國立臺灣大學
National Taiwan University

Qualcomm

Continual Personalization for Diffusion Models

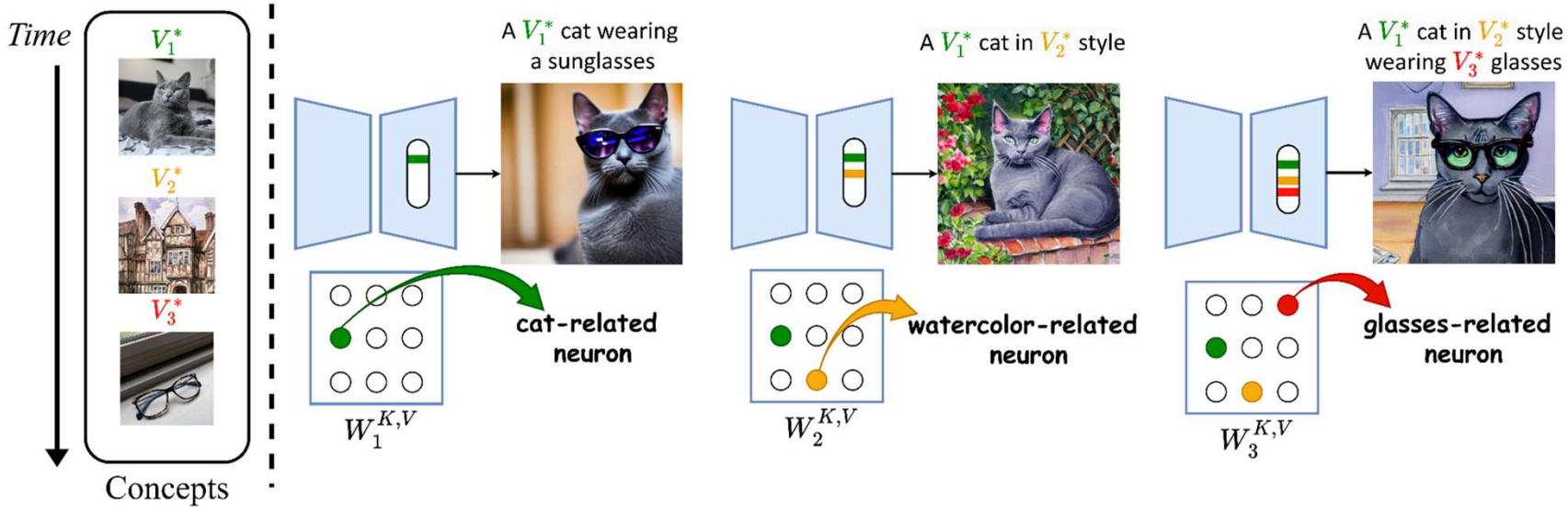
Yu-Chien Liao* Jr-Jen Chen* Chi-Pin Huang
Ci-Siang Lin Meng-Lin Wu Yu-Chiang Frank Wang

Continual Personalizaion

- Learn multiple concepts within a single model **without extra module.**
  - Capable of combining multiple concepts in a single image.
- Personalize each concept in **a sequential and never-ending manner.**
  - Being able to dynamically add new concepts without extra operation.

