eVAE：进化变分自编码器

Zhangkai Wu,[1] **Longbing Cao,** [1] **Lei Qi** [2]

[1] 悉尼科技大学 [2] 东南大学 berenwu1938@gmail.com,
Longbing.Cao@uts.edu.au，qilei@seu.edu.cn

arXiv:2301.00011v1 [cs.NE] 1 Jan 2023

摘要

变分自编码器（VAE）的替代损失给其训练带来了诸多挑战，导致任务拟合与表示推断之间失衡。为避免这一问题，现有针对 VAE 的策略着重于通过引入超参数来调整权衡、在一些温和假设下推导更紧的下界，或根据特定的神经网络设置分解损失分量。VAE 仍面临不确定性权衡学习的问题。我们基于变分信息瓶颈（VIB）理论和集成进化神经学习，提出了一种新颖的进化变分自编码器（eVAE）。eVAE 将变分遗传算法融入 VAE，并采用变分变异、交叉和进化等变分演化算子。其内外联合训练机制在无需额外约束的情况下，以协同且动态的方式生成和更新证据下界（ELBO）中的不确定性权衡学习。除了在变分信息瓶颈假设下学习数据的有损压缩和表示之外，eVAE 还提出了一种用于调节 VAE 和深度神经网络关键因素的进化范式，并通过将进化优化融入深度学习，解决早熟收敛和随机搜索问题。实验表明，eVAE 能以较低的重构损失解决文本生成中的 KL 消失问题，以清晰的图像生成所有解耦因素，并提升图像生成质量。与其他方法相比，eVAE 在重构损失、解耦和生成-推断平衡方面表现更佳。

引言

变分自编码器（Kingma 和 Welling，2013）（VAE）凭借融合概率原理与深度神经学习原理，能够从观测数据中学习连续且平滑的分布，因此备受关注。然而，尽管近年来出现了各种VAE机制和变体，VAE仍面临一些重大问题，包括变分推断过程中的动态不确定性学习，以及表示压缩与任务拟合之间的权衡。下文将简要分析这些问题及现有解决方案的不足。随后，我们将介绍一种新颖的进化变分自编码器（eVAE）框架，以解决其中一些问题和不足。

变分自编码器与差距分析变分自编码器在融入先验知识、将输入映射为概率表示以及近似输出似然方面展现出显著优势，。随机梯度变分贝叶斯（SGVB）估计器（Kingma 和 Welling 2013）与神经网络设置相结合，能够学习一个狭窄的概率潜在空间，从而推断隐藏空间中更具代表性的属性。变分自编码器已应用于多个领域，包括时间序列预测（Fortuin 等，2019年）、图像域外检测（Nalisnick 等，2019年；Ran 等，2022年；Havtorn 等，2021年；Xiao、Yan 和 Amit，2020年）、利用脉冲信号生成图像（Kamata、Mukuta 和 Harada，2022年），以及通过语言建模生成文本（Zhang 等，2019年）。

从理论上讲，变分推断中的界优化旨在用代理函数替代对数似然函数，使其能够通过梯度下降进行优化。在实践中，当后验分布与先验分布之间的差距较小时，证据下界（ELBO）仍无法接近条件似然的最大值。这会导致推断过程中出现动态不确定性，并使表示鲁棒性与重建效果之间的权衡失效。KL散度过弱可能导致KL消失，而KL散度过强则可能造成似然较差。这种权衡对后验分布的不相交性、数据特征和网络架构十分敏感（Bozkurt等人（2021））。

近来，人们提出了多种技术来应对这些问题。第一类方法旨在调整目标函数中的平衡。例如，β-VAE 引入超参数 β（Higgins 等人，2017）；InfoVAE 在 KL 项中添加缩放参数（Zhao、Song 和 Ermon，2019）；SA-VAE 采用循环退火计划，逐步增大 β，以减少 KL 消失（Fu 等，2019）；ControlVAE 则引入比例-积分-微分（PID）控制来调整超参数（Shao 等，2020）。这些方法都只是通过调整目标函数的一部分来提供局部解决方案，无法在动态情境下权衡并解决平衡问题。另一类方法则是调整网络架构或变分自编码器设置，例如 VQ-VAE 及其变体（Razavi、Van den Oord 和 Vinyals，2019；Berliner 等人，2022）、脉冲变分自编码器（Kamata、Mukuta 和 Harada，2022）、多模态变分自编码器（Sutter、Daunhawer 和 Vogt，2021）、专家混合多模态变分自编码器（Shi 等人，2019），以及基于专家乘积结构对其证据下界进行调整（Kviman 等人，2022；Jankowiak 和 Phan，2022）。然而，这些模型受限于特定的

任务或数据设置。

eVAE从信息瓶颈的角度来看，变分自编码器可以视为一种有损信息压缩过程。在一定范围内调整 KL散度，可以控制从潜变量表示到样本重构的信息瓶颈，从而有利于权衡压缩与重构（Alemi 等人，2017；Burgess 等人，2018）。这一变分信息瓶颈启发我们改进变分自编码器机制和目标函数。为了在不限制网络和输入的情况下，直接应对动态演化场景中的生成-推断平衡问题，我们提出了一种新颖的进化变分自编码器（eVAE）。eVAE 将变分演化学习和变分信息瓶颈融入并整合到变分自编码器中，以实现更好的最优探索，并权衡表示压缩与生成拟合。然而，将进化学习融入变分自编码器仍是一个开放课题。我们将变分遗传算法引入 eVAE，以进化的方式优化变分自编码器目标及其模型探索。

因此，eVAE 以进化且概率化的方式动态优化变分自编码器的推断（KL 项），进而调整变分自编码器的生成过程，使生成与推断达到平衡。为避免过早收敛，eVAE 引入概率染色体选择，以实现平滑搜索。为避免穷举式随机搜索，我们采用模拟二进制交叉（Deb 和 Beyer，2001年）和柯西分布变异，引导训练稳定收敛。本研究的主要贡献如下：

- 我们提出 eVAE 模型，以平衡生成与推断之间的权衡。具体而言，我们提出了基于柯西分布的变分变异算子和基于 SBX 的变分交叉算子，并采用演化式内外联合训练算法训练模型，从而解决借助变分遗传算法优化深度模型时出现的早熟收敛和随机搜索问题。

- 我们基于信息瓶颈理论，推导 β-VAE 相关模型的下界，以此阐释 eVAE。我们依据率失真理论分析相应下界中的信息流，表明只有 eVAE 中的迭代特定下界能够捕捉优化趋势，从而有效训练模型。

- 我们在三项任务上评估 eVAE：分别是在 dSprites 上进行解耦表示学习、在 CelebA 上进行图像生成，以及在 PTB语料库上进行文本生成。

eVAE 首次构建了一种进化变分自编码器框架，不对变分自编码器架构、输入设置和目标函数（证据下界）施加限制。它将变分编码与解码、信息瓶颈和进化学习融入深度神经网络框架。与先进的变分自编码器相比，评估结果表明，eVAE 在解耦和重构损失方面表现更好，并能有效解决 KL 消失问题。

VAE 背景与问题

在此，我们简要介绍变分自编码器的背景，重点关注其目标以及实现推断与生成平衡的方法，因为这些内容与本文最为相关。鉴于调整变分自编码器网络结构、输入设置和学习约束等其他相关进展与本文无关，因此不予讨论。

变分自编码器 变分自编码器通过学习观测数据 $\mathcal{X}$关于潜变量 $\boldsymbol{z}$ 的连续平滑表示分布 $p(x)$，缓解了自编码器的稀疏表示等问题（Tolstikhin 等人 2017）。在编码神经网络中学习到编码分布 $q_\phi(\boldsymbol{z} \mid \boldsymbol{x})$ 后，变分自编码器会应用变分推断来近似后验分布 $p_\theta(\boldsymbol{x} \mid \boldsymbol{z})$。随后，可在生成过程中从该学习到的分布中采样，以完成重构和生成等学习任务。借助 SGVB估计器和重参数化技巧，梯度变得易于计算，生成参数 θ 和推断参数 ϕ 也能够得以学习。变分自编码器的目标可转化为证据下界，其中对数据的经验分布 p_{data} 取期望，以兼顾重构 $\mathcal{L}_{\text{R}}$ 与推断 $\mathcal{L}_{\text{I}}$（ Esmaeili 等人 2019；Ghosh 等人 2019）。

$$\begin{aligned}\mathcal{L}_{ELBO} &= E_{x\sim p_{\text{data}}}\left[E_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log p_\theta(\boldsymbol{x} \mid \boldsymbol{z})\right]\right. \\ &\quad \left.-\mathrm{KL}\left(q_\phi(\boldsymbol{z} \mid \boldsymbol{x}) \| p(\boldsymbol{z})\right)\right] \\ &= E_{x\sim p_{\text{data}}}\mathcal{L}_{\text{R}} + \mathcal{L}_{\text{I}}\end{aligned} \tag{1}$$

