Exercise 2 Report

資工四 陳昱瑋 409410118

Introduction

這次的作業主要是熟悉Diffusion model的實作與應用，透過Huggingface Hub的Pretrained Model取得訓練好的模型，使用python實際使用這些模型，對其下關鍵字(prompt)並取得生成圖片。這次作業有三個任務:

1. Text2Img主要是利用10個prompts產出10張圖。
2. Img2Img則是使用Cityscapes datasets中的10張real images，透過Dreambooth產生1000張fake images，並使用FID進行評分。
3. Generation Annotations則是使用這1000 fake images，透過於Exercise 1 trained model (我使用Resnet-50)，使用detectron2進行object detection，將inference後的annotatoins以COCO-format的形式儲存下來。

Text2Img

- How to run Diffusers

  Python code text2img.py:

  ```
  from diffusers import DiffusionPipeline
  import torch

  pipe = DiffusionPipeline.from_pretrained(
      "pretrained model path",
      torch_dtype=torch.float16,
      use_safetensors=True
  )

  image = base("prompt", height=512, width=512).images[0]
  image.save("result.png")
  ```

  Then run `python text2img.py`

- Prompts
  - "A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, daytime. The pedestrians pass by the street. the cars stop in front of the pedestrians. The traffic light is red, so the pedestrians can pass through the street. Some modern architectures along the street."
  - "A view to a street from a car's front window. ISO400, high-quality, utra-hd, utra-detail, realistic, daytime, hyperrealistic. The pedestrians walk along the pavement. The buildings, street lamps and trees are along the street. The sky is blue. There're some road signs beside the traffic light and street."
  - "A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, hyperrealistic, daytime. The buildings, street lamps and trees are along the street. The traffic light is green. The scene is at the intersection, cars, trucks, vans are cross the intersection."

- "A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, hyperrealistic, daytime. A mejastic museum at the right side of the street. There's few pedestrians on the sidewalk. Some parking slot along the raod. A police man is on the sidewalk. There's a bus stop at the bus stand, people get off the bus.”
- "A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, hyperrealistic, daytime. There's multiple shop, supermarket, convience store, etc. Some road sign along the sidewalk. the pedestrians fill the sidewalk. the shop sign is colorful.",
- "A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, hyperrealistic, daytime. A lot of trees, and the road is pave with leaves. The street lamp is along the road, no pedestrians."
- "A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, hyperrealistic, daytime. Person is going to ride on the taxi. Some cars, hotel, buildings. Trees along the road, and the railing along the sidewalk. Some pedestrians walk on the sidewalk"
- "A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, hyperrealistic, daytime. There're a vanding machine, and an ATM. The motocycle stop at the intersection, and the traffic light is green. A little graffiti on the building's wall, not too much. Few pedestrians."
- "A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, hyperrealistic, daytime. The bicycle is on the sidewalk. The traffic light is red. The trees are along the street. The buildings at both side of street. The cars stop before the pedestrians",
- "A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, hyperrealistic, daytime. The bus stop at the bus stand, people ride on a bus. A lots of pedestsrians crowd on the sidewalk. The shop sign is on building. Some trash can along the sidewalk."

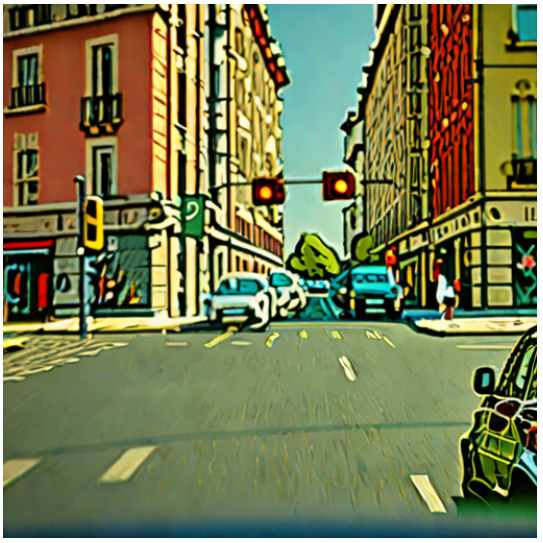
- Different experiment settings
  - Model
    - ***stable-diffusion-v1-5***
    - ***stable-diffusion-2-1***
    - ***sdxl-turbo***
- Experiment results & CLIP Score
  - 由於stable-diffusion-v1-5與sdxl-turbo的結果在CLIP Score分數上較高，故選擇這兩個Model作為結果比較
  - CLIP Score
    - ***stable-diffusion-v1-5***

