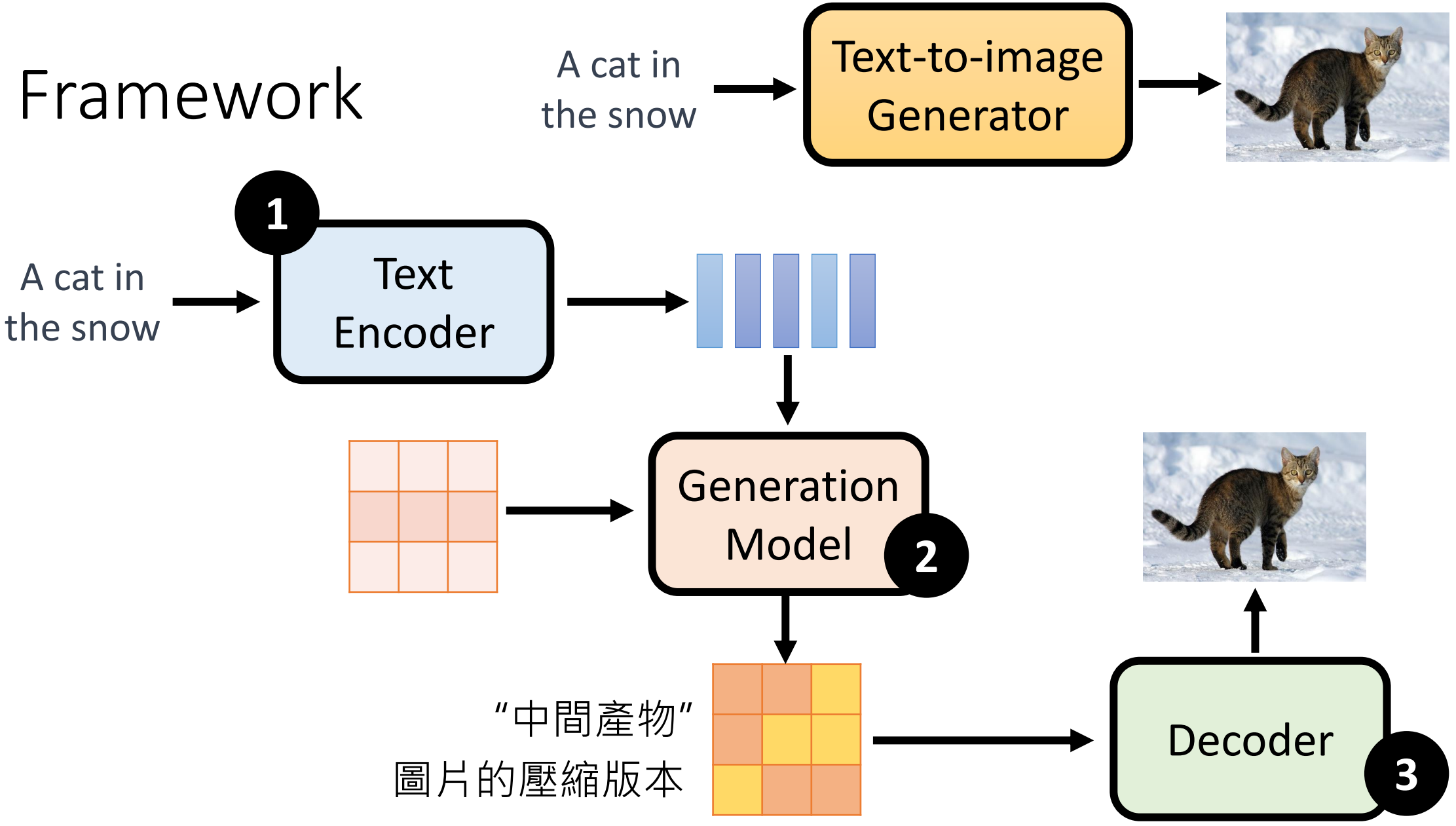
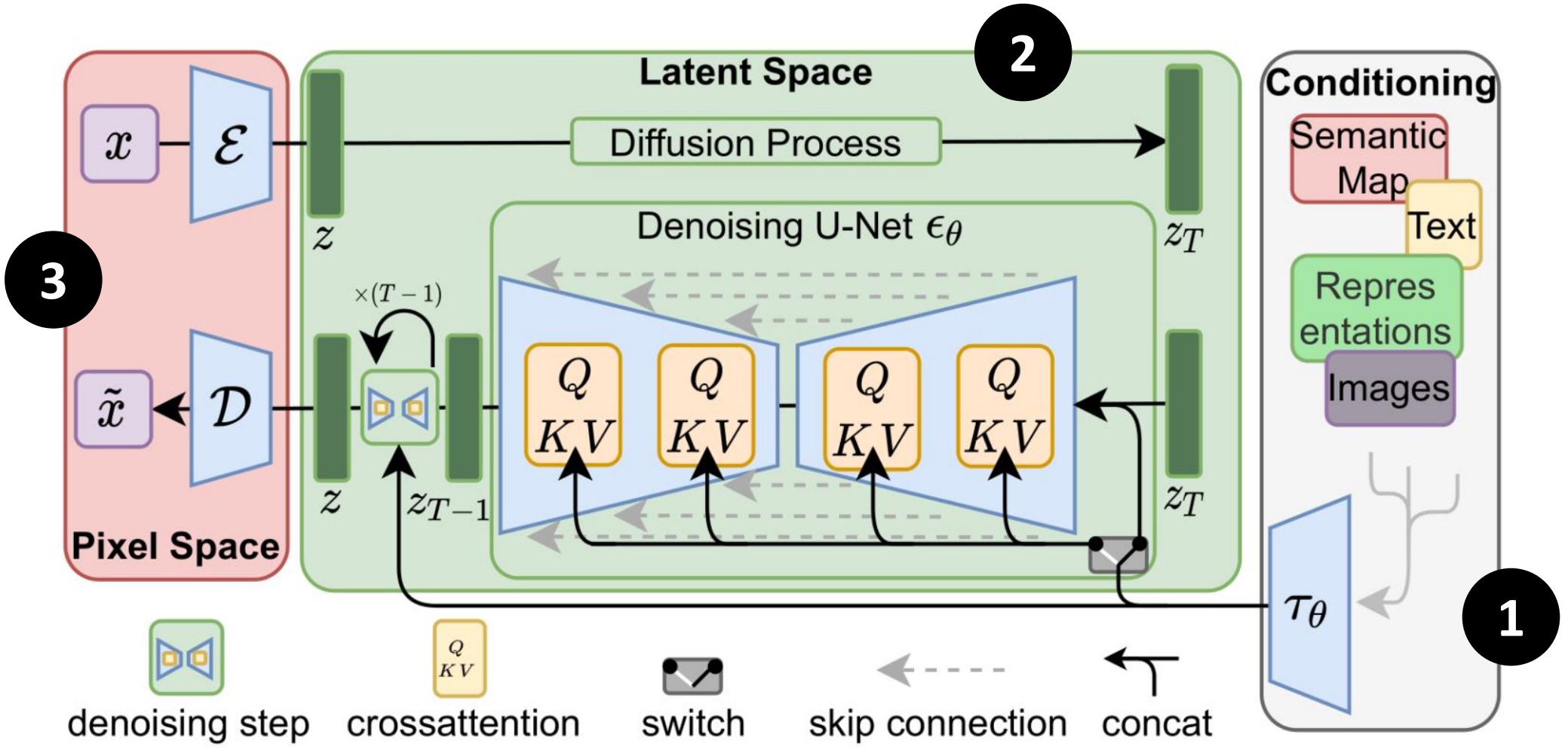
Stable Diffusion

