2023 SAMSUNG AI CHALLENGE :
IMAGE QUALITY ASSESSMENT – CAPTIONING SOLUTION

인하대학교 허건혁

목차

- Model - Blip

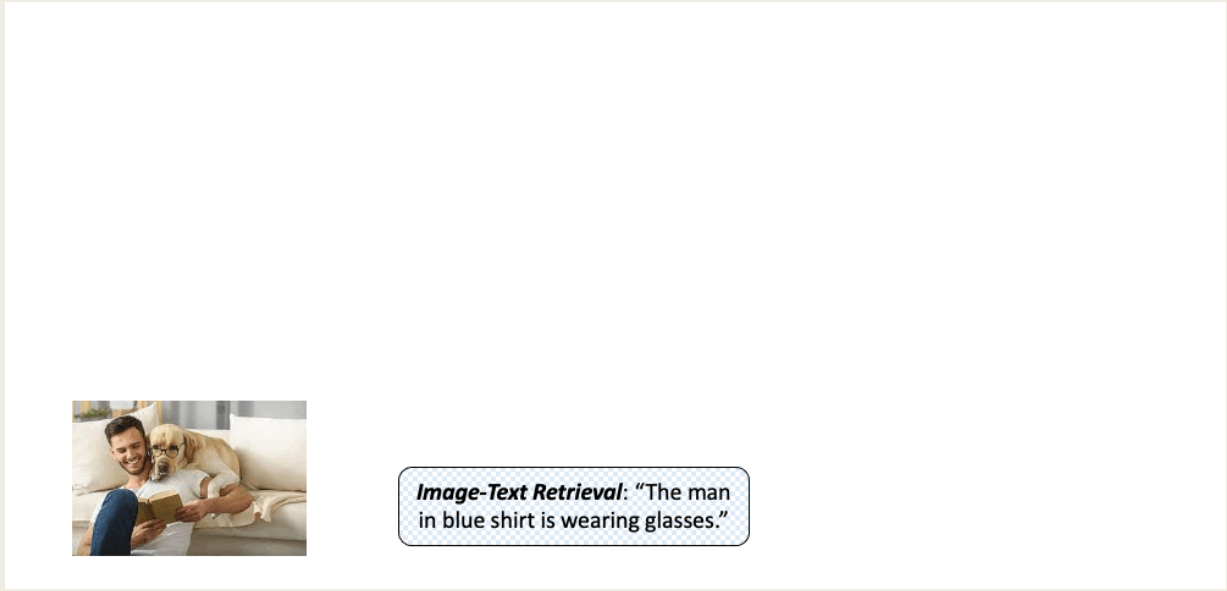
- Strategy

1. Weight Sampling
2. Diffusion Image
3. LoRa
4. Voting Ensemble

Model : Blip

Pretrain Model

- Base & Large : pretrained on COCO dataset



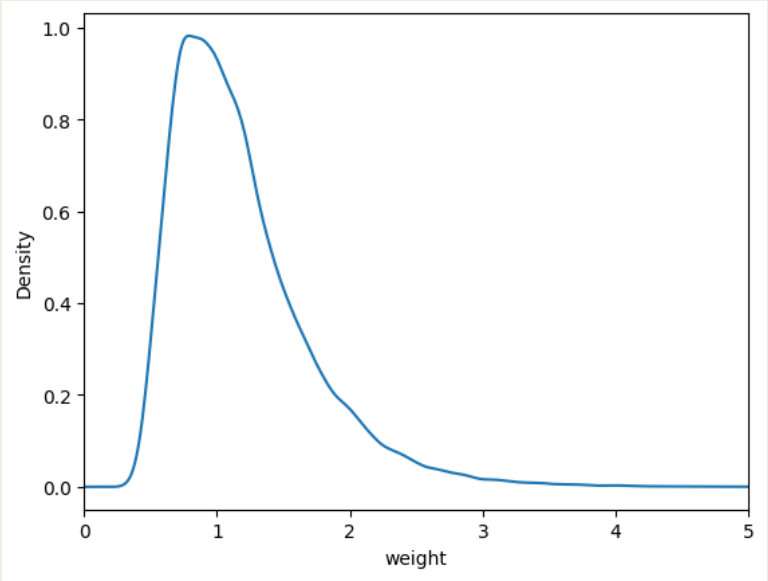
[참조]
https://huggingface.co/Salesforce/blip-image-captioning-base
https://huggingface.co/Salesforce/blip-image-captioning-large

Strategy - 1 Weight Sampling

Weight는 훈련 데이터의 연속적인 단어의 빈도 N-gram을 통하여
큰 값일 수록 좋은 comments라 판단하였다.