      CLIP score: 25.2806
    - ***sdxl-turbo***

      CLIP score: 26.3205

- stable-diffusion-v1-5
- ***sdxl-turbo***

"A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, daytime. The pedestrians pass by the street. the cars stop in front of the pedestrians. The traffic light is red, so the pedestrians can pass through the street. Some modern architectures along the street."



"A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, daytime. The pedestrians pass by the street. the cars stop in front of the pedestrians. The traffic light is red, so the pedestrians can pass through the street. Some modern architectures along the street."

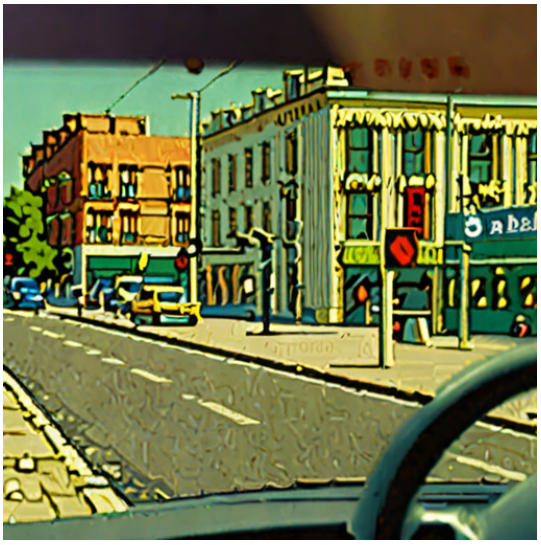
"A view to a street from a car's front window. ISO400, high-quality, utra-hd, utra-detail, realistic, daytime, hyperrealistic. The pedestrians walk along the pavement. The buildings, street lamps and trees are along the street. The sky is blue. There're some road signs beside the traffic light and street."

"A view to a street from a car's front window. ISO400, high-quality, utra-hd, utra-detail, realistic, daytime, hyperrealistic. The pedestrians walk along the pavement. The buildings, street lamps and trees are along the street. The sky is blue. There're some road signs beside the traffic light and street."

"A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, hyperrealistic, daytime. The buildings, street lamps and trees are along the street. The traffic light is green. The scene is at the intersection, cars, trucks, vans are cross the intersection."

"A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, hyperrealistic, daytime. The buildings, street lamps and trees are along the street. The traffic light is green. The scene is at the intersection, cars, trucks, vans are cross the intersection."

"A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, hyperrealistic, daytime. A mejastic museum at the right side of the street. There's few pedestrians on the sidewalk. Some parking slot along the raod. A police man is on the sidewalk. There's a bus stop at the bus stand, people get off the bus.”



"A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, hyperrealistic, daytime. A mejastic museum at the right side of the street. There's few pedestrians on the sidewalk. Some parking slot along the raod. A police man is on the sidewalk. There's a bus stop at the bus stand, people get off the bus.”

"A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, hyperrealistic, daytime. There's multiple shop, supermarket, convience store, etc. Some road sign along the sidewalk. the pedestrians fill the sidewalk. the shop sign is colorful.",

"A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, hyperrealistic, daytime. There's multiple shop, supermarket, convience store, etc. Some road sign along the sidewalk. the pedestrians fill the sidewalk. the shop sign is colorful.",

"A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, hyperrealistic, daytime. A lot of trees, and the road is pave with leaves. The street lamp is along the road, no pedestrians."

"A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, hyperrealistic, daytime. A lot of trees, and the road is pave with leaves. The street lamp is along the road, no pedestrians."

