Your Diffusion Model is Secretly a Zero-Shot Classifier

Dmitry Bylinkin

MIPT

2024

Introduction

- Diffusion models are state-of-the-art for sampling from unknown distribution.
- Proposed paper demonstrates their ability to perform zero-shot classification.
- Key approach: leveraging density estimates from diffusion models.

Methodology

- Diffusion models use a noising-denoising process:

$$p_\theta(x_0|c) = \int p(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t, c) dx_{1:T},$$

where $p(x_T) \in \mathcal{N}(0, 1)$.

- Noising:

$$x_t = \sqrt{\alpha_{t_i}} x + \sqrt{1 - \alpha_{t_i}} \epsilon_i.$$

- ELBO approximation

$$-\mathbb{E}_\epsilon \left[\sum_{t=1}^{T} w_t \|\epsilon - \epsilon_\theta(x_t, c)\|^2 - \log p_\theta(x_0 \mid x_1, c) \right] + C.$$

Methodology

- Classification with Bayes' theorem:

$$p_\theta(c_i \mid \mathbf{x}) = \frac{p(c_i)p_\theta(\mathbf{x} \mid c_i)}{\sum_j p(c_j)p_\theta(\mathbf{x} \mid c_j)}.$$

- Monte Carlo estimation for prediction error:

$$\mathbb{E}_{t,\epsilon}[\|\epsilon - \epsilon_\theta(x_t, c)\|^2] \approx \frac{1}{N}\sum_{i=1}^{N} \left\|\epsilon_i - \epsilon_\theta\left(\sqrt{\alpha_{t_i}}\mathbf{x} + \sqrt{1-\alpha_{t_i}}\epsilon_i, c_j\right)\right\|^2.$$

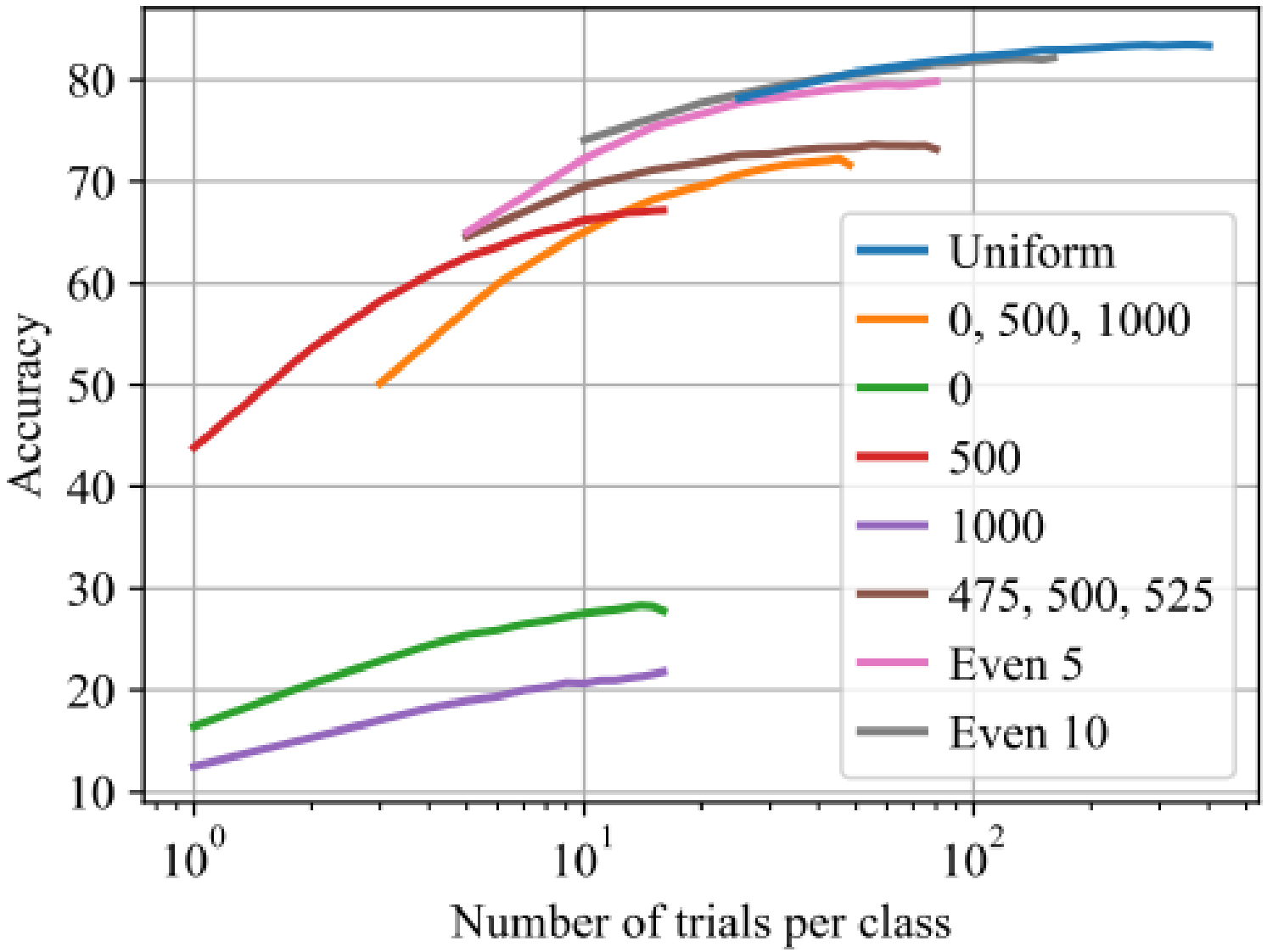
- Final result:

