Diffusion art or Digital Forgery?
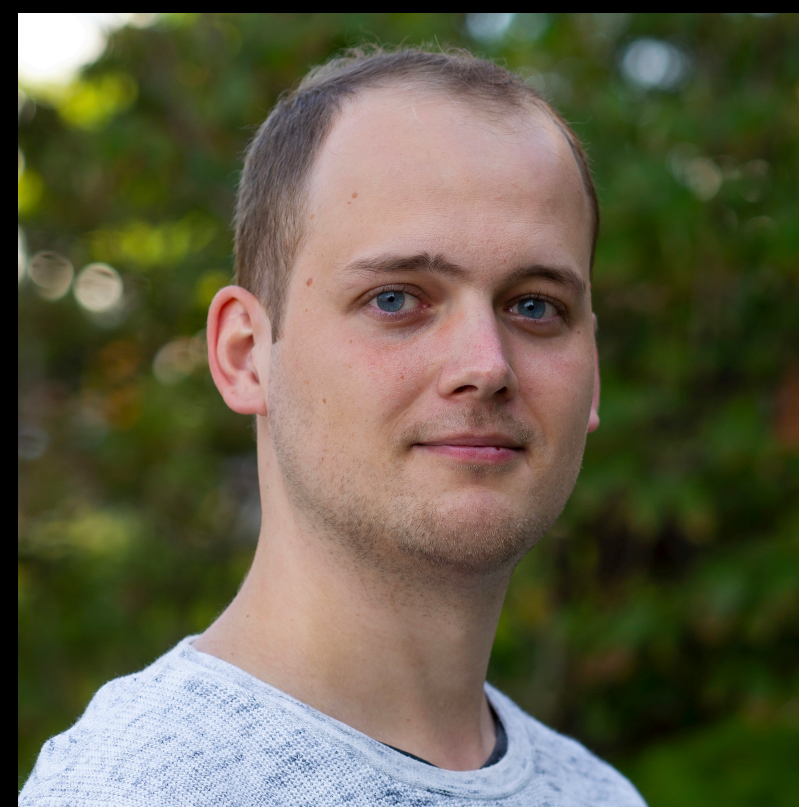
Investigating data replication in diffusion models

Gowthami Somepalli Vasu Singla Jonas Geiping Micah Goldblum Tom Goldstein
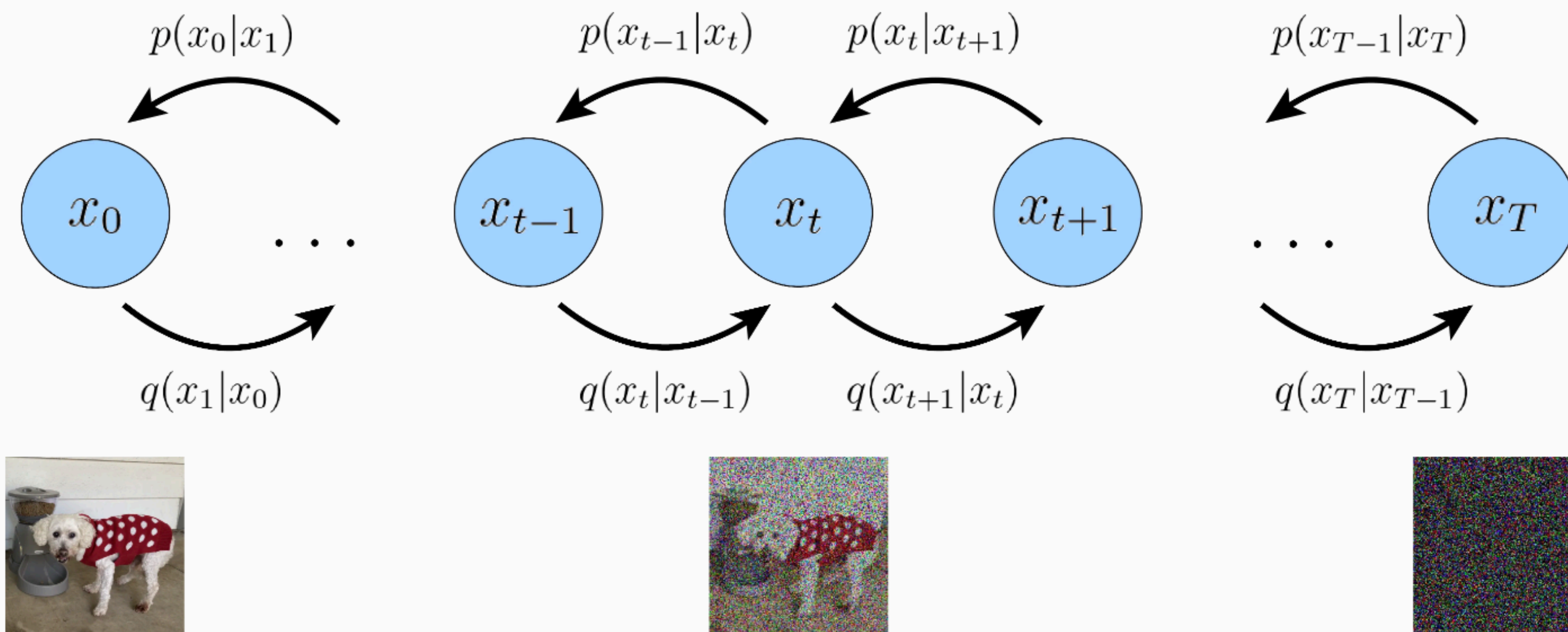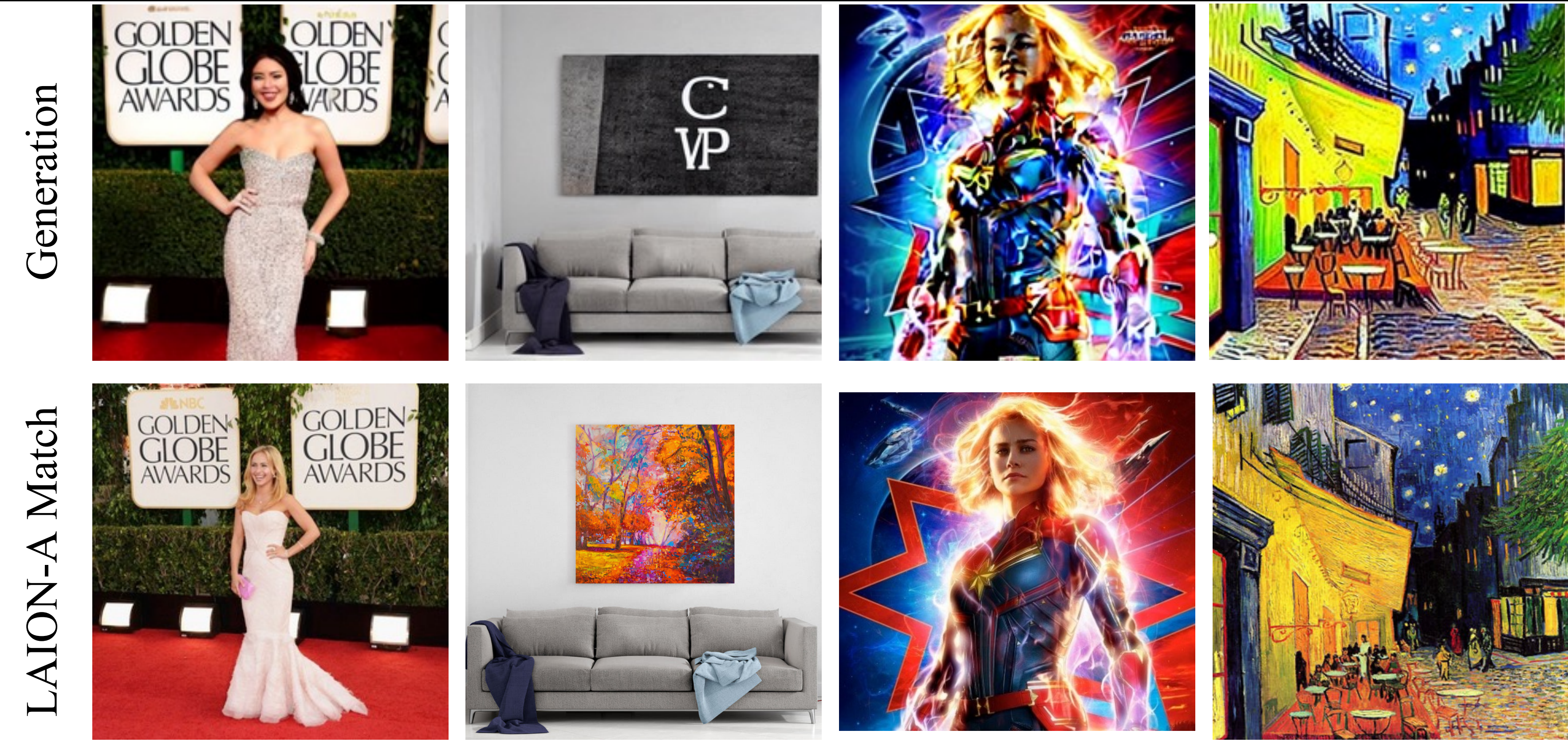
JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

https://somepago.github.io/diffrep.html

Diffusion models

$$p(x_0|x_1) \qquad p(x_{t-1}|x_t) \qquad p(x_t|x_{t+1}) \qquad p(x_{T-1}|x_T)$$

$$x_0 \quad \dots \quad x_{t-1} \quad x_t \quad x_{t+1} \quad \dots \quad x_T$$

$$q(x_1|x_0) \qquad q(x_t|x_{t-1}) \qquad q(x_{t+1}|x_t) \qquad q(x_T|x_{T-1})$$

Visual representation of a Variational Diffusion Model. An input is steadily noised over time until it becomes identical to Gaussian noise; a diffusion model learns to reverse this process.

