Joint Diffusion Sampling via Positive-Unlabeled Guidance for Multi-Modal Data

Matt Raymond [1] **Yilun Zhu** [1] **Jianxin Zhang** [1] **Angela Violi** [1 2] **Clayton Scott** [1]

Abstract

Multi-modal generative models typically require abundant training data from multi-modal joint distributions, which is often unavailable in the life sciences. We propose to treat each modality as a marginal distribution and correct their independent diffusion processes to sample from their joint distribution. Specifically, we introduce "joint diffusion sampling," a method that generates a sample from joint distributions using pre-trained models for individual (uni-modal) marginal distributions and minimal data from the (multi-modal) joint distribution. We demonstrate preliminary uni- and multi-modal results for images, molecules, and Boolean values, and discuss multi-modal applications of our approach.

1. Introduction

In the life sciences, multi-modal generative foundation models face a critical bottleneck: acquiring high-quality multi-modal datasets is often challenging due to ethical, regulatory, or practical barriers (Lvovs et al., 2025; Liu et al., 2025). While foundation models for individual modalities (*e.g.*, for proteins or molecules) are becoming increasingly common (Xu et al., 2023; Liu et al., 2024; Li et al., 2024; Guo et al., 2024; Abramson et al., 2024; Wang et al., 2024a; Bachimanchi & Volpe, 2025; Wang et al., 2025), data-scarcity remains a challenge for multi-modal models.

Data-limited problems with compositional structures have been successfully modeled using diffusion models by composing pre-trained models (Liu et al., 2021; Du et al., 2023; Geng et al., 2024a;b; Wu et al., 2024; Skreta et al., 2025) using simple transformations (*e.g.*, addition, subtraction). In contrast to existing methods, which apply multiple diffusion models to the same feature vector (*e.g.*, $s_1(\boldsymbol{x}) + s_2(\boldsymbol{x})$), our work coordinates pretrained foundation models for different feature vectors (*e.g.*, $[s_1(\boldsymbol{x}_1)^\top \; s_2(\boldsymbol{x}_2)^\top]^\top$).

We introduce *joint diffusion sampling* (JDS), where the goal is to generate a pair (A, B) of variables following a particular joint distribution. For each of these two variables, a pre-trained diffusion foundation model is available, and a limited amount of training data for the joint distribution is also available. We propose a solution that guides a sample from this compositional model toward the target distribution using a classifier trained via positive-unlabeled learning. Unlike prior works, we do not require an estimate of the prior class probabilities, which is frequently difficult to estimate.

1.1. Joint sampling with positive-unlabeled data

Let $\mathcal{X}_1, \mathcal{X}_2$ denote two feature spaces of potentially-different modalities, and let $\mathcal{X} \coloneqq \mathcal{X}_1 \times \mathcal{X}_2$ denote the product feature space. Furthermore, let a bold symbol (*e.g.* $\boldsymbol{x}$) indicate a feature vector and let $f_1(\boldsymbol{x}_1), f_2(\boldsymbol{x}_2)$ be marginal probability densities on $\mathcal{X}_1, \mathcal{X}_2$ from which it is possible to draw samples efficiently. Let $f(\boldsymbol{x})$ be a joint probability density on $\mathcal{X}$. Given the densities f_1, f_2 and a training sample drawn from f, *joint sampling* refers to the problem of generating a new sample from f. We are particularly interested in the setting where the training sample size from f is small, so that directly training a generative model for f is impractical.

In this work, f_1, f_2 are accessible through pre-trained diffusion models, in which case we refer to the problem as *joint diffusion sampling* (JDS). Performance may be measured using a metric on a suitable feature space, such as the CLIP-MMD score (Jayasumana et al., 2024).

1.2. Contributions and outline

Our contributions are that we **1)** solve the JDS problem with classifier guidance and **2)** demonstrate that our Positive-Unlabeled (PU) guidance is competitive with oracle classifier guidance *even with a label rates as low as 0.0012%*.

We first detail the theoretical basis for our work, then discuss our contribution in detail, before evaluating our approach and discussing our limitations and related work.

[1]Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, United States [2]Department of Chemical Engineering, University of Michigan, Ann Arbor, MI, United States. Correspondence to: Angela Violi <avioli@umich.edu>.

Proceedings of the ICML 2025 Workshop on Multi-modal Foundation Models and Large Language Models for Life Sciences, Vancouver, Canada. 2025.

2. Background and related work

We briefly introduce copulas, diffusion models, posterior guidance, and PU classification to set up our contributions.

2.1. Copulas

Our work is inspired by the theory of copulas. A copula $C : [0,1]^2 \to [0,1]$ is a joint cumulative distribution function (CDF) with uniform marginals (Nelsen, 2006). Let an un-bolded symbol (*e.g.*, x) denote a scalar. Furthermore, suppose that $\boldsymbol{x} \coloneqq [x_1 \ x_2]^\top$, and let $F_1(x_1), F_2(x_2)$, and $F(\boldsymbol{x})$ be the CDFs corresponding to $f_1(x_1), f_2(x_2)$, and $f(\boldsymbol{x})$. By Sklar's theorem (Sklar, 1959), there exists a copula C such that $F(\boldsymbol{x}) = C(F(x_1), F_2(x_2))$. Similarly,

$$f(\boldsymbol{x}) = f_1(x_1) f_2(x_2) c(F(x_1), F(x_2)) \ . \tag{1}$$

Taking the logarithm of both sides, we get

$$\log f(\boldsymbol{x}) = \log f_1(x_1) f_2(x_2) + \log c(F(x_1), F(x_2)) \ ,$$

which is the inspiration for our approach.

2.2. Diffusion models

A *diffusion model* is a generative model defined in terms of two random processes, the forward and reverse diffusion processes. The forward process is a random process $\{\mathbf{X}(t)\}_{t=0}^{T}$ that is governed by a stochastic differential equation (SDE) that stochastically transforms the target data distribution $\mathbf{X}(0)$ to a distribution $\mathbf{X}(T)$ that is easily sampled from, by means of a simple mechanism. Let capital letters (*e.g.*, X) indicate random variables, and let $\mathbf{X}(t)$ have density $f(\boldsymbol{x}; t)$ at time t. If the *score* $s(\boldsymbol{x}; t) \coloneqq \nabla_{\boldsymbol{x}} \log f(\boldsymbol{x}; t)$ is known, one can sample from $f(\boldsymbol{x}; T)$ and stochastically transform back to $\mathbf{X}(0)$ by means of the reverse diffusion process (Anderson, 1982). Typically in practice, the interval $[0, T]$ is discretized, the forward process increasingly corrupts data with Gaussian noise, and the score is estimated from data using score-matching (Song & Ermon, 2019). For $t = 0$, the training data is given, and for larger t, it is synthetically generated by propagating the training data through the forward process.

2.3. Classifier guidance

Often a diffusion model is used to generate data from several classes. Classifier guided diffusion is a technique for sampling from a specific class that does not require retraining the model (Song et al., 2021b). In the context of two classes, let $Y \in \mathcal{Y} \coloneqq \{+1, -1\}$ be a binary label, and supposed that labeled training data $(\mathbf{X}^i, Y^i)$, $i \in \{1, \ldots, n\}$ is given. The idea behind classifier guided diffusion is to replace $f(\boldsymbol{x}; t)$ with $f(\boldsymbol{x}|Y = +1; t)$ in the definition of the score. By Bayes' rule, the "adjusted" score function is

$$\nabla_{\boldsymbol{x}} \log f(\boldsymbol{x}|y; t) = s(\boldsymbol{x}; t) + \nabla_{\boldsymbol{x}} \log \eta(\boldsymbol{x}; t) \ , \tag{2}$$

where $\eta(\boldsymbol{x}; t) \coloneqq \mathbb{P}(Y = +1|\mathbf{X}(t) = \boldsymbol{x}, t)$. The problem of estimating $\eta(\boldsymbol{x}; t)$ is known as *class probability estimation* (CPE) and can be solved by empirical risk minimization (ERM) with cross entropy loss using the synthetically-generated training data for $(\mathbf{X}(t), Y)$ at each time-step t.

2.4. Positive-unlabeled class probability estimation

As mentioned above, given jointly distributed $(\mathbf{X}, Y)$, CPE is the problem of estimating $\eta(\boldsymbol{x}) \coloneqq \mathbb{P}(Y = +1|\mathbf{X} = \boldsymbol{x})$ from labeled training data. In classifier-guided diffusion, this problem is solved at each t, but here we suppress the dependence on t. The problem of positive-unlabeled (PU) CPE is to estimate $\eta(\boldsymbol{x})$ from positive and unlabeled data.

Following Ivanov (2020), PU-CPE can be solved as follows. Let $f_{\mathrm{p}}(\boldsymbol{x})$ and $f_{\mathrm{n}}(\boldsymbol{x})$ denote the class-conditional densities $\mathbb{P}(\boldsymbol{x}|Y = +1)$ and $\mathbb{P}(\boldsymbol{x}|Y = -1)$. Then the marginal density of $\mathbf{X}$ is

$$f_{\mathrm{u}}(\boldsymbol{x}) \coloneqq \pi f_{\mathrm{p}}(\boldsymbol{x}) + (1 - \pi) f_{\mathrm{n}}(\boldsymbol{x}) \ , \tag{3}$$

where $\pi = \mathbb{P}(Y = +1)$ is the prior probability of the positive class. Furthermore, by Bayes rule,

$$\eta(\boldsymbol{x}) = \Pr(Y = 1|\mathbf{X} = \boldsymbol{x}) = \pi \frac{f_{\mathrm{p}}(\boldsymbol{x})}{f_{\mathrm{u}}(\boldsymbol{x})} \ . \tag{4}$$

Ivanov (2020)'s insight is that

$$\eta(\boldsymbol{x}) = \pi \frac{g(\boldsymbol{x})}{1 - g(\boldsymbol{x})} \ , \tag{5}$$

where

$$g(\boldsymbol{x}) \coloneqq \frac{f_{\mathrm{p}}(\boldsymbol{x})}{f_{\mathrm{p}}(\boldsymbol{x}) + f_{\mathrm{u}}(\boldsymbol{x})} \ . \tag{6}$$

Furthermore, $g(\boldsymbol{x})$, referred to as the *balanced positive-unlabeled (BPU) posterior*[1] can be learned by balanced ERM (Appendix A). π can be estimated under distributional assumptions (Elkan & Noto, 2008; Blanchard et al., 2010).

3. PU guidance for joint diffusion sampling

In this section, we show how to reduce JDS to PU-guided sampling, while also introducing a general technique for PU classifier guidance that avoids estimation of π.

3.1. The fundamental invariance of PU guidance

Classifier-guided diffusion requires estimation of $f_{\mathrm{p}}(\boldsymbol{x}; t)$ for each of the discrete values of t in the diffusion model. By combining (4) and (5), and re-introducing t, observe that

$$\frac{f_{\mathrm{p}}(\boldsymbol{x}; t)}{f_{\mathrm{u}}(\boldsymbol{x}; t)} = \frac{g(\boldsymbol{x}; t)}{1 - g(\boldsymbol{x}; t)} \ , \tag{7}$$

[1] Elkan & Noto (2008); Bekker & Davis (2020) call this the "nontraditional classifier"

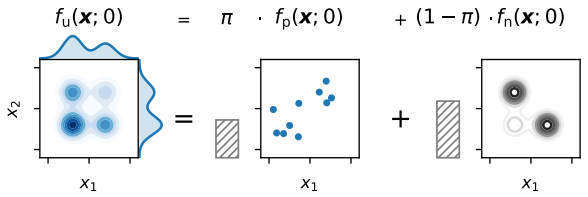

Figure 1. An illustration of how the JDS problem can be reformulated as a BPU posterior estimation problem, with a sample from $f_p(\cdot;0)$ and with $f_u(\boldsymbol{x};0) \coloneqq f_1(\boldsymbol{x}_1;0)f(\boldsymbol{x}_2;0)$. The bars represent the prior probabilities (*e.g.*, π). Blue indicates known quantities, and gray indicates unknown quantities.

and therefore

$$\nabla_{\boldsymbol{x}} \log f_p(\boldsymbol{x};t) = \nabla_{\boldsymbol{x}} \log \left[f_u(\boldsymbol{x};t) \frac{g(\boldsymbol{x};t)}{1-g(\boldsymbol{x};t)} \right] . \quad (8)$$

Thus, the guidance term does not require estimation of π, unlike existing PU methods (Na et al., 2024). We call this approach *prior-free PU guidance*. Additional advantages of this approach are discussed in Section 6.