变分自编码器的问题 为求解证据下界，推断模型 $q_\phi(z|x)$ 可以通过联合训练来最大化证据下界，从而获得适用于任务拟合的合理压缩。然而，解码器能力不足以及数据的多样性，可能导致表达能力强的后验分布偏向任务拟合，而非最优推断（Zhao、Song和Ermon（2019））。例如，在变分语言生成中，基于自回归模型（如长短期记忆网络和PixelCNN）构建的解码器，可以利用自回归特性生成语言样本，而非依赖基于后验分布的潜变量（Wu、Wang和Zhang 2019）。变分自编码器会退化为自回归模型，在训练过程中，后验分布与先验分布之间的KL散度会迅速降至零。这会导致KL消失，并因缺乏多样性而在测试中泛化能力较差。通过解码器学习具有相同分布的先验分布正交变换的方法（Khemakhem 等人 2020；Mita、Filippone 和 Michiardi 2021），可能会牺牲最优表示中的准确推断能力以及对数据拟合的泛化能力。其他研究则将训练冲突归因于界优化的固有属性。例如，在对后验分布作出明确的因子化假设时（Locatello 等人 2019），即，

$$p(\mathbf{x}, \mathbf{z}) = p(\mathbf{x} \mid \mathbf{z}) \prod_i p\left(z_i\right), \tag{2}$$

约束变分样本的证据下界有利于数据拟合（Burda、Grosse 和 Salakhutdinov 2016），但无法最大化对数似然上的概率质量。此外，原始VAE优化器会加剧 $q_\phi(z|x_i)$ 之间的分离，即 $\mu_i \to \infty$，$\sigma_i \to 0^+$，以分开

集中于每个样本的对数似然，从而最大化联合分布的质量（Zhao、Song和Ermon（2019））。

VAE增强 解决上述问题的一种方法是在后验分布（Alemi 等人，2018；Wang 和 Wang，2019；Domke 和 Sheldon，2019；Jankowiak和Phan（2022）；Ruiz 等人，2021；Kviman等人（2022））、先验分布（Tomczak 和 Welling，2018；Klushyn 等人，2019）以及证据下界的分解（Zhao、Song和Ermon（2019）；Esmaeili 等人，2019）中，在一些温和假设下收紧对数似然下界，从而获得正确的变分近似。例如，β-VAE（Higgins 等人，2017）引入超参数β，用于对 $\mathcal{L}_{KL}$ 项加权。随后，通过改变 β，证据下界会最小化 $\mathcal{L}_R$，以使数据拟合与正则化 $\mathcal{L}_I$ 共同收敛:

$$\begin{aligned}\mathcal{L}_{ELBO} =& E_{x\sim p_{data}}\left[E_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x}\mid\mathbf{z})\right]\right.\\ &\left.-\beta D_{KL}\left(q_\phi(\mathbf{z}\mid\mathbf{x})\|p(\mathbf{z})\right)\right]\\ =& E_{x\sim p_{\text{data}}}\left[\mathcal{L}_{\text{R}}+\beta\mathcal{L}_{\text{I}}\right]\end{aligned} \tag{3}$$

β-VAE 引入了一些根本性限制，由此催生了各种后续研究。InfoVAE 在 KL项上引入缩放参数 λ，并将目标函数改写为（Zhao、Song和Ermon（2019）:

$$\begin{aligned}\mathcal{L}_{ELBO} =& \alpha I_q(\mathbf{x};\mathbf{z}) - D_{KL}\left(q_\phi(\mathbf{z}\mid\mathbf{x})\|p(\mathbf{z})\right)\\ &- E_{q(z)}\left[D_{KL}\left(q_\phi(\mathbf{z}\mid\mathbf{x})\|p_\theta(\mathbf{x}\|\mathbf{z})\right)\right]\end{aligned} \tag{4}$$

其中，I_q 是互信息，权重为 α，且 $\alpha+\lambda-1=0$。这种对 KL 散度进行线性调节的方式表明，其应对动态不确定性时存在局限。

此外，条件变分自编码器（CVAE）（Sohn、Lee 和 Yan，2015年）将初始猜测作为条件变量引入目标函数，以处理多模态数据。SA-VAE采用循环退火计划，将训练拆分为多个周期，从 $\beta=0$开始，并逐步增大 β，直至 $\beta=1$，以减少 KL消失（Fu 等，2019年）。在 ControlVAE 中，PID控制将 KL 散度与设定点进行比较，并将两者的差值作为反馈传给控制器，以调节超参数 β（t）（Shao 等，2020年）。因此，ControlVAE 能够动态优化 KL 散度，但受限于 PID 控制器；该控制器采用的调节机制独立于变分自编码器本身。

现有研究在以下方面仍有不足：如何在重构与推断之间近似分配权重，如何在变分自编码器的工作机制内调节外部超参数，以及如何在进化学习过程中动态处理这些问题。我们的 eVAE 通过引入变分遗传学习来平衡推断与生成，并以进化方式将二者的影响纳入调节变分自编码器学习行为的过程，从而实现更好的不确定性权衡学习，弥补了这些不足。

eVAE模型

eVAE 还通过共同解决公式（3）和（4）中的以下问题，来调整公式（1）中的表征-生成平衡：在外循环中调整与变分自编码器行为无关的超参数、非动态优化，以及对数据和网络的设置受限。图1展示了用于改善变分自编码器平衡的变分演化学习 eVAE 框架。

eVAE - 演化式内外联合训练

如图 2 所示，eVAE 通过内外联合迭代训练过程，演化式地学习并优化 VAE 及其进化参数。在时间 t，给定输入 $x_t\in\mathcal{X}$，VAE 内层训练过程会使用某种初始化 β_0 训练一个 VAE 模型 $\mathcal{V}$（为确保通用性，我们不限制变分自编码器框架；在本文的案例研究中使用 β-VAE），以学习编码器 $q_\phi(\mathbf{z}_t\mid\mathbf{x}_t)$，以及带有先验分布$p(\mathbf{z})$的后验分布 $p_\theta(\mathbf{x}_t\mid\mathbf{z}_t)$。由此得到具有初始权重 β_0的 VAE 内层目标函数:

$$\begin{aligned}\mathcal{L}_{ELBO_t} = E_{x_t\sim p_{\mathcal{X}}}\left[E_{q_\phi(\mathbf{z}_t|\mathbf{x}_t)}\left[\log p_\theta\left(\mathbf{x}_t\mid\mathbf{z}_t\right)\right]-\right.\\ \left.\beta_0 D_{KL}\left(q_\phi\left(\mathbf{z}_t\mid\mathbf{x}_t\right)\|p(\mathbf{z})\right)\right]\end{aligned} \tag{5}$$

通过应用随机梯度变分贝叶斯和重参数化技巧，$\mathcal{V}$ 的生成过程将后验分布估计为

$$p_\theta(\mathbf{x}\mid\mathbf{z})=\prod_t p_\theta(\mathbf{x}_t\mid\mathbf{z}_t) \tag{6}$$

其中 $\mathbf{x}_t\mid\mathbf{z}_t\sim\mathcal{N}(\mu(\mathbf{z}_t;\theta)\ \Sigma(\mathbf{z}_t;\theta))$，其中 $\mathbf{z}_t=\mu_{x_t}+\epsilon\Sigma_{x_t}$和 $\epsilon\sim\mathcal{N}(0,I)$。

接下来，我们引入外层变分进化学习器 $\mathcal{E}$（下文“变分演化”部分将通过变分遗传算法对其进行说明），在外层过程中优化公式（5）中的目标。$\mathcal{E}$适应度函数为 f，从演化分布 $\beta_t\sim\mathcal{R}$ 中采样一个染色体 β_t，并通过变分交叉和变异使其进化，从而生成并改进参数 β_{t+1}。

随后，将演化得到的参数纳入 $\mathcal{V}_t$，用于下一轮 $t+1$ 训练。将相关证据下界信息反向传播，以调整变分自编码器网络并优化相关参数后，我们便得到更新后的重构和推断性能。因此，公式（5）演化为:

$$\begin{aligned}\mathcal{L}_{ELBO_{t+1}} = E_{x_{t+1}\sim p_{\mathcal{X}}}\left[E_{q_\phi(\mathbf{z}_{t+1}|\mathbf{x}_{t+1})}\left[\log p_\theta\left(\mathbf{x}_{t+1}\right.\right.\right.\\ \left.\left.\left.\mid\mathbf{z}_{t+1}\right)\right]-\beta_{t+1}D_{KL}\left(q_\phi\left(\mathbf{z}_{t+1}\mid\mathbf{x}_{t+1}\right)\|p(\mathbf{z})\right)\right]\end{aligned} \tag{7}$$

