也找到了针对角度的最优编码。对于 z_1 和 z_2，的编码，由于每个子图像分别位于两个独立潜空间中的一个，因此必须结合类别标签来理解这些维度上的变异性。图 16d 使用预测的类别标签为原子坐标着色，其中较大的绿色圆点对应检测到的硅原子。图 16e（改变 z_1）和图 16f（改变 z_2）展示了遍历流形。图 16d-f 表明，第零类（图 16d 中的红色圆点）对应理想的石墨烯晶格，而第一类（图 16e 中的蓝色圆点）对应缺陷以及与非晶区域的接近程度。在这些类别的条件下，潜变量 z_1 和 z_2 对特征向量中的额外变异性进行编码。整个分析都可以在随手稿附带的“jrvae_graphene.ipnyb”Jupyter 笔记本中重现和可视化。

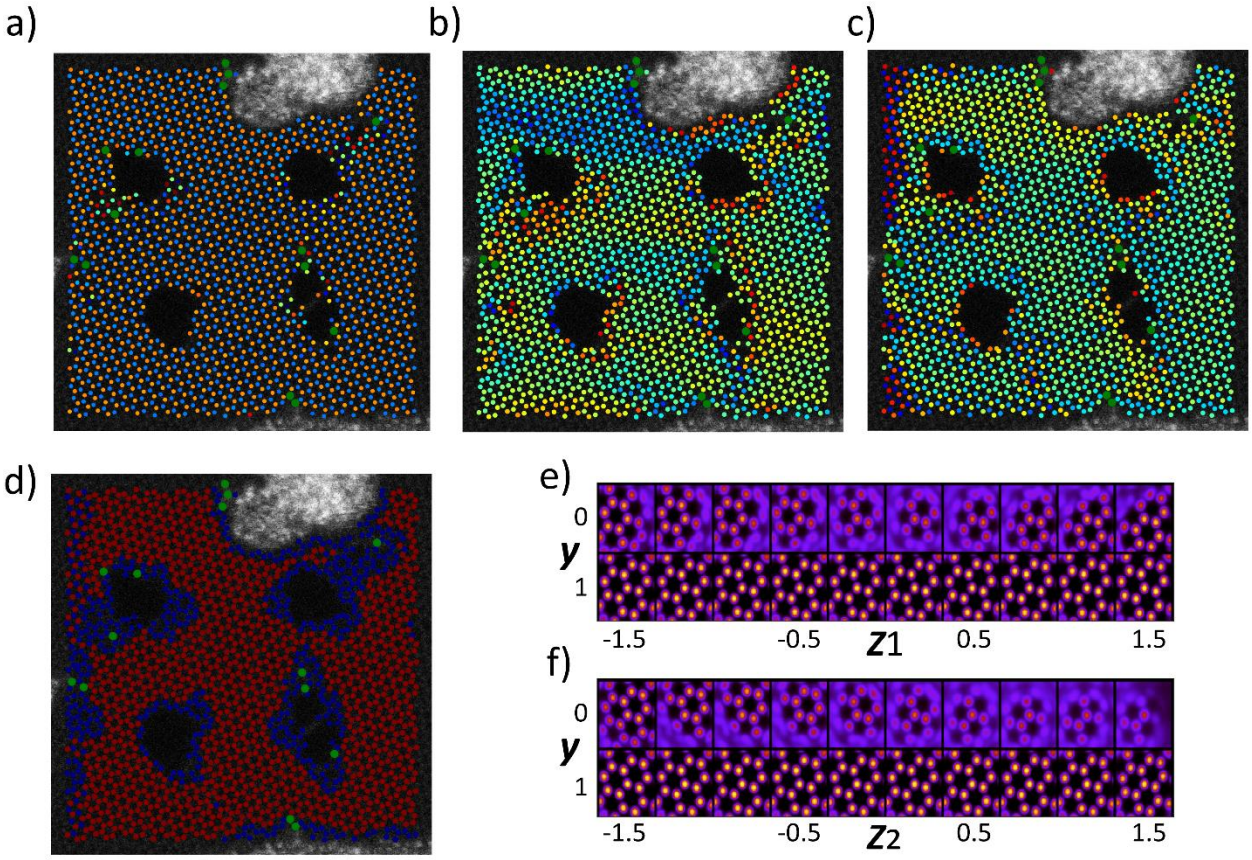


图 16。 语义分割网络在第零帧中识别出的原子分别使用以下方式着色：(a) 角度编码 (θ)；(b) 第一潜变量的值（z_1）；(c) 第二潜变量的值（z_2）； 以及 (d) 在石墨烯数据集上训练的 jrVAE 网络预测的类别标签（y）。当 (e) z_1 变化和 (f) z_2 变化时，遍历 jrVAE 的流形。