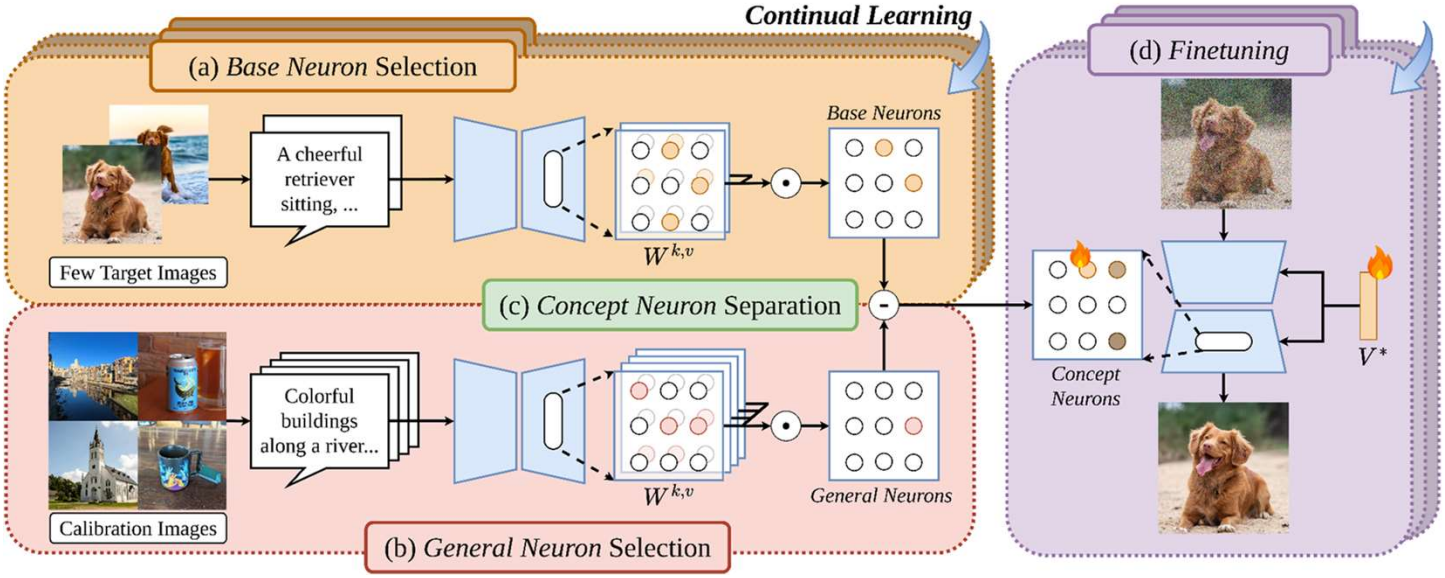
Problem Settings

- **Find concept-related neurons** for each concept as training target.
  - We aim to find **disjoint neurons** for each concept.
- Consider several user-provided concepts (1-6 images per concpet).



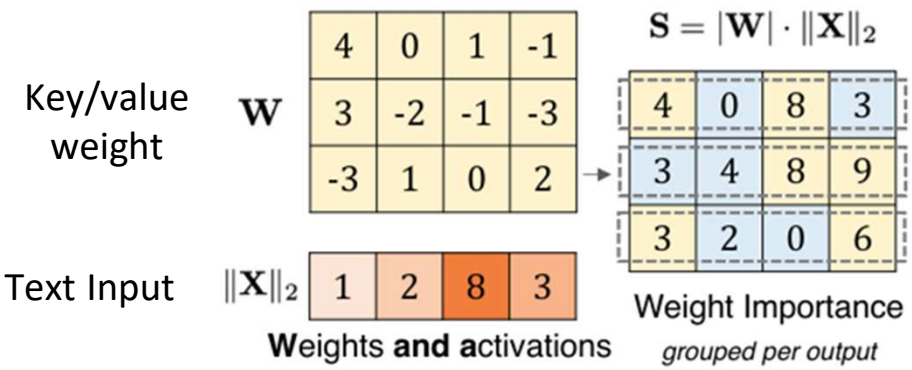
Methods - Overview

- Classify neurons to different categories.
- Find ***Concept neurons*** by excluding ***General neurons*** from ***Base neurons***.
  - Finetune different ***Concept neurons*** for different concepts to prevent forgetting.