Epoch 마다 전체 데이터 셋의 90%를 weight를 가중치로 샘플링 하였다

$$weight = \frac{len(s)}{\sum_{n\in[2,3,4]} \log(n\text{-}gram\text{-}freq)},\ s \in \text{Data}$$



- Comments Sample

High_Weight

weight - 2.86 : i dont know how i missed this photo. it is amazing. both animals look so noble and sure of themselves, and so comfortable with the other being around. i love photography like this. and i love all that white!

weight - 2.03 : on balance i think you achieved what you were after, and i concur with the others that it was a bold statement certainly against the grain.

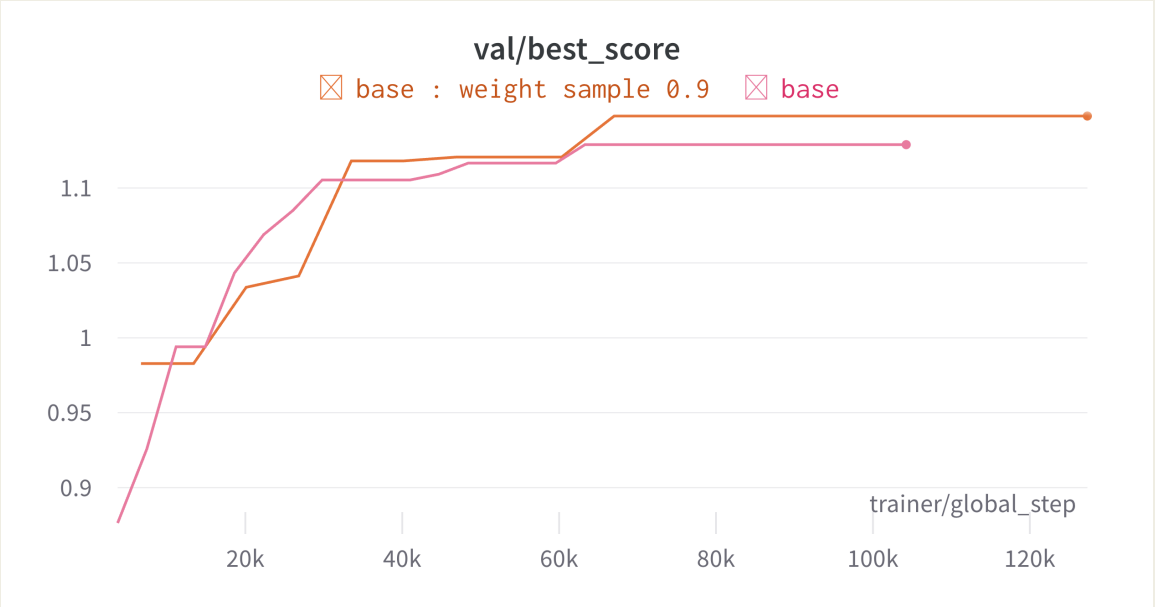
Low_Weight

weight - 0.49 : nice use of colour and lines.

weight - 0.97 : nice pastel tones. pleasant photo.

Strategy - 1 Weight Sampling

Base model 기준
1.129 -> 1.148로 점수가 향상되었다.