Framework


A cat in the snow
Text-to-image Generator
A cat in the snow
1
Text Encoder
Generation Model
2
"中間產物"
圖片的壓縮版本
Decoder
3

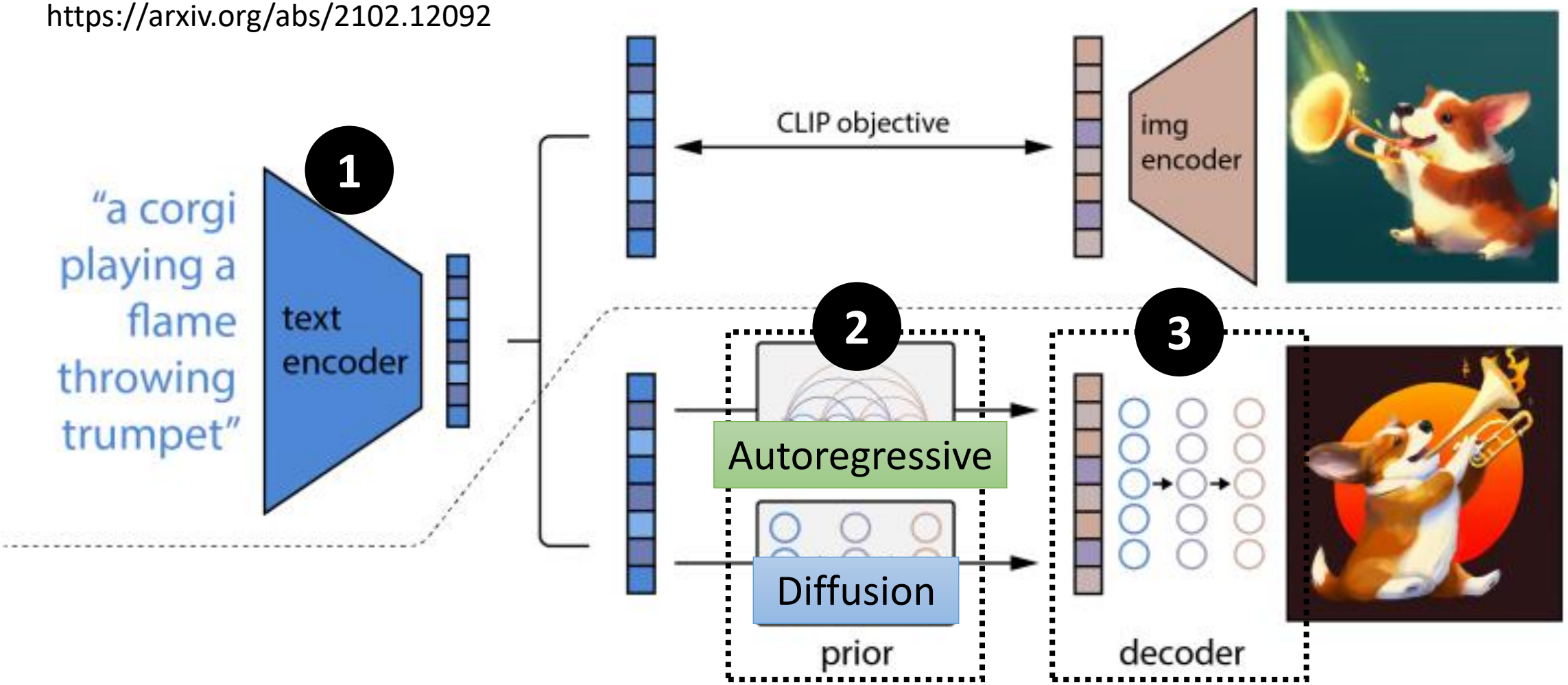
Stable Diffusion

https://arxiv.org/abs/2112.10752



DALL-E series

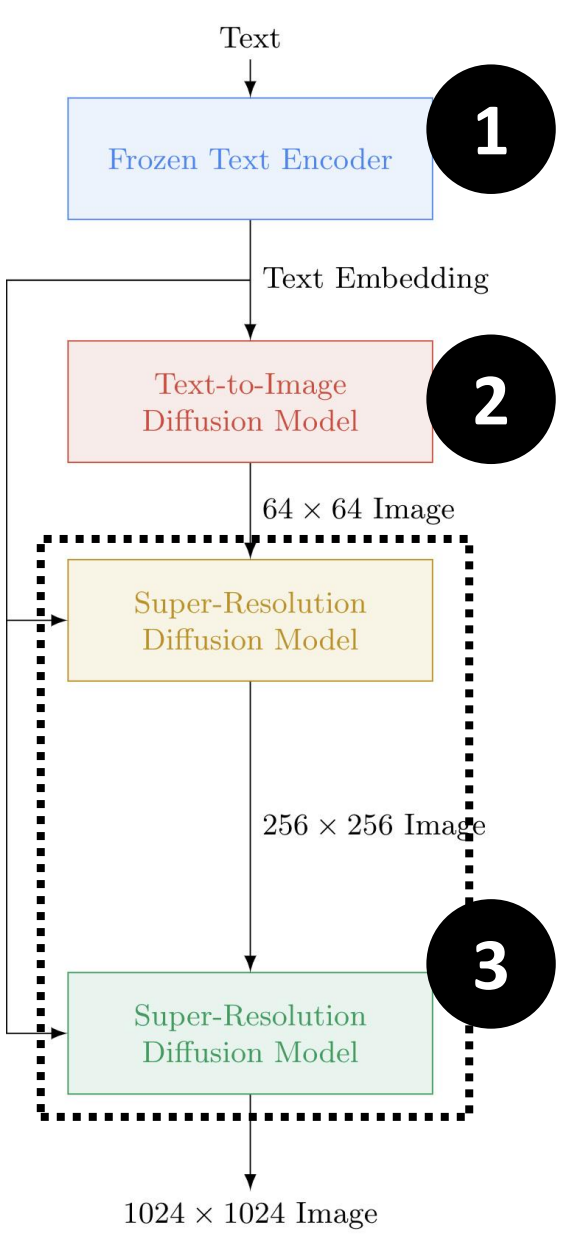
https://arxiv.org/abs/2204.06125
https://arxiv.org/abs/2102.12092