"A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, hyperrealistic, daytime. Person is going to ride on the taxi. Some cars, hotel, buildings. Trees along the road, and the railing along the sidewalk. Some pedestrians walk on the sidewalk"

"A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, hyperrealistic, daytime. Person is going to ride on the taxi. Some cars, hotel, buildings. Trees along the road, and the railing along the sidewalk. Some pedestrians walk on the sidewalk"

"A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, hyperrealistic, daytime. There're a vanding machine, and an ATM. The motocycle stop at the intersection, and the traffic light is green. A little graffiti on the building's wall, not too much. Few pedestrians."

"A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, hyperrealistic, daytime. There're a vanding machine, and an ATM. The motocycle stop at the intersection, and the traffic light is green. A little graffiti on the building's wall, not too much. Few pedestrians."

"A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, hyperrealistic, daytime. The bicycle is on the sidewalk. The traffic light is red. The trees are along the street. The buildings at both side of street. The cars stop before the pedestrians",

"A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, hyperrealistic, daytime. The bicycle is on the sidewalk. The traffic light is red. The trees are along the street. The buildings at both side of street. The cars stop before the pedestrians",

"A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, hyperrealistic, daytime. The bus stop at the bus stand, people ride on a bus. A lots of pedestsrians crowd on the sidewalk. The shop sign is on building. Some trash can along the sidewalk."

"A view to a street from a car's front window. ISO400, high-quality, utra-hd, realistic, hyperrealistic, daytime. The bus stop at the bus stand, people ride on a bus. A lots of pedestsrians crowd on the sidewalk. The shop sign is on building. Some trash can along the sidewalk."

Img2Img

- Methods
  - ***Dreambooth***
- Experiment settings
  - Model

    - ***stable-diffusion-v1-5***
  - Using accelerate to run the ***Dreambooth***

```
export MODEL_NAME="runwayml/stable-diffusion-v1-5"
export INSTANCE_DIR="../../../cityscapes"
export OUTPUT_DIR="../../../model_v3"
export DREAMBOOTH_OUTPUT="../../../dreambooth_output_v3"

accelerate launch train_dreambooth.py \
  --pretrained_model_name_or_path=$MODEL_NAME  \
  --instance_data_dir=$INSTANCE_DIR \
  --output_dir=$OUTPUT_DIR \
  --instance_prompt="A view to a street from a car's front window. ISO4
00, high-quality, utra-hd, realistic, hyperrealistic" \
  --resolution=512 \
  --train_batch_size=1 \
  --gradient_accumulation_steps=1 \
  --learning_rate=5e-6 \
  --lr_scheduler="constant" \
  --lr_warmup_steps=0 \
  --max_train_steps=400 \
  --with_prior_preservation \
  --prior_loss_weight=1.0 \
  --class_data_dir=$DREAMBOOTH_OUTPUT \
  --class_prompt="The pedestians passby the street. the cars stop in fr
ont of the pedestrains. The traffic light is red or green. Some modern
architectures along the street." \
  --snr_gamma=5.0
```

    - 使用 **Prior preservation loss** 讓model可以根據generative image來知道自己已經產生了什麼影像對Model進行微調
    - 使用 **Min-SNR weighting** 可以透過rebalancing讓training process更快的收斂
- Experiment results & FID Score
  - According to the documentation page

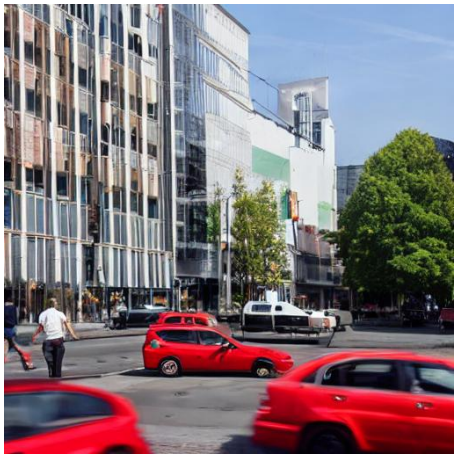
    > *A good practice to run the evaluation across different seeds and inference steps, and then report an average result.*

