

# CLASSIFIER-FREE DIFFUSION GUIDANCE

Нигматзянов Тимур БПМИ211

# План доклада

- Вспомним диффузионную модель
- Обучение с использованием классификатора
- Минусы такого подхода
- Безклассификаторное обучение
- Эксперименты
- Заключение

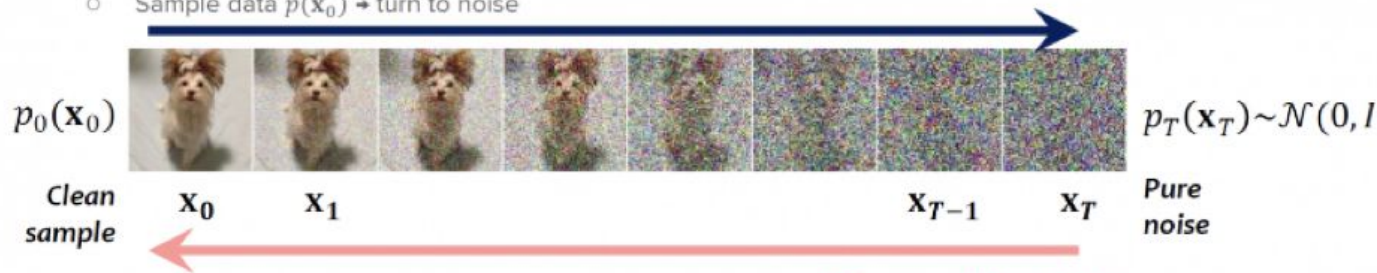
# Диффузионная модель

**Диффузионные модели** - одни из видов генеративных моделей, она стали крайне популярны последнее время благодаря своему высокому качеству генерации объектов из заданного распределения

## Denoising diffusion models

- **Forward / noising process**

- Sample data  $p(\mathbf{x}_0)$  → turn to noise



- **Reverse / denoising process**

# Модель с использованием классификатора

Для повышения качества генерации изображений был предложен новый метод с использованием классификатора

Основные этапы:

- Тренировка диффузионной модели
- Тренировка классификатора
- Совмещение моделей в процессе генерации

# Введение



Figure 2: The effect of guidance on a mixture of three Gaussians, each mixture component representing data conditioned on a class. The leftmost plot is the non-guided marginal density. Left to right are densities of mixtures of normalized guided conditionals with increasing guidance strength.

На этом изображении показан эффект направления на смеси трех гауссовских распределений, где каждый компонент смеси представляет данные, обусловленные на классе

# Модель с использованием классификатора

В Nichol & Dhariwal (2021) был предложен метод условной генерации, который повышает качество генерируемых картинок, при этом уменьшая их разнообразие. Для этого предобучается «шумный» классификатор на зашумленных картинках. Затем он используется во время сэмплирования, корректируя предсказанное среднее. Важно, что исходная диффузионная модель никак не меняется, что делает трюк еще более привлекательным.

$$p_\phi(y|x_t) \longrightarrow \nabla_x \log p_\phi(y|x_t) =$$

$$p_\theta(x_{i-1}|x_i) \text{ в } p_\theta(x_{i-1}|x_i, y)$$

# Алгоритм обучения

---

**Algorithm 1** Classifier guided diffusion sampling, given a diffusion model  $(\mu_\theta(x_t), \Sigma_\theta(x_t))$ , classifier  $p_\phi(y|x_t)$ , and gradient scale  $s$ .

---

Input: class label  $y$ , gradient scale  $s$

$x_T \leftarrow$  sample from  $\mathcal{N}(0, \mathbf{I})$

**for all**  $t$  from  $T$  to 1 **do**

$\mu, \Sigma \leftarrow \mu_\theta(x_t), \Sigma_\theta(x_t)$

$x_{t-1} \leftarrow$  sample from  $\mathcal{N}(\mu + s\Sigma \nabla_{x_t} \log p_\phi(y|x_t), \Sigma)$

**end for**

**return**  $x_0$

---

# Минусы такого подхода

Основные проблемы метода с использованием классификатора:

- Сложность:** Требуется тренировка дополнительного классификатора
- Совместимость:** Классификатор должен быть обучен на шумных данных, что усложняет его тренировку



# CLASSIFIER-FREE DIFFUSION GUIDANCE

Авторы статьи предлагают новый метод, который исключает необходимость использования классификатора.

