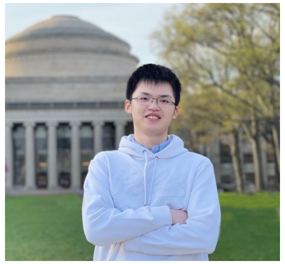
Return of Unconditional Generation:
A Self-supervised Representation Generation Method

Tianhong Li

Joint work with Dina Katabi and Kaiming He



Conditional Generation Prevails

Conditional Generation Prevails

Parrot

Conditional Generation Prevails

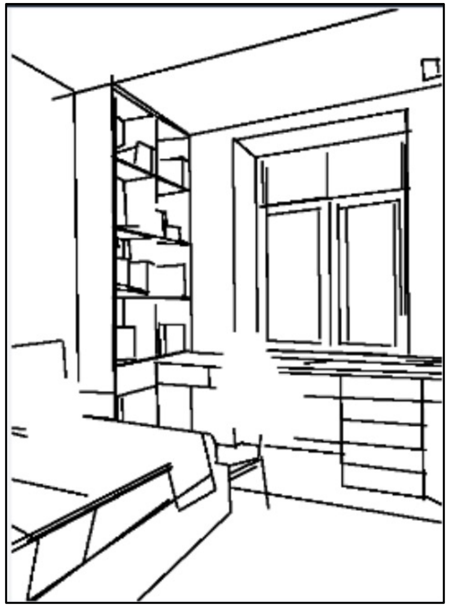
Parrot

"A rabbit playing violin"

Conditional Generation Prevails

Parrot

"A rabbit playing violin"



ControlNet

Unconditional Generation Lags Behind

Unconditional Generation Lags Behind

Otter

Ballon

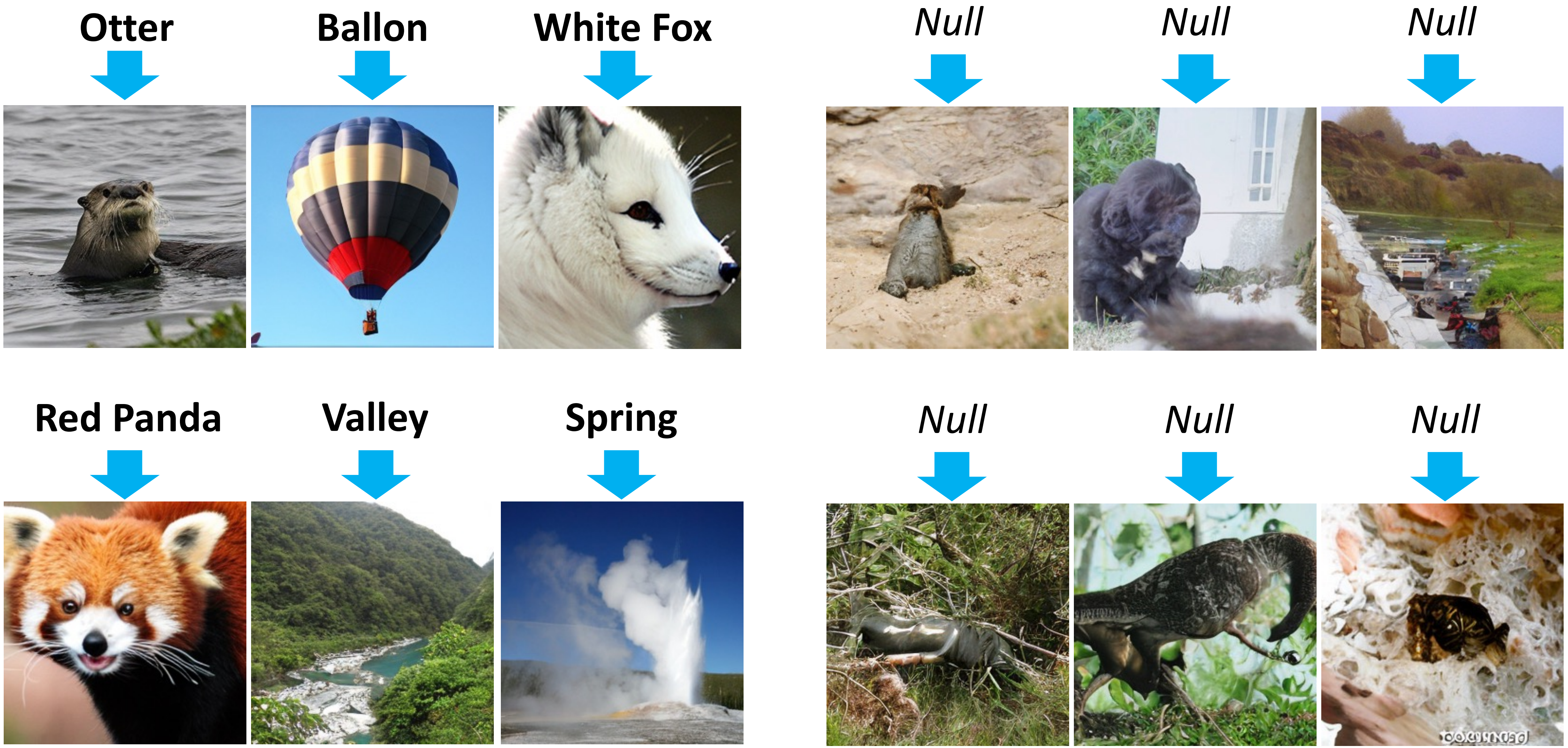
White Fox

Red Panda

Valley

Spring

Unconditional Generation Lags Behind



Even SD3…

“A beautiful mushroom”

“A cartoon woman”

“A rabbit playing violin”

“A cat holding a sign that says hello world”

“A tiger wearing t-shirt on the street”

“Two hands in heart shape with ‘love’ in it”

Even SD3…

“A beautiful mushroom”

“A cartoon woman”

“A rabbit playing violin”

“ ”

“ ”

“ ”

“A cat holding a sign that says hello world”

“A tiger wearing t-shirt on the street”

“Two hands in heart shape with ‘love’ in it”

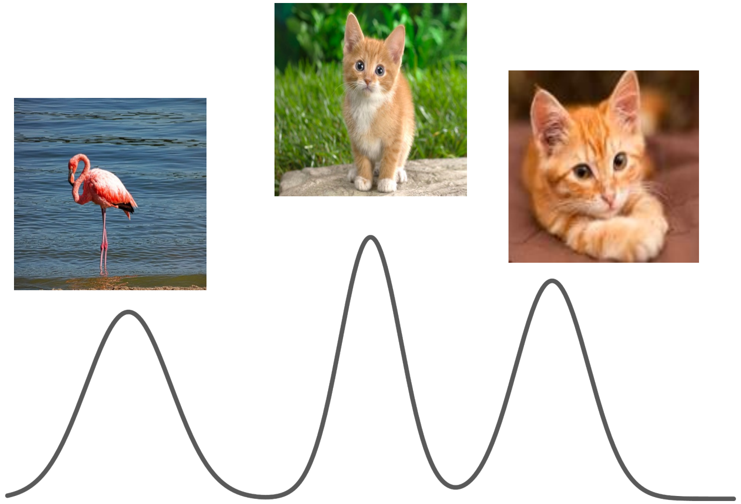
“ ”

“ ”

“ ”

Unconditional is Harder than Conditional

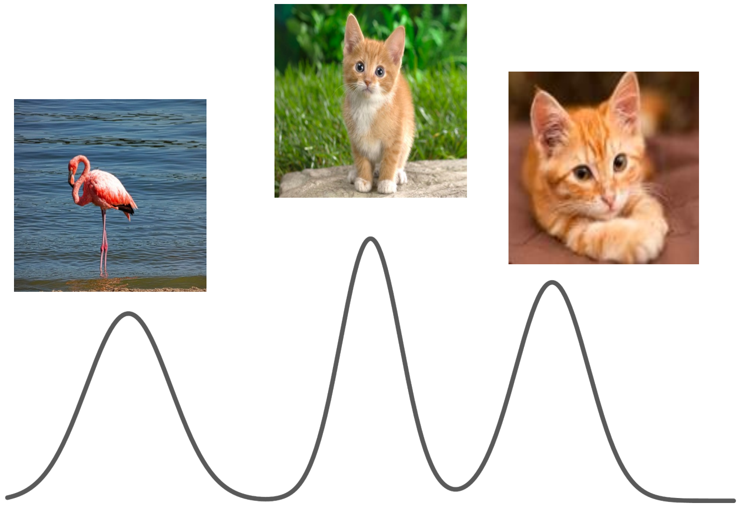
Unconditional is Harder than Conditional

