```
gpu_info = !nvidia-smi
gpu_info = '\n'.join(gpu_info)
if gpu_info.find('failed') >= 0:
  print('Not connected to a GPU')
else:
  print(gpu_info)
```

```
Wed May  7 04:06:46 2025
+-----------------------------------------------------------------------------
--------------------+
| NVIDIA-SMI 550.54.15              Driver Version: 550.54.15
CUDA Version: 12.4     |
|-----------------------------------------+------------------------
+----------------------+
| GPU  Name                 Persistence-M | Bus-Id          Disp.A |
Volatile Uncorr. ECC |
| Fan  Temp   Perf          Pwr:Usage/Cap |           Memory-Usage |
GPU-Util  Compute M. |
|                                         |                        |
MIG M. |
|
=========================================+========================+===
====================|
|   0  NVIDIA A100-SXM4-40GB          Off |   00000000:00:04.0 Off |
0 |
| N/A   30C    P0             46W /  400W |       0MiB /  40960MiB |
0%      Default |
|                                         |                        |
Disabled |
+-----------------------------------------+------------------------
+----------------------+


+-----------------------------------------------------------------------------
--------------------+
| Processes:
|
|  GPU   GI   CI        PID   Type   Process name
GPU Memory |
|        ID   ID
Usage      |
|
=============================================================================
====================|
|  No running processes found
|
+-----------------------------------------------------------------------------
--------------------+
```

```
from psutil import virtual_memory
ram_gb = virtual_memory().total / 1e9
print('Your runtime has {:.1f} gigabytes of available RAM\
n'.format(ram_gb))

if ram_gb < 20:
  print('Not using a high-RAM runtime')
else:
  print('You are using a high-RAM runtime!')
```

```
Your runtime has 89.6 gigabytes of available RAM

You are using a high-RAM runtime!
```

```
!git clone https://github.com/salaniz/pycocoevalcap.git
%cd pycocoevalcap
!pip install .
%cd ..
```

```
Cloning into 'pycocoevalcap'...
remote: Enumerating objects: 821, done.ote: Counting objects: 100%
(12/12), done.ote: Compressing objects: 100% (9/9), done.ote: Total
821 (delta 4), reused 3 (delta 3), pack-reused 809 (from 2)etadata
(setup.py) ... ent already satisfied: pycocotools>=2.0.2 in
/usr/local/lib/python3.11/dist-packages (from pycocoevalcap==1.2)
(2.0.8)
Requirement already satisfied: matplotlib>=2.1.0 in
/usr/local/lib/python3.11/dist-packages (from pycocotools>=2.0.2-
>pycocoevalcap==1.2) (3.10.0)
Requirement already satisfied: numpy in
/usr/local/lib/python3.11/dist-packages (from pycocotools>=2.0.2-
>pycocoevalcap==1.2) (2.0.2)
Requirement already satisfied: contourpy>=1.0.1 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools>=2.0.2->pycocoevalcap==1.2) (1.3.2)
Requirement already satisfied: cycler>=0.10 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools>=2.0.2->pycocoevalcap==1.2) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools>=2.0.2->pycocoevalcap==1.2) (4.57.0)
Requirement already satisfied: kiwisolver>=1.3.1 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools>=2.0.2->pycocoevalcap==1.2) (1.4.8)
Requirement already satisfied: packaging>=20.0 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools>=2.0.2->pycocoevalcap==1.2) (24.2)
Requirement already satisfied: pillow>=8 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools>=2.0.2->pycocoevalcap==1.2) (11.2.1)
```

```
Requirement already satisfied: pyparsing>=2.3.1 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools>=2.0.2->pycocoevalcap==1.2) (3.2.3)
Requirement already satisfied: python-dateutil>=2.7 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools>=2.0.2->pycocoevalcap==1.2) (2.9.0.post0)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.7-
>matplotlib>=2.1.0->pycocotools>=2.0.2->pycocoevalcap==1.2) (1.17.0)
Building wheels for collected packages: pycocoevalcap
  Building wheel for pycocoevalcap (setup.py) ... e=pycocoevalcap-1.2-
py3-none-any.whl size=104312245
sha256=97150f49da2a49ac7f54cb9c585f2f19123ba3ca09fce2c2325373241338d23
2
  Stored in directory:
/tmp/pip-ephem-wheel-cache-z6bnfjgd/wheels/0e/98/9f/b6578f2310a0adf702
387edf950a2ba69dbf680c0b6830b312
Successfully built pycocoevalcap
Installing collected packages: pycocoevalcap
Successfully installed pycocoevalcap-1.2
/content
```