Основные этапы:

- Совместная тренировка: Обучаются две модели – условная и безусловная
- Комбинация скоринга: Результаты этих моделей комбинируются для создания итогового изображения

# Алгоритм обучения

---

**Algorithm 1** Joint training a diffusion model with classifier-free guidance

---

**Require:**  $p_{\text{uncond}}$ : probability of unconditional training

- 1: **repeat**
  - 2:    $(\mathbf{x}, \mathbf{c}) \sim p(\mathbf{x}, \mathbf{c})$  ▷ Sample data with conditioning from the dataset
  - 3:    $\mathbf{c} \leftarrow \emptyset$  with probability  $p_{\text{uncond}}$  ▷ Randomly discard conditioning to train unconditionally
  - 4:    $\lambda \sim p(\lambda)$  ▷ Sample log SNR value
  - 5:    $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  - 6:    $\mathbf{z}_\lambda = \alpha_\lambda \mathbf{x} + \sigma_\lambda \epsilon$  ▷ Corrupt data to the sampled log SNR value
  - 7:   Take gradient step on  $\nabla_\theta \|\epsilon_\theta(\mathbf{z}_\lambda, \mathbf{c}) - \epsilon\|^2$  ▷ Optimization of denoising model
  - 8: **until** converged
-

# Генерация выборки

---

**Algorithm 2** Conditional sampling with classifier-free guidance

---

**Require:**  $w$ : guidance strength

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

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

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

2: **for**  $t = 1, \dots, T$  **do**

$\triangleright$  Form the classifier-free guided score at log SNR  $\lambda_t$

3:  $\tilde{\epsilon}_t = (1 + w)\epsilon_\theta(\mathbf{z}_t, \mathbf{c}) - w\epsilon_\theta(\mathbf{z}_t)$

$\triangleright$  Sampling step (could be replaced by another sampler, e.g. DDIM)

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

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

6: **end for**

7: **return**  $\mathbf{z}_{T+1}$

---

# Мотивация такого подхода

Наше предсказание:

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

Мотивация этой формулы следовала из формулы Байеса:

$$p(y|x_t) \propto \frac{p(x_t|y)}{p(x_t)} \implies \log p(y|x_t) \propto \log p(x_t|y) - \log p(x_t) \implies \nabla_{x_t} \log p(y|x_t)$$

# Метрики

FID (Frechet Inception Distance):

$$\mathbf{FID} = |\mu - \mu_w|^2 + \text{tr}(\Sigma + \Sigma_w - 2(\Sigma\Sigma_w)^{1/2})$$

IS (Inception Score):

$$IS(p_{gen}, p_{dis}) := \exp\left(\mathbb{E}_{x \sim p_{gen}} \left[ D_{KL} \left( p_{dis}(\cdot|x) \parallel \int p_{dis}(\cdot|x) p_{gen}(x) dx \right) \right]\right)$$

# Плюсы такого подхода

- Упрощение архитектуры
- Снижение вычислительных затрат
- Устойчивость и стабильность
- Более гибкое управление генерацией:

