无分类器扩散引导

Jonathan Ho & Tim Salimans
Google Research, Brain 团队
{jonathanho,salimans}@google.com

摘要

分类器引导（Classifier Guidance）是最近提出的方法，用于在条件扩散模型训练完成后调节模态覆盖与样本保真度之间的权衡，其思想与低温采样或截断（truncation）等技术类似。分类器引导将扩散模型的得分估计与图像分类器的梯度相结合，因而需要在扩散模型之外额外训练一个图像分类器。这自然引出一个问题：能否在没有分类器的情况下进行引导？我们证明，纯生成式模型确实可以在不使用分类器的情况下实现引导——我们将此方法称为无分类器引导（Classifier-Free Guidance）。具体做法是联合训练一个条件扩散模型和一个无条件扩散模型，然后将两者的条件得分估计和无条件得分估计组合，从而得到与分类器引导相似的样本质量-多样性权衡效果。

1 引言

Diffusion models have recently emerged as an expressive and flexible family of generative models, delivering competitive sample quality and likelihood scores on image and audio synthesis tasks (??????). These models have delivered audio synthesis performance rivaling the quality of autoregressive models with substantially fewer inference steps (??), and they have delivered Im-

图 1: 64x64 ImageNet 扩散模型在 malamute 类别上的无分类器引导效果。从左到右：无分类器引导强度逐步增大，最左侧为无引导样本。

本文为 AI 辅助翻译版本，仅供学习交流；引用请以英文原文为准：arXiv:2207.12598。
本文的短版本发表于 NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications: https://openreview.net/pdf?id=qw8AKxfYbI

图 2: 引导对三个高斯分布混合模型的作用效果，每个混合分量代表以类别为条件的数据。最左侧为无引导的边缘密度。从左到右依次为不同引导强度下归一化条件密度的混合。

ageNet generation results outperforming BigGAN-deep (?) and VQ-VAE-2 (?) in terms of FID score and classification accuracy score (??).

扩散模型（Diffusion Models）近来已成为表达能力强且灵活的生成式模型家族，在图像和音频合成任务上取得了具有竞争力的样本质量和似然分数 (??????)。这些模型的音频合成性能已能匹敌自回归模型，且推理步数大幅减少 (??)；在 ImageNet 生成任务上，其 FID 分数和分类准确率得分已超越 BigGAN-deep (?) 和 VQ-VAE-2 (?) (??)。

? proposed *classifier guidance*, a technique to boost the sample quality of a diffusion model using an extra trained classifier. Prior to classifier guidance, it was not known how to generate "low temperature" samples from a diffusion model similar to those produced by truncated BigGAN (?) or low temperature Glow (?): naive attempts, such as scaling the model score vectors or decreasing the amount of Gaussian noise added during diffusion sampling, are ineffective (?). Classifier guidance instead mixes a diffusion model's score estimate with the input gradient of the log probability of a classifier. By varying the strength of the classifier gradient, ? can trade off Inception score (?) and FID score (?) (or precision and recall) in a manner similar to varying the truncation parameter of BigGAN.

? 提出了分类器引导，一种利用额外训练的分类器来提升扩散模型样本质量的技术。在分类器引导出现之前，人们尚不清楚如何从扩散模型中生成类似截断 BigGAN (?) 或低温 Glow (?) 那样的" 低温" 样本：简单的尝试——如缩放模型得分向量或减少扩散采样时加入的高斯噪声——都被证明无效 (?)。分类器引导的做法是将扩散模型的得分估计与分类器对数概率的输入梯度相混合。通过改变分类器梯度的强度，? 能够在 Inception Score (?) 和 FID 分数 (?)（或精度与召回）之间进行权衡，效果类似于调节 BigGAN 的截断参数。

We are interested in whether classifier guidance can be performed without a classifier. Classifier guidance complicates the diffusion model training pipeline because it requires training an extra classifier, and this classifier must be trained on noisy data so it is generally not possible to plug in a pre-trained classifier. Furthermore, because classifier guidance mixes a score estimate with a classifier gradient during sampling, classifier-guided diffusion sampling can be interpreted as attempting to confuse an image classifier with a gradient-based adversarial attack. This raises the question of whether classifier guidance is successful at boosting classifier-based metrics such as FID and Inception score (IS) simply because it is adversarial against such classifiers. Stepping in direction of classifier gradients also bears some resemblance to GAN training, particularly with nonparameteric generators; this also raises the question of whether classifier-guided diffusion models perform well on classifier-based metrics because they are beginning to resemble GANs, which are already known to perform well on such metrics.

我们感兴趣的问题是：能否在没有分类器的情况下实现引导？分类器引导要求额外训练一个分类器，这使扩散模型的训练流程变得更复杂；而且该分类器必须在含噪数据上训练，通常无法直接使用预训练分类器。此外，分类器引导在采样时将得分估计与分类器梯度混合，因此分类器引导的扩散采样可以理解为试图用基于梯度的对抗攻击来迷惑图像分类器。这引出一个疑问：分类器引导之所以能提升 FID 和 Inception Score 这类基于分类器的指标，是否仅仅因为它对这些分类器构成了对抗性攻击？沿着分类器梯度方向前进也与 GAN 训练有相似之处，尤其是与非参数化生成器的训练类似；这也让人怀疑，分类器引导扩散模型在基于分类器的指标上表现良好，是否因为它们开始与 GAN 趋同——而 GAN 在这类指标上本就表现优异。

To resolve these questions, we present *classifier-free guidance*, our guidance method which avoids any classifier entirely. Rather than sampling in the direction of the gradient of an image classifier, classifier-free guidance instead mixes the score estimates of a conditional diffusion model and a jointly trained unconditional diffusion model. By sweeping over the mixing weight, we attain a FID/IS tradeoff similar to that attained by classifier guidance. Our classifier-free guidance results demonstrate that pure generative diffusion models are capable of synthesizing extremely high fidelity samples possible with other types of generative models.

为了回答上述问题，我们提出*无分类器引导*，该方法完全避免了分类器的使用。无分类器引导不沿图像分类器的梯度方向采样，而是将条件扩散模型与联合训练的无条件扩散模型的得分估计进行混合。通过调节混合权重，我们获得了与分类器引导相似的 FID/IS 权衡曲线。我们的无分类器引导结果表明，纯生成式扩散模型能够合成极高保真度的样本，这一能力此前仅有其他类型生成式模型才具备。

2 背景

We train diffusion models in continuous time (**???**): letting $\mathbf{x} \sim p(\mathbf{x})$ and $\mathbf{z} = \{\mathbf{z}_\lambda \mid \lambda \in [\lambda_{\min}, \lambda_{\max}]\}$ for hyperparameters $\lambda_{\min} < \lambda_{\max} \in \mathbb{R}$, the forward process $q(\mathbf{z}|\mathbf{x})$ is the variance-preserving Markov process (**?**):

我们在连续时间下训练扩散模型 (**???**)：设 $\mathbf{x} \sim p(\mathbf{x})$，$\mathbf{z} = \{\mathbf{z}_\lambda \mid \lambda \in [\lambda_{\min}, \lambda_{\max}]\}$，其中 $\lambda_{\min} < \lambda_{\max} \in \mathbb{R}$ 为超参数。前向过程 $q(\mathbf{z}|\mathbf{x})$ 为保方差马尔可夫过程 (**?**)：

$$q(\mathbf{z}_\lambda|\mathbf{x}) = \mathcal{N}(\alpha_\lambda \mathbf{x}, \sigma_\lambda^2 \mathbf{I}), \text{ 其中 } \alpha_\lambda^2 = 1/(1+e^{-\lambda}),\ \sigma_\lambda^2 = 1 - \alpha_\lambda^2 \tag{1}$$

$$q(\mathbf{z}_\lambda|\mathbf{z}_{\lambda'}) = \mathcal{N}((\alpha_\lambda/\alpha_{\lambda'})\mathbf{z}_{\lambda'}, \sigma_{\lambda|\lambda'}^2 \mathbf{I}), \text{ where } \lambda < \lambda',\ \sigma_{\lambda|\lambda'}^2 = (1 - e^{\lambda-\lambda'})\sigma_\lambda^2 \tag{2}$$

We will use the notation $p(\mathbf{z})$ (or $p(\mathbf{z}_\lambda)$) to denote the marginal of $\mathbf{z}$ (or $\mathbf{z}_\lambda$) when $\mathbf{x} \sim p(\mathbf{x})$ and $\mathbf{z} \sim q(\mathbf{z}|\mathbf{x})$. Note that $\lambda = \log \alpha_\lambda^2/\sigma_\lambda^2$, so λ can be interpreted as the log signal-to-noise ratio of $\mathbf{z}_\lambda$, and the forward process runs in the direction of decreasing λ.