PU-guidance may be applied to JDS as follows. Recall the notation from Section 1.1, and consider the PU learning problem where

$$f_u(\boldsymbol{x};t) \coloneqq f_1(\boldsymbol{x}_1;t)f_2(\boldsymbol{x}_2;t) \quad f_p(\boldsymbol{x};t) \coloneqq f(\boldsymbol{x};t), \quad (9)$$

so that

$$f_1(\boldsymbol{x}_1;t)f_2(\boldsymbol{x}_2;t) = \pi f(\boldsymbol{x};t) + (1-\pi) f_n(\boldsymbol{x};t) . \quad (10)$$

This means that the product distribution $f_1(\boldsymbol{x}_1)f_2(\boldsymbol{x}_2)$ is a mixture of the joint distribution $f(\boldsymbol{x})$ (which we care about) and some other density $f_n(\boldsymbol{x})$ (which we don't) for a suitable prior probability π. We provide necessary and sufficient conditions for the existence of such π and f_n in Appendix B, and an illustration in Figure 1.

This perspective directly enables PU diffusion guidance in the context of JDS. Let $\{\mathbf{X}_1(t)\}_{t=1}^T$ and $\{\mathbf{X}_2(t)\}_{t=1}^T$ be the forward processes for the two given diffusion models, and set $\mathbf{X}(t) = [\mathbf{X}_1(t)^\top \ \mathbf{X}_2(t)^\top]^\top$. Further let s_1, s_2 be the scores associated with the probability densities f_1, f_2. Plugging this into (8), we get an additive correction of independent diffusion processes:

$$\nabla_{\boldsymbol{x}} \log f(\boldsymbol{x}|y;t) = \begin{bmatrix} s_1(\boldsymbol{x}_1;t) \\ s_2(\boldsymbol{x}_2;t) \end{bmatrix} + \nabla_{\boldsymbol{x}} \log \frac{g(\boldsymbol{x};t)}{1-g(\boldsymbol{x};t)} . \quad (11)$$

Note the similarities between (2) and (11). The first term in (11) may be interpreted as the score of a product distribution, which corresponds to the unconditional diffusion process in (2). The second term in (11) is the score of the posterior from (2), expressed using g. Furthermore, this additive update is reminiscent of a log-copula, which transforms a product distribution to a joint distribution (see Section 2.1). Available diffusion models provide for the first term, and the second term can be learned via PU CPE, using the small joint dataset as the positive data, and paired, independent examples from f_1 and f_2 (or the corresponding diffusion models) as the unlabeled data.

4. Experiments and results

We demonstrate experimentally the application of prior-free (PF) PU guidance JDS. Specifically, we show that the sample quality of PF guidance is equivalent to that of fully-supervised guidance for uni- and multi-modal data. Further details and results are in Appendices C to I.

4.1. Dataset creation

Each experiment considers two data types (modalities) which are in some cases the same. To generate the "jointly distributed" data, we start with existing datasets for each modality, and apply certain criteria to select which pairs of data points (one from each modality) are draws from the joint distribution f_p. This strategy inherently also gives us pairs from f_n, namely, the pairs that do not meet the criteria.

4.2. Sampling approaches and diffusion models

We compare our prior-free PU guidance against fully-supervised positive-negative (PN) classifier guidance. Using PN and PU data, the PN model is trained by ERM with cross entropy loss, and the BPU posterior is estimated using balanced ERM with cross entropy. We use Denoising Diffusion Probabilistic Models (DDPM) (Ho et al., 2020) for images and Boolean values, and Geometric Latent Diffusion Models (Geo-LDM) (Xu et al., 2023) for molecules. Further training details are available in Appendix C.

4.3. Datasets and evaluation metrics

We systematically introduce the generation tasks and metrics that we use to compare PN and PF PU guidance.

4.3.1. Uni-modal: handwritten digits

Given a diffusion model for handwritten digits and a small set of same-parity pairs of digits, the goal is to generate more same-parity pairs of digits. Specifically, same-parity digits comprise the positive distribution, different-parity digits comprise the negative distribution, and random pairs of digits comprise the unlabeled distribution. We first train a diffusion model to generate a single image from the MNIST

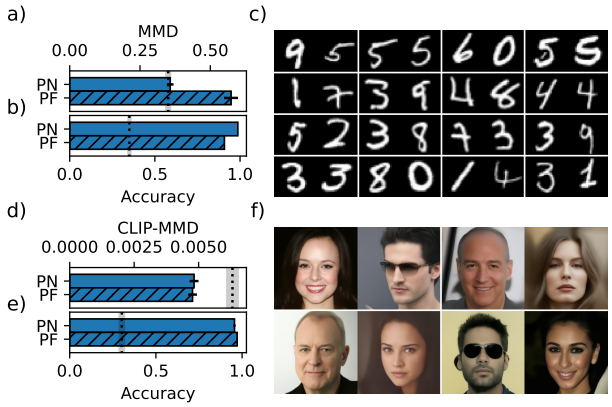

Figure 2. Jointly sampled images. **a)** and **b)** contain the MMD and accuracy for same-parity numbers. **c)** Random sample using PF PU guidance. **d)** and **e)** contain the MMD and accuracy for pairs of differently-gendered faces. **f)** Random sample using PF PU guidance. The vertical dotted line indicates the performance of an unconditional diffusion model, and the black bars and gray region indicate the bootstrapped standard deviation.

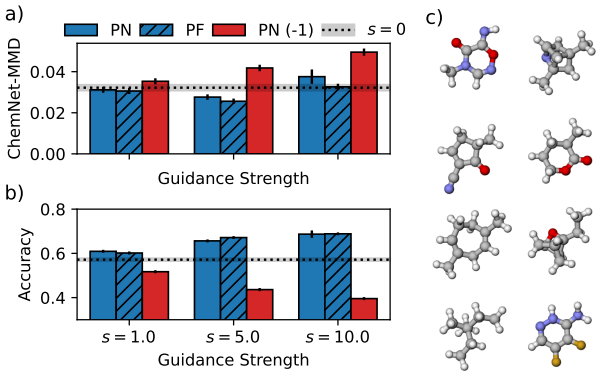

Figure 3. Jointly sampled molecules. **a)** and **b)** contain the ChemNet-MMD and accuracy for pairs of similar-dipole-moment molecules. **c)** Random sample of molecule pairs from PF PU guidance. "PN (-1)" indicates guidance towards the negative class. The horizontal dotted line indicates the performance of an unconditional diffusion model, and the black bars and gray region indicate the bootstrapped standard deviation. "Guidance strength" is a scalar that is used to up-weight the contribution of the guidance $\nabla \log \eta(\boldsymbol{x}; t)$.

dataset (LeCun et al., 1998), and train g using $60{,}000^2$ unlabeled and 10,000 positive image pairs, which results in a 0.0012% label rate for positive samples.

We compute the accuracy as the fraction of $\sim$10,000 generated image pairs that are same-parity digits. We also evaluate the maximum mean discrepancy (MMD) (Gretton et al., 2012) between $\sim$5,000 same-parity pairs of MNIST test digits and a generated sample, which is computed using the concatenation of flattened images.

In Figures 2a to 2c, we find that prior-free PU guidance is not quite as good as PN guidance, particularly with respect to the MMD. This may be because the positive class has many modes (25) relative to the sample size (10,000).

4.3.2. Uni-modal: human faces

Given a diffusion model for human faces and a small set of different-gender pairs of faces, the goal is to generate more different-gender pairs of faces. Specifically, different-gender pairs comprise the positive distribution, same-gender pairs comprise the negative distribution, and random pairs of faces comprise the unlabeled distribution. We use a pre-trained diffusion model (Ho et al., 2020) to generate a single face from the CelebA-HQ (256) dataset (Karras et al., 2018), and then train g using $24{,}183^2$ unlabeled and 1,000 positive image pairs, which results in a 0.0012% label rate for positive samples.

We compute the accuracy as the fraction of $\sim$1,000 generated image pairs that are different-gender faces. We also evaluate the MMD between $\sim$3,000 differently-gendered pairs of CelebA-HQ validation faces and the generated sample, which is computed using the concatenation of CLIP features from each face (Jayasumana et al., 2024).

In Figures 2d to 2f, we find that PF PU guidance is comparable to PN guidance, and even achieves higher accuracy.

4.3.3. Uni-modal: small molecules

Given a diffusion model for generating small molecules and a small set of molecule pairs that both have a dipole moment above or below 2, the goal is to generate more pairs of small molecules with similar dipole moments. We say that such molecules have "same-parity dipole moments." Specifically, molecule pairs with same-parity dipole moments comprise the positive distribution, molecule pairs with different-parity dipole moments comprise the negative distribution, and random pairs of molecules comprise the unlabeled distribution. We use a pre-trained diffusion model (Hoogeboom et al., 2022) to generate a single small molecule from the QM9 dataset (Ramakrishnan et al., 2014), and then train g using $100{,}000^2$ unlabeled and 100,000 positive molecule pairs, which results in a 0.0026% label rate for positive samples.

We compute the accuracy as the fraction of $\sim$10,000 generated molecule pairs that have same-parity dipole moments. We also evaluate the MMD between 10,000 pairs of same-parity dipole moment QM9 validation pairs and a generated sample, which is computed using the concatenation of ChemNet features from each molecule (Preuer et al., 2018).

Figures 3a to 3b shows that prior-free PU guidance generally

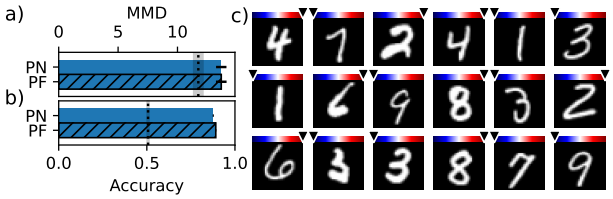

Figure 4. Jointly sampled digits and Boolean. **a)** and **b)** contain the MMD and accuracy for pairs of digits and Booleans. The vertical dotted line indicates the performance of an unconditional diffusion model, and the black bars and gray region indicate the bootstrapped standard deviation. **c)** Random sample of digit-Boolean pairs from PF PU guidance. The unit range $[0, 1]$ (a relaxation of $\{0, 1\}$) and value are visualized using the color bar and black triangle.

performs as well as PN guidance, particularly for accuracy.

4.3.4. Multi-modal: digits and Boolean values

Given a diffusion model for generating MNIST digits, a diffusion model for generating Boolean values (which essentially simulates a Bernoulli trial), and a small set of multi-modal digit-Boolean pairs, the goal is to generate more pairs of such multi-modal data. Specifically, the joint data contains samples where even (resp. odd) numbers are paired with the true (resp. false) Boolean, the not-joint has the opposite Booleans, and the unlabeled distribution is random combinations. We train a 1-dimensional diffusion model to generate Boolean values, and then train g using $59{,}000 \times 2$ unlabeled and 1,000 positive digit-Boolean pairs, which results in a 1.6% label rate for positive samples.

Unlike other plots, Figure 4a shows the marginal distribution of images. Here, we note that an unconditional model achieves a significantly lower MMD than either PN and PF PU guidance, meaning that classifier guidance distorts the marginal distribution when sampling from the joint distribution. However, the amount that the guidance distorts the marginal distribution is comparable between PN and PF PU guidance. Furthermore, both methods similarly lead to significantly improved accuracy in Figure 4.

5. Limitations and future work

One major limitation of classifier guidance is its test-time compute cost, which is due to backpropagation through the classifier. Future work may seek to mitigate this.

We draw our conclusions from semi-synthetic datasets. Although we believe that our tests are sufficient to demonstrate the effectiveness of our method, future work should investigate real-world compositional datasets. It is also an open question whether our PF PU guidance can be applied to other score- or flow-based models (*e.g.* Liu et al. (2023b)).

This study lays the foundation for future applications to multi-modal data in the life sciences, particularly when joint data is scarce. Future work may apply JDS to multi-modal datasets in nanochemistry, molecular cell biology, materials design, multi-omics, and medical imaging (Takeda et al., 2023; Kim & Park, 2024; Luo et al., 2025).