# Эксперименты

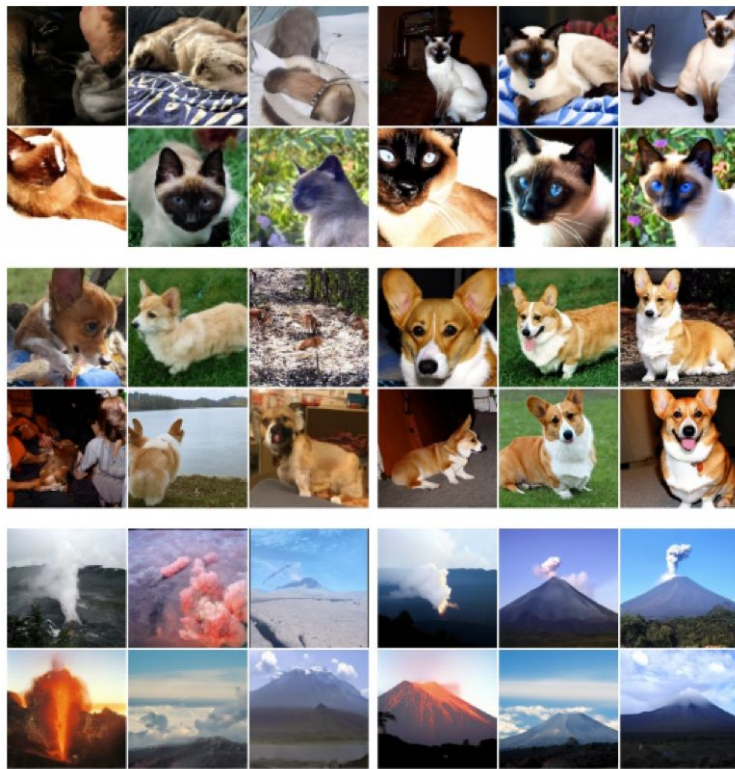


Figure 3: Classifier-free guidance on 128x128 ImageNet. Left: non-guided samples, right: classifier-free guided samples with  $w = 3.0$ . Interestingly, strongly guided samples such as these display saturated colors. See Fig. 8 for more.

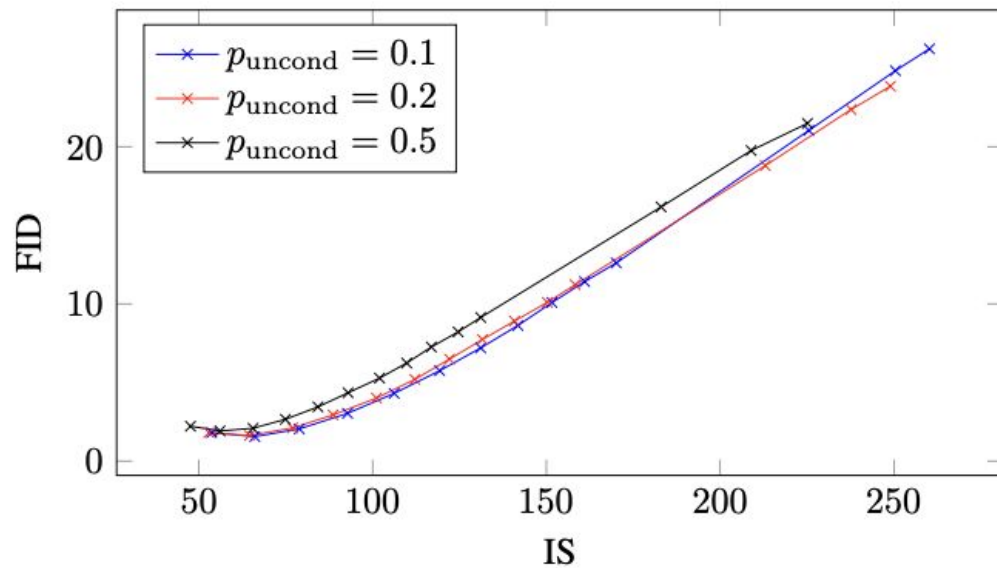
# Эксперименты

Model	FID (↓)	IS (↑)
ADM (Dhariwal & Nichol, 2021)	2.07	-
CDM (Ho et al., 2021)	<b>1.48</b>	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	<b>260.2</b> / 248.97 / 225.1

Table 1: ImageNet 64x64 results ( $w = 0.0$  refers to non-guided models).