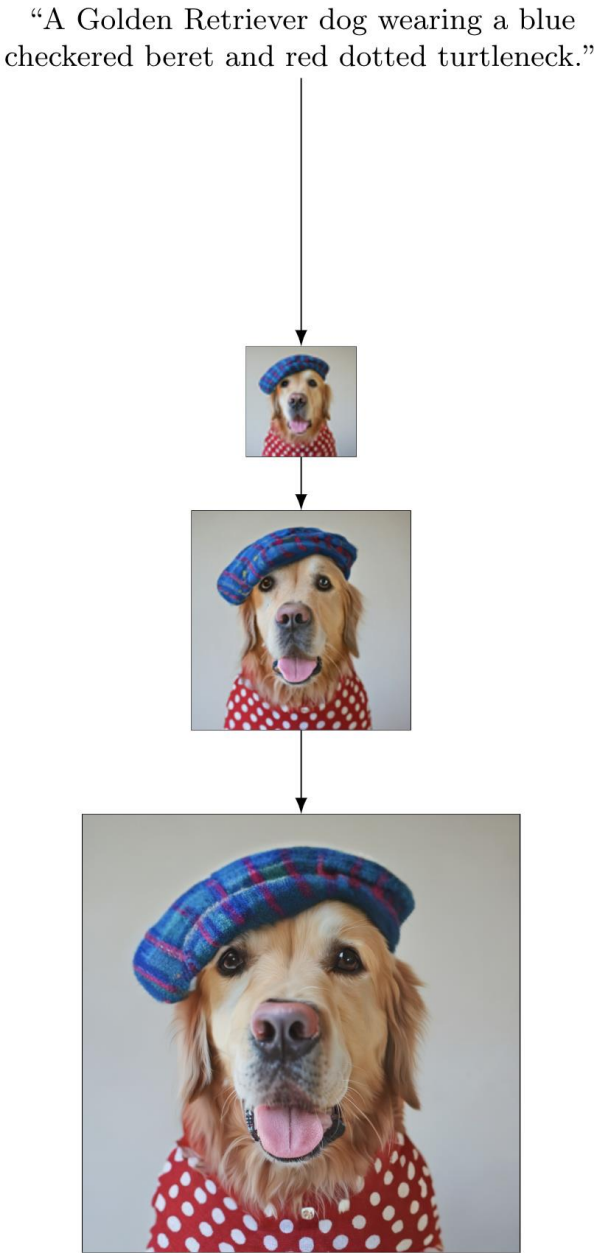


Imagen

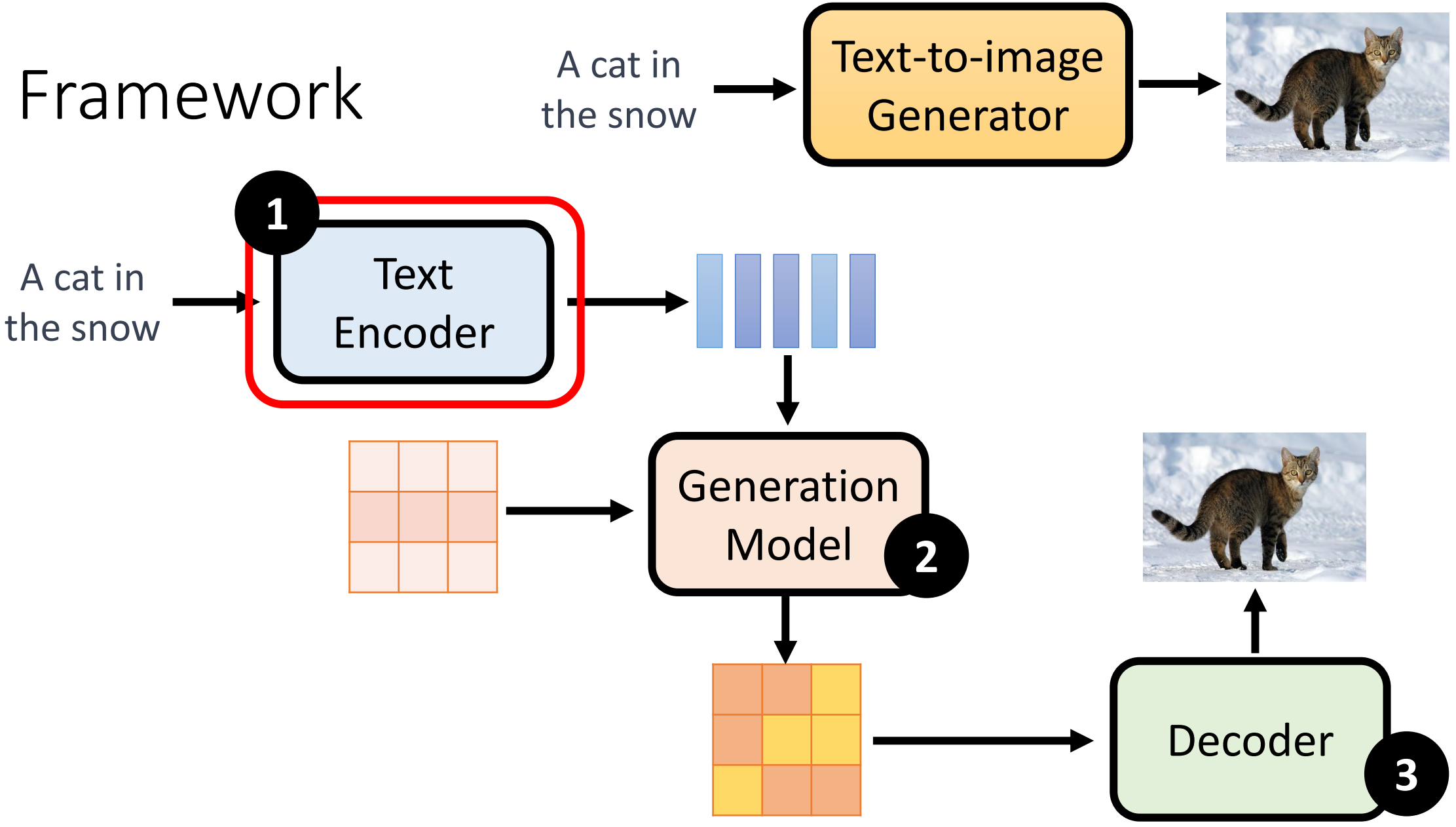
https://imagen.research.google/

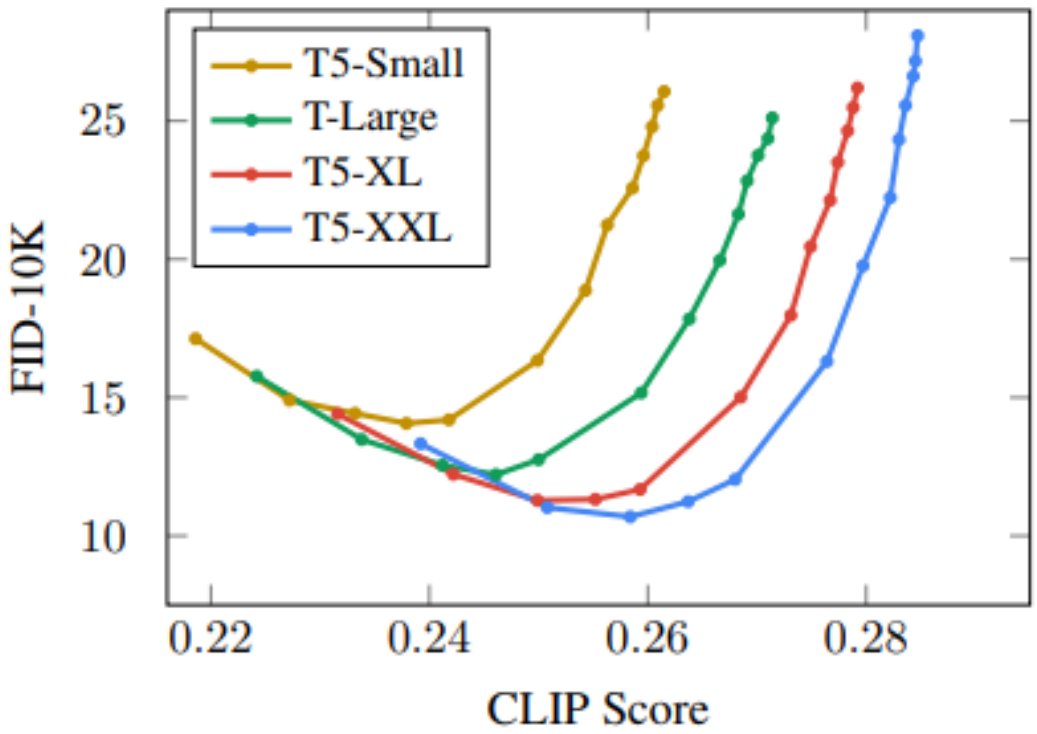
https://arxiv.org/abs/2205.11487



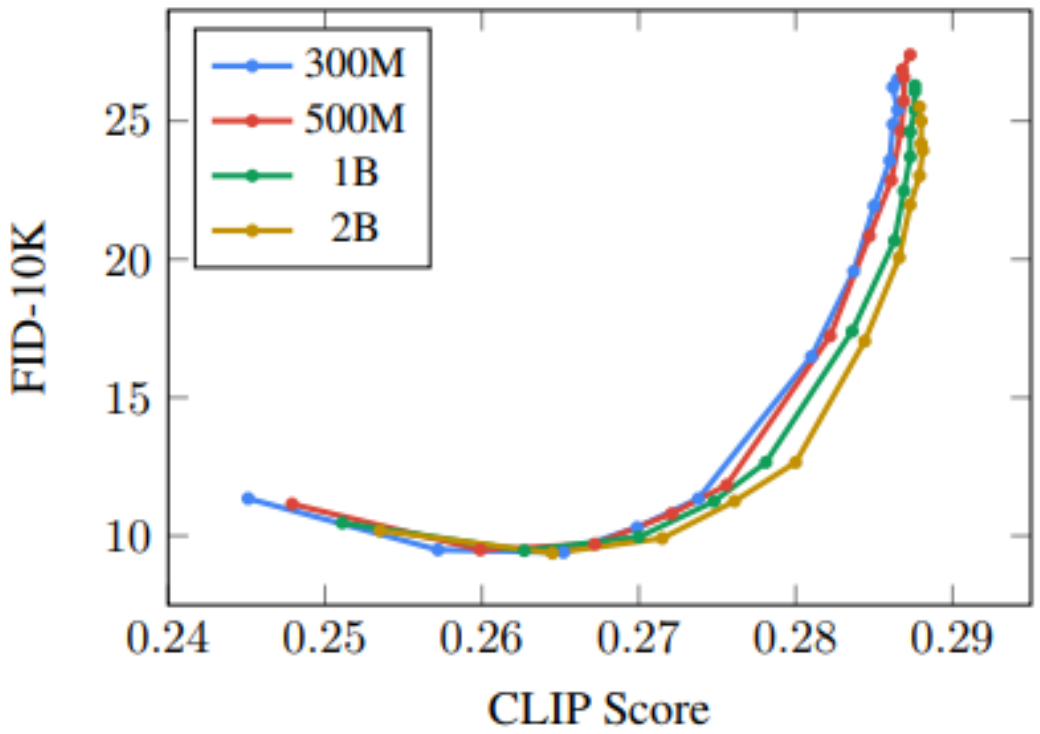


Framework


A cat in the snow
Text-to-image Generator
A cat in the snow
1
Text Encoder
Generation Model
2
Decoder