Source: https://calvinyluo.com/2022/08/26/diffusion-tutorial.html

Do you think diffusion models replicate training data?

https://somepago.github.io/diffrep.html

Short answer: Yes.

Generation

LAION-A Match

https://somepago.github.io/diffrep.html

What counts as replication?

https://somepago.github.io/diffrep.html

What counts as replication?

Query
image

Perceptual + Semantic

Semantic

https://somepago.github.io/diffrep.html

What counts as replication?

Query image

Perceptual + Semantic
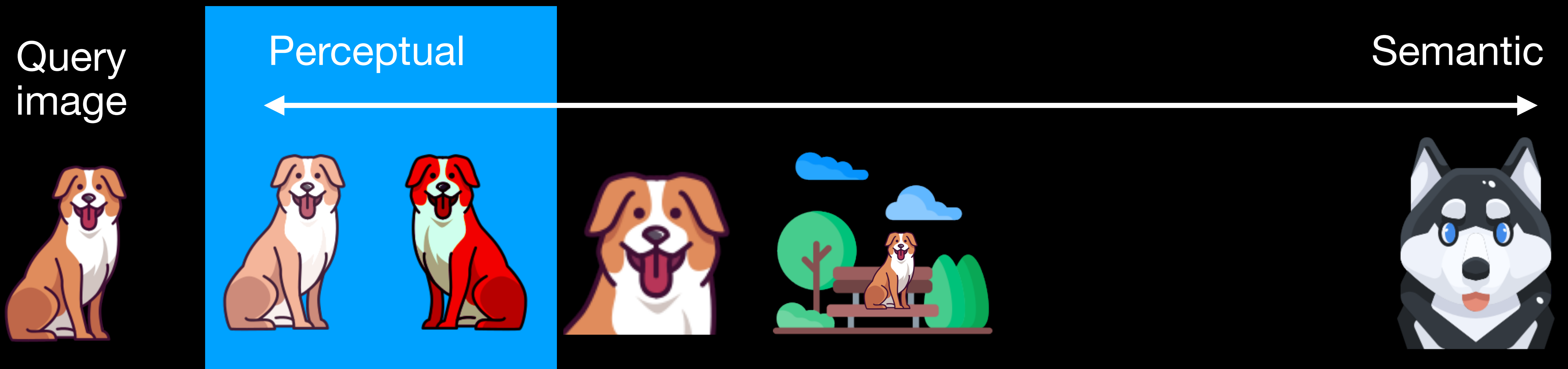
Semantic

https://somepago.github.io/diffrep.html

"We say that a generated image has replicated content if it contains an object (either in the foreground or background) that appears identically in a training image, neglecting minor variations in appearance that could result from data augmentation."

Can we detect replication with existing methods?

https://somepago.github.io/diffrep.html

Current methods

- Instance retrieval / Copy detection

- Self-supervised/ Supervised learning methods

Query image

Perceptual

Semantic

https://somepago.github.io/diffrep.html

Current datasets

- Oxford

- Paris

- INSTRE

- CUB-200

- GPR1200

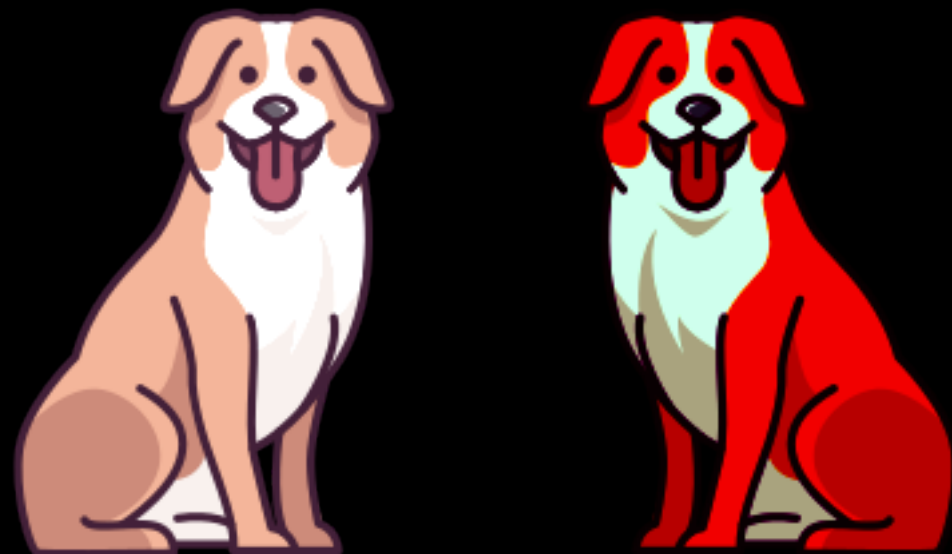Query image

Perceptual

Semantic

https://somepago.github.io/diffrep.html
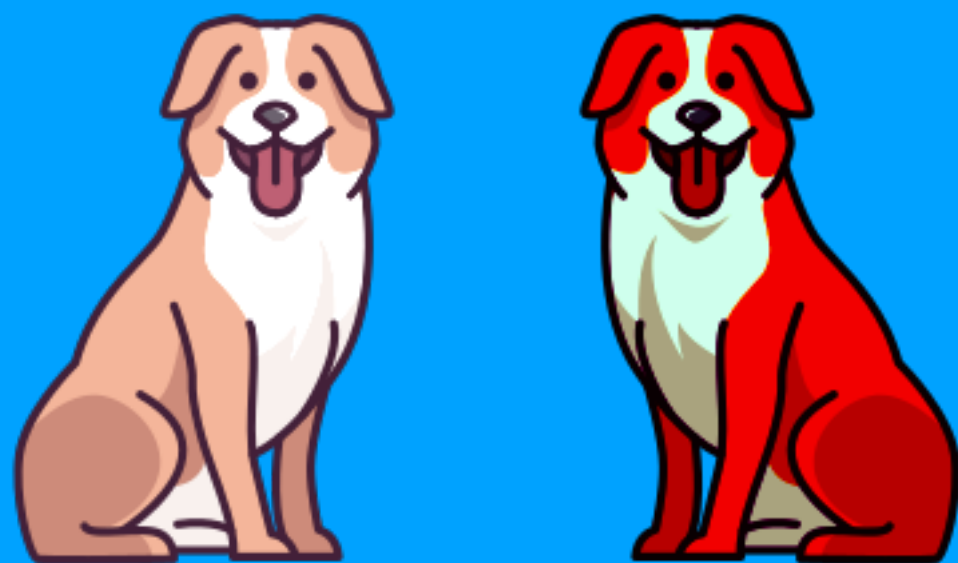
No partial retrieval datasets!