我们用 $p(\mathbf{z})$（或 $p(\mathbf{z}_\lambda)$）表示当 $\mathbf{x} \sim p(\mathbf{x})$、$\mathbf{z} \sim q(\mathbf{z}|\mathbf{x})$ 时 $\mathbf{z}$（或 $\mathbf{z}_\lambda$）的边缘分布。注意 $\lambda = \log \alpha_\lambda^2/\sigma_\lambda^2$，因此 λ 可解释为 $\mathbf{z}_\lambda$ 的对数信噪比（log SNR），前向过程沿 λ 递减方向运行。

给定 $\mathbf{x}$，前向过程的反向转移动态为 $q(\mathbf{z}_{\lambda'}|\mathbf{z}_\lambda, \mathbf{x}) = \mathcal{N}(\tilde{\boldsymbol{\mu}}_{\lambda'|\lambda}(\mathbf{z}_\lambda, \mathbf{x}), \tilde{\sigma}_{\lambda'|\lambda}^2 \mathbf{I})$，其中

$$\tilde{\boldsymbol{\mu}}_{\lambda'|\lambda}(\mathbf{z}_\lambda, \mathbf{x}) = e^{\lambda-\lambda'}(\alpha_{\lambda'}/\alpha_\lambda)\mathbf{z}_\lambda + (1 - e^{\lambda-\lambda'})\alpha_{\lambda'}\mathbf{x}, \quad \tilde{\sigma}_{\lambda'|\lambda}^2 = (1 - e^{\lambda-\lambda'})\sigma_{\lambda'}^2 \tag{3}$$

反向过程生成模型从 $p_\theta(\mathbf{z}_{\lambda_{\min}}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$ 开始。我们指定转移：

$$p_\theta(\mathbf{z}_{\lambda'}|\mathbf{z}_\lambda) = \mathcal{N}(\tilde{\boldsymbol{\mu}}_{\lambda'|\lambda}(\mathbf{z}_\lambda, \mathbf{x}_\theta(\mathbf{z}_\lambda)), (\tilde{\sigma}_{\lambda'|\lambda}^2)^{1-v}(\sigma_{\lambda|\lambda'}^2)^v) \tag{4}$$

采样时，我们沿递增序列 $\lambda_{\min} = \lambda_1 < \cdots < \lambda_T = \lambda_{\max}$ 共 T 个时间步；换言之，我们遵循 **??** 的离散时间祖先采样器。若模型 $\mathbf{x}_\theta$ 正确，则当 $T \to \infty$ 时，我们得到的样本来自某 SDE，其样本路径分布为 $p(\mathbf{z})$ (**?**)，我们以 $p_\theta(\mathbf{z})$ 表示连续时间模型分布。方差为 $\tilde{\sigma}^2_{\lambda'|\lambda}$ 和 $\sigma^2_{\lambda|\lambda'}$ 的对数空间插值，如 **?** 所建议；我们发现使用常量超参数 v 而非学习依赖 $\mathbf{z}_\lambda$ 的 v 效果很好。注意当 $\lambda' \to \lambda$ 时方差简化为 $\tilde{\sigma}^2_{\lambda'|\lambda}$，因此 v 仅在以非无穷小时间步采样时才起作用。

反向过程均值来自将估计 $\mathbf{x}_\theta(\mathbf{z}_\lambda) \approx \mathbf{x}$ 代入 $q(\mathbf{z}_{\lambda'}|\mathbf{z}_\lambda, \mathbf{x})$ (**??**)（$\mathbf{x}_\theta$ 也接收 λ 作为输入，但为简洁起见省略）。我们用 $\boldsymbol{\epsilon}$ 预测参数化 $\mathbf{x}_\theta$ (**?**)：$\mathbf{x}_\theta(\mathbf{z}_\lambda) = (\mathbf{z}_\lambda - \sigma_\lambda \boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda))/\alpha_\lambda$，并训练以下目标：

$$\mathbb{E}_{\boldsymbol{\epsilon},\lambda}\left[\|\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda) - \boldsymbol{\epsilon}\|_2^2\right] \tag{5}$$

其中 $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$，$\mathbf{z}_\lambda = \alpha_\lambda \mathbf{x} + \sigma_\lambda \boldsymbol{\epsilon}$，$\lambda$ 从 $[\lambda_{\min}, \lambda_{\max}]$ 上的分布 $p(\lambda)$ 中采样。该目标为多噪声尺度上的去噪得分匹配 (**??**) (**?**)，当 $p(\lambda)$ 为均匀分布时，该目标正比于潜变量模型 $\int p_\theta(\mathbf{x}|\mathbf{z})p_\theta(\mathbf{z})d\mathbf{z}$ 的边缘对数似然的变分下界，忽略未指定的解码器 $p_\theta(\mathbf{x}|\mathbf{z})$ 项和 $\mathbf{z}_{\lambda_{\min}}$ 处的先验项 (**?**)。

If $p(\lambda)$ is not uniform, the objective can be interpreted as weighted variational lower bound whose weighting can be tuned for sample quality (**??**). We use a $p(\lambda)$ inspired by the discrete time cosine noise schedule of **?**: we sample λ via $\lambda = -2\log\tan(au + b)$ for uniformly distributed $u \in [0, 1]$, where $b = \arctan(e^{-\lambda_{\max}/2})$ and $a = \arctan(e^{-\lambda_{\min}/2}) - b$. This represents a hyperbolic secant distribution modified to be supported on a bounded interval. For finite timestep generation, we use λ values corresponding to uniformly spaced $u \in [0, 1]$, and the final generated sample is $\mathbf{x}_\theta(\mathbf{z}_{\lambda_{\max}})$.

若 $p(\lambda)$ 非均匀，则该目标可解释为加权变分下界，其权重可针对样本质量进行调节 (**??**)。我们使用的 $p(\lambda)$ 受 **?** 的离散时间余弦噪声调度启发：通过 $\lambda = -2\log\tan(au + b)$ 采样 λ，其中 $u \in [0, 1]$ 均匀分布，$b = \arctan(e^{-\lambda_{\max}/2})$，$a = \arctan(e^{-\lambda_{\min}/2}) - b$。这表示修改为有界区间支撑的双曲正割分布。对于有限时间步生成，我们使用对应均匀间隔 $u \in [0, 1]$ 的 λ 值，最终生成样本为 $\mathbf{x}_\theta(\mathbf{z}_{\lambda_{\max}})$。

Because the loss for $\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda)$ is denoising score matching for all λ, the score $\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda)$ learned by our model estimates the gradient of the log-density of the distribution of our noisy data $\mathbf{z}_\lambda$, that is $\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda) \approx -\sigma_\lambda \nabla_{\mathbf{z}_\lambda} \log p(\mathbf{z}_\lambda)$; note, however, that because we use unconstrained neural networks to define $\boldsymbol{\epsilon}_\theta$, there need not exist any scalar potential whose gradient is $\boldsymbol{\epsilon}_\theta$. Sampling from the learned diffusion model resembles using Langevin diffusion to sample from a sequence of distributions $p(\mathbf{z}_\lambda)$ that converges to the conditional distribution $p(\mathbf{x})$ of the original data $\mathbf{x}$.

由于 $\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda)$ 的损失对所有 λ 均为去噪得分匹配，模型学到的得分 $\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda)$ 估计了含噪数据 $\mathbf{z}_\lambda$ 分布的对数密度梯度，即 $\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda) \approx -\sigma_\lambda \nabla_{\mathbf{z}_\lambda} \log p(\mathbf{z}_\lambda)$；但需注意，由于我们使用无约束神经网络定义 $\boldsymbol{\epsilon}_\theta$，可能不存在任何标量势使其梯度为 $\boldsymbol{\epsilon}_\theta$。从学到的扩散模型中采样类似于使用 Langevin 扩散从一系列分布 $p(\mathbf{z}_\lambda)$ 中采样，这些分布收敛到原始数据 $\mathbf{x}$ 的条件分布 $p(\mathbf{x})$。

In the case of conditional generative modeling, the data $\mathbf{x}$ is drawn jointly with conditioning information $\mathbf{c}$, i.e. a class label for class-conditional image generation. The only modification to the model is that the reverse process function approximator receives $\mathbf{c}$ as input, as in $\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda, \mathbf{c})$.

在条件生成建模的场景中，数据 $\mathbf{x}$ 与条件信息 $\mathbf{c}$ 联合采样，例如类别条件图像生成中的类别标签。对模型的唯一修改是反向过程函数近似器接收 $\mathbf{c}$ 作为输入，如 $\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda, \mathbf{c})$。

3 引导

An interesting property of certain generative models, such as GANs and flow-based models, is the ability to perform truncated or low temperature sampling by decreasing the variance or range of noise inputs to the generative model at sampling time. The intended effect is to decrease the diversity of the samples while increasing the quality of each individual sample. Truncation in BigGAN (**?**), for example, yields a tradeoff curve between FID score and Inception score for low and high amounts of truncation, respectively. Low temperature sampling in Glow (**?**) has a similar effect.

GAN 和基于流的模型等生成式模型具备一项有趣的性质：在采样时降低噪声输入的方差或取值范围，即可执行截断（truncated）或低温采样（low temperature sampling）。其预期效果是降低样本多样性，同时提升每个独立样本的质量。例如，BigGAN (**?**) 中的截断在低截断量和高截断量下分别产生 FID 与 Inception Score 之间的权衡曲线。Glow (**?**) 中的低温采样也有类似效果。

Unfortunately, straightforward attempts of implementing truncation or low temperature sampling in diffusion models are ineffective. For example, scaling model scores or decreasing the variance of Gaussian noise in the reverse process cause the diffusion model to generate blurry, low quality samples (**?**).