$$p_\theta(c_i|x) = \frac{1}{\sum_j \exp\{\mathbb{E}_{t,\epsilon}[\|\epsilon - \epsilon_\theta(x_t, c_i)\|^2 - \|\epsilon - \epsilon_\theta(x_t, c_j)\|^2]\}}.$$

Diffusion Classifier Overview

- Uses pre-trained diffusion models like Stable Diffusion.
- Estimates conditional densities $p(x|c)$ for each class c.
- Key steps:
  1. Add noise to the input image.
  2. Predict noise for each class.
  3. Compare prediction errors to identify the best class.
- No additional training required.

Practical Insights



Experiments

	Zero-shot?	Food	CIFAR10	Aircraft	Pets	Flowers	STL10	ImageNet	ObjectNet
Synthetic SD Data	✓	12.6	35.3	9.4	31.3	22.1	38.0	18.9	5.2
SD Features	✗	73.0	84.0	**35.2**	75.9	**70.0**	87.2	56.6	10.2
Diffusion Classifier (ours)	✓	**77.9**	**87.1**	24.3	**86.2**	59.4	**95.3**	**58.9**	**38.3**
CLIP ResNet-50	✓	81.1	75.6	19.3	85.4	65.9	94.3	58.2	40.0
OpenCLIP ViT-H/14	✓	92.7	97.3	42.3	94.6	79.9	98.3	76.8	69.2

Experiments

Method	ID	OOD		
	IN	IN-v2	IN-A	ObjectNet
ResNet-18	70.6	58.3	0.3	26.6
ResNet-34	73.1	61.3	0.4	31.6
ResNet-50	77.6	63.2	0.0	35.6
ResNet-101	77.6	66.8	1.6	38.2
ViT-L/32	78.0	65.6	5.2	29.9
ViT-L/16	80.3	68.7	7.5	36.7
ViT-B/16	**81.5**	**69.7**	9.4	**37.8**
Diffusion Classifier	77.3	64.5	**19.6**	32.3

Conclusion

- Generative models like diffusion models can perform discriminative tasks.
- Diffusion Classifier achieves competitive results in zero-shot and supervised classification.
- Future directions:
  - Reducing inference time.
  - Optimizing models for classification tasks.

Questions?