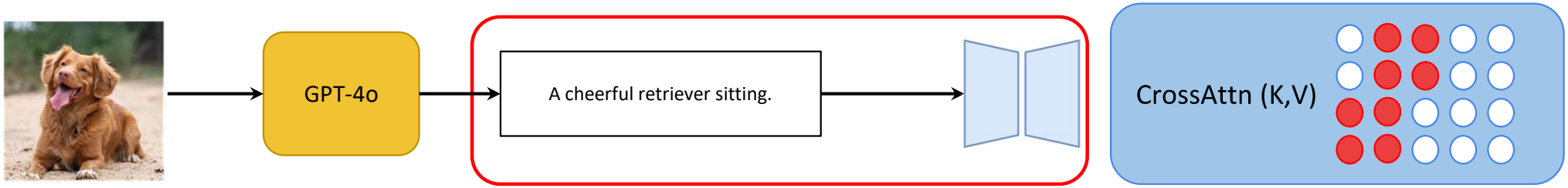
Methods - Scoring Each Neurons

- Tuning the specific **Key/Value neurons in CrossAttn** only.
  - Which handle the text information in diffusion model.
- Finding which neurons contribute to the generation.
  - Assign **important score** for all neurons with **Wanda**[1].
  - Pick high-scoring neurons as ***Base neurons.***



[1] *A Simple and Effective Pruning Approach for Large Language Models*

Target Image Set



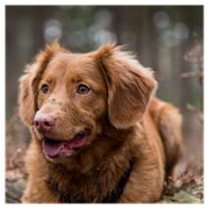
Target Image Set

GPT-4o

A cheerful retriever sitting.

CrossAttn (K,V)

and



GPT-4o

A retriever relaxes in a forest, its eyes alert and tongue slightly out

CrossAttn (K,V)

and

GPT-4o

A retriever is running in the ocean.

CrossAttn (K,V)

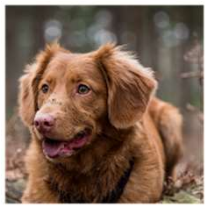
Target Image Set

GPT-4o

A cheerful retriever sitting.

CrossAttn (K,V)

and



CrossAttn (K,V)
(*General neurons*
+
"dog" *Concept neurons*)

CrossAttn (K,V)