不幸的是，在扩散模型中直接实现截断或低温采样的尝试均不奏效。例如，缩放模型得分或减小反向过程中高斯噪声的方差都会导致扩散模型生成模糊、低质量的样本 (**?**)。

3.1 分类器引导

为了在扩散模型中实现类似截断的效果，**?** 引入了分类器引导（*classifier guidance*），将扩散得分 $\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda, \mathbf{c}) \approx -\sigma_\lambda \nabla_{\mathbf{z}_\lambda} \log p(\mathbf{z}_\lambda|\mathbf{c})$ 修改为包含辅助分类器模型 $p_\theta(\mathbf{c}|\mathbf{z}_\lambda)$ 的对数似然梯度，如下所示：

$$\tilde{\boldsymbol{\epsilon}}_\theta(\mathbf{z}_\lambda, \mathbf{c}) = \boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda, \mathbf{c}) - w\sigma_\lambda \nabla_{\mathbf{z}_\lambda} \log p_\theta(\mathbf{c}|\mathbf{z}_\lambda) \approx -\sigma_\lambda \nabla_{\mathbf{z}_\lambda} [\log p(\mathbf{z}_\lambda|\mathbf{c}) + w \log p_\theta(\mathbf{c}|\mathbf{z}_\lambda)],$$

其中 w 是控制分类器引导强度的参数。这一修改后的得分 $\tilde{\boldsymbol{\epsilon}}_\theta(\mathbf{z}_\lambda, \mathbf{c})$ 随后替代 $\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda, \mathbf{c})$ 用于扩散模型采样，产生近似服从以下分布的样本：

$$\tilde{p}_\theta(\mathbf{z}_\lambda|\mathbf{c}) \propto p_\theta(\mathbf{z}_\lambda|\mathbf{c}) p_\theta(\mathbf{c}|\mathbf{z}_\lambda)^w.$$

其效果是提升数据中被分类器 $p_\theta(\mathbf{c}|\mathbf{z}_\lambda)$ 赋予正确标签高似然的那些部分的概率：能被良好分类的数据在感知质量 Inception Score (**?**) 上得分较高，而该指标本身的设计就是奖励生成式模型的这一行为。**?** 因此发现，通过设置 $w > 0$，他们可以提升扩散模型的 Inception Score，但代价是样本多样性下降。

图 2 illustrates the effect of numerically solved guidance $\tilde{p}_\theta(\mathbf{z}_\lambda|\mathbf{c}) \propto p_\theta(\mathbf{z}_\lambda|\mathbf{c}) p_\theta(\mathbf{c}|\mathbf{z}_\lambda)^w$ on a toy 2D example of three classes, in which the conditional distribution for each class is an isotropic Gaussian. The form of each conditional upon applying guidance is markedly non-Gaussian. As guidance strength is increased, each conditional places probability mass farther away from other classes and towards directions of high confidence given by logistic regression, and most of the mass becomes concentrated in smaller regions. This behavior can be seen as a simplistic manifestation of the Inception score boost and sample diversity decrease that occur when classifier guidance strength is increased in an ImageNet model.

图 2 展示了在包含三个类别的玩具二维示例上，数值求解引导 $\tilde{p}_\theta(\mathbf{z}_\lambda|\mathbf{c}) \propto p_\theta(\mathbf{z}_\lambda|\mathbf{c}) p_\theta(\mathbf{c}|\mathbf{z}_\lambda)^w$ 的效果，其中每个类别的条件分布为各向同性高斯分布。施加引导

算法 1 使用无分类器引导联合训练扩散模型

Require: p_{uncond}: probability of unconditional training

repeat
  $(\mathbf{x}, \mathbf{c}) \sim p(\mathbf{x}, \mathbf{c})$ ▷ 从数据集采样带条件的数据
  $\mathbf{c} \leftarrow \varnothing$ with probability p_{uncond} ▷ 随机丢弃条件信息以进行无条件训练
  $\lambda \sim p(\lambda)$ ▷ 采样对数信噪比值
  $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  $\mathbf{z}_\lambda = \alpha_\lambda \mathbf{x} + \sigma_\lambda \boldsymbol{\epsilon}$ ▷ 将数据加噪至采样的对数信噪比值
  Take gradient step on $\nabla_\theta \left\| \boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda, \mathbf{c}) - \boldsymbol{\epsilon} \right\|^2$ ▷ 去噪模型的优化
until converged

后，各条件分布的形状明显偏离高斯分布。随着引导强度增大，每个条件分布的概率质量远离其他类别，向逻辑回归给出的高置信度方向集中，且大部分质量聚集在更小的区域内。这一行为可视为在 ImageNet 模型上增大分类器引导强度时所出现的 Inception Score 提升和样本多样性下降现象的简化体现。

对无条件模型施加权重为 $w+1$ 的分类器引导，理论上应与对条件模型施加权重为 w 的分类器引导得到相同结果，因为 $p_\theta(\mathbf{z}_\lambda|\mathbf{c})p_\theta(\mathbf{c}|\mathbf{z}_\lambda)^w \propto p_\theta(\mathbf{z}_\lambda)p_\theta(\mathbf{c}|\mathbf{z}_\lambda)^{w+1}$；或用得分表示为：

$$\begin{aligned}\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda) - (w+1)\sigma_\lambda \nabla_{\mathbf{z}_\lambda} \log p_\theta(\mathbf{c}|\mathbf{z}_\lambda) &\approx -\sigma_\lambda \nabla_{\mathbf{z}_\lambda} [\log p(\mathbf{z}_\lambda) + (w+1) \log p_\theta(\mathbf{c}|\mathbf{z}_\lambda)] \\ &= -\sigma_\lambda \nabla_{\mathbf{z}_\lambda} [\log p(\mathbf{z}_\lambda|\mathbf{c}) + w \log p_\theta(\mathbf{c}|\mathbf{z}_\lambda)],\end{aligned}$$

但有趣的是，**?** 将分类器引导施加到已有的类别条件模型上时取得了最佳结果，而非施加到无条件模型上。因此，我们也将保持在已有条件模型上做引导的设置。

3.2 无分类器引导

While classifier guidance successfully trades off IS and FID as expected from truncation or low temperature sampling, it is nonetheless reliant on gradients from an image classifier and we seek to eliminate the classifier for the reasons stated in 节 1. Here, we describe *classifier-free guidance*, which achieves the same effect without such gradients. Classifier-free guidance is an alternative method of modifying $\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda, \mathbf{c})$ to have the same effect as classifier guidance, but without a classifier. 算法 1 and 2 describe training and sampling with classifier-free guidance in detail.

虽然分类器引导能像截断或低温采样一样，成功地在 IS 和 FID 之间做出权衡，但它仍依赖图像分类器的梯度，而我们出于 节 1 中所述的原因希望消除分类器。这里我们描述*无分类器引导*（*classifier-free guidance*），它无需分类器梯度即可实现相同效果。无分类器引导是另一种修改 $\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda, \mathbf{c})$ 的方法，能达到与分类器引导相同的效果，但无需分类器。算法 1 and 2 详细描述了无分类器引导的训练和采样过程。

与其训练单独的分类器模型，我们选择联合训练无条件去噪扩散模型 $p_\theta(\mathbf{z})$（通过得分估计器 $\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda)$ 参数化）和条件模型 $p_\theta(\mathbf{z}|\mathbf{c})$（通过 $\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda, \mathbf{c})$ 参数化）。我们使用单一神经网络来参数化这两个模型，其中对于无条件模型，我们可以在预测得分时将类别标识符 $\mathbf{c}$ 简单输入为空标记 $\varnothing$，即 $\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda) = \boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda, \mathbf{c} = \varnothing)$。我们联合训练无条件和条件模型，方法很简单：以某种概率 p_{uncond} 随机将 $\mathbf{c}$ 设为无条件类别标识符 $\varnothing$，该概率设为超参数。（当然，也可以分别训练两个模型而不是联合训练，但我们选择联合训练是因为它实现起来极简单，不会使训练流程复杂化，也不会增加总参数量。）随后，我们使用以下条件得分估计和无条

算法 2 使用无分类器引导的条件采样

Require: w: guidance strength

Require: $\mathbf{c}$: conditioning information for conditional sampling

Require: $\lambda_1, \ldots, \lambda_T$: increasing log SNR sequence with $\lambda_1 = \lambda_{\min}$, $\lambda_T = \lambda_{\max}$

$\mathbf{z}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

for $t = 1, \ldots, T$ **do**

▷ 在对数信噪比 λ_t 处构造无分类器引导得分

$\tilde{\boldsymbol{\epsilon}}_t = (1 + w)\boldsymbol{\epsilon}_\theta(\mathbf{z}_t, \mathbf{c}) - w\boldsymbol{\epsilon}_\theta(\mathbf{z}_t)$

▷ 采样步骤（可替换为其他采样器，如 DDIM）

$\tilde{\mathbf{x}}_t = (\mathbf{z}_t - \sigma_{\lambda_t}\tilde{\boldsymbol{\epsilon}}_t)/\alpha_{\lambda_t}$