Query
image

Perceptual
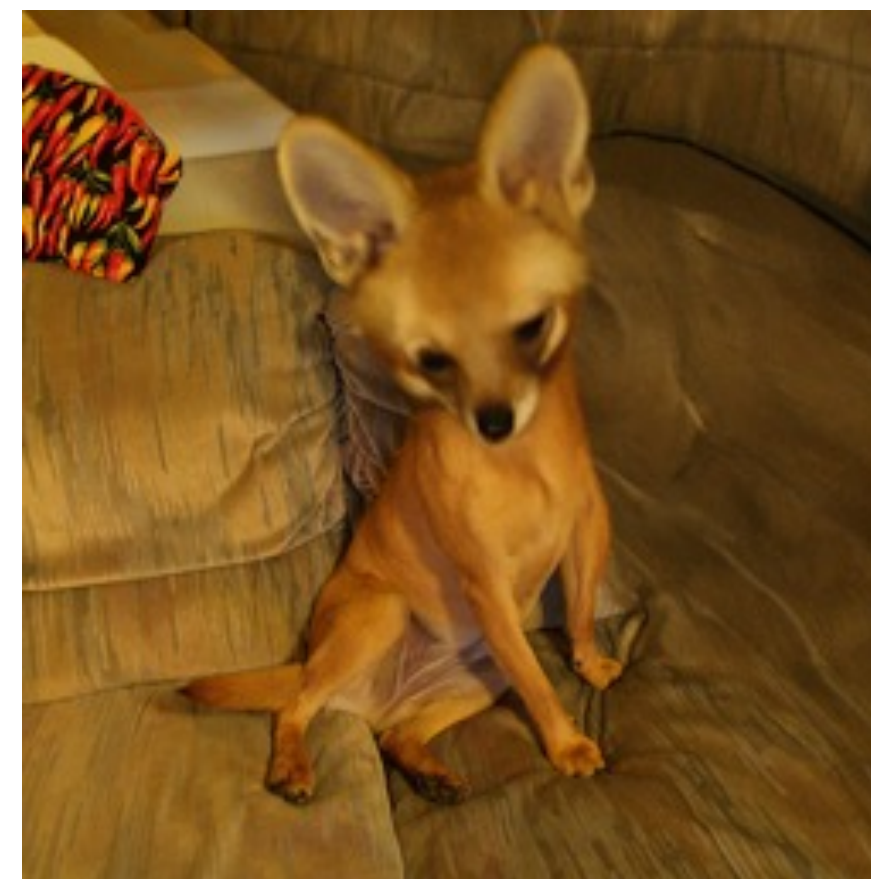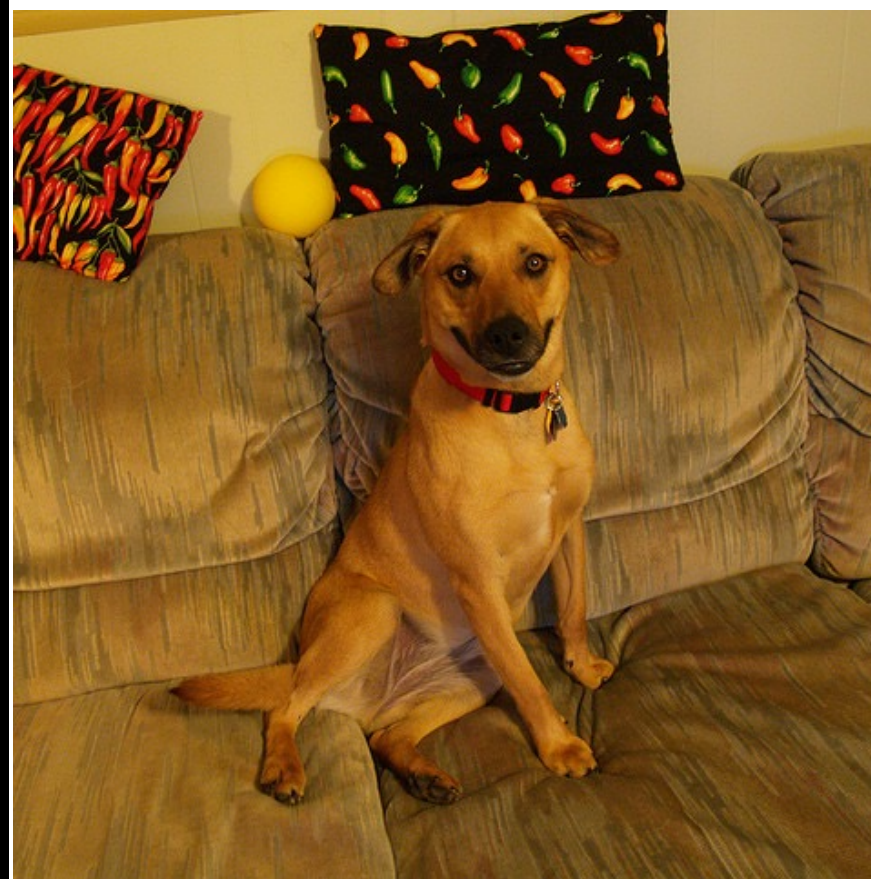
Semantic