如果满足适应度公式（17），我们就通过 β_{t+1}更新 β。这样便会得到更新后的 $\mathcal{V}_{t+1}$ 及其参数 ϕ_{t+1}、θ_{t+1}和 β_{t+1}，用于第 $t+1$轮迭代。我们进一步重复上述 VGA $\mathcal{E}$，以获得另一组参数并重新训练模型$\mathcal{V}_{t+1}$，直至其收敛。

上述内外迭代训练会随着时间推移，在生成过程中的样本上迭代优化重构（即使用重构的 $\hat{}x_t$ 最小化 $\mathcal{L}_R=\|x_t-\hat{}x_t\|$）与推断（即寻找合适的 $\mathcal{L}_I$）之间的平衡，以及变分自编码器的学习行为。随后对变分自编码器 $\mathcal{V}$ 进行优化，直至收敛。

因此，eVAE 生成过程优化以下目标函数:

$$\begin{aligned}\mathcal{L}_{eVAE}=\min_{\boldsymbol{\theta}_t,\boldsymbol{\phi}_t}\textstyle\sum_{t=1}^{N}\mathcal{L}_{ELBO_{t+1}}\left(\boldsymbol{\theta}_t,\boldsymbol{\phi}_t\right.;\\ \left.f\left(\mathcal{E}\left(\phi_t,\theta_t,\{\beta\},\mathcal{L}_{ELBO_t}\right)\right),\mathbf{x}_t\right)\end{aligned} \tag{8}$$

以实现表示与构造之间的最佳平衡 β^*:

$$\beta^*\sim f\left(\mathcal{E}\left(\{\beta\}\mid\phi_t,\theta_t,\mathcal{L}_{ELBO_t}\right)\right) \tag{9}$$

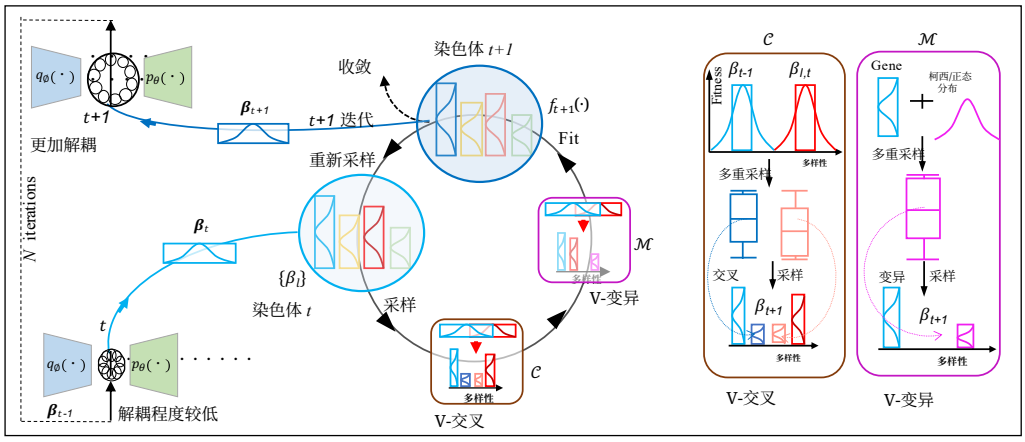


图1：eVAE 的框架。变分自编码器的结果为染色体采样提供依据。随后通过变分 V-交叉和 V-变异更新基因。根据适应度检查进化结果，以决定是否 $t+1$ 重新训练、放弃或收敛。

因此，eVAE 的整体证据下界可表示为：

$$\begin{aligned}\mathcal{L}_{eVAE} &= E_{\mathbf{z}\sim q_\phi(\mathbf{z}_t|\mathbf{x}_t)} \log p_\theta(\mathbf{x}_t \mid \mathbf{z}_t) + \\ & f_{\beta_t\sim R(\beta_t,\mathcal{E})} D_{KL}(q_\phi(\mathbf{z}_t \mid \mathbf{x}_t) \| p(\mathbf{z})) \\ &= E_{x_{t+1}\sim p_{\mathcal{X}}}[\mathcal{L}_R + f(\beta)\mathcal{L}_I + \mathcal{E}(\Delta\mathcal{L}_{ELBO})]\end{aligned} \tag{10}$$

eVAE 中的变分演化

在此，我们介绍变分进化学习器 $\mathcal{E}$，用于优化参数。我们通过变分遗传算法（VGA）实例化 $\mathcal{E}$，并采用其变分交叉和变异操作。这样可以避免在开发与探索失衡时常见的问题，包括早熟收敛和随机搜索。

变分遗传算法 eVAE 将基于 VGA 的外部参数优化与 VAE 内部基于梯度下降的优化相结合。因此，VGA 包含若干变分步骤和操作：初始化、变分交叉（V-交叉）、变分变异（V-变异）和变分评估（V-评估）。如下方图2所示，下面将分别介绍这些内容。

染色体选择：为使 GA 搜索空间更大、更平滑，并与 VAE 兼容，染色体通过连续变量 β 嵌入，该变量从不断演化的分布 $\mathcal{R}$ 中采样（例如，在交叉过程中）。假设 L 个独立染色体构成候选组 $\{\beta_l\} = \{\beta_1, \ldots, \beta_L\}$，从中选择一个染色体来训练 VAE；经过 VGA 操作和 VAE 重新训练后，该染色体会获得一个适应度值 f。因此，在跨代 VGA 进化和子代选择过程中，我们得到嵌入后的染色体—适应度配对，例如：

$$\{\beta, f\} = ((\beta_1, f_1), \ldots, (\beta_L, f_L)) \tag{11}$$

用于优化内外层联合训练。

V-交叉：在时间 $t-1$ 和 t 对配对染色体进行交叉，以生成新的基因，从而提高子代的遗传多样性。遵循模拟二进制交叉（SBX）（Deb和Beyer，2001年）方法，我们实施以下交叉操作 $\mathcal{C}$，以在策略 $\mathcal{C}$ 中，将时间 $t-1$的染色体 β_{t-1} 与时间 t 从染色体候选组 $\{\beta_l\}$ 中选出的染色体 β_t 进行演化，形成一个较大的搜索空间，从而获得下一时间 $t+1$ 的子代候选染色体 β_{t+1}：

$$\mathcal{C}: \begin{cases} \beta_{t+1} = \frac{1}{2}[(1+r_c)\beta_{l,t} + (1-r_c)\beta_{t-1}], or \\ \beta_{t+1} = \frac{1}{2}[(1-r_c)\beta_{l,t} + (1+r_c)\beta_{t-1}] \end{cases} \tag{12}$$

其中，r_c 是交叉率，它从概率密度函数 $P_c(r_c)$中抽取：

$$P_c(r_c) = \begin{cases} 0.5(\eta+1)r_c^n, & \text{if } r_c \le 1 \\ 0.5(\eta+1)\frac{1}{r_c^{\eta+2}}, & \text{otherwise} \end{cases} \tag{13}$$

通过集中其对应的父代 β_{t-1} 和$\beta_{l,\ t}$。通过从以下函数中采样 r_c：

$$r_c = \begin{cases} (2u)^{\frac{1}{\eta+1}}, & \text{if } u \le 0.5 \\ \left(\frac{1}{2(1-u)}\right)^{\frac{1}{\eta+1}}, & \text{otherwise} \end{cases} \tag{14}$$

其中，u是随机变量，η 是用于缩放的超参数。随后，可以通过选择更优的 $\mathcal{C}$ 策略（即更好地满足公式（17）的策略）来创建子代 β_{t+1}。