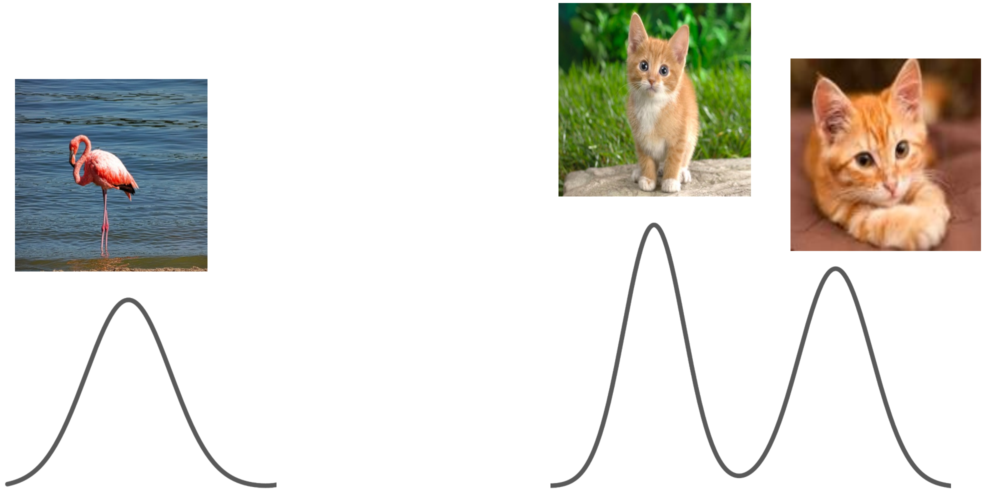


$P(X)$

Hard

Unconditional is Harder than Conditional





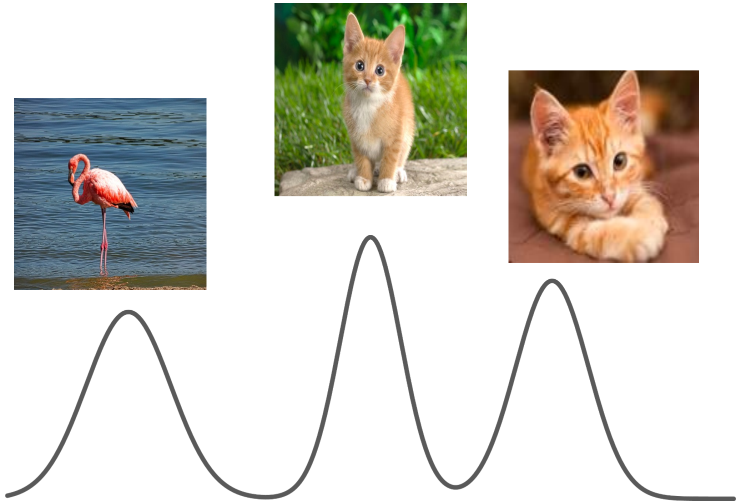
$$P(X) = P(X|\mathrm{bird}) \cdot P(\mathrm{bird}) + P(X|\mathrm{cat}) \cdot P(\mathrm{cat})$$

Hard

Easy

Easy

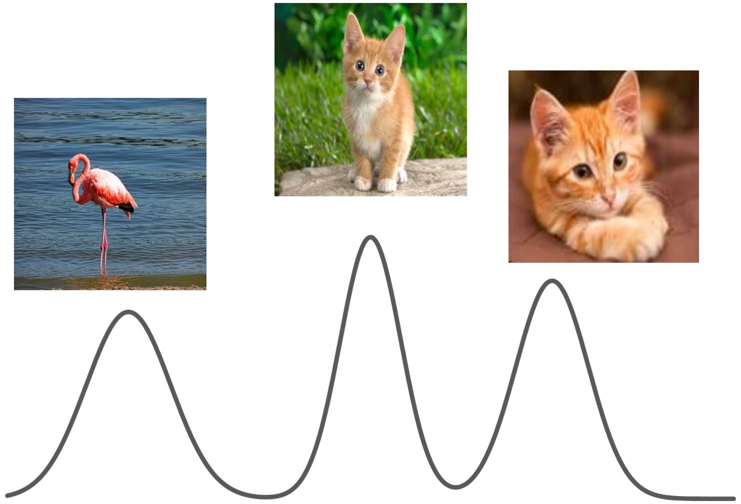
Decompose Hard Distribution!



More Generally, for any function f:

$$P(X) = P(X|f(X)) \cdot P(f(X))$$

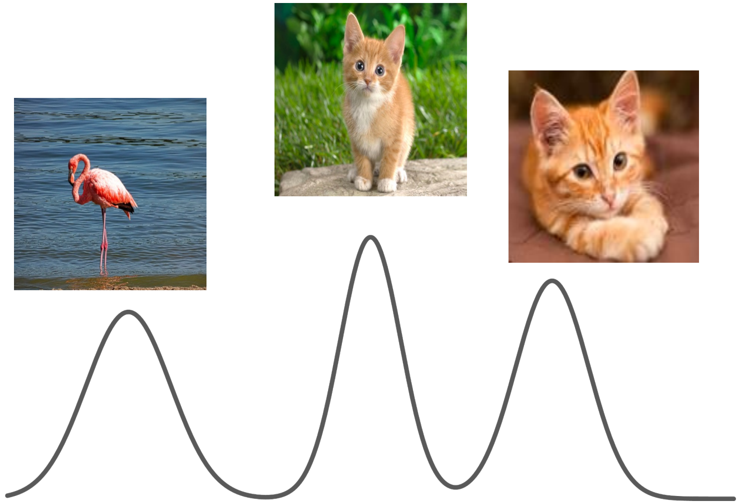
Decompose Hard Distribution!



More Generally, for any function f:

$$P(X) = P(X|f(X)) \cdot P(f(X))$$

Decompose Hard Distribution!

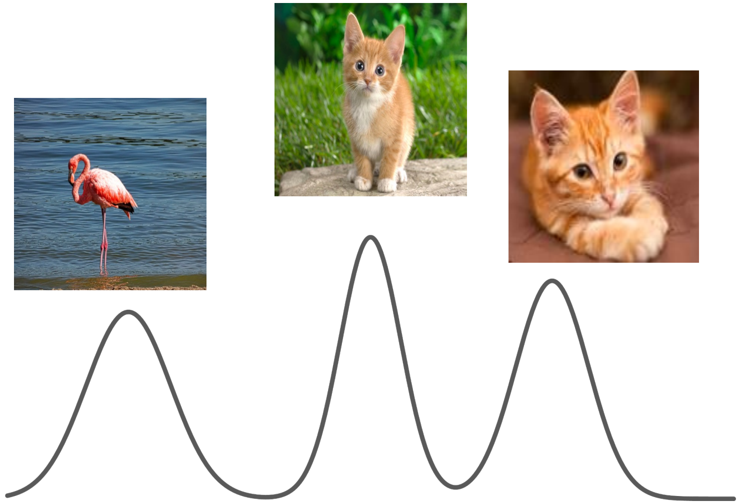


More Generally, for any function f:

$$P(X) = P(X|f(X)) \cdot P(f(X))$$

- $P(f(X))$ should be easy to model

Decompose Hard Distribution!

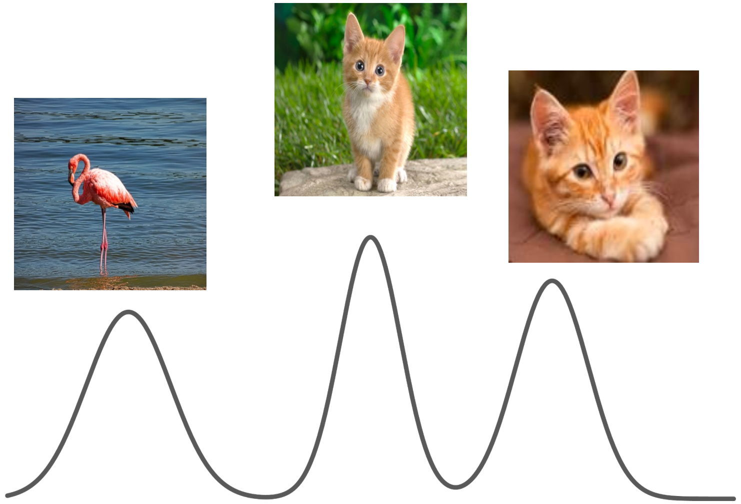


More Generally, for any function f:

$$P(X) = P(X|f(X)) \cdot P(f(X))$$

- $P(f(X))$ should be easy to model
- $f(X)$ should provide rich semantics

Decompose Hard Distribution!



