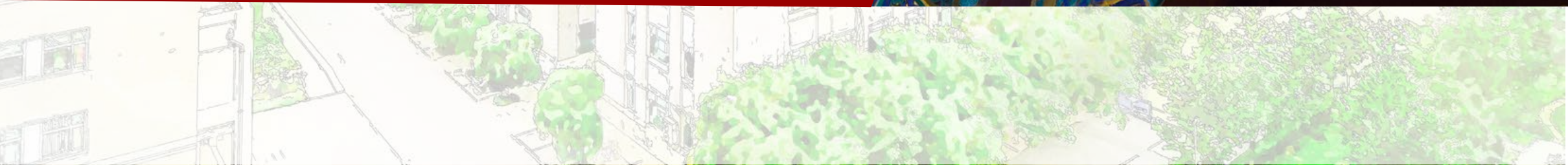




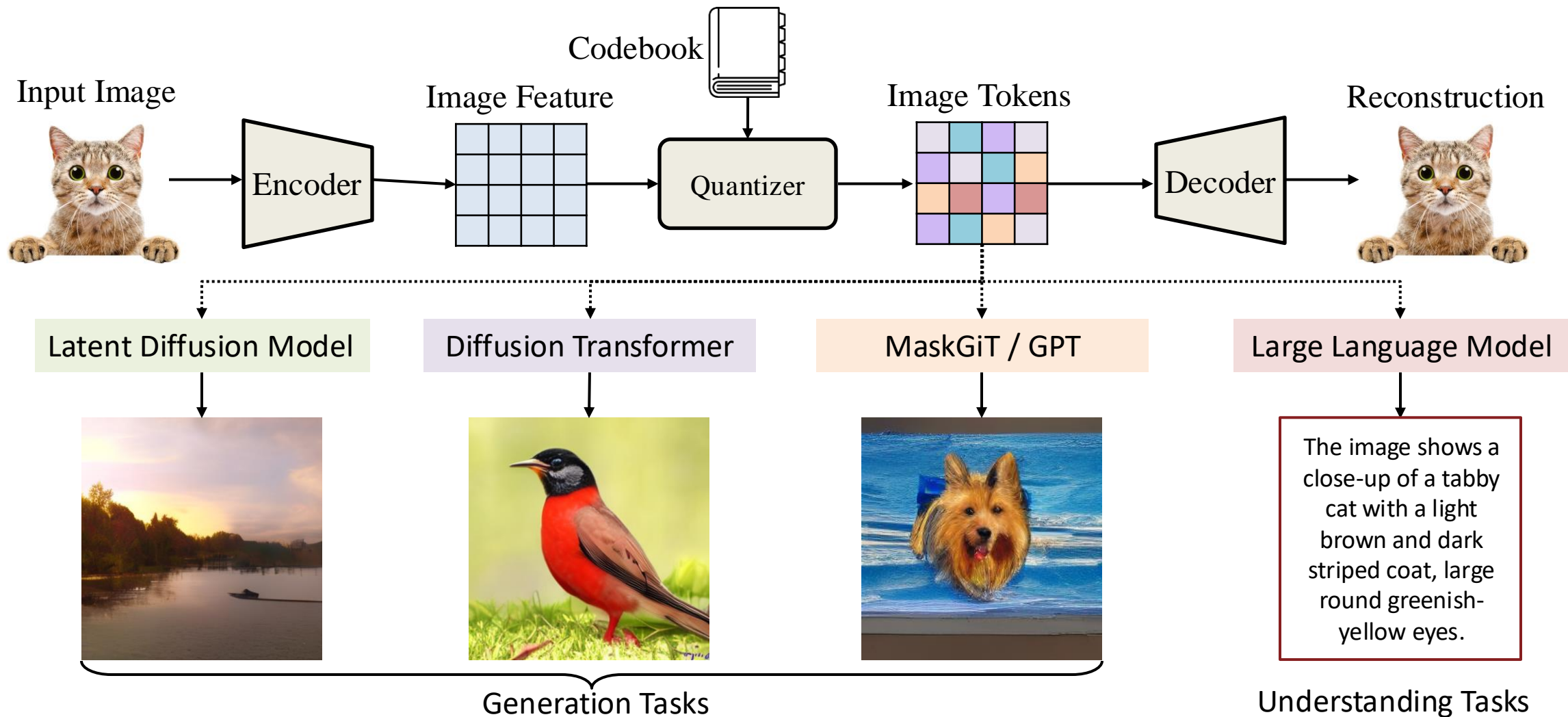
# Scaling the Codebook Size of VQGAN to 100,000 with a Utilization Rate of 99%

