Homework 10 - Diffusion

TAs: 林熙哲、袁紹翔、劉建蘴
ntu-ml-2025-spring-ta@googlegroups.com
Deadline: 2025/06/13 23:59:59 (UTC+8)

Links

- ML 2025 Spring
- NTU COOL
- JudgeBoi
- Sample Code

Outline

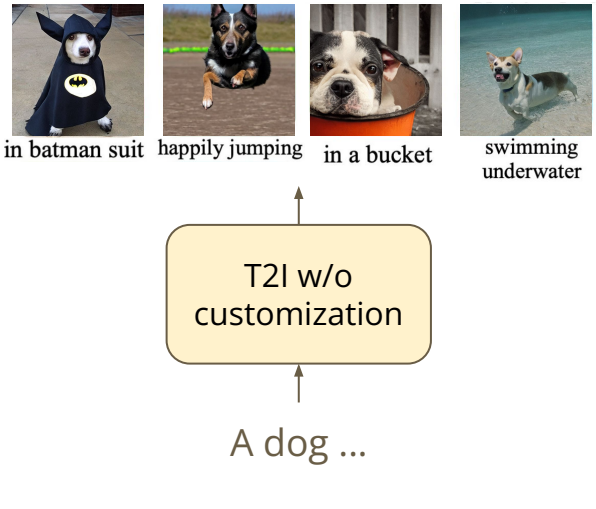
- Task description
- Eval Metric
- TODOs
- Dataset
- Grading
- Hints
- Submission
- Regulations
- References and Appendices

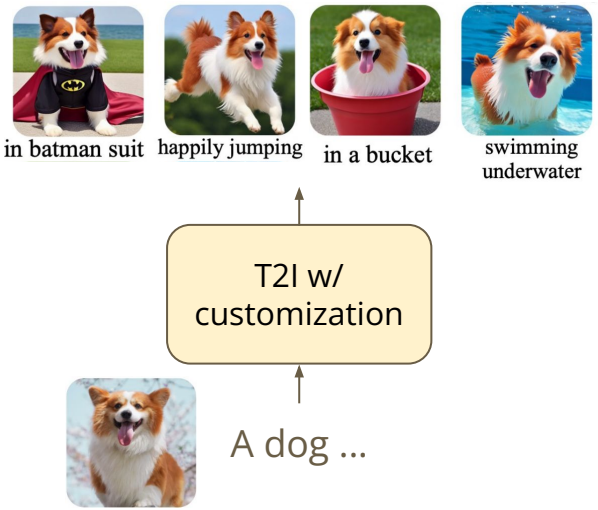
Previous Course

- ML 2023 Spring
- 【生成式AI】速覽圖像生成常見模型
- Diffusion Model Playlist

Task description

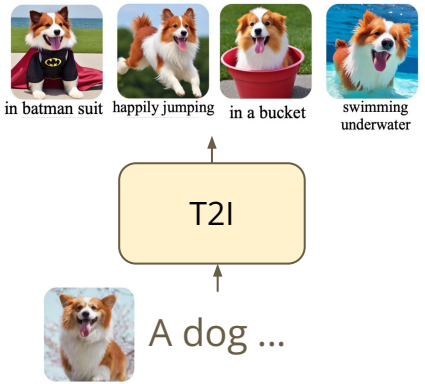
- **Task**: Customization of text-to-image (T2I) diffusion model
- **Goal**: Given few images of specific object (reference) we want make diffusion model manipulate the object in the generated images