Strategy - 2 Diffusion Image

- Text : "that white and pink is so gorgeous, awesome picture”
- Image :

Diffusion 모델을 활용하여
Text comments로부터 Train Image를 생성

Case	Valid	Public
Base	1.129	1.102
Base + diffusion	1.121	1.149

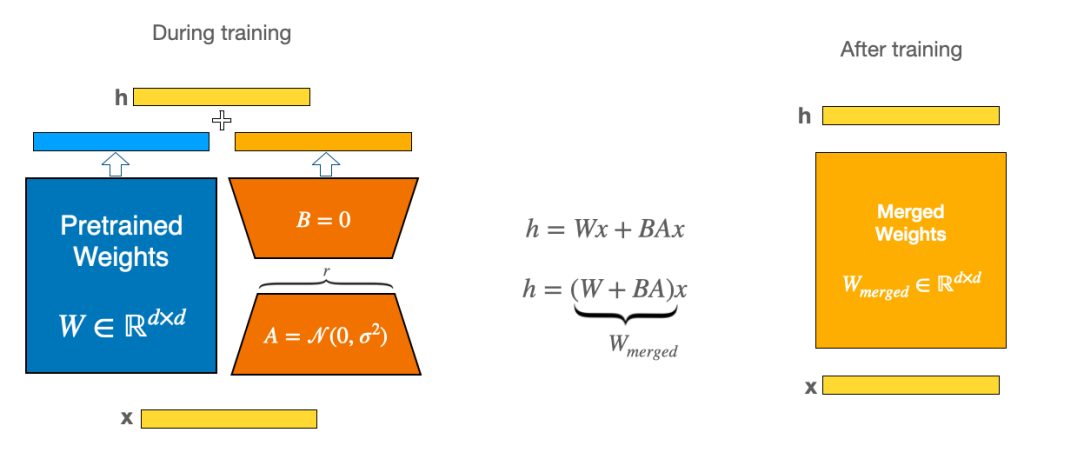
Diffusion image 9642장을 생성하여 모델을 훈련 결과
valid와 public 평균 점수가 소폭 향상

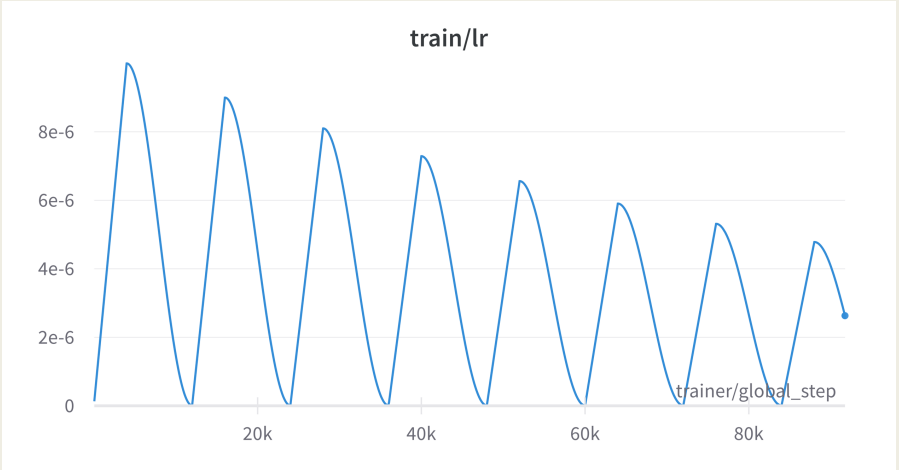
[참조]
https://huggingface.co/runwayml/stable-diffusion-v1-5

Strategy - 3 LoRa

사전 학습 파라미터를 보존하기 위하여
LoRa & Warmup cosine annealing 사용

학습 속도 및 Batch_Size를 키우기 위해 FP16 학습






```
train_module: ["crossattention", "text_decoder.cls"]
lora_module: [vision_model, attention.self.query, attention.self.value]
```

[참조]
https://github.com/huggingface/peft

Strategy - 4 Voting Ensemble

```
j00zs3u6dr,6.282756328585,"hdr is way overdone, in my opinion"
pzhsjj4uxo,7.126460552215001,"a little too dark for my taste, but still a great image."
```

```
j00zs3u6dr,2.2018399325078972,"hdr is way overdone, in my opinion"
pzhsjj4uxo,4.9868235476749945,"a reasonable use of hdr, and an engaging photo, in my opinion"
```

서로 다른 모델에 대하여 동일한 캡션을 생성한 이미지에 대하여
Score값을 Valid, Public 점수의 평균을 가중치로 Voting Ensemble

Public 점수가 1.3126 -> 1.3595 점수 향상