3



(a) Impact of encoder size.

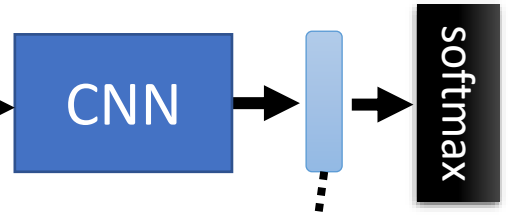


(b) Impact of U-Net size.

https://arxiv.org/abs/2205.11487

Fréchet Inception Distance (FID)

https://arxiv.org/abs/1706.08500



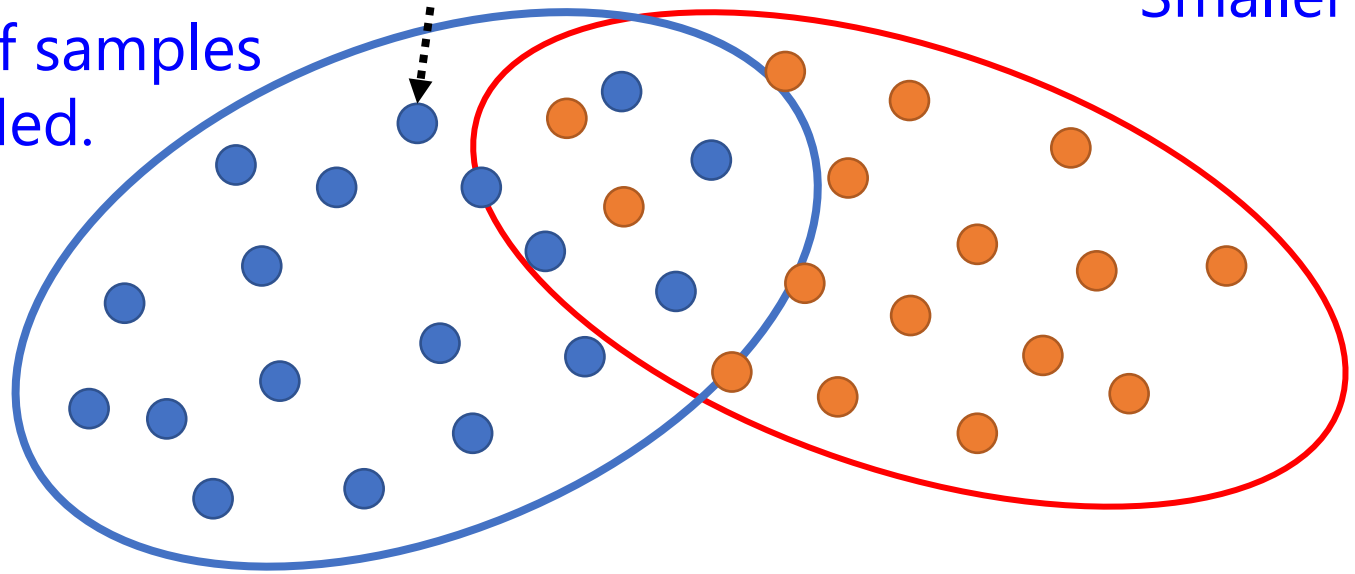
red points: real images

blue points: generated images

FID = Fréchet distance between the two **Gaussians** ???

