嵌入到控制：局部线性延迟原始图像控制动力学模型

曼努埃尔·瓦特* Jost Tobias Springenberg*
Joschka Boedecker
德国弗赖堡大学
{Watterm，Springj，jboedeck}@cs.unifreiburg.de

马丁·里德米勒
谷歌深度思维
伦敦，英国
riedmiller@google.com

摘要

介绍了一种基于原始像素图的非线性动力系统模型学习与控制的嵌入控制方法。E2C 由一个深度生成模型组成，该模型属于变分自动编码器家族，它学习从一个潜在空间生成图像轨迹，其中动力学被约束为局部线性。我们的模型是直接从潜在空间中的最优控制公式导出的，支持对图像序列的长期预测，并在各种复杂控制问题上表现出很强的性能。

1 导言

具有连续状态和动作空间的非线性动力系统的控制是机器人技术中的关键问题之一，在更广泛的背景下，是增强自主代理学习的关键问题之一。针对这一问题的一类突出的算法是基于模型的局部最优（随机）控制算法，如 iLQG 控制，它通过局部线性化逼近一般非线性控制问题。当结合后退视界控制和机器学习方法来学习近似系统模型时，这些算法是解决复杂控制问题的有力工具；然而，它们要么依赖于已知的系统模型，要么需要设计相对低维的状态表示。为了使真正的自治代理成功，我们最终需要能够控制来自原始感觉输入的复杂动力系统的算法(例如。只有图像)。[1, 2][3][3,4, 5] 在本文中，我们解决了这个难题。

如果将随机最优控制(SOC)方法直接应用于原始图像数据的控制，它们将面临两个主要障碍。首先，感官数据通常是高维的。具有数千像素的图像在计算上不可行地呈现朴素 SOC 解。其次，图像内容通常是观测系统动力学的高度非线性函数；因此，模型识别和控制这种动力学是不平凡的。

[6, 7] 虽然这两个问题原则上都可以通过设计更先进的 SOC 算法来解决，但我们对“原始图像的最优控制”问题的处理方式是不同的：将高维非线性系统中的局部最优控制问题转化为识别低维潜态空间的问题，其中局部最优控制可以容易地执行。为了学习这样一个潜在空间，我们提出了一个新的深生成模型，属于变分自动编码器，它是从潜在空间中的 ILQG 公式导出的。由此产生的嵌入到控制(E2C)系统是一种概率生成模型，它对感官空间中可行的轨迹持有信念，允许在潜在空间中进行准确的长期规划，并且被完全无监督地训练。我们证明了我们的方法在四个具有挑战性的任务上的成功，从原始图像控制，并将其与一系列无监督表示学习的方法进行比较。此外，我们还验证了深层卷积网络是大图像的强大生成模型。[8, 9]

作者的贡献是一样的。

2 嵌入到控制(E2C)模型

简要回顾了动力系统 SOC 问题，介绍了潜在空间近似局部最优控制，并对模型进行了推导。

2.1 问题的提出

我们考虑了形式未知动力系统的控制

$$^{s}\ 不+我=/^{(s)}\ t，m)+\&£\backsim^{n}(0,习£)， \quad (1)$$

其中 t 表示时间步骤，stGR^{ns} 系统状态、应用控制和£系统噪声。 函数 f(st，ut)是一个任意的、平滑的系统动力学。 我们用表示法 P(st)等效地引用方程(1)。我|st，uj，我们假设它是一个多元正态分布 N(f(st，ut)，习 Q。 我们进一步假设，我们只允许访问视觉描绘 xt 电子　在圣州。这种限制需要解决一个联合状态识别和控制问题。

为了简单起见，我们将在下面假设 xt 是对 st 的完全观察到的描述，但稍后放松这个假设。

然后，我们的目标是推断一个低维潜态空间模型，其中可以执行最优控制。也就是说，我们寻求学习一个函数 m，将高维图像 xt 映射到具有 n 的低维向量 zte$_{z}$《$_{x}$ 从而可以用 zt 代替 Xt：来解决控制问题

$$zt=m(xt)+3，3n(0，：S3)， \quad (2)$$

其中 3 解释了系统噪声；或等效的 ztn(m(Xt)，S3)。 假设目前可以学习（或近似）这样的函数，我们将首先在一个潜在空间中定义 SOC，然后介绍我们的模型。

2.2　潜空间中的随机局部最优控制.

设 zte 是从状态 st 和 f 的图像 xt 中推断出的潜在状态^{拉特}(zt，ut)潜在空间中的跃迁动力学，即 zt。我$=f^{拉特}$(zt，ut)。 因此 f^{拉特}模拟 z 中发生的变化$_{t}$当控制你的时候$_{t}$作为 f(s)的潜在空间模拟应用于底层系统$_{t}$，U$_{t}$)。 假设是 f^{拉特}已知，可以通过最小化函数 J(Z)来导出动力系统中长度为 T 的轨迹的最优控制$_{t:t}$，U$_{t:t}$ 它给出了预期的未来成本时，以下(Z1：T，U1：T)：

$$J(zi：T，UI：T)=E_{z}\overline{CT(Z)_{t}^{T1}}，\overline{U_{t})}+£c(zt，叫)， \quad (3)$$

其中 c(zt，ut)是瞬时成本，ct(zt，ut)表示终端成本，zi：t=(zi，...，zt}和 ui：t=(ui，...，ut}分别是状态序列和动作序列。如果 zt 包含关于 st 的足够信息，即 st 可以单独从 zt 和 f 推断^{拉特}[10] 是可微的，成本最小化控制可以通过 SOC 算法从 J(ZI：T，UI：T)中计算。 这些最优控制算法在每个时间步长 t 上用局部线性动力学近似全局非线性动力学。然后可以以封闭的形式找到局部最优动作。正式地，给定一个参考轨迹 zi：t，最优轨迹的当前估计与相应的控制 UI：t，系统被线性化为

$$+我=A(Zt)zt+B(Zt)ut+o(Zt)+3，3n(0，S3)， \quad (4)$$

其中 A(zQ=^{f}，B(zQ=^{f}崇？—是当地的雅可比人，o(Zt)是一个偏移量。 为了有效地计算局部控件，我们假设成本是潜在表示的二次函数

$$c(zt，ut)=(ztzgoal)^{t}你+Rz(ztzgoal)?军士， \quad (5)$$

在那里 Rze ^{xnz}和 R$_{a}$e ^{xnu}是成本加权矩阵和 z ^{果阿}我是推断目标状态的表示。 我们也假设 c？？(z) 你呢？)=c(z)？ 你呢？(4)[11][12])在这篇论文中。 结合方程，给出了每个时间步长 t 的局部线性二次高斯公式，该公式可以用 SOC 算法求解，如迭代线性二次调节(ILQR)或近似推理控制(AICO)。 该轨迹优化步骤的结果是具有相应控制序列(z[？，E：)的局部最优轨迹？)*Arg Minzi$_{U1：T}$J(zi：？，UI：？)。

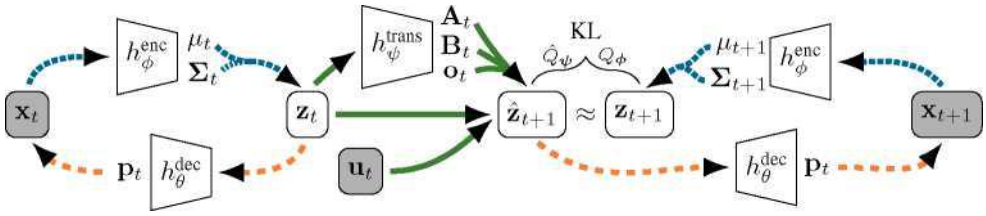


图 1：E2C 模型中的信息流。从左到右，我们编码和解码图像 x(与网络 h#ᶜ 和 h*ᶜ 其中我们使用潜在代码 z（用于转换步骤）。h 捫ˢ 网络计算局部矩阵 At，Bt，o，我们可以用它来预测 z 不+1 从 ZT 强制 KL 散度对它们的分布和

和你 ₜ。 与编码 z 相似不+我 重建再次由 hQ 执行 EC.