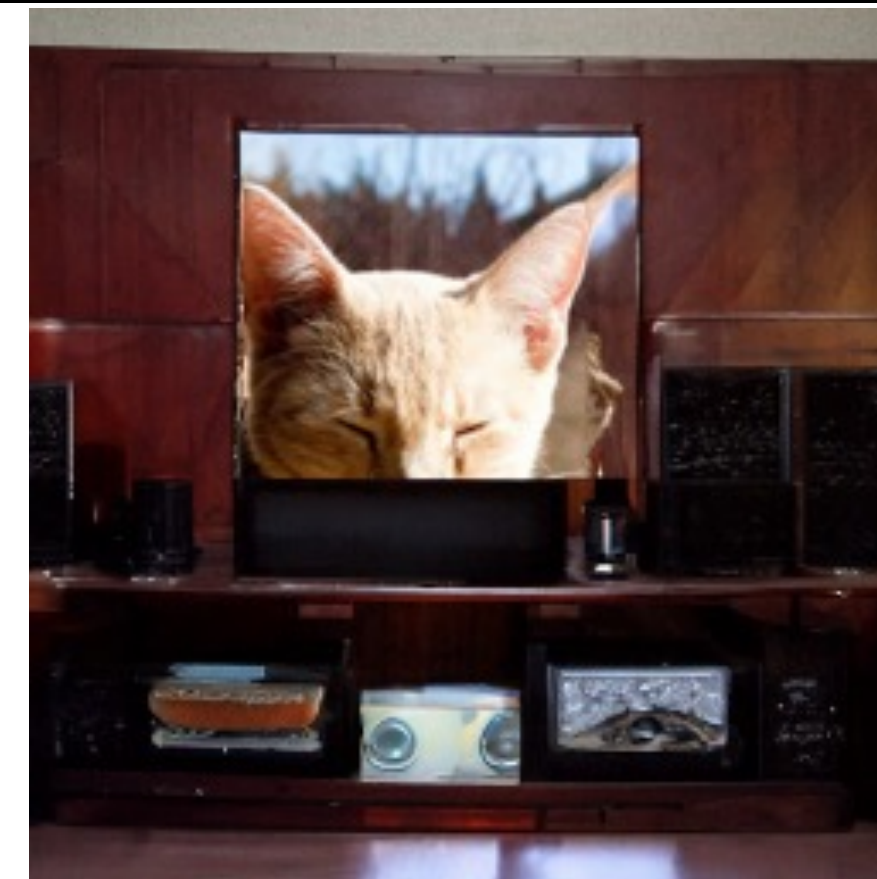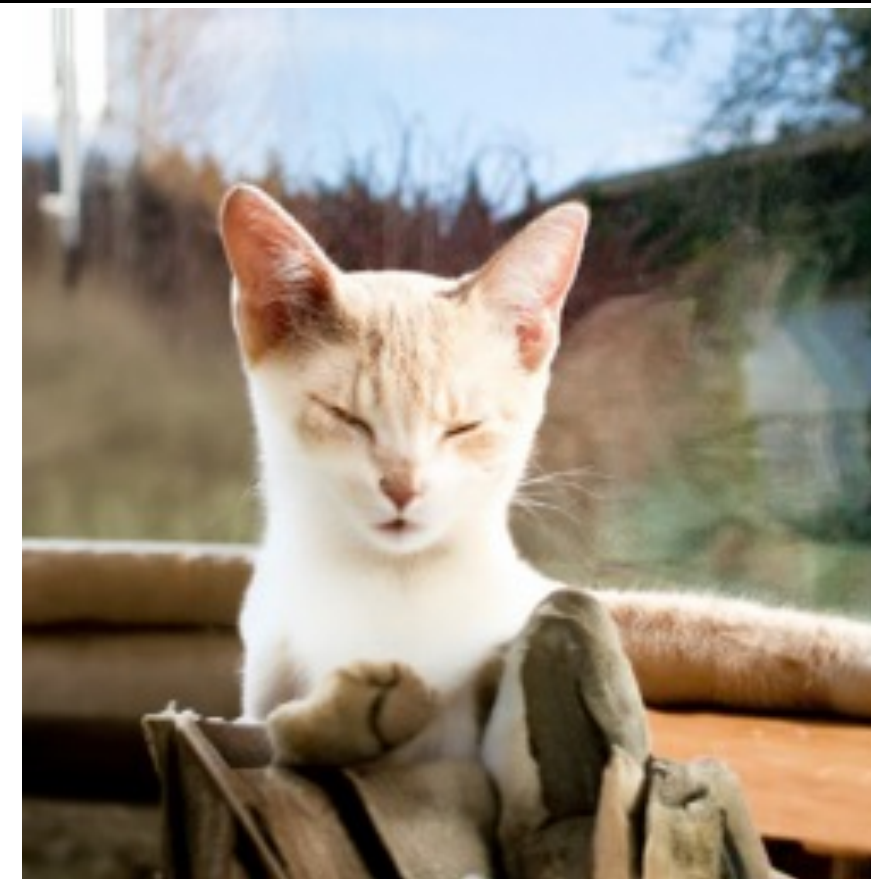https://somepago.github.io/diffrep.html
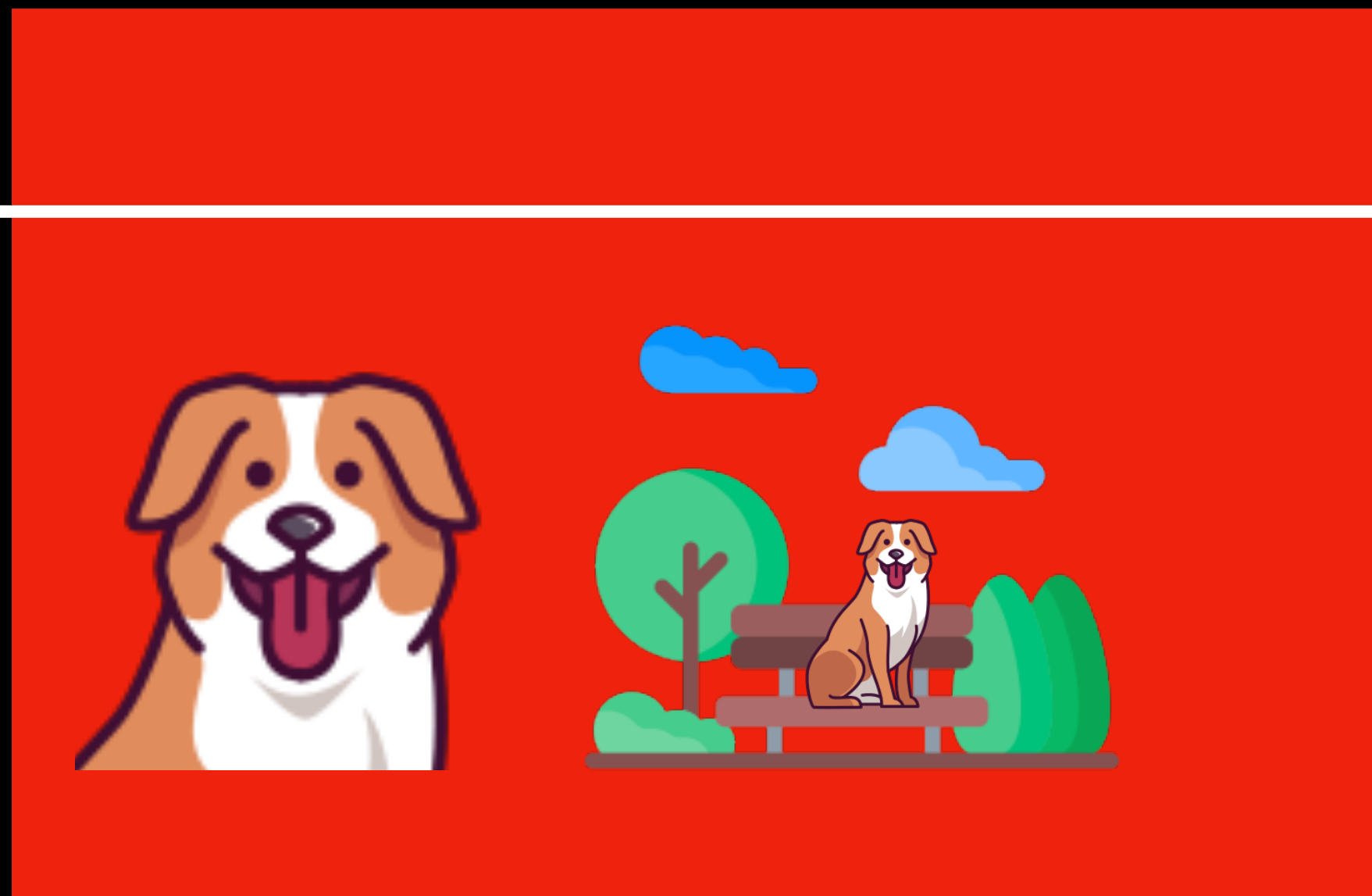
Synthetic datasets
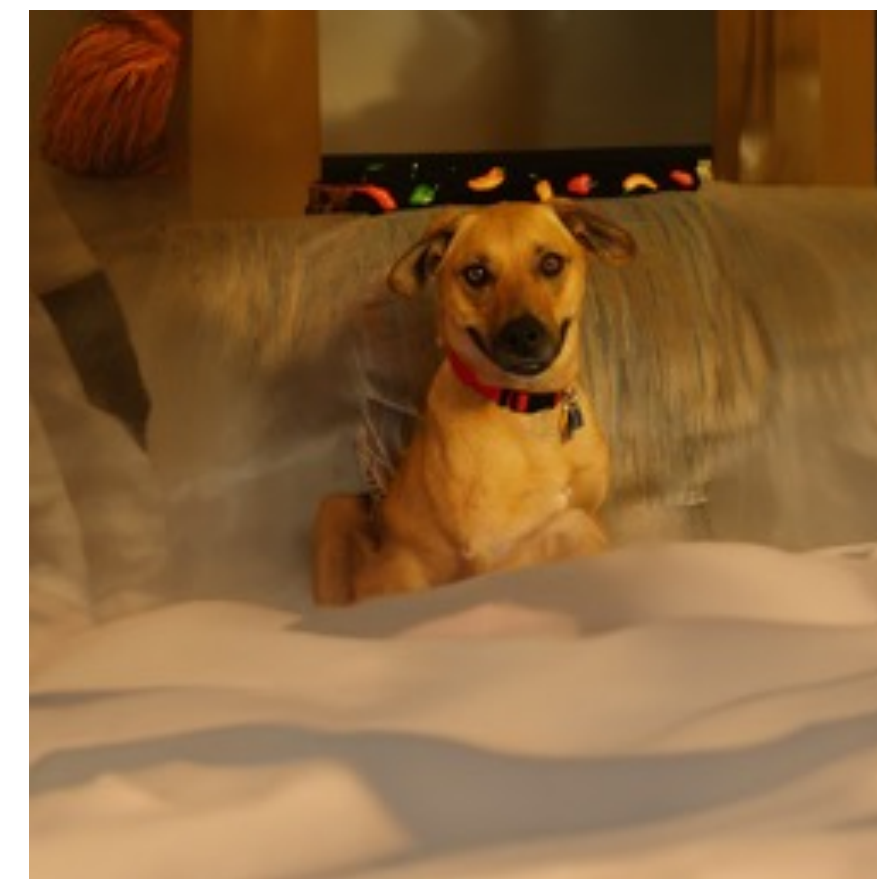
Ground truth   Segmix   Diagonal in-painting   Patch out-painting

15

How good are these methods?

Type	Method	Average Rank ↓
CD/IR	Multigrain [6], ResNet-50	5.9
	SSCD [47] , ResNet-50	**4.2**
PT	ViT [16] S/16, IN1k	5.5
	ViT-B/16, IN12k	9.2
	ViT-B/16, CLIP [51] on LAION [57]	5.4
	Swin Transformer [36], Base, IN1k	**4.1**
SSL	MoCo [15], ViT-B/16	5.1
	MoCo, ViT-B/16 + CutMix [68]	6.5
	VicRegL [5], ResNet-50	5.0
	DINO [12], ViT-B/16, split-product	**4.1**

https://somepago.github.io/diffrep.html

Can diffusion models copy?