# Эксперименты

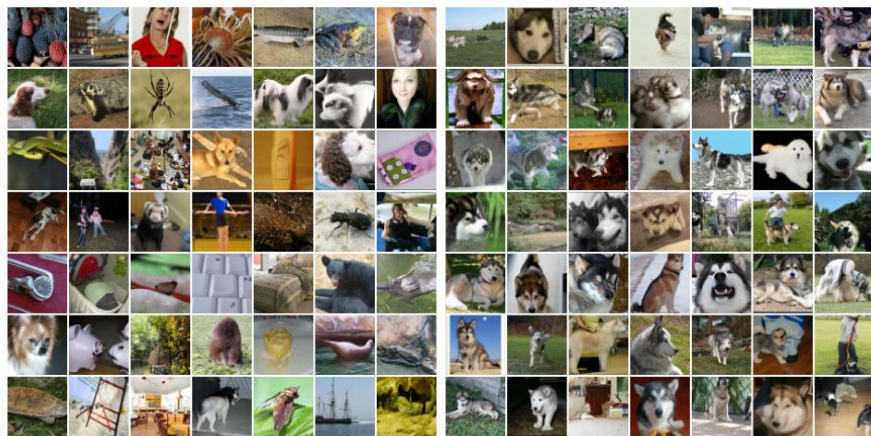


# Генерация

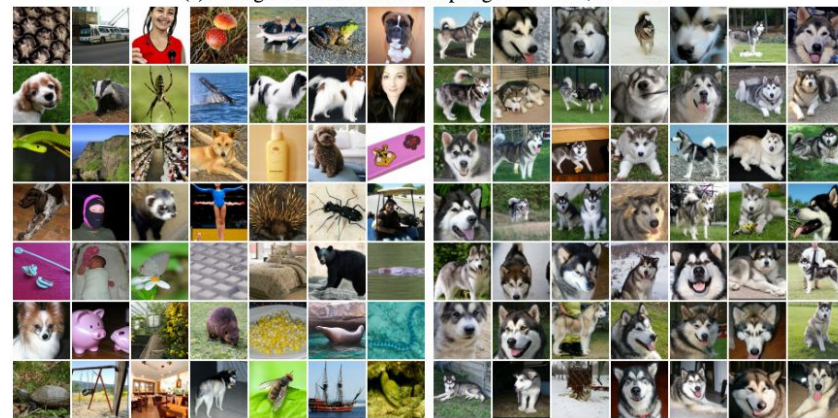
Здесь представлены результаты генерации на ImageNet 64x64.