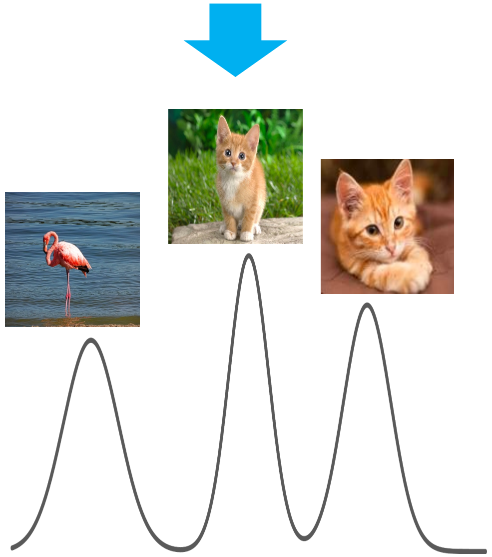
More Generally, for any function f:

$$P(X) \quad = \quad P(X|{\color{red}f(X)}) \cdot P({\color{red}f(X)})$$

- $P(f(X))$ should be easy to model
- $f(X)$ should provide rich semantics
- f should be unsupervised for unconditional generation

Representation-Conditioned Generation (RCG)

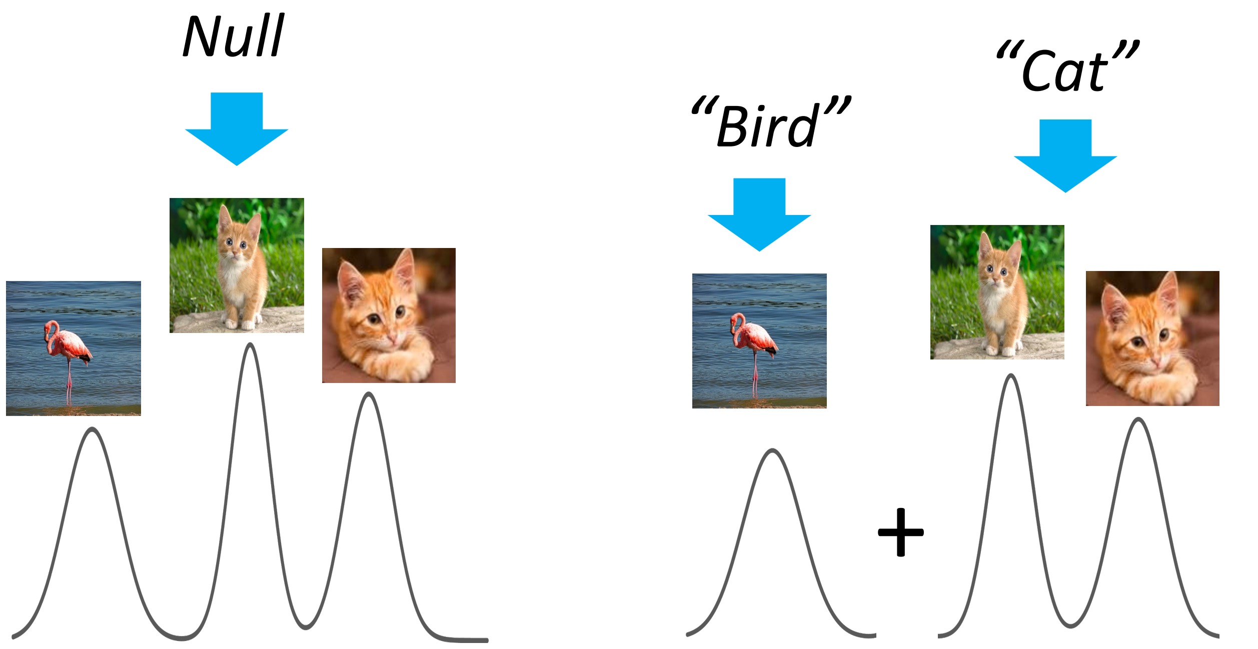
Null



Unconditional

- Unsupervised
- Too complex
- Bad performance

Representation-Conditioned Generation (RCG)



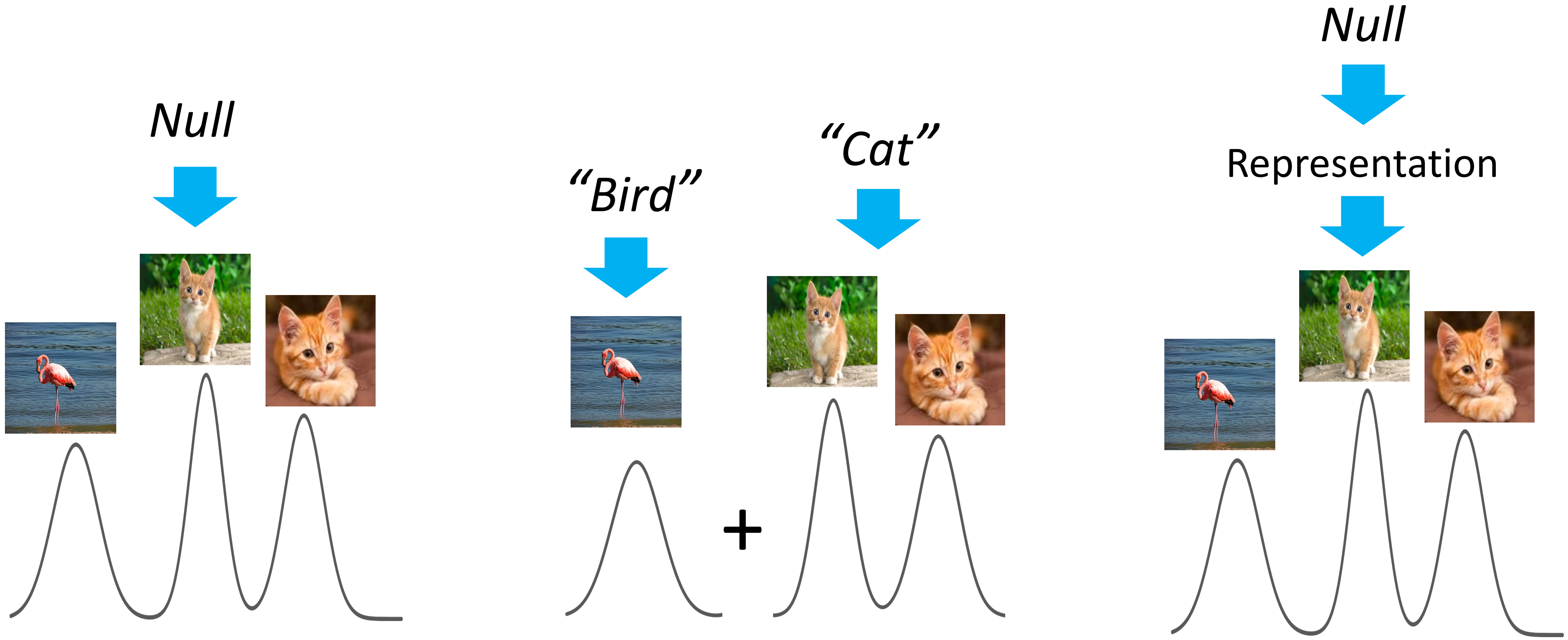
Unconditional

- Unsupervised
- Too complex
- Bad performance

Conditional

- Require labels
- Easy to model
- Good performance

Representation-Conditioned Generation (RCG)