https://somepago.github.io/diffrep.html
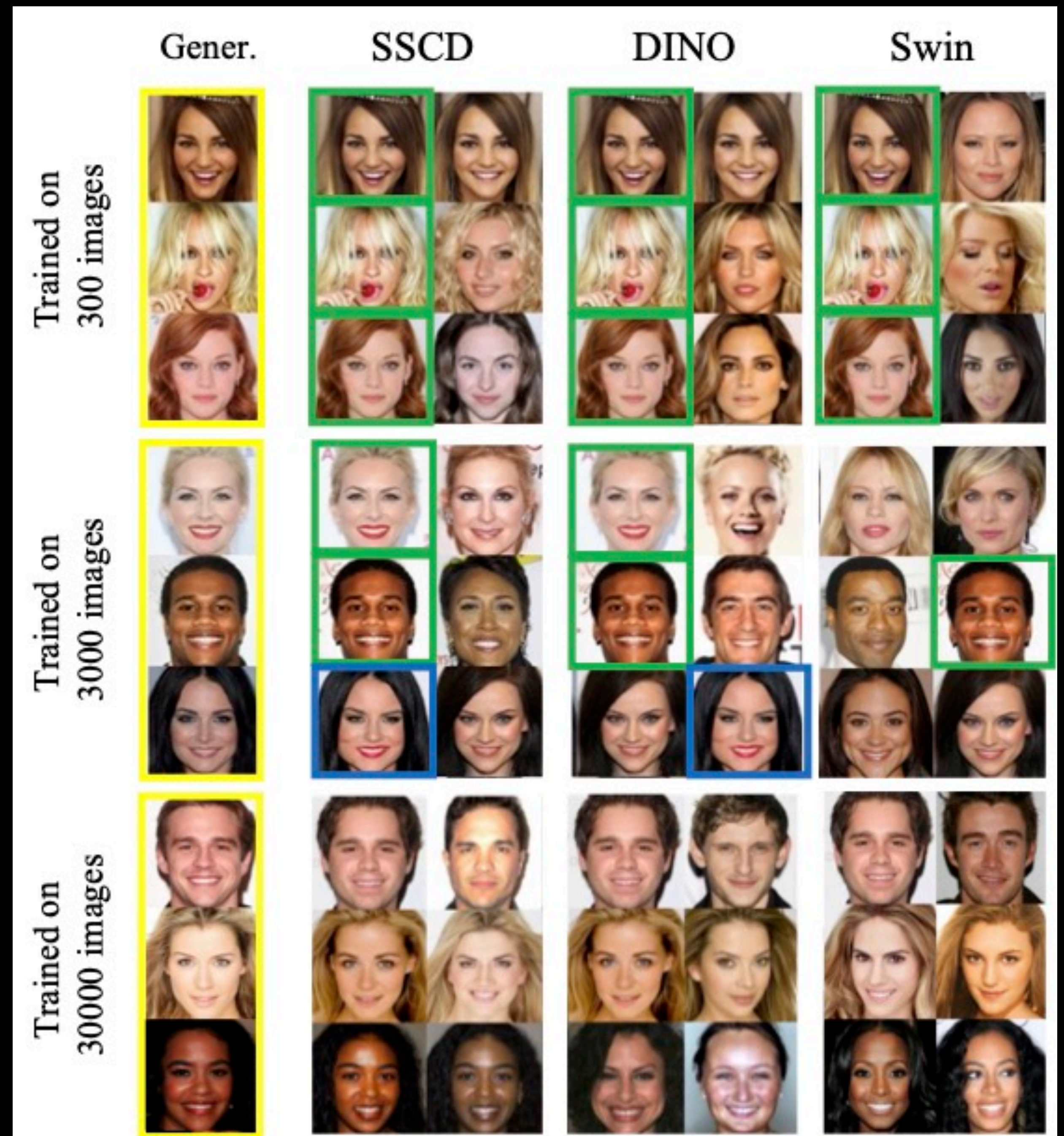
Can diffusion models copy?

* Overfit a small dataset and see if copying exists.

* Setup: Train a diffusion model (DDPM) on Celeb-A HQ with varying amounts of data (300 | 3000 | 30,000)
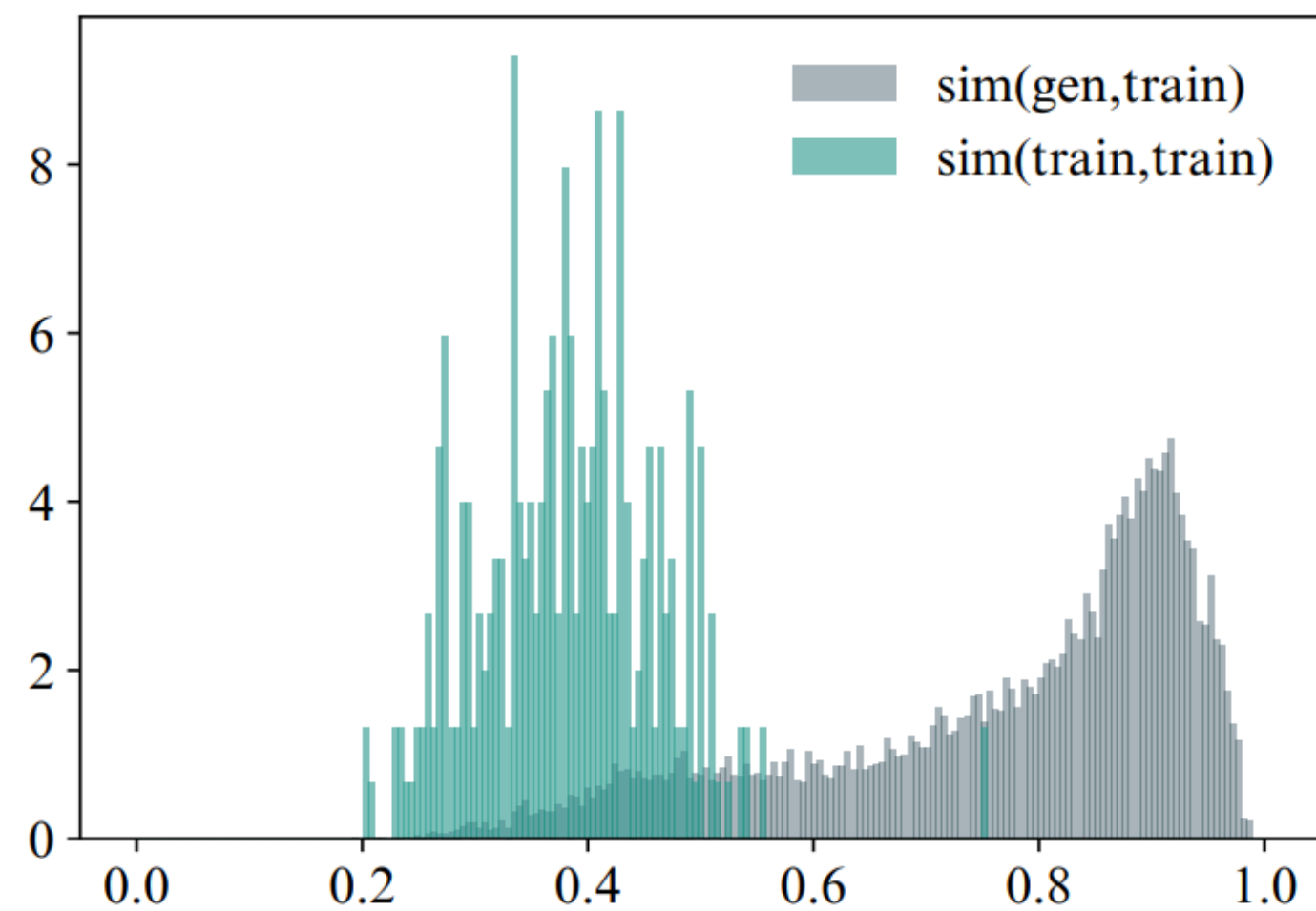
https://somepago.github.io/diffrep.html

Can diffusion models copy?