6. Related works

PU Learning Most of the PU learning literature either assumes knowledge of π (du Plessis et al., 2014; 2015; Kiryo et al., 2017; Na et al., 2024; Takahashi et al., 2025), or estimates π under distributional assumptions, such as disjoint supports of f_{p} and f_{n} (Elkan & Noto, 2008; Lee & Liu, 2003; Elkan & Noto, 2008; Ivanov, 2020) or irreducibility (Blanchard et al., 2010), which may not hold for life-sciences data. For example, the supports of f_{p} and f_{n} are likely to overlap in biological datasets. Recent works make weaker assumptions (Zhu et al., 2023; Garg et al., 2021), but estimation of π still incurs estimation error that our approach avoids.

PU diffusion models Takahashi et al. (2025) train a diffusion model on PU data and Na et al. (2024)'s method could be used for PU data under the label-noise perspective. Both approaches assume that the prior probability π is known.

Compositional diffusion Composing pre-trained diffusion models achieves significant out-of-distribution generalization (Liu et al., 2021; Du et al., 2023; Liu et al., 2023a; Geng et al., 2024a;b; Wu et al., 2024; Skreta et al., 2025). However, these approaches assume a known compositional structure, or that an expressive latent variable (*e.g.*, text) is available. Yang et al. (2024) use a small diffusion model to adjust a large one, but they do not consider JDS.

Multi-modal generative models Some multi-modal diffusion models learn a unified model for all combinations of modalities (Chen et al., 2023; Ruan et al., 2023; Luo et al., 2025), which is expensive to retrain when adding new modalities. Other works learn a joint embedding for all modalities (Takeda et al., 2023; Tang et al., 2023; Zhan et al., 2024; Wang et al., 2024b). If the joint embedding is sufficiently general, new modalities may be added without retraining, but this may not hold in practice (Liu et al., 2024)

References

Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016):493–500, June 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07487-w.

Anderson, B. D. Reverse-time diffusion equation models.

Stochastic Processes and their Applications, 12(3):313–326, 1982. ISSN 0304-4149. doi: 10.1016/0304-4149(82)90051-5.

Bachimanchi, H. and Volpe, G. Diffusion models for super-resolution microscopy: a tutorial. *Journal of Physics: Photonics*, 7(1):013001, January 2025. doi: 10.1088/2515-7647/ada101.

Bansal, A., Chu, H.-M., Schwarzschild, A., Sengupta, S., Goldblum, M., Geiping, J., and Goldstein, T. Universal guidance for diffusion models. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 843–852, 2023. doi: 10.1109/CVPRW59228.2023.00091.

Bekker, J. and Davis, J. Learning from positive and unlabeled data: a survey. *Machine Learning*, 109(4):719–760, April 2020. ISSN 1573-0565. doi: 10.1007/s10994-020-05877-5.

Blanchard, G., Lee, G., and Scott, C. Semi-supervised novelty detection. *Journal of Machine Learning Research*, 11(99):2973–3009, 2010. URL http://jmlr.org/papers/v11/blanchard10a.html.

Chen, C., Ding, H., Sisman, B., Xu, Y., Xie, O., Yao, B. Z., Tran, S. D., and Zeng, B. Diffusion models for multi-task generative modeling. In *The Twelfth International Conference on Learning Representations*, October 2023. URL https://openreview.net/forum?id=cbv0sBIZh9.

Chung, H., Kim, J., Mccann, M. T., Klasky, M. L., and Ye, J. C. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=OnD9zGAGT0k.

Darcet, T., Oquab, M., Mairal, J., and Bojanowski, P. Vision transformers need registers. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=2dnO3LLiJ1.

Du, Y., Durkan, C., Strudel, R., Tenenbaum, J. B., Dieleman, S., Fergus, R., Sohl-Dickstein, J., Doucet, A., and Grathwohl, W. S. Reduce, reuse, recycle: Compositional generation with energy-based diffusion models and MCMC. In Krause, A., Brunskill, E., Cho, K., Engelhardt, B., Sabato, S., and Scarlett, J. (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 8489–8510. PMLR, July 2023. URL https://proceedings.mlr.press/v202/du23a.html.

du Plessis, M. C., Niu, G., and Sugiyama, M. Analysis of learning from positive and unlabeled data. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., and Weinberger, K. (eds.), *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL https://proceedings.neurips.cc/paper_files/paper/2014/file/f032bc3f1eb547f716df87edb523b8f0-Paper.pdf.

du Plessis, M. D., Niu, G., and Sugiyama, M. Convex formulation for learning from positive and unlabeled data. In *Proceedings of the 32nd International Conference on Machine Learning*, pp. 1386–1394. PMLR, June 2015. URL https://proceedings.mlr.press/v37/plessis15.html.

Elkan, C. and Noto, K. Learning classifiers from only positive and unlabeled data. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '08, pp. 213–220. Association for Computing Machinery, August 2008. ISBN 978-1-60558-193-4. doi: 10.1145/1401890.1401920.

Falcon, W. et al. PyTorch Lightning, March 2019. URL https://www.pytorchlightning.ai.

Garg, S., Wu, Y., Smola, A. J., Balakrishnan, S., and Lipton, Z. Mixture proportion estimation and PU learning: A modern approach. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 8532–8544. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/47b4f1bfdf6d298682e610ad74b37dca-Paper.pdf.

Geng, D., Park, I., and Owens, A. Factorized diffusion: Perceptual illusions by noise decomposition. In *European Conference on Computer Vision*, pp. 366–384. Springer, 2024a. doi: 10.1007/978-3-031-72998-0_21.

Geng, D., Park, I., and Owens, A. Visual anagrams: Generating multi-view optical illusions with diffusion models. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24154–24163, 2024b. doi: 10.1109/CVPR52733.2024.02280.

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A. A kernel two-sample test. *Journal of Machine Learning Research*, 13(25):723–773, 2012. URL http://jmlr.org/papers/v13/gretton12a.html.

Guo, Z., Liu, J., Wang, Y., Chen, M., Wang, D., Xu, D., and Cheng, J. Diffusion models in bioinformatics and computational biology. *Nature Reviews Bioengineering*,

2(2):136–154, February 2024. ISSN 2731-6092. doi: 10.1038/s44222-023-00114-9.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H. (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf.

Hoogeboom, E., Satorras, V. G., Vignac, C., and Welling, M. Equivariant diffusion for molecule generation in 3D. In Chaudhuri, K., Jegelka, S., Song, L., Szepesvari, C., Niu, G., and Sabato, S. (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 8867–8887. PMLR, July 2022. URL https://proceedings.mlr.press/v162/hoogeboom22a.html.

Ivanov, D. DEDPUL: Difference-of-estimated-densities-based positive-unlabeled learning. In *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 782–790, 2020. doi: 10.1109/ICMLA51294.2020.00128.

Jayasumana, S., Ramalingam, S., Veit, A., Glasner, D., Chakrabarti, A., and Kumar, S. Rethinking FID: Towards a better evaluation metric for image generation. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9307–9315, 2024. doi: 10.1109/CVPR52733.2024.00889.

Karras, T., Aila, T., Laine, S., and Lehtinen, J. Progressive growing of GANs for improved quality, stability, and variation. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=Hk99zCeAb.

Kim, J. and Park, H. Adaptive latent diffusion model for 3D medical image to image translation: Multi-modal magnetic resonance imaging study. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 7604–7613, 2024. URL https://openaccess.thecvf.com/content/WACV2024/html/Kim_Adaptive_Latent_Diffusion_Model_for_3D_Medical_Image_to_Image_WACV_2024_paper.html.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. In Bengio, Y. and LeCun, Y. (eds.), *3rd International Conference on Learning Representations*, 2015. URL http://arxiv.org/abs/1412.6980.

Kiryo, R., Niu, G., du Plessis, M. C., and Sugiyama, M. Positive-unlabeled learning with non-negative risk estimator. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/7cce53cf90577442771720a370c3c723-Paper.pdf.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, November 1998. URL http://yann.lecun.com/exdb/mnist/.

Lee, W. S. and Liu, B. Learning with positive and unlabeled examples using weighted logistic regression. In *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, Washington, DC, 2003. URL https://aaai.org/papers/icml03-060-learning-with-positive-and-unlabeled-examples-using-weighted-logistic-regression/.

Li, K., Li, J., Tao, Y., and Wang, F. stDiff: A diffusion model for imputing spatial transcriptomics through single-cell transcriptomics. *Briefings in Bioinformatics*, 25(3): bbae171, May 2024. ISSN 1477-4054. doi: 10.1093/bib/bbae171.

Liu, J., Cen, X., Yi, C., Wang, F.-a., Ding, J., Cheng, J., Wu, Q., Gai, B., Zhou, Y., He, R., Gao, F., and Li, Y. Challenges in AI-driven biomedical multimodal data fusion and analysis. *Genomics, Proteomics & Bioinformatics*, pp. qzaf011, February 2025. ISSN 1672-0229. doi: 10.1093/gpbjnl/qzaf011.

Liu, N., Li, S., Du, Y., Tenenbaum, J., and Torralba, A. Learning to compose visual relations. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 23166–23178. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/c3008b2c6f5370b744850a98a95b73ad-Paper.pdf.

Liu, N., Du, Y., Li, S., Tenenbaum, J. B., and Torralba, A. Unsupervised compositional concepts discovery with text-to-image generative models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2085–2095, 2023a. URL https://openaccess.thecvf.com/content/ICCV2023/html/Liu_Unsupervised_Compositional_Concepts_Discovery_with_Text-to-Image_Generative_Models_ICCV_2023_paper.html.

Liu, X., Gong, C., and Liu, Q. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2023b. URL https://openreview.net/forum?id=XVjTT1nw5z.

Liu, Y., Wang, S., Dong, J., Chen, L., Wang, X., Wang, L., Li, F., Wang, C., Zhang, J., Wang, Y., Wei, S., Chen, Q., and Liu, H. De novo protein design with a denoising diffusion network independent of pretrained structure prediction models. *Nature Methods*, 21(11):2107–2116, November 2024. ISSN 1548-7105. doi: 10.1038/s41592-024-02437-w.

Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., and Van Gool, L. RePaint: Inpainting using denoising diffusion probabilistic models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11451–11461, 2022. doi: 10.1109/CVPR52688.2022.01117.

Luo, E., Liu, Q., Hao, M., Wei, L., and Zhang, X. Multimodal diffusion model with dual-cross-attention for multi-omics data generation and translation, March 2025.

Lvovs, D., Creason, A. L., Levine, S. S., Noble, M., Mahurkar, A., White, O., and Fertig, E. J. Balancing ethical data sharing and open science for reproducible research in biomedical data science. *Cell Reports Medicine*, 6(4):102080, April 2025. ISSN 2666-3791. doi: 10.1016/j.xcrm.2025.102080.

Mayr, A., Klambauer, G., Unterthiner, T., Steijaert, M., Wegner, J. K., Ceulemans, H., Clevert, D.-A., and Hochreiter, S. Large-scale comparison of machine learning methods for drug target prediction on ChEMBL. *Chemical Science*, 9:5441–5451, 2018. doi: 10.1039/C8SC00148K.

Na, B., Kim, Y., Bae, H., Hyun, J. L., Jung, S. K., Kang, W., and Moon, I.-C. Label-noise robust diffusion models. In *International conference on machine learning*, April 2024. URL https://openreview.net/forum?id=HXWTXXtHNl.

Nelsen, R. B. *An Introduction to Copulas*. Springer Series in Statistics. Springer, New York, NY, 2 edition, 2006. ISBN 978-0-387-28659-4. doi: 10.1007/0-387-28678-0.

Oquab, M. et al. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=a68SUt6zFt.

Pavasovic, K. L., Verbeek, J., Biroli, G., and Mezard, M. Classifier-free guidance: From high-dimensional analysis to generalized guidance forms, 2025. URL https://arxiv.org/abs/2502.07849.

Preuer, K., Renz, P., Unterthiner, T., Hochreiter, S., and Klambauer, G. Fréchet ChemNet distance: A metric for generative models for molecules in drug discovery. *Journal of Chemical Information and Modeling*, 58(9):1736–1741, September 2018. ISSN 1549-9596. doi: 10.1021/acs.jcim.8b00234.

Ramakrishnan, R., Dral, P. O., Rupp, M., and von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, 1(1):140022, August 2014. ISSN 2052-4463. doi: 10.1038/sdata.2014.22.

Ruan, L., Ma, Y., Yang, H., He, H., Liu, B., Fu, J., Yuan, N. J., Jin, Q., and Guo, B. MM-Diffusion: Learning multimodal diffusion models for joint audio and video generation. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10219–10228, 2023. doi: 10.1109/CVPR52729.2023.00985.

