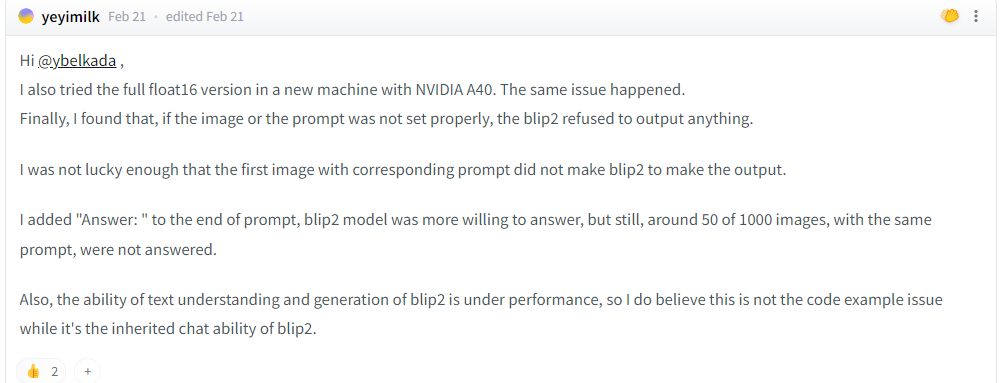
Homework #3: Data Augmentation

Image Captioning:

因為GPU內存的關係這邊我使用Salesforce/blip2-opt-2.7b以及Salesforce/blip2-flan-t5-xl生成每個圖像的描述。這邊我設的問題如下:

**Please describe the content of this image in detail.**

我最後選擇blip2-flan-t5-xl模型輸出的因為它生成的文字比較穩定且正確，在使用Salesforce/blip2-opt-2.7b時我也發現到如果不給他Answer:的結尾或開頭設定Question他會不輸出文字，這個問題我也有在他的hugging face討論發現



後來我參考她的方式就能成功輸出了，但猶豫blip2-flan-t5-xl還是比較穩定且正確，所以我最後選擇他生成的描述。

Prompt Design:

這邊我參考助教PPT上的prompt以及一點點小小的更改製作prompt\_w\_label以及prompt\_w\_suffix，prompt\_w\_label主要就是把generated\_text後面加上you should have 那張圖的label，prompt\_w\_suffix就是把prompt\_w\_label加上一些東西，範例如下。

範例:

generated\_text:woman in warehouse holding clipboard

**Prompt Template #1(**prompt\_w\_label**)**:

woman in warehouse holding clipboard. imagesize:512x512.Image should have Head, Person, Ear, Face, Hands

**Prompt Template #2(**prompt\_w\_suffix**)**:

woman in warehouse holding clipboard. imagesize:512x512.Image should have Head, Person, Ear, Face, Hands.HD quality, highly detailed and real.

Text-to-Image Generation

這邊生成圖片我使用masterful/gligen-1-4-generation-text-box生成圖片，使用剛剛製作的prompt\_w\_label以及prompt\_w\_suffix只用text給他生成圖片，然後用FID衡量生成圖像與真實圖像之間的相似度，結果如下表格

結果發現Template #1的FID是比較好的，所以我使用Template #1加上layout重新生成圖片，最後結果如下表格:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Text grounding** | | **Layout-to-Image** |
| **prompt** | Template #1 | Template #2 | Template #1 |
| **FID** | 54.931862 | 55.7761929541 | 60.06578 |