Smaller is better

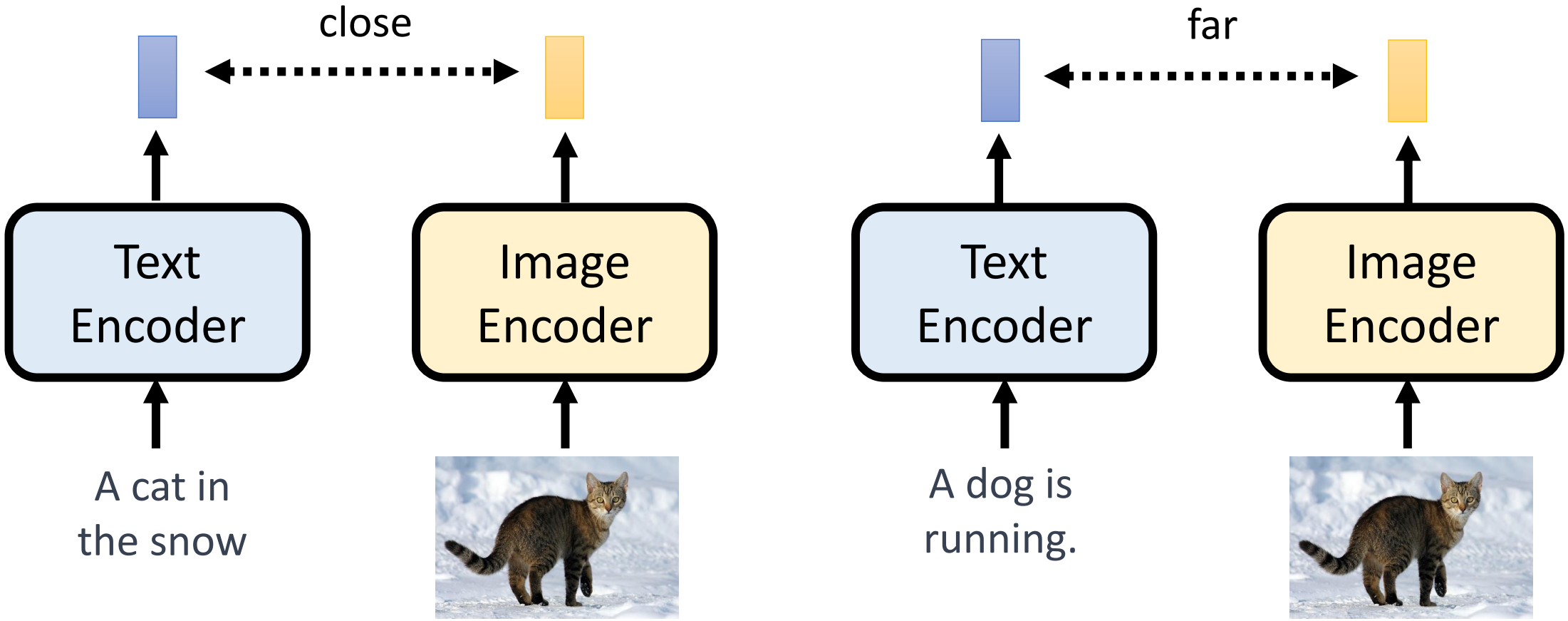
A lot of samples is needed.



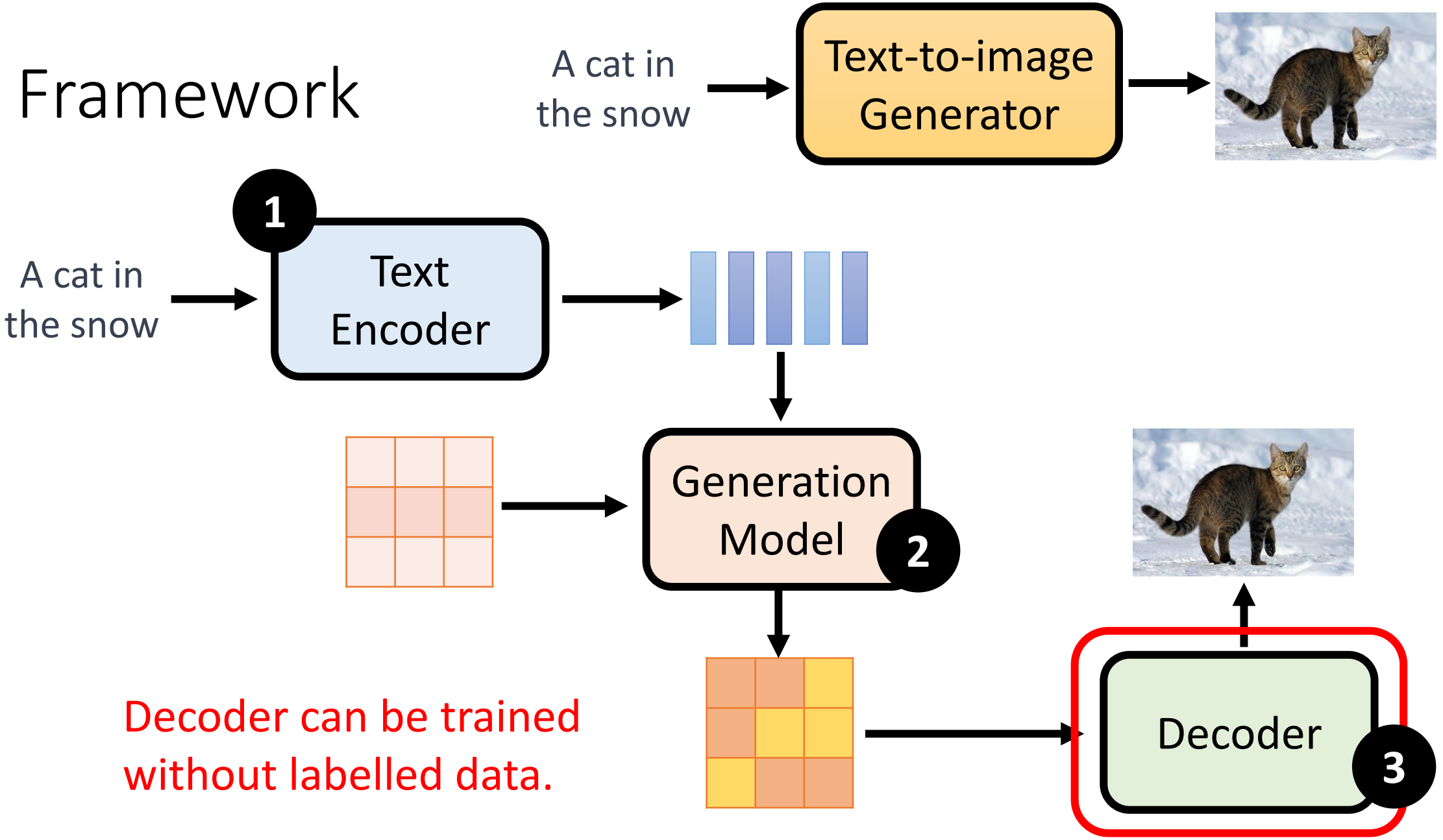
Contrastive Language-Image Pre-Training (CLIP)

https://arxiv.org/abs/2103.00020

400 million image-text pairs



Framework


A cat in the snow
Text-to-image Generator
A cat in the snow
1
Text Encoder
Generation Model
2
Decoder
3
Decoder can be trained without labelled data.

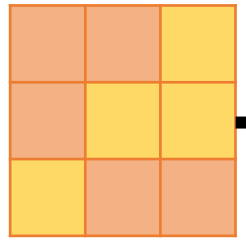
「中間產物」為小圖

Decoder

(Images are generated by Midjourney)