2.3 动力系统局部线性潜态空间模型

从 SOC 公式开始，我们现在讨论学习 XT 的适当的低维潜在表示 ztP(ZJm(XT))的问题。 表示 zt 必须满足三个属性：（1）它必须捕获关于 xt 的足够信息（足以实现重建)；(2）它必须允许对下一个潜在状态 zt 进行精确预测.因此，隐式地，下一个观测的 XT+I；(III)预测 f 拉特 对于所有有效的控制幅度 ut，下一个潜在状态必须是局部可线性的。给定一些表示 zt，属性(II)和(III)特别要求我们捕获潜在表示的可能高度非线性的变化，因为由控制命令引起的观测场景的转换。 关键的是，这些都是特别难建模和随后线性化。我们通过采取更直接的方法来规避这个问题：而不是学习潜在空间 z 和过渡模型 f 拉特 然后将其线性化并与 SOC 算法相结合，我们在学习过程中直接将期望的变换属性施加到表示 ZT 上。 我们将选择这些性质，这样在潜在空间中的预测以及根据方程对下一次观测的局部线性推断是容易的。(4)

我们希望从潜在表示中得到的转换属性可以直接从本节中给出的 iLQG 公式中形式化。 形式上，遵循方程，让潜在表示为高斯 P(Z\X)=N(M(XJ，)。 为了从 xt 中推断 z，我们首先需要一种采样潜态的方法。 理想情况下，我们将直接从未知的真实后验 P(Z\X)生成样本，然而，我们无法访问该样本。 遵循变分 Bayes 方法(见 Jordan 等人)。 2.2(2)[13]为了概述)，我们采用白参数的近似后验分布 Q^(Z\X)采样 z。

Q©的推理模型。 在我们的工作中，这总是一个对角高斯分布 Q©(Z\X)=N(口 t，diag(B2)，其平均值门 t E 协方差 ST=diag(B2)e xnz 是计算出来的通过具有输出的编码神经网络

$$太=WSF(XT)+b_m, \tag{6}$$

$$日志\ g=WBh\ 源(X\ T)+歸, \tag{7}$$

在哪里©ᶜ*e* R^{Nc} 是最后一个隐藏层的激活，其中©由编码网络的所有可学习参数集给出，包括权重矩阵 W_m，w_{ct} 和偏见 b_m，bb。基于神经网络对高斯分布的均值和方差进行参数化，为我们的潜在空间提供了一个自然和非常有表现力的模型。此外，它还带来了这样的好处，即我们可以利用再参数化技巧，通过潜在分布，基于样本的损失函数的反向传播梯度。[6, 7]

P 的生成模型。。 利用近似后验分布 Q©根据方程在潜在空间中强制执行局部线性关系，从潜在样本 z 和 zt+i 中生成观察到的样本（图像）x 和 xt+i，得到如下生成模型(4)

$$\begin{aligned} ZT\ \ q©(z\quad x) &= N(Mt,\ St), \\ z_{不+我} \sim q_v(z) \quad \backslash^{Z,\ u)} &= {}^{n(a)}太+了\ {}^{b}t^{u}t \quad +{}^{o}t,\ CT), \\ XT,\ XT+I \quad Pq(X)\quad z) &= Bernoulli(pt), \end{aligned} \tag{8}$$

Q 年是下一个潜在的地方 状态后验分布，它完全遵循线性形式 Re 用于随机最优控制。 用 3tN(0，HT)作为系统噪声的估计，

可将 C 分解为 CT=. 请注意，当过渡动态在我们生成模型操作在推断的潜在空间上，它考虑到未转换的控件。 也就是说，我们的目标是学习一个潜在的空间，使得 z 中的过渡动力学将 x 中的非线性观测动力学线性化，并且在应用控制中是局部线性的。从 zt 中重建图像是通过将样本通过解码神经网络的多个隐藏层进行的，该网络计算生

成的 Bernoulli 分布的平均 ptn[1]*(X\Z) 作为

$$Pt=Wph^{*c}\ (zt)+bp, \tag{9}$$

在那里嗡。 G 是解码网络中最后一个隐藏层的响应。 解码网络的一组参数，包括权重矩阵 W_p 和偏见 b_p 然后组成所学习的生成参数 0。

Q 伞的过渡模型。剩下的是指定如何在 GR 处的线性化矩阵新西兰 xnz，Bt G R 新西兰 XNU 和偏移的 GR㎡是预测的。 遵循与分布均值和协方差矩阵相同的方法，我们根据隐藏的表示 h 捫预测样本 ZT 的所有局部变换参数 s(zt)G R 新罗具有参数 W 的第三神经网络，我们称之为变换网络。 具体地，我们将变换矩阵和偏移量参数化为

$$\begin{aligned} \text{vec[at]} &= \text{华 h}^{\wedge ns}(zt)+b_{a}, \\ \text{维克[Bt]} &= {}^{\wedge ns}(zt)+b\ B, \\ \text{不} &= \text{Woh}^{\wedge}a^{ns}(zt)+bo, \end{aligned} \tag{10}$$

其中 vec 表示矢量化，因此 vec[AJGR(n Z)和 VEC[Bt]GR(nz)' nu)。 绕过估计尺寸 n 的全矩阵 。 *x n*。我们可以选择它是=(I+vtrf)恒等矩阵的扰动，它将 AT 的估计参数降低到 2n。.

1. 完整体系结构的草图如图所示，它还可视化了一个额外的约束，这对于学习长期预测的表示是必不可少的：我们要求来自状态转换分布的样本 zt+1Q 伞类似于 xt 的编码.我通过 Q©。虽然它看起来只是学习了一个完美的重建 xt。我来自 ZT。我已经足够了，我们需要在 Z 中进行多步预测，这必须对应于观测到的空间 X 中的有效轨迹。在不强制 Q 伞和 Q©样品之间的相似性的情况下，在潜在空间中发生转变，从 zt 到作用 ut 可能导致点 Zt.我，从哪个重建 XT.我是可能的，但这不是一个有效的编码(即。 该模型永远不会将任何图像编码为 ZT。i)。 在 Zt 执行另一个操作.然后，我不会导致一个有效的潜在状态，因为过渡模型取决于来自推理网络的样本-因此长期预测失败。 简而言之，编码和过渡模型之间的这种差异导致了一个生成模型，该模型不能准确地模拟观测形成的马尔可夫链。

2.4 通过随机梯度变分贝叶斯学习

为了训练模型，我们使用数据集 D=(x1，u1，x？)，.....t1，Ut1，xt)}包含观测元组和从与动力系统的相互作用中获得的相应控制。利用该数据集，我们通过最小化真数据负对数似然对数 P(xt，ut，xt+i)上的变分界，并加上对潜在表示的附加约束来学习推理、过渡和生成模型的参数。 完全损失的功能 n[2]

$$l(d)= \underset{(十)_{t}\mathbf{u}_{t}\mathbf{x}_{t+1})e\ D}{£} l^{束缚}(xt,\ ut,\ xt+i)+入 KL(Q 半(Z\m \|Q©(Z5))。 \tag{11}$$

这个损失的第一部分是对数似然上的每个样本变分界

£绑定(xt，ut，xt+1)=Eztq，[logPe(xt\zt)logP。2.3(11) 2.3(xt+i\zt+i)]+KL(Q©\\P(Z)，（12），其中 Q©、Pe 和 Q 仲是参数推断、生成和过渡分布，P(Zt)是近似后验 Q©的先验分布，我们总是选择它是平均零和单位方差的各向同性高斯分布。 方程中的第二个 KL 散度是一个具有权重入的附加收缩项，它强制了过渡模型和推理模型之间的一致性。这个术语对于在与实际系统动力学相对应的潜在空间中建立马尔可夫链是必不可少的（深入讨论见上文）。 这种 KL 散度也可以看作是潜在过渡模型的先验。 请注意，所有 KL 项都可以解析地计算我们的模型（详见补充）。

在训练过程中，我们通过抽样逼近 L(D)中的期望。 具体来说，我们为每个输入取一个样本，并使用方程对该样本进行转换，以给出一个有效的样本 Zt+我来自仲。然后，我们通过使用 SGD 最小化 L(D)来联合学习模型的所有参数。(10)