Unconditional

- Unsupervised
- Too complex
- Bad performance

Conditional

- Require labels
- Easy to model
- Good performance

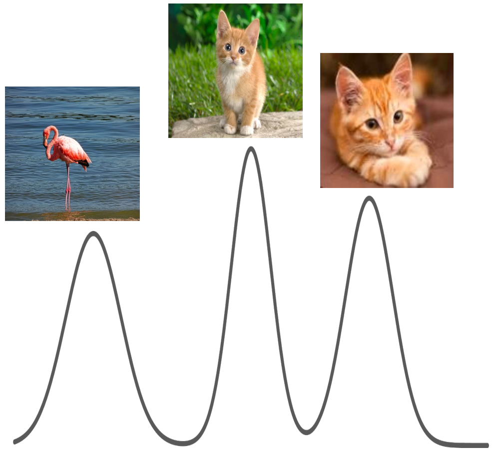
Rep. Conditioned

- Unsupervised
- Easy to model
- Good performance

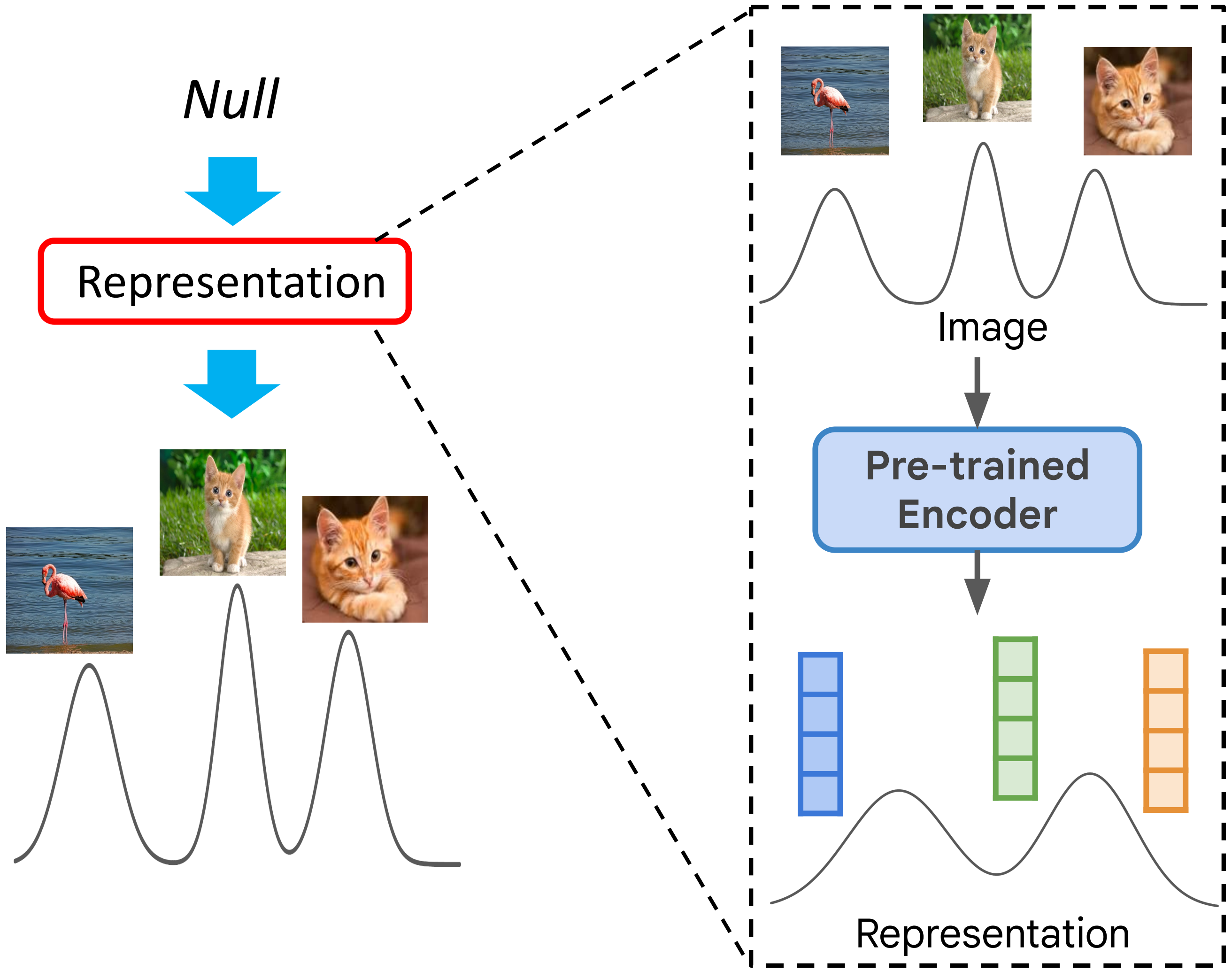
Representation-Conditioned Generation (RCG)