$\mathbf{z}_{t+1} \sim \mathcal{N}(\tilde{\boldsymbol{\mu}}_{\lambda_{t+1}|\lambda_t}(\mathbf{z}_t, \tilde{\mathbf{x}}_t), (\tilde{\sigma}^2_{\lambda_{t+1}|\lambda_t})^{1-v}(\sigma^2_{\lambda_t|\lambda_{t+1}})^v)$ if $t < T$ else $\mathbf{z}_{t+1} = \tilde{\mathbf{x}}_t$

end for

return $\mathbf{z}_{T+1}$

件得分估计的线性组合进行采样：

$$\tilde{\boldsymbol{\epsilon}}_\theta(\mathbf{z}_\lambda, \mathbf{c}) = (1 + w)\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda, \mathbf{c}) - w\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda) \tag{6}$$

公式 (6) 中不存在分类器梯度，因此在 $\tilde{\boldsymbol{\epsilon}}_\theta$ 方向上移动一步不能解释为对图像分类器进行基于梯度的对抗攻击。此外，$\tilde{\boldsymbol{\epsilon}}_\theta$ 是由得分估计构造的，由于使用了无约束神经网络，这些得分估计是非保守向量场，因此通常不存在一个标量势（如分类器对数似然）使得 $\tilde{\boldsymbol{\epsilon}}_\theta$ 成为对应的分类器引导得分。

Despite the fact that there in general may not exist a classifier for which 公式 (6) is the classifier-guided score, it is in fact inspired by the gradient of an implicit classifier $p^i(\mathbf{c}|\mathbf{z}_\lambda) \propto p(\mathbf{z}_\lambda|\mathbf{c})/p(\mathbf{z}_\lambda)$. If we had access to exact scores $\boldsymbol{\epsilon}^*(\mathbf{z}_\lambda, \mathbf{c})$ and $\boldsymbol{\epsilon}^*(\mathbf{z}_\lambda)$ (of $p(\mathbf{z}_\lambda|\mathbf{c})$ and $p(\mathbf{z}_\lambda)$, respectively), then the gradient of this implicit classifier would be $\nabla_{\mathbf{z}_\lambda} \log p^i(\mathbf{c}|\mathbf{z}_\lambda) = -\frac{1}{\sigma_\lambda}[\boldsymbol{\epsilon}^*(\mathbf{z}_\lambda, \mathbf{c}) - \boldsymbol{\epsilon}^*(\mathbf{z}_\lambda)]$, and classifier guidance with this implicit classifier would modify the score estimate into $\tilde{\boldsymbol{\epsilon}}^*(\mathbf{z}_\lambda, \mathbf{c}) = (1 + w)\boldsymbol{\epsilon}^*(\mathbf{z}_\lambda, \mathbf{c}) - w\boldsymbol{\epsilon}^*(\mathbf{z}_\lambda)$. Note the resemblance to 公式 (6), but also note that $\tilde{\boldsymbol{\epsilon}}^*(\mathbf{z}_\lambda, \mathbf{c})$ differs fundamentally from $\tilde{\boldsymbol{\epsilon}}_\theta(\mathbf{z}_\lambda, \mathbf{c})$. The former is constructed from the scaled classifier gradient $\boldsymbol{\epsilon}^*(\mathbf{z}_\lambda, \mathbf{c}) - \boldsymbol{\epsilon}^*(\mathbf{z}_\lambda)$; the latter is constructed from the estimate $\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda, \mathbf{c}) - \boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda)$, and this expression is not in general the (scaled) gradient of any classifier, again because the score estimates are the outputs of unconstrained neural networks.

尽管通常不存在分类器使得 公式 (6) 成为对应的分类器引导得分，但它实际上源于隐式分类器 $p^i(\mathbf{c}|\mathbf{z}_\lambda) \propto p(\mathbf{z}_\lambda|\mathbf{c})/p(\mathbf{z}_\lambda)$ 的梯度启发。如果我们能获取精确得分 $\boldsymbol{\epsilon}^*(\mathbf{z}_\lambda, \mathbf{c})$ 和 $\boldsymbol{\epsilon}^*(\mathbf{z}_\lambda)$（分别为 $p(\mathbf{z}_\lambda|\mathbf{c})$ 和 $p(\mathbf{z}_\lambda)$ 的得分），则该隐式分类器的梯度为 $\nabla_{\mathbf{z}_\lambda} \log p^i(\mathbf{c}|\mathbf{z}_\lambda) = -\frac{1}{\sigma_\lambda}[\boldsymbol{\epsilon}^*(\mathbf{z}_\lambda, \mathbf{c}) - \boldsymbol{\epsilon}^*(\mathbf{z}_\lambda)]$，对该隐式分类器施加分类器引导会将得分估计修改为 $\tilde{\boldsymbol{\epsilon}}^*(\mathbf{z}_\lambda, \mathbf{c}) = (1+w)\boldsymbol{\epsilon}^*(\mathbf{z}_\lambda, \mathbf{c}) - w\boldsymbol{\epsilon}^*(\mathbf{z}_\lambda)$。注意这与 公式 (6) 形式相似，但也请注意 $\tilde{\boldsymbol{\epsilon}}^*(\mathbf{z}_\lambda, \mathbf{c})$ 与 $\tilde{\boldsymbol{\epsilon}}_\theta(\mathbf{z}_\lambda, \mathbf{c})$ 有本质区别。前者由缩放后的分类器梯度 $\boldsymbol{\epsilon}^*(\mathbf{z}_\lambda, \mathbf{c}) - \boldsymbol{\epsilon}^*(\mathbf{z}_\lambda)$ 构造；后者则由估计值 $\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda, \mathbf{c}) - \boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda)$ 构造，而该表达式通常不是任何分类器的（缩放）梯度，这同样是因为得分估计是无约束神经网络的输出。

It is not obvious a priori that inverting a generative model using Bayes' rule yields a good classifier that provides a useful guidance signal. For example, **?** find that discriminative models generally outperform implicit classifiers derived from generative models, even in artificial cases where the specification of those generative models exactly matches the data distribution. In cases such as ours, where we expect the model to be misspecified, classifiers derived by Bayes' rule can

be inconsistent (?) and we lose all guarantees on their performance. Nevertheless, in 节 4, we show empirically that classifier-free guidance is able to trade off FID and IS in the same way as classifier guidance. In 节 5 we discuss the implications of classifier-free guidance in relation to classifier guidance.

用贝叶斯规则反转生成式模型，是否就能得到一个优良分类器来提供有用的引导信号? 这并非显然。例如，? 发现判别式模型通常优于从生成式模型导出的隐式分类器，即便在那些生成式模型的设定与数据分布精确匹配的人工场景中也是如此。在类似我们这种预期模型存在错误设定（misspecified）的情形中，由贝叶斯规则导出的分类器可能不一致 (?)，其性能便失去任何保证。尽管如此，在 节 4 中我们通过实验证明，无分类器引导能够以和分类器引导相同的方式在 FID 和 IS 之间做出权衡。在 节 5 中我们将讨论无分类器引导与分类器引导的关系及其含义。

4 实验

We train diffusion models with classifier-free guidance on area-downsampled class-conditional ImageNet (?), the standard setting for studying tradeoffs between FID and Inception scores starting from the BigGAN paper (?).

我们在区域降采样的类别条件 ImageNet (?) 上训练了带无分类器引导的扩散模型，这是自 BigGAN 论文 (?) 以来研究 FID 与 Inception Score 之间权衡的标准设定。

The purpose of our experiments is to serve as a proof of concept to demonstrate that classifier-free guidance is able to attain a FID/IS tradeoff similar to classifier guidance and to understand the behavior of classifier-free guidance, not necessarily to push sample quality metrics to state of the art on these benchmarks. For this purpose, we use the same model architectures and hyperparameters as the guided diffusion models of ? (apart from continuous time training as specified in 节 2); those hyperparameter settings were tuned for classifier guidance and hence may be suboptimal for classifier-free guidance. Furthermore, since we amortize the conditional and unconditional models into the same architecture without an extra classifier, we in fact are using less model capacity than previous work. Nevertheless, our classifier-free guided models still produce competitive sample quality metrics and sometimes outperform prior work, as can be seen in the following sections.

我们的实验旨在作为概念验证，证明无分类器引导能够达到与分类器引导相似的 FID/IS 权衡，并理解无分类器引导的行为特性，而非一定要在这些基准上将样本质量指标推到当前最优。为此，我们使用了与 ? 引导扩散模型相同的模型架构和超参数（除了 节 2 中指定的连续时间训练）；那些超参数是为分类器引导调优的，因此对无分类器引导而言可能并非最优。此外，由于我们将条件模型和无条件模型分摊到同一架构中而不使用额外的分类器，实际上我们使用的模型容量比先前工作更少。尽管如此，我们的无分类器引导模型仍然产生了有竞争力的样本质量指标，有时甚至超越了先前工作，如下文所示。

4.1 改变无分类器引导强度

这里我们通过实验验证本文的主要主张：无分类器引导能够以类似于分类器引导或 GAN 截断的方式在 IS 和 FID 之间进行权衡。我们将所提出的无分类器引导应用于 64×64 和 128×128 类别条件 ImageNet 生成任务。在 表 1 和 图 4 中，我们展示了在 64×64 ImageNet 模型上遍历引导强度 w 对样本质量的影响；表 2 和 图 5 展示了 128×128 模型的对应结果。我们考虑 $w \in \{0, 0.1, 0.2, \ldots, 4\}$，并遵循 ? 和 ? 的流程，对每个 w 值使用 50000 个样本计算 FID 和 Inception Score。所有模型使用对数信噪比端点 $\lambda_{\min} = -20$ 和 $\lambda_{\max} = 20$。