Blip model(Base code)

```
# Install Hugging Face Transformers (for BLIP-2) and other libraries
!pip install transformers
!pip install torch torchvision
!pip install datasets
!pip install sentencepiece
!pip install evaluate
!pip install git+https://github.com/jmhessel/clipscore.git

Requirement already satisfied: transformers in
/usr/local/lib/python3.11/dist-packages (4.51.3)
Requirement already satisfied: filelock in
/usr/local/lib/python3.11/dist-packages (from transformers) (3.18.0)
Requirement already satisfied: huggingface-hub<1.0,>=0.30.0 in
/usr/local/lib/python3.11/dist-packages (from transformers) (0.30.2)
Requirement already satisfied: numpy>=1.17 in
/usr/local/lib/python3.11/dist-packages (from transformers) (2.0.2)
Requirement already satisfied: packaging>=20.0 in
/usr/local/lib/python3.11/dist-packages (from transformers) (24.2)
Requirement already satisfied: pyyaml>=5.1 in
/usr/local/lib/python3.11/dist-packages (from transformers) (6.0.2)
Requirement already satisfied: regex!=2019.12.17 in
/usr/local/lib/python3.11/dist-packages (from transformers)
(2024.11.6)
Requirement already satisfied: requests in
```

```
/usr/local/lib/python3.11/dist-packages (from transformers) (2.32.3)
Requirement already satisfied: tokenizers<0.22,>=0.21 in
/usr/local/lib/python3.11/dist-packages (from transformers) (0.21.1)
Requirement already satisfied: safetensors>=0.4.3 in
/usr/local/lib/python3.11/dist-packages (from transformers) (0.5.3)
Requirement already satisfied: tqdm>=4.27 in
/usr/local/lib/python3.11/dist-packages (from transformers) (4.67.1)
Requirement already satisfied: fsspec>=2023.5.0 in
/usr/local/lib/python3.11/dist-packages (from huggingface-
hub<1.0,>=0.30.0->transformers) (2025.3.2)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
/usr/local/lib/python3.11/dist-packages (from huggingface-
hub<1.0,>=0.30.0->transformers) (4.13.2)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
(3.4.1)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
(3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
(2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
(2025.4.26)
Requirement already satisfied: torch in
/usr/local/lib/python3.11/dist-packages (2.6.0+cu124)
Requirement already satisfied: torchvision in
/usr/local/lib/python3.11/dist-packages (0.21.0+cu124)
Requirement already satisfied: filelock in
/usr/local/lib/python3.11/dist-packages (from torch) (3.18.0)
Requirement already satisfied: typing-extensions>=4.10.0 in
/usr/local/lib/python3.11/dist-packages (from torch) (4.13.2)
Requirement already satisfied: networkx in
/usr/local/lib/python3.11/dist-packages (from torch) (3.4.2)
Requirement already satisfied: jinja2 in
/usr/local/lib/python3.11/dist-packages (from torch) (3.1.6)
Requirement already satisfied: fsspec in
/usr/local/lib/python3.11/dist-packages (from torch) (2025.3.2)
Collecting nvidia-cuda-nvrtc-cu12==12.4.127 (from torch)
  Downloading nvidia_cuda_nvrtc_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cuda-runtime-cu12==12.4.127 (from torch)
  Downloading nvidia_cuda_runtime_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cuda-cupti-cu12==12.4.127 (from torch)
  Downloading nvidia_cuda_cupti_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-cudnn-cu12==9.1.0.70 (from torch)
```

```
  Downloading nvidia_cudnn_cu12-9.1.0.70-py3-none-
manylinux2014_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-cublas-cu12==12.4.5.8 (from torch)
  Downloading nvidia_cublas_cu12-12.4.5.8-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cufft-cu12==11.2.1.3 (from torch)
  Downloading nvidia_cufft_cu12-11.2.1.3-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-curand-cu12==10.3.5.147 (from torch)
  Downloading nvidia_curand_cu12-10.3.5.147-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cusolver-cu12==11.6.1.9 (from torch)
  Downloading nvidia_cusolver_cu12-11.6.1.9-py3-none-
manylinux2014_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-cusparse-cu12==12.3.1.170 (from torch)
  Downloading nvidia_cusparse_cu12-12.3.1.170-py3-none-
manylinux2014_x86_64.whl.metadata (1.6 kB)
Requirement already satisfied: nvidia-cusparselt-cu12==0.6.2 in
/usr/local/lib/python3.11/dist-packages (from torch) (0.6.2)
Requirement already satisfied: nvidia-nccl-cu12==2.21.5 in
/usr/local/lib/python3.11/dist-packages (from torch) (2.21.5)
Requirement already satisfied: nvidia-nvtx-cu12==12.4.127 in
/usr/local/lib/python3.11/dist-packages (from torch) (12.4.127)
Collecting nvidia-nvjitlink-cu12==12.4.127 (from torch)
  Downloading nvidia_nvjitlink_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Requirement already satisfied: triton==3.2.0 in
/usr/local/lib/python3.11/dist-packages (from torch) (3.2.0)
Requirement already satisfied: sympy==1.13.1 in
/usr/local/lib/python3.11/dist-packages (from torch) (1.13.1)
Requirement already satisfied: mpmath<1.4,>=1.1.0 in
/usr/local/lib/python3.11/dist-packages (from sympy==1.13.1->torch)
(1.3.0)
Requirement already satisfied: numpy in
/usr/local/lib/python3.11/dist-packages (from torchvision) (2.0.2)
Requirement already satisfied: pillow!=8.3.*,>=5.3.0 in
/usr/local/lib/python3.11/dist-packages (from torchvision) (11.2.1)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.11/dist-packages (from jinja2->torch) (3.0.2)
Downloading nvidia_cublas_cu12-12.4.5.8-py3-none-
manylinux2014_x86_64.whl (363.4 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 363.4/363.4 MB 2.9 MB/s eta
0:00:00
anylinux2014_x86_64.whl (13.8 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 13.8/13.8 MB 128.5 MB/s eta
0:00:00
anylinux2014_x86_64.whl (24.6 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 24.6/24.6 MB 103.0 MB/s eta
0:00:00
```

```
e_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl (883 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 883.7/883.7 kB 56.9 MB/s eta
0:00:00
anylinux2014_x86_64.whl (664.8 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 664.8/664.8 MB 1.7 MB/s eta
0:00:00
anylinux2014_x86_64.whl (211.5 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 211.5/211.5 MB 11.3 MB/s eta
0:00:00
anylinux2014_x86_64.whl (56.3 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 56.3/56.3 MB 40.3 MB/s eta
0:00:00
anylinux2014_x86_64.whl (127.9 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 127.9/127.9 MB 18.9 MB/s eta
0:00:00
anylinux2014_x86_64.whl (207.5 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 207.5/207.5 MB 5.0 MB/s eta
0:00:00
anylinux2014_x86_64.whl (21.1 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 21.1/21.1 MB 96.4 MB/s eta
0:00:00
e-cu12, nvidia-cuda-nvrtc-cu12, nvidia-cuda-cupti-cu12, nvidia-cublas-
cu12, nvidia-cusparse-cu12, nvidia-cudnn-cu12, nvidia-cusolver-cu12
  Attempting uninstall: nvidia-nvjitlink-cu12
    Found existing installation: nvidia-nvjitlink-cu12 12.5.82
    Uninstalling nvidia-nvjitlink-cu12-12.5.82:
      Successfully uninstalled nvidia-nvjitlink-cu12-12.5.82
  Attempting uninstall: nvidia-curand-cu12
    Found existing installation: nvidia-curand-cu12 10.3.6.82
    Uninstalling nvidia-curand-cu12-10.3.6.82:
      Successfully uninstalled nvidia-curand-cu12-10.3.6.82
  Attempting uninstall: nvidia-cufft-cu12
    Found existing installation: nvidia-cufft-cu12 11.2.3.61
    Uninstalling nvidia-cufft-cu12-11.2.3.61:
      Successfully uninstalled nvidia-cufft-cu12-11.2.3.61
  Attempting uninstall: nvidia-cuda-runtime-cu12
    Found existing installation: nvidia-cuda-runtime-cu12 12.5.82
    Uninstalling nvidia-cuda-runtime-cu12-12.5.82:
      Successfully uninstalled nvidia-cuda-runtime-cu12-12.5.82
  Attempting uninstall: nvidia-cuda-nvrtc-cu12
    Found existing installation: nvidia-cuda-nvrtc-cu12 12.5.82
    Uninstalling nvidia-cuda-nvrtc-cu12-12.5.82:
      Successfully uninstalled nvidia-cuda-nvrtc-cu12-12.5.82
  Attempting uninstall: nvidia-cuda-cupti-cu12
    Found existing installation: nvidia-cuda-cupti-cu12 12.5.82
    Uninstalling nvidia-cuda-cupti-cu12-12.5.82:
      Successfully uninstalled nvidia-cuda-cupti-cu12-12.5.82
  Attempting uninstall: nvidia-cublas-cu12
    Found existing installation: nvidia-cublas-cu12 12.5.3.2
```

```
    Uninstalling nvidia-cublas-cu12-12.5.3.2:
      Successfully uninstalled nvidia-cublas-cu12-12.5.3.2
  Attempting uninstall: nvidia-cusparse-cu12
    Found existing installation: nvidia-cusparse-cu12 12.5.1.3
    Uninstalling nvidia-cusparse-cu12-12.5.1.3:
      Successfully uninstalled nvidia-cusparse-cu12-12.5.1.3
  Attempting uninstall: nvidia-cudnn-cu12
    Found existing installation: nvidia-cudnn-cu12 9.3.0.75
    Uninstalling nvidia-cudnn-cu12-9.3.0.75:
      Successfully uninstalled nvidia-cudnn-cu12-9.3.0.75
  Attempting uninstall: nvidia-cusolver-cu12
    Found existing installation: nvidia-cusolver-cu12 11.6.3.83
    Uninstalling nvidia-cusolver-cu12-11.6.3.83:
      Successfully uninstalled nvidia-cusolver-cu12-11.6.3.83
Successfully installed nvidia-cublas-cu12-12.4.5.8 nvidia-cuda-cupti-
cu12-12.4.127 nvidia-cuda-nvrtc-cu12-12.4.127 nvidia-cuda-runtime-
cu12-12.4.127 nvidia-cudnn-cu12-9.1.0.70 nvidia-cufft-cu12-11.2.1.3
nvidia-curand-cu12-10.3.5.147 nvidia-cusolver-cu12-11.6.1.9 nvidia-
cusparse-cu12-12.3.1.170 nvidia-nvjitlink-cu12-12.4.127
Collecting datasets
  Downloading datasets-3.6.0-py3-none-any.whl.metadata (19 kB)
Requirement already satisfied: filelock in
/usr/local/lib/python3.11/dist-packages (from datasets) (3.18.0)
Requirement already satisfied: numpy>=1.17 in
/usr/local/lib/python3.11/dist-packages (from datasets) (2.0.2)
Requirement already satisfied: pyarrow>=15.0.0 in
/usr/local/lib/python3.11/dist-packages (from datasets) (18.1.0)
Collecting dill<0.3.9,>=0.3.0 (from datasets)
  Downloading dill-0.3.8-py3-none-any.whl.metadata (10 kB)
Requirement already satisfied: pandas in
/usr/local/lib/python3.11/dist-packages (from datasets) (2.2.2)
Requirement already satisfied: requests>=2.32.2 in
/usr/local/lib/python3.11/dist-packages (from datasets) (2.32.3)
Requirement already satisfied: tqdm>=4.66.3 in
/usr/local/lib/python3.11/dist-packages (from datasets) (4.67.1)
Collecting xxhash (from datasets)
  Downloading xxhash-3.5.0-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (12 kB)
Collecting multiprocess<0.70.17 (from datasets)
  Downloading multiprocess-0.70.16-py311-none-any.whl.metadata (7.2