3 实验结果

我们在四个视觉任务上评估我们的模型：一个有障碍物的平面上的代理，一个经典倒立摆摆动任

[1]在建模黑白图像时，Pq 的 Bernoulli 分布是一个常见的选择。

[2]请注意，这是潜在状态空间模型的损失，与 SOC 成本不同。

务的视觉版本，平衡一个推车-杆系统，以及控制一个具有更大图像的三连杆臂.. 下文将详细说明这些问题。

3.1 实验装置

模型训练。我们考虑了我们的模型的两种不同的网络类型：标准的全连通神经网络，最多有三层，对中等大小的图像很好地工作，用于平面和摆动实验；编码器的深卷积网络与卷积网络相结合，作为解码器，根据文献的最新发现，我们发现这是一个适合较大图像的模型。 训练是在所有实验中使用亚当进行的。 所有任务的训练数据集 D 是通过随机抽样 N 状态观测和具有相应继承国的动作生成的。 对于平面，我们使用 N=3000 个样本，对于倒立摆和推车杆系统，我们使用 N=15，000，对于臂 N=30，000。 补充材料中列出了体系结构参数和超参数选择的完整列表以及对卷积网络的深入解释。 我们将制作我们的代码和一个包含控制轨迹的视频，所有系统都可以在 http://ml.informatik.uni-freiburg.de/research/e2c 下使用。[8, 9][14]

模型变体。 除了上面导出的嵌入到控制(E2C)动力学模型外，我们还考虑了两个变体：通过删除潜在动力学网络 h^ns ，即 将其输出设置为方程中的一个，我们得到一个变体，其中 A(，(10)

并被估计为全球

线性矩阵(全局 E2C)。 如果我们代替过渡模型，用一个网络估计动力学作为一个非线性函数_f 拉特 2.2， 并且只有在规划过程中线性化，估计 At，Bt，ot 为 Jacobians，如本节所述，我们得到了一个具有非线性潜在动力学的变体。 **基线模型。** 为了进行彻底的比较，并表现出任务的复杂性，我们还测试了平面上的一组基线模型和倒立摆任务(使用与 E2C 模型相同的体系结构)：在视觉问题的自动编码子任务上训练标准变分自动编码器(VAE)和深度自动编码器(AE)。 也就是说，给定一个用于训练模型的数据集 D，我们从 D 中的元组中删除所有操作，并且忽略图像之间的时间上下文。在自动编码器训练后，我们学习了一个潜在空间的动力学模型，近似于 f 拉特 2.2. 从章节中，我们还考虑了一个 VAE 变体，在潜在的表示上有一个慢项-在补充材料中给出了这个变体的完整描述。

最优控制算法.. 为了在不同模型的潜在空间中执行最优控制，我们采用了两种轨迹优化算法：迭代线性二次调节(iLQR)（用于平面和倒立摆）和近似推理控制(AICO)（所有其他实验）。 对于所有 VAE，这两种方法都对分布 Q 和分布的平均值起作用。此外，AICO 还利用了局部高斯协方差 ST 和 CT。 除了在平面系统上的实验外，控制是以模型预测控制的方式进行的，采用了引入的后退视界方案。 以获得给定图像 x 的闭环控制[11][12][3],首先通过编码器获得潜在状态 ZT。 然后通过优化(z；)找到局部最优轨迹。+T，u`_t+t)*ArgMinz，+T 具有固定、小视界■J-T 的 J(Ztt+T，ut：t+T)

t+T 丿 t1+T/.　　o　　1　　\　i 丿 03i/　　7

Ut：不+T

T(与 T=10，除非另有说明)。 控件 u；应用于系统，并观察到向 zt+i 的转换(通过编码下一个图像 xt+i)。 然后是一个新的控制序列与地平线

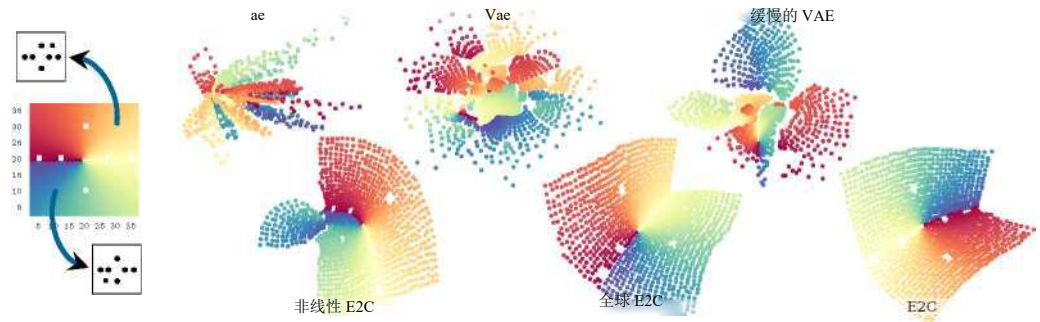


图 2：平面系统的真实状态空间（左)与例子(障碍编码为圆圈)和推断空间(右）的不同模型。 通过为代理的每个有效位置生成图像并将它们嵌入到相应的编码器中，空间被跨越。

从 z 开始$_{t+1}$ 是用最后一个估计的轨迹作为引导找到的。请注意，规划完全是在潜在状态下执行的，除了对当前状态的描述外，没有访问任何观察。 计算成本函数 c(z)$_t$ 对于 z 中轨迹优化所需的 UJ，我们假设目标状态 Sgoal 的观测 xgoal 的知识。 然后将这一观测转化为潜在空间，并根据方程计算成本。(5)

3.2 平面系统中的控制

平面系统中的代理可以通过在 x 和 y 方向上选择连续偏移来在有界二维平面上移动。 状态的高维表示是 40x40 黑白图像.. 由六个圆形障碍物阻碍，任务是移动到图像的右下角，从图像顶部的随机 x 位置开始。 障碍的编码是在规划之前获得的，另一个二次成本项是惩罚接近它们。

2. 图中显示了控制的观测结果及其相应的状态值，并将其嵌入到潜在空间中。该图还清楚地显示了 E2C 模型相对于其竞争对手的一个基本优势：虽然单独训练的自动编码器为美观的图片制作，但这些模型未能发现状态空间的底层结构，使动力学估计复杂化，并且根据所述空间的距离在很大程度上使成本无效。 另一方面，在这些端到端模型中包括潜在动力学约束，产生接近最优平面嵌入的潜在空间。

1 我们通过积累潜在和真实的轨迹成本来检验长期精度，以量化想象的轨迹是否反映现实。 所有模型从顶部的随机位置开始并执行 40 个预先计算的操作时的结果汇总在表中，使用单独的测试集来评估重建。 虽然所有方法都实现了较低的重建损失，但每个轨迹累积实际成本的差异表明了 E2C 模型的优越性。利用全局或局部线性 E2C 模型，潜在空间规划的轨迹与实际状态规划的轨迹一样好。 除了 E2C 之外，所有的模型都没有给出长期的预测，从而导致良好的性能。

3.3 学习倒立摆的摆动

[15] 3(接下来，我们将从图像中控制经典倒立摆系统。 我们通过绘制一条固定长度的线，从图像的中心开始，以与摆位对应的角度来描绘状态。 这项任务的目标是摆动和平衡一个欠驱动的钟摆从一个静止的位置（钟摆悬挂下来）。 图中给出了该系统的示例观测和重建)。 在视觉倒立摆任务中，我们的算法面临两个额外的困难：观测到的空间是非马尔可夫的，因为角速度不能从单个图像中推断出来，第二，由于渲染摆角的离散化误差作为小的 48x48 像素图像使得精确控制变得困难.. 为了恢复马尔可夫属性，我们叠加了两个图像（作为输入通道），从而观察了一步历史。