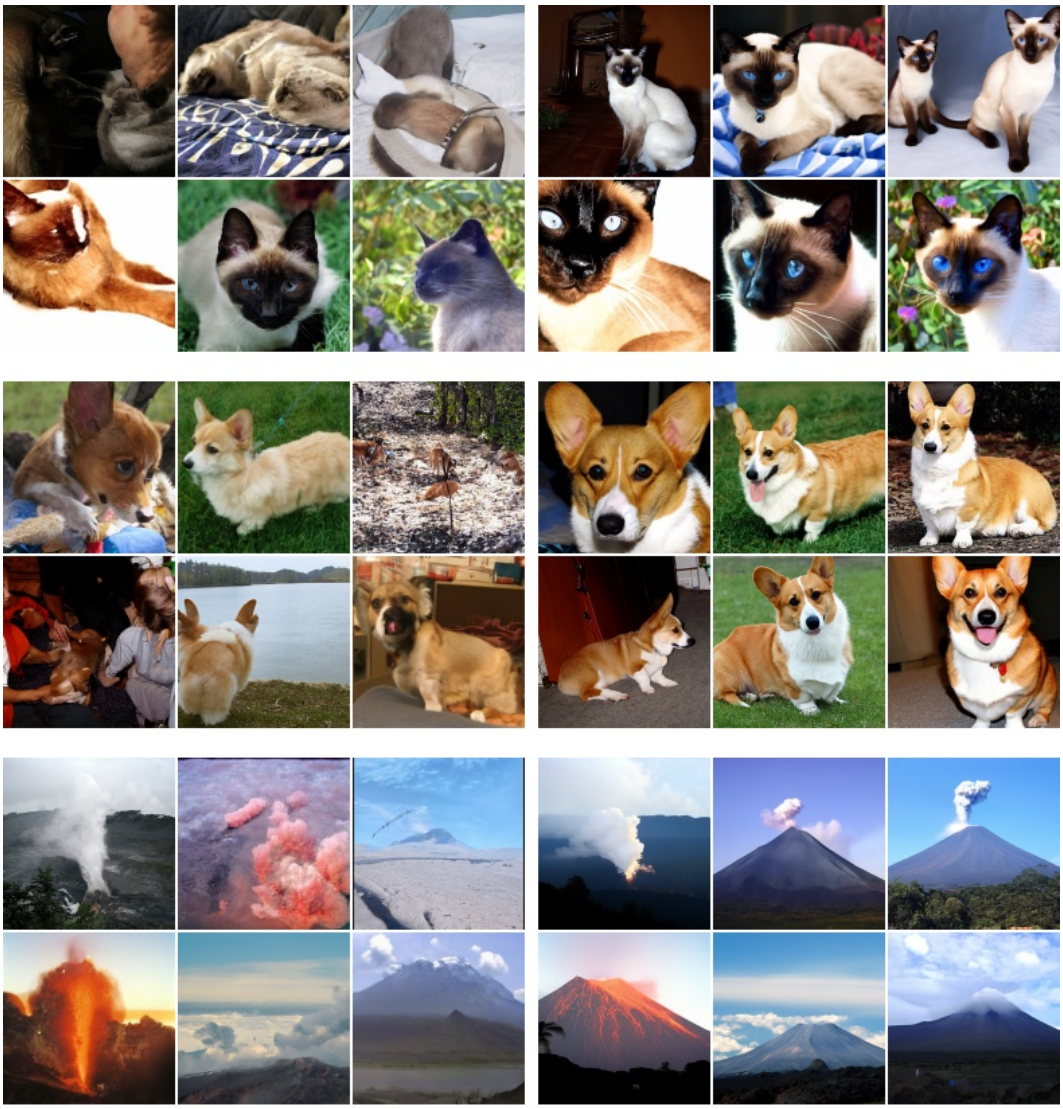


图 3: 128x128 ImageNet 上的无分类器引导。左：无引导样本，右：$w = 3.0$ 的无分类器引导样本。有趣的是，此类强引导样本呈现出饱和的色彩。更多示例见 图 8。

64×64 模型使用采样器噪声插值系数 $v = 0.3$，训练 40 万步；128×128 模型使用 $v = 0.2$，训练 270 万步。

We obtain the best FID results with a small amount of guidance ($w = 0.1$ or $w = 0.3$, depending on the dataset) and the best IS result with strong guidance ($w \geq 4$). Between these two extremes we see a clear trade-off between these two metrics of perceptual quality, with FID monotonically decreasing and IS monotonically increasing with w. Our results compare favorably to **?** and **?**, and in fact our 128×128 results are the state of the art in the literature. At $w = 0.3$, our model's FID score on 128×128 ImageNet outperforms the classifier-guided ADM-G, and at $w = 4.0$, our model outperforms BigGAN-deep at both FID and IS when BigGAN-deep is evaluated its best-IS truncation level.

我们在少量引导（$w = 0.1$ 或 $w = 0.3$，取决于数据集）下获得了最佳 FID 结果，在强引导（$w \geq 4$）下获得了最佳 IS 结果。在这两个极端之间，我们看到了这两个感知质量指标之间清晰的权衡：FID 随 w 单调递减，IS 随 w 单调递增。我们的结果优于 **?** 和 **?**，而且实际上我们的 128×128 结果是文献中的当前最优。在 $w = 0.3$ 时，我们的模型在 128×128 ImageNet 上的 FID 分数超越了分类器引导的 ADM-G；在 $w = 4.0$ 时，我们的模型在 FID 和 IS 上均超越了在其最佳 IS 截断水平下评估的 BigGAN-deep。

图 1, 3 and 6 to 8 show randomly generated samples from our model for different levels of guidance: here we clearly see that increasing classifier-free guidance strength has the expected effect of decreasing sample variety and increasing individual sample fidelity.

图 1, 3 and 6 to 8 展示了我们的模型在不同引导强度下的随机生成样本：这里我们可以清楚地看到，增加无分类器引导强度带来了预期的效果——降低样本多样性，同时提升单个样本的保真度。

Model	FID (↓)	IS (↑)
ADM (**?**)	2.07	-
CDM (**?**)	**1.48**	67.95
Ours	$p_{\text{uncond}} = 0.1/0.2/0.5$	
$w = 0.0$	1.8 / 1.8 / 2.21	53.71 / 52.9 / 47.61
$w = 0.1$	1.55 / 1.62 / 1.91	66.11 / 64.58 / 56.1
$w = 0.2$	2.04 / 2.1 / 2.08	78.91 / 76.99 / 65.6
$w = 0.3$	3.03 / 2.93 / 2.65	92.8 / 88.64 / 74.92
$w = 0.4$	4.3 / 4 / 3.44	106.2 / 101.11 / 84.27
$w = 0.5$	5.74 / 5.19 / 4.34	119.3 / 112.15 / 92.95
$w = 0.6$	7.19 / 6.48 / 5.27	131.1 / 122.13 / 102
$w = 0.7$	8.62 / 7.73 / 6.23	141.8 / 131.6 / 109.8
$w = 0.8$	10.08 / 8.9 / 7.25	151.6 / 140.82 / 116.9
$w = 0.9$	11.41 / 10.09 / 8.21	161 / 150.26 / 124.6
$w = 1.0$	12.6 / 11.21 / 9.13	170.1 / 158.29 / 131.1
$w = 2.0$	21.03 / 18.79 / 16.16	225.5 / 212.98 / 183
$w = 3.0$	24.83 / 22.36 / 19.75	250.4 / 237.65 / 208.9
$w = 4.0$	26.22 / 23.84 / 21.48	**260.2** / 248.97 / 225.1

表 1: ImageNet 64x64 结果（$w = 0.0$ 表示无引导模型）。

4.2 改变无条件训练概率

The main hyperparameter of classifier-free guidance at training time is p_{uncond}, the probability of training on unconditional generation during joint training of the conditional and unconditional diffusion models. Here, we study the effect of training models on varying p_{uncond} on 64×64 ImageNet.

无分类器引导在训练阶段的主要超参数是 p_{uncond}，即在条件扩散模型和无条件扩散模型联合训练期间用于无条件生成训练的概率。这里，我们在 64×64 ImageNet 上研究不同 p_{uncond} 对训练模型的影响。

表 1 and 图 4 show the effects of p_{uncond} on sample quality. We trained models with $p_{\text{uncond}} \in \{0.1, 0.2, 0.5\}$, all for 400 thousand training steps, and evaluated sample quality across various

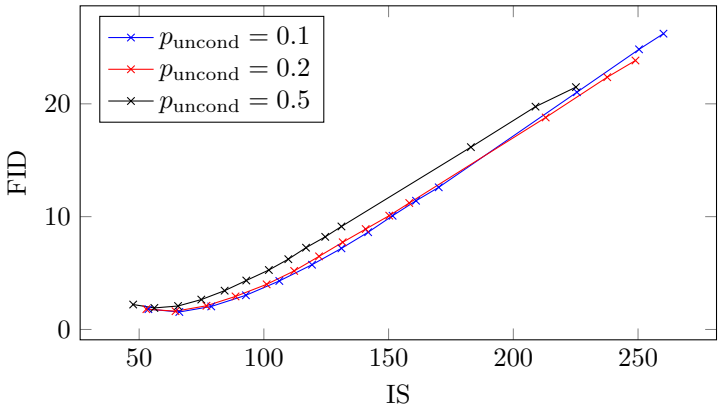


图 4: ImageNet 64x64 模型在不同引导强度下的 IS/FID 曲线。每条曲线代表一个具有不同无条件训练概率 p_{uncond} 的模型。与 表 1 配套。

guidance strengths. We find $p_{\text{uncond}} = 0.5$ consistently performs worse than $p_{\text{uncond}} \in \{0.1, 0.2\}$ across the entire IS/FID frontier; $p_{\text{uncond}} \in \{0.1, 0.2\}$ perform about equally as well as each other.