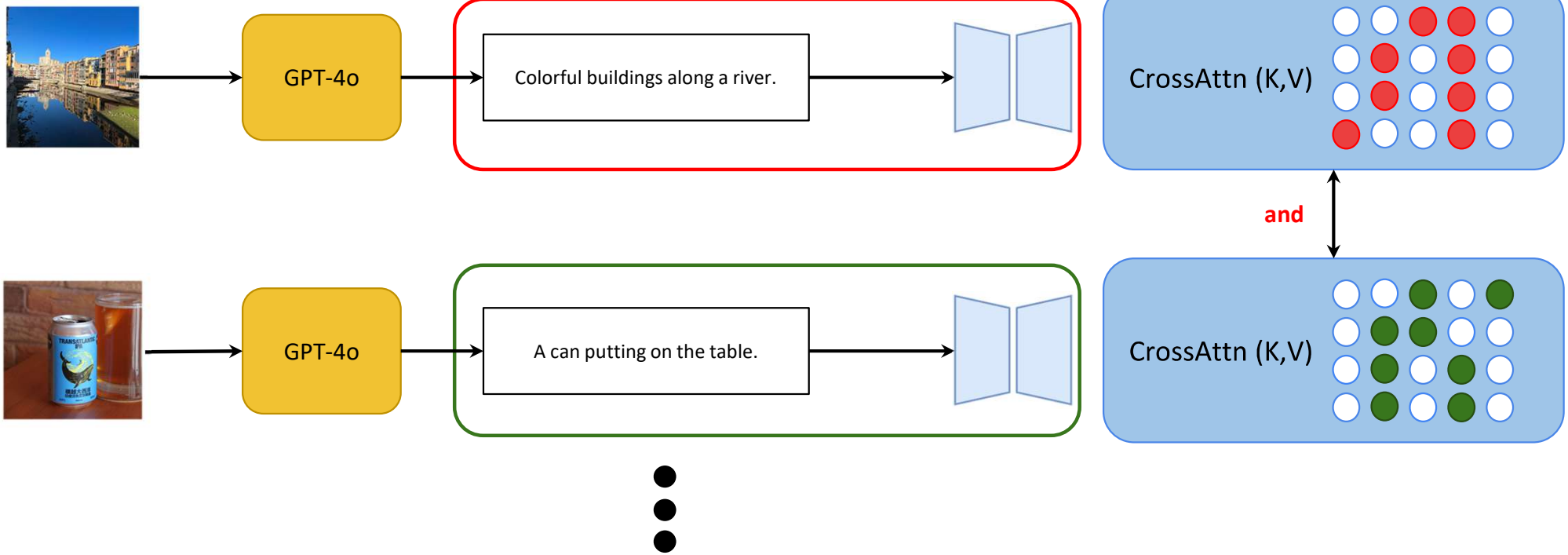
and

GPT-4o

A retriever is running in the ocean.

CrossAttn (K,V)

Calibration Image Set



Calibration Image Set

GPT-4o

Colorful buildings along a river.

CrossAttn (K,V)