Слева: случайные классы

Справа: один класс

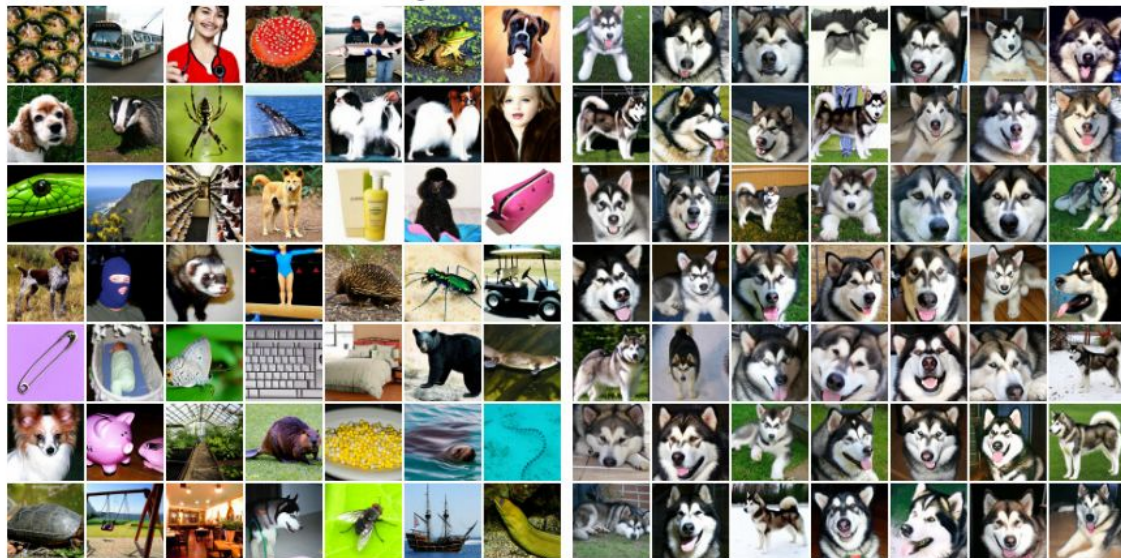


(a) Non-guided conditional sampling: FID=1.80, IS=53.71



(b) Classifier-free guidance with  $w = 1.0$ : FID=12.6, IS=170.1

# Генерация



(c) Classifier-free guidance with  $w = 3.0$ : FID=24.83, IS=250.4

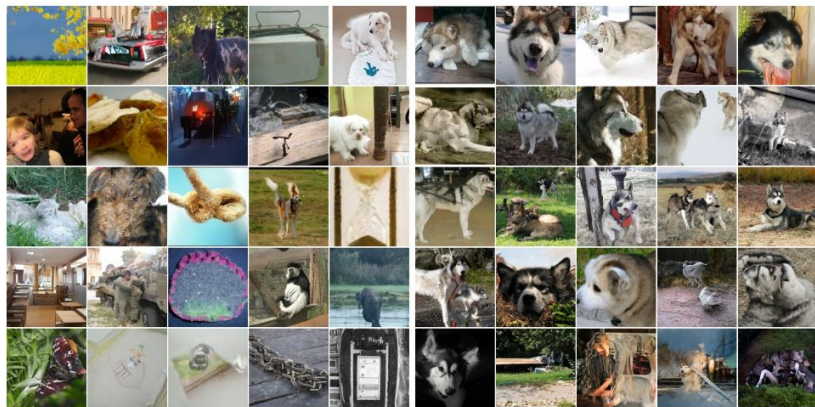


# Генерация

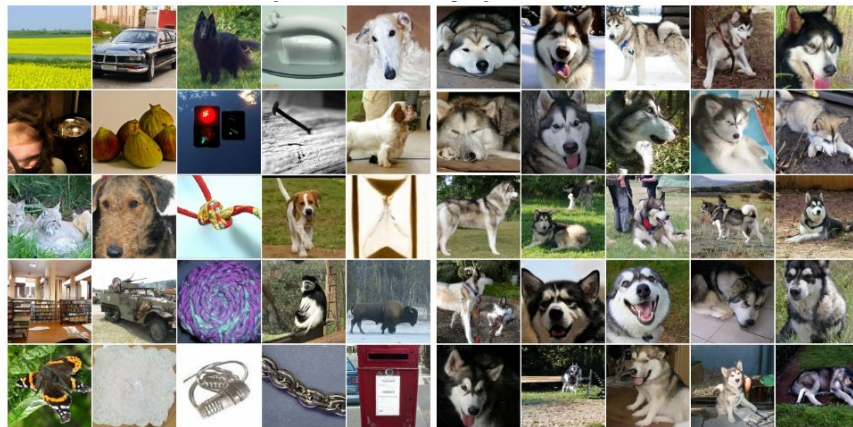
Здесь представлены результаты генерации на ImageNet  
128x128.

Слева: случайные классы

Справа: один класс

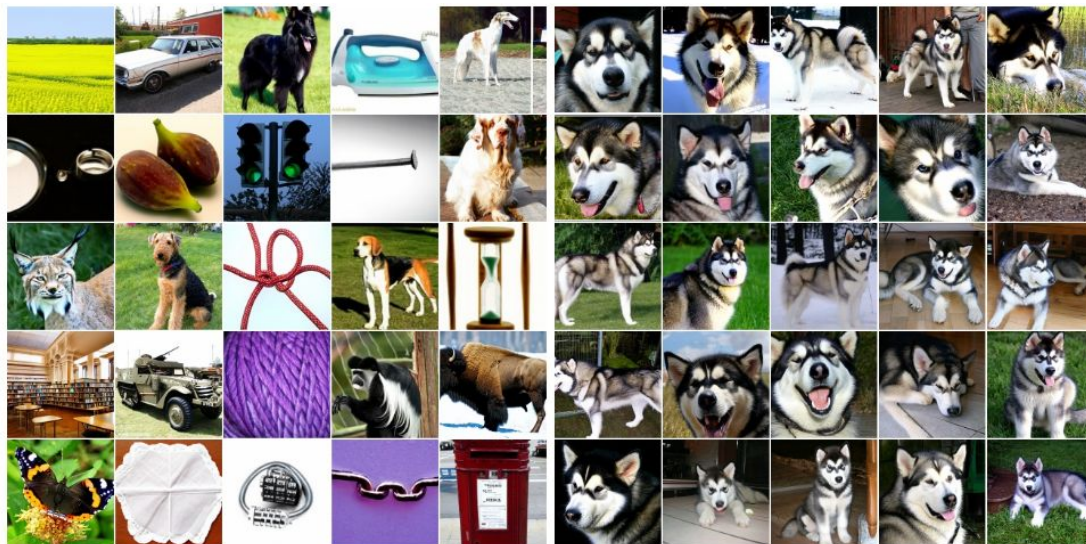


(a) Non-guided conditional sampling: FID=7.27, IS=82.45



(b) Classifier-free guidance with  $w = 1.0$ : FID=7.86, IS=297.98

# Генерация



(c) Classifier-free guidance with  $w = 4.0$ : FID=21.53, IS=421.03

# Заключение

Самым практическим преимуществом нашего метода управления без классификатора является его чрезвычайная простота: это всего лишь однострочное изменение кода во время обучения. Также большим преимуществом является качество генерации.