Null

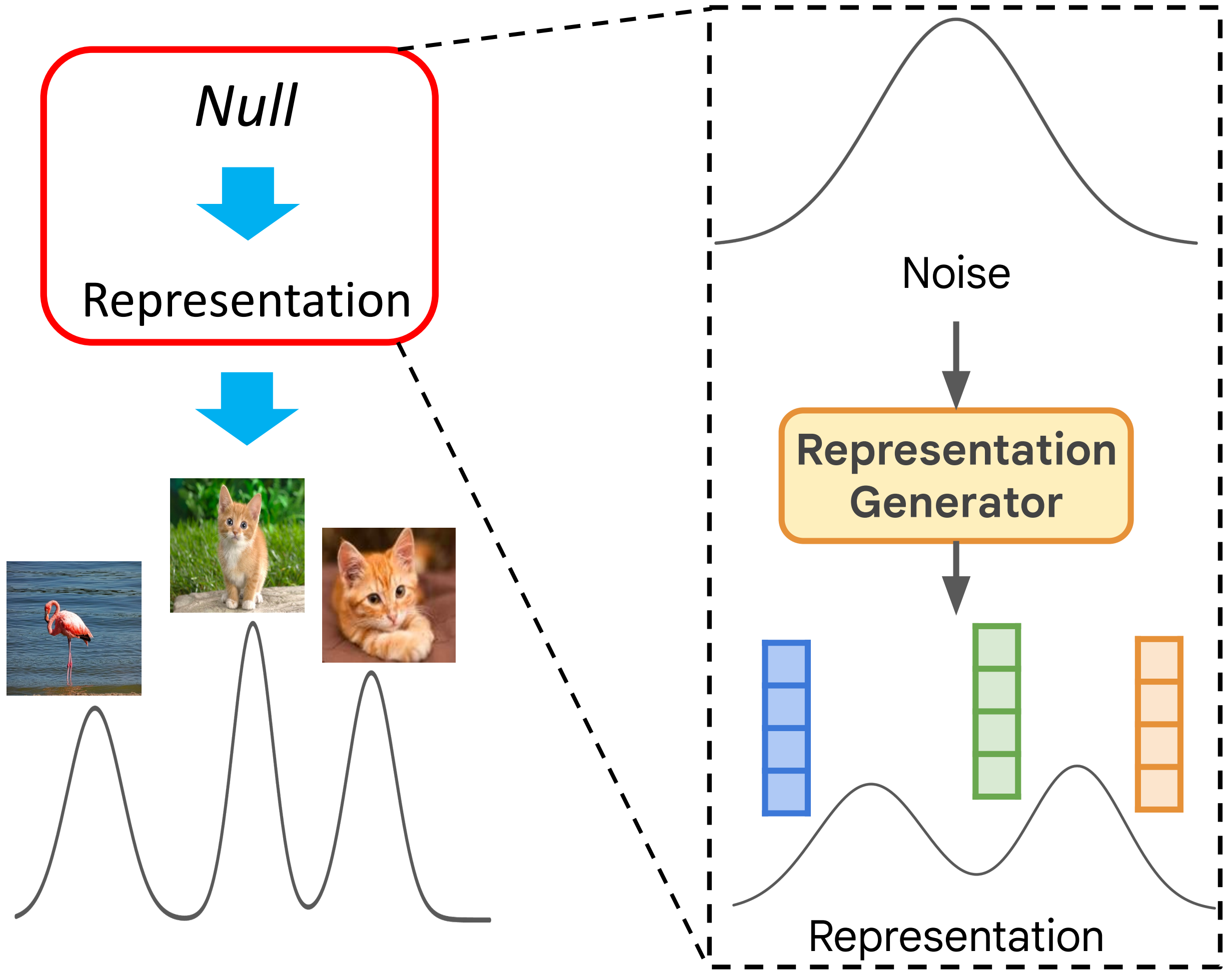
Representation



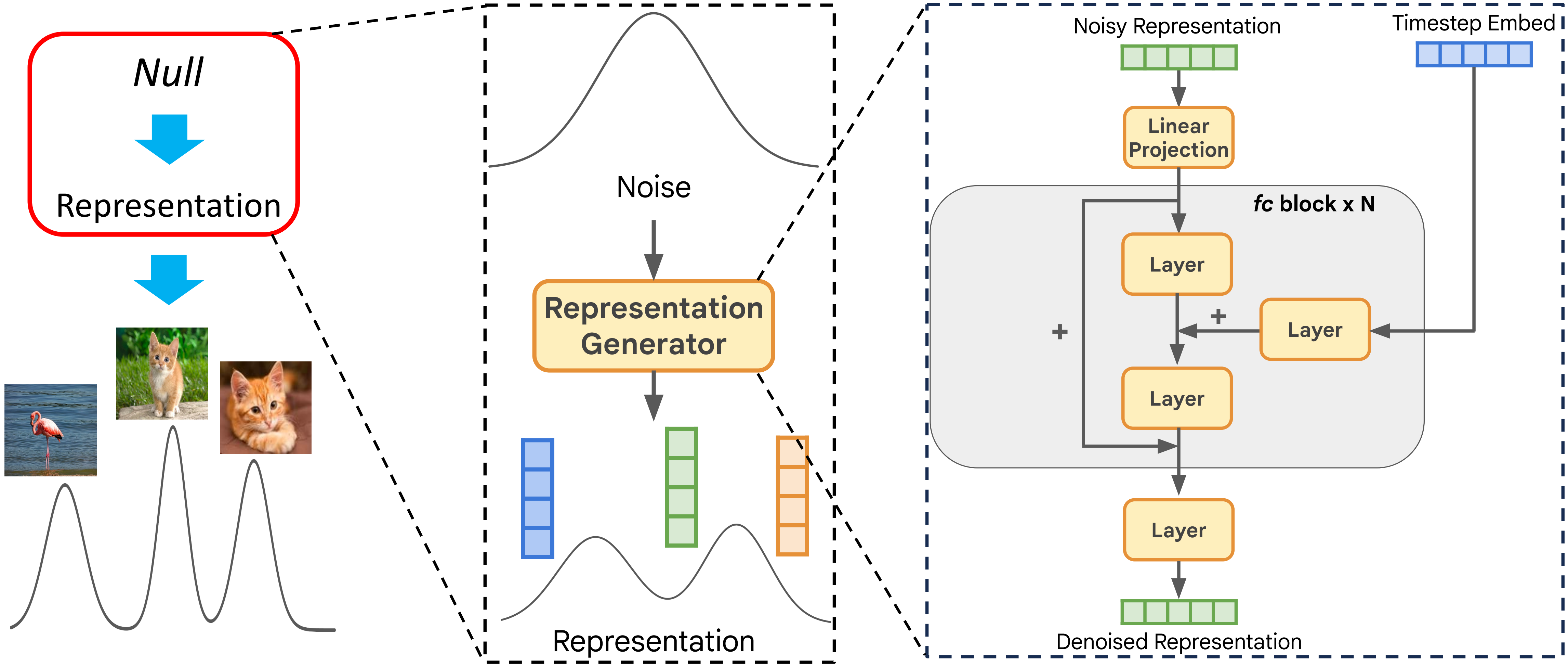
Representation Extraction



Representation Generation



Representation Generation



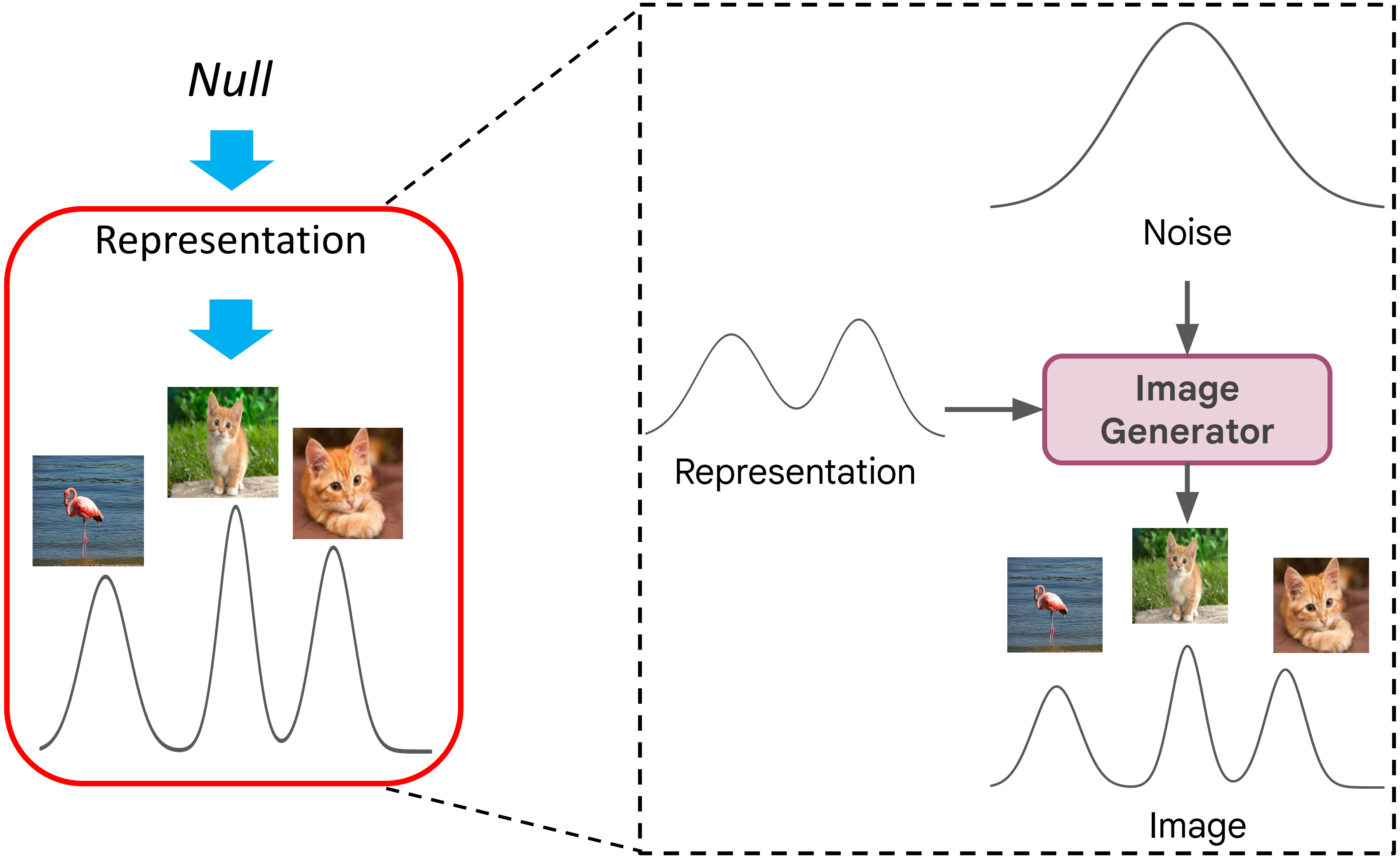
Representation Generation

#Blocks	rep FD↓
3	0.71
6	0.53
12	**0.48**
18	0.50
24	0.49

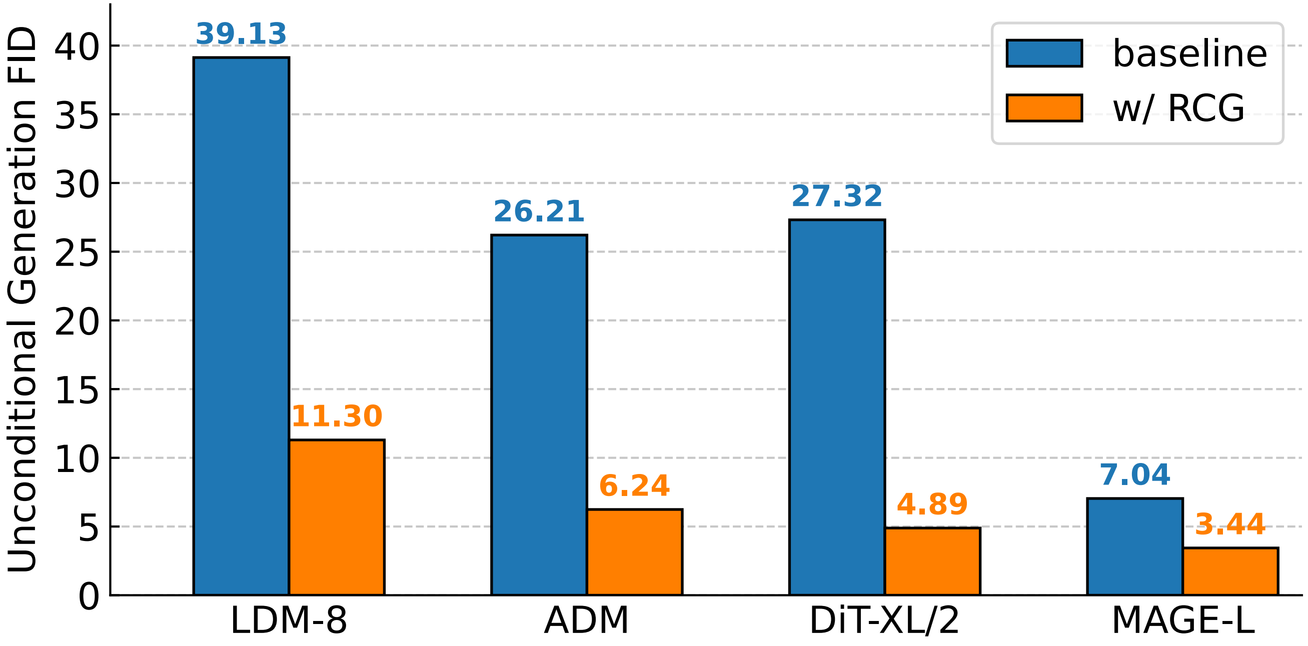
Hidden Dim	rep FD↓
256	5.98
512	1.19
1024	0.56
1536	**0.48**
2048	0.48

- Light-weight model (12 blocks, 1536 channels)
- Accurate representation generation