    我對1000張圖做random取10張圖進行評估，得到以下10比FID Scores

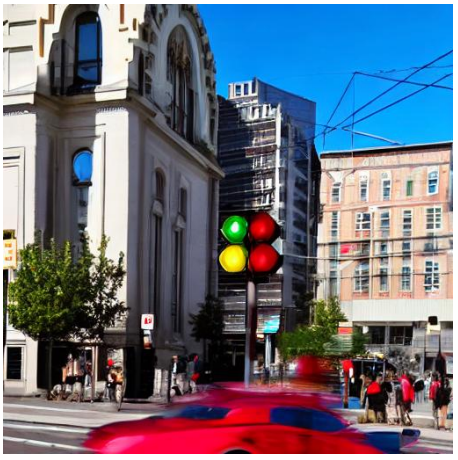
    1. FID: 287.7755126953125
    2. FID: 274.1945495605469
    3. FID: 287.4744567871094
    4. FID: 277.3009948730469
    5. FID: 264.44195556640625
    6. FID: 277.9429016113281

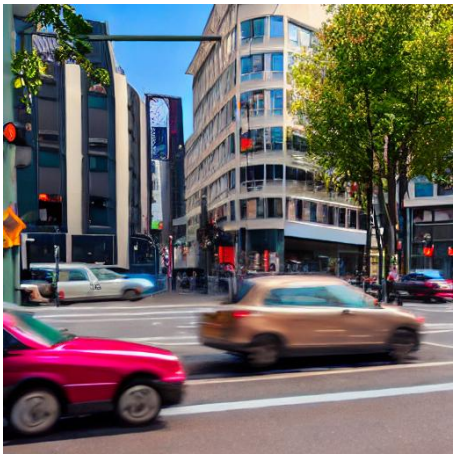
7. FID: 245.8429718017578832519531
8. FID: 273.6518859863281
9. FID: 310.86737060546875
10. FID: 293.3287658691406
    - highest: 318.34, lowest: 264.44
    - **Average FID≈304.74**

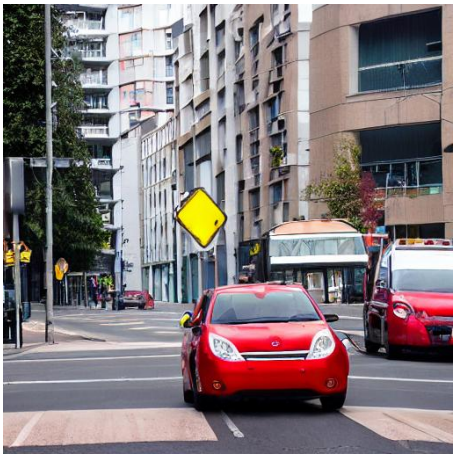
- 318.34's fake image

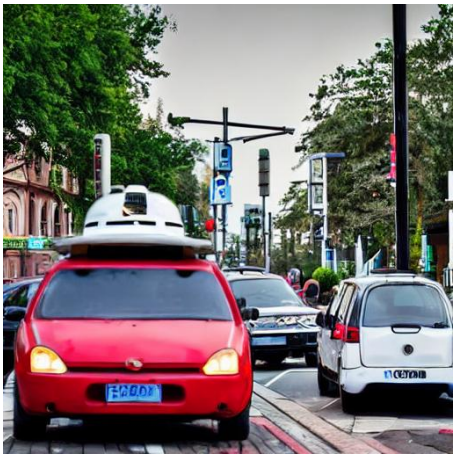


- 264.44's fake image









- If we generate the fake images by the ***Dreambooth*** trained model (base model: "*runwayml/stable-diffusion-v1-5*"), the **FID become 210.27438354492188**

- real images

A view to a street from a car's front window. A majestic, big, classic museum in front the view. Some flower decorated in front of the museum. A black car stop in front of the museum. Two statues stand in the front door of museum.

A view to a street from a car's front window. The scene is at the intersection. However, in the view in front of us, is the pavement for pedestrians. both side are the buildings. Some graffiti on the building at the right side. Some cars stop at the pavement.