表 I. 自编码器和变分自编码器

1	AE	自编码器是一种通过学习重构 输入数据来实现这一目标的神经网络，它将数据压缩到低维潜空间中 然后将其扩展回原始维度。
2	编码器	自编码器中的神经网络组件，用于将输入 数据映射为低维潜空间表示。
3	解码器	自编码器中的神经网络组件，用于重构 来自编码表示的原始输入数据。

4	重构 loss	衡量输入与重构 输出数据之间的差异。通常使用均方误差或 二元交叉熵损失来计算图像数据，并使用平均绝对误差 或余弦相似度来计算文本数据。
5	潜在 分布	数据编码到潜在空间后形成的分布。 潜在分布的结构有助于得出结论 涉及系统物理特性、类别数量等方面。
6	解码后的潜在 空间	通过解码在以下区域采样的潜在向量所生成的图像， 这些向量采样自点网格，并以图像表格的形式呈现。潜在 表示展示了数据特征如何随 潜在空间变化，从而使潜变量具有可解释性。
7	VAE	变分自编码器是一类贝叶斯生成模型， 利用概率编码器和解码器来学习潜在 数据分布的表示，并能够估计 潜变量的后验概率。
8	编码器- 解码器网络	编码和解码后的 物体具有不同的维度。变分编码器- 解码器可用于构建结构-性质 关系。[38,57,58]
8	先验分布 分布	变分自编码器中潜变量的先验分布。通常假设 为高斯分布。
9	KL散度	Kullback-Leibler散度，用于衡量 两个概率分布之间的差异，通常用于变分 推断，以衡量目标分布与 近似分布之间的距离。
10	因子 变异性	数据中的特征。例如，在 MNIST 中是手写风格，而 在卡片中，则是应用于图像的旋转和剪切
11	解耦 of 表示	变分自编码器发现数据中变化因素的能力。 尽管缺乏严格的定义，但这是其中最显著的 变分自编码器优势之一，而且可以直接从数据中观察到。理想情况下， 表示将发现数据中的变化因素。

12	iVAE	不变变分自编码器会将变化的物理因素分离出来，例如旋转、平移或膨胀/收缩，并分别编码到专用的潜变量中。
13	rVAE	旋转不变变分自编码器是一类不变变分自编码器，它会显式地将旋转和/或平移解缠到专用的潜变量中。通常用于图像数据。
14	cVAE	条件变分自编码器，已知的物理行为（例如物体类别或连续变量）可用于对数据进行条件化。对于离散标签，我们为每个类别构建独立的潜在空间。
15	半监督变分自编码器	半监督变分自编码器适用于仅有部分数据集可获得目标标签/类别标签。半监督变分自编码器会传播从有监督数据集中学到的信息，并将其用于无监督数据集。不变半监督变分自编码器对于构建花粉、恒星、分子等图像的图像库非常有用。
16	jVAE	联合变分自编码器可同时将物体分离到各个子空间中（分类），并解耦每个子空间中的表示空间。
17	遍历表示	联合变分自编码器或半监督变分自编码器的解码后的潜空间表示，作为离散类别和一个连续潜变量的函数在其余潜变量保持为常数时零。
18	信道容量	每次迭代中分配给证据下界各信道的权重在训练期间。这些容量决定了训练阶段主导证据下界的项。

8. 总结

总而言之，本文讨论了用于成像数据的不变变分自编码器的基本原理、实现方法及应用。所引入定义的完整列表总结于表 I 中。变分自编码器的核心概念是对数据表示进行解耦

，即寻找简约变化因素。在 MNIST 等经典数据集中，变化因素指的是手写风格的特征，例如字母的倾斜角度和宽度。在成像数据中，变化因素包括纳米颗粒的形状变化、应变或铁电极化导致的晶胞畸变、石墨烯中原子的化学邻域，或四维扫描透射电子显微镜中局部衍射图样的变异性。相应地，变分自编码器为从观测数据中学习这些行为提供了一条潜在途径。

然而，简单变分自编码器在实验数据中的直接应用受到限制。首先，在观测研究中，同一类型的多个物体可能以不同的观测角度被观测到。其次，用于识别子图像的采样过程可能伴随着随机平移。最后，系统可能存在仿射畸变，这些畸变可能源于材料内禀结构，也可能由成像系统导致。本文介绍的平移和旋转不变变分自编码器能够直接分离这些变化的物理因素，并在潜在表示中解耦其余的变化因素。