「中間產物」為「Latent Representation」

Latent Representation



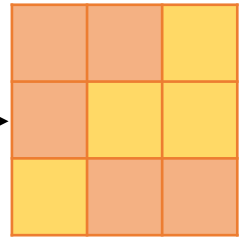
Decoder

Auto-encoder

$H \times W \times 3$

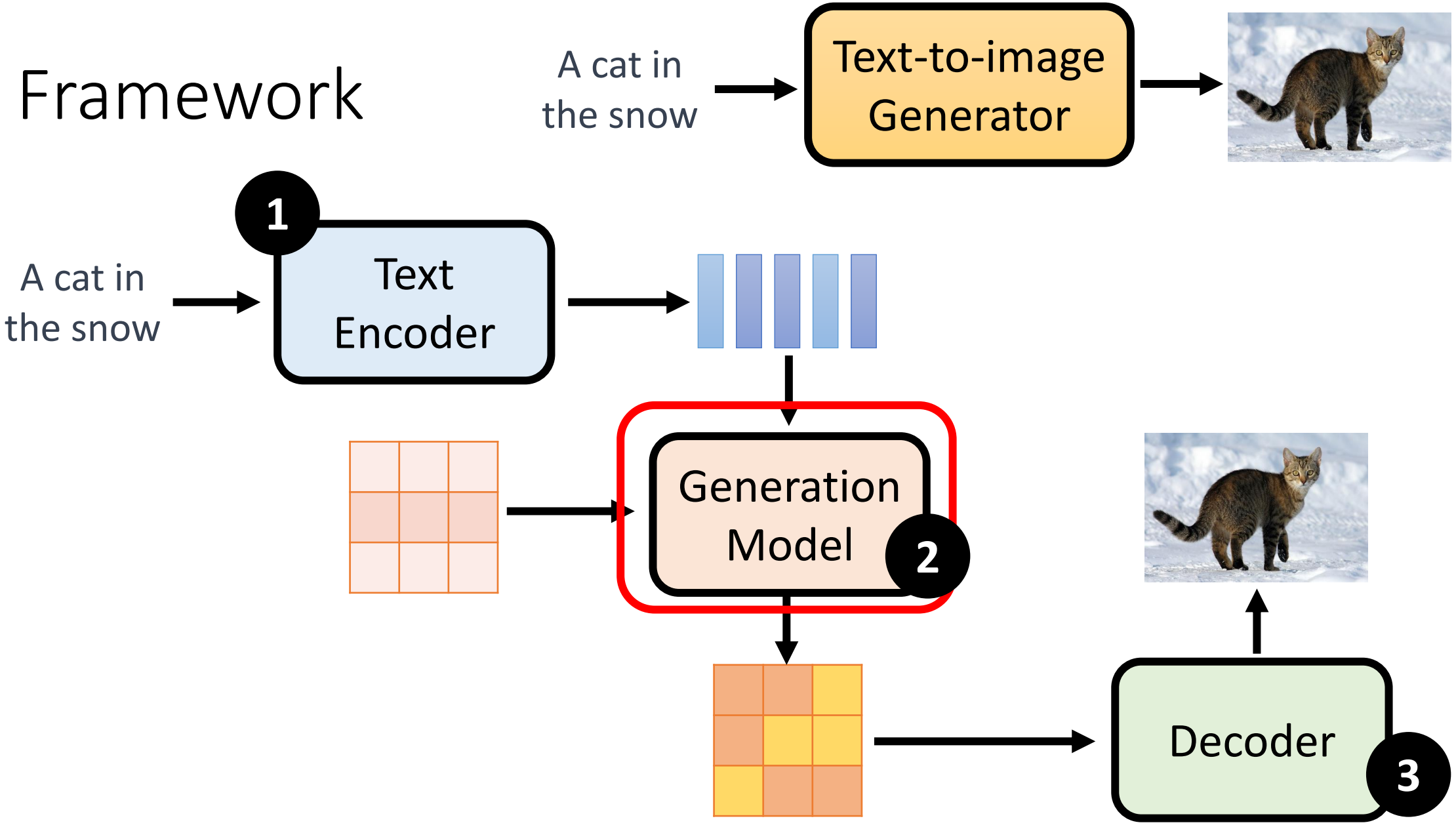
Encoder

$h \times w \times c$