Satorras, V. G., Hoogeboom, E., and Welling, M. E(n) equivariant graph neural networks. In Meila, M. and Zhang, T. (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9323–9332. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/satorras21a.html.

Sklar, A. Fonctions de répartition à n dimensions et leurs marges. *Annales de l'Institut de Statistique de l'Université de Paris*, 8(3):229–231, 1959. URL https://hal.science/hal-04094463.

Skreta, M., Atanackovic, L., Bose, J., Tong, A., and Neklyudov, K. The superposition of diffusion models using the Itô density estimator. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=2o58Mbqkd2.

Song, J., Meng, C., and Ermon, S. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021a. URL https://openreview.net/forum?id=St1giarCHLP.

Song, Y. and Ermon, S. Generative modeling by estimating gradients of the data distribution. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/3001ef257407d5a371a96dcd947c7d93-Paper.pdf.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-based generative modeling through stochastic differential equations. In *International*

Conference on Learning Representations, 2021b. URL https://openreview.net/forum?id=PxTIG12RRHS.

Takahashi, H., Iwata, T., Kumagai, A., Yamanaka, Y., and Yamashita, T. Positive-unlabeled diffusion models for preventing sensitive data generation. In *Proceedings of the 13th International Conference on Learning Representations*, April 2025. URL https://openreview.net/forum?id=jKcZ4hF4s5.

Takeda, S., Priyadarsini, I., Kishimoto, A., Shinohara, H., Hamada, L., Masataka, H., Fuchiwaki, J., and Nakano, D. Multi-modal foundation model for material design. In *AI for Accelerated Materials Design - NeurIPS 2023 Workshop*, November 2023. URL https://openreview.net/forum?id=EiT2bLsfM9.

Tang, Z., Yang, Z., Zhu, C., Zeng, M., and Bansal, M. Any-to-any generation via composable diffusion. In Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., and Levine, S. (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 16083–16099. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/33edf072fe44f19079d66713a1831550-Paper-Conference.pdf.

Wang, L., Song, C., Liu, Z., Rong, Y., Liu, Q., Wu, S., and Wang, L. Diffusion models for molecules: A survey of methods and tasks, 2025. URL https://arxiv.org/abs/2502.09511.

Wang, X., Zhu, H., Terashi, G., Taluja, M., and Kihara, D. DiffModeler: large macromolecular structure modeling for cryo-EM maps using a diffusion model. *Nature Methods*, 21(12):2307–2317, December 2024a. ISSN 1548-7105. doi: 10.1038/s41592-024-02479-0.

Wang, Y., Yu, J., and Zhang, J. Zero-shot image restoration using denoising diffusion null-space model. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=mRieQgMtNTQ.

Wang, Y., Liu, X., Huang, F., Xiong, Z., and Zhang, W. A multi-modal contrastive diffusion model for therapeutic peptide generation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(1):3–11, March 2024b. ISSN 2374-3468. doi: 10.1609/aaai.v38i1.27749.

Wu, T., Maruyama, T., Wei, L., Zhang, T., Du, Y., Iaccarino, G., and Leskovec, J. Compositional generative inverse design. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=wmX0CqFSd7.

Xu, M., Powers, A. S., Dror, R. O., Ermon, S., and Leskovec, J. Geometric latent diffusion models for 3D molecule generation. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 38592–38610. PMLR, July 2023. URL https://proceedings.mlr.press/v202/xu23n.html.

Yang, S., Du, Y., Dai, B., Schuurmans, D., Tenenbaum, J. B., and Abbeel, P. Probabilistic adaptation of black-box text-to-video models. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=pjtIEgscE3.

Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R. R., and Smola, A. J. Deep sets. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/f22e4747da1aa27e363d86d40ff442fe-Paper.pdf.

Zhan, C., Lin, Y., Wang, G., Wang, H., and Wu, J. MedM2G: Unifying medical multi-modal generation via cross-guided diffusion with visual invariant. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11502–11512, 2024. doi: 10.1109/CVPR52733.2024.01093.

Zhu, Y., Fjeldsted, A., Holland, D., Landon, G., Lintereur, A., and Scott, C. Mixture proportion estimation beyond irreducibility. In Krause, A., Brunskill, E., Cho, K., Engelhardt, B., Sabato, S., and Scarlett, J. (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 42962–42982. PMLR, July 2023. URL https://proceedings.mlr.press/v202/zhu23c.html.

A. Balanced empirical risk

Here, we provide a definition of the balanced empirical risk from Section 2.4.

$$\frac{1}{m}\sum_{i=1}^{m}\ell_{\mathrm{CE}}(g(\boldsymbol{x}_{\mathrm{p}}^{i};t),+1)+\frac{1}{k}\sum_{i=1}^{k}\ell_{\mathrm{CE}}(g(\boldsymbol{x}_{u}^{i};t),-1)\ , \tag{12}$$

where ℓ_{CE} is the cross-entropy/logistic loss. Intuitively, this is just the empirical risk where each class (positive and unlabeled) is weighted as if the class priors were equal.

B. An alternate framing of PU learning

Our approach to joint diffusion sampling features an application of PU learning in which $f_{\mathrm{u}}(\boldsymbol{x};t)\coloneqq f_1(\boldsymbol{x}_1;t)f_2(\boldsymbol{x}_2;t)$, and $f_{\mathrm{p}}(\boldsymbol{x};t)\coloneqq f(\boldsymbol{x};t)$, where $f(\boldsymbol{x};0)$ is the joint distribution to be sampled from (Section 1.1). The following theorem characterizes when the mixture in (10) is well-posed, where U, P, and Q are measures corresponding to the densities f_{u}, f_{p}, and f_{n}.

Theorem B.1. *Let $(X,\mathcal{F})$ be a measurable space, let U be a probability measure on $(X,\mathcal{F})$ and let P be a probability measure on $(X,\mathcal{F})$. Then there exists some $\pi\in(0,1)$ and a probability measure Q on $(X,\mathcal{F})$ such that*

$$U=\pi P+(1-\pi)Q$$

if and only if:

1. *$P\ll U$ (i.e., P is absolutely continuous with respect to U),*
2. *The Radon–Nikodym derivative $f=\frac{dP}{dU}$ is essentially bounded, i.e., if $L\coloneqq\|f\|_\infty$, then one can choose π such that $\pi\le\frac{1}{L}$.*

Proof. **(Necessity).** Assume that there exist $\pi\in(0,1)$ and a probability measure Q such that $U=\pi P+(1-\pi)Q$. For any measurable set $A\in\mathcal{F}$, we have

$$U(A)=\pi P(A)+(1-\pi)Q(A)\ge\pi P(A).$$

Thus,

$$P(A)\le\frac{1}{\pi}U(A)\quad\text{for all } A\in\mathcal{F}.$$

So $U(A)=0\implies P(A)=0$, *i.e.*,, $P\ll U$. By the Radon–Nikodym theorem, there exists a U-measurable function $f=\frac{dP}{dU}$ such that

$$P(A)=\int_A f\,dU\quad\text{for all } A\in\mathcal{F}.$$

Now, suppose by way of contradiction that there exists a measurable set E with $U(E)>0$ on which

$$f(x)>\frac{1}{\pi}\quad\text{for all } x\in E.$$

Then,

$$P(E)=\int_E f\,dU>\frac{1}{\pi}U(E),$$

which contradicts the earlier inequality $P(E)\le\frac{1}{\pi}U(E)$. Therefore, it must be that

$$f(x)\le\frac{1}{\pi}\quad\text{for } U\text{-almost every } x\in X.$$

Defining $L := \|f\|_\infty$, we deduce that

$$\pi \leq \frac{1}{L}.$$

(Sufficiency). Conversely, assume that $P \ll U$ and that the Radon–Nikodym derivative $f = \frac{dP}{dU}$ is essentially bounded with $L = \|f\|_\infty < \infty$. Choose any $\pi \in (0, 1)$ satisfying

$$\pi \leq \frac{1}{L}.$$

Then, for U-almost every $x \in X$,

$$f(x) \leq L \leq \frac{1}{\pi},$$

so that

$$1 - \pi f(x) \geq 0 \quad \text{for } U\text{-almost every } x.$$

Define Q on $\mathcal{F}$ by

$$Q(A) := \frac{1}{1-\pi}\Big(U(A) - \pi P(A)\Big) \quad \text{for all } A \in \mathcal{F}.$$

Since $1 - \pi f(x) \geq 0$ U-a.e., it follows that for every $A \in \mathcal{F}$,

$$U(A) - \pi P(A) \geq 0,$$

so $Q(A) \geq 0$. Moreover,

$$Q(X) = \frac{1}{1-\pi}\Big(U(X) - \pi P(X)\Big) = \frac{1}{1-\pi}\Big(1 - \pi\Big) = 1,$$

so that Q is a probability measure.

□

By Theorem B.1, there exists π and $f_{\mathrm{n}}(\cdot; t)$ such that

$$f_{\mathrm{u}}(\boldsymbol{x}; t) \coloneqq f_1(\boldsymbol{x}_1; t) f_2(\boldsymbol{x}_2; t) = \pi f_{\mathrm{p}}(\boldsymbol{x}; t) + (1 - \pi) f_{\mathrm{n}}(\boldsymbol{x}; t)\ , \tag{13}$$

so that Equation (10) is justified.

C. Experimental details

Here we discuss additional experimental details that did not fit in the main text.

C.1. Diffusion model architecture

We use two types of diffusion models, Denoising Diffusion Probabilistic Models (DDPM) (Ho et al., 2020) and Geometric Latent Diffusion Models (Geo-LDM) (Xu et al., 2023).

DDPMs define a diffusion process over the space of random images. We train our own DDPM for sampling MNIST digits, and use two pre-trained DDPMs for the CelebA-HQ dataset, one trained using exponential moving average (EMA) and one without (Ho et al., 2020). As in Takahashi et al. (2025), we use 50 Denoising Diffusion Implicit Model (DDIM) steps (Song et al., 2021a), with noise, for image sampling.

We treat Booleans as two points ($\{-1, 1\}$) in the real numbers and perform diffusion on the reals using DDPMs as we do for images.

Geo-LDM defines a diffusion process over an E(3)-equivariant latent molecule structure with a atoms. Let $\mathbf{M} \in \mathbb{R}^{(3+d+c)\times a}$ be a random vector that represents each of a atoms using its (x, y, z) position, a d-dimensional one-hot-vector indicating the chemical element, and a c-dimensional electronic charge. GeoLDM uses an E(3)-equivariant autoencoder $\mathcal{E}_D(\mathcal{E}_E(\cdot)) : \mathbb{R}^{(3+d+c)\times a} \to \mathbb{R}^{(3+2)\times a} \to \mathbb{R}^{(3+d+c)\times a}$ to convert the mixed random variable $\mathbf{M}$ to a continuous random variable $\mathcal{E}_E(\mathbf{M})$. E(3)-equivariant diffusion is implemented using an E(3)-equivariant score network (Hoogeboom et al., 2022), and we use the pre-trained models without modifications (Xu et al., 2023).

C.2. Classifier guidance

Now, we discuss the classifier model architectures used for each image dataset. For MNIST, we train a small convolutional network, and for CelebA-HQ we add a two-layer head to the DinoV2 feature extractor (small, with registers) (Oquab et al., 2024; Darcet et al., 2024). For pairs of images we implement a Siamese neural network with a two-layer Deep Sets (Zaheer et al., 2017) model with mean aggregation to combine the extracted features. For CelebaA-HQ, we extracted all the features prior to training to increase efficiency. SiLU was used for all activation functions to ensure non-constant activations during guidance. We trained the BPU posterior model using a high learning rate and weight decay to prevent overfitting, with 1–10 epochs for the MNIST dataset and 1–100 epochs for the CelebA-HQ dataset depending on whether individual or pairs of images were being classified. The PN classifier was trained using the cross-entropy loss until convergence using a learning rate of 0.001. Optimization was implemented using PyTorch Lightning (Falcon et al., 2019) with the Adam optimizer (Kingma & Ba, 2015), and hyperparameters were manually optimized to minimize the validation loss.