表 1 和 图 4 展示了 p_{uncond} 对样本质量的影响。我们训练了 $p_{\text{uncond}} \in \{0.1, 0.2, 0.5\}$ 的模型，均训练 40 万步，并在不同引导强度下评估样本质量。我们发现 $p_{\text{uncond}} = 0.5$ 在整个 IS/FID 前沿上的表现始终不如 $p_{\text{uncond}} \in \{0.1, 0.2\}$；$p_{\text{uncond}} \in \{0.1, 0.2\}$ 的表现大致相当。

Based on these findings, we conclude that only a relatively small portion of the model capacity of the diffusion model needs to be dedicated to the unconditional generation task in order to produce classifier-free guided scores effective for sample quality. Interestingly, for classifier guidance, **?** report that relatively small classifiers with little capacity are sufficient for effective classifier guided sampling, mirroring this phenomenon that we found with classifier-free guided models.

基于这些发现，我们得出结论：只需将扩散模型相对较小的一部分模型容量分配给无条件生成任务，就能产生对样本质量有效的无分类器引导得分。有趣的是，对于分类器引导，**?** 报告称容量很小的分类器就足以进行有效的分类器引导采样，这与我们在无分类器引导模型中发现的现象相呼应。

4.3 改变采样步数

Since the number of sampling steps T is known to have a major impact on the sample quality of a diffusion model, here we study the effect of varying T on our 128×128 ImageNet model. 表 2 and 图 5 show the effect of varying $T \in \{128, 256, 1024\}$ over a range of guidance strengths. As expected, sample quality improves when T is increased, and for this model $T = 256$ attains a good balance between sample quality and sampling speed.

由于采样步数 T 对扩散模型的样本质量有重大影响，这里我们研究改变 T 对 128×128 ImageNet 模型的影响。表 2 和 图 5 展示了在一系列引导强度下 $T \in \{128, 256, 1024\}$ 的效果。如预期所示，样本质量随 T 增大而提升，对于该模型，$T = 256$ 在样本质量和采样速度之间取得了良好的平衡。

Note that $T = 256$ is approximately the same number of sampling steps used by ADM-G (**?**), which is outperformed by our model. However, it is important to note that each sampling step for our method requires evaluating the denoising model twice, once for the conditional $\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda, \mathbf{c})$ and once for the unconditional $\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda)$. Because we used the same model architecture as ADM-G, the

fair comparison in terms of sampling speed would be our $T = 128$ setting, which underperforms compared to ADM-G in terms of FID score.

注意 $T = 256$ 与 ADM-G (**?**) 使用的采样步数大致相同，而我们的模型优于 ADM-G。不过需要指出，我们的方法每个采样步需要评估去噪模型两次，一次用于条件 $\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda, \mathbf{c})$，一次用于无条件 $\boldsymbol{\epsilon}_\theta(\mathbf{z}_\lambda)$。由于我们使用了与 ADM-G 相同的模型架构，从采样速度的公平比较来看，应该对应我们的 $T = 128$ 设定，而在该设定下我们的 FID 分数不及 ADM-G。

Model	FID (↓)	IS (↑)
BigGAN-deep, max IS (**?**)	25	253
BigGAN-deep (**?**)	5.7	124.5
CDM (**?**)	3.52	128.8
LOGAN (**?**)	3.36	148.2
ADM-G (**?**)	2.97	-
Ours	$T = 128/256/1024$	
$w = 0.0$	8.11 / 7.27 / 7.22	81.46 / 82.45 / 81.54
$w = 0.1$	5.31 / 4.53 / 4.5	105.01 / 106.12 / 104.67
$w = 0.2$	3.7 / 3.03 / 3	130.79 / 132.54 / 130.09
$w = 0.3$	3.04 / **2.43** / **2.43**	156.09 / 158.47 / 156
$w = 0.4$	3.02 / 2.49 / 2.48	183.01 / 183.41 / 180.88
$w = 0.5$	3.43 / 2.98 / 2.96	206.94 / 207.98 / 204.31
$w = 0.6$	4.09 / 3.76 / 3.73	227.72 / 228.83 / 226.76
$w = 0.7$	4.96 / 4.67 / 4.69	247.92 / 249.25 / 247.89
$w = 0.8$	5.93 / 5.74 / 5.71	265.54 / 267.99 / 265.52
$w = 0.9$	6.89 / 6.8 / 6.81	280.19 / 283.41 / 281.14
$w = 1.0$	7.88 / 7.86 / 7.8	295.29 / 297.98 / 294.56
$w = 2.0$	15.9 / 15.93 / 15.75	378.56 / 377.37 / 373.18
$w = 3.0$	19.77 / 19.77 / 19.56	409.16 / 407.44 / 405.68
$w = 4.0$	21.55 / 21.53 / 21.45	**422.29** / 421.03 / 419.06

表 2: ImageNet 128x128 结果（$w = 0.0$ 表示无引导模型）。

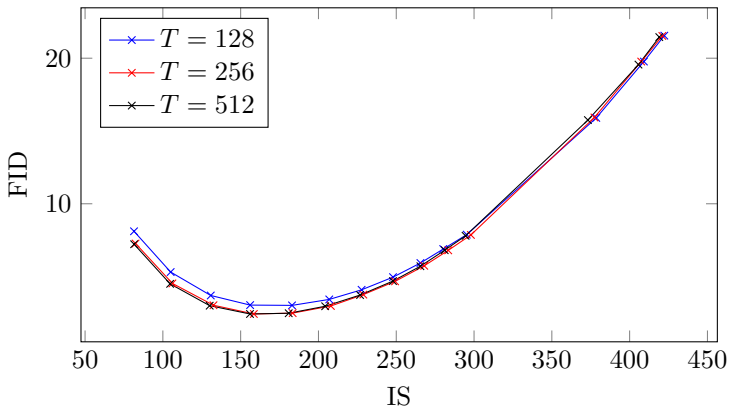


图 5: ImageNet 128x128 模型在不同引导强度下的 IS/FID 曲线。每条曲线代表以不同时间步数 T 进行采样。与 表 2 配套。

5 讨论

The most practical advantage of our classifier-free guidance method is its extreme simplicity: it is only a one-line change of code during training—to randomly drop out the conditioning—and during sampling—to mix the conditional and unconditional score estimates. Classifier guidance, by contrast, complicates the training pipeline since it requires training an extra classifier. This classifier must be trained on noisy $\mathbf{z}_\lambda$, so it is not possible to plug in a standard pre-trained classifier.

无分类器引导方法最实际的优点是其极端的简洁性：训练时只需一行代码改动——随机丢弃条件信息；采样时也只需一行——将条件得分估计和无条件得分估计混合。相比之下，分类器引导需要额外训练一个分类器，这使训练流程变得更复杂。该分类器必须在含噪 $\mathbf{z}_\lambda$ 上训练，因此无法直接使用标准的预训练分类器。

Since classifier-free guidance is able to trade off IS and FID like classifier guidance without needing an extra trained classifier, we have demonstrated that guidance can be performed with a pure generative model. Furthermore, our diffusion models are parameterized by unconstrained neural networks and therefore their score estimates do not necessarily form conservative vector fields, unlike classifier gradients (**?**). Therefore, our classifier-free guided sampler follows step directions that do not resemble classifier gradients at all and thus cannot be interpreted as a gradient-based adversarial attack on a classifier, and hence our results show that boosting the classifier-based IS and FID metrics can be accomplished with pure generative models with a sampling procedure that is not adversarial against image classifiers using classifier gradients.

由于无分类器引导无需额外训练分类器就能像分类器引导一样在 IS 和 FID 之间进行权衡，我们证明了引导可以仅凭纯生成式模型来实现。此外，我们的扩散模型由无约束神经网络参数化，因此其得分估计不一定形成保守向量场，这与分类器梯度不同 (**?**)。因此，我们的无分类器引导采样器所遵循的步方向与分类器梯度完全不同，不能解释为对分类器的基于梯度的对抗攻击。我们的结果由此表明，提升基于分类器的 IS 和 FID 指标可以通过纯生成式模型来实现，其采样过程并不利用分类器梯度对图像分类器进行对抗性攻击。

We also have arrived at an intuitive explanation for how guidance works: it decreases the unconditional likelihood of the sample while increasing the conditional likelihood. Classifier-free guidance accomplishes this by decreasing the unconditional likelihood with a *negative* score term, which to our knowledge has not yet been explored and may find uses in other applications.