3 图显示了模型的潜在空间的拓扑结构，以及真实状态和潜在空间中的一个样本轨迹。 事实上，模型可以学习一个有意义的嵌入，分离

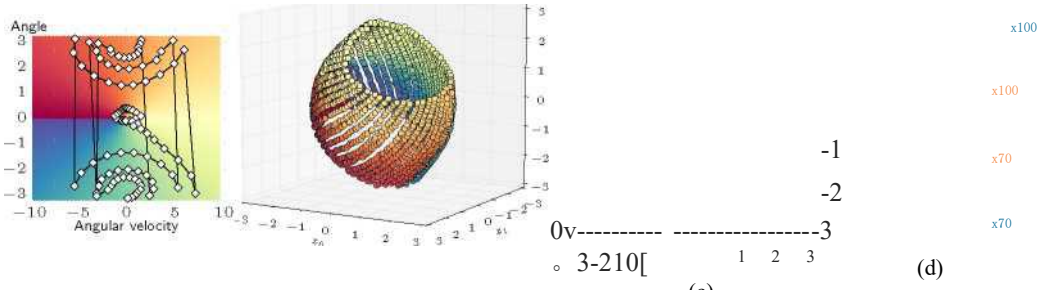
表 1：平面和摆系统从原始像素学习模型的不同方法的比较。 我们将所有模型与它们在抽样过渡测试集上的预测质量进行比较，并将它们与 SOC（不同启动状态下控制的轨迹成本）相结合时的性能进行比较。 请注意，潜在空间中的轨迹成本不一定具有可比性。 在执行计划动作的同时，根据仿真器的动力学计算出“真实”轨迹成本。 为 s 的真实模型。 实际轨迹成本为 20。 24 士 4。 平面系统 15 个，9 个。 8 士 2. 钟摆的 4。 成功被定义为达到目标状态，并在轨迹的其余部分（如果不是终止）保持电子接近。 所有统计数字量化超过 5/30（平面/摆）不同的起始位置。 一个 f 标记单独训练的动力学网络。

算法	状态损失日志 p(xt|xt)	下一个州损失日志 p(xt+1|xt，ut)	弹道费用 暂时的	是真的	成功 百分比
		飞机系统			
AEt	11.5 士 97.8	3538.9 士 1395.2	1325.6 士 81.2	273.3 士 16.4	0%
Vae*	3.6 士 18.9	652.1 士 930.6	43.1 士 20.8	91.3 士 16.4	0%
VAE+缓慢*	10.5 士 22.8	104.3 士 235.8	47.1 士 20.5	89.1 士 16.4	0%
非线性 E2C	8.3 士 5.5	11.3 士 10.1	19.8 士 9.8	42.3 士 16.4	96.6%
全球 E2C	**6.9 士 3.2**	**9.3 士 4.6**	12.5 士 3.9	27.3 士 9.7	**100%**
e2c	7.7 士 2.0	9.7 士 3.2	10.3 士 2.8	**25.1 士 5.3**	**100%**
		倒立摆摆动			
ae*	8.9 士 100.3	13433.8 士 6238.8	1285.9 士 355.8	194.7 士 44.8	0%
Vae*	7.5 士 47.7	8791.2 士 17356.9	497.8 士 129.4	237.2 士 41.2	0%
VAE+缓慢*	26.5 士 18.0	779.7 士 633.3	419.5 士 85.8	188.2 士 43.6	0%
E2C 无潜在 KL	64.4 士 32.8	87.7 士 64.2	489.1 士 87.5	213.2 士 84.3	0%
非线性 E2C	**59.6 士 25.2**	**72.6 士 34.5**	313.3 士 65.7	37.4 士 12.4	63.33%
全球 E2C	115.5 士 56.9	125.3 士 62.6	628.1 士 45.9	125.1 士 10.7	0%
e2c	84.0 士 50.8	89.3 士 42.9	275.0 士 16.6	**15.4 士 3.4**	**90%**

1 速度和位置，从这个数据是显著的（没有其他模型恢复这个形状）。 表再次定量地比较了不同的模型。虽然 E2C 模型在重建性能方面不是最好的，但它是唯一导致稳定摆动和平衡行为的模型。我们用这样一个事实来解释其他模型的失败，即非线性潜在动力学模型不能保证对所有控制幅度都是线性化的，从而导致实际系统动力学的不稳定固定点周围的不理想行为，并且对于这一任务，全局线性动力学模型是不够的。

3.4 平衡推车杆和控制模拟机器人手臂

最后，我们考虑使用六层卷积推理和六层上卷积生成网络来控制两个更复杂的动态系统，从而产生从输入到重建的 12 层深路径。 具体来说，我们从两个 80x80 像素图像的历史以及基于两个 128x128 像素图像的历史的三连杆平面机器人臂中控制经典 Cartpole 系统的视觉版本。 在这两个



(a) (b)

实验中，潜在空间被设置为 8 维。 车杆的实际状态维数为四，用一控制[16]

图 3：(A)倒立摆任务的真实状态空间覆盖了 E2C 代理所采取的成功轨迹。(b)学习的潜在空间。(c)(a)在潜在空间中追踪的轨迹。 (d)显示当前位置（右)和历史(左）的图像 x 和重建 x。

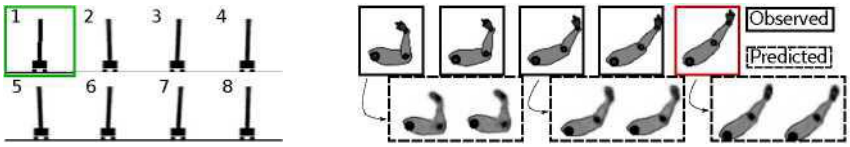


图 4：左：来自购物车极域的轨迹。 只有第一个图像（绿色）是“真实的”，所有其他图像都是由我们的模型“梦想”的。注意实际图像中存在的离散化工件。右：在视觉机器人臂域中观察到（历史图像省略)和预测图像(包括历史图像）的轨迹，目标标记为红色。

动作，而对于手臂，真实状态可以用 6 维（关节角度和速度）来描述，并使用与电机扭矩对应的三维动作矢量来控制。

与以前的实验一样，E2C 模型似乎没有问题，可以在潜在空间中找到可以执行控制的图像的局部线性嵌入。 从我们的系统执行的轨迹中绘制出两个问题的示例性图像。 这些轨迹的成本（11）。 4 手推车杆 13，85。对于手臂)，仅比在从相同的起始状态（7）开始的真实系统动力学上操作的 AICO 获得的轨迹略差。 28 和 60。74 分别)。 补充材料包含使用这些领域的额外实验。

4 与最近的工作比较

[17][18] 在表示学习的背景下进行控制(见 Bohmer 等人)。为了回顾)，类似于我们的基线模型的深度自动编码器（忽略状态转换）以前已经被应用过，例如。 兰格和里德米勒。 最近 Mnih 等人关于 Atari 游戏（模型免费）深度端到端 Q 学习的工作采取了一条更直接的基于图像流的控制途径。[19][20] 以及基于内核和深度策略的机器人控制学习。[21]

[22]与我们的方法接近的是 Wahlstrom 等人最近的一篇论文。 其中，自动编码器用于从图像中提取用于控制的潜在表示，在此基础上学习了正向动力学的非线性模型。它们的模型是联合训练的，因此在我们的比较中类似于非线性 E2C 变体。与我们的模型相反，它们的公式需要 PCA 预处理，既不确保潜在空间中的长期预测不会发散，也不确保它们是可线性的。

如上所述，我们的系统属于 VAE 家族，通常类似于最近的工作，如 Kingma 和 Welling，Rezende 等人。 [6][7]，Gregor 等人。 [23][24] ，拜耳和奥森多夫。 我们的工作和最近的进展之间的两个额外的平行训练深度神经网络可以观察到。 首先，在学习过程中在潜在空间中强制执行所需的转换的想法-使得数据变得易于建模-已经在文献中多次出现。 这包括开发转换自动编码器和最近的图像概率模型。 第二，在过去的几年里，图像对之间的学习关系虽然没有控制，但受到了社区的相当关注。 [25][26,27][28, 29] 在更广泛的背景下，我们的模型与马尔可夫决策过程中的状态估计工作有关(见 Langford 等人)。 [30] 讨论)通过，例如，隐马尔可夫模型和卡尔曼滤波器。[31, 32]

5 结论

提出了一种高维图像流随机最优控制系统 E2C 嵌入。该方法的关键是提取潜在动力学模型，该模型在状态转换中被限制为局部线性。 对四个具有挑战性的基准的评估表明，E2C 可以找到可以轻松执行控制的嵌入，通过对实际系统模型的最优控制达到接近可实现的性能。

致谢

我们感谢 A.Radford、L.Metz 和 T.De Wolf 分享代码，并感谢 A.Dosovitskiy 进行有益的讨论。 这项工作的部分资金来自 DFG 在优先项目“自主学习”(SPP1597)和大脑链接-大脑工具卓越集群(赠款编号 EXC1086)中的赠款。 M.Watter 是通过巴登-符腾堡州研究生资助计划资助的。

