Classifier-Free Diffusion Guidance

Обзор-рецензия на статью

Рахманов Сергей, БПМИ192

Text2Image: Авторегрессионные модели

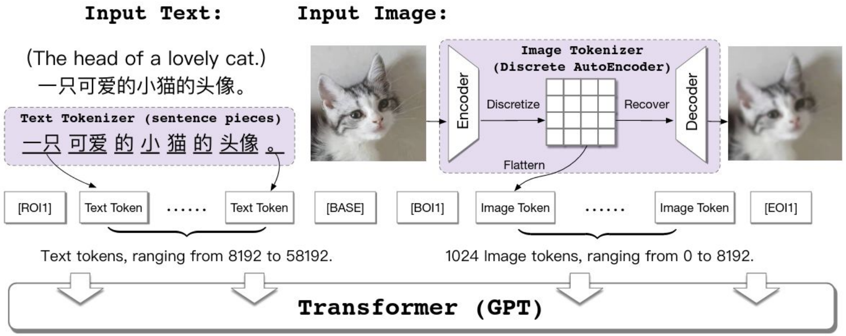


Figure 3: The framework of CogView. [ROI1], [BASE1], etc., are seperator tokens.

CogView (Ding et al.)

Zero-Shot Text-to-Image Generation

(a) a tapir made of accordion. a tapir with the texture of an accordion. (b) an illustration of a baby hedgehog in a christmas sweater walking a dog (c) a neon sign that reads "backprop". a neon sign that reads "backprop". backprop neon sign (d) the exact same cat on the top as a sketch on the bottom

Figure 2. With varying degrees of reliability, our model appears to be able to combine distinct concepts in plausible ways, create anthropomorphized versions of animals, render text, and perform some types of image-to-image translation.

DALL-E 1 (OpenAI)

Text2Image: Classifier Guidance (Nichol & Dhariwal)

Figure 3: Samples from an unconditional diffusion model with classifier guidance to condition on the class "Pembroke Welsh corgi". Using classifier scale 1.0 (left; FID: 33.0) does not produce convincing samples in this class, whereas classifier scale 10.0 (right; FID: 12.0) produces much more class-consistent images.

Text2Image: Classifier Guidance (Nichol & Dhariwal)

Figure 3: Samples from an unconditional diffusion model with classifier guidance to condition on the class "Pembroke Welsh corgi". Using classifier scale 1.0 (left; FID: 33.0) does not produce convincing samples in this class, whereas classifier scale 10.0 (right; FID: 12.0) produces much more class-consistent images.

Но можно и без классификатора!

Text2Image: Classifier-free Guidance

OpenReview.net   Search OpenReview...   Login

← Go to **NeurIPS 2021 Workshop DGMs Applications** homepage

Classifier-Free Diffusion Guidance

Jonathan Ho, Tim Salimans

27 Sept 2021 (modified: 27 Nov 2021)   DGMs and Applications @ NeurIPS 2021 Poster   Readers: Everyone   Show Bibtex   Show Revisions

Keywords: diffusion, score

TL;DR: Classifier guidance without a classifier

Abstract: Classifier guidance is a recently introduced method to trade off mode coverage and sample fidelity in conditional diffusion models post training, in the same spirit as low temperature sampling or truncation in other types of generative models. This method combines the score estimate of a diffusion model with the gradient of an image classifier and thereby requires training an image classifier separate from the diffusion model. We show that guidance can be performed by a pure generative model without such a classifier: we jointly train a conditional and an unconditional diffusion model, and find that it is possible to combine the resulting conditional and unconditional scores to attain a trade-off between sample quality and diversity similar to that obtained using classifier guidance.

Reply Type: all   Author: everybody   Visible To: all readers   Hidden From: nobody   **1 Reply**

[-] **Paper Decision**

NeurIPS 2021 Workshop DGMs Applications Program Chairs

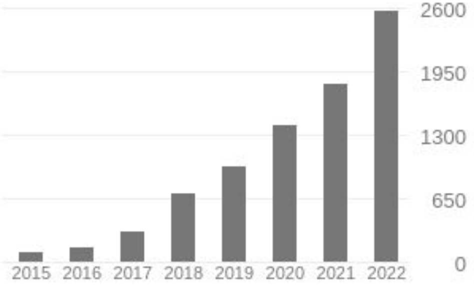
21 Oct 2021   NeurIPS 2021 Workshop DGMs Applications Paper30 Decision   Readers: Everyone

Decision: Accept (Poster)

Classifier-free Guidance: авторы

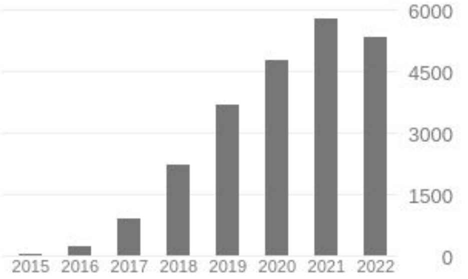
- **Jonathan Ho**
  - Google Brain
  - PhD, UC Berkeley

	All	Since 2017
Citations	8226	7841
h-index	26	26
i10-index	28	28



- **Tim Salimans**
  - Google Brain
  - PhD, Erasmus University

	All	Since 2017
Citations	23262	22752
h-index	29	29
i10-index	42	38



Суть статьи

- Совместно обучаем безусловную и условную диффузионные модели.
- С вероятностью p проводим обучение с условием.
- Во время inference используем условную модель для контролируемой генерации.