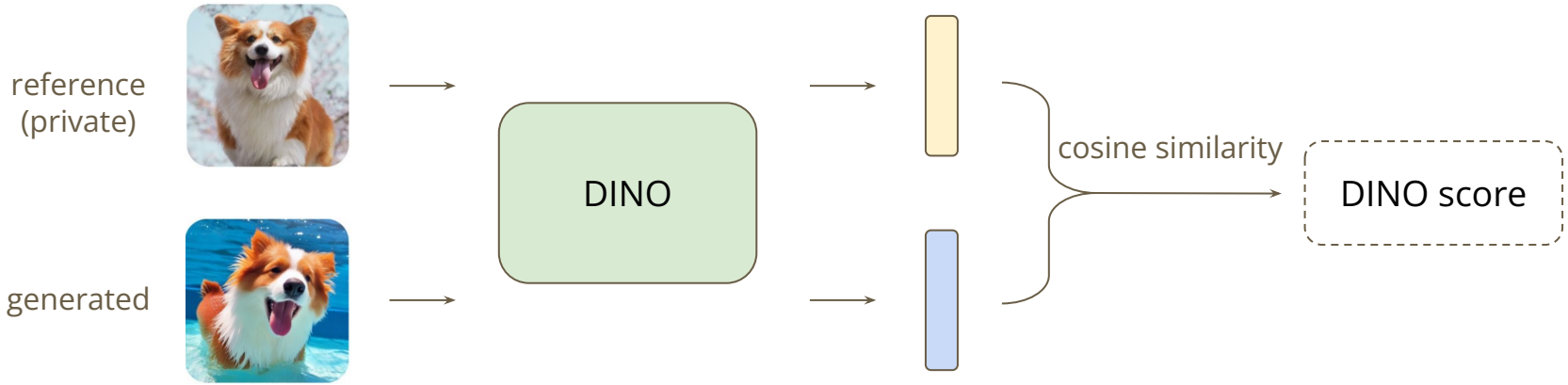
Task description

- You will be given
  - Images of reference objects to customize
  - Text prompts to generate images of reference objects
- You needs
  - Generate images that following the text prompts with reference objects in it
- We will evaluation
  - Whether objects in generated images match the reference
  - Whether the generated images follow the text prompts



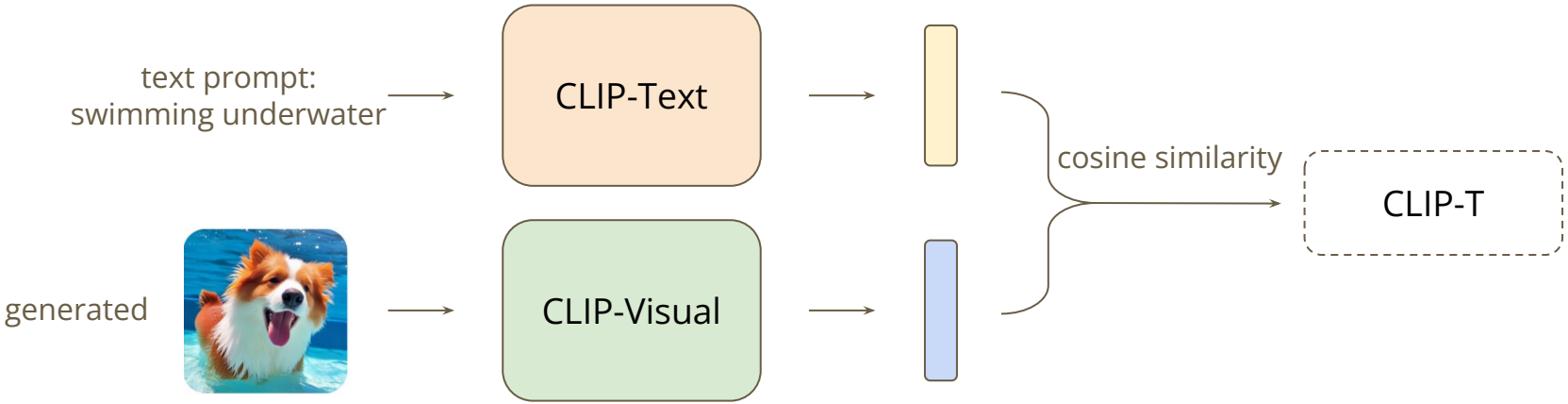
Eval Metric: DINO Score

- To measure “Whether objects in generated images match the reference”
- Use DINO to extract features for generated and private reference images
- Compute the average cosine similarity between the extracted features (the higher the better)