V-变异：交叉生成的子代还可以进一步变异，以提升其基因组多样性和进化能力。与 V-交叉相对应，采用变分变异策略 $\mathcal{M}$ 来增加子代的多样性，以更新从交叉继承的 β_{t+1} 或当前时刻的 β_t。以较小的概率 p_m，从群体中选取一个染色体 β_{t+1} 进行变异，如下所示：

$$\mathcal{M}: \beta_{t+1} = \beta_{t+1} + r_m \tag{15}$$

其中 r_m 是从柯西分布 p_m中采样得到的随机变量：

$$P_m = \frac{1}{\pi}\left(\frac{1}{1+\beta_{t+1}^2}\right) \tag{16}$$

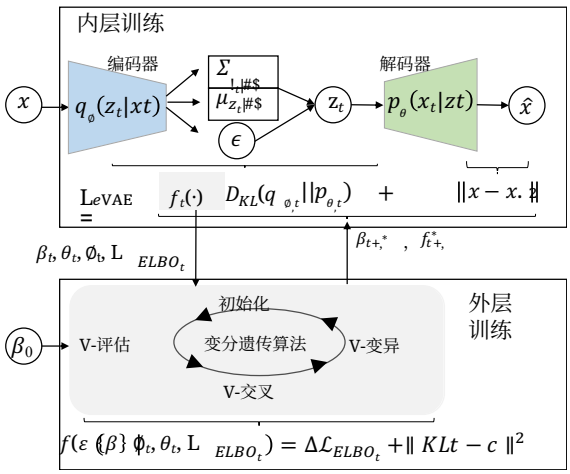


图2：eVAE——内外层联合进化训练过程。上半部分展示了时刻 t 的变分自编码器训练，随后将目标纳入下半部分，即通过变分遗传算法进行外层训练，并将基于适应度优化得到的结果反馈给变分自编码器，以进行进一步训练。

V-评估与VGA适应度函数通过V-交叉和V-变异操作更新的染色体，会根据变分自编码器目标进行评估，以确定其是否能够传递到下一代。在变分自编码器框架中，我们采用以下启发式适应度函数 f 来引导染色体的进化，使其适应变分自编码器目标。该适应度函数综合了面向变分自编码器的随机梯度下降方向，以及与最优KL散度之间的距离，从而实现多目标优化。

$$f_{t+1} = \Delta\mathcal{L}_{ELBO_{t+1}} + ||KL_{t+1}(\beta_{t+1}) - c|| \quad (17)$$

其中$\Delta\mathcal{L}_{ELBO_{t+1}}$ 为:

$$\Delta\mathcal{L}_{ELBO_{t+1}} = \mathcal{L}_{ELBO_{t+1}}(\beta_{t+1}) - \mathcal{L}_{ELBO_t}(\beta_t) \quad (18)$$

表示应用进化得到的 β后，变分自编码器的证据下界差值。c 是任务特定的信息瓶颈，用于确保 eVAE 在进化优化过程中实现有界优化。

V-评估和适应度 f 引导 eVAE 收敛，并通过增强的进化参数化平衡重构与推断。适应度值越大，染色体在其候选组中的表现越强，从而形成更优的配对 $\{\beta^*, f^*\}$。

eVAE模型的完整训练过程见补充材料 A 中的算法1。

理论分析

本文以 β-VAE 为例，研究进化变分自编码器。因此，我们将在下文分析 eVAE 在调整 β-VAE 参数、平衡重构与正则化，以及训练性能方面的影响。

β-VAE 直接调整超参数 β，以设定与任务相关的比例，平衡推断损失 $\mathcal{L}_{\mathrm{I}}$ 和生成损失 $\mathcal{L}_{\mathrm{R}}$。本质上，通过 β 逐步对 $\mathcal{L}_{\mathrm{I}}$ 进行正则化等价于在特定实验的阈值（即常数 C）下最大化似然（Higgins 等人，2017）：

$$\max_{\phi,\theta} E_{x\sim p_{\text{data}}}\mathcal{L}_R \quad \text{s.t. } \beta\mathcal{L}_I < C \quad (19)$$

这一学习过程可以用信息瓶颈（IB）理论来解释（Alemi 等人，2018）。$\beta\mathcal{L}_I$ 约束充当瓶颈，对由输入 X 构建的潜变量 Z 的容量进行压缩，使其能够有效表示目标 X （Alemi 等人，2017）。因此，IB 最大化 $I(Z,\ X)$，以获得简洁的表示：

$$\max I(Z, \hat{X}) \quad \text{s.t. } I(X, Z) \le I_c \quad (20)$$

此外，变分信息瓶颈（VIB）引入了一个针对特定实验的优化下界：

$$\mathcal{L}_{\beta-VAE}(\theta,\phi) = -D - \beta_C R \le -\beta_C I - D \quad (21)$$

其中，R 表示压缩率，D 表示衡量表示与任务相关性的失真，β_C 是针对特定实验的常数。因此，变分自编码器可以被视为实现率失真权衡的过程。

ControlVAE 是 β-VAE 的一种变体，通过 PID 控制将针对 KL散度 的 β_{KL} 调节至期望的 KL 点：

$$\mathcal{L}_{ControlVAE}(\theta,\phi) = -D - \beta_{KL} R \le -\beta_{KL} I - D \quad (22)$$

eVAE 扩展了 β-VAE，通过变分进化学习器 $\mathcal{E}$ 调节信息瓶颈 βI（X，Z）。外层训练经过第 t 次迭代后，β_t 会在适应度函数 f 的引导下，进化以适应任务和训练阶段。eVAE 会生成一个下界，以优化内层训练阶段第 $t+1$ 次迭代：

$$\mathcal{L}_{eVAE} \le -\mathcal{E}(\beta_t) I_t - \mathcal{E}(\beta_t) D_t \quad (23)$$

其中，随迭代而变化的下界可在 $-\mathcal{E}$（β_t）D_t 中的任务拟合与 $\mathcal{E}$（β_t）I_t 中的推断质量之间取得平衡。式 (21) 和式 (23) 的推导见补充材料 B 和 C。

我们进一步通过图 3 中的率失真（R-D）曲线，将 eVAE 的 VIB 效果与 VAE、β-VAE 和 ControlVAE 进行比较。VAE（黄色）在早期倾向于牺牲经验误差最小化来进行表征学习，此时率在四分之一的迭代次数内被优化至下降。至于 β-VAE（绿色），由于其针对实验的下界（$-\beta_C I - D$），失真无法最小化。ControlVAE（粉色）最初专注于优化失真以拟合数据，最终优化率以获得平滑的表示，这依赖于针对 KL散度 的下界（$-\beta_{KL} I - D$）。然而，直接通过给定的 PID控制器 修改 β，会导致推断能力与重构质量在迭代过程中出现波动。eVAE 并不监控优化过程，而是生成一个随迭代变化的下界（$-\mathcal{E}$（β_t）$I_t - \mathcal{E}$（β_t）D_t），以实现最小化失真与控制率之间的平衡。

实验

我们在三项典型任务中评估 eVAE：解耦表示、图像生成和语言建模。

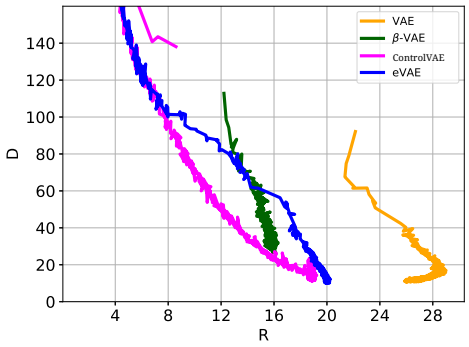


图 3：VAE、 β-VAE、ControlVAE 和 eVAE 在 dSprites 上的信息平面及其 $R-D$ 曲线。

数据集与基线方法

下面列出了三种不同的学习任务及其数据集，用于将 eVAE 与 VAE 基线进行比较。

解耦表示 通过学习相互独立的潜变量来生成图像。我们在 dSprites 上验证 eVAE。dSprites 是一个二维形状数据集，包含 737,280 张由六个真实因素生成的二值图像：形状（正方形、椭圆、心形）、尺度、方向、位置 x 和位置 y。为了同等全面地考察解耦性能和图像生成能力，我们将 β-VAE 与 ControlVAE 进行比较。

图像生成 根据给定的数据点重建图像样本。我们使用 CelebA 数据集（裁剪版），该数据集是一个真实世界的名人面部数据集，包含 202,599 张 RGB 图像（Liu 等，2015）。基线方法为 β-VAE 和 ControlVAE。

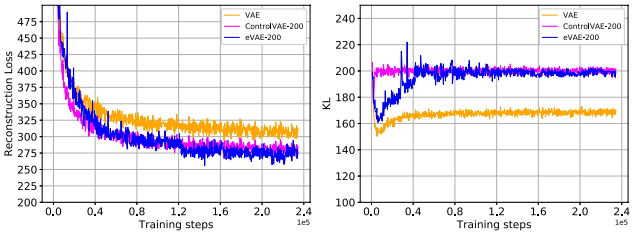
语言建模进行词级文本生成。我们使用包含英语语料库的宾夕法尼亚树库数据集（PTB）（Marcinkiewicz，1994）。为展示抑制 KL消失的效果，我们将 eVAE 与三种启发式方法进行比较：代价退火（记作 Cost-10k）（Bowman 等，2016）、循环退火（Cyc-8）（Fu 等，2019）以及 PID控制（PID-3）（Shao 等，2020）。