Representation-conditioned Image Generation

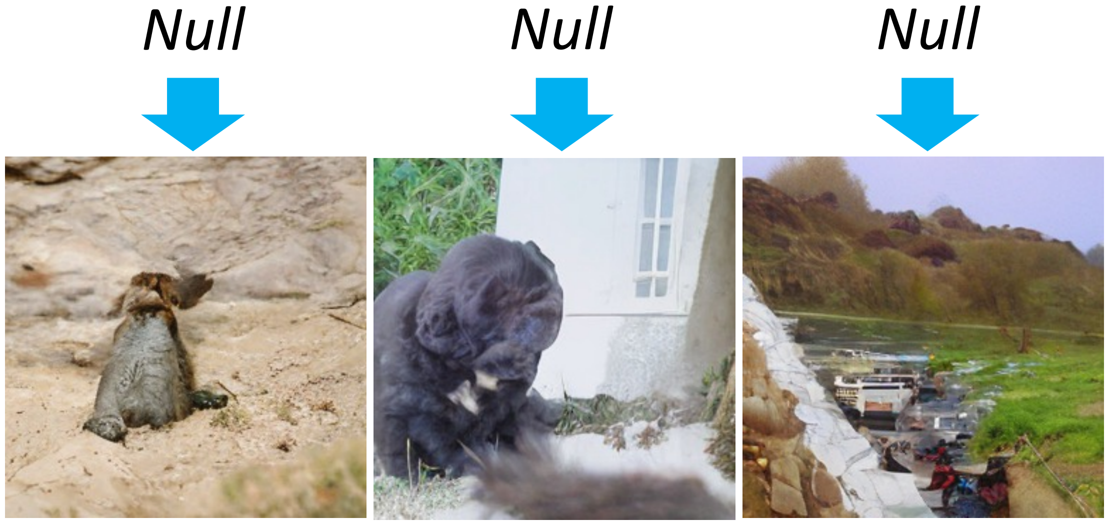


Representation-conditioned Image Generation

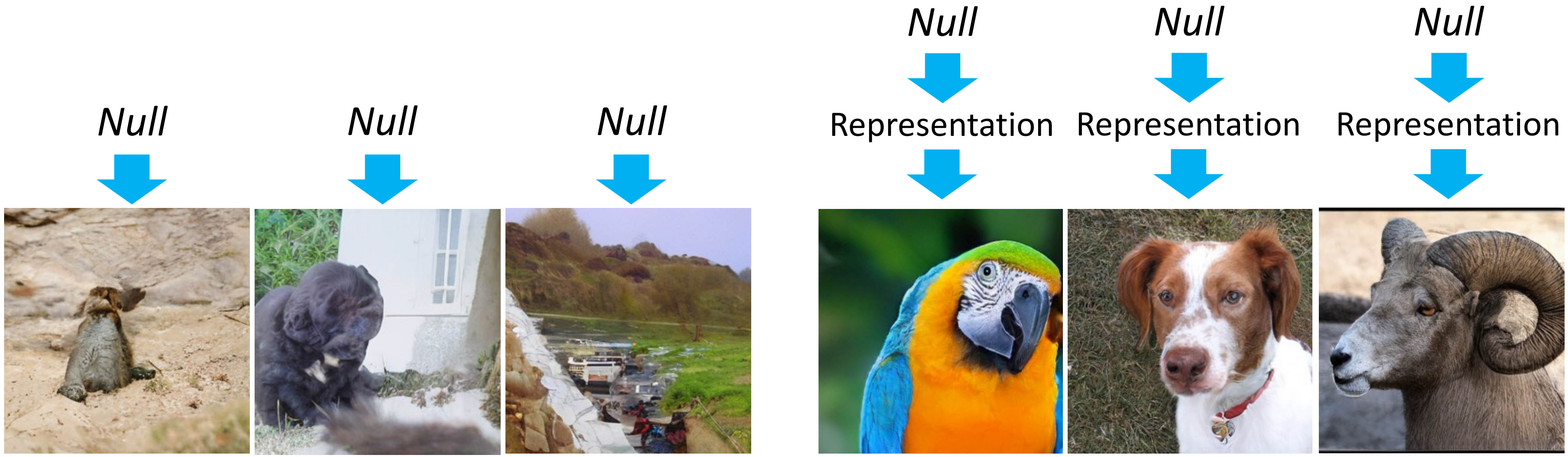


- RCG consistently improves different image generators

Representation-conditioned Image Generation



Representation-conditioned Image Generation



New SOTA in Unconditional Generation

New SOTA in Unconditional Generation

Unconditional generation	#params	FID↓	IS↑
BigGAN [19]	$\sim$70M	38.61	24.7
ADM [18]	554M	26.21	39.7
MaskGIT [10]	227M	20.72	42.1

- SOTA models are poor at unconditional generation

New SOTA in Unconditional Generation

Unconditional generation	#params	FID↓	IS↑
BigGAN [19]	~70M	38.61	24.7
ADM [18]	554M	26.21	39.7
MaskGIT [10]	227M	20.72	42.1
RCDM† [5]	-	19.0	51.9
IC-GAN† [9]	~75M	15.6	59.0
ADDP [61]	176M	8.9	95.3
MAGE-B [41]	176M	8.67	94.8
MAGE-L [41]	439M	7.04	123.5
RDM-IN† [4]	400M	5.91	158.8

- Most prior works focus on retrieval-based generation which require ground-truth images during generation

New SOTA in Unconditional Generation

Unconditional generation	#params	FID↓	IS↑
BigGAN [19]	∼70M	38.61	24.7
ADM [18]	554M	26.21	39.7
MaskGIT [10]	227M	20.72	42.1
RCDM† [5]	-	19.0	51.9
IC-GAN† [9]	∼75M	15.6	59.0
ADDP [61]	176M	8.9	95.3
MAGE-B [41]	176M	8.67	94.8
MAGE-L [41]	439M	7.04	123.5
RDM-IN† [4]	400M	5.91	158.8
RCG (MAGE-B)	239M	3.98	177.8
RCG (MAGE-L)	502M	3.44	186.9

- RCG largely improves SOTA

New SOTA in Unconditional Generation

Unconditional generation	#params	FID↓	IS↑
BigGAN [19]	∼70M	38.61	24.7
ADM [18]	554M	26.21	39.7
MaskGIT [10]	227M	20.72	42.1
RCDM† [5]	-	19.0	51.9
IC-GAN† [9]	∼75M	15.6	59.0
ADDP [61]	176M	8.9	95.3
MAGE-B [41]	176M	8.67	94.8
MAGE-L [41]	439M	7.04	123.5
RDM-IN† [4]	400M	5.91	158.8
RCG (MAGE-B)	239M	3.98	177.8
RCG (MAGE-L)	502M	3.44	186.9
RCG-G (MAGE-B)	239M	3.19	212.6
RCG-G (MAGE-L)	502M	**2.15**	**253.4**

- RCG further rivals SOTA class-conditional generation

What’s in the Representation Space?