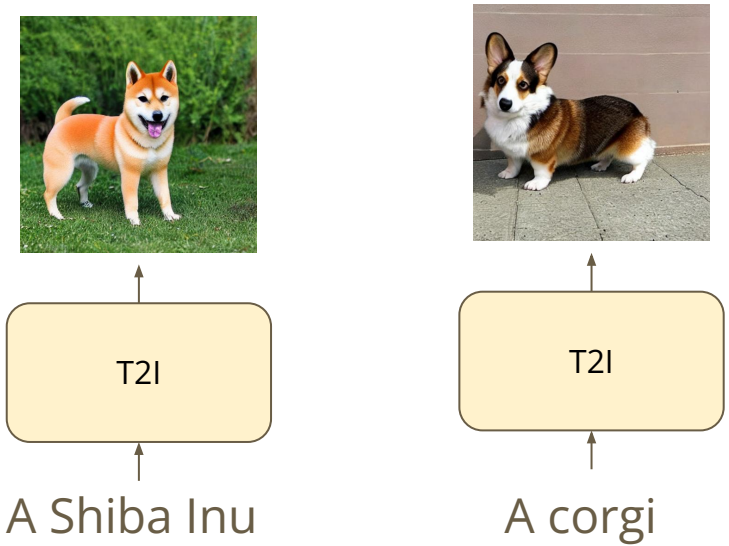
Eval Metric: CLIP-T

- To measure “Whether the generated images follow the text prompts”
- Use CLIP to create features for text prompts and generated images
- Compute the average cosine similarity between the extracted features (the higher the better)



Reason Why Normal T2I Model Can't DO Customization

- T2I model actually can generate object of specific characteristics



Reason Why Normal T2I Model Can't DO Customization

- T2I model actually can generate object of specific characteristics
- BUT it is difficult to describe the object we want to customize

Reference

T2I

A corgi

T2I

A Pembroke Welsh Corgi. It has a reddish-brown and white coat. The reddish-brown fur covers the ears, upper face, and much of the body, while the white fur is prominent on the chest, underbelly, front legs, and part of the face. The fur appears soft, fluffy, and well-groomed...