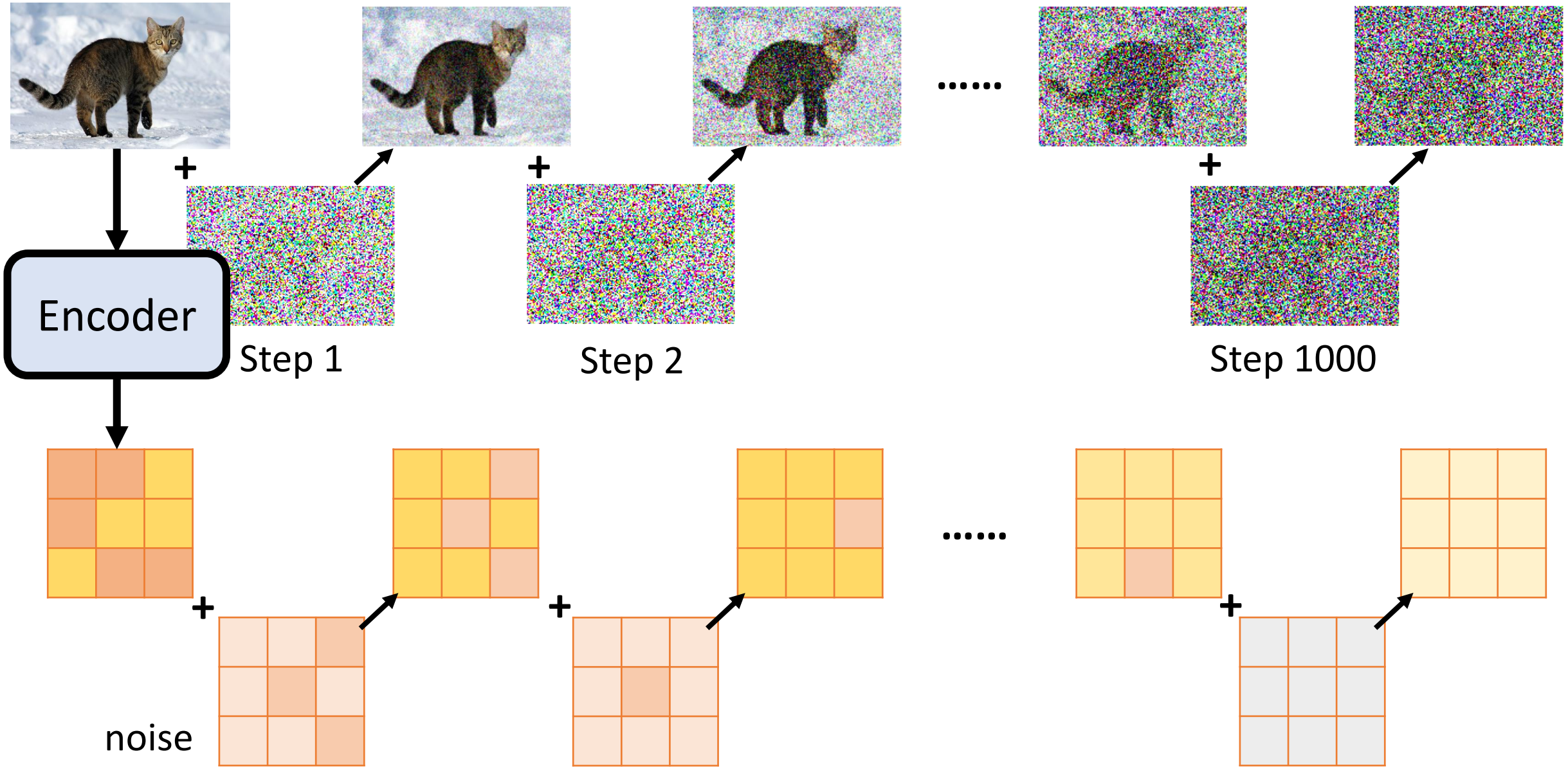


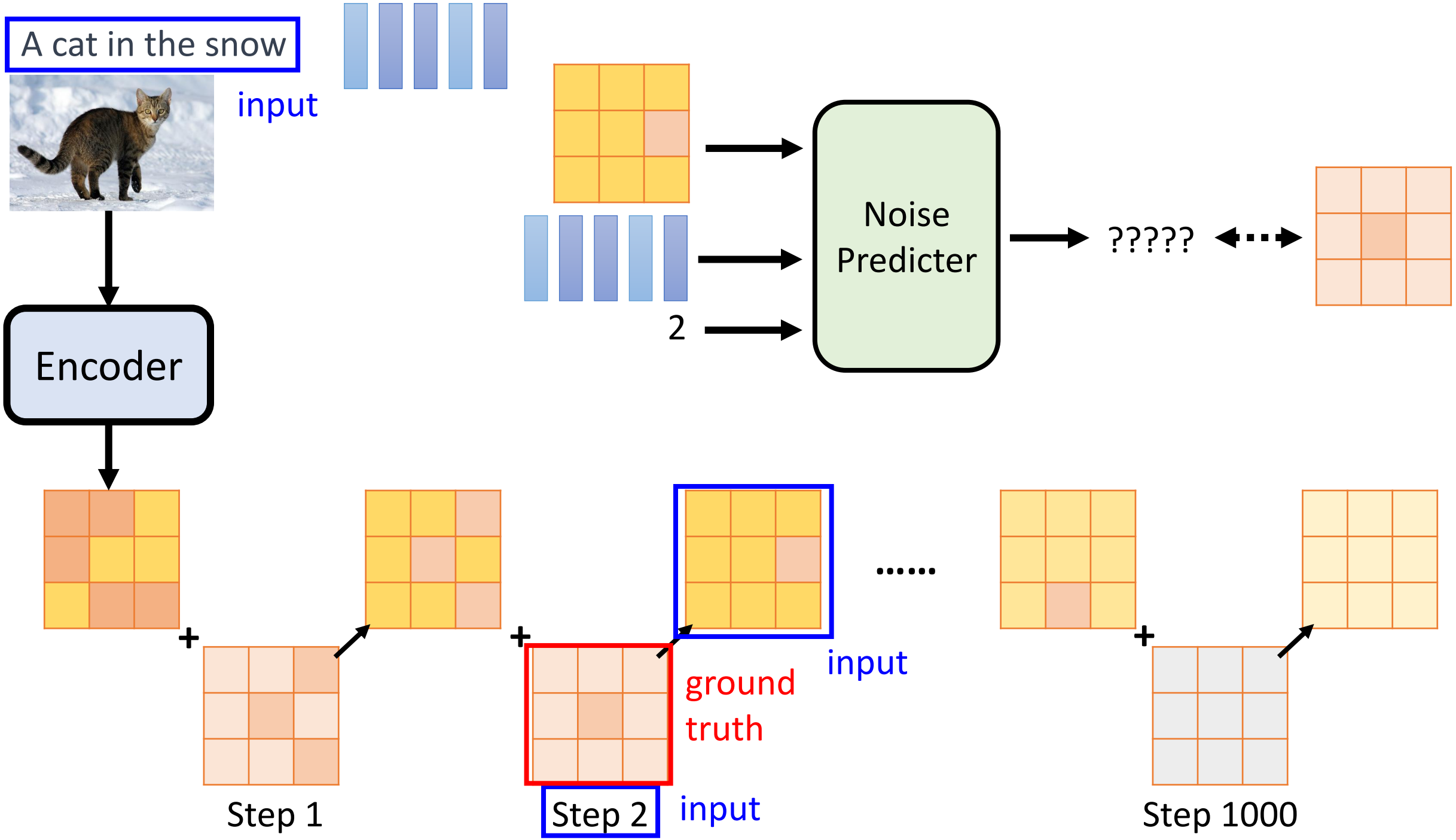
Decoder

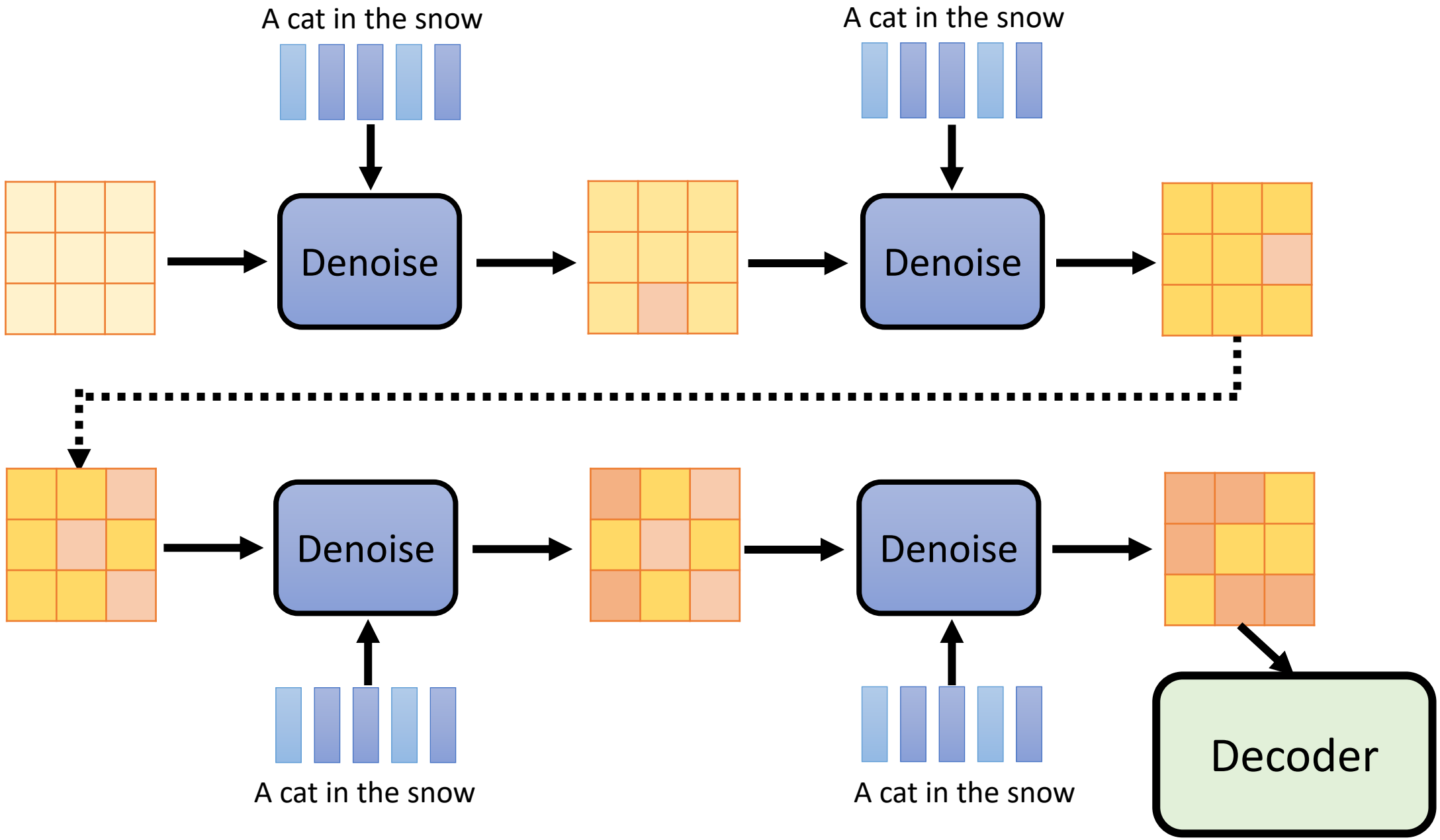
Framework


A cat in the snow
Text-to-image Generator
A cat in the snow
1
Text Encoder
Generation Model
2
Decoder
3