teddy bear


valley


Tibetan mastiff


tench

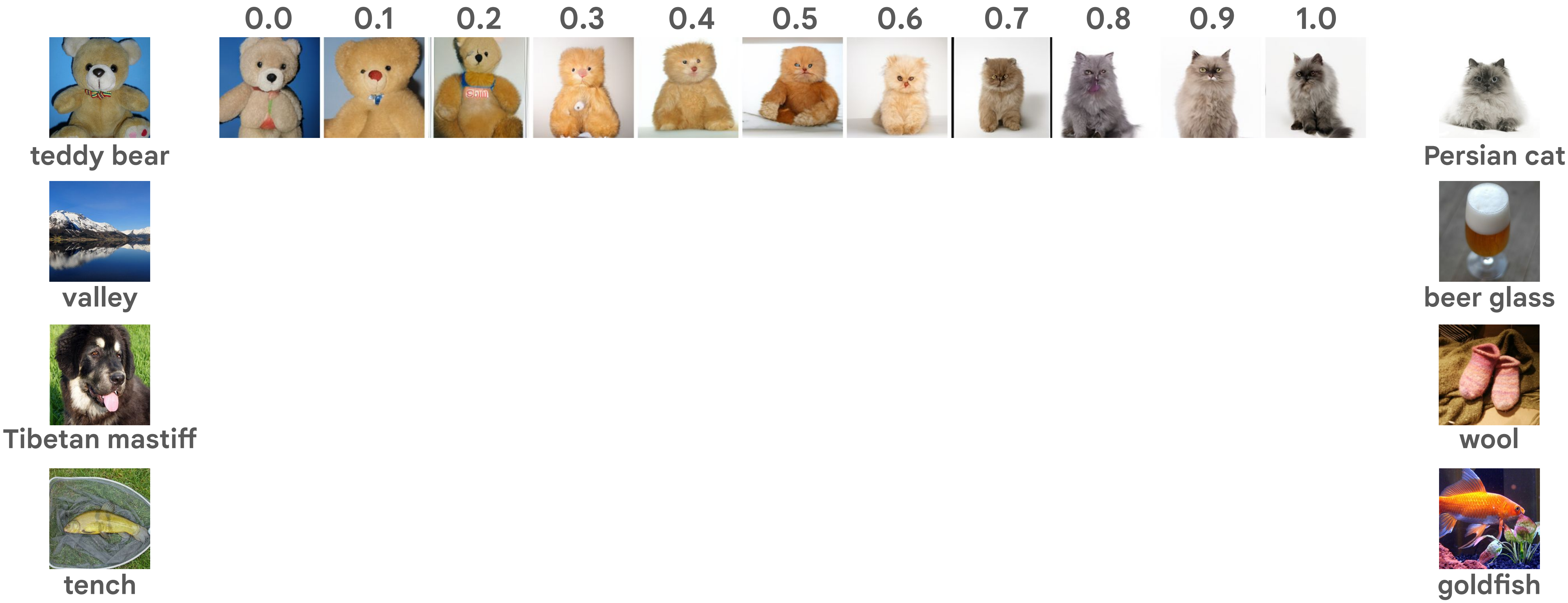

Persian cat


beer glass

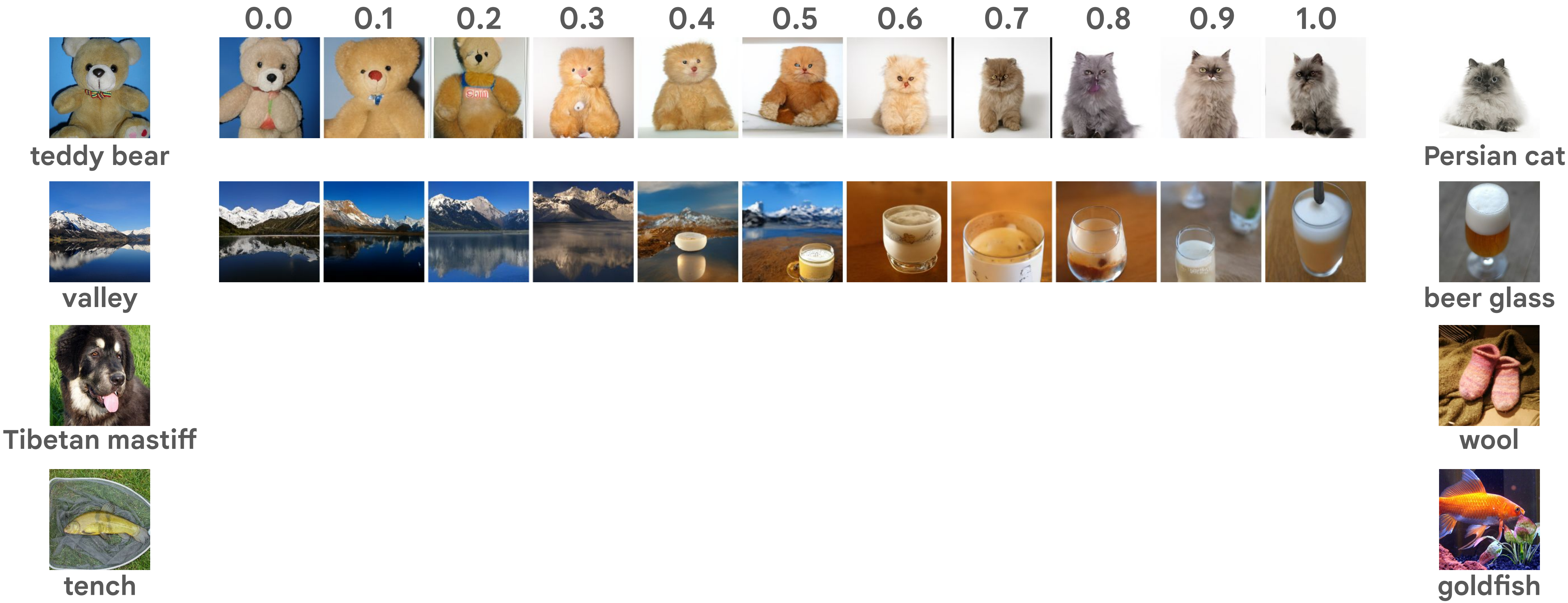

wool


goldfish

What’s in the Representation Space?

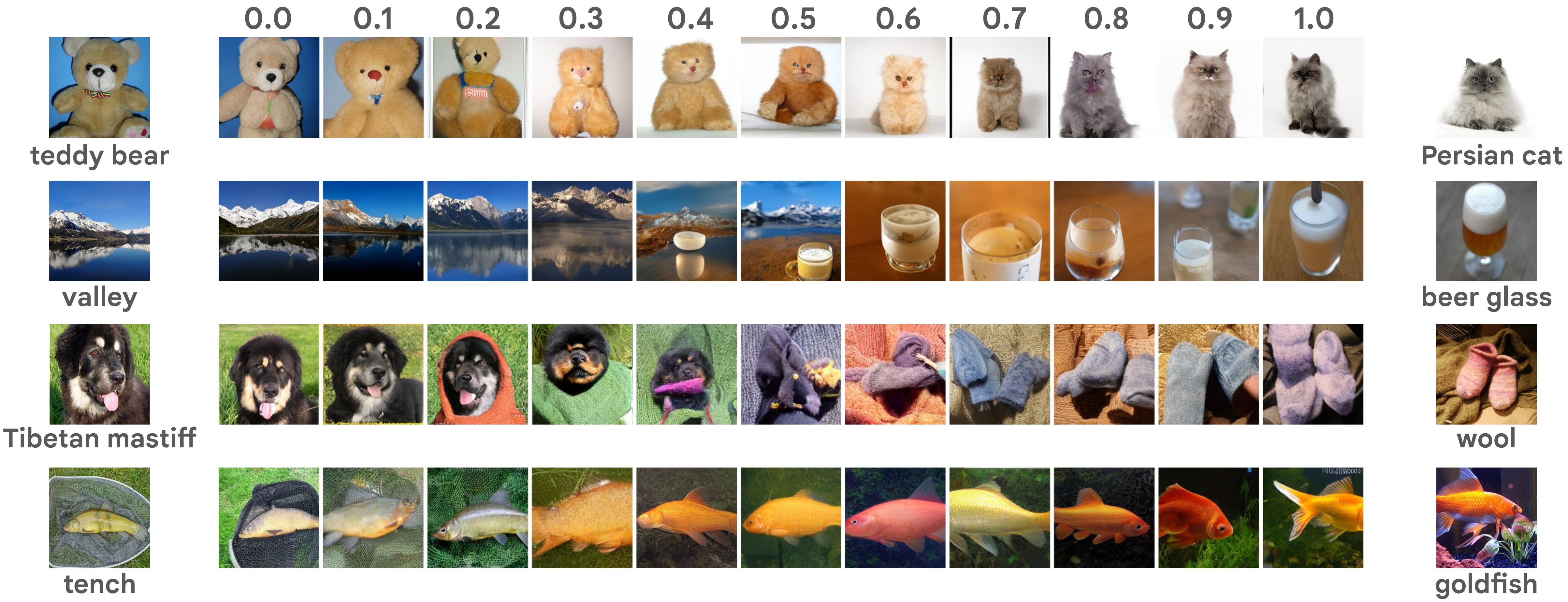


What's in the Representation Space?



What’s in the Representation Space?

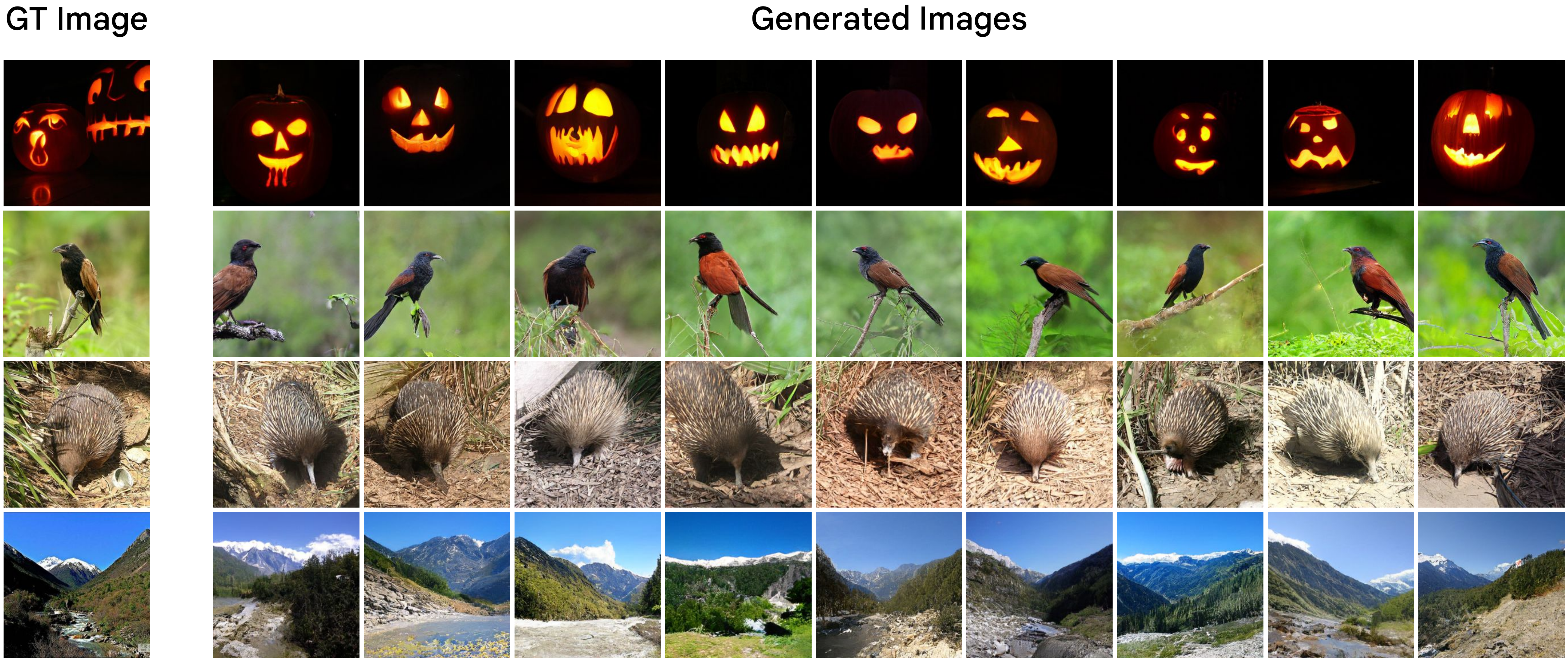
What’s in the Representation Space?