类似地，成像数据通常采集自包含多个离散对象类别的系统，而这些类别可能是已知的、部分已知的（有示例），或未知的。我们讨论条件变分自编码器、半监督变分自编码器和联合变分自编码器，将其作为应对这些任务的框架。值得注意的是，条件变分自编码器还可以扩展到连续条件向量，从而能够纳入已知的变化物理因素。

总体而言，该框架能够对成像数据集进行全面分析。我们进一步讨论将变分自编码器方法扩展到表征学习，引入基于物理的损失函数和因果表示。我们还讨论了潜在空间和潜在流形的形成及其性质背后的一些直觉。最后，我们指出，变分自编码器方法可以扩展到构建结构-性质关系（refs）和主动学习等更复杂的任务，但这些内容将留待未来的出版物中讨论。

本综述中阐释相关概念的笔记本均已完整公开 https://github.com/saimani5/VAE-tutorials。我们鼓励读者探索这些概念，并将 VAE 应用于自己的数据。

致谢：

本研究（工作流程开发、手稿撰写）得到美国能源部科学办公室基础能源科学办公室的支持，作为能源前沿

研究中心计划的一部分：CSSAS——跨尺度合成科学中心——由编号为 DE-SC0019288 的资助项目支持，位于华盛顿大学特区。该工作还得到三维铁电微电子学中心（3 DFeM）的部分支持；该中心是由美国能源部（DOE）科学办公室基础能源科学资助的能源前沿研究中心，资助编号为 DE-SC0021118。作者感谢 Maxim Ziatdinov 博士编写了用于构建文章中讨论的变分自编码器网络的 Python 软件包 AtomAI 和 pyroved。我们还感谢他在整个研究过程中提供的深刻见解。

参考文献

1 van Zuylen, J. 安东尼·范·列文虎克的显微镜。《显微镜学杂志》121, 309-328（1981）。2 Masters, B. R. 细胞生物学与医学中光学显微镜的历史。eLS（2008）。3 Fan, Z. 等。用于能源材料与器件的原位透射电子显微镜。Advanced Materials 31, 1900608（2019）。4 Bruma, A. 扫描透射电子显微镜：材料科学应用的先进表征方法。（CRC Press，2020）。5 Wang, Y., Skaanvik, S. A., Xiong, X., Wang, S. 和 Dong, M. 用于电催化的扫描探针显微镜。《Matter》4, 3483-3514（2021）。6 Wen, H., Cherukara, M. J. 和 Holt, M. V. 用于材料科学的时间分辨X射线显微镜。《材料研究年度评论》49, 389-415（2019）。7 Hui, F. 和 Lanza, M. 用于先进纳米电子学的扫描探针显微镜。《自然电子学》2, 221-229（2019）。8 Gerber, C. 和 Lang, H. P. 纳米世界之门如何开启。《自然·纳米技术》1, 3-5，doi:10.1038/nnano.2006.70（2006）。9 Binnig, G. 和 Rohrer, H. 扫描隧道显微术。《Helvetica Physica Acta》55, 726-735（1982）。10 Stroscio, J. A., Feenstra, R. M. 和 Fein, A. P. 通过扫描隧道显微术研究Si(111)2X1表面的电子结构。《物理评论快报》57, 2579-2582，doi:10.1103/PhysRevLett.57.2579（1986）。11 Asenjo, A., Gomezrodriguez, J. M. 和 Baro, A. M. 硅上金属沉积物的电流成像隧道谱学。《超显微镜》42, 933-939，doi:10.1016/0304-3991(92)90381-s（1992）。12 Grutter, P., Liu, Y., LeBlanc, P. 和 Durig, U. 磁耗散力显微镜。Applied Physics Letters 71, 279-281（1997）。13 Martin, Y. 和 Wickramasinghe, H. K. 通过力显微镜进行磁成像，分辨率达1000 Å。Applied Physics Letters 50, 1455-1457，doi:10.1063/1.97800（1987）。14 Noy, A., Vezenov, D. V. 和 Lieber, C. M. 化学力显微镜。《材料科学年度评论》27, 381-421，doi:10.1146/annurev.matsci.27.1.381（1997）。15 Gruverman, A., Auciello, O. 和 Tokumoto, H. 通过扫描力显微镜对铁电薄膜中的畴结构进行成像与调控。《材料科学年度评论》28, 101 123, 10 1146 281 101 1998 - doi: . /annurev.matsci. . . ( )。16 Kim, M. 等。SrTiO3中晶界的非化学计量性与电活性。《物理评论快报》86, 4056-4059，doi:10.1103/PhysRevLett.86.4056（2001）。17 Pennycook, S. J., Varela, M., Lupini, A. R., Oxley, M. P. 和 Chisholm, M. F. 原子分辨率光谱成像：过去、现在与未来。《电子显微镜杂志》58, 87-97，doi:10.1093/jmicro/dfn030（2009）。18 Sohlberg, K., Rashkeev, S., Borisevich, A. Y., Pennycook, S. J. 和 Pantelides, S. T. Pt/氧化铝催化体系中异常Pt-Pt间距的起源。《化学物理化学》5, 1893-1897，doi:10.1002/cphc.200400212（2004）。19 Chisholm, M. F., Luo, W. D., Oxley, M. P., Pantelides, S. T. 和 Lee, H. N. 极性界面处的原子尺度补偿现象。《物理评论快报》105，doi:197602 10.1103/PhysRevLett.105.197602（2010）。20 Browning, N. D. 等。原子结构对SrTiO3晶界处电势垒形成的影响。Applied Physics Letters 74, 2638-2640，doi:10.1063/1.123922（1999）。