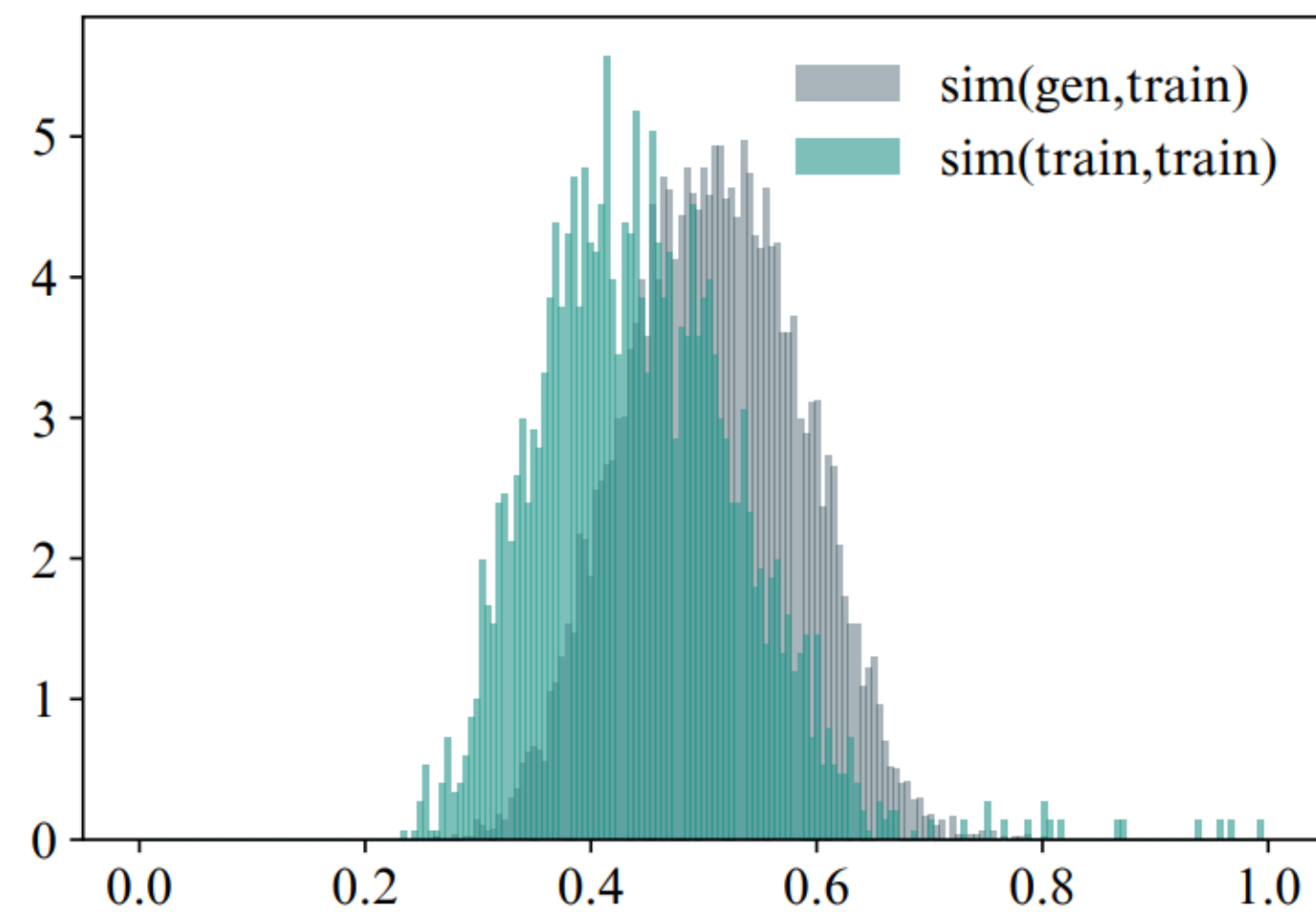https://somepago.github.io/diffrep.html

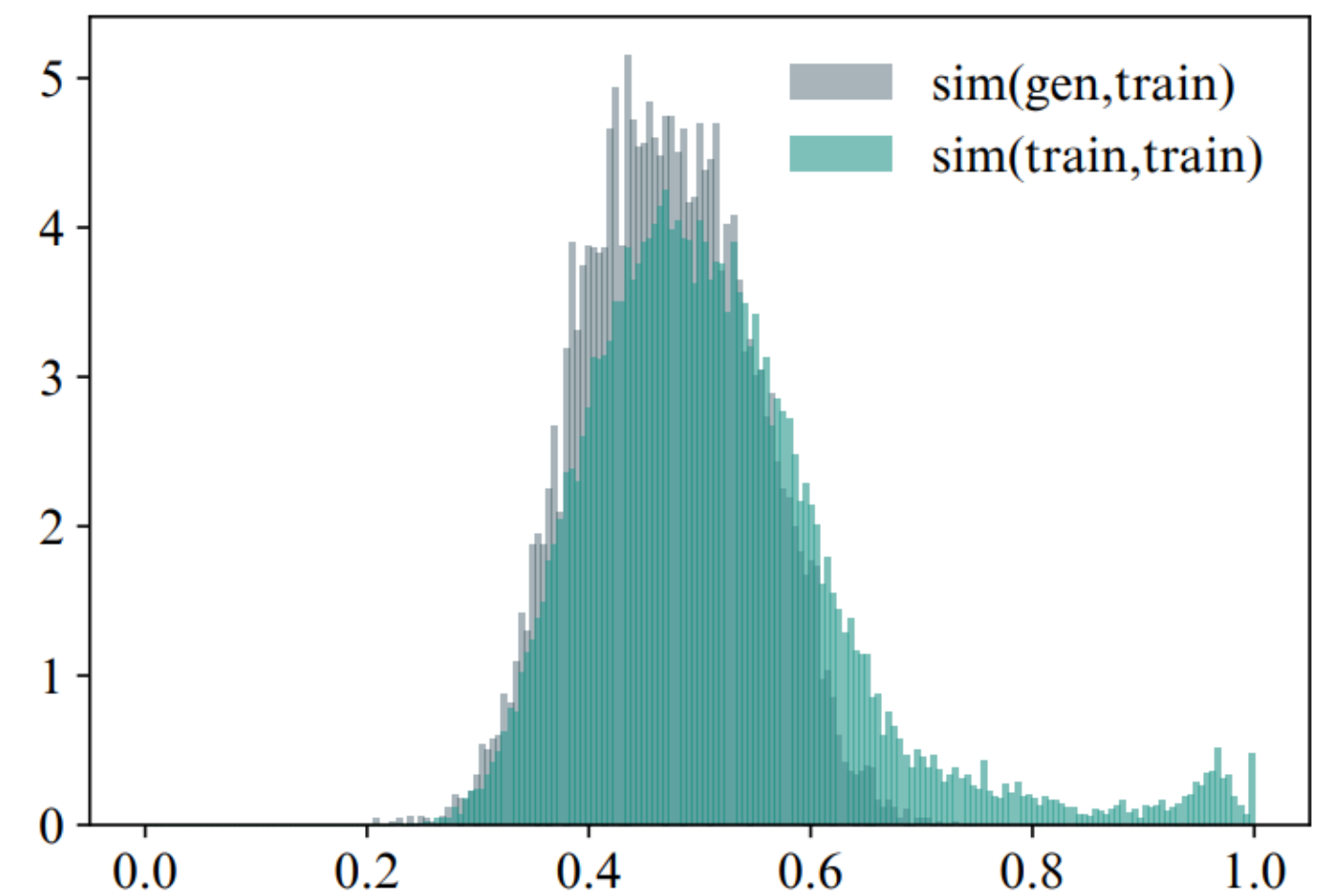# Can diffusion models copy?

- Setup:

  - Train a diffusion model (DDPM) on Celeb-A HQ with varying amounts of data (300 | 3000 | 30,000)

  - Eval on 10,000 generations

(a) 300 points   (b) 3000 points   (c) All points

https://somepago.github.io/diffrep.html

Do larger diffusion models copy?

- Latent diffusion model (class-conditioned) trained on ImageNet.

https://somepago.github.io/diffrep.html

Do larger diffusion models copy?

- Latent diffusion model (class-conditioned) trained on ImageNet.

https://somepago.github.io/diffrep.html

Do larger diffusion models copy?

- Latent diffusion model (class-conditioned) trained on ImageNet.