The noiseless data estimate did not work with GeoLDM, so we instead fine-tuned a pretrained model. We modify the E(3)-equivariant networks from Satorras et al. (2021) to operate on the latent space of $\mathcal{E}_E$, and pre-train a noise-aware model. We train this model to predict all 12 QM9 properties and the counts of 5 atomic elements (C, H, O, N, F) using multi-output regression with the squared error and Poisson regression. We use all but the last layer to extract features for each molecule and pass the concatenated outputs to a two-layer multi-layer perceptron (MLP) with SiLU activation functions. We fine-tune the entire model for PU and PN classification on noisy data to learn $\tilde{g}(\cdot; t)$. Our PU learning used a mixed learning rate of 0.001 and 0.0001 for the MLP and extractor to prevent overfitting

For MNIST-Boolean pairs, we train a noise-aware classifier using the pretrained diffusion models as feature extractors. Specifically, we use the downsampling feature extractor from the score U-Net and we treat all but the last layer of the 1-dimensional diffusion model as a feature extractor. During model inference, we concatenate the extracted features and pass them to a two-layer MLP with dropout. We freeze the weights on the feature extractors, and train the two-layer MLP for 10 epochs. During inference, we use the rescaled guidance of "power-law CFG" from Pavasovic et al. (2025) but with our classifier score. The PN and PU classifiers were trained using the same hyperparameters.

All classifier guidance models were trained on an NVIDIA RTX 3080, with training times ranging from 1–60 minutes long. For comparison, each PUDM trained for approximately six hours on an NVIDIA A40 GPU. Inference was performed on both GPUs.

C.3. Training the BPU posterior estimate

Our BPU posterior $\widehat{g}$ was trained using the objective in (12), implement by bootstrap resampling the positive data for each mini-batch of the unlabeled data. Hyperparameters were manually tuned to minimize the balanced positive-unlabeled cross-entropy and zero-one loss of the PU classifier. When training a PU classifier, we randomly select n positive data points (*e.g.*, images) from the training dataset and treat all remaining data points as unlabeled.

C.4. Evaluation

For each pair of images or molecules, we extract the features for each member independently, then concatenate them to form a feature for the pair. Because our tasks are permutation invariant, we augment the extracted pairs of features with both permutations before computing the MMD. We extract the image features as above, and we compute the ChemNet embeddings for our generated molecules (Mayr et al., 2018).

Because it is not clear how to compute the MMD for multi-modal data, we only compute the MMD of the handwritten digits. The prior probability of the Boolean `is_even` being true is 0.5, and our unconditional binary diffusion model produces a prior probability of 0.48. In comparison, PN classifier guidance results in a prior probability of 0.46, while PF PU guidance results in a prior probability of 0.48, which preserves the prior probability from the marginal model.

D. Evaluating prior-free PU guidance for individual diffusion models

Previously, we have compared prior-free (PF) PU guidance against a fully-supervised positive-negative (PN) classifier for solving the joined diffusion sampling (JDS) problem. Here, we compare PF and PN guidance against a PU supervised diffusion model (Takahashi et al., 2025) in the single-foundation model setting. This approach either fine-tunes existing diffusion models or trains a model from scratch, and it is not straightforward to use it to solve the JDS problem.

D.1. PUDM implementation

The positive-unlabeled supervised diffusion model (Takahashi et al., 2025) attempts to minimize reconstruction error for the positive class and maximize the reconstruction error for the negative class using a variant of the nnPU loss (Kiryo et al., 2017) and an estimate for π.

The original positive-unlabeled diffusion model (PUDM) was designed to sample from negative classes and not sample from positive classes. We addressed this by just flipping the sign of the class labels that we passed to the existing implementation. For clarity, we reword the PUDM approach here.

Instead of directly minimizing the ℓ_2 norm of the reconstruction error, PUDM also tries to maximize the norm of the reconstruction error for the positive class. Define $\mathbf{X}(t)|\boldsymbol{x}$ as shorthand for $\mathbf{X}(t)|\mathbf{X}(0) = \boldsymbol{x}$, and let N and M denote the number of positive and unlabeled samples, respectively. In the positive-negative setting, we can use a supervised diffusion model to perform the optimization

$$\widehat{s} = \arg\min_{\zeta} \frac{1}{N} \sum_{n=1}^{N} \underbrace{\mathbb{E}_{\mathbf{X}(t)|\boldsymbol{x}^n} \left[\|\zeta(\mathbf{X}(t); t) - \nabla_{\mathbf{X}(t)} \log f(\mathbf{X}(t)|\boldsymbol{x}^n; t)\|_2^2 \right]}_{\ell(\boldsymbol{x}^n, \zeta; +1)} \tag{14}$$

$$+ \frac{1}{M} \sum_{m=1}^{M} \underbrace{\mathbb{E}_{\mathbf{X}(t)|\boldsymbol{x}^m} \left\{ \log(1 - \exp\left[\|\zeta(\mathbf{X}(t); t) - \nabla_{\mathbf{X}(t)} \log f(\mathbf{X}(t)|\boldsymbol{x}^m; t)\|_2^2 \right] \right\}}_{\ell(\boldsymbol{x}^m, \zeta; -1)} , \tag{15}$$

which can be interpreted as empirical risk minimization with the loss ℓ. This approach learns to sample according to the positive density f_{p} and not the negative density f_{n}.

To perform PU diffusion, (Takahashi et al., 2025) suggest minimizing the non-negative risk estimator from Kiryo et al. (2017), which can be written as

$$\mathcal{L}_{\mathrm{p}}(s; y) \coloneqq \frac{1}{N} \sum_{n=1}^{N} \ell(\boldsymbol{x}^n, s; y) \tag{16}$$

$$\mathcal{L}_{\mathrm{u}}(s; y) \coloneqq \frac{1}{M} \sum_{m=1}^{M} \ell(\boldsymbol{x}^m, s; y) \tag{17}$$

$$\mathcal{L}_{\mathrm{nnPU}}(s) \coloneqq \pi \mathcal{L}_{\mathrm{p}}(s; +1) + \max\left\{0, \mathcal{L}_{\mathrm{u}}(s; -1) - \pi \mathcal{L}_{\mathrm{p}}(s; -1)\right\} . \tag{18}$$

This approach removes the fraction of the risk that is contributed by the unlabeled positive data, while preventing the risk from becoming negative. Note that this approach requires π to be known *a priori*, and (Takahashi et al., 2025) state that "analyzing the optimal value of [π] in the proposed method is our important future work."

The original PUDM implementation (Takahashi et al., 2025) did not use noise in their DDIM (Song et al., 2021a) sampling procedure. However, we find that PUDM performs better when non-deterministic sampling is performed, so we use this version to ensure rigorous comparison.

In our PU guidance experiments, the true $\pi \approx 0.5$, but we evaluate PUDM models trained with several different values of π.

D.2. Results

We find that PUDM, PN, and PF PU guidance achieve similar performance for MNIST data (Appendix Figure 5 **a)–b**), but that PUDM outperforms both classifier guidance approaches for CelebA-HQ (Appendix Figure 6 **a)–b**). However, we find that this performance gap can be significantly decreased by instead using a DDPM model trained using exponential moving average (EMA). We believe that this performance gap is, in part, due to a poor estimate of the noiseless data (see Appendix C.2).

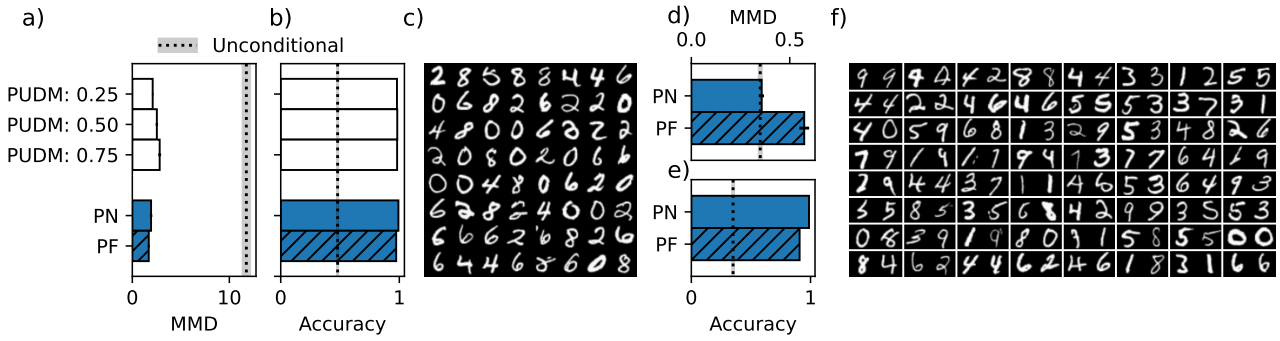

Figure 5. Single and paired samples for MNIST. **a)** MMD for even numbers, "PF" means "prior-free." **b)** The fraction of generated images that are even numbers. **c)** Random samples using PU guidance. **d)** MMD for pairs of even or odd numbers. **e)** The fraction of pairs that are both even or odd. **f)** Random samples using PU guidance. The vertical dotted line indicates the performance of an unconditional diffusion model, and the black bars and gray region indicate the bootstrapped standard deviation.

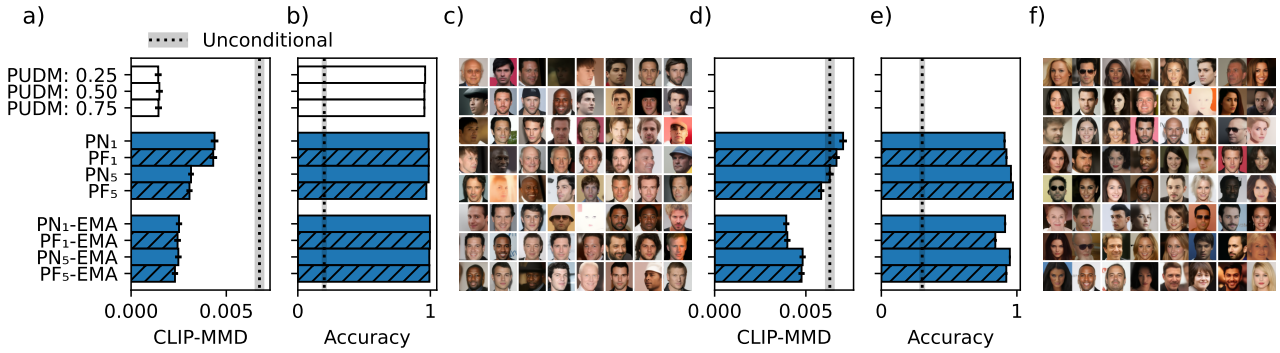

Figure 6. Single and paired samples for CelebA-HQ. **a)** CLIP-MMD for male faces, "PF" means "prior-free." The subscript indicates the number of recurrence steps. **b)** The fraction of generated images that are male faces. **c)** Random samples using prior-free PU guidance. **e)** CLIP-MMD for different-gendered faces. **f)** The fraction of generated faces that are different-gendered. **g)** Random samples using prior-free PU guidance. The vertical dotted line indicates the performance of an unconditional diffusion model, and the black bars and gray region indicate the bootstrapped standard deviation.

E. Recurrence steps

Previous work shows that repeating diffusion steps, called 'recurrence steps," leads to better diffusion guidance (Bansal et al., 2023; Chung et al., 2023; Lugmayr et al., 2022; Wang et al., 2023). The figures in the main text use 5 recurrence steps for CelebA-HQ and QM9, but here we demonstrate the results of using both 1 and 5 recurrence steps.

We find that the number of recurrence steps significantly improves the performance of the non-EMA diffusion model on CelebA-HQ (Appendix Figure 6), but the improvement is less pronounced for the EMA diffusion model.

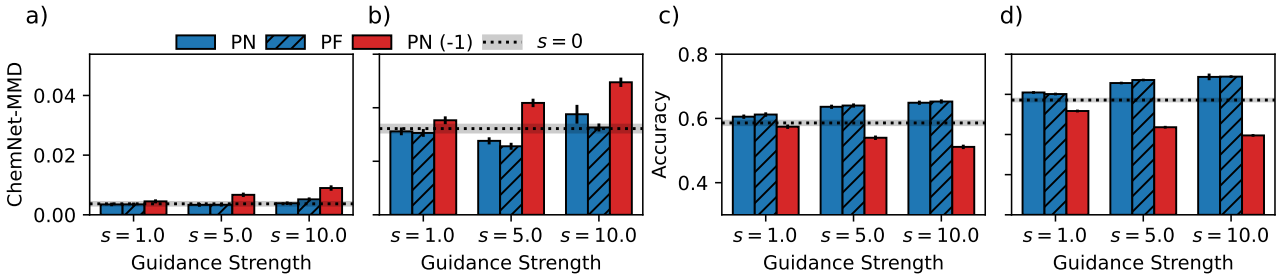

Figure 7. Molecule pairs for QM9. **a)**,**b)** ChemNet-MMD for pairs of similar-dipole-moment molecules. **c)**,**d)** Fraction of molecules that are predicted to belong to the joint distribution. "PN (-1)" indicates guidance towards the negative class. Guidance towards the positive class is blue, and guidance towards the negative class is red. The horizontal dotted line indicates the performance of an unconditional diffusion model, and the black bars and gray region indicate the bootstrapped standard deviation.