参考资料

[1] D.雅各布森和 D.梅恩。 差分动态规划。 美国 Elsevier，1970 年。

[2] E.托多罗夫和李伟。约束非线性随机系统局部最优反馈控制的广义迭代 LQG 方法。在行政协调会。IEEE，2005。

[3] 塔萨 Y.，艾雷兹和 W.D.斯马特。 后退视界差分动态规划。 在检察官办公室。 国家统计局，2008 年。

[4] 潘 Y.和西奥多罗。 概率微分动态规划.. 在检察官办公室。 国家统计局，2014 年。

[5] S.Levine 和 V.Koltun。 通过轨迹优化进行变量策略搜索。 在检察官办公室。 国家统计局，2013 年。

[6] D.P·金马和 M·韦林。 自动编码变分贝叶斯。 在检察官办公室。 ICLR，2014 年。

[7] D.J.Rezende、S.Mohamed 和 D.Wierstra。 深层生成模型中的随机反向传播和近似推理。 在检察官办公室。 ICML，2014 年。

[8] M.D.Zeiler，D.Krishnan，G.W.Taylor 和 R.Fergus。 反褶积网络。 在 CVPR，2010 年。

[9] A.Dosovitskiy、J.T.Springenberg 和 T.Brox。 学习用卷积神经网络生成椅子。 在检察官办公室。 CVPR，2015 年。

[10] *R.F.斯坦格尔。 最佳控制和估计。 多佛出版物，1994 年。*

[11] 李 W.和 E.Todorov。 非线性生物运动系统的迭代线性二次型调节器设计。 在检察官办公室。 ICINCO，2004 年。

[12] M.杜桑。 基于近似推理的机器人轨迹优化。 在检察官办公室。 ICML，2009 年。

[13] M.I.Jordan、Z.Ghahramani、T.S.Jaakkola 和 L.K.Saul。 图形模型的变分方法介绍。 《机器学习》，1999 年。

[14] D.金马和巴。 亚当：一种随机优化的方法。 在检察官办公室。 ICLR，2015 年。

[15] H.王，田中和格里芬。 非线性系统模糊控制的一种方法；稳定性和设计问题。 IEEE Trans. 关于模糊系统，4（1），1996 年。

[16] R.萨顿和巴托。强化学习导论。麻省理工学院出版社，剑桥，美国，1stedition，1998 年。ISBN0262193981。

[17] W.Bohmer、J.T.Springenberg、J.Boedecker、M.Riedmiller 和 K.Obermayer。 用于控制的状态表示的自主学习。 KI Kunstliche Intelligenz，2015 年。

[18] S.兰格和里德米勒先生。深度自动编码器神经网络在强化学习中的应用。在检察官办公室。国际 JCNN，2010 年。

[19] V.Mnih，K.Kavukcuoglu，D.Silver，A. A.Rusu、J.Veness、M.G.Bellemare、A.Graves、M.Riedmiller、A.K.Fidjeland、G.Ostrovski、S.Petersen、C.Beattie、A.Sadik、I.Antonoglou、H.King、D.Kumaran、D.Wierstra、S.Legg 和 D.Hassabis。 人级控制通过深度强化学习.. 自然，518（7540），2015 年 02 月。

[20] 范霍夫 H.，彼得斯和诺依曼。 具有高维状态特征的非参数控制策略的学习.. 在检察官办公室。 澳大利亚，2015 年。

[21] S.Levine，C.Finn，T.Darrell 和 P.Abbeel。 深度视觉运动政策的端到端培训。 共同 RR，abs/1504.00702，2015 年。 网址 http://arxiv.org/abs/1504.00702.

[22] N.Wahlstrom、T.B.Schon 和 M.P.Deisenroth。 从像素到扭矩：具有深度动力学模型的策略学习。 共同 RR，abs/1502.02251，2015 年。 网址 http://arxiv.org/abs/1502.02251.

[23] K.Gregor、I.Danihelka、A.Graves、D.Rezende 和 D.Wierstra。 DRAW：用于图像生成的递归神经网络。 在检察官办公室。 ICML，2015 年。

[24] J.拜耳和 C·奥森多夫。 学习随机递归网络.. 2014 年在 NIPS 举行的差异推断进展讲习班。

[25] G.Hinton，A.Krizhevsky 和 S.Wang。 转换自动编码器。 在检察官办公室。 ICANN，2011。

[26] L.Dinh，D.Krueger 和 Y.Bengio。 尼斯：非线性独立分量估计。 共同 RR，abs/1410.8516，2015 年。 网址 http://arxiv.org/abs/1410.8516.

[27] 科恩 T.和韦林先生。 学习的视觉表示的转换属性。 在 ICLR，2015 年。

[28] G.W.Taylor，L.Sigal，D.J. 舰队和辛顿将军。 三维人体姿态跟踪的动态二元潜在变量模型。 在检察官办公室。 CVPR，2010 年。

[29] R.Memisevic。 学会联系图像。 IEEE Trans. 关于 PAMI，35（8）：1829-1846，2013 年。

[30] J.朗福德，萨拉胡特迪诺夫和张。 学习非线性动力学模型.. 在 ICML，2009 年。

[31] *M.和哈里森。 贝叶斯预测和动态模型(统计中的 Springer 系列)*。 斯普林格-维拉格，1997 年 2 月。 ISBN0387947256。

[32] T.Matsubara，V.Gomez 和 H.J.Kappen。使用概率图形模型的连续状态系统的潜在 KullbackLeibler 控制。阿联酋，2014 年。

[33] T.D.Kulkarni，W.Whitney，P.Kohli 和 J. B.Tenbaum。 深卷积逆图形网络.. 共同 RR，abs/1503.03167，2015 年。 网址 http://arxiv.org/abs/1503.03167.

[34] C.Osendorfer，H.Soyer 和 P.van der Smagt。 具有快速近似卷积稀疏编码的图像超分辨率。 在检察官办公室。 计算机科学讲座笔记。 斯普林格国际出版社，2014 年。

[35] R.Jonschkowski 和 O.Brock。机器人学中的状态表示学习：使用关于物理交互的先验知识。在检察官办

公室。 RSS，2014 年。

[36] 莱根斯坦 R.，威尔伯特和 Wiskott。 强化学习高维输入流的慢特征。 PLOS 计算生物学，2010 年。

[37] 邹 W.、吴国和余国强。 具有时间相干性的视觉不变性的无监督学习。 在 NIPS*2011 深度学习和无监督特征学习讲习班，2011 年。

[38] A.M.Saxe，J.L.Mc Clelland 和 S.Ganguli。 深度线性神经网络学习非线性动力学的精确解。 在检察官办公室。 ICLR，2014 年。

[39] X.格洛洛，波德斯和本戈。 深度稀疏整流神经网络.. 在澳大利亚。 机器学习研究杂志-研讨会和会议记录，2011 年。

对 E2C 描述的补充

A. 1 状态转移矩阵分解和 KL 发散

如本文所述，在 E 处估计全局部态过渡矩阵 xnz
从方程（8）需要过渡网络来预测 n_z x n_z参数。 使用任意状态转换矩阵也不方便地需要反演上述矩阵来计算方程（11）中的 KL 散度惩罚（通过它很难反向传播）。 我们开始使用一个完整的矩阵(并且只近似于所有 KL 散度项)进行实验，但很快发现，在我们的任何基准中，可以使用身份矩阵的秩一扰动来代替，而不会失去性能。 相反，由此产生的网络参数较少，因此更容易训练。 我们在这里给出了这个过程的推导，以及如何计算方程（11）的 KL 散度。 对于我们所表示的 ATASA=I+vtrf，因此只需要由过渡网络估计 vt 和 RT，将 AT 的输出数量从 nZ 减少到 2nz。

两个多元高斯分布之间的 KL 散度由.