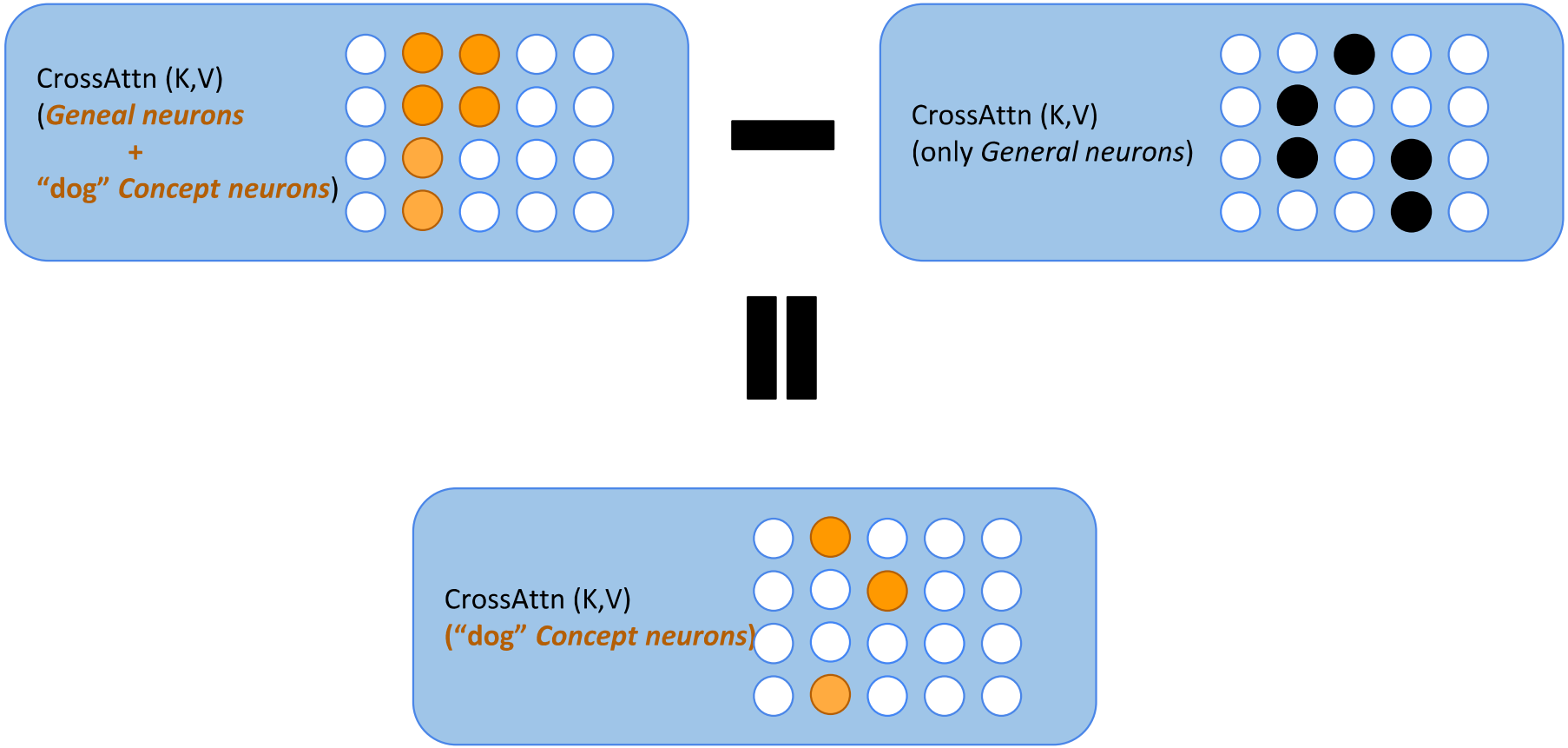
and

GPT-4o

A can putting on the table.

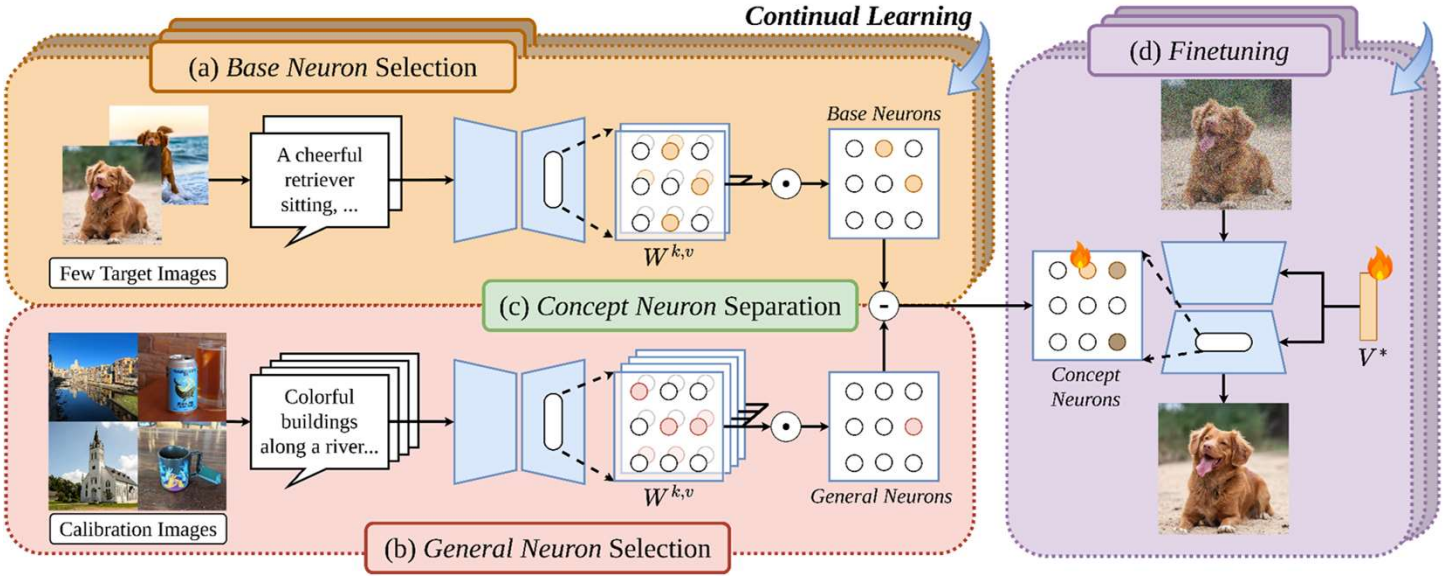
CrossAttn (K,V)

CrossAttn (K,V)
(only ***General neurons***)


CrossAttn (K,V)
(*Geneal neurons*
+
"dog" *Concept neurons*)
CrossAttn (K,V)
(only *General neurons*)
CrossAttn (K,V)
("dog" *Concept neurons*)