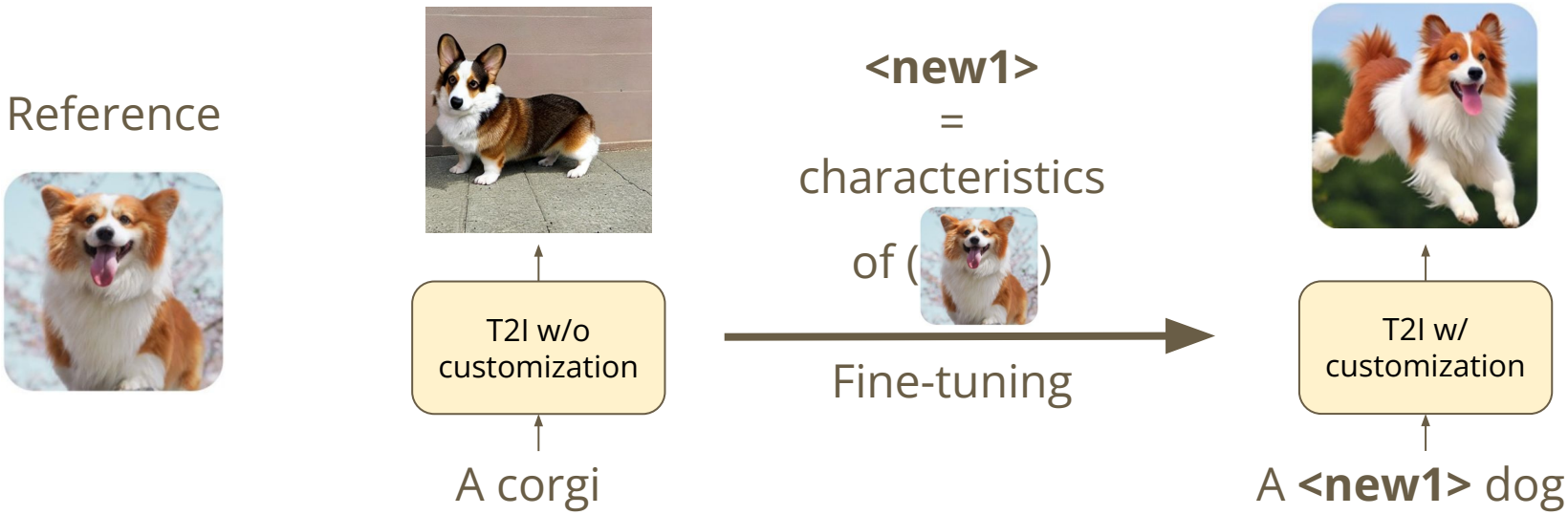
T2I

A **"special"** dog

Kumari et al. Multi-Concept Customization of Text-to-Image Diffusion (CVPR' 23)

Custom Diffusion

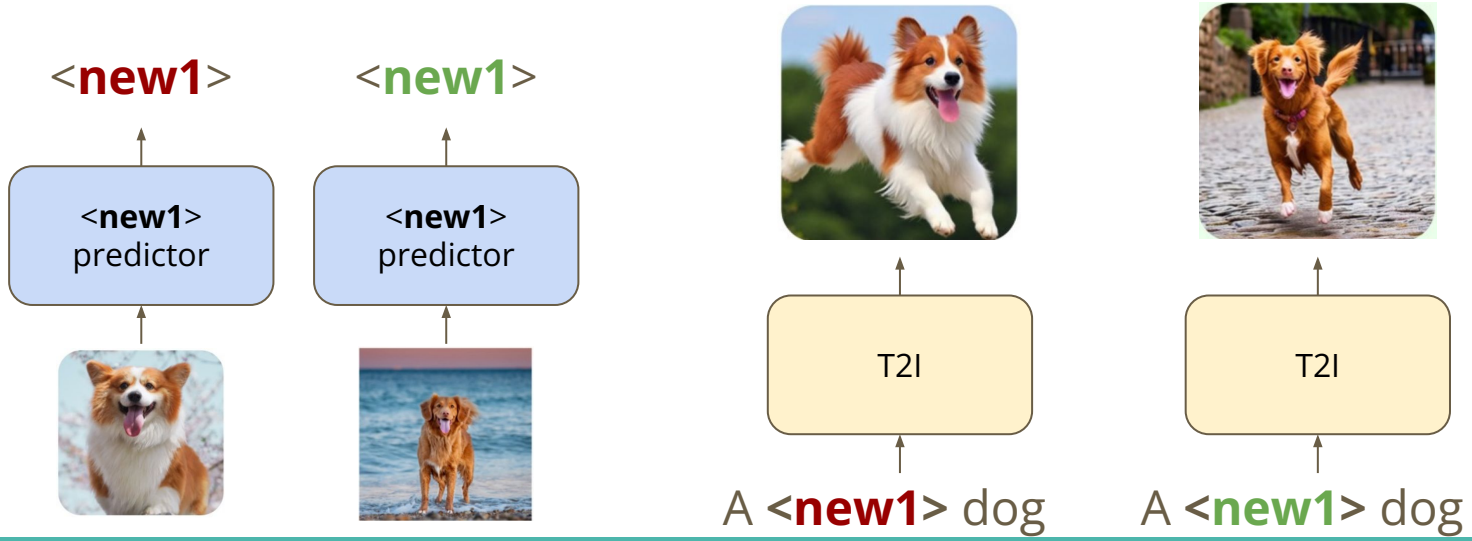
- Fine-tune to make T2I model learn a word "**<new1>**" that represent the object characteristics (Recall: Homework 5)



Li et al. Multi-Concept Customization of Text-to-Image Diffusion (arXiv' 2305)

BLIP-Diffusion

- Using custom diffusion, we need to fine-tune once for each object
- Why not train a model to predict the embedding of the "<**new1**>"
- **Don't need to fine-tune!**



TODOs

- Use the above two method in sample code
- Observe the generated image to tune hyperparameters
- Submit the generated images

Dataset

- **6** objects to customize each has
  - Images (one or many) as reference for customization
  - Text prompt as model input to generate images for evaluation
- For each object, you need to generate **15** images.