21 Clausen-Schaumann, H., Seitz, M., Krautbauer, R. 和 Gaub, H. E. 单个生物分子的力谱学。《化学生物学当前观点》 **4**, 524-530，doi:10.1016/s1367- 5931(00)00126-5（2000）。22 Rief, M., Oesterhelt, F., Heymann, B. 和 Gaub, H. E. 通过原子力显微镜对多糖进行单分子力谱学。《科学》 **275**, 1295-1297，doi:10.1126/science.275.5304.1295（1997）。23 Balke, N., Bdikin, I., Kalinin, S. V. 和 Kholkin, A. L. 铁电与压电材料的机电成像与谱学：最新进展及未来展望。《美国陶瓷学会志》 **92**, 1629-1647，doi:10.1111/j.1551-2916.2009.03240.x（2009）。24 Van der Walt, S. 等 scikit-image：Python中的图像处理。*PeerJ* **2**, e453（2014）。25 Krizhevsky, A., Sutskever, I. 和 Hinton, G. E. 使用深度卷积神经网络进行ImageNet分类。《*ACM*通讯》 **60**, 84-90（2017）。26 Opitz, D. 和 Maclin, R. 常用集成方法：一项实证研究。《人工智能研究杂志》 **11**, 169-198（1999）。27 Vyas, A. 等 载于《欧洲计算机视觉会议（*ECCV*）论文集》。550-564。28 Xia, G. 和 Bouganis, C.-S. 深度集成多样性对分布外检测的效用。*arXiv*预印本 *arXiv:2207.07517*（2022）。29 Liu, Y., Kelley, K. P., Funakubo, H., Kalinin, S. V. 和 Ziatdinov, M. 实时探索铁电畴壁的物理特性：深度学习赋能的扫描探针显微镜。《*Advanced Science*》**9**, 2203957（2022）。30 Liu, Y. 等 通过自动化扫描探针显微镜解析混合钙钛矿中单个晶界处的电子输运与滞回。*arXiv*预印本 *arXiv:2210.14138*（2022）。31 Kingma, D. P. 和 Welling, M. 自编码变分贝叶斯。*arXiv*预印本 *arXiv:1312.6114*（2013）。32 Kingma, D. P. 和 Welling, M. 变分自编码器简介。《机器学习基础与趋势》 **12**, 307-392（2019）。33 Rezende, D. J., Mohamed, S. 和 Wierstra, D. 载于《国际机器学习会议》。1278-1286（PMLR）。34 Taud, H. 和 Mas, J. 多层感知器（MLP）。用于土地变化情景建模的地学信息学方法, 451-455（2018）。35 Goodfellow, I., Bengio, Y. 和 Courville, A. 载于《深度学习》第2016卷，330-372页（麻省理工学院出版社，马萨诸塞州剑桥，美国，2016）。36 Vaswani, A. 等 注意力就是你所需要的一切。神经信息处理系统进展**30**（2017）。37 de Haan, P., Cohen, T. S. 和 Welling, M. 自然图网络。神经信息处理系统进展 **33**, 3636-3646（2020）。38 Kalinin, S. V., Kelley, K., Vasudevan, R. K. 和 Ziatdinov, M. 借助隐藏潜变量，探寻铁电体中畴结构与功能性的关系。《*ACS*应用材料与界面》 **13**, 1693-1703（2021）。39 Lecun, Y., Bottou, L., Bengio, Y. 和 Haffner, P. 应用于文档识别的基于梯度的学习。《IEEE学报》《*IEEE*学报》 **86**, 2278-2324，doi:10.1109/5.726791（1998）。40 Duvenaud, D. K. 等 用于学习分子指纹的图上的卷积网络。神经信息处理系统进展 **28**（2015）。41 Weininger, D. SMILES：一种化学语言和信息系统。1. 方法论与编码规则简介。《化学信息与计算机科学杂志》 **28**, 31 36, 101021 00057 005 1988 - doi: . /ci a ( )。