KL(M|I1)=2 ^Tr（斗 ¹ &)⁺（何一总部）T 莫) k+日志 （刃十尸 1))(13)
2\ ⊥ q ノ ノ

对于简化的表示法，使 KL(NQ||N1)=KL(Q||Q)，让我们假设

无=N(MQ，AXQA)T)=N(Mt，at^t Af)=Q，
n 1=" M1，习 i)="太+我，习 t+我)=Q。

下面给出的推导的要点是使上述 KL 散度的偏导数有效地可计算。 为此，我们不能通过数值算法来取迹或行列式，因为我们必须能够以符号形式取梯度。 除此之外，我们喜欢处理一批样本，因此计算应该有一个方便的形式，而不需要过多的张量积在两者之间。 我们从简化开始

导致的跟踪项

tr OS：，。)=Tr(X¹AXO A¹)
=Tr(S.⁻¹(我+VR¹所以(我+VR¹)¹)
=Tr(Si+S.¹vr¹所以+¹)¹))
=Tr(S.⁻¹所以+S⁻¹所以(vr.¹)¹ +⁻¹vr¹所以+S⁻¹vr¹所以(vr.¹)¹)
Tr(A+B)=Tr(A)+Tr(B)=Tr(S⁻¹ 所以)+Tr(S.⁻¹所以(vr.¹)¹)+Tr(S⁻¹vr¹ 所以)+Tr(S.⁻¹vrS°RV¹(美国广播公司=美国广播公司

=V 吳+V 竺 y 竺 V 竺察.¹ +Tr(v¹s⁻¹vrS°r)
v%v%v%▽ᵃ哦，我+2ˆ哦，我ᵛ1ⁱⁱ
v_死 i_

最后一个方程易于实现，只需要对非批量维度进行求和。用相同的求和方案可以很快地导出均值的差：.

(Mi Mo)¹ 莫)=五化”.

它仍然是行列式的比率，我们将用矩阵行列式引理来简化

)=log det si-log det(AS° A¹)
(德西
ᵒᵍ (Det ASO A¹ =日志 Ua！！ 我-日志(DetAdetA)¹) Det A¹ =DETA

2V 日志 ai，我日志(DETA)²n 吨]
两个 flogai，我日志(1+v¹r)² 2£记录一个° 矩阵行列式引理
二

² （五）⁽¹ᵒ gᵃ我-¹ᵒ gᵃ3，i)-¹ᵒg⁽¹ +v%ri)

把上面的公式放在一起，最终得到了结果

KL(MIIM) $\int$ 吨+V 哼 v 1 $\int$ $\int$ (14)
我，我
+v%-v²k
+² （五）⁽¹⁾。 gᵃ 2，i⁻洛 gᵃ不，我⁾ 托格(1+Vg))。

补充实验装置

B. 1 卷积

我们使用卷积推理网络进行购物车极和三链臂任务.. 虽然这些网络帮助我们克服了输入维度大的问题(即。 在三链臂任务中的 2x128x128 像素图像)，我们仍然必须用解码器网络生成全分辨率图像。对于高维图像生成，完全连接的神经网络根本不是一种选择。因此，我们决定使用上卷积网络，这些网络最近被证明是图像生成的强大模型。[8, 9， 33]

为了建立这些模型，我们基本上“镜像”了编码器使用的卷积体系结构。更具体地说，对于编码器网络中的每个 5x5 卷积，然后是 2x2 最大池化步骤，我们在解码器网络中引入了 2x2 上采样和 5x5 卷积步骤。下面给出完整的网络架构.. 它类似于 Dosovitskiy 等人使用的卷积网络。 [9] 。 我们使用的上采样策略是简单的“穿孔”上采样，如所述。[34]

b. 2 具有慢速的变分自动编码器

[35, 36][37] 在学习过程中强制时间缓慢以前被发现是学习表示在强化学习和表示学习视频中的一个很好的代理。 我们还考虑了一个 VAE 变体，在潜在表示上有一个慢项，通过强制执行时间接近图像编码的相似性。这可以通过增加标准的 VAE 目标 L 来实现束缚 在潜在后验 Q©上附加 KL 散度项：

$$l^{\text{慢一点}}(xt, xt_{+1}) = KL(Q©(Zt+i|xt+i)||Q©(Zt|xt))。 \quad (15)$$

8 事实上，在潜在的空间中，类似的状态似乎有一个稍微好的一致性，例如。 如主要论文中的图所示。 然而，我们的实验表明，仅仅一个缓慢项并不足以构造潜在空间，因此局部线性预测和控制变得可行。

b. 3 评价标准

为了比较 E2C 的所有变体和基线的性能，以下标准是重要的：

- 自动编码。 能够重建给定的观测值是模型工作的基本必要性。 重建成本驱动模型从其观测结果中识别单个状态。
- 解码下一个状态。 为了使任何计划完全可能，解码器必须能够从执行的动力学模型的转换中生成正确的图像。 如果不是这样，我们知道编码的潜在状态和过渡模型不一致，从而阻止了任何规划。
- 优化潜在轨迹成本。 实现指定目标的动作序列将完全由潜在空间中的局部线性化动力学决定。 因此，在潜在空间中最小化轨迹成本再次是成功控制的必要条件。
- 优化实际轨迹成本。 虽然已经确定了潜在动力学的动作序列，但决定的标准是这是否反映了真实的状态轨迹成本。因此，在现实中实施“梦想”计划是每个模型的最优性标准。为了使不同的模型具有可比性，我们使用相同的成本矩阵进行评估，这些矩阵不一定与优化相同。

我们在论文的评估表中反映了这四个标准。 对于当前和下一个状态的重建，我们指定了平均对数损失，在 Bernoulli 分布的情况下，交叉熵误差函数：

$$日志\ p(x|x) = 页 \underset{=1i=0}{\overset{1Nn_x}{}} \log\ x_n\ ，我+(1Xn，i)日志(1-£_{n}（一）。 \quad (16)$$

对于一个模型想象和真正实现的成本，我们从不同的起始状态采样，并根据 SOC 方法积累潜在和真实状态空间中的距离。

三连杆机器人臂

我们在上一次实验中使用的机器人手臂是用 MapleSim http://www.maplesoft.com/products/maplesim/模拟器在 Python 中生成的动力学模拟的，并可视化地使用 Py Game 生成对 E2C 的输入。 我们模拟了一个相当标准的机器人手臂，有三个环节。 链接的长度被设置为 2,1。 2 和 0。 7（单位定为米）。 相应环节的群众均设置为 10kg..

b. 5 评估真实的系统模型

为了比较不同模型与最优控制算法相结合的有效性，我们总是报告潜在空间的成本（如最优控制算法所使用的）以及“真实”轨迹成本。 为了计算这一实际成本，我们评估了与潜在空间相同的成本函数（二次成本对给定目标状态的偏差），但使用执行过程中的实际系统状态和不同的成本矩阵进行了公平的比较。

作为任何模型可以实现控制的性能的上限，我们还通过将 iLQR/AICO 应用于实际系统动力学模型来计算真实的系统成本。 我们有这个模型，因为所有的实验都是在模拟中进行的。

b. 6 神经网络训练

B.6.1 实验装置

所有数据集都是作为 D=((x1，x2)，.，(x)预先创建的 $_{t}u_{t}x_{t}$ }用于培训、验证和测试拆分。 虽然 E2C 模型是关于 D 的培训，但没有包含任何过渡信息的模型(即。对 AE，VAE)进行图像 Dimages=(x1，...，x 的训练 $_{t}$ 从原始数据集 D 中提取}。在图像子集 D 对=((X1，X2)，...，(X)上训练慢 VAE_{t_1}，x_{t} }和

我们的 E2C 模型上的全 D。

为了学习图像自动编码器的动力学预测，我们提取了潜在的表示，并将它们与从 D 到 D 的动作相结合 ₄动力学=(Z1，U1，Z2)，...，(zt_i，ut_i，ZT)}。在这些低维表示上，我们训练了动力学 MLPs，从而确保所有方法都在完全相同的数据上进行了训练。

B.6.2 实施细节

[38][14] 我们对每一层都使用正交权重初始化。 正如本文所描述的，亚当被用作所有网络的学习规则。我们发现这两种技术对于稳定训练和实现所有方法的良好重建都是至关重要的。这两种方法也显然有助于将所有方法所需的超参数搜索减少到最低限度。 在训练过程中，我们可以分为三个阶段：潜在空间的展开、平凡解的克服（数据集的平均图像）和潜在 KL 项的最小化。 用于我们的实验的体系结构如下(其中 ReLU 代表校正线性单元和卷积。 对于卷积)：[39]

飞机

- 投入：40² 图像尺寸，2 个动作尺寸
- 潜在空间维数：2
- 编码器：150ReLU150ReLU4 线性(2 用于 AE)
- 解码器：200ReLU200ReLU1600 线性(乙状结肠为 AE)
- 动力学：100ReLU100ReLU+输出层(全球 E2C 除外)
  AE，VAE，VAE，慢，非线性 E2C：2 线性
  E2C：8 线性(22 对 A(，21 对 B(，2 对 O（)，入=0.25）
- 亚当：=10_.⁴危 2=0。1
- 评估费用：R.₂ =0. 11，rᵤ=我，r。=一