on the grass

on a cobble-stone street

in the jungle

with sunglasses

in the snow

on a plate

Grading

- 10 points in total, submit before deadline: 2025/06/13 23:59:59 (UTC+8)
- **No late submission is allowed**
- Successfully customize 1 object = +1pt

Baseline	Score	Estimated Time	Hints
Simple	2pts	20 mins	Simply run BLIP Diffusion
Medium	2pts	60 mins	Tune hyperparameters for BLIP Diffusion
Strong	2pts	4 hours	Use Custom Diffusion
Code submission	4pts	-	Submit code to NTU COOL

Grading

“Successfully customize a object” means DINO score and CLIP-T of the generated images are **both** higher than the following thresholds

	DINO Score	CLIP-T
Object 1	68	18
Object 2	60	17
Object 3	61	18
Object 4	68	19
Object 5	60	19
Object 6	57	17

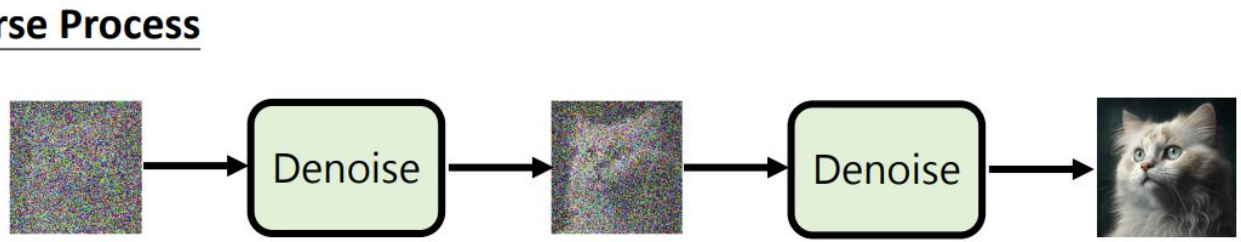
Hints: Disclaimer

- It is not guarantee to surpass the baseline if you follow all of the hints
- A reasonable range of value is given for you to adjust hyperparameters
  - You may get better result using a value out of the range
  - You may get worse result using a value within the range
- DINO score and CLIP-T may not fully align with human feelings
  - You may find that high scores don't always lead to satisfying results (and sometimes the reverse is true)
  - Human evaluation is usually needed but we don’t have efforts to do that

figure: ML 2023 Spring

Hints: BLIP Diffusion

- BLIP Diffusion works quite well for most cases so **try it first**
- There are mainly two hyperparameter you can tune for BLIP Diffusion
  - **num_inference_steps**: The number of denoising steps. More denoising steps usually lead to a higher quality image at the expense of slower inference.
  - **guidance_scale**: Higher guidance scale encourages to generate images that are closely linked to the text prompt, usually at the expense of lower image quality.



Hints: BLIP Diffusion

- Suggestion on tuning hyperparameters
  - **num_inference_steps**: The default value in sample code is relative low. Keep increasing until resulting quality no longer improves. (reasonable value: less than 100)
  - **guidance_scale**: The default value in sample code is relative low. Try to increase it and observe the resulting images. If you see the artifacts like the following, it usually means the value is too large. (reasonable value: less than 20)

Hints: Custom Diffusion

- For more difficult cases (e.g., **object 5, 6**), you'll need to fine-tune
- There are few things you can try to adjust to pass the strong baseline
  - **instance_prompt**: The text prompt used for training (teaching model **<new1>**)
  - **parameter_to_train**: What (and how many) parameters to fine-tune
  - **learning_rate**
  - **max_train_steps**
  - **num_inference_steps**
  - **guidance_scale**

Hints: Custom Diffusion

- Suggestion on tuning hyperparameters
  - **instance_prompt:**
    - You must include “<new1>” in the sentence
    - You would like to include information related to the object to customize
      e.g., If object to customize is a black dog, try to use “photo of a <new1> **black dog**”
    - When inference, the input text prompts should be changed correspondingly
  - **parameter_to_train**: More parameters to train usually leads to much similar generated object to reference (DINO score↑). But less control during inference (CLIP-T↓).
  - **max_train_steps**: More training steps usually leads to much similar generated object to reference (DINO score↑). But less control during inference (CLIP-T↓).