https://somepago.github.io/diffrep.html
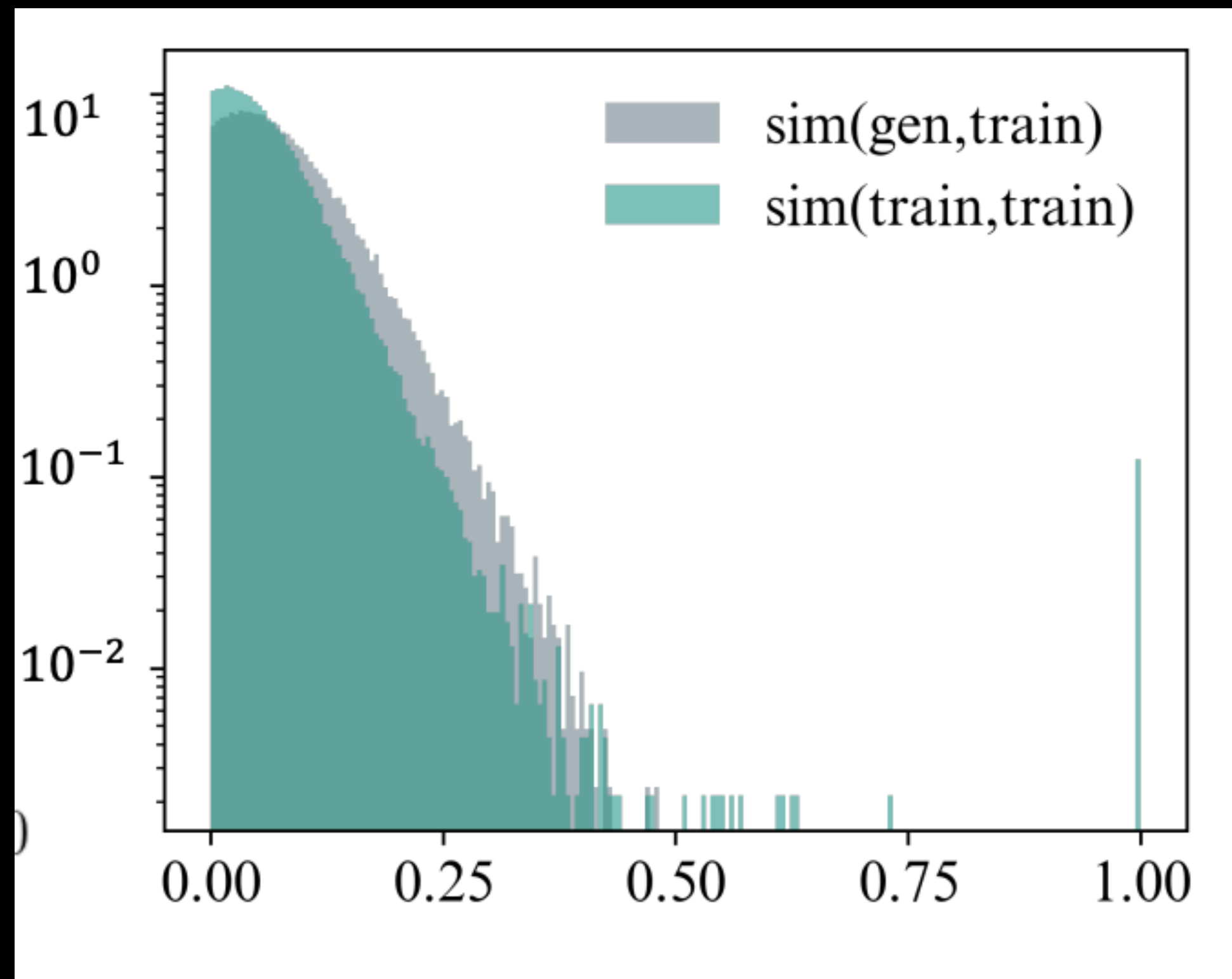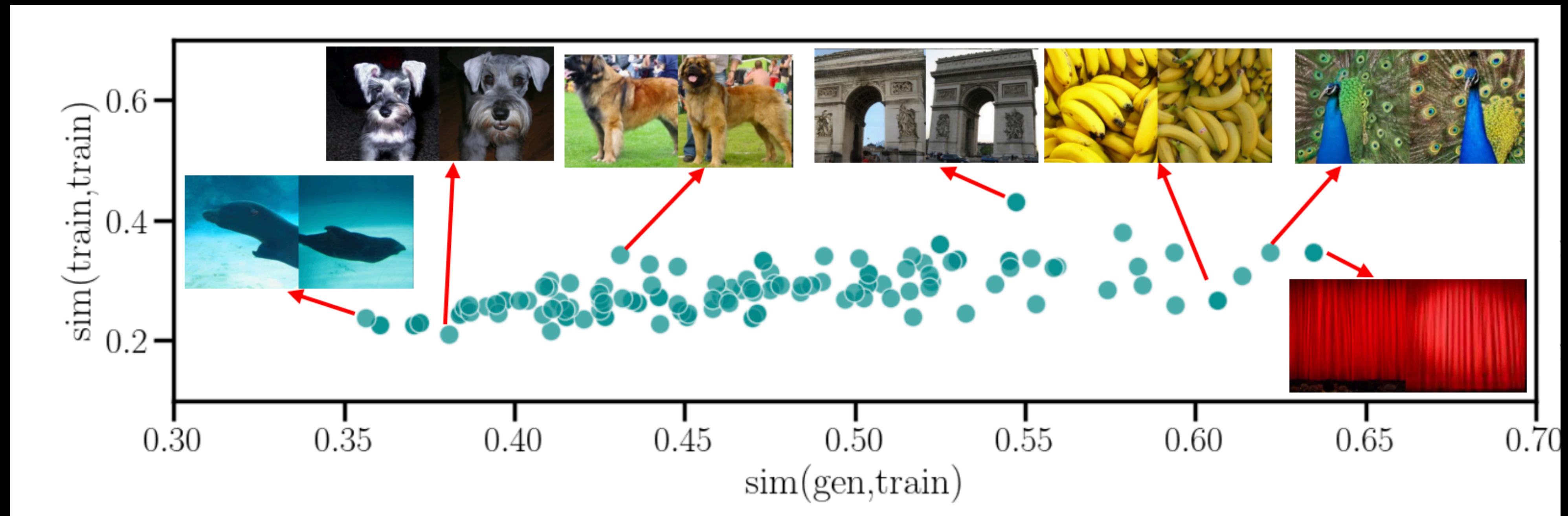
Diffusion models trained on larger datasets <u>do not</u> copy then? 🤔

24

Stable diffusion copies from training data!

Generation

LAION-A Match

https://somepago.github.io/diffrep.html

Experimental setup

- SD v1.4 which is trained on LAION 2B dataset and finetuned on LAION-Aesthetics V5+ (~400M images)

- 10,000 generations with captions randomly sampled from LAION-A V6+ (~12M pairs)

- All generations are compared against LAION-A V6+ (a subset of whole training data)

https://somepago.github.io/diffrep.html

~1.88%

Of generated data are potential copies.

https://somepago.github.io/diffrep.html
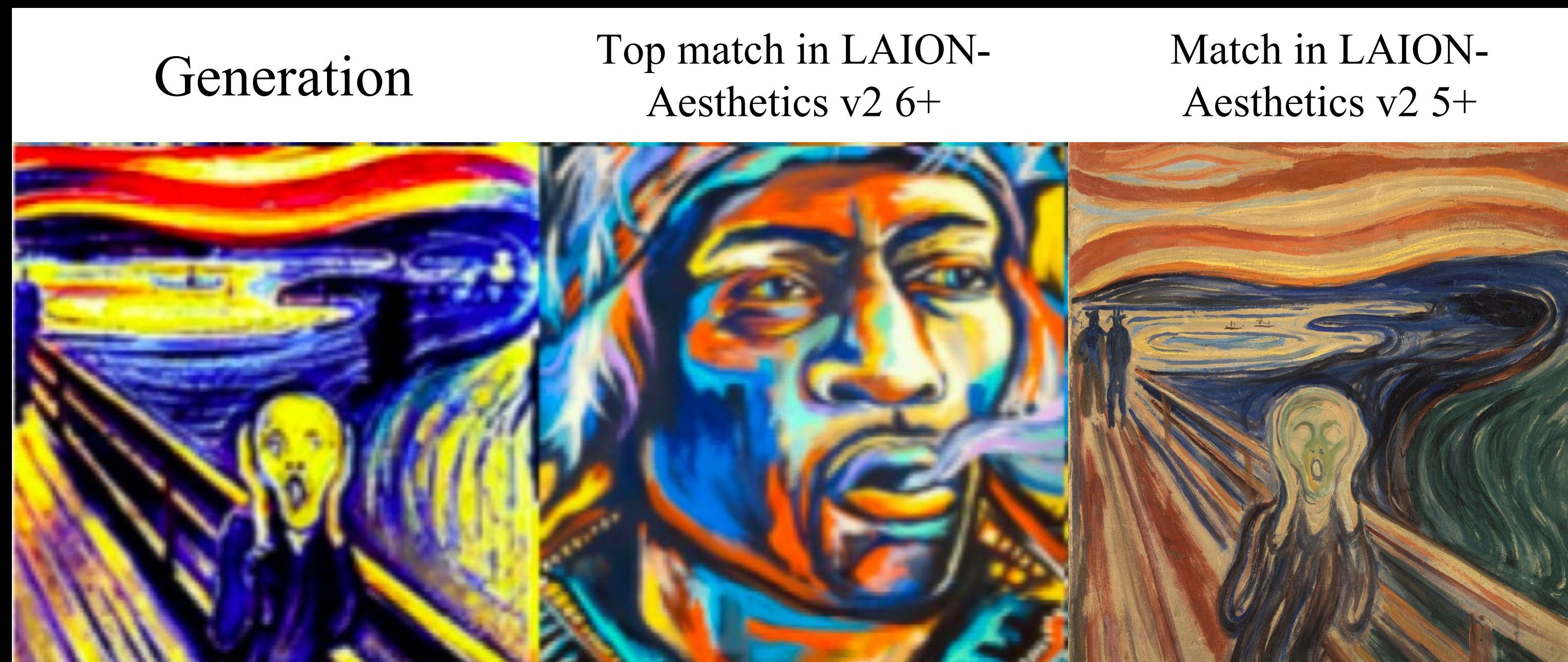
How good is this number?

- A lower bound? Since search happened over a subset.

| Generation | Top match in LAION-Aesthetics v2 6+ | Match in LAION-Aesthetics v2 5+ |

https://somepago.github.io/diffrep.html

How good is this number?