As expected, molecular stability and validity are higher under 5 recurrence steps than 1 (Appendix H). We note that the unconditional fraction of similar-dipole-molecules shifts when the number of recurrence steps is increased. To demonstrate that our PN/PU guidance was not simply reversing the benefit of our recurrence steps, we also demonstrate PN guidance towards the negative class using $1 - \widehat{\eta}(\boldsymbol{x}(t); t)$, which has similar molecular metrics (Appendix H) while receiving a worse ChemNet-MMD and accuracy. Because (5) may lead to probabilities greater than 1 and π is assumed to be unknown, we do not perform such negative guidance with PU learning.

F. Runtime comparisons

In Table 1, we provide extensive runtime comparisons of each method that we benchmarked. Because GeoLDM does not provide the model training time, we use the estimate from Hoogeboom et al. (2022) which uses a similar diffusion process on the same dataset.

Table 1. Runtime comparisons of each model. All image models use 50 inference steps with one recurrence step. "CelebA" is the CelebA-HQ (256) dataset, "PN" indicates positive-negative guidance, "PF" indicates prior-free positive-unlabeled guidance, and "Extractor" indicates the pretrained feature extractor for the QM9 dataset. The "Featurize" column indicates the amount of time used to pre-process datasets that utilize pre-trained feature extractors or autoencoders. Starred values are provided by other papers.

Method	Dataset	Featurize	Training	Inference	Batch size	Hardware
DDPM	MNIST	-	003:31:03	00:04	128	A40
DDPM	CelebA	-	063:00:00*	00:04	2	TPU v3-8
DDPM	$\{0, 1\}$	-	000:00:19	00:00	1,000	RTX 3080
GeoLDM	QM9	-	305:00:00*	01:22	64	GTX 1080 Ti
PUDM	MNIST even	-	005:52:26	00:04	128	A40
PUDM (FT)	CelebA	-	006:34:08	00:04	2	A40
PN	MNIST even	-	000:01:49	00:11	128	RTX 3080
PF	MNIST even	-	000:00:18	00:11	128	RTX 3080
PN	CelebA male	00:27	000:00:13	00:11	2	RTX 3080
PF	CelebA male	00:27	000:00:01	00:11	2	RTX 3080
PN	MNIST SP	-	000:01:48	00:11	64	RTX 3080
PF	MNIST SP	-	000:02:42	00:11	64	RTX 3080
PN	CelebA DG	00:27	000:01:07	00:11	1	RTX 3080
PF	CelebA DG	00:27	000:00:06	00:11	1	RTX 3080
PN	CelebA SG	00:27	000:00:20	00:11	1	RTX 3080
PF	CelebA SG	00:27	000:00:02	00:11	1	RTX 3080
Extractor	QM9	00:36	009:23:48	-	-	A40
PN	QM9 dipole	00:36	000:57:19	02:09	64	RTX 3080
PF	QM9 dipole	00:36	001:12:40	02:09	64	RTX 3080
PN	MNIST + $\{0, 1\}$	-	000:02:57	00:09	32	RTX 3080
PF	MNIST + $\{0, 1\}$	-	000:02:57	00:09	32	RTX 3080

G. Samples from each model

Here, we provide unfiltered samples from each method to provide a qualitative comparison.

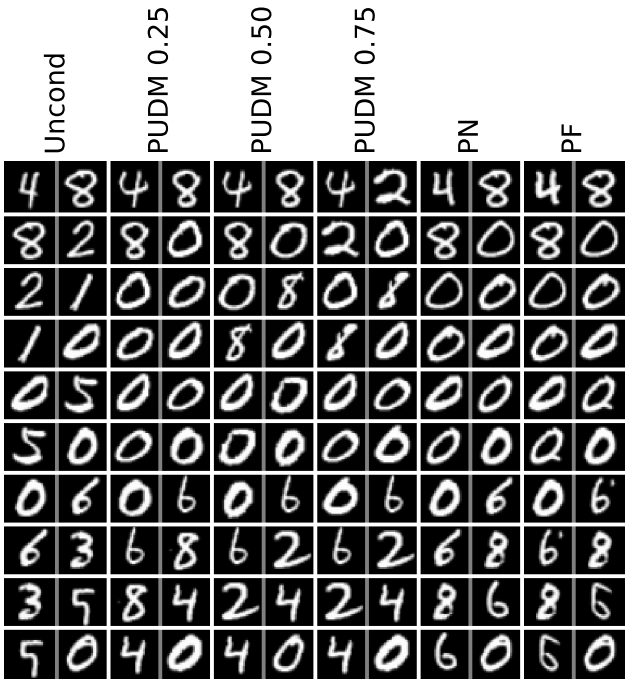

Figure 8. **MNIST even:** Non-cherry-picked samples from each generation method. Each row contains pairs of samples, and each column holds samples from a given method. For MNIST even, the pairs are independent of each other, but are included to facilitate comparison with the pairwise generation process. "Uncond" means "unconditional generation."

Figure 9. **CelebA-HQ (256) male (1 recurrence step):** Non-cherry-picked samples from each generation method using 1 recurrence step. Each row contains pairs of samples, and each column holds samples from a given method. For CelebA-HQ (256) male, the pairs are independent of each other, but are included to facilitate comparison with the pairwise generation process. "Uncond" means "unconditional generation."

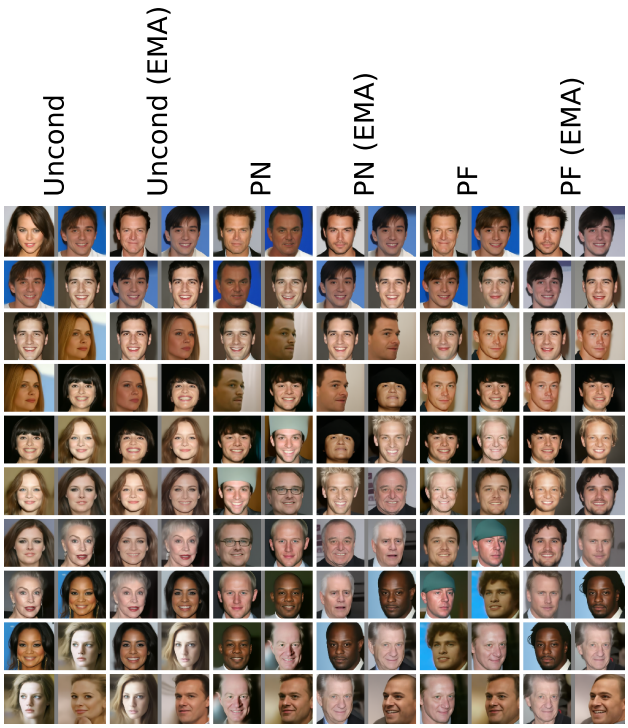

Figure 10. **CelebA-HQ (256) male (5 recurrence steps):** Non-cherry-picked samples from each generation method using 5 recurrence steps. Each row contains pairs of samples, and each column holds samples from a given method. For CelebA-HQ (256) male, the pairs are independent of each other, but are included to facilitate comparison with the pairwise generation process. "Uncond" means "unconditional generation."

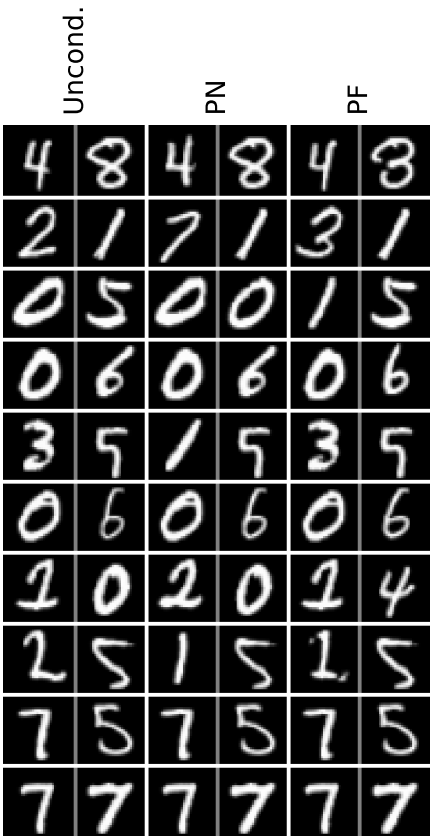

Figure 11. **MNIST same-parity:** Non-cherry-picked samples from each generation method. Each row contains pairs of samples, and each column holds samples from a given method. "Uncond" means "unconditional generation."

Figure 12. **CelebA-HQ (256) same gender (1 recurrence step):** Non-cherry-picked samples from each generation method using 1 recurrence step. Each row contains pairs of samples, and each column holds samples from a given method. "Uncond" means "unconditional generation."

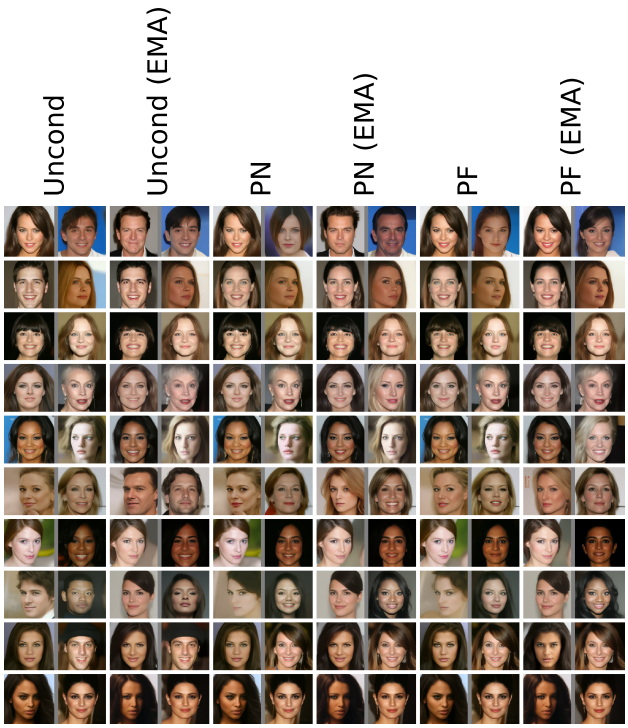

Figure 13. **CelebA-HQ (256) same gender (5 recurrence steps):** Non-cherry-picked samples from each generation method using 5 recurrence steps. Each row contains pairs of samples, and each column holds samples from a given method. "Uncond" means "unconditional generation."

Figure 14. **CelebA-HQ (256) different gender (1 recurrence step):** Non-cherry-picked samples from each generation method using 1 recurrence step. Each row contains pairs of samples, and each column holds samples from a given method. "Uncond" means "unconditional generation."

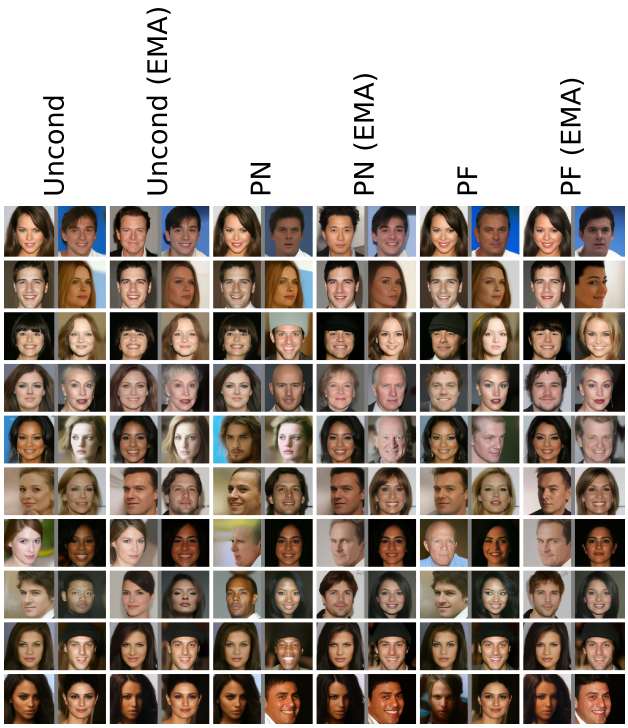

Figure 15. **CelebA-HQ (256) different gender (5 recurrence steps):** Non-cherry-picked samples from each generation method using 5 recurrence steps. Each row contains pairs of samples, and each column holds samples from a given method. "Uncond" means "unconditional generation."