我们还得到了关于引导工作机制的直观解释：它降低样本的无条件似然，同时提高其条件似然。无分类器引导通过一个负得分项来降低无条件似然，据我们所知，这种做法尚未被探索，或许能在其他应用中找到用途。

Classifier-free guidance as presented here relies on training an unconditional model, but in some cases this can be avoided. If the class distribution is known and there are only a few classes, we can use the fact that $\sum_{\mathbf{c}} p(\mathbf{x}|\mathbf{c})p(\mathbf{c}) = p(\mathbf{x})$ to obtain an unconditional score from conditional scores without explicitly training for the unconditional score. Of course, this would require as many forward passes as there are possible values of $\mathbf{c}$ and would be inefficient for high dimensional conditioning.

本文所呈现的无分类器引导依赖于训练无条件模型，但在某些情况下可以避免。如果类别分布已知且只有少量类别，我们可以利用 $\sum_{\mathbf{c}} p(\mathbf{x}|\mathbf{c})p(\mathbf{c}) = p(\mathbf{x})$ 从条件得分中获取无条件得分，而无需显式训练无条件得分。当然，这需要执行与 $\mathbf{c}$ 可能取值数量同样多的前向传播，对于高维条件来说效率低下。

A potential disadvantage of classifier-free guidance is sampling speed. Generally, classifiers can be smaller and faster than generative models, so classifier guided sampling may be faster than classifier-free guidance because the latter needs to run two forward passes of the diffusion model, one for conditional score and another for the unconditional score. The necessity to run multiple passes of the diffusion model might be mitigated by changing the architecture to inject conditioning late in the network, but we leave this exploration for future work.

无分类器引导的一个潜在缺点是采样速度。通常，分类器可以比生成式模型更小更快，因此分类器引导采样可能比无分类器引导更快，因为后者需要运行两次扩散模型的前向传播，一次用于条件得分，另一次用于无条件得分。通过修改架构将条件注入推迟到网络靠后的位置，或许可以缓解多次运行扩散模型的需求，但我们将此探索留待未来工作。

Finally, any guidance method that increases sample fidelity at the expense of diversity must face the question of whether decreased diversity is acceptable. There may be negative impacts in deployed models, since sample diversity is important to maintain in applications where certain parts of the data are underrepresented in the context of the rest of the data. It would be an interesting avenue of future work to try to boost sample quality while maintaining sample diversity.

最后，任何以牺牲多样性为代价来提高样本保真度的引导方法都必须面对一个问题：多样性降低是否可接受。在实际部署的模型中，这可能会带来负面影响，因为在数据的某些部分相对于其他部分代表性不足的应用场景中，保持样本多样性至关重要。尝试在保持样本多样性的同时提升样本质量，将是未来工作中一个有趣的研究方向。

6 结论

We have presented classifier-free guidance, a method to increase sample quality while decreasing sample diversity in diffusion models. Classifier-free guidance can be thought of as classifier guidance without a classifier, and our results showing the effectiveness of classifier-free guidance confirm that pure generative diffusion models are capable of maximizing classifier-based sample quality metrics while entirely avoiding classifier gradients. We look forward to further explorations of classifier-free guidance in a wider variety of settings and data modalities.

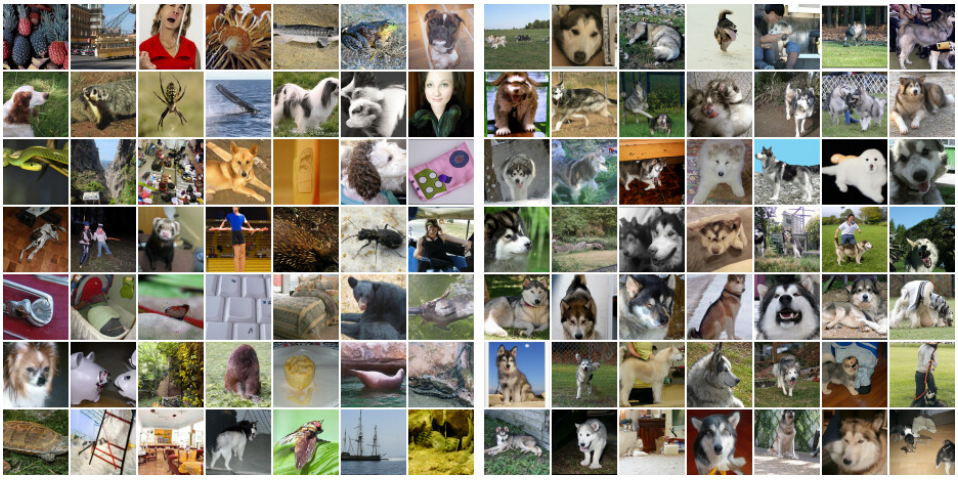
我们提出了无分类器引导，一种在扩散模型中提高样本质量同时降低样本多样性的方法。无分类器引导可以理解为不带分类器的分类器引导，而我们展示其有效性的结果证实了纯生成式扩散模型完全有能力在最大化基于分类器的样本质量指标的同时，彻底避开分类器梯度。我们期待在更广泛的设定和数据模态中进一步探索无分类器引导。