Submission: NTU COOL

- 4 points in total, submit before deadline: 2025/06/13 23:59:59 (UTC+8)
- **No late submission is allowed**
- Submit your **code** to NTU COOL. (4 points)
  - We can only see your last submission
  - Compress your code into **<student_ID>_hw10.zip** (e.g. b13901001_hw10.zip)
  - After TAs unzip your **<student_ID>_hw10.zip**, all your files should locate under a directory called **<student_ID>_hw10**
  - The directory **MUST** include a **README** and **codes** that can reproduce your results. **Do NOT** submit the model checkpoint or dataset and other unrelated files.

Submission: NTU COOL

- Structure of the zipped file:
  - <student_ID>_hw10
    - <student_ID>_hw10_1.ipynb or .py or .sh
    - <student_ID>_hw10_2.ipynb or .py or .sh
    - ...
    - README.md or .txt
- Examples for valid structure of the zipped file:
  - b13901001_hw10
    - b13901001_hw10_1.ipynb
    - README.md
  - b13901001_hw10
    - b13901001_hw10_1.py
    - b13901001_hw10_2.py
    - b13901001_hw10_3.sh
    - README.txt

Submission: NTU COOL

- How to write a README?
  - Specify your environment(colab, kaggle...) and GPU(T4, T4*2, P100...)
  - List all references used to finish the homework.
    - Which part of code is generated by which model(GPT, Gemini, Grok...). Shared link for the chat is better.
    - Website link, NTU Cool discussion, Offline discussion with classmates(Student IDs)...
  - If you run the code in your environment instead of colab or kaggle.
    - Specify the python version.
    - Provide a requirements.txt for additional installed packages.
  - If you decompose sample code into multiple scripts.
    - Specify the function of each file.
    - Provide a step-by-step instruction for running your scripts with correct commands and execution order.
  - If you have no idea.
    - Ask README Generator.

Submission: JudgeBoi

- 6 points in total, submit before deadline: 2025/06/13 23:59:59 (UTC+8)
- **No late submission is allowed**
- Submit your **generated images** to JudgeBoi (6 points)
  - See **Eval Metrics** and **Grading** sections for details about how grading is done
  - **5** submission quota per day, reset at **23:59 (UTC+8)**
  - Compress generated images into a **zipped file (< 8MB)**
  - After unzip, there should be 6 directories (object-1, object-2, ...,object-6) each contains 15 images (0.jpg, 1.jpg, ..., 14.jpg) each has resolution 512x512

Submission: JudgeBoi

- Structure of the zipped file:
  - zip_file_name
    - object-1
      - 0.jpg
      - 1.jpg
      - ...
      - 14.jpg
    - object-2
      - 0.jpg
      - ...
    - ...
    - object-6
      - 0.jpg
      - ...

Regulations

- You should NOT use data that is not provided by TA
- You should NOT plagiarize, if you use any other resource, you should cite it in the reference.
- You should NOT modify your prediction files manually.
- Do NOT share codes or prediction files with any living creatures.
- Do NOT use any approaches to submit your results more than 5 times a day.
- Please protect your own work and ensure that your answers are not accessible to others. If your work is found to have been copied by others, you will be subject to the same penalties.
- Your final grade x 0.9 and get a score 0 for that homework if you violate any of the above rules first time (within a semester)
- Your will get F for the final grade if you violate any of the above rules multiple times (within a semester).
- Prof. Lee & TAs preserve the rights to change the rules & grades

If any questions, you can ask us via...

- NTU COOL (Recommended)
- Email
  - ntu-ml-2025-spring-ta@googlegroups.com
  - The title should begin with “[HW10]”
- TA hours (Online)
  - Link: https://meet.google.com/vqe-crcb-bjf
  - Each Friday before 2025/06/13
  - 13:30 - 14:10; 17:00 - 18:00