- ~~A lower bound? Since search happened over a subset.~~

  We analysed only a few captions.

https://somepago.github.io/diffrep.html

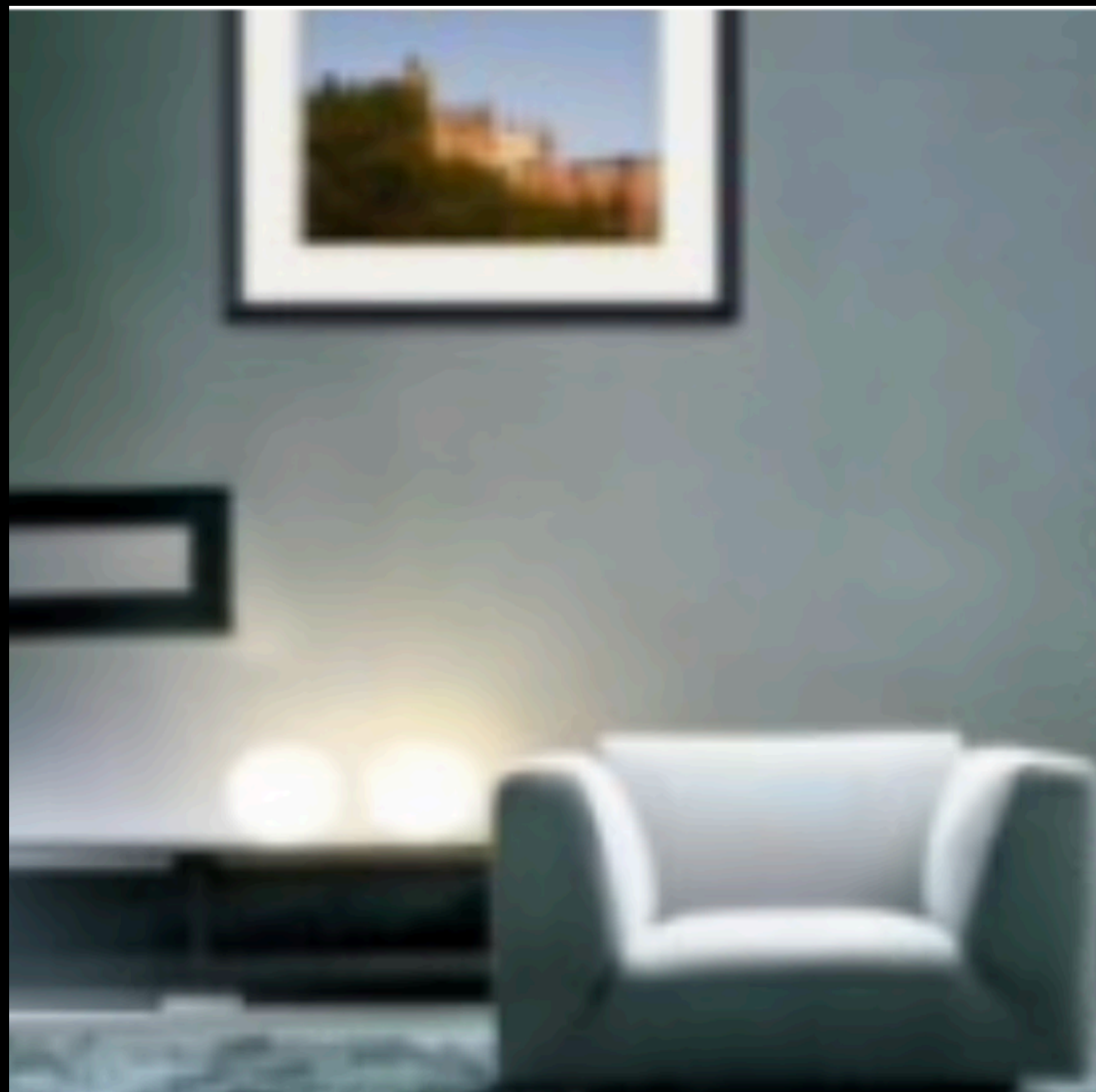# How good is this number?

- ~~A lower bound? Since search happened over a subset.~~

  We analysed only a few captions.

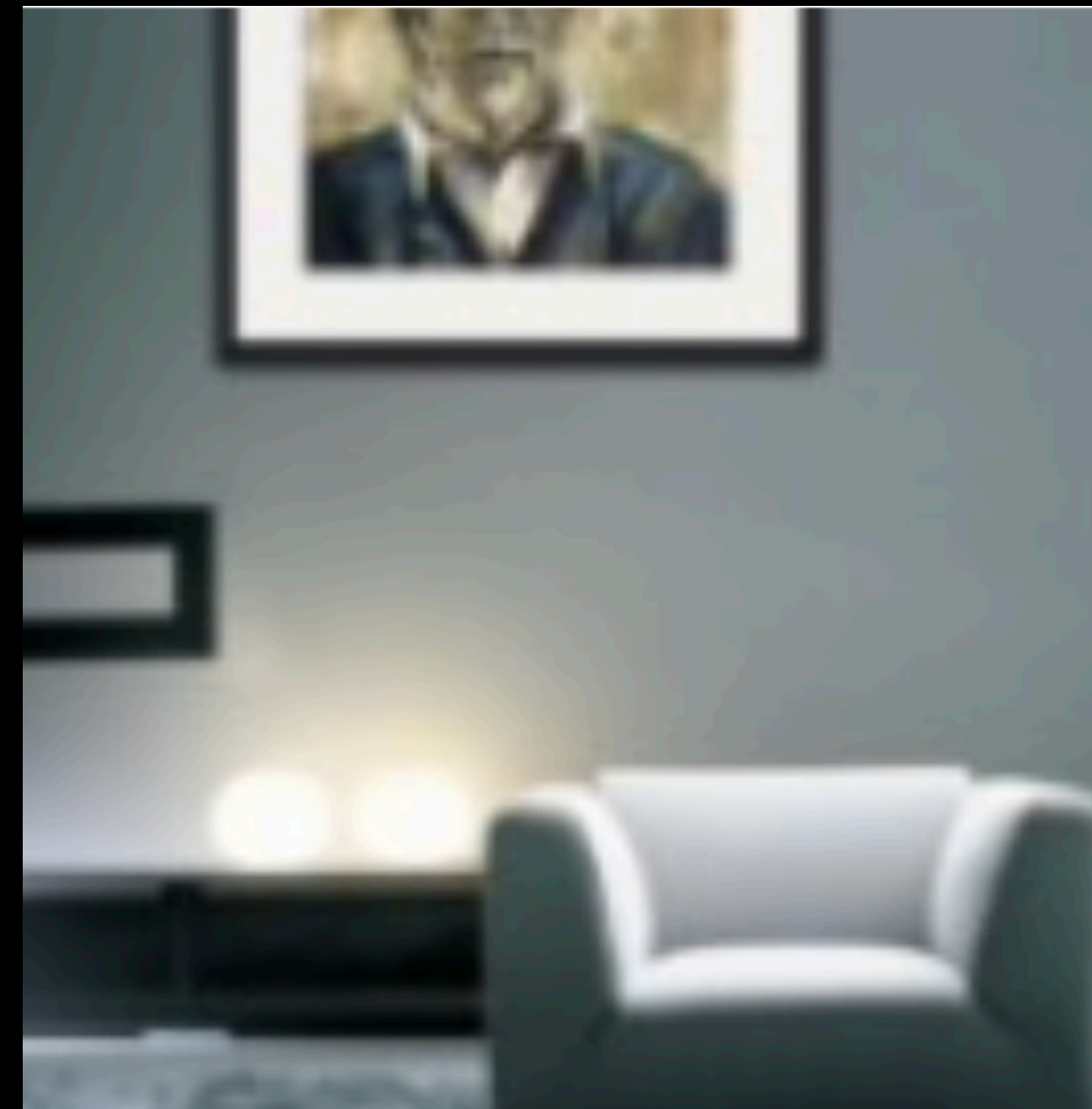- An upper bound? Since captions are sampled from training data.

30

Caption dependance - NONE

Generation

LAION-A match

"Hill Country Castle by R Del Angel"

"Ben Hogan Portrait Golf Legend&quot (2014) by GinetteCallaway"

https://somepago.github.io/diffrep.html

Caption dependance - NONE

Generation

LAION-A match

"New Orleans House Galaxy Case"

"iPhone Cover - Viking - EURO Trailer"

https://somepago.github.io/diffrep.html

Caption dependance - word/phrase match

Generation

LAION-A match

"Sony Boss Confirms Bloodborne Expansion is Coming"

"Bloodborne 32 Minute Speedrun"

https://somepago.github.io/diffrep.html

Caption dependance - word/phrase match

Prompt: <The description of the wall art> Canvas Wall Art Print

Prompt: A painting of the Great Wave off Kanagawa by Katsushika Hokusai.

How good is this number?

- A̶ l̶o̶w̶e̶r̶ b̶o̶u̶n̶d̶? S̶i̶n̶c̶e̶ s̶e̶a̶r̶c̶h̶ h̶a̶p̶p̶e̶n̶e̶d̶ o̶v̶e̶r̶ a̶ s̶u̶b̶s̶e̶t̶.

  We only analysed only a few captions.

- A̶n̶ u̶p̶p̶e̶r̶ b̶o̶u̶n̶d̶? S̶i̶n̶c̶e̶ c̶a̶p̶t̶i̶o̶n̶s̶ a̶r̶e̶ s̶a̶m̶p̶l̶e̶d̶ f̶r̶o̶m̶ t̶r̶a̶i̶n̶i̶n̶g̶ d̶a̶t̶a̶.

  Certain words and phrases can trigger copying, we do not need whole caption!

https://somepago.github.io/diffrep.html

Possible causes of replication

* Data duplication

* Text conditioning

https://somepago.github.io/diffrep.html

Replication spectrum

Content and style copied   Style copied

Generation

Top Match

https://somepago.github.io/diffrep.html

Thank you!

Project page: https://somepago.github.io/diffrep.html