摆摆摆

- 输入：2 个 48² 图像维度，1 个动作维度
- 潜在空间维数：3
- 编码器：800ReLU800ReLU6 线性(AE3)
- 解码器：800ReLU800ReLU4608 线性(乙状结肠用于 AE)
- 动力学：100ReLU100ReLU+输出层(全球 E2C 除外)
  AE，VAE，VAE，慢，非线性 E2C：3 线性
  E2C：12 线性(23 表示在=(I+寸曷)，31 表示 B(，3 表示 b()，入=0.25
- 亚当：=3。 10⁴危 2=0。1
- 评估费用：R.₂ =我，rᵤ=0. 1i

纸箱-杆平衡

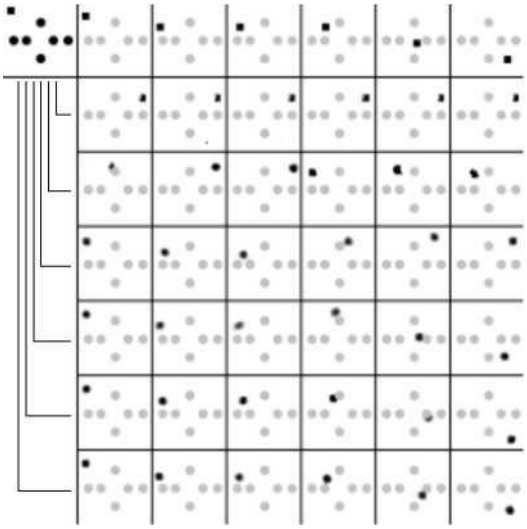
- 投入：280.² 图像维度，1 个动作维度
- 潜在空间维数：8
- 编码器：32x5x5ReLU32x5x5ReLU32x5x5ReLU512ReLU
- 解码器：512ReLU512ReLU2x2 上采样 32x5x5ReLU2x2 上采样 32x5x5ReLU2x2 上采样 32x5x5 卷积。 勒卢
- 动力学：200ReLU200ReLU+32 线性(2，8 用于=(I+vtrf)，8，1 用于 B(，8 用于 bt)，入=1
- 亚当：a=10.⁴危 2=0。1
- 评估费用：Rz=I，R.ᵤ =一

三连臂

- 输入：2 个 128² 图像尺寸，3 个动作尺寸
- 潜在空间维数：8
- 编码器：64x5x5conv.. 重新 LU2x2 最大池 32x5x5Conv.. 重新 LU2x2 最大池 32x5x5Conv.. 再路 2x2 最大池-512 再路-512 再路
- 解码器：512ReLU512ReLU2x2 上采样 32x5x5ReLU2x2 上采样 32x5x5ReLU2x2 上采样

64x5x5 卷积。 勒卢

- 动力学：200ReLU200ReLU+48 线性(28 在=(我+VQ？)，BT 的 83，BT 的 8)，入=1



- 亚当：a=10.⁴危 2=0。1
- 评估费用：Rz=I，R.ᵤ=0. 001i

真正的国家

ae

Vae

缓慢的 VAE

非线性 E2C

全球 E2C

e2c

图 5：为平面任务生成不同模型的“梦”轨迹（从左到右）。 在这种描述中，障碍物的不透明度已经降低，以提高代理的可见度。

C.补充评价

C. 平面和钟摆的 1 轨迹

为了定性地测量预测精度，对轨迹的起始状态进行编码，并将动作应用于潜在表示。每次转换后，对预测的潜在位置进行解码和可视化。 这样，可以对图中的平面系统和图中的倒立摆产生多步预测 5 6 7.

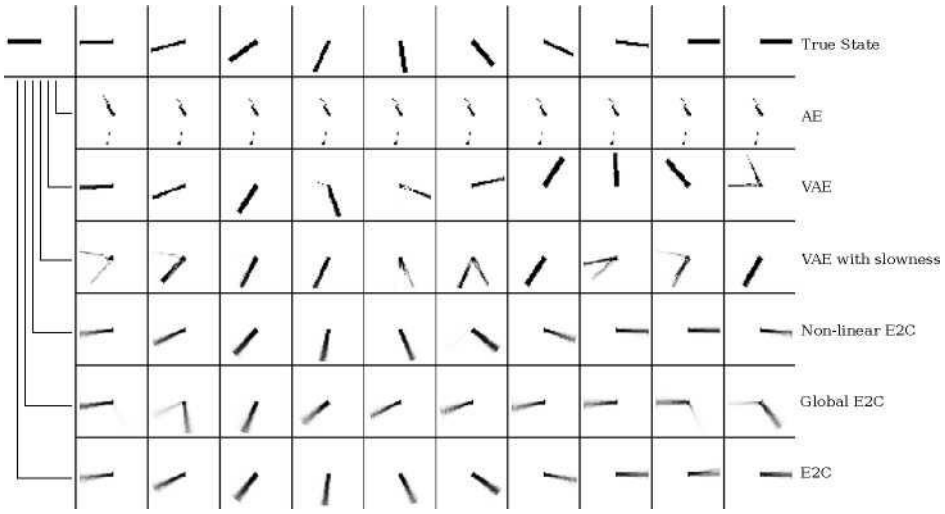


图 6：产生被动动力学的“梦想”轨迹（从左到右）：摆从 0 角开始=母没有速度。 模型必须预测动力学，而不施加力。

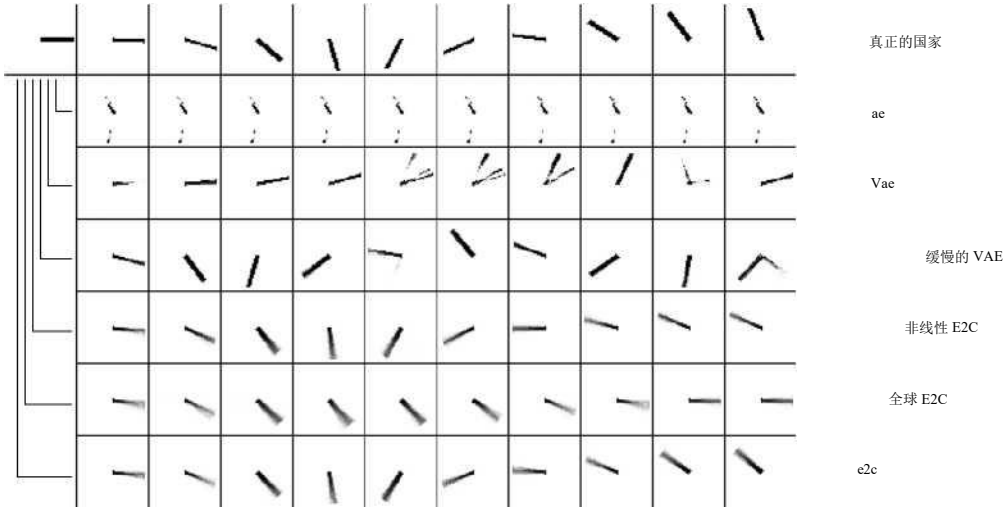


图 7：有梦想的轨迹（从左到右）用于控制动力学：钟摆以 0=n 角开始，没有速度。 对于 6 个时间步长，向右施加全力，然后向左施加 4 个全力的时间步长。

C.2 倒立摆潜伏空间.