A cat in the snow

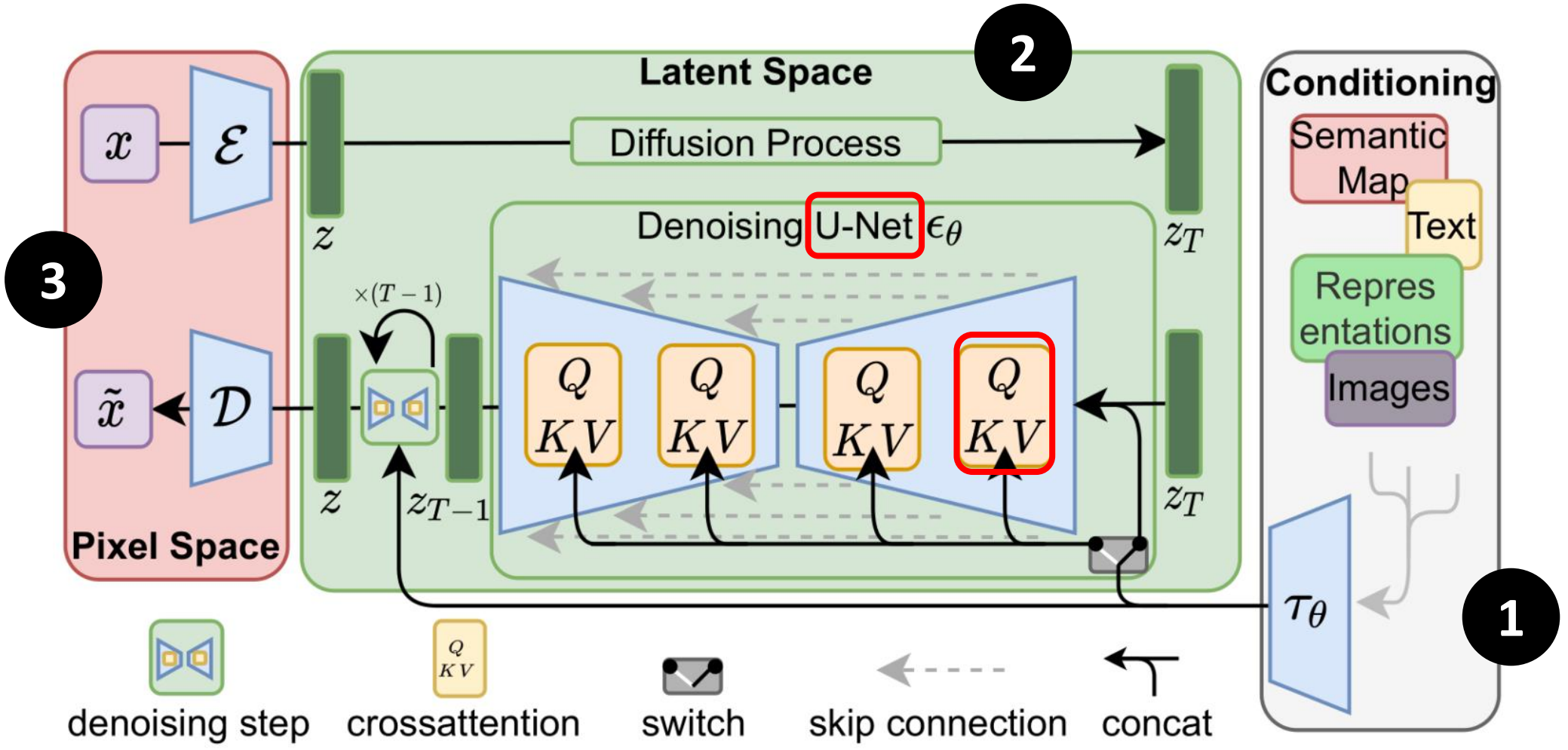



A cat in the snow
input
Encoder
2
Noise Predicter
?????
+
Step 1
+
ground truth
input
Step 2
input
.....
+
Step 1000

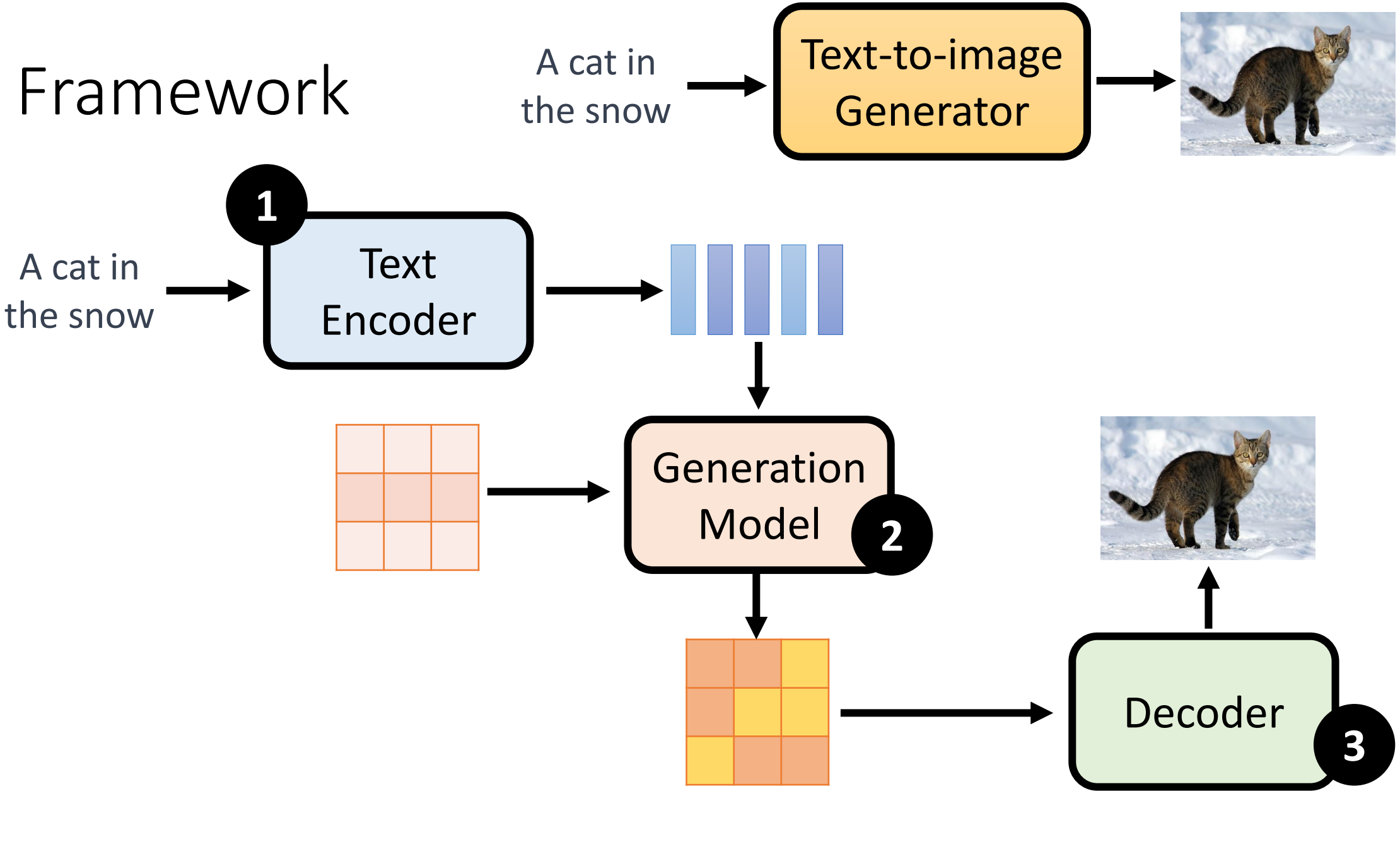

A cat in the snow
A cat in the snow
Denoise
Denoise
Denoise
Denoise
A cat in the snow
A cat in the snow
Decoder

Stable Diffusion

https://arxiv.org/abs/2112.10752



Framework


A cat in the snow
Text-to-image Generator
A cat in the snow
1
Text Encoder
Generation Model
2
Decoder
3