G.1. QM9 pair

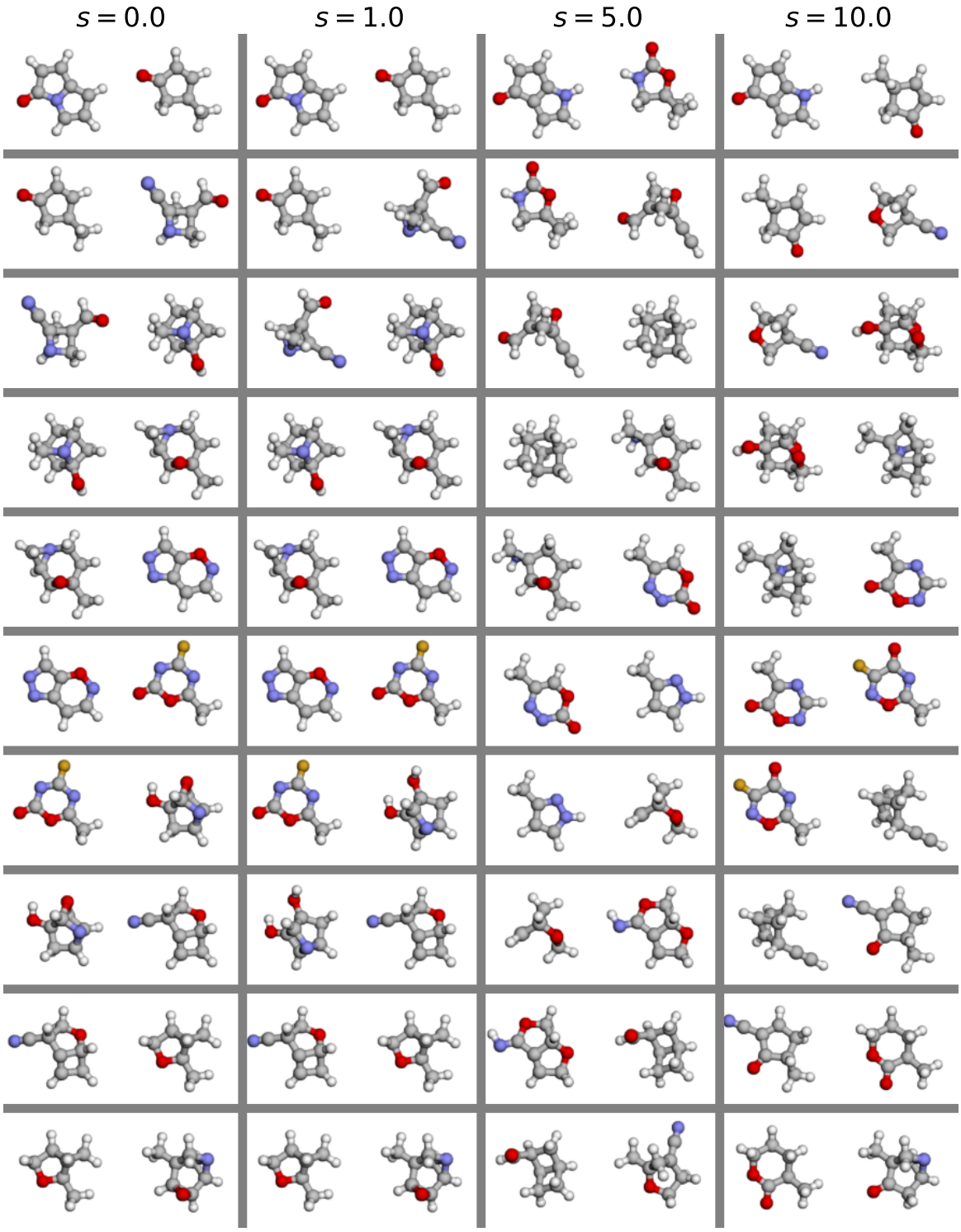

Figure 16. Non-cherry-picked samples from different guidance strengths (s) using 1 recurrence step. Each row contains pairs of samples, and each column holds samples from a given method. $s = 0$ indicates unconditional generation (*i.e.*,, no guidance).

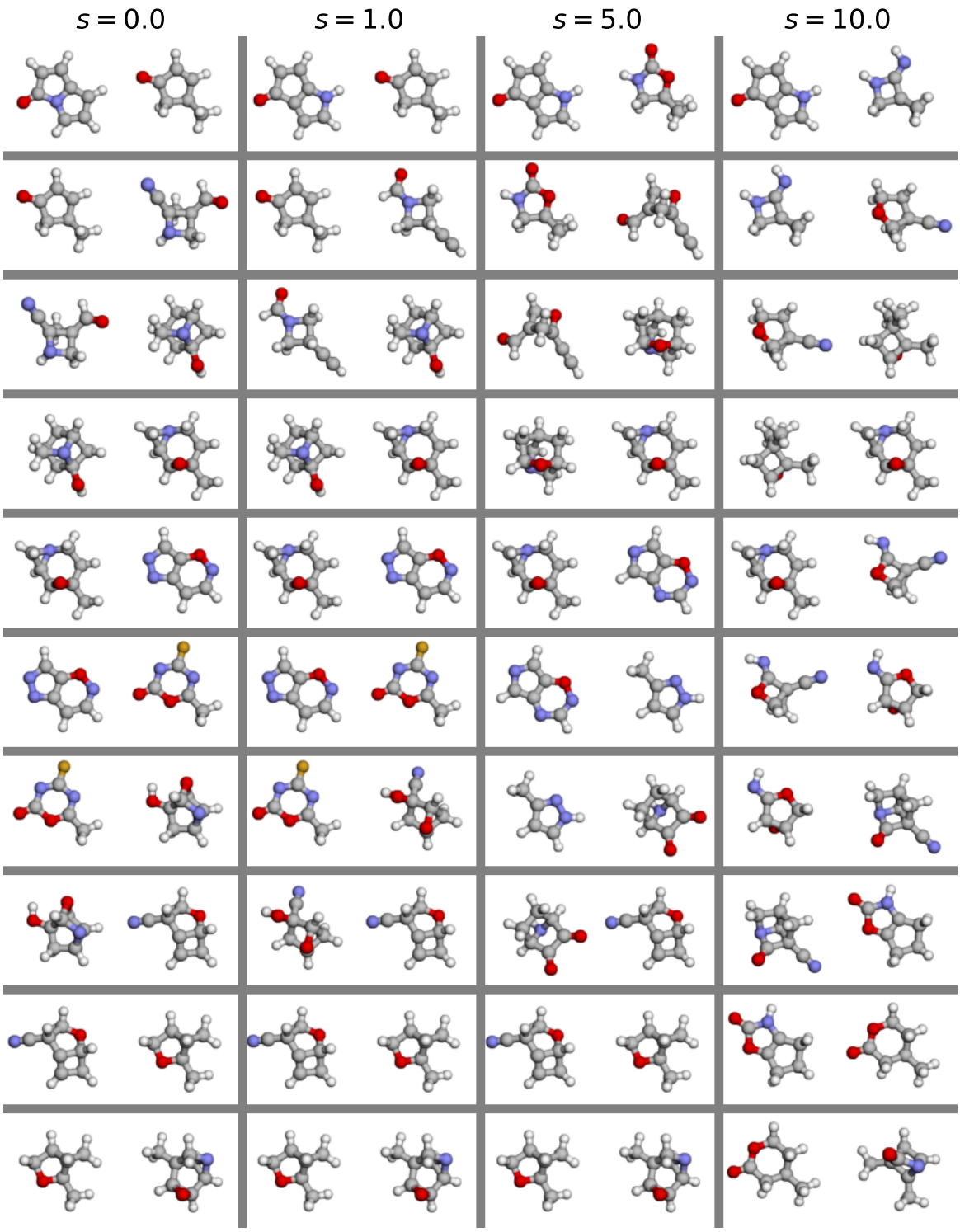

Figure 17. Non-cherry-picked samples from different guidance strengths (s) using 5 recurrence steps. Each row contains pairs of samples, and each column holds samples from a given method. $s = 0$ indicates unconditional generation (*i.e.*,, no guidance).

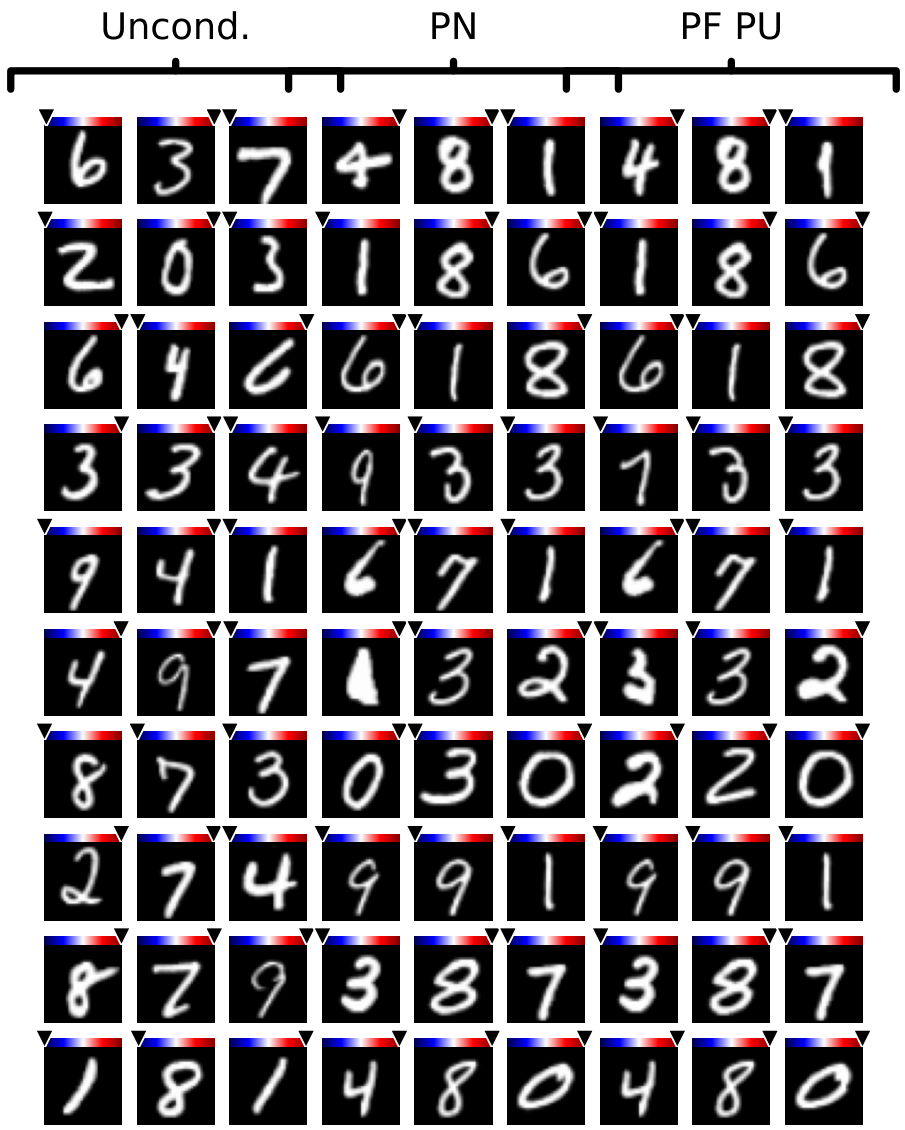

Figure 18. **Digits and Booleans (5 recurrence steps:** Non-cherry-picked samples from each generation method using 5 recurrence steps. The unit range $[0, 1]$ (a relaxation of a Boolean value) and value are visualized using the color bar and black triangle.

H. QM9 metrics

Here we provide common molecular metrics computed on 10,000 samples from our generated QM9 molecules. See the Appendix, "Limitations of RDKit-based metrics" for an analysis of these RDKit metrics.