为确保公平比较，我们采用与 β-VAE 相同的嵌入网络，并调整 eVAE，使其达到与 ControlVAE 相同的 KL 设定点。VAE 架构和超参数设置见补充材料 D。

性能评估

解耦表示学习的性能这里，我们验证了利用解耦特征重构清晰图像的性能。图 5(a) 显示，在预期 KL 散度相同（KL = 19）的情况下，eVAE 的重构损失更低，为 9.2，优于 ControlVAE。此外，图 5(a) 和 (b) 显示，与 β-VAE 和 ControlVAE 直接修改权重不同，eVAE 通过 VGA 适应度函数引导的动态加权，实现了比其他变分自编码器更稳定的训练曲线。图 5(c) 展示了各个解耦因子的变化。图 6 进一步证明，在 KL = 19 设定点下，eVAE 能够可靠地解耦所有因子，并实现清晰的重构质量。相比之下，ControlVAE 捕获了尺度、 x 和y 这些生成因子，而 β-VAE 除尺度之外，将其他四个因子纠缠在一起。

图像生成性能 我们在 CelebA 上评估 eVAE 的图像生成性能。为进行公平且简洁的比较，我们将 KL 设定点设为 200，这一数值在（Shao 等，2020）中可获得最佳重构质量。我们调整 eVAE，使其达到与 ControlVAE 类似的目标 KL 点。图 4(a) 显示，与原始变分自编码器和 ControlVAE 相比，eVAE 的重构误差最低。在不断演化的 β 的引导下，eVAE 中的 KL 散度能够动态达到预期值，如图 4(b) 所示。



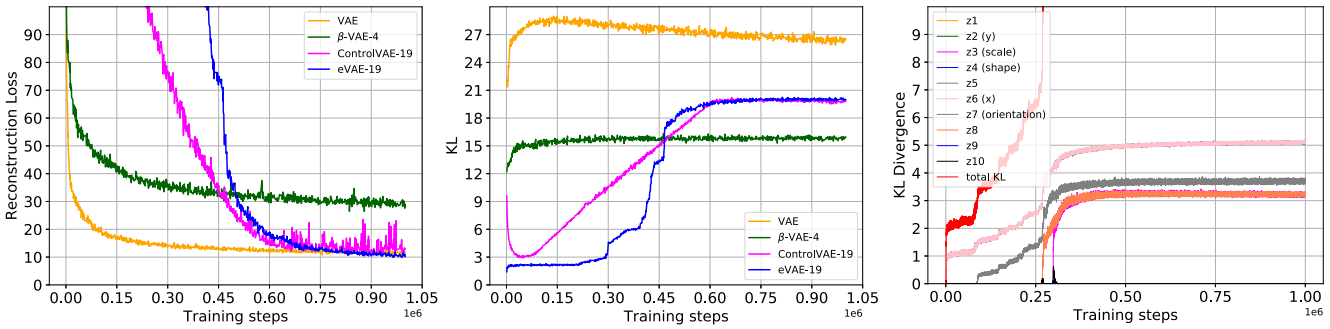
(a) 重构损失　　(b) KL散度

图 4：CelebA 上不同变分自编码器在图像生成任务中的性能比较。

语言建模的性能 为了展示 eVAE 在防止变分语言模型退化为标准语言模型方面的有效性，我们绘制了 KL散度、重构损失以及 KL 权重随迭代次数变化的学习曲线。图 7(a) 展示了 β 的进化，表明 eVAE 能够在迭代过程中自动学习调整 β，而不是依赖启发式设定。图 7(b) 显示，代价退火（Cost-10k）模型仍然受到 KL消失问题的困扰，KL散度在迭代结束时趋近于 0。相比之下，循环退火（Cyc-8）、PID 和遗传算法均能收敛至非零散度，而 eVAE 在词生成任务中的重构损失最低，达到 70，如图 7(c) 所示。此外，与 PID控制相比，VGA 能够找到通向更稳定阶段的路径，即使非零的已学习 β 固定不变，训练过程中也不会出现重构波动。

结论

我们提出了进化变分自编码器（eVAE），以动态学习变分自编码器中重建效果与推理鲁棒性之间的权衡。eVAE 引入了变分进化算子，以解决早熟收敛和随机搜索问题。在 eVAE 中，随机梯度下降优化与遗传算法协同遵循内外联合训练机制。基于变分信息瓶颈理论，eVAE 推导出迭代特定下界，在迭代过程中平衡压缩与解压缩的能力。多项任务实验表明，与最先进的变分自编码器相比，eVAE 能以更解耦的特征遍历样本，生成重构损失更低的清晰图像，并在变分语言建模中自动解决 KL 消失问题。未来我们将把 eVAE 扩展到其他变分自编码器模型和任务中。



(a) 重构损失　(b) KL散度　(c) 逐元素 KL 散度

图 5：dSprites[1]上的学习曲线。(a、b)表明，在固定 KL 点 KL=19下，与 VAE（$\beta = 1$）、β-VAE（$\beta = 4$）和 ControlVAE（KL = 19）相比，eVAE 的重构损失最低。(c)展示了 eVAE 中逐元素 KL 散度随迭代次数的变化。可以观察到，eVAE 在每个生成器维度（因子）上都保持了稳定且合理的 KL 散度：位置-y （z2）、尺度（z3）、形状（z4）、位置-x （z6）、方向（z7）。关于生成 KL 散度的更多比较见补充材料 E。关于生成 KL 散度的更多比较见补充材料 E。

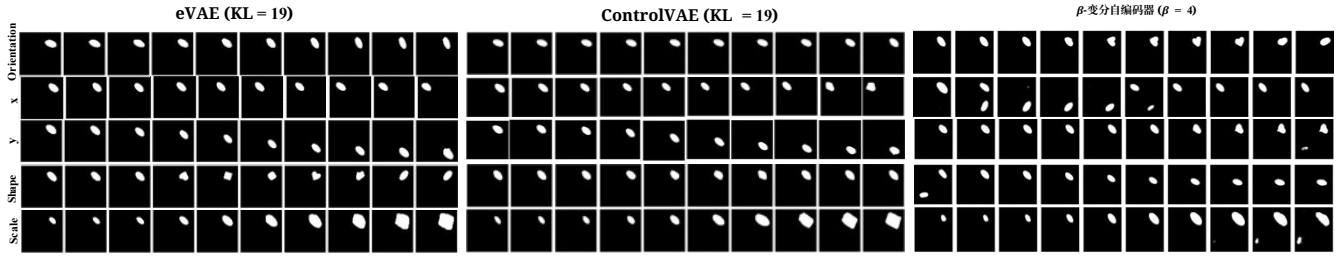
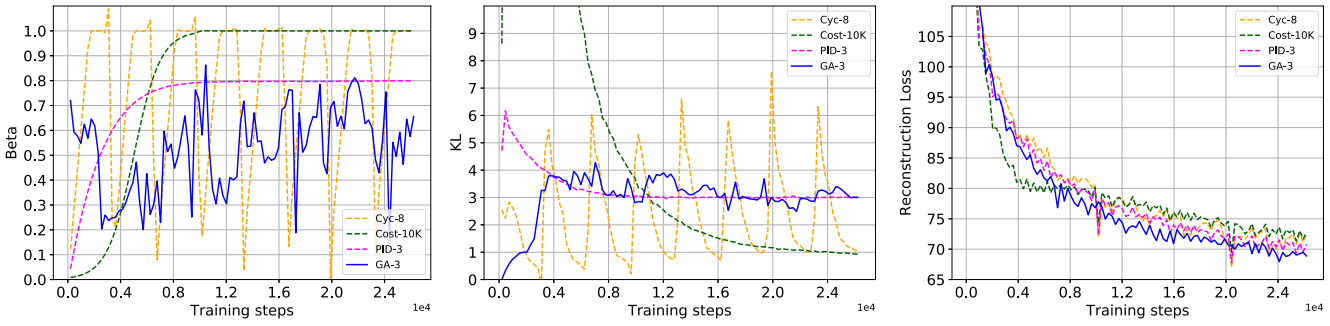


图 6：dSprites 上的潜变量遍历。为了直观地区分潜在因子，我们以椭圆为例进行说明。每一行表示遍历中的一个潜在因子，同时保持其他因子不变。第一列是用于初始化的种子图像，随后我们在范围 [-3, 3] 内操控潜变量维度 z 。



(a) Beta　(b) KL散度　(c) 重构损失

图 7：PTB 上语言建模的学习曲线。PID 和 VGA 的后缀表示 KL 设定点，即 KL = 3；Cyc-8 表示 8 个迭代周期，权重在此期间从 0 线性增加到 1；Cost-10k 表示 10k 次迭代，在此期间权重通过 Sigmoid 函数反复从 0 增加到 1。