将钟摆描绘成一个三维潜在空间，可以在图中进行视觉比较。 8

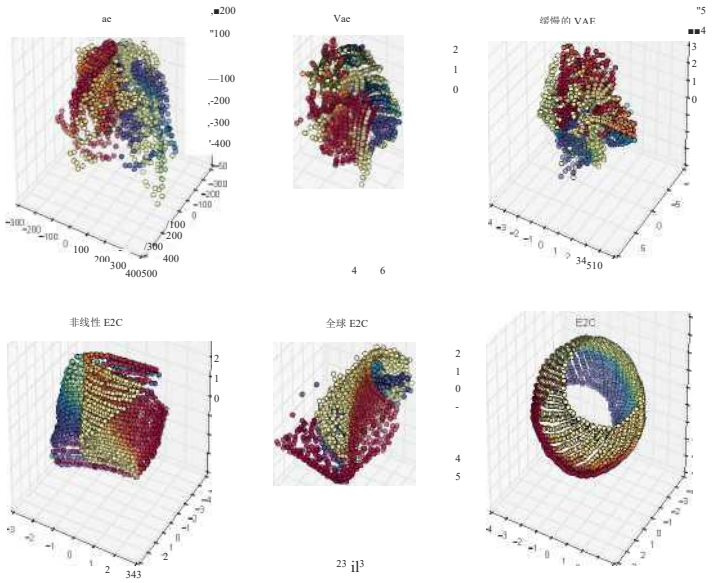


图 8：倒立摆所有基线模型和 E2C 变体的潜在空间。

第三 C.车杆和三连杆臂的轨迹

c.1 9 最后，与图中的图像相似，给出了购物车系统的多步预测。 我们描述了重要的情况：（1)一个长期的预测，小车-极静止(本质上是底层动力学的不稳定固定点)；(2)小车-极向右移动，改变极点角速度的方向(中间柱)；(3）极向右移动最远。 E2C 模型的长期预测都是高质量的。 请注意，对于不受控制的动力学，预测显示极向右移动的轻微偏差（我们在购物车极的训练模型中一直看到的效果）。 我们将这一问题归因于极角图像绘制过程中的离散化误差使得很难准确地预测小速度。

第四 C.为三连杆臂任务所采取的示例轨迹

10 图显示了由 E2C 系统执行的三连杆臂的受控轨迹的一段。 请注意，与本补充材料中的其他数字相比，它没有显示长期预测，而是显示了 E2C 系统在与模型预测控制相结合时所采取的轨迹的 10 个步骤（连同一步预测）。 对于所有任务的附加可视化和受控轨迹，我们参考补充视频。

第五 C.推车杆和机器人臂不同模型的比较

2 在 Tablewe 中，我们比较了我们的各种模型在实际轨迹成本和任务成功百分比的购物车-杆和机器人手臂。 所有结果平均超过 30 个不同的起始状态与一个固定的目标状态。

小车极总是以目标状态（零角和零速度）开始，具有小的加性高斯噪声(a=0。01)。 成功被定义为防止极点低于±0 的角度。 85 雷达。 三连杆臂系统以随机配置开始，目标是打开所有关节(例如。使所有角度为零)，并保持电子接近该位置。

结果表明，只有 E2C 及其非线性变体才能成功地执行这一任务，尽管两者之间仍然存在很大的性能差距。 我们得出结论，在训练相应模型后，线性化非线性动力学的误差增加到不再允许对系统进行精确控制的程度。

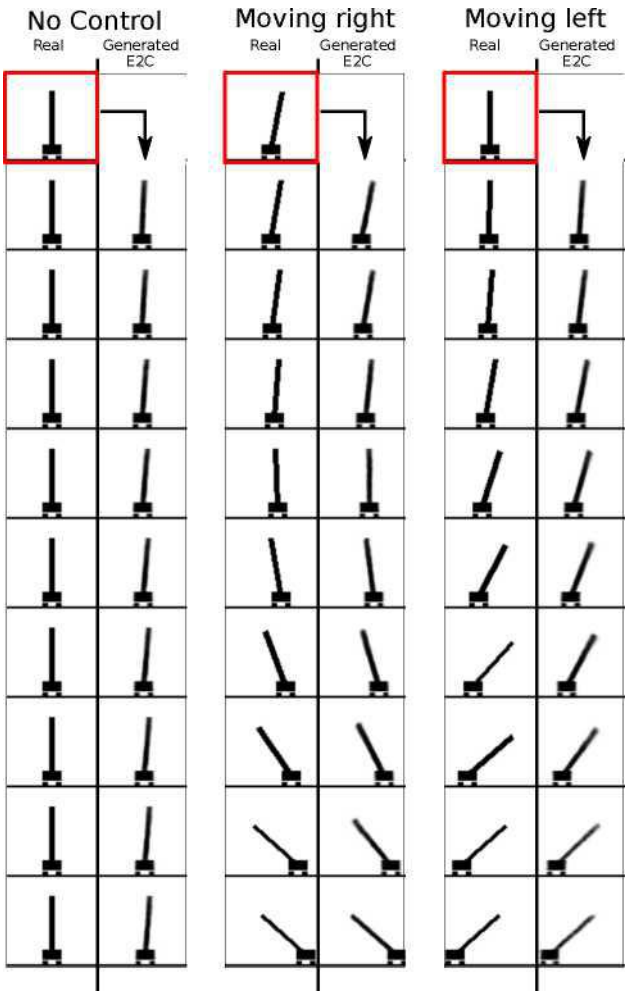


图 9：车杆系统中不受控制（左列)和受控(中/右列)动力学的梦想轨迹(从上到下）。红色图像显示初始配置，该配置被编码，导致 z1。 然后，通过跟踪潜在空间中的动力学，在没有额外输入的情况下生成每个列右半部分的图像。 左列描述了不受控制的情况(u=所有步骤的 0)。 中间柱显示在每个步骤中施加扭矩 20 的受控轨迹，右柱显示在每个步骤中施加扭矩 20 的轨迹。 在这些描述中省略了历史图像的预测。

表 2：购物车极和 threelink 任务不同方法的轨迹成本比较。 表中省略了标准自动编码器、变分自动编码器和全局 E2C 模型，因为它们在此任务上失败(性能类似于速度慢的 VAE)。

算法	真正的模型	VAE+慢速	E2C 无潜在 KL	非线性 E2C	e2c
特拉伊。费田			**纸箱-鞋底平衡**		
	15.33 ± 7.70	49.12 ± 16.94	48.90 ± 17.88	31.96 ± 13.26	**22.23 ± 14.89**
成功%	100%	0%	0%	63%	**93%**
特拉伊。费田			**三连臂**		
	59.46	1275.53 ± 864.66	1246.69 ± 262.6	460.40 ± 82.18	**90.23 ± 47.38**
成功%	100%	0%	0%	40%	**90%**

表 3：基于“真实”成本的 AICO 和 ILQR 之间的比较，用于使用卷积网络控制购物车极和三连杆机器人臂。

方法	I LQR	艾科
卡特尔		
e2c	14. 56 十 4。12	12.56 十 2.47
真正的模型	7. 45 士 1。22	7. 03 士 1。07
三环机器人手臂		
e2c	93. 78 十 32。9892.99 十 20。12	
真正的模型	53.59±9.74	56.34 10.82

第六 C.车杆和机器人臂轨迹优化器的比较

为了比较 AICO 如何处理潜在空间中估计的协方差矩阵，我们在购物车极和三连杆机器人臂任务上进行了一个额外的实验，并将其与 iLQR 进行了比较。 我们使用局部线性 E2C 模型进行模型预测控制，从 10 个不同的开始状态开始。 其余设置如本节所示 c.5.

3， 正如表中所报告的，这两种方法对这些任务执行的是相同的，表明我们的模型估计的协方差矩阵不会“伤害”规划，但考虑到它们也不会提高性能。

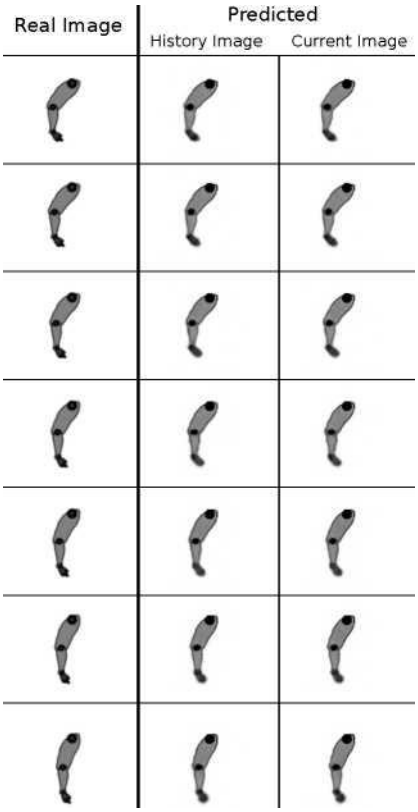


图 10：从轨迹（从上到下）中提取的框架，由嵌入到控制系统执行。 左列显示与 MDP 中的转换对应的真实图像。 中，右列显示了基于前两幅图像的历史图像和当前图像的预测..