Lei Zhu<sup>1</sup>, Fangyun Wei<sup>2</sup>, Yanye Lu<sup>1</sup>, Dong Chen<sup>2</sup>  
<sup>1</sup>Peking University, <sup>2</sup>Microsoft Research Asia



# Image Quantization

- Encoding an image into image tokens, which plays a pivotal role in various applications.





# Codebook Collapse

➤ VQ-based image quantization models face codebook collapse problem, limiting its effectiveness.

Codebook Size (1024)      Codebook Size (16384)      Codebook Size (50,000)      Codebook Size (100,000)

VQGAN-FC



100% Utilization Rate



11.2% Utilization Rate



3.6% Utilization Rate

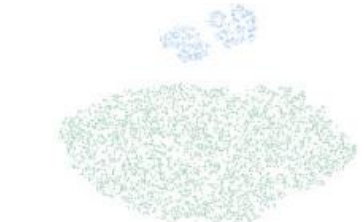


1.9% Utilization Rate

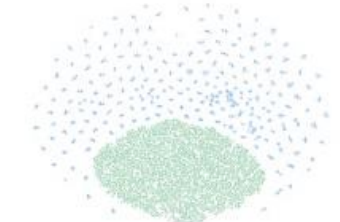
VQGAN-EMA



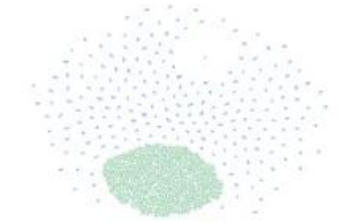
100% Utilization Rate



83.2% Utilization Rate



40.2% Utilization Rate

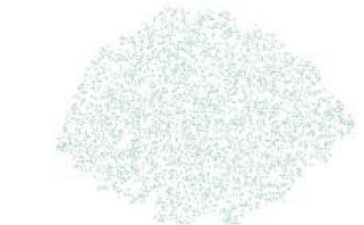


24.2% Utilization Rate

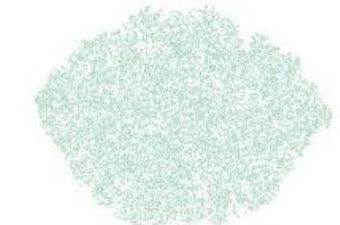
VQGAN-LC (Ours)



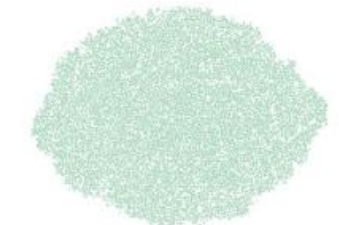
100% Utilization Rate



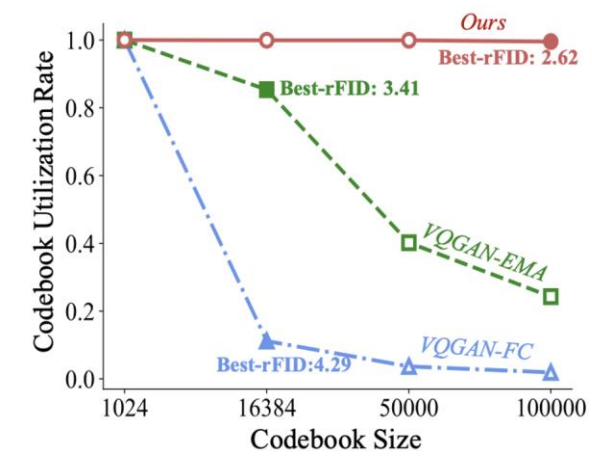
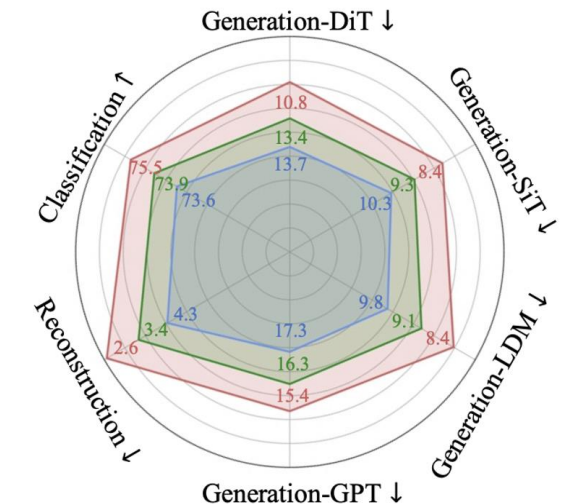
99.9% Utilization Rate