参考文献

Alemi, A.; Poole, B.; Fischer, I.; Dillon, J.; Saurous, R. A.; and Murphy, K. 2018. 修复有缺陷的证据下界。载于*ICML*。Alemi, A. A.; Fischer, I.; Dillon, J. V.; and Murphy, K. 2017. 深度变分信息瓶颈。载于*ICLR*。Berliner, A.; Rotman, G.; Adi, Y.; Reichart, R.; and Hazan, T. 2022. 使用自然进化策略学习离散结构化变分自编码器。载于*ICLR*。Bowman, S. R.; Vilnis, L.; Vinyals, O.; Dai, A. M.; Jozefowicz, R.; and Bengio, S. 2016. 从连续空间生成句子。载于计算自然语言学习会议。Bozkurt, A.; Esmaeili, B.; Tristan, J.-B.; Brooks, D. H.; Dy, J. G.; and van de Meent, J.-W. 2021. 变分自编码器中的率正则化与泛化。*AISTATS*。Burda, Y.; Grosse, R.; and Salakhutdinov, R. 2016. 重要性加权自编码器。载于*ICLR*。Burgess, C. P.; Higgins, I.; Pal, A.; Matthey, L.; Watters, N.; Desjardins, G.; and Lerchner, A. 2018. 理解 β-VAE中的解耦。载于*NeurIPS*。Deb, K.; and Beyer, H.-G. 2001. 采用模拟二进制交叉的自适应遗传算法。进化计算，9(2)：197–221。Domke, J.; and Sheldon, D. R. 2019. 分而耦合：使用蒙特卡罗变分目标进行后验近似。载于*NeurIPS*。Esmaeili, B.; Wu, H.; Jain, S.; Bozkurt, A.; Siddharth, N.; Paige, B.; Brooks, D. H.; Dy, J.; and Meent, J.-W. 2019. 结构化解耦表示。载于*AISTATS*。Fortuin, V.; H¨user, M.; Locatello, F.; Strathmann, H.; and R¨atsch, G. 2019. SOM-变分自编码器：面向时间序列的可解释离散表示学习。载于*ICLR*。Fu, H.; Li, C.; Liu, X.; Gao, J.; Celikyilmaz, A.; and Carin, L. 2019. 循环退火计划：缓解KL消失的一种简单方法。载于北美计算语言学协会年会。Ghosh, P.; Sajjadi, M. S.; Vergari, A.; Black, M.; and Sch¨olkopf, B. 2019. 从变分自编码器到确定性自编码器。载于*ICLR*。Havtorn, J. D. D.; Frellsen, J.; Hauberg, S.; and Maaløe, L. 2021. 分层变分自编码器知道自己不知道什么。载于*ICML*。Higgins, I.; Matthey, L.; Pal, A.; Burgess, C.; Glorot, X.; Botvinick, M.; Mohamed, S.; and Lerchner, A. 2017. β-变分自编码器：在受约束的变分框架下学习基本视觉概念。载于*ICLR*。Jankowiak, M.; and Phan, D. 2022. 用于变分退火重要性采样的替代似然。载于*ICML*。Kamata, H.; Mukuta, Y.; and Harada, T. 2022. 全脉冲变分自编码器。*AAAI*。Khemakhem, I.; Kingma, D.; Monti, R.; and Hyvarinen, A. 2020. 变分自编码器与非线性独立成分分析：统一框架。载于*AISTATS*。Kingma, D. P.; and Welling, M. 2013. 自动编码变分贝叶斯。*arXiv*。Klushyn, A.; Chen, N.; Kurle, R.; Cseke, B.; and van der Smagt, P. 2019. 在变分自编码器中学习层次先验。载于*NeurIPS*。Kviman, O.; Melin, H.; Koptagel, H.; Elvira, V.; and Lager- gren, J. 2022. 变分近似的多重重要性采样证据下界及深度集成。载于*AISTATS*。Liu, Z.; Luo, P.; Wang, X.; and Tang, X. 2015. 在自然环境中进行人脸属性深度学习。载于*ICCV*。Locatello, F.; Abbati, G.; Rainforth, T.; Bauer, S.; Sch¨olkopf, B.; and Bachem, O. 2019. 关于解耦表示的公平性。载于*NeurIPS*。Marcinkiewicz, M. A. 1994. 构建大型英语标注语料库：宾州树库。《使用大型语料库》。Mita, G.; Filippone, M.; and Michiardi, P. 2021. 一种用于解耦表示的可识别双重变分自编码器。载于*ICML*。Nalisnick, E.; Matsukawa, A.; Teh, Y. W.; Gorur, D.; and Lakshminarayanan, B. 2019. 深度生成模型是否知道自己不知道什么？载于*ICLR*。Ran, X.; Xu, M.; Mei, L.; Xu, Q.; and Liu, Q. 2022. 通过具有可靠不确定性估计的变分自编码器检测分布外样本。《神经网络》。Razavi, A.; Van den Oord, A.; and Vinyals, O. 2019. 使用VQ-VAE-2 生成多样化的高保真图像。载于*NeurIPS*。Ruiz, F. J.; Titsias, M. K.; Cemgil, T.; and Doucet, A. 2021. 使用耦合马尔可夫链对变分自编码器进行无偏梯度估计。载于*UAI*。Shao, H.; Yao, S.; Sun, D.; Zhang, A.; Liu, S.; Liu, D.; Wang, J.; and Abdelzaher, T. 2020. Controlvae：可控变分自编码器。载于*ICML*。Shi, Y.; Paige, B.; Torr, P.; et al. 2019. 用于多模态深度生成模型的变分混合专家自编码器。载于*NeurIPS*。Sohn, K.; Lee, H.; and Yan, X. 2015. 使用深度条件生成模型学习结构化输出表示。载于*NeurIPS*。Sutter, T. M.; Daunhawer, I.; and Vogt, J. E. 2021. 广义多模态证据下界。*arXiv*。Tolstikhin, I.; Bousquet, O.; Gelly, S.; and Schoelkopf, B. 2017. Wasserstein自编码器。*arXiv*。Tomczak, J.; and Welling, M. 2018. 使用 Vamp先验的变分自编码器。载于*AISTATS*。Wang, P. Z.; and Wang, W. Y. 2019. 用于变分自编码器的神经高斯Copula。载于*EMNLP*。Wu, D.; Wang, L.; and Zhang, P. 2019. 使用变分自回归网络求解统计力学问题。《物理评论快报》，122(8)：080602。Xiao, Z.; Yan, Q.; and Amit, Y. 2020. 似然遗憾：一种用于变分自编码器的分布外检测分数。*NeurIPS*。Zhang, Y.; Wang, Y.; Zhang, L.; Zhang, Z.; and Gai, K. 2019. 通过自标注条件变分自编码器改进多样化文本生成。载于*ICASSP*。Zhao, S.; Song, J.; and Ermon, S. 2019. InfoVAE：平衡变分自编码器中的学习与推断。载于*AAAI*。