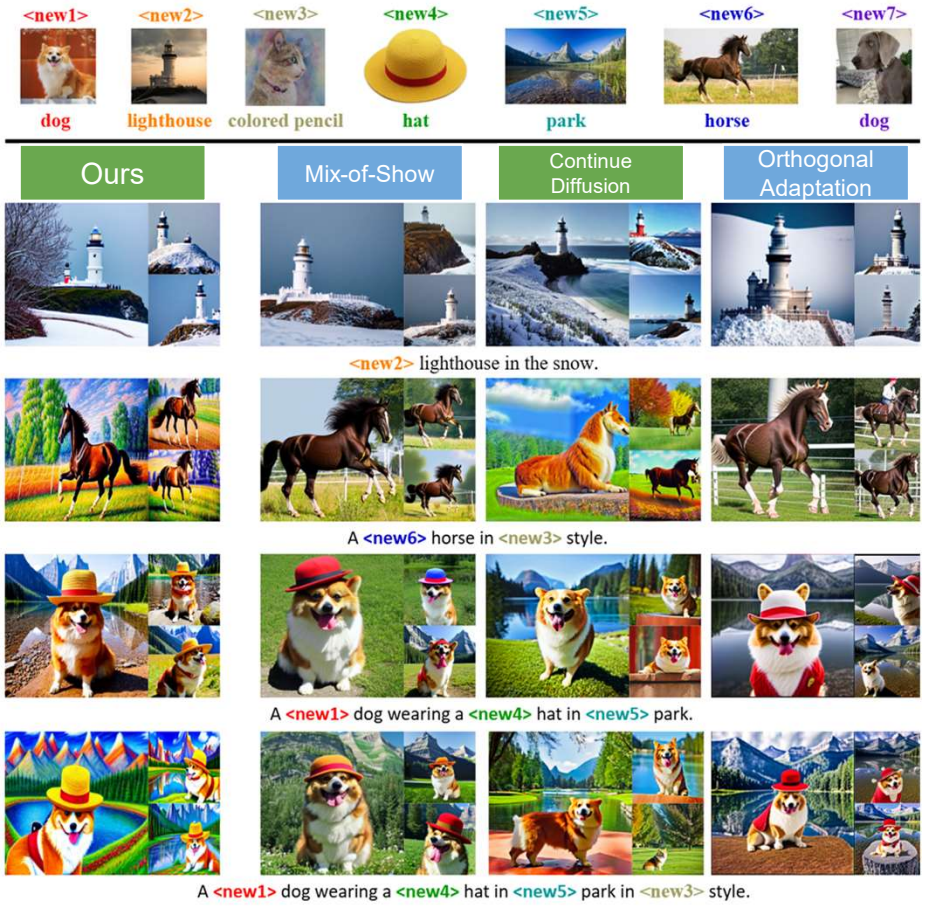
Methods - Overview

- Find ***Concept neurons*** by excluding ***General neurons*** from ***Base neurons***.
  - Finetune different ***Concept neurons*** for different concepts to prevent forgetting.



Experiments - Qualitative Result

- For **continuously learning method**
  - Continuously learned 7 concepts.
- For **multiple concepts learned method**
  - Each concept learn a LoRA, and combined all 7 LoRAs.



Experiment - Quantitative Result

Methods	Single Concept		Multiple Concepts		Computational Resources	
	CLIP-I ↑	CLIP-T↑	CLIP-I↑	CLIP-T↑	Memory(MB)↓	Time(s)↓
Textual Inversion [11]	72.76	72.69	65.30	65.00	**0 / 0**	**0**
Custom Diffusion [18]	67.88	74.92	65.87	68.70	3547 / **0**	10
Mix-of-Show [13]	**75.86**	75.75	65.26	70.62	62852 / **0**	727
Orthogonal Adaption [27]	74.67	74.87	66.37	69.20	5663 / 3167	42
Continual Diffusion [32]	71.82	66.12	66.15	60.30	2461 / 4747	10
CNS	74.88	**76.95**	**67.21**	**79.22**	**0 / 0**	**0**

Evaluation metrics - ***CLIP-I*** and ***CLIP-T*** for the performance of the method,
cost of ***Memory*** and ***Time*** for the efficiency of the method.

Thanks for your attention!