Krenn, M., Häse, F., Nigam, A., Friederich, P. & Aspuru-Guzik, A. 自指嵌入字符串（SELFIES）：一种100%稳健的分子字符串表示。Machine Learning: Science and Technology 1, 045024, doi:10.1088/2632-2153/aba947 (2020年)。Liu, Y., Proksch, R., Wong, C. Y., Ziatdinov, M. & Kalinin, S. V. 通过深度学习解耦铁电畴壁动力学并识别钉扎机制。Advanced Materials 33, 2103680, doi:https://doi.org/10.1002/adma.202103680 (2021年)。Kalinin, S. V., Steffes, J. J., Liu, Y., Huey, B. D. & Ziatdinov, M. 通过无监督机器学习解耦动态压电响应力显微镜中的铁电畴壁几何形态与演化路径。Nanotechnology 33, 055707 (2021年)。Liu, Y., Ziatdinov, M. & Kalinin, S. V. 通过非高斯线性模型和深度核学习探索因果物理机制：在铁电畴结构中的应用。ACS Nano 16, 1250-1259 (2021年)。Valleti, S. M. P., Ignatans, R., Kalinin, S. V. & Tileli, V. 从原位显微术观测中解码相变机制。Small 18, 2104318, doi:https://doi.org/10.1002/smll.202104318 (2022年)。Liu, Y. et al. 解码平移不变数据：在带激励扫描探针显微术中的应用。Machine Learning: Science and Technology 2, 045028 (2021年)。Liu, Y., Huey, B. D., Ziatdinov, M. A. & Kalinin, S. V. 通过以先验知识为条件的表征学习实现物理发现：在铁电畴动力学中的应用。arXiv预印本 arXiv:2203.03122 (2022年)。Bepler, T., Zhong, E., Kelley, K., Brignole, E. & Berger, B. 利用空间变分自编码器显式解耦图像内容与平移和旋转。神经信息处理系统进展 32 (2019年)。Kalinin, S. V., Dyck, O., Jesse, S. & Ziatdinov, M. 通过变分自编码器探索无序系统中的序参量和动态过程。《科学》进展 7, eabd5084, doi:doi:10.1126/sciadv.abd5084 (2021年)。Ziatdinov, M. et al. 构建并探索石墨烯中的原子缺陷库：扫描透射电子显微镜和扫描隧道显微术研究。Science Advances 5, eaaw8989, doi:doi:10.1126/sciadv.aaw8989 (2019年)。Dyck, O., Kim, S., Kalinin, S. V. & Jesse, S. 利用扫描透射电子显微镜在石墨烯中放置单原子。Applied Physics Letters 111, 113104, doi:10.1063/1.4998599 (2017年)。Dyck, O., Kim, S., Kalinin, S. V. & Jesse, S. 利用像差校正扫描透射电子显微镜对石墨烯边缘的硅原子进行电子束操控。Nano Research 11, 6217-6226, doi:10.1007/s12274-018-2141-6 (2018年)。Dyck, O. et al. 通过电子束操控逐原子构建结构。Small 14, 1801771, 101002 201801771 2018 doi:https://doi.org/ . /smll. ( )。Ziatdinov, M. et al. 原子分辨扫描透射电子显微图像的深度学习：化学识别和局部转变追踪。ACS Nano 11, 12742-12752 (2017年)。Biswas, A., Vasudevan, R., Ziatdinov, M. & Kalinin, S. V. 通过潜空间贝叶斯优化方法优化变分自编码器的训练轨迹*。Machine Learning: Science and Technology 4, 015011, doi:10.1088/2632-2153/acb316 (2023年)。Roccapriore, K. M., Ziatdinov, M., Cho, S. H., Hachtel, J. A. & Kalinin, S. V. 纳米颗粒组装体中局域等离激元响应的可预测性。Small 17, 2100181, doi:https://doi.org/10.1002/smll.202100181 (2021年)。

58 Ziatdinov, M., Ghosh，A., Wong, C. Y. & Kalinin, S. V. 用于电子显微成像和扫描探针显微镜中图像与光谱数据深度学习分析的 AtomAI框架。*NatureMachine Intelligence* **4**, 1101-1112, doi:10.1038/s42256-022-00555-8 (2022年)。