参考文献

Alemi, A.; Poole, B.; Fischer, I.; Dillon, J.; Saurous, R. A.; and Murphy, K. 2018. 修正有缺陷的证据下界。载于*ICML*。Alemi, A. A.; Fischer, I.; Dillon, J. V.; and Murphy, K. 2017. 深度变分信息瓶颈。载于*ICLR*。Berliner, A.; Rotman, G.; Adi, Y.; Reichart, R.; and Hazan, T. 2022. 使用自然进化策略学习离散结构化变分自编码器。载于*ICLR*。Bowman, S. R.; Vilnis, L.; Vinyals, O.; Dai, A. M.; Jozefow- icz, R.; and Bengio, S. 2016. 从连续空间生成句子。载于计算自然语言学习会议。Bozkurt, A.; Esmaeili, B.; Tristan, J.-B.; Brooks, D. H.; Dy, J. G.; and van de Meent, J.-W. 2021. 变分自编码器中的率正则化与泛化。*AISTATS*。Burda, Y.; Grosse, R.; and Salakhutdinov, R. 2016. 重要性加权自编码器。载于*ICLR*。Burgess, C. P.; Higgins, I.; Pal, A.; Matthey, L.; Watters, N.; Desjardins, G.; and Lerchner, A. 2018. 理解 β-VAE中的解耦。载于*NeurIPS*。Deb, K.; and Beyer, H.-G. 2001. 采用模拟二进制交叉的自适应遗传算法。《进化计算》，9(2)：197–221。Domke, J.; and Sheldon, D. R. 2019. 分而耦之：使用蒙特卡洛变分目标进行后验近似。载于*NeurIPS*。Esmaeili, B.; Wu, H.; Jain, S.; Bozkurt, A.; Siddharth, N.; Paige, B.; Brooks, D. H.; Dy, J.; and Meent, J.-W. 2019. 结构化解耦表示。载于*AISTATS*。Fortuin, V.; H¨user, M.; Locatello, F.; Strathmann, H.; and R¨atsch, G. 2019. SOM-变分自编码器：时间序列上的可解释离散表示学习。载于*ICLR*。Fu, H.; Li, C.; Liu, X.; Gao, J.; Celikyilmaz, A.; and Carin, L. 2019. 循环退火计划：缓解KL消失的一种简单方法。载于北美计算语言学协会年会。Ghosh, P.; Sajjadi, M. S.; Vergari, A.; Black, M.; and Sch¨olkopf, B. 2019. 从变分自编码器到确定性自编码器。载于*ICLR*。Havtorn, J. D. D.; Frellsen, J.; Hauberg, S.; and Maaløe, L. 2021. 层次变分自编码器知道自己不知道什么。载于*ICML*。Higgins, I.; Matthey, L.; Pal, A.; Burgess, C.; Glorot, X.; Botvinick, M.; Mohamed, S.; and Lerchner, A. 2017. β-VAE：通过受约束的变分框架学习基本视觉概念。载于*ICLR*。Jankowiak, M.; and Phan, D. 2022. 用于变分退火重要性采样的替代似然。载于*ICML*。Kamata, H.; Mukuta, Y.; and Harada, T. 2022. 全脉冲变分自编码器。*AAAI*。Khemakhem, I.; Kingma, D.; Monti, R.; and Hyvarinen, A. 2020. 变分自编码器与非线性 ICA：一个统一框架。载于*AISTATS*。Kingma, D. P.; and Welling, M. 2013. 自编码变分贝叶斯。*arXiv*。Klushyn, A.; Chen, N.; Kurle, R.; Cseke, B.; and van der Smagt, P. 2019年. 在变分自编码器中学习层次先验分布。载于*NeurIPS*. Kviman, O.; Melin, H.; Koptagel, H.; Elvira, V.; and Lager- gren, J. 2022年. 变分近似的多重重要性采样证据下界与深度集成。载于*AISTATS*. Liu, Z.; Luo, P.; Wang, X.; and Tang, X. 2015年. 在自然环境中进行人脸属性深度学习。载于*ICCV*. Locatello, F.; Abbati, G.; Rainforth, T.; Bauer, S.; Sch¨olkopf, B.; and Bachem, O. 2019年. 关于解耦表示的公平性。载于*NeurIPS*. Marcinkiewicz, M. A. 1994年. 构建大型英语标注语料库：宾州树库。《使用大型语料库》. Mita, G.; Filippone, M.; and Michiardi, P. 2021年. 一种用于解耦表示的可识别双重变分自编码器。载于*ICML*. Nalisnick, E.; Matsukawa, A.; Teh, Y. W.; Gorur, D.; and Lakshminarayanan, B. 2019年. 深度生成模型知道自己不知道什么吗？载于*ICLR*. Ran, X.; Xu, M.; Mei, L.; Xu, Q.; and Liu, Q. 2022年. 通过具有可靠不确定性估计的变分自编码器检测分布外样本。《神经网络》. Razavi, A.; Van den Oord, A.; and Vinyals, O. 2019年. 使用 VQ-VAE-2 生成多样化的高保真图像。载于*NeurIPS*. Ruiz, F. J.; Titsias, M. K.; Cemgil, T.; and Doucet, A. 2021年. 使用耦合马尔可夫链对变分自编码器进行无偏梯度估计。载于*UAI*. Shao, H.; Yao, S.; Sun, D.; Zhang, A.; Liu, S.; Liu, D.; Wang, J.; and Abdelzaher, T. 2020年. ControlVAE：可控变分自编码器。载于*ICML*. Shi, Y.; Paige, B.; Torr, P.; et al. 2019年. 用于多模态深度生成模型的变分混合专家自编码器。载于*NeurIPS*. Sohn, K.; Lee, H.; and Yan, X. 2015年. 使用深度条件生成模型学习结构化输出表示。载于*NeurIPS*. Sutter, T. M.; Daunhawer, I.; and Vogt, J. E. 2021年. 广义多模态证据下界。*arXiv*. Tolstikhin, I.; Bousquet, O.; Gelly, S.; and Schoelkopf, B. 2017年. Wasserstein自编码器。*arXiv*. Tomczak, J.; and Welling, M. 2018年. 带有 Vamp 先验的变分自编码器。载于 *AISTATS*. Wang, P. Z.; and Wang, W. Y. 2019年. 用于变分自编码器的神经高斯 Copula。载于*EMNLP*. Wu, D.; Wang, L.; and Zhang, P. 2019年. 使用变分自回归网络求解统计力学问题。《物理评论快报》, 122(8): 080602. Xiao, Z.; Yan, Q.; and Amit, Y. 2020年. 似然遗憾：一种用于变分自编码器的分布外检测分数。*NeurIPS*. Zhang, Y.; Wang, Y.; Zhang, L.; Zhang, Z.; and Gai, K. 2019年. 通过条件变分自编码器的自标注改进多样化文本生成。载于*ICASSP*. Zhao, S.; Song, J.; and Ermon, S. 2019年. InfoVAE：平衡变分自编码器中的学习与推断。载于*AAAI*.

补充材料 - eVAE：进化变分自编码器

这里我们提供正文的补充材料，包括 eVAE算法、实验设置、正方形和心形上的潜变量遍历，以及变分下界和上界的推导。

A eVAE算法

下面，算法 1 展示了 eVAE 的建模过程，该过程通过迭代方式实现变分自编码器的内外联合训练与进化学习。

算法 1：演化式内外联合训练 1：输入：交叉率 Pr_c，变异率 Pr_m，批量数据 x。2：输出：$\{\beta^*, f^*\}$3：初始化：解码器参数 θ，编码器参数 ϕ，$\{\beta_l\} = \{\beta_1, \ldots, \beta_L\}$ // 初始化群体 4：while $t < T$ do 5：从 $N(0, 1)$中采样 $Pr_t \leftarrow$ //进化概率 6：if $Pr_t < Pr_m$ then 7：θ_t，ϕ_t // 保存当前变分参数 8：通过 V-交叉操作$\mathcal{C}(\beta_{t-1}, \beta_t)$生成 β_{t+1} ，依据公式 (12)、(13) 和 (14) 9：通过 $f(\mathcal{E}(\phi_t, \theta_t, \{\beta\}, \mathcal{L}_{ELBO_t}))$评估 β_{t+1}，依据公式 (17) 和 (18) 10：else if $Pr_t < Pr_c$ then 11：θ_t，ϕ_t //保存当前变分参数 12：通过 V-变异操作 $\mathcal{M}(\beta_t)$生成 β_{t+1}，依据公式 (15) 和 (16) 13：通过适应度函数$f(\mathcal{E}(\phi_t, \theta_t, \{\beta\}, \mathcal{L}_{ELBO_t}))$评估$\beta_{t+1}$ ，依据公式 (17) 和 (18) 14：else 15：$\{\beta^*, f^*\} \leftarrow \beta_l$ //选择适应度最高的一对 16：θ_{t+1}，$\phi_{t+1} \leftarrow \mathcal{L}_{eVAE}(\theta_t, \phi_t, \beta^* | \mathbf{x}_t)$ //更新变分参数 17：end if 18：返回：$\{\beta^*, f^*\}$19：end while

B β-变分自编码器下界的推导

根据变分信息瓶颈理论，证据下界可由失真 $D = -E_{q_\phi(x,z)}[\log p_\theta(x \mid z)]$ 和率 $R = E_{q_\phi(x,z)}[D_{KL}(q_\phi(z|x)||p(z))]$，限定，其中 $I(X, Z)$ 的上界为 $I \leq R$:

$$\begin{aligned} R &= \int q(x) \int q_\theta(z \mid x) \log \frac{q_\theta(z \mid x)}{p(z)} \\ &= I(X, Z) + TC(Z) \\ &\geq I(X, Z) \end{aligned} \tag{24}$$

其下界为 $I \geq H - D$:

$$\begin{aligned} I &= \int dxdzp(x, z) \log \frac{p(x \mid z)}{p(x)} \\ &\geq \int dxdzp(x, z) \log \frac{q(x \mid z)}{p(x)} \\ &= \int dxdzp(x, z) \log q(x \mid z) \\ &- \int dxp(x) \log p(x) \\ &= \int dxdzp(x, z) \log q(x \mid z) \\ &+ H(X) \\ &= H - D \end{aligned} \tag{25}$$