RCG to Diversify Real Images

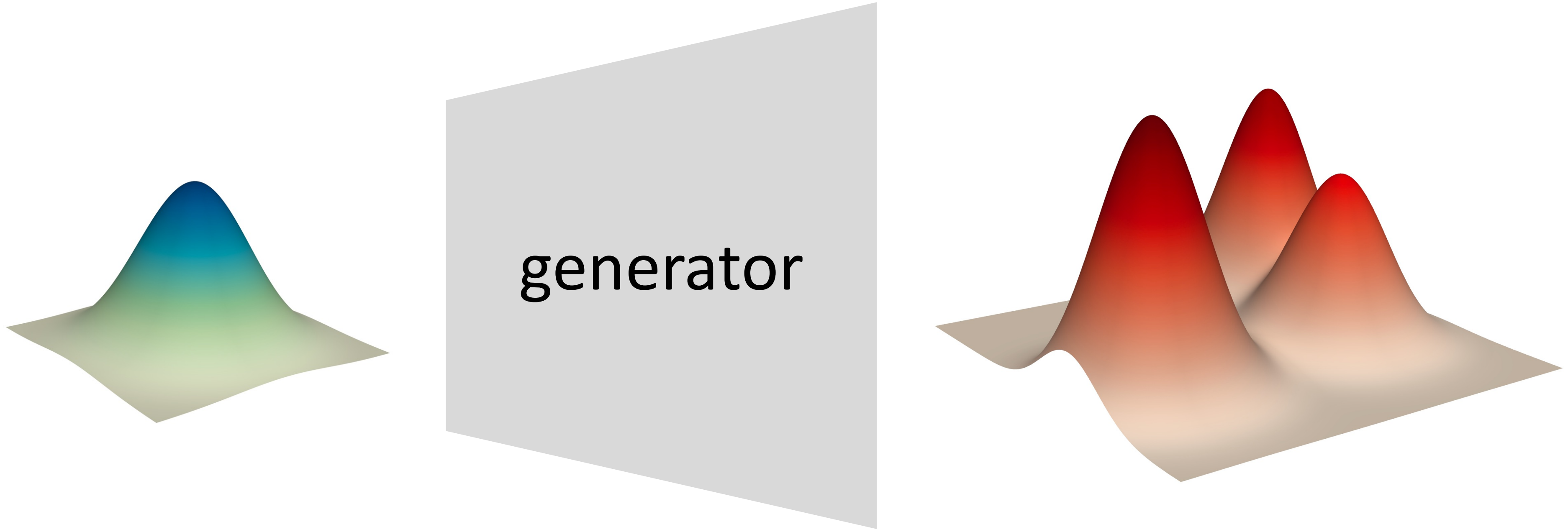
GT Image

RCG to Diversify Real Images

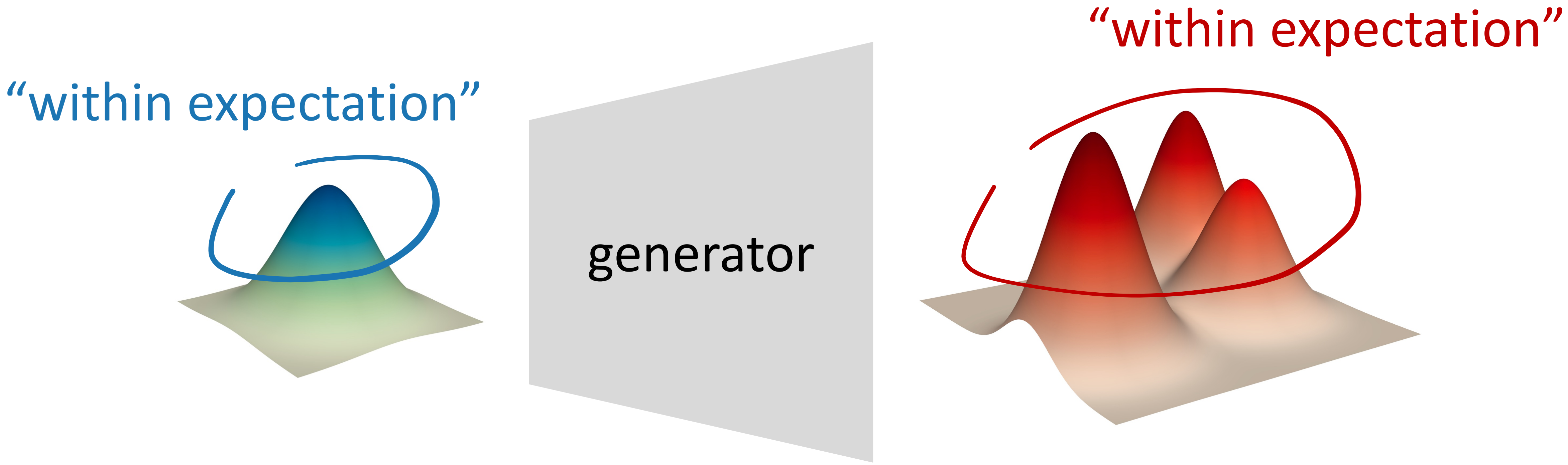


Return of Unconditional Generation

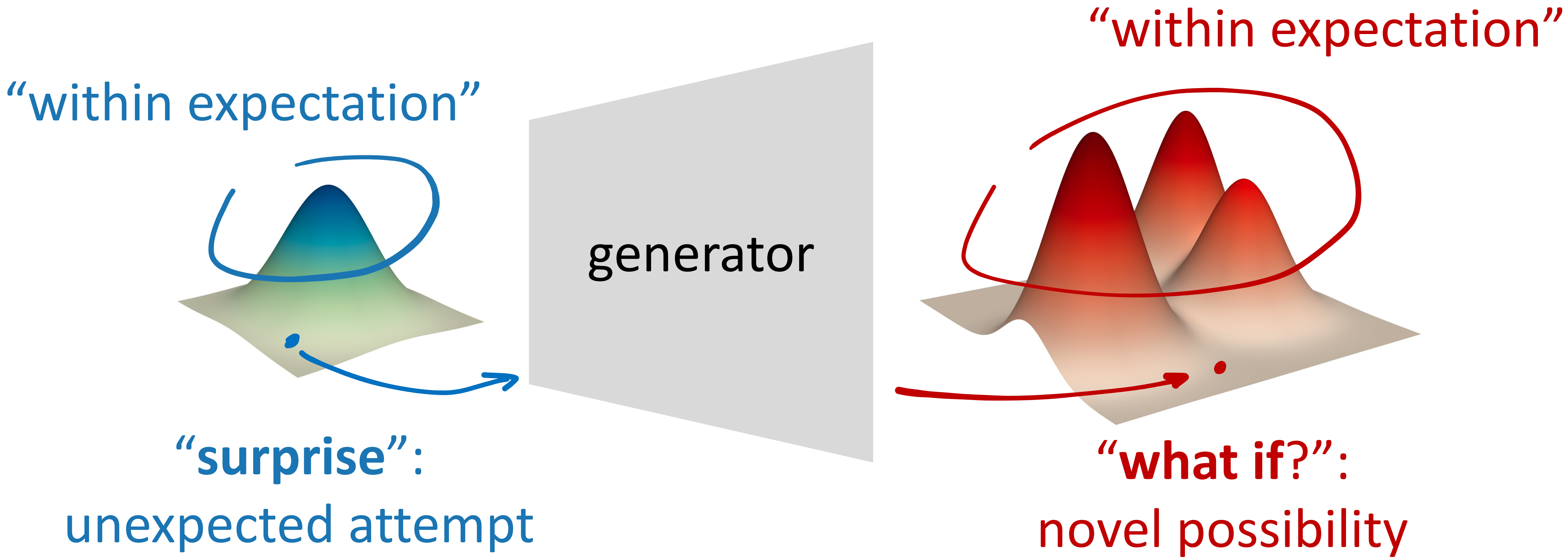
Return of Unconditional Generation



Return of Unconditional Generation



Return of Unconditional Generation



Takeaways

- Unconditional generation is behind, but it matters
- RCG: decompose distribution and generate representation
- Many new possibilities with unconditional generation!
- Codes are available at https://github.com/LTH14/rcg.
- Poster: Friday afternoon East Exhibit Hall A-C #1603
- Also check our other Spotlight paper MAR: Thursday noon East Exhibit Hall A-C #1505