- fake images (generated images)

Generate annotations from the previous object detector

- Method
  - 根據 https://haobin-tan.netlify.app/ai/computer-vision/object-detection/coco-dataset-format/, the COCO format:

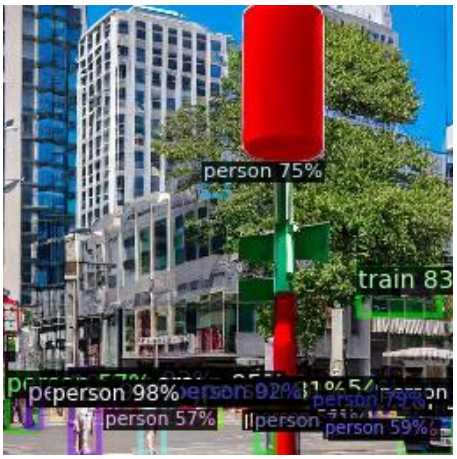
```
{
    "info": {
        "Attach any information you want"
    },
    "categories": [
        {
            "id": 0,
            "name": "Person",
            "supercategory": "Object"
        }, ...
    ],
    "images": [
        {
            "id": 0,
            "file_name": "0001.jpg",
            "height": 275,
            "width": 490,
        }
    ],
    "annotations": [
        {
            "id": 0,
            "image_id": 0,
            "category_id": 2,
            "bbox": [45, 2, 85, 85],
            "area": 7225,
            "iscrowd": 0
        },...
    ]
}
```

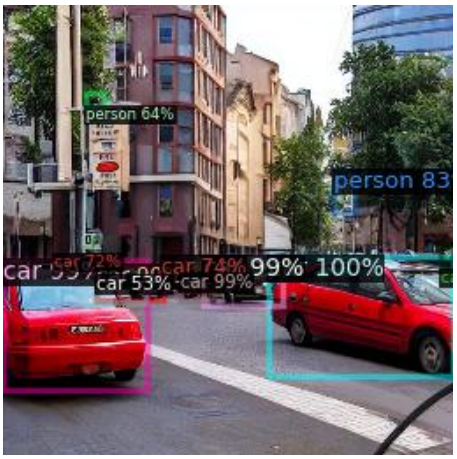
在inference.py中，將img2img2的1000張output為input進行inference。

在這之後使用 `def convert_to_coco` 將model prediction output轉成COCO format。

> Detectron2 model output format: https://detectron2.readthedocs.io/en/latest/tutorials/models.html#model-output-format

- few sample images with bounding box annotations.





Discussion

About text2img

- Memory Usage

  執行diffuser的過程很常會碰到CUDA out of memory，故需要根據documentation page去將資源進行分配，reduce memory usage。在這裡使用了 `enable_sequential_cpu_offload()` ，主要是將weight offloading到CPU上面，只有當在forward pass的時候才將這些weight loading到GPU上面。在如此方面可以讓memory usage降低到3GB。

  然而 trade-off 就是在inference的時候速度會很慢，由於Diffusers是透過iterative去產出圖，weight在CPU與GPU之間的轉換將會降低效率。

  在這樣的情況下可以嘗試使用 `enable_model_cpu_offload()` ，透過full-model offloading到CPU上，只有main components of pipeline會loading到GPU。此方法可以兼顧memory usage與inference速度問題。

- `torch_dtype=torch.float16` produce blank black image

  為了加速產圖的速度，於本機時嘗試使用 `float16` 進行加速，然而不管怎麼試總是產出全黑的圖。經過查詢，發現是nvidia cuda 對 GTX 1650 Ti 優化問題。之後到國網中心 jupyter 運行就沒有問題了。

  - https://github.com/huggingface/diffusers/issues/2153
  - Only can solved by using `float32` (default setting)

About img2img

- Memory Usage

  由於img2img的記憶體要求量十分的高，只能透過 `enable_sequential_offload()` 對memory進行優化，同時必須開啟 `accelerate` 的優化項目 `deepspeed` (https://huggingface.co/docs/transformers/deepspeed)。

  根據Documentation:

  > ***DeepSpeed*** *is a PyTorch optimization library that makes distributed training memory-efficient and fast.*
  > *- ZeRO-1, optimizer state partioning across GPUs*
  > *- ZeRO-2, gradient partitioning across GPUs*
  > *- ZeRO-3, parameteter partitioning across GPUs*

  在這邊使用了 ***ZeRO-2*** 進行了優化。