7 致谢

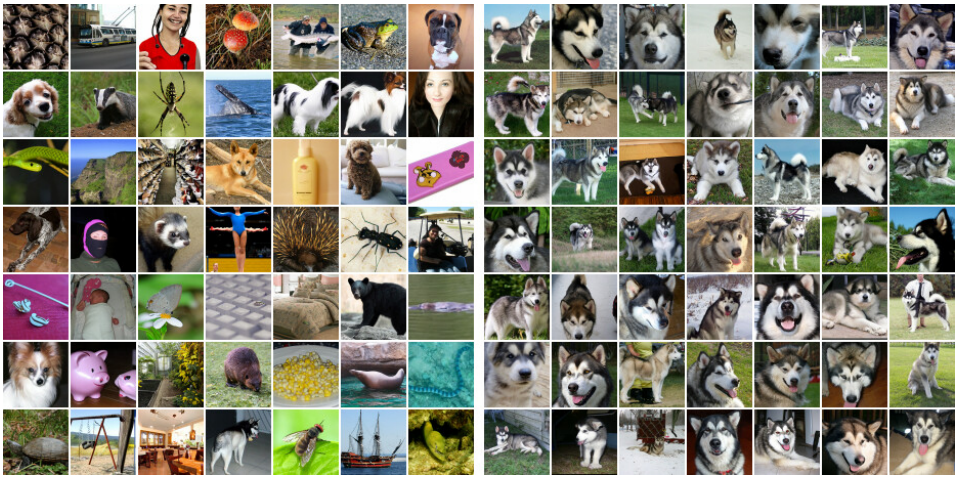
感谢 Ben Poole 和 Mohammad Norouzi 的讨论。

参考文献

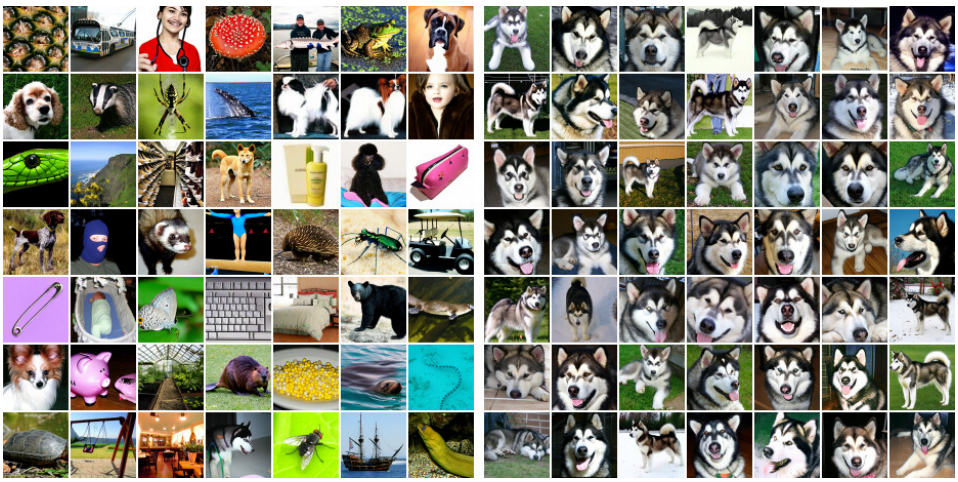
A 更多样本



(a) 无引导条件采样：FID=1.80，IS=53.71

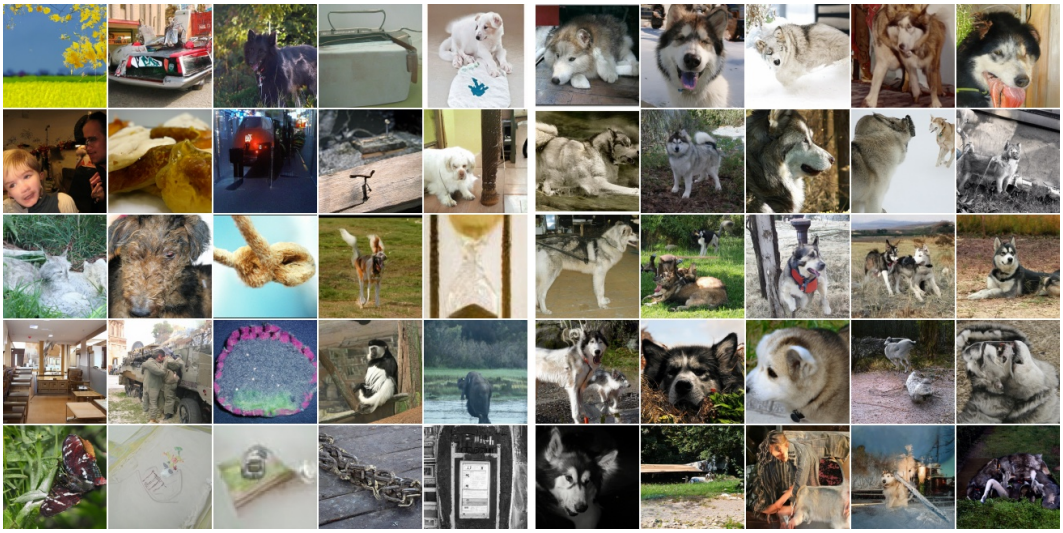


(b) 无分类器引导（$w = 1.0$）：FID=12.6，IS=170.1

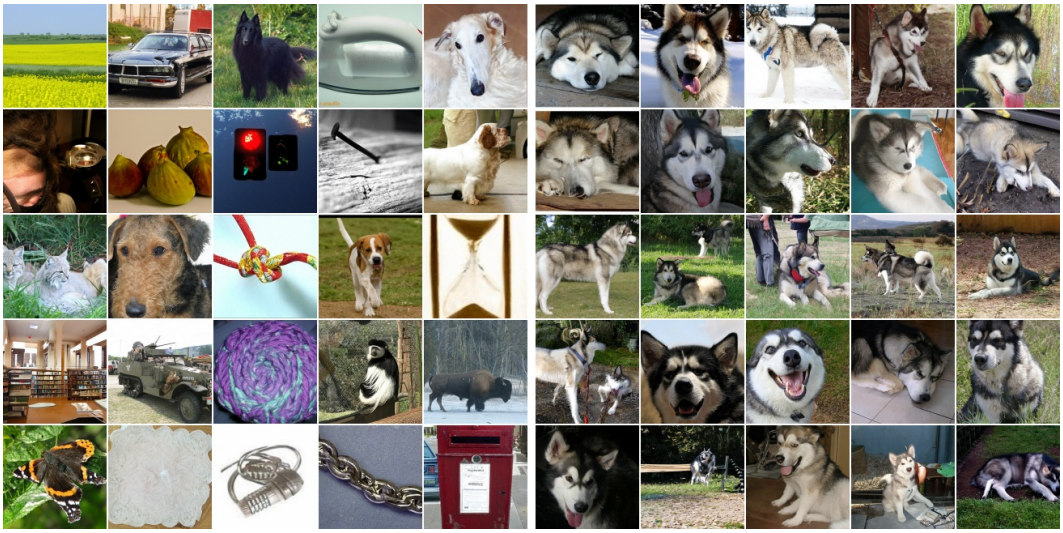


(c) 无分类器引导（$w = 3.0$）：FID=24.83，IS=250.4

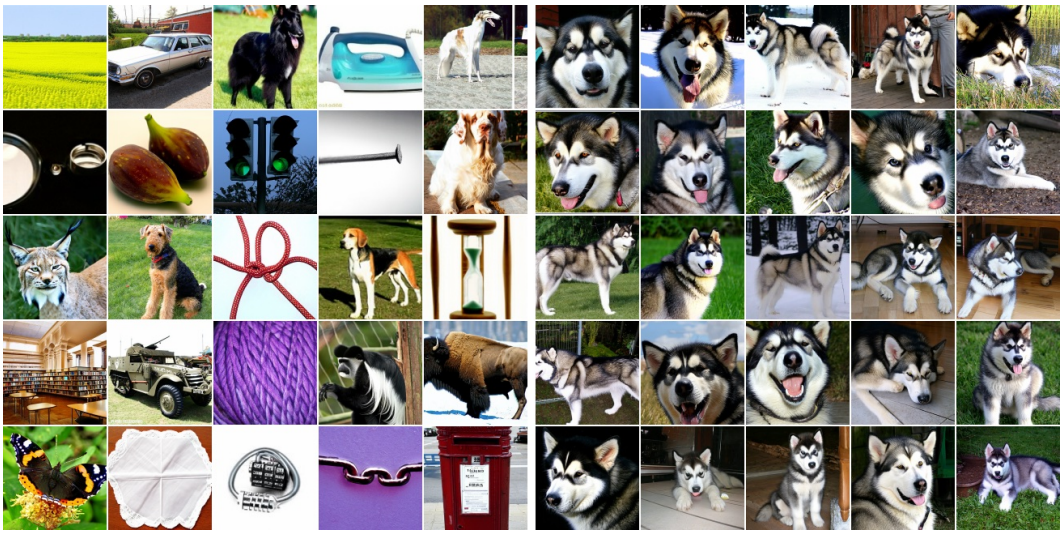
图 6: ImageNet 64x64 上的无分类器引导。左：随机类别。右：单一类别（malamute）。各子图采样使用相同的随机种子。



(a) 无引导条件采样：FID=7.27，IS=82.45



(b) 无分类器引导（$w = 1.0$）：FID=7.86，IS=297.98



(c) 无分类器引导（$w = 4.0$）：FID=21.53，IS=421.03

图 7: ImageNet 128x128 上的无分类器引导。左：随机类别。右：单一类别（malamute）。各子图采样使用相同的随机种子。

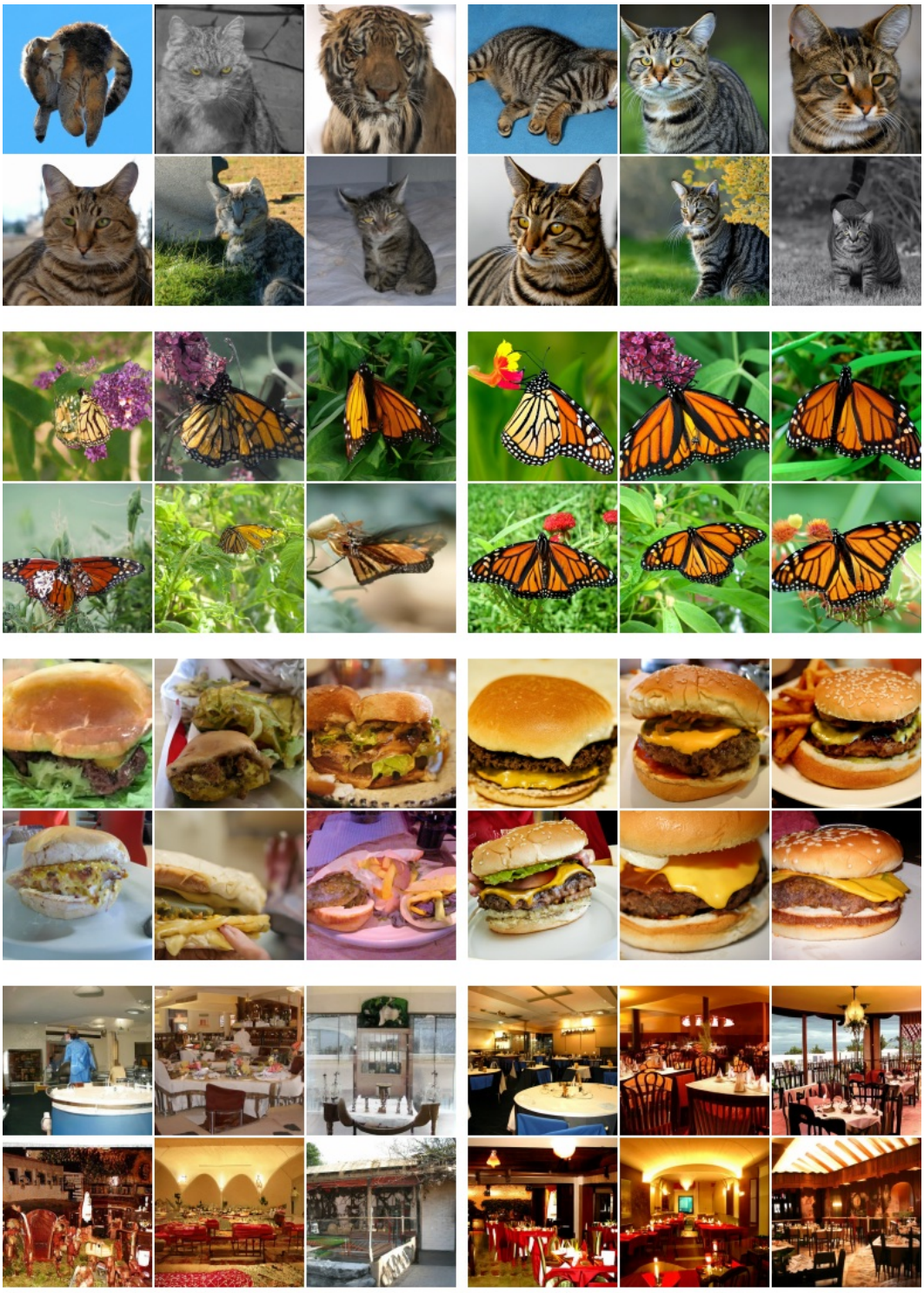


图 8: 128x128 ImageNet 上更多无分类器引导示例。左：无引导样本，右：无分类器引导样本（$w = 3.0$）。