

# UBAI PRACTICE - Captioning

hs.hwang

2024-11-04

# Table of contents

<b>1</b>	<b>Image Captioning</b>	<b>3</b>
1.	.....	3
	.....	3
	.....	4
2.	.....	4
3.	.....	6

# 1 Image Captioning

Python , (Image Captioning) .  
 ,  
 , “A black dog sitting among leaves in a forest, surrounded by trees.(  
.)” .



A black dog sitting among  
leaves in a forest,  
surrounded by trees.

1.

,  
 ,  
 . captioning . Terminal , cd (captioning  
) .

```
pwd
cd
```

```
• (captioning) [ssu@gate1 ~]$ pwd
/home1/ssu
• (captioning) [ssu@gate1 ~]$ cd captioning
• (captioning) [ssu@gate1 captioning]$ pwd
/home1/ssu/captioning
```

```
conda create -n captioning python=3.8
conda activate captioning
```

```
• (base) [ssu@gate1 captioning]$ conda activate captioning
○ (captioning) [ssu@gate1 captioning]$
```

```
pip install torch torchvision transformers matplotlib jupyter
```

## 2.

### Microsoft COCO ( MS COCO)

MS COCO Object detection( ), Segmentation( ), Captioning ,

MS COCO shell

```
#!/bin/bash

# COCO dataset directory
mkdir -p /data/coco

# Download COCO Train2014 images and captions
cd /data/coco
wget http://images.cocodataset.org/zips/train2014.zip
wget http://images.cocodataset.org/zips/val2014.zip
wget http://images.cocodataset.org/annotations/annotations_trainval2014.zip

# Unzip the dataset
unzip train2014.zip
unzip val2014.zip
unzip annotations_trainval2014.zip
```

```
mkdir      data          . mkdir          .
cd          .
wget       MS COCO dataset          ,          .
unzip      dataset zip          ,          .
          , shell          .
```

```
❌ (base) [ssu@gate1 captioning]$ ./dataset_download.sh
bash: ./dataset_download.sh: Permission denied
✅ (base) [ssu@gate1 captioning]$ chmod +x dataset_download.sh
```

```
dataset_download.sh      shell      , “permission denied ( )”      .
chmod                      .
chmod                      ,          .
```

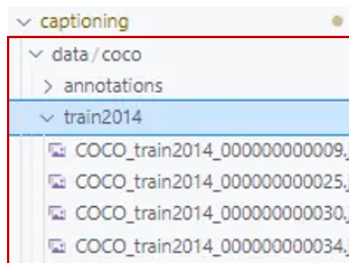
```
chmod [references] [operator] [modes] file1 ...
```

operator	role
r	(read)
w	(write)
x	(execute)

```
chmod +x [file_name.sh] +x [file_name.sh] .
```

```
(base) [ssu@gate1 captioning]$ ./dataset_download.sh
bash: ./dataset_download.sh: Permission denied
(base) [ssu@gate1 captioning]$ chmod +x dataset_download.sh
```

```
.
,
```



### 3.

```
,
```

**Transformer** . Transformer 2017 Google , .

transformer.py .

slurm . , HPC slurm transformer.sh .

```
#!/bin/bash
#SBATCH --job-name=captioning
#SBATCH --output=./output/training_captioning_%n_%j.out
#SBATCH --error=./output/training_captioning_%n_%j.err
#SBATCH --nodes=2
#SBATCH --partition=gpu3
#SBATCH --gres=gpu:4
#SBATCH --cpus-per-task=16
#SBATCH --mem=128G
#SBATCH --time=24:00:00

echo "start at:" `date` #
echo "node: $HOSTNAME" #
echo "jobid: $SLURM_JOB_ID" # jobid

# Load modules
module load cuda/11.8

# Train the transformer-based image captioning model
python transformer.py
```

```
#SBATCH --job-name=captioning job-name captioning .
output error output training_captioning .
#SBATCH --nodes=2 , node 2 . , node 2 .
#SBATCH --gres=gpu:4 gpu 4 .
module load cuda/11.8 module cuda 11.8 version .
python transformer.py transformer.py .
sbatch transformer.sh (job) .
tqdm , error .
```

```
14080 Epoch 1: 54% ██████████ | 7039/12942 [1:01:55<50:41, 1.94it/s, loss=0.05]
14081 Epoch 1: 54% ██████████ | 7039/12942 [1:01:55<50:41, 1.94it/s, loss=0.00199]
14082 Epoch 1: 54% ██████████ | 7040/12942 [1:01:55<51:08, 1.92it/s, loss=0.00199]
14083 Epoch 1: 54% ██████████ | 7040/12942 [1:01:56<51:08, 1.92it/s, loss=0.0475]
14084 Epoch 1: 54% ██████████ | 7041/12942 [1:01:56<51:23, 1.91it/s, loss=0.0475]
14085 Epoch 1: 54% ██████████ | 7041/12942 [1:01:56<51:23, 1.91it/s, loss=0.00561]
```

out log , .

“a shoe rack with some shoes and a dog sleeping on them” caption .



Captioning .