99.9% Utilization Rate

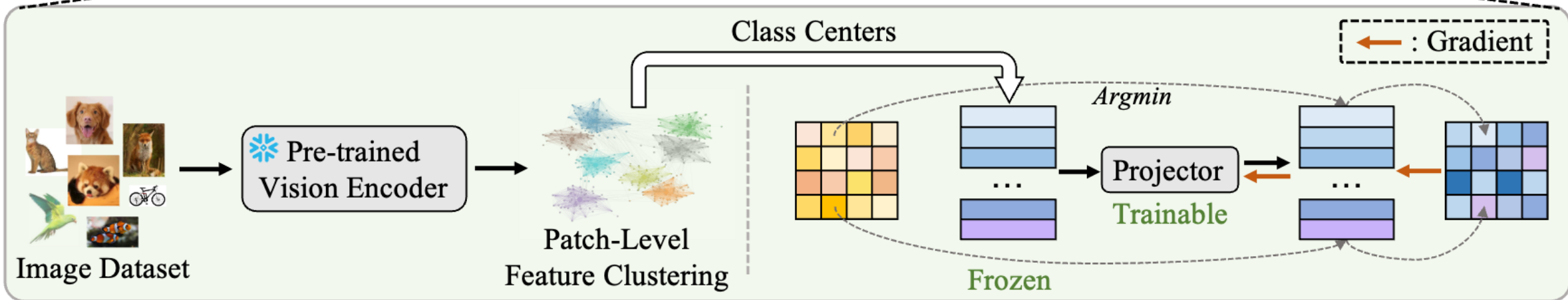
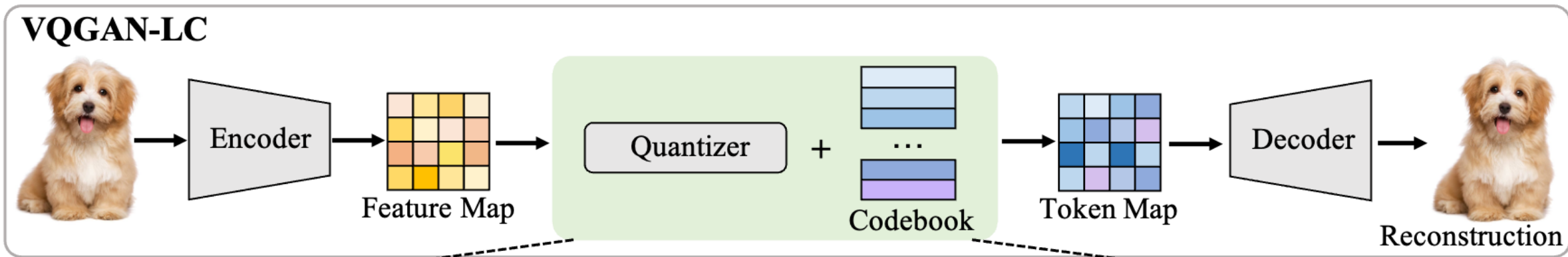


99.9% Utilization Rate



Visualization of the *active* and *inactive* codes for VQ-based models using t-SNE

- Learning a codebook projector to align the distribution between codebook and image features.





# Experiments

➤ Scaling up codebook size to pursue better performance for quantization and downstream tasks.

Method	Reconstruction (rFID)				Generation with LDM [3] (FID)			
	1,024	16,384	50K	100K	1,024	16,384	50K	100K
VQGAN-FC	4.82	4.29	4.96	4.65	10.81	9.78	10.37	10.12
VQGAN-EMA	4.93	3.41	3.88	3.46	10.16	9.13	9.29	9.50
VQGAN-LC (Ours)	4.97	3.01	2.75	2.62	9.93	8.84	8.61	8.36





# Thank You!