kB)
Collecting fsspec<=2025.3.0,>=2023.1.0 (from
fsspec[http]<=2025.3.0,>=2023.1.0->datasets)
  Downloading fsspec-2025.3.0-py3-none-any.whl.metadata (11 kB)
Requirement already satisfied: huggingface-hub>=0.24.0 in
/usr/local/lib/python3.11/dist-packages (from datasets) (0.30.2)
Requirement already satisfied: packaging in
/usr/local/lib/python3.11/dist-packages (from datasets) (24.2)
Requirement already satisfied: pyyaml>=5.1 in
```

```
/usr/local/lib/python3.11/dist-packages (from datasets) (6.0.2)
Requirement already satisfied: aiohttp!=4.0.0a0,!=4.0.0a1 in
/usr/local/lib/python3.11/dist-packages (from
fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (3.11.15)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
/usr/local/lib/python3.11/dist-packages (from huggingface-hub>=0.24.0-
>datasets) (4.13.2)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests>=2.32.2-
>datasets) (3.4.1)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.11/dist-packages (from requests>=2.32.2-
>datasets) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.11/dist-packages (from requests>=2.32.2-
>datasets) (2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.11/dist-packages (from requests>=2.32.2-
>datasets) (2025.4.26)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.11/dist-packages (from pandas->datasets)
(2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in
/usr/local/lib/python3.11/dist-packages (from pandas->datasets)
(2025.2)
Requirement already satisfied: tzdata>=2022.7 in
/usr/local/lib/python3.11/dist-packages (from pandas->datasets)
(2025.2)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (2.6.1)
Requirement already satisfied: aiosignal>=1.1.2 in
/usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (1.3.2)
Requirement already satisfied: attrs>=17.3.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (25.3.0)
Requirement already satisfied: frozenlist>=1.1.1 in
/usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (1.6.0)
Requirement already satisfied: multidict<7.0,>=4.5 in
/usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (6.4.3)
Requirement already satisfied: propcache>=0.2.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (0.3.1)
Requirement already satisfied: yarl<2.0,>=1.17.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (1.20.0)
```

```
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2-
>pandas->datasets) (1.17.0)
Downloading datasets-3.6.0-py3-none-any.whl (491 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 491.5/491.5 kB 10.3 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 116.3/116.3 kB 12.8 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 193.6/193.6 kB 17.5 MB/s eta
0:00:00
ultiprocess-0.70.16-py311-none-any.whl (143 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 143.5/143.5 kB 13.5 MB/s eta
0:00:00
anylinux_2_17_x86_64.manylinux2014_x86_64.whl (194 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 194.8/194.8 kB 19.5 MB/s eta
0:00:00
ultiprocess, datasets
  Attempting uninstall: fsspec
    Found existing installation: fsspec 2025.3.2
    Uninstalling fsspec-2025.3.2:
      Successfully uninstalled fsspec-2025.3.2
ERROR: pip's dependency resolver does not currently take into account
all the packages that are installed. This behaviour is the source of
the following dependency conflicts.
gcsfs 2025.3.2 requires fsspec==2025.3.2, but you have fsspec 2025.3.0
which is incompatible.
Successfully installed datasets-3.6.0 dill-0.3.8 fsspec-2025.3.0
multiprocess-0.70.16 xxhash-3.5.0
Requirement already satisfied: sentencepiece in
/usr/local/lib/python3.11/dist-packages (0.2.0)
Collecting evaluate
  Downloading evaluate-0.4.3-py3-none-any.whl.metadata (9.2 kB)
Requirement already satisfied: datasets>=2.0.0 in
/usr/local/lib/python3.11/dist-packages (from evaluate) (3.6.0)
Requirement already satisfied: numpy>=1.17 in
/usr/local/lib/python3.11/dist-packages (from evaluate) (2.0.2)
Requirement already satisfied: dill in /usr/local/lib/python3.11/dist-
packages (from evaluate) (0.3.8)
Requirement already satisfied: pandas in
/usr/local/lib/python3.11/dist-packages (from evaluate) (2.2.2)
Requirement already satisfied: requests>=2.19.0 in
/usr/local/lib/python3.11/dist-packages (from evaluate) (2.32.3)
Requirement already satisfied: tqdm>=4.62.1 in
/usr/local/lib/python3.11/dist-packages (from evaluate) (4.67.1)
Requirement already satisfied: xxhash in
/usr/local/lib/python3.11/dist-packages (from evaluate) (3.5.0)
Requirement already satisfied: multiprocess in
/usr/local/lib/python3.11/dist-packages (from evaluate) (0.70.16)
Requirement already satisfied: fsspec>=2021.05.0 in
```

```
/usr/local/lib/python3.11/dist-packages (from fsspec[http]>=2021.05.0-
>evaluate) (2025.3.0)
Requirement already satisfied: huggingface-hub>=0.7.0 in
/usr/local/lib/python3.11/dist-packages (from evaluate) (0.30.2)
Requirement already satisfied: packaging in
/usr/local/lib/python3.11/dist-packages (from evaluate) (24.2)
Requirement already satisfied: filelock in
/usr/local/lib/python3.11/dist-packages (from datasets>=2.0.0-
>evaluate) (3.18.0)
Requirement already satisfied: pyarrow>=15.0.0 in
/usr/local/lib/python3.11/dist-packages (from datasets>=2.0.0-
>evaluate) (18.1.0)
Requirement already satisfied: pyyaml>=5.1 in
/usr/local/lib/python3.11/dist-packages (from datasets>=2.0.0-
>evaluate) (6.0.2)
Requirement already satisfied: aiohttp!=4.0.0a0,!=4.0.0a1 in
/usr/local/lib/python3.11/dist-packages (from fsspec[http]>=2021.05.0-
>evaluate) (3.11.15)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
/usr/local/lib/python3.11/dist-packages (from huggingface-hub>=0.7.0-
>evaluate) (4.13.2)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests>=2.19.0-
>evaluate) (3.4.1)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.11/dist-packages (from requests>=2.19.0-
>evaluate) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.11/dist-packages (from requests>=2.19.0-
>evaluate) (2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.11/dist-packages (from requests>=2.19.0-
>evaluate) (2025.4.26)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.11/dist-packages (from pandas->evaluate)
(2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in
/usr/local/lib/python3.11/dist-packages (from pandas->evaluate)
(2025.2)
Requirement already satisfied: tzdata>=2022.7 in
/usr/local/lib/python3.11/dist-packages (from pandas->evaluate)
(2025.2)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]>=2021.05.0->evaluate) (2.6.1)
Requirement already satisfied: aiosignal>=1.1.2 in
/usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]>=2021.05.0->evaluate) (1.3.2)
Requirement already satisfied: attrs>=17.3.0 in
```

```
/usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]>=2021.05.0->evaluate) (25.3.0)
Requirement already satisfied: frozenlist>=1.1.1 in
/usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]>=2021.05.0->evaluate) (1.6.0)
Requirement already satisfied: multidict<7.0,>=4.5 in
/usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]>=2021.05.0->evaluate) (6.4.3)
Requirement already satisfied: propcache>=0.2.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]>=2021.05.0->evaluate) (0.3.1)
Requirement already satisfied: yarl<2.0,>=1.17.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]>=2021.05.0->evaluate) (1.20.0)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2-
>pandas->evaluate) (1.17.0)
Downloading evaluate-0.4.3-py3-none-any.whl (84 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 84.0/84.0 kB 2.4 MB/s eta
0:00:00
/jmhessel/clipscore.git
  Cloning https://github.com/jmhessel/clipscore.git to /tmp/pip-req-
build-7dwfjok3
  Running command git clone --filter=blob:none --quiet
https://github.com/jmhessel/clipscore.git /tmp/pip-req-build-7dwfjok3
  Resolved https://github.com/jmhessel/clipscore.git to commit
10364652765136 21f77f1c2208d742e4a430781f
ERROR: git+https://github.com/jmhessel/clipscore.git does not appear
to be a Python project: neither 'setup.py' nor 'pyproject.toml' found.
```

```
from transformers import BlipProcessor, BlipForConditionalGeneration
from PIL import Image
import torch, os, json
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from tqdm import tqdm
import evaluate



# Set the device
device = "cuda" if torch.cuda.is_available() else "cpu"

# Load BLIP processor and model (using the 'blip-image-captioning-
base' model)
processor = BlipProcessor.from_pretrained("Salesforce/blip-image-
captioning-base")
model = BlipForConditionalGeneration.from_pretrained("Salesforce/blip-
image-captioning-base").to(device).eval()
```

```python
# Set paths to the image and annotation directories
images_path = "/content/drive/MyDrive/val2017/"
annotations_path =
"/content/drive/MyDrive/annotations/captions_val2017.json"

# Load annotations
with open(annotations_path, "r") as f:
    coco_data = json.load(f)

image_id_to_filename = {img["id"]: img["file_name"] for img in
coco_data["images"]}
gt_captions = {}
for ann in coco_data["annotations"]:
    img_id = ann["image_id"]
    if img_id not in gt_captions:
        gt_captions[img_id] = []
    gt_captions[img_id].append(ann["caption"])
```

```
Using a slow image processor as `use_fast` is unset and a slow
processor was saved with this model. `use_fast=True` will be the
default behavior in v4.52, even if the model was saved with a slow
processor. This will result in minor differences in outputs. You'll
still be able to use a slow processor with `use_fast=False`.
/usr/local/lib/python3.11/dist-packages/huggingface_hub/utils/_auth.py
:94: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your
settings tab (https://huggingface.co/settings/tokens), set it as
secret in your Google Colab and restart your session.
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to
access public models or datasets.
  warnings.warn(
```


```
{"model_id":"cb4e9aa0969d43b9b001c68b1f61ffdd","version_major":2,"vers
ion_minor":0}

{"model_id":"8f05f2497f5f4e188d84ba7f605d5d86","version_major":2,"vers
ion_minor":0}

{"model_id":"3ff7c517c2c14bcfbae482673a79a08f","version_major":2,"vers
ion_minor":0}

{"model_id":"041bb340773d4beb8b5e3c136dc68a79","version_major":2,"vers
ion_minor":0}

{"model_id":"1c565768bd584cb892256029e11bd315","version_major":2,"vers
ion_minor":0}
```

{"model_id":"21fb9395b9c14bc990d2f8fc11dacd10","version_major":2,"version_minor":0}

{"model_id":"6900c3b1ee1d44c09d50ee59334fac43","version_major":2,"version_minor":0}

{"model_id":"0b04db6c912443acb53f15a9b8435320","version_major":2,"version_minor":0}

```
# Filter valid images
image_ids = list(gt_captions.keys())[:2000]  # Limit to 2000 for your dataset
image_files = [image_id_to_filename[i] for i in image_ids]

# Generate captions
results = []
gen_caps = []
ref_caps = []
ann_id = 0

for idx, (img_id, img_file) in enumerate(tqdm(zip(image_ids, image_files), total=len(image_ids))):
    img_path = os.path.join(images_path, img_file)
    image = Image.open(img_path).convert("RGB")

    # Preprocess and generate captions
    inputs = processor(images=image, return_tensors="pt").to(device)
    generated_ids = model.generate(**inputs, max_new_tokens=30)
    caption = processor.batch_decode(generated_ids, skip_special_tokens=True)[0].strip()

    # Save results
    references = gt_captions[img_id]
    results.append({
        "Image_File": img_file,
        "Generated_Caption": caption,
        "Ground_Truth_Captions": references
    })
    gen_caps.append({"image_id": idx, "caption": caption})
    for ref in references:
        ref_caps.append({"image_id": idx, "id": ann_id, "caption": ref})
        ann_id += 1
```

```
100%|██████████| 2000/2000 [19:18<00:00,  1.73it/s]
```

```
# Save results to CSV
#pd.DataFrame(results).to_csv("/content/drive/MyDrive/generated_results_baseline.csv", index=False)
```

```python
import torch
from transformers import BlipProcessor, BlipForConditionalGeneration
from PIL import Image
import os
import pandas as pd
from tqdm import tqdm
import tempfile  # Import tempfile module
import json
from pycocotools.coco import COCO
from pycocoevalcap.eval import COCOEvalCap

# COCO-style Evaluation
pred_file = tempfile.NamedTemporaryFile(delete=False,
suffix='.json').name
ref_file = tempfile.NamedTemporaryFile(delete=False,
suffix='.json').name

with open(pred_file, "w") as f:
    json.dump(gen_caps, f)

image_list = [{"id": i} for i in range(len(image_ids))]
with open(ref_file, "w") as f:
    json.dump({"annotations": ref_caps, "images": image_list}, f)

coco = COCO(ref_file)
res = coco.loadRes(pred_file)
cocoEval = COCOEvalCap(coco, res)
cocoEval.evaluate()
```

```
loading annotations into memory...
Done (t=0.01s)
creating index...
index created!
Loading and preparing results...
DONE (t=0.00s)
creating index...
index created!
tokenization...
setting up scorers...
Downloading stanford-corenlp-3.6.0 for SPICE ...
Progress: 384.5M / 384.5M (100.0%)
Extracting stanford-corenlp-3.6.0 ...
Done.
computing Bleu score...
{'testlen': 14419, 'reflen': 17704, 'guess': [14419, 12419, 10419,
8419], 'correct': [11716, 6573, 3231, 1510]}
ratio: 0.814448712155399
Bleu_1: 0.647
Bleu_2: 0.522
Bleu_3: 0.407
```

```
Bleu_4: 0.313
computing METEOR score...
METEOR: 0.241
computing Rouge score...
ROUGE_L: 0.535
computing CIDEr score...
CIDEr: 1.011
computing SPICE score...
SPICE: 0.184

# Print evaluation metrics
metrics = {k: round(v, 4) for k, v in cocoEval.eval.items()}
print(metrics)

{'Bleu_1': 0.647, 'Bleu_2': 0.5222, 'Bleu_3': 0.4068, 'Bleu_4':
0.3131, 'METEOR': 0.2407, 'ROUGE_L': np.float64(0.5353), 'CIDEr':
np.float64(1.0105), 'SPICE': np.float64(0.1841)}

#testing

# Set device to CPU
device = "cpu"

# Load BLIP-1 model (Image Captioning model)
processor = BlipProcessor.from_pretrained("Salesforce/blip-image-
captioning-base")
model = BlipForConditionalGeneration.from_pretrained("Salesforce/blip-
image-captioning-base").to(device).eval()

# Path to the testing folder containing your images
testing_images_path = "/content/drive/MyDrive/test2017"

# Get all image files in the testing folder (assuming .jpg or .jpeg
format)
image_files = [f for f in os.listdir(testing_images_path) if
f.endswith('.jpg') or f.endswith('.jpeg')]

# Generate captions
results = []

for idx, img_file in enumerate(tqdm(image_files,
total=len(image_files))):
    img_path = os.path.join(testing_images_path, img_file)
    image = Image.open(img_path).convert("RGB")

    # Preprocess image and generate caption
    inputs = processor(images=image, return_tensors="pt").to(device)
    generated_ids = model.generate(**inputs, max_new_tokens=30)
    caption = processor.batch_decode(generated_ids,
skip_special_tokens=True)[0].strip()
```

```
    # Save results
    results.append({
        "Image_File": img_file,
        "Generated_Caption": caption
    })

# Save results as a CSV file
generated_results_df = pd.DataFrame(results)
generated_results_df.to_csv("/content/drive/MyDrive/generated_testing_
results_blip1.csv", index=False)

# Optionally, you can print the generated captions
for result in results:
    print(f"Image: {result['Image_File']}")
    print(f"Generated Caption: {result['Generated_Caption']}")
    print("-" * 50)
```

```
100%|██████████| 30/30 [00:32<00:00,  1.08s/it]

Image: 000000047056.jpg
Generated Caption: a table with a vase and a vase of flowers
--------------------------------------------------
Image: 000000054498.jpg
Generated Caption: a man riding a motorcycle
--------------------------------------------------
Image: 000000102209.jpg
Generated Caption: a man is playing with a frc
--------------------------------------------------
Image: 000000040965.jpg
Generated Caption: a tray of food
--------------------------------------------------
Image: 000000026680.jpg
Generated Caption: two zebras walking in a field
--------------------------------------------------
Image: 000000181543.jpg
Generated Caption: a surfer is riding a wave in the ocean
--------------------------------------------------
Image: 000000241446.jpg
Generated Caption: a teddy bear sitting on a rail
--------------------------------------------------
Image: 000000023271.jpg
Generated Caption: two women in a field
--------------------------------------------------
Image: 000000488045.jpg
Generated Caption: a bench in the snow
--------------------------------------------------
Image: 000000114063.jpg
Generated Caption: a toilet with a green lid
--------------------------------------------------
Image: 000000151122.jpg
```

```
Generated Caption: a plate of food with a bowl of macaro and a bowl of
macaro
--------------------------------------------------
Image: 000000526846.jpg
Generated Caption: a white and pink carpet
--------------------------------------------------
Image: 000000505458.jpg
Generated Caption: a pan of vegetables
--------------------------------------------------
Image: 000000379473.jpg
Generated Caption: a bird is flying
--------------------------------------------------
Image: 000000110820.jpg
Generated Caption: a blur of a person walking with an umbrella
--------------------------------------------------
Image: 000000526327.jpg
Generated Caption: a snowboarder is doing a trick on a ramp
--------------------------------------------------
Image: 000000304597.jpg
Generated Caption: a man on a surfboard
--------------------------------------------------
Image: 000000130611.jpg
Generated Caption: a zebra standing in the shade
--------------------------------------------------
Image: 000000471220.jpg
Generated Caption: a woman holding a wii game controller
--------------------------------------------------
Image: 000000297369.jpg
Generated Caption: a stone patio with a picnic table and benches
--------------------------------------------------
Image: 000000468714.jpg
Generated Caption: a busy street with a lot of cars and a no parking
sign
--------------------------------------------------
Image: 000000122286.jpg
Generated Caption: a bedroom with a bed, desk, and a computer
--------------------------------------------------
Image: 000000557330.jpg
Generated Caption: a man playing tennis
--------------------------------------------------
Image: 000000004336.jpg
Generated Caption: a cat is sitting on a bed with a pile of papers
--------------------------------------------------
Image: 000000084457.jpg
Generated Caption: a woman playing a video game
--------------------------------------------------
Image: 000000509670.jpg
Generated Caption: a wooden table
--------------------------------------------------
```

```
Image: 000000006069.jpg
Generated Caption: a cat looking at a bird on a feeder
--------------------------------------------------
Image: 000000098660.jpg
Generated Caption: a man sitting at a table with a cell phone
--------------------------------------------------
Image: 000000326419.jpg
Generated Caption: a train on the tracks
--------------------------------------------------
Image: 000000248460.jpg
Generated Caption: a man kneeling down to pick a frck
--------------------------------------------------
```

Finetuning Blip Model

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
#NEW
import os
import json
from PIL import Image
from transformers import BlipProcessor, BlipForConditionalGeneration
from torch.utils.data import Dataset
from transformers import Seq2SeqTrainer, Seq2SeqTrainingArguments
from pycocotools.coco import COCO
from pycocoevalcap.eval import COCOEvalCap
import tempfile
import torch
from pathlib import Path
import pandas as pd
from tqdm import tqdm
import random

# Set the device
device = "cuda" if torch.cuda.is_available() else "cpu"

# Set paths to the data directories
train_images_path = "/content/drive/MyDrive/train2017/"
train_captions_path =
"/content/drive/MyDrive/annotations/captions_train2017.json"
val_images_path = "/content/drive/MyDrive/val2017/"
val_captions_path =
```

```
"/content/drive/MyDrive/annotations/captions_val2017.json"
output_dir = "/content/drive/MyDrive/finetuned/final_model"

import os
os.environ["WANDB_DISABLED"] = "true"

# Load the training annotations
with open(train_captions_path, "r") as f:
    train_coco_captions = json.load(f)

# Load the validation annotations
with open(val_captions_path, "r") as f:
    val_coco_captions = json.load(f)

# Map image IDs to filenames and captions for training data
train_image_id_to_filename = {img['id']: img['file_name'] for img in
train_coco_captions['images']}
train_gt_captions = {}
for annot in train_coco_captions['annotations']:
    image_id = annot['image_id']
    caption = annot['caption']
    if image_id not in train_gt_captions:
        train_gt_captions[image_id] = []
    train_gt_captions[image_id].append(caption)

# Map image IDs to filenames and captions for validation data
val_image_id_to_filename = {img['id']: img['file_name'] for img in
val_coco_captions['images']}
val_gt_captions = {}
for annot in val_coco_captions['annotations']:
    image_id = annot['image_id']
    caption = annot['caption']
    if image_id not in val_gt_captions:
        val_gt_captions[image_id] = []
    val_gt_captions[image_id].append(caption)

# Filter valid image files with captions for training
train_image_files = sorted([f for f in os.listdir(train_images_path)
if f.endswith(".jpg")])
train_valid_image_files, train_valid_image_ids = [], []
for img_id, captions in train_gt_captions.items():
    filename = train_image_id_to_filename.get(img_id)
    if filename and filename in train_image_files:
        train_valid_image_files.append(filename)
        train_valid_image_ids.append(img_id)

# Filter valid image files with captions for validation
val_image_files = sorted([f for f in os.listdir(val_images_path) if
f.endswith(".jpg")])
val_valid_image_files, val_valid_image_ids = [], []
```

```
for img_id, captions in val_gt_captions.items():
    filename = val_image_id_to_filename.get(img_id)
    if filename and filename in val_image_files:
        val_valid_image_files.append(filename)
        val_valid_image_ids.append(img_id)

# Define Dataset Class for Training and Validation
class CocoCaptionDataset(Dataset):
    def __init__(self, image_dir, image_ids, image_filenames,
gt_captions, processor):
        self.image_dir = image_dir
        self.image_ids = image_ids
        self.image_filenames = image_filenames
        self.gt_captions = gt_captions
        self.processor = processor

    def __len__(self):
        return len(self.image_ids)

    def __getitem__(self, idx):
        img_id = self.image_ids[idx]
        img_file = self.image_filenames[idx]
        img_path = os.path.join(self.image_dir, img_file)
        image = Image.open(img_path).convert("RGB")
        caption = random.choice(self.gt_captions[img_id])
        inputs = self.processor(images=image, text=caption,
return_tensors="pt", padding="max_length", truncation=True,
max_length=30)
        inputs = {k: v.squeeze() for k, v in inputs.items()}
        inputs["labels"] = inputs["input_ids"]
        return inputs

# Load the pre-trained model and processor
processor = BlipProcessor.from_pretrained("Salesforce/blip-image-
captioning-base")
model = BlipForConditionalGeneration.from_pretrained("Salesforce/blip-
image-captioning-base").to(device)
```

```
Using a slow image processor as `use_fast` is unset and a slow
processor was saved with this model. `use_fast=True` will be the
default behavior in v4.52, even if the model was saved with a slow
processor. This will result in minor differences in outputs. You'll
still be able to use a slow processor with `use_fast=False`.
/usr/local/lib/python3.11/dist-packages/huggingface_hub/utils/_auth.py
:94: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your
settings tab (https://huggingface.co/settings/tokens), set it as
secret in your Google Colab and restart your session.
You will be able to reuse this secret in all of your notebooks.
```

```
Please note that authentication is recommended but still optional to
access public models or datasets.
  warnings.warn(

{"model_id":"477e3e5f01f94ef9973fd4cca8635fa0","version_major":2,"vers
ion_minor":0}

{"model_id":"21d115bd811d488a9d724a2e88694f59","version_major":2,"vers
ion_minor":0}

{"model_id":"d2f8c5546e68488799b008b4a2f02233","version_major":2,"vers
ion_minor":0}

{"model_id":"1afeec00b26c43059db4f06995bc3576","version_major":2,"vers
ion_minor":0}

{"model_id":"c33ec4eda0344d58b451c922fc731456","version_major":2,"vers
ion_minor":0}

{"model_id":"0ccf04192bf142588ef4e4d665898bff","version_major":2,"vers
ion_minor":0}

{"model_id":"68cbc4ed865a482192ef403267943fe4","version_major":2,"vers
ion_minor":0}

{"model_id":"aca7fd5650bd43bdbfe8d98f91e51e63","version_major":2,"vers
ion_minor":0}

# Prepare the datasets
train_dataset = CocoCaptionDataset(train_images_path,
train_valid_image_ids, train_valid_image_files, train_gt_captions,
processor)
val_dataset = CocoCaptionDataset(val_images_path, val_valid_image_ids,
val_valid_image_files, val_gt_captions, processor)

# Training Arguments
training_args = Seq2SeqTrainingArguments(
    output_dir=output_dir,
    per_device_train_batch_size=16,  # You can increase if GPU memory
allows
    per_device_eval_batch_size=32,   # Large eval batch size for
inference
    num_train_epochs=3,
    learning_rate=3e-5,              # Good learning rate for fine-
tuning
    warmup_steps=500,                # Stabilize training with warmup
    save_steps=1000,                 # Save checkpoints every 1000
steps
    logging_steps=100,               # Log every 100 steps
    eval_steps=500,                  # Evaluate every 500 steps
(adjust if necessary)
```

```
    eval_strategy="steps",     # Evaluate after certain steps
    save_total_limit=2,              # Limit number of saved
checkpoints
    fp16=True,                       # Mixed precision training for
faster and more memory-efficient training
    gradient_accumulation_steps=1,   # No need for gradient
accumulation unless memory is an issue
    dataloader_num_workers=4,       # Adjust based on available CPU
cores
    remove_unused_columns=True,      # Clean up unused columns in the
dataset
    predict_with_generate=True,      # Ensure generation for image
captioning tasks
)

# Trainer Setup
trainer = Seq2SeqTrainer(
    model=model,
    args=training_args,
    train_dataset=train_dataset,
    eval_dataset=val_dataset,
    tokenizer=processor,
)
```

```
<ipython-input-23-6566318c9472>:22: FutureWarning: `tokenizer` is
deprecated and will be removed in version 5.0.0 for
`Seq2SeqTrainer.__init__`. Use `processing_class` instead.
  trainer = Seq2SeqTrainer(
```

```
# Train the model
trainer.train()
print("Training completed.")
```

```
<IPython.core.display.HTML object>
```

```
Training completed.
```

```
# Save the trained model and processor
model.save_pretrained(output_dir)
processor.save_pretrained(output_dir)
```

```
[]
```

```
import torch

model_path = os.path.join(output_dir, 'pytorch_model.bin')
torch.save(model.state_dict(), model_path)
print(f"Model weights saved to {model_path}")
```

```
Model weights saved to
/content/drive/MyDrive/finetuned/final_model/pytorch_model.bin
```

```
# Evaluate on the validation set
results = []
gen_caps = []
ref_caps = []
ann_id = 0
num_samples = min(2000, len(val_valid_image_files))  # Adjust the
number of samples
for idx in tqdm(range(num_samples)):
    img_file = val_valid_image_files[idx]
    img_id = val_valid_image_ids[idx]
    img_path = os.path.join(val_images_path, img_file)
    image = Image.open(img_path).convert("RGB")
    inputs = processor(image, return_tensors="pt").to(device)
    output = model.generate(**inputs)
    caption = processor.decode(output[0], skip_special_tokens=True)
    references = val_gt_captions[img_id]

    results.append({
        "Image_File": img_file,
        "Generated_Caption": caption,
        "Ground_Truth_Captions": references
    })
    gen_caps.append({"image_id": idx, "caption": caption})
    for ref in references:
        ref_caps.append({"image_id": idx, "id": ann_id, "caption":
ref})
        ann_id += 1

100%|██████████| 2000/2000 [09:14<00:00,  3.61it/s]

# Save results to CSV
df = pd.DataFrame(results)
df.to_csv("/content/drive/MyDrive/generated_eval_results.csv",
index=False)

# Run COCOEvalCap for evaluation
pred_file = tempfile.NamedTemporaryFile(delete=False,
suffix='.json').name
ref_file = tempfile.NamedTemporaryFile(delete=False,
suffix='.json').name
with open(pred_file, "w") as f:
    json.dump(gen_caps, f)
image_ids = list({ann["image_id"] for ann in ref_caps})
image_list = [{"id": i} for i in image_ids]
with open(ref_file, "w") as f:
    json.dump({"annotations": ref_caps, "images": image_list}, f)

coco = COCO(ref_file)
res = coco.loadRes(pred_file)
cocoEval = COCOEvalCap(coco, res)
```

```
cocoEval.evaluate()

# Print evaluation scores
scores = cocoEval.eval
scores = {k: round(v, 4) for k, v in scores.items()}
print(scores)
```

```
loading annotations into memory...
Done (t=0.01s)
creating index...
index created!
Loading and preparing results...
DONE (t=0.00s)
creating index...
index created!
tokenization...
setting up scorers...
computing Bleu score...
{'testlen': 19851, 'reflen': 19523, 'guess': [19851, 17851, 15851,
13851], 'correct': [15264, 8677, 4390, 2185]}
ratio: 1.0168006966141978
Bleu_1: 0.769
Bleu_2: 0.611
Bleu_3: 0.470
Bleu_4: 0.357
computing METEOR score...
METEOR: 0.287
computing Rouge score...
ROUGE_L: 0.575
computing CIDEr score...
CIDEr: 1.227
computing SPICE score...
SPICE: 0.220
{'Bleu_1': 0.7689, 'Bleu_2': 0.6114, 'Bleu_3': 0.4695, 'Bleu_4':
0.3575, 'METEOR': 0.2869, 'ROUGE_L': np.float64(0.5754), 'CIDEr':
np.float64(1.2265), 'SPICE': np.float64(0.2202)}
```

```
import os
from PIL import Image
from transformers import BlipProcessor, BlipForConditionalGeneration

# Define the path to the testing folder
testing_images_path = "/content/drive/MyDrive/test2017"  # Replace
with the actual path to your testing folder

# Load the fine-tuned model and processor
model =
BlipForConditionalGeneration.from_pretrained(output_dir).to(device)
processor = BlipProcessor.from_pretrained(output_dir)
```

```python
# List all image files in the testing folder
image_files = [f for f in os.listdir(testing_images_path) if
f.endswith(('.jpg', '.png', '.jpeg'))]

# Function to generate caption for a single image
def generate_caption(image_path):
    # Load and preprocess the image
    image = Image.open(image_path).convert("RGB")
    inputs = processor(image, return_tensors="pt").to(device)

    # Generate a caption
    output = model.generate(**inputs)
    caption = processor.decode(output[0], skip_special_tokens=True)
    return caption

# Iterate over the images in the testing folder and generate captions
for img_file in image_files:
    img_path = os.path.join(testing_images_path, img_file)
    caption = generate_caption(img_path)
    print(f"Caption for {img_file}: {caption}")
```

```
Caption for 000000023271.jpg: two women with umbrellas walk down a
dirt road.
Caption for 000000505458.jpg: a close up of a vegetable dish with
carrots and broccoli.
Caption for 000000526327.jpg: a man in a blue jacket is snowboarding
on a ramp.
Caption for 000000488045.jpg: a bench sitting in the snow near a lake.
Caption for 000000471220.jpg: a man and a woman playing a video game.
Caption for 000000526846.jpg: a woman is walking past a display of
furniture.
Caption for 000000509670.jpg: a yellow vase and a yellow candle on a
table.
Caption for 000000557330.jpg: a man holding a tennis racket on a
court.
Caption for 000000468714.jpg: a busy city street with a lot of
traffic.
Caption for 000000326419.jpg: a red and white train traveling down the
tracks.
Caption for 000000379473.jpg: a bird is sitting on a branch with a sky
background.
Caption for 000000304597.jpg: a man riding a surfboard on top of a
wave.
Caption for 000000114063.jpg: a small toilet with a blue bucket next
to it.
Caption for 000000181543.jpg: a surfer riding a wave in the ocean.
Caption for 000000110820.jpg: a man walking down a street holding an
umbrella.
Caption for 000000130611.jpg: a zebra standing next to another zebra
in a wooded area.
```

```
Caption for 000000248460.jpg: two people are playing with a frisbee in
a field.
Caption for 000000241446.jpg: a teddy bear sitting on a rail with a
train in the background.
Caption for 000000297369.jpg: a park bench in a gravel area with trees
in the background.
Caption for 000000122286.jpg: a bed room with a bed a desk and a
window
Caption for 000000151122.jpg: a plate with a hot dog and a bowl of
macaroni and cheese.
Caption for 000000098660.jpg: a man sitting at a table with a cell
phone.
Caption for 000000102209.jpg: a man standing on a dirt area holding a
frisbee.
Caption for 000000054498.jpg: a man riding a motorcycle down a street.
Caption for 000000084457.jpg: a woman standing next to a man holding a
wii remote.
Caption for 000000006069.jpg: a couple of birds standing on top of a
wooden fence.
Caption for 000000040965.jpg: a tray of food on a table with a cup of
water.
Caption for 000000026680.jpg: two zebras walking in a field with other
zebras in the background.
Caption for 000000004336.jpg: a cat is sitting on the floor next to a
bed.
Caption for 000000047056.jpg: a vase with flowers and other items on a
table.
```

Blip 2 model (Salesforce/blip2-flan-t5-xl-coco) Finetuned on coco

```
!pip install transformers accelerate datasets
!pip install pycocotools pycocoevalcap

import os, json, tempfile
from PIL import Image
import torch
import pandas as pd
from tqdm import tqdm
from pycocotools.coco import COCO
from pycocoevalcap.eval import COCOEvalCap
from torch.utils.data import Dataset


device = "cuda" if torch.cuda.is_available() else "cpu"
```

```
Requirement already satisfied: transformers in
/usr/local/lib/python3.11/dist-packages (4.51.3)
Requirement already satisfied: accelerate in
/usr/local/lib/python3.11/dist-packages (1.6.0)
Collecting datasets
  Downloading datasets-3.5.1-py3-none-any.whl.metadata (19 kB)
Requirement already satisfied: filelock in
/usr/local/lib/python3.11/dist-packages (from transformers) (3.18.0)
Requirement already satisfied: huggingface-hub<1.0,>=0.30.0 in
/usr/local/lib/python3.11/dist-packages (from transformers) (0.30.2)
Requirement already satisfied: numpy>=1.17 in
/usr/local/lib/python3.11/dist-packages (from transformers) (2.0.2)
Requirement already satisfied: packaging>=20.0 in
/usr/local/lib/python3.11/dist-packages (from transformers) (24.2)
Requirement already satisfied: pyyaml>=5.1 in
/usr/local/lib/python3.11/dist-packages (from transformers) (6.0.2)
Requirement already satisfied: regex!=2019.12.17 in
/usr/local/lib/python3.11/dist-packages (from transformers)
(2024.11.6)
Requirement already satisfied: requests in
/usr/local/lib/python3.11/dist-packages (from transformers) (2.32.3)
Requirement already satisfied: tokenizers<0.22,>=0.21 in
/usr/local/lib/python3.11/dist-packages (from transformers) (0.21.1)
Requirement already satisfied: safetensors>=0.4.3 in
/usr/local/lib/python3.11/dist-packages (from transformers) (0.5.3)
Requirement already satisfied: tqdm>=4.27 in
/usr/local/lib/python3.11/dist-packages (from transformers) (4.67.1)
Requirement already satisfied: psutil in
/usr/local/lib/python3.11/dist-packages (from accelerate) (5.9.5)
Requirement already satisfied: torch>=2.0.0 in
/usr/local/lib/python3.11/dist-packages (from accelerate)
(2.6.0+cu124)
Requirement already satisfied: pyarrow>=15.0.0 in
/usr/local/lib/python3.11/dist-packages (from datasets) (18.1.0)
Collecting dill<0.3.9,>=0.3.0 (from datasets)
  Downloading dill-0.3.8-py3-none-any.whl.metadata (10 kB)
Requirement already satisfied: pandas in
/usr/local/lib/python3.11/dist-packages (from datasets) (2.2.2)
Collecting xxhash (from datasets)
  Downloading xxhash-3.5.0-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (12 kB)
Collecting multiprocess<0.70.17 (from datasets)
  Downloading multiprocess-0.70.16-py311-none-any.whl.metadata (7.2
kB)
Collecting fsspec<=2025.3.0,>=2023.1.0 (from
fsspec[http]<=2025.3.0,>=2023.1.0->datasets)
  Downloading fsspec-2025.3.0-py3-none-any.whl.metadata (11 kB)
Requirement already satisfied: aiohttp in
/usr/local/lib/python3.11/dist-packages (from datasets) (3.11.15)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in
```

```
/usr/local/lib/python3.11/dist-packages (from aiohttp->datasets)
(2.6.1)
Requirement already satisfied: aiosignal>=1.1.2 in
/usr/local/lib/python3.11/dist-packages (from aiohttp->datasets)
(1.3.2)
Requirement already satisfied: attrs>=17.3.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp->datasets)
(25.3.0)
Requirement already satisfied: frozenlist>=1.1.1 in
/usr/local/lib/python3.11/dist-packages (from aiohttp->datasets)
(1.6.0)
Requirement already satisfied: multidict<7.0,>=4.5 in
/usr/local/lib/python3.11/dist-packages (from aiohttp->datasets)
(6.4.3)
Requirement already satisfied: propcache>=0.2.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp->datasets)
(0.3.1)
Requirement already satisfied: yarl<2.0,>=1.17.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp->datasets)
(1.20.0)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
/usr/local/lib/python3.11/dist-packages (from huggingface-
hub<1.0,>=0.30.0->transformers) (4.13.2)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
(3.4.1)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
(3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
(2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
(2025.4.26)
Requirement already satisfied: networkx in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (3.4.2)
Requirement already satisfied: jinja2 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (3.1.6)
Collecting nvidia-cuda-nvrtc-cu12==12.4.127 (from torch>=2.0.0-
>accelerate)
  Downloading nvidia_cuda_nvrtc_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cuda-runtime-cu12==12.4.127 (from torch>=2.0.0-
>accelerate)
  Downloading nvidia_cuda_runtime_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cuda-cupti-cu12==12.4.127 (from torch>=2.0.0-
```

```
>accelerate)
  Downloading nvidia_cuda_cupti_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-cudnn-cu12==9.1.0.70 (from torch>=2.0.0->accelerate)
  Downloading nvidia_cudnn_cu12-9.1.0.70-py3-none-
manylinux2014_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-cublas-cu12==12.4.5.8 (from torch>=2.0.0-
>accelerate)
  Downloading nvidia_cublas_cu12-12.4.5.8-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cufft-cu12==11.2.1.3 (from torch>=2.0.0->accelerate)
  Downloading nvidia_cufft_cu12-11.2.1.3-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-curand-cu12==10.3.5.147 (from torch>=2.0.0-
>accelerate)
  Downloading nvidia_curand_cu12-10.3.5.147-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cusolver-cu12==11.6.1.9 (from torch>=2.0.0-
>accelerate)
  Downloading nvidia_cusolver_cu12-11.6.1.9-py3-none-
manylinux2014_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-cusparse-cu12==12.3.1.170 (from torch>=2.0.0-
>accelerate)
  Downloading nvidia_cusparse_cu12-12.3.1.170-py3-none-
manylinux2014_x86_64.whl.metadata (1.6 kB)
Requirement already satisfied: nvidia-cusparselt-cu12==0.6.2 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (0.6.2)
Requirement already satisfied: nvidia-nccl-cu12==2.21.5 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (2.21.5)
Requirement already satisfied: nvidia-nvtx-cu12==12.4.127 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (12.4.127)
Collecting nvidia-nvjitlink-cu12==12.4.127 (from torch>=2.0.0-
>accelerate)
  Downloading nvidia_nvjitlink_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Requirement already satisfied: triton==3.2.0 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (3.2.0)
Requirement already satisfied: sympy==1.13.1 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (1.13.1)
Requirement already satisfied: mpmath<1.4,>=1.1.0 in
/usr/local/lib/python3.11/dist-packages (from sympy==1.13.1-
>torch>=2.0.0->accelerate) (1.3.0)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.11/dist-packages (from pandas->datasets)
(2.9.0.post0)
```

```
Requirement already satisfied: pytz>=2020.1 in
/usr/local/lib/python3.11/dist-packages (from pandas->datasets)
(2025.2)
Requirement already satisfied: tzdata>=2022.7 in
/usr/local/lib/python3.11/dist-packages (from pandas->datasets)
(2025.2)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2-
>pandas->datasets) (1.17.0)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.11/dist-packages (from jinja2->torch>=2.0.0-
>accelerate) (3.0.2)
Downloading datasets-3.5.1-py3-none-any.whl (491 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 491.4/491.4 kB 10.3 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 116.3/116.3 kB 10.7 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 193.6/193.6 kB 18.3 MB/s eta
0:00:00
ultiprocess-0.70.16-py311-none-any.whl (143 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 143.5/143.5 kB 13.6 MB/s eta
0:00:00
anylinux2014_x86_64.whl (363.4 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 363.4/363.4 MB 3.2 MB/s eta
0:00:00
anylinux2014_x86_64.whl (13.8 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 13.8/13.8 MB 118.5 MB/s eta
0:00:00
anylinux2014_x86_64.whl (24.6 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 24.6/24.6 MB 99.1 MB/s eta
0:00:00
e_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl (883 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 883.7/883.7 kB 55.5 MB/s eta
0:00:00
anylinux2014_x86_64.whl (664.8 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 664.8/664.8 MB 1.7 MB/s eta
0:00:00
anylinux2014_x86_64.whl (211.5 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 211.5/211.5 MB 11.6 MB/s eta
0:00:00
anylinux2014_x86_64.whl (56.3 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 56.3/56.3 MB 41.7 MB/s eta
0:00:00
anylinux2014_x86_64.whl (127.9 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 127.9/127.9 MB 19.2 MB/s eta
0:00:00
anylinux2014_x86_64.whl (207.5 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 207.5/207.5 MB 4.1 MB/s eta
0:00:00
anylinux2014_x86_64.whl (21.1 MB)
```

```
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 21.1/21.1 MB 104.9 MB/s eta
0:00:00
anylinux_2_17_x86_64.manylinux2014_x86_64.whl (194 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 194.8/194.8 kB 15.5 MB/s eta
0:00:00
e-cu12, nvidia-cuda-nvrtc-cu12, nvidia-cuda-cupti-cu12, nvidia-cublas-
cu12, fsspec, dill, nvidia-cusparse-cu12, nvidia-cudnn-cu12,
multiprocess, nvidia-cusolver-cu12, datasets
  Attempting uninstall: nvidia-nvjitlink-cu12
    Found existing installation: nvidia-nvjitlink-cu12 12.5.82
    Uninstalling nvidia-nvjitlink-cu12-12.5.82:
      Successfully uninstalled nvidia-nvjitlink-cu12-12.5.82
  Attempting uninstall: nvidia-curand-cu12
    Found existing installation: nvidia-curand-cu12 10.3.6.82
    Uninstalling nvidia-curand-cu12-10.3.6.82:
      Successfully uninstalled nvidia-curand-cu12-10.3.6.82
  Attempting uninstall: nvidia-cufft-cu12
    Found existing installation: nvidia-cufft-cu12 11.2.3.61
    Uninstalling nvidia-cufft-cu12-11.2.3.61:
      Successfully uninstalled nvidia-cufft-cu12-11.2.3.61
  Attempting uninstall: nvidia-cuda-runtime-cu12
    Found existing installation: nvidia-cuda-runtime-cu12 12.5.82
    Uninstalling nvidia-cuda-runtime-cu12-12.5.82:
      Successfully uninstalled nvidia-cuda-runtime-cu12-12.5.82
  Attempting uninstall: nvidia-cuda-nvrtc-cu12
    Found existing installation: nvidia-cuda-nvrtc-cu12 12.5.82
    Uninstalling nvidia-cuda-nvrtc-cu12-12.5.82:
      Successfully uninstalled nvidia-cuda-nvrtc-cu12-12.5.82
  Attempting uninstall: nvidia-cuda-cupti-cu12
    Found existing installation: nvidia-cuda-cupti-cu12 12.5.82
    Uninstalling nvidia-cuda-cupti-cu12-12.5.82:
      Successfully uninstalled nvidia-cuda-cupti-cu12-12.5.82
  Attempting uninstall: nvidia-cublas-cu12
    Found existing installation: nvidia-cublas-cu12 12.5.3.2
    Uninstalling nvidia-cublas-cu12-12.5.3.2:
      Successfully uninstalled nvidia-cublas-cu12-12.5.3.2
  Attempting uninstall: fsspec
    Found existing installation: fsspec 2025.3.2
    Uninstalling fsspec-2025.3.2:
      Successfully uninstalled fsspec-2025.3.2
  Attempting uninstall: nvidia-cusparse-cu12
    Found existing installation: nvidia-cusparse-cu12 12.5.1.3
    Uninstalling nvidia-cusparse-cu12-12.5.1.3:
      Successfully uninstalled nvidia-cusparse-cu12-12.5.1.3
  Attempting uninstall: nvidia-cudnn-cu12
    Found existing installation: nvidia-cudnn-cu12 9.3.0.75
    Uninstalling nvidia-cudnn-cu12-9.3.0.75:
      Successfully uninstalled nvidia-cudnn-cu12-9.3.0.75
  Attempting uninstall: nvidia-cusolver-cu12
    Found existing installation: nvidia-cusolver-cu12 11.6.3.83
```

```
    Uninstalling nvidia-cusolver-cu12-11.6.3.83:
      Successfully uninstalled nvidia-cusolver-cu12-11.6.3.83
ERROR: pip's dependency resolver does not currently take into account
all the packages that are installed. This behaviour is the source of
the following dependency conflicts.
gcsfs 2025.3.2 requires fsspec==2025.3.2, but you have fsspec 2025.3.0
which is incompatible.
Successfully installed datasets-3.5.1 dill-0.3.8 fsspec-2025.3.0
multiprocess-0.70.16 nvidia-cublas-cu12-12.4.5.8 nvidia-cuda-cupti-
cu12-12.4.127 nvidia-cuda-nvrtc-cu12-12.4.127 nvidia-cuda-runtime-
cu12-12.4.127 nvidia-cudnn-cu12-9.1.0.70 nvidia-cufft-cu12-11.2.1.3
nvidia-curand-cu12-10.3.5.147 nvidia-cusolver-cu12-11.6.1.9 nvidia-
cusparse-cu12-12.3.1.170 nvidia-nvjitlink-cu12-12.4.127 xxhash-3.5.0
```


```
{"id":"0f7eb37c5dc5496a9db7670d93b5f7e1","pip_warning":{"packages":
["nvidia"]}}
```


```
Requirement already satisfied: pycocotools in
/usr/local/lib/python3.11/dist-packages (2.0.8)
Requirement already satisfied: pycocoevalcap in
/usr/local/lib/python3.11/dist-packages (1.2)
Requirement already satisfied: matplotlib>=2.1.0 in
/usr/local/lib/python3.11/dist-packages (from pycocotools) (3.10.0)
Requirement already satisfied: numpy in
/usr/local/lib/python3.11/dist-packages (from pycocotools) (2.0.2)
Requirement already satisfied: contourpy>=1.0.1 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools) (1.3.2)
Requirement already satisfied: cycler>=0.10 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools) (4.57.0)
Requirement already satisfied: kiwisolver>=1.3.1 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools) (1.4.8)
Requirement already satisfied: packaging>=20.0 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools) (24.2)
Requirement already satisfied: pillow>=8 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools) (11.2.1)
Requirement already satisfied: pyparsing>=2.3.1 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools) (3.2.3)
Requirement already satisfied: python-dateutil>=2.7 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools) (2.9.0.post0)
Requirement already satisfied: six>=1.5 in
```

```
/usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.7-
>matplotlib>=2.1.0->pycocotools) (1.17.0)

!pip install git+https://github.com/salesforce/LAVIS.git --no-deps

ERROR: Operation cancelled by user
^C

!pip install iopath
!pip install decord


Collecting iopath
  Downloading iopath-0.1.10.tar.gz (42 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 0.0/42.2 kB ? eta -:--:--
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 42.2/42.2 kB 2.0 MB/s eta
0:00:00
etadata (setup.py) ... ent already satisfied: tqdm in
/usr/local/lib/python3.11/dist-packages (from iopath) (4.67.1)
Requirement already satisfied: typing_extensions in
/usr/local/lib/python3.11/dist-packages (from iopath) (4.13.2)
Collecting portalocker (from iopath)
  Downloading portalocker-3.1.1-py3-none-any.whl.metadata (8.6 kB)
Downloading portalocker-3.1.1-py3-none-any.whl (19 kB)
Building wheels for collected packages: iopath
  Building wheel for iopath (setup.py) ... e=iopath-0.1.10-py3-none-
any.whl size=31527
sha256=7c617a4b8172f09ada20ebb97d7dd889cc3dabb34134eb7984d47d5a6be22df
0
  Stored in directory:
/root/.cache/pip/wheels/ba/5e/16/6117f8fe7e9c0c161a795e10d94645ebcf301
ccbd01f66d8ec
Successfully built iopath
Installing collected packages: portalocker, iopath
Successfully installed iopath-0.1.10 portalocker-3.1.1
Collecting decord
  Downloading decord-0.6.0-py3-none-manylinux2010_x86_64.whl.metadata
(422 bytes)
Requirement already satisfied: numpy>=1.14.0 in
/usr/local/lib/python3.11/dist-packages (from decord) (2.0.2)
Downloading decord-0.6.0-py3-none-manylinux2010_x86_64.whl (13.6 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 13.6/13.6 MB 116.2 MB/s eta
0:00:00

!pip install webdataset

Collecting webdataset
  Downloading webdataset-0.2.111-py3-none-any.whl.metadata (15 kB)
Collecting braceexpand (from webdataset)
  Downloading braceexpand-0.1.7-py2.py3-none-any.whl.metadata (3.0 kB)
Requirement already satisfied: numpy in
```

```
/usr/local/lib/python3.11/dist-packages (from webdataset) (2.0.2)
Requirement already satisfied: pyyaml in
/usr/local/lib/python3.11/dist-packages (from webdataset) (6.0.2)
Downloading webdataset-0.2.111-py3-none-any.whl (85 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 85.5/85.5 kB 2.5 MB/s eta
0:00:00

!pip install transformers accelerate datasets
!pip install pycocotools pycocoevalcap

Requirement already satisfied: transformers in
/usr/local/lib/python3.11/dist-packages (4.51.3)
Requirement already satisfied: accelerate in
/usr/local/lib/python3.11/dist-packages (1.6.0)
Requirement already satisfied: datasets in
/usr/local/lib/python3.11/dist-packages (3.5.1)
Requirement already satisfied: filelock in
/usr/local/lib/python3.11/dist-packages (from transformers) (3.18.0)
Requirement already satisfied: huggingface-hub<1.0,>=0.30.0 in
/usr/local/lib/python3.11/dist-packages (from transformers) (0.30.2)
Requirement already satisfied: numpy>=1.17 in
/usr/local/lib/python3.11/dist-packages (from transformers) (2.0.2)
Requirement already satisfied: packaging>=20.0 in
/usr/local/lib/python3.11/dist-packages (from transformers) (24.2)
Requirement already satisfied: pyyaml>=5.1 in
/usr/local/lib/python3.11/dist-packages (from transformers) (6.0.2)
Requirement already satisfied: regex!=2019.12.17 in
/usr/local/lib/python3.11/dist-packages (from transformers)
(2024.11.6)
Requirement already satisfied: requests in
/usr/local/lib/python3.11/dist-packages (from transformers) (2.32.3)
Requirement already satisfied: tokenizers<0.22,>=0.21 in
/usr/local/lib/python3.11/dist-packages (from transformers) (0.21.1)
Requirement already satisfied: safetensors>=0.4.3 in
/usr/local/lib/python3.11/dist-packages (from transformers) (0.5.3)
Requirement already satisfied: tqdm>=4.27 in
/usr/local/lib/python3.11/dist-packages (from transformers) (4.67.1)
Requirement already satisfied: psutil in
/usr/local/lib/python3.11/dist-packages (from accelerate) (5.9.5)
Requirement already satisfied: torch>=2.0.0 in
/usr/local/lib/python3.11/dist-packages (from accelerate)
(2.6.0+cu124)
Requirement already satisfied: pyarrow>=15.0.0 in
/usr/local/lib/python3.11/dist-packages (from datasets) (18.1.0)
Requirement already satisfied: dill<0.3.9,>=0.3.0 in
/usr/local/lib/python3.11/dist-packages (from datasets) (0.3.8)
Requirement already satisfied: pandas in
/usr/local/lib/python3.11/dist-packages (from datasets) (2.2.2)
Requirement already satisfied: xxhash in
/usr/local/lib/python3.11/dist-packages (from datasets) (3.5.0)
```

```
Requirement already satisfied: multiprocess<0.70.17 in
/usr/local/lib/python3.11/dist-packages (from datasets) (0.70.16)
Requirement already satisfied: fsspec<=2025.3.0,>=2023.1.0 in
/usr/local/lib/python3.11/dist-packages (from
fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (2025.3.0)
Requirement already satisfied: aiohttp in
/usr/local/lib/python3.11/dist-packages (from datasets) (3.11.15)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp->datasets)
(2.6.1)
Requirement already satisfied: aiosignal>=1.1.2 in
/usr/local/lib/python3.11/dist-packages (from aiohttp->datasets)
(1.3.2)
Requirement already satisfied: attrs>=17.3.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp->datasets)
(25.3.0)
Requirement already satisfied: frozenlist>=1.1.1 in
/usr/local/lib/python3.11/dist-packages (from aiohttp->datasets)
(1.6.0)
Requirement already satisfied: multidict<7.0,>=4.5 in
/usr/local/lib/python3.11/dist-packages (from aiohttp->datasets)
(6.4.3)
Requirement already satisfied: propcache>=0.2.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp->datasets)
(0.3.1)
Requirement already satisfied: yarl<2.0,>=1.17.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp->datasets)
(1.20.0)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
/usr/local/lib/python3.11/dist-packages (from huggingface-
hub<1.0,>=0.30.0->transformers) (4.13.2)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
(3.4.1)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
(3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
(2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
(2025.4.26)
Requirement already satisfied: networkx in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (3.4.2)
Requirement already satisfied: jinja2 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (3.1.6)
```

```
Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.4.127 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (12.4.127)
Requirement already satisfied: nvidia-cuda-runtime-cu12==12.4.127
in /usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (12.4.127)
Requirement already satisfied: nvidia-cuda-cupti-cu12==12.4.127 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (12.4.127)
Requirement already satisfied: nvidia-cudnn-cu12==9.1.0.70 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (9.1.0.70)
Requirement already satisfied: nvidia-cublas-cu12==12.4.5.8 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (12.4.5.8)
Requirement already satisfied: nvidia-cufft-cu12==11.2.1.3 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (11.2.1.3)
Requirement already satisfied: nvidia-curand-cu12==10.3.5.147 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (10.3.5.147)
Requirement already satisfied: nvidia-cusolver-cu12==11.6.1.9 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (11.6.1.9)
Requirement already satisfied: nvidia-cusparse-cu12==12.3.1.170 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (12.3.1.170)
Requirement already satisfied: nvidia-cusparselt-cu12==0.6.2 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (0.6.2)
Requirement already satisfied: nvidia-nccl-cu12==2.21.5 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (2.21.5)
Requirement already satisfied: nvidia-nvtx-cu12==12.4.127 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (12.4.127)
Requirement already satisfied: nvidia-nvjitlink-cu12==12.4.127 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (12.4.127)
Requirement already satisfied: triton==3.2.0 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (3.2.0)
Requirement already satisfied: sympy==1.13.1 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (1.13.1)
Requirement already satisfied: mpmath<1.4,>=1.1.0 in
/usr/local/lib/python3.11/dist-packages (from sympy==1.13.1-
>torch>=2.0.0->accelerate) (1.3.0)
Requirement already satisfied: python-dateutil>=2.8.2 in
```

```
/usr/local/lib/python3.11/dist-packages (from pandas->datasets)
(2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in
/usr/local/lib/python3.11/dist-packages (from pandas->datasets)
(2025.2)
Requirement already satisfied: tzdata>=2022.7 in
/usr/local/lib/python3.11/dist-packages (from pandas->datasets)
(2025.2)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2-
>pandas->datasets) (1.17.0)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.11/dist-packages (from jinja2->torch>=2.0.0-
>accelerate) (3.0.2)
Requirement already satisfied: pycocotools in
/usr/local/lib/python3.11/dist-packages (2.0.8)
Requirement already satisfied: pycocoevalcap in
/usr/local/lib/python3.11/dist-packages (1.2)
Requirement already satisfied: matplotlib>=2.1.0 in
/usr/local/lib/python3.11/dist-packages (from pycocotools) (3.10.0)
Requirement already satisfied: numpy in
/usr/local/lib/python3.11/dist-packages (from pycocotools) (2.0.2)
Requirement already satisfied: contourpy>=1.0.1 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools) (1.3.2)
Requirement already satisfied: cycler>=0.10 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools) (4.57.0)
Requirement already satisfied: kiwisolver>=1.3.1 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools) (1.4.8)
Requirement already satisfied: packaging>=20.0 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools) (24.2)
Requirement already satisfied: pillow>=8 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools) (11.2.1)
Requirement already satisfied: pyparsing>=2.3.1 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools) (3.2.3)
Requirement already satisfied: python-dateutil>=2.7 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools) (2.9.0.post0)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.7-
>matplotlib>=2.1.0->pycocotools) (1.17.0)
```

```
!git clone https://github.com/salaniz/pycocoevalcap.git
%cd pycocoevalcap
!pip install .
%cd ..
```

```
fatal: destination path 'pycocoevalcap' already exists and is not an
empty directory.
/content/pycocoevalcap
Processing /content/pycocoevalcap
  Preparing metadata (setup.py) ... ent already satisfied:
pycocotools>=2.0.2 in /usr/local/lib/python3.11/dist-packages (from
pycocoevalcap==1.2) (2.0.8)
Requirement already satisfied: matplotlib>=2.1.0 in
/usr/local/lib/python3.11/dist-packages (from pycocotools>=2.0.2-
>pycocoevalcap==1.2) (3.10.0)
Requirement already satisfied: numpy in
/usr/local/lib/python3.11/dist-packages (from pycocotools>=2.0.2-
>pycocoevalcap==1.2) (2.0.2)
Requirement already satisfied: contourpy>=1.0.1 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools>=2.0.2->pycocoevalcap==1.2) (1.3.2)
Requirement already satisfied: cycler>=0.10 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools>=2.0.2->pycocoevalcap==1.2) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools>=2.0.2->pycocoevalcap==1.2) (4.57.0)
Requirement already satisfied: kiwisolver>=1.3.1 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools>=2.0.2->pycocoevalcap==1.2) (1.4.8)
Requirement already satisfied: packaging>=20.0 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools>=2.0.2->pycocoevalcap==1.2) (24.2)
Requirement already satisfied: pillow>=8 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools>=2.0.2->pycocoevalcap==1.2) (11.2.1)
Requirement already satisfied: pyparsing>=2.3.1 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools>=2.0.2->pycocoevalcap==1.2) (3.2.3)
Requirement already satisfied: python-dateutil>=2.7 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools>=2.0.2->pycocoevalcap==1.2) (2.9.0.post0)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.7-
>matplotlib>=2.1.0->pycocotools>=2.0.2->pycocoevalcap==1.2) (1.17.0)
Building wheels for collected packages: pycocoevalcap
  Building wheel for pycocoevalcap (setup.py) ... e=pycocoevalcap-1.2-
py3-none-any.whl size=593642242
sha256=46d7cc1635732b07fae784a422d5bad77b7db803ac418c55073b9a7bdceca65
9
```

```
  Stored in directory:
/tmp/pip-ephem-wheel-cache-rag0_zww/wheels/0e/98/9f/b6578f2310a0adf702
387edf950a2ba69dbf680c0b6830b312
Successfully built pycocoevalcap
Installing collected packages: pycocoevalcap
  Attempting uninstall: pycocoevalcap
    Found existing installation: pycocoevalcap 1.2
    Uninstalling pycocoevalcap-1.2:
      Successfully uninstalled pycocoevalcap-1.2
Successfully installed pycocoevalcap-1.2
/content
```

```
import torch
from transformers import Blip2Processor, Blip2ForConditionalGeneration
from PIL import Image
import os
import json
import pandas as pd
from tqdm import tqdm
from pycocotools.coco import COCO
from pycocoevalcap.eval import COCOEvalCap
import tempfile

device = "cuda" if torch.cuda.is_available() else "cpu"

# Load BLIP-2 FLAN-T5-XL
processor = Blip2Processor.from_pretrained("Salesforce/blip2-flan-t5-
xl-coco")
model =
Blip2ForConditionalGeneration.from_pretrained("Salesforce/blip2-flan-
t5-xl-coco", device_map="auto", torch_dtype=torch.float16).eval()


images_path = "/content/drive/MyDrive/val2017/"
annotations_path =
"/content/drive/MyDrive/annotations/captions_val2017.json"

# Load annotations
with open(annotations_path, "r") as f:
    coco_data = json.load(f)

image_id_to_filename = {img["id"]: img["file_name"] for img in
coco_data["images"]}
gt_captions = {}
for ann in coco_data["annotations"]:
    img_id = ann["image_id"]
    if img_id not in gt_captions:
        gt_captions[img_id] = []
    gt_captions[img_id].append(ann["caption"])
```

```
# Filter valid images
image_ids = list(gt_captions.keys())[:2000]  # limit to 100 for demo
image_files = [image_id_to_filename[i] for i in image_ids]

# Generate captions
results = []
gen_caps = []
ref_caps = []
ann_id = 0

for idx, (img_id, img_file) in enumerate(tqdm(zip(image_ids,
image_files), total=len(image_ids))):
    img_path = os.path.join(images_path, img_file)
    image = Image.open(img_path).convert("RGB")

    inputs = processor(images=image, return_tensors="pt").to(device,
torch.float16)
    generated_ids = model.generate(**inputs, max_new_tokens=30)
    caption = processor.batch_decode(generated_ids,
skip_special_tokens=True)[0].strip()

    # Save results
    references = gt_captions[img_id]
    results.append({
        "Image_File": img_file,
        "Generated_Caption": caption,
        "Ground_Truth_Captions": references
    })
    gen_caps.append({"image_id": idx, "caption": caption})
    for ref in references:
        ref_caps.append({"image_id": idx, "id": ann_id, "caption":
ref})
        ann_id += 1

# Save CSV
pd.DataFrame(results).to_csv("/content/drive/MyDrive/generated_results
.csv", index=False)

# COCO-style Evaluation
pred_file = tempfile.NamedTemporaryFile(delete=False,
suffix='.json').name
ref_file = tempfile.NamedTemporaryFile(delete=False,
suffix='.json').name

with open(pred_file, "w") as f:
    json.dump(gen_caps, f)

image_list = [{"id": i} for i in range(len(image_ids))]
with open(ref_file, "w") as f:
    json.dump({"annotations": ref_caps, "images": image_list}, f)
```

```
coco = COCO(ref_file)
res = coco.loadRes(pred_file)
cocoEval = COCOEvalCap(coco, res)
cocoEval.evaluate()

# Print scores
metrics = {k: round(v, 4) for k, v in cocoEval.eval.items()}
print(metrics)
```

{"model_id":"9db952dcec244cf9b4d716211409682a","version_major":2,"version_minor":0}

```
100%|██████████| 2000/2000 [29:50<00:00,  1.12it/s]

loading annotations into memory...
Done (t=0.01s)
creating index...
index created!
Loading and preparing results...
DONE (t=0.00s)
creating index...
index created!
tokenization...
setting up scorers...
computing Bleu score...
{'testlen': 19822, 'reflen': 19562, 'guess': [19822, 17822, 15822, 13822], 'correct': [14774, 8217, 4016, 1869]}
ratio: 1.0132910745322046
Bleu_1: 0.745
Bleu_2: 0.586
Bleu_3: 0.443
Bleu_4: 0.330
computing METEOR score...
METEOR: 0.272
computing Rouge score...
ROUGE_L: 0.560
computing CIDEr score...
CIDEr: 1.118
computing SPICE score...
SPICE: 0.216
{'Bleu_1': 0.7453, 'Bleu_2': 0.5862, 'Bleu_3': 0.4435, 'Bleu_4': 0.3295, 'METEOR': 0.2725, 'ROUGE_L': np.float64(0.5598), 'CIDEr': np.float64(1.1175), 'SPICE': np.float64(0.2159)}

import torch
from transformers import Blip2Processor, Blip2ForConditionalGeneration
from PIL import Image
import os
import pandas as pd
```

```
from tqdm import tqdm

# Set device to CPU
device = "cpu"

# Load BLIP-2 FLAN-T5-XL model
processor = Blip2Processor.from_pretrained("Salesforce/blip2-flan-t5-
xl-coco")
model =
Blip2ForConditionalGeneration.from_pretrained("Salesforce/blip2-flan-
t5-xl-coco").to(device).eval()

# Path to the testing folder containing your images
testing_images_path = "/content/drive/MyDrive/test2017"

# Get all image files in