总而言之，变分信息瓶颈可以表示为:

$$H - D \leq I \leq R \tag{26}$$

其中， H 是数据的熵。

在 β-变分自编码器框架中， β_C 对压缩器和解压器施加了特定于实验的约束，从而得到变分下界:

$$\begin{aligned} H - D \leq I \leq R \\ H - D \leq \beta I \leq \beta R \\ H \leq \beta I + D \leq \beta R + D \end{aligned} \tag{27}$$

总之，下界可表示为:

$$\mathcal{L}_{\beta-VAE}(\theta, \phi) = -D - \beta_C R \leq -\beta_C I - D \tag{28}$$

C eVAE 下界的推导

根据公式（10）和公式（17），我们可以得到 eVAE 在第 t 次迭代中的损失函数，如下所示:

$$\begin{aligned} \mathcal{L}_{eVAE} &= E_{\mathbf{z} \sim q_\phi(\mathbf{z}_t|\mathbf{x}_t)} \log p_\theta(\mathbf{x}_t \mid \mathbf{z}_t) + \\ &f_{\beta_t \sim R}(\beta_t, \mathcal{E}) D_{KL}(q_\phi(\mathbf{z}_t \mid \mathbf{x}_t) \| p(\mathbf{z})) \\ &= E_{x_{t+1} \sim p_\mathcal{X}}[\mathcal{L}_R + f(\beta_t)\mathcal{L}_I + \mathcal{E}(\Delta\mathcal{L}_{ELBO})] \\ &= E_{x_{t+1} \sim p_\mathcal{X}}[\mathcal{L}_R + f(\beta_t)\mathcal{L}_I \\ &+ \mathcal{E}(\Delta\mathcal{L}_{ELBO_{t+1}} + ||KL_{t+1}(\beta_{t+1}) - c||)] \end{aligned} \tag{29}$$

根据变分信息瓶颈理论， R_t 设定了第 t 次迭代中信息瓶颈 $I_t(X_t, Z_t)$ 的上界:

$$\begin{aligned} R_t &= \int q(x_t) \int q_\theta(z_t \mid x_t) \log \frac{q_\theta(z_t \mid x_t)}{p(z_t)} \\ &= I_t(X_t, Z_t) + TC(Z_t) \\ &\geq I_t(X_t, Z_t) \end{aligned} \tag{30}$$

而 D_t 设定了下界:

$$\begin{aligned} I_t &= \int dxdzp(x_t, z_t) \log \frac{p(x_t \mid z_t)}{p(x_t)} \\ &\geq \int dxdzp(x_t, z_t) \log \frac{q(x_t \mid z_t)}{p(x_t)} \\ &= \int dxdzp(x_t, z_t) \log q(x_t \mid z_t) \\ &- \int dxp(x_t) \log p(x_t) \\ &= \int dxdzp(x_t, z_t) \log q(x_t \mid z_t) \\ &+ H(X) \\ &= H - D_t \end{aligned} \tag{31}$$

经过变分进化学习器 $\mathcal{E}$优化后，受动态下界约束的 eVAE 损失函数如下:

$$\begin{aligned} \mathcal{L}_{eVAE} &= E_{\mathbf{z} \sim q_\phi(\mathbf{z}_t|\mathbf{x}_t)} \log p_\theta(\mathbf{x}_t \mid \mathbf{z}_t) + \\ & f_{\beta_t \sim R}\left(\beta_t, \mathcal{E}\right) D_{KL}(q_\phi(\mathbf{z}_t \mid \mathbf{x}_t) \| p(\mathbf{z})) \\ & = E_{x_{t+1} \sim p_{\mathcal{X}}}[\mathcal{L}_R + f(\beta_t)\mathcal{L}_I + \mathcal{E}(\Delta\mathcal{L}_{ELBO})] \\ & = E_{x_{t+1} \sim p_{\mathcal{X}}}[\mathcal{L}_R + f(\beta_t)\mathcal{L}_I \\ & + \mathcal{E}(\Delta\mathcal{L}_{ELBO_{t+1}} + ||KL_{t+1}(\beta_{t+1}) - c||)] \\ & = -D_t - \beta_t R_t - \mathcal{E}(-D_{t+1} - \beta_{t+1}R_{t+1} + D_t + c) \\ & \leq -\mathcal{E}(\beta_t)D_t - \mathcal{E}(\beta_t)I_t \end{aligned} \tag{32}$$

其中， c 是期望的 KL 点， β_t 是第 t次迭代中的一个染色体样本。

D 实验设置

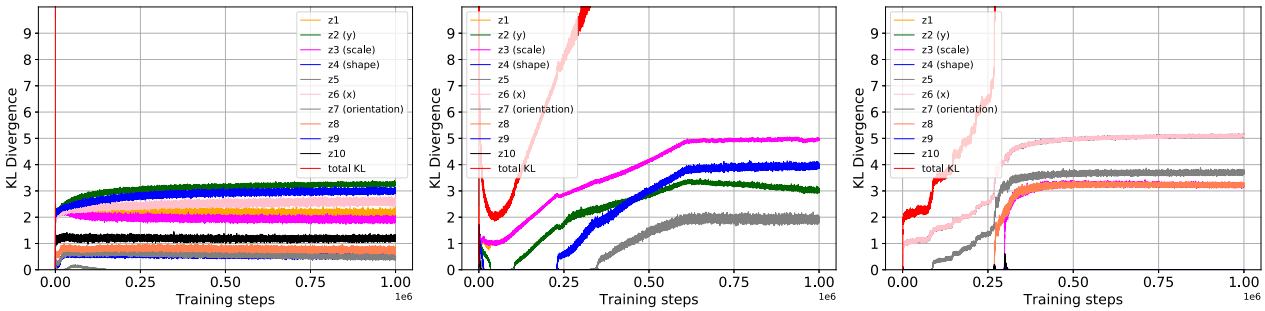
表 1 展示了用于评估 eVAE 的三个学习任务的实验设置。我们给出了在三个数据集 dSprites、CelebA 和 PTB 上的编码器-解码器网络架构设置、进化学习参数及其对应的优化器。

E 解耦表示学习中的逐元素 KL 散度分析

图 8 进一步展示了在数据集 dSprites 上，三种 VAE 模型在迭代过程中的逐元素 KL 散度演化情况： β-VAE 和 ControlAVE，以及作为对比的 eVAE。

数据集	优化器		架构	eVAE初始化	
dSprites	Adam(1e-4)	输入 编码器	(64,64,1) 卷积层(32, 32, 32), ReLU激活函数, 卷积层(32,16,16), ReLU激活函数, 卷积层(32,8,8), ReLU激活函数,卷积层(32,4,4), ReLU激活函数, 线性层(256), ReLU激活函数, 线性层(256), ReLU激活函数, 线性层(20)	变异率 交叉率 个体数量	0.001 0.04 20
		潜变量	10		
		解码器	线性层(256)， ReLU激活函数， 线性层(256), ReLU激活函数， 线性层(512)， ReLU激活函数, 卷积层(32,4,4)， ReLU激活函数， 卷积层(32,8,8), ReLU激活函数， 卷积层(32,16,16)， ReLU激活函数, 卷积层(32,32,32)， ReLU激活函数，(1, 64,64)		
CelebA （KL散度 = 200）	Adam(1e-4)	输入 编码器	(128,128,3) 卷积层(32, 32, 32), ReLU激活函数, 卷积层(32,16,16), ReLU激活函数, 卷积层(32,8,8),ReLU激活函数, 卷积层(32,4,4), ReLU激活函数, 线性层(256), ReLU激活函数, 线性层(256), ReLU激活函数, 线性层(20)	变异率 交叉率 个体数量	0.001 0.03 20
		潜变量	10		
		解码器	线性层(256), ReLU激活函数, 线性层(256), ReLU激活函数, 线性层(512), ReLU激活函数, 卷积层(32,4,4), ReLU激活函数, 卷积层(32,8,8)， ReLU激活函数, 卷积层(32,16,16)， ReLU激活函数， 卷积层(32,32,32), ReLU激活函数，(3, 64,64)		
PTB (KL散度 = 3)	初始学习率：0.001 – 阈值：100 衰减因子：0.5 最大衰减：5	输入 编码器 潜变量 解码器		变异率 交叉率 个体数量	0.006 0.07 10

表 1：在 dSprites 上进行解耦表示学习、在 CelebA 上进行图像生成以及在 PTB 上进行语言生成的实验设置。



(a) β-变分自编码器　(b) ControlVAE　(c) eVAE

图 8： β-VAE、ControlAVE 和 eVAE 在 dSprites 上逐次迭代的逐元素 KL 散度。图中显示，eVAE 的各项 KL 散度均大于 β-VAE，为生成解耦表示提供了充足空间。此外，与 ControlVAE 相比，eVAE 生成的 KL 散度更加稳定。