Сильные стороны статьи

- Простая, но практичная и элегантная идея
- Подробно описан механизм работы идеи
- Проведена серия экспериментов, показаны результаты и приведены примеры генераций
- Идея значима для последующих исследований

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.

Слабые стороны статьи

- Нет кода, нельзя воспроизвести результаты :(
- Сравнение с Classifier-guided приведено с одним w

Model	FID (↓)	IS (↑)
BigGAN-deep, max IS (Brock et al., 2019)	25	253
BigGAN-deep (Brock et al., 2019)	5.7	124.5
CDM (Ho et al., 2021)	3.52	128.8
LOGAN (Wu et al., 2019)	3.36	148.2
ADM-G (Dhariwal & Nichol, 2021)	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

Table 2: ImageNet 128x128 results ($w = 0.0$ refers to non-guided models).

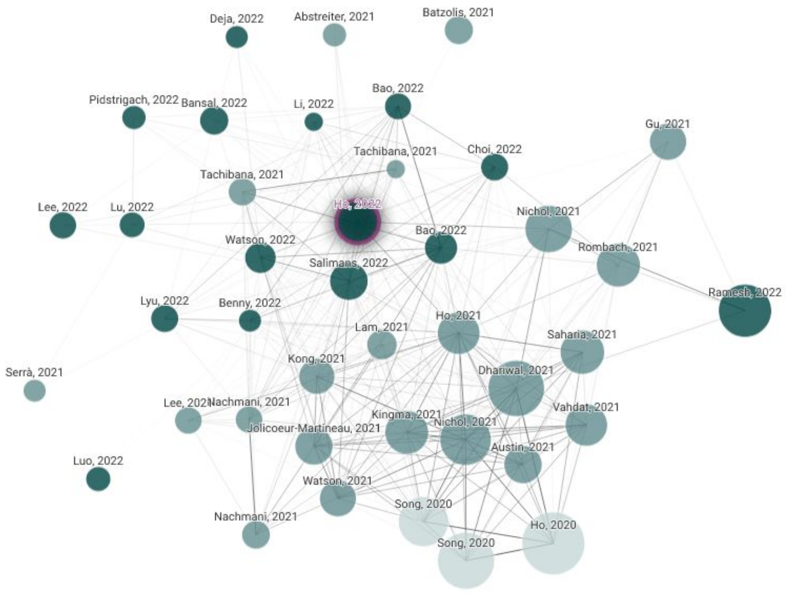
Использование статьи

Imagen, Imagen Video
(Google Research)

DALL·E 2 (OpenAI)

Использование статьи

- Image2Image задачи (Couairon et al.)
- Дифференциально-приватные диффузионные сети (Dockhorn et al.)
- Мета-обучение (Nava et al.)
- Всего 97 цитирований



Обратный guidance, разные модальности

- Что будет, если не проводить guidance, а наоборот штрафовать повышение условной вероятности. Получим ли семантически противоположное изображение?

- Classifier-free guidance в DiffusionLM (языковые модели), в DiffWave (генерация звука)

Обратный guidance, разные модальности

- Classifier-free guidance в ***DiffusionLM (языковые модели)***, в DiffWave (генерация звука)

SELF-CONDITIONED EMBEDDING DIFFUSION
FOR TEXT GENERATION

Robin Strudel [1*] Corentin Tallec [2] Florent Altché [2] Yilun Du [3*]
Yaroslav Ganin [2] Arthur Mensch [2] Will Grathwohl [2] Nikolay Savinov [2]
Sander Dieleman [2] Laurent Sifre [2] Rémi Leblond [2]

ABSTRACT

Can continuous diffusion models bring the same performance breakthrough on natural language they did for image generation? To circumvent the discrete nature of text data, we can simply project tokens in a continuous space of embeddings, as is standard in language modeling. We propose Self-conditioned Embedding Diffusion (SED), a continuous diffusion mechanism that operates on token embeddings and allows to learn flexible and scalable diffusion models for both conditional and unconditional text generation. Through qualitative and quantitative evaluation, we show that our text diffusion models generate samples comparable with those produced by standard autoregressive language models — while being in theory more efficient on accelerator hardware at inference time. Our work paves the way for scaling up diffusion models for text, similarly to autoregressive models, and for improving performance with recent refinements to continuous diffusion.