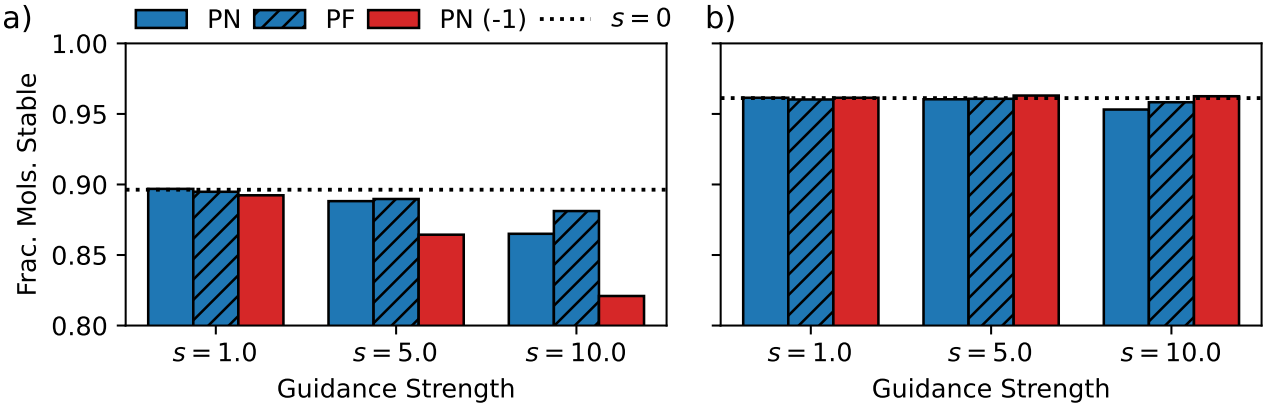

Figure 19. Molecular stability on the QM9 dataset. **a)** uses 1 recurrence step and **b)** uses 5 recurrence steps. "PN" is positive-negative classifier guidance, "PF" is prior-free PU guidance, "PN (1)" is PN guidance towards the negative class, and $s = 0$ indicates the result when using no guidance.

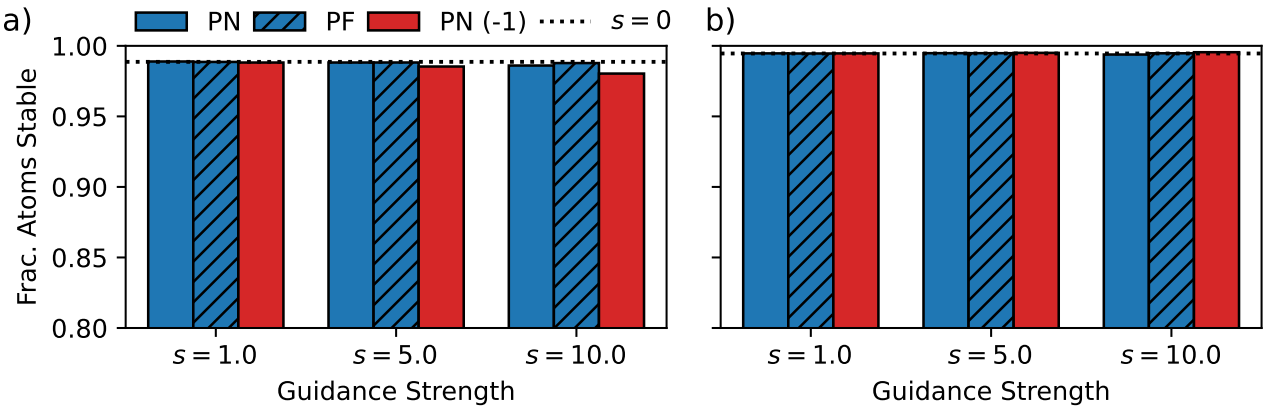

Figure 20. Atom/bond stability on the QM9 dataset. **a)** uses 1 recurrence step and **b)** uses 5 recurrence steps. "PN" is positive-negative classifier guidance, "PF" is prior-free PU guidance, "PN (1)" is PN guidance towards the negative class, and $s = 0$ indicates the result when using no guidance.

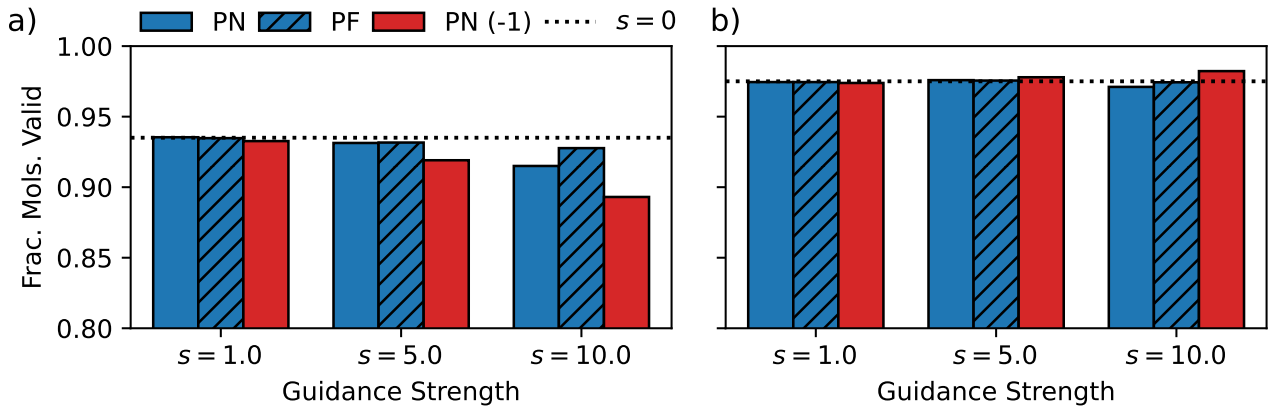

Figure 21. Validity of generated molecules. **a)** uses 1 recurrence step and **b)** uses 5 recurrence steps. "PN" is positive-negative classifier guidance, "PF" is prior-free PU guidance, "PN (1)" is PN guidance towards the negative class, and $s = 0$ indicates the result when using no guidance.

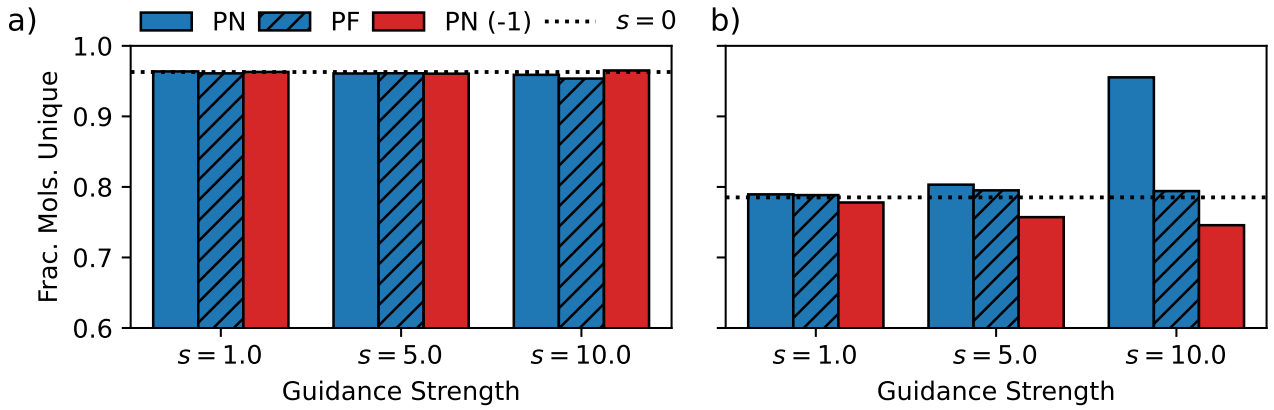

Figure 22. Uniqueness of generated molecules. **a)** uses 1 recurrence step and **b)** uses 5 recurrence steps. "PN" is positive-negative classifier guidance, "PF" is prior-free PU guidance, "PN (-1)" is PN guidance towards the negative class, and $s = 0$ indicates the result when using no guidance.

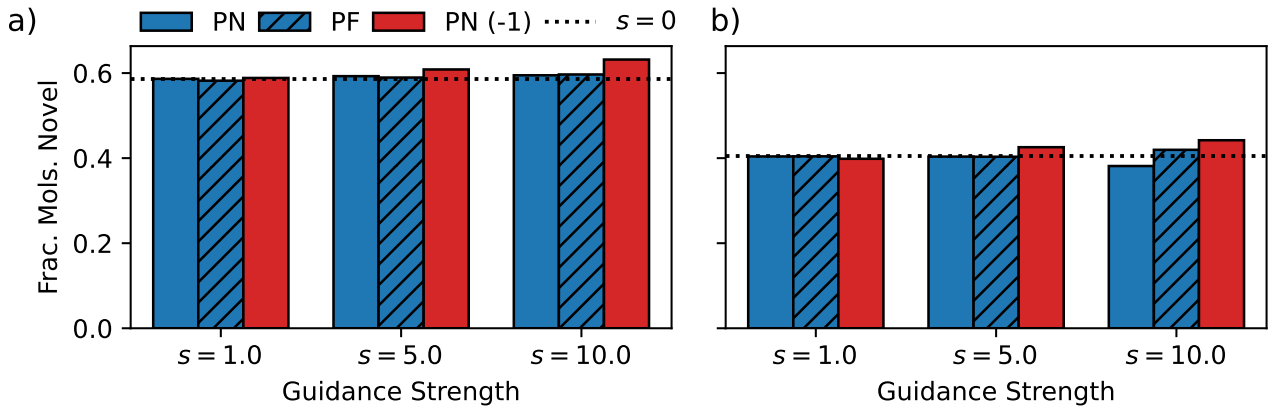

Figure 23. Novelty of generated molecules. **a)** uses 1 recurrence step and **b)** uses 5 recurrence steps. "PN" is positive-negative classifier guidance, "PF" is prior-free PU guidance, "PN (-1)" is PN guidance towards the negative class, and $s = 0$ indicates the result when using no guidance.

I. CelebA-HQ (256) mode collapse

We find that both PN and PU models may suffer from mode collapse, as evidenced by the model's tendency to generate all women. See Appendix G for more examples.

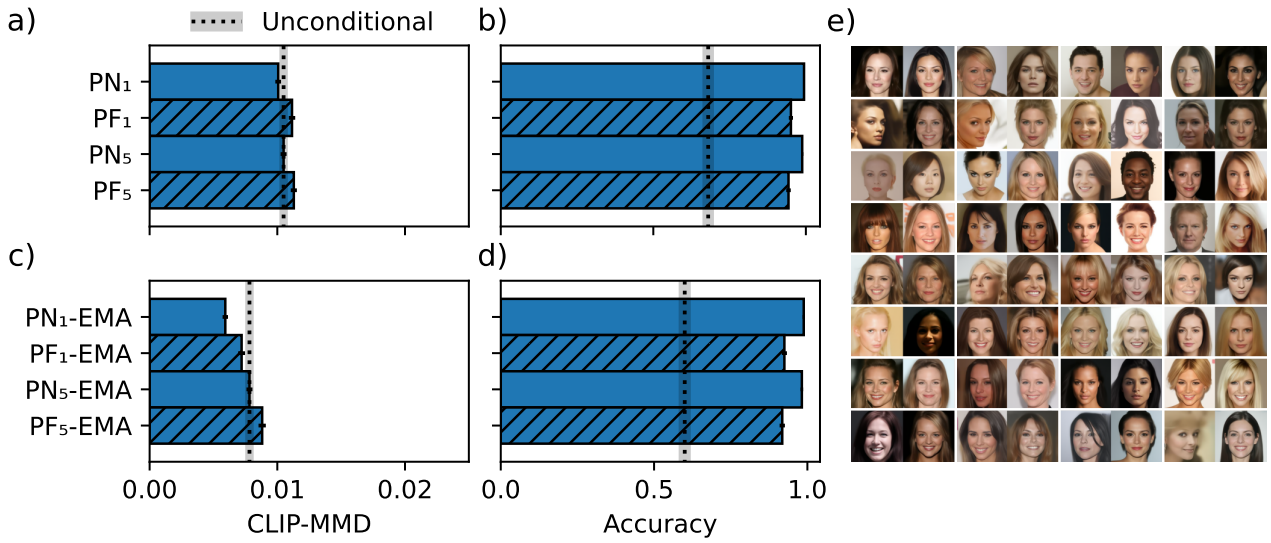

Figure 24. **CelebA-HQ (256): a)** CLIP-MMD for different-gendered faces ($\sigma = 100$). **b)** The fraction of generated faces that are same-gendered. **c)** Random samples using PU guidance with EMA and 5 recurrence steps. All error bars are the standard deviation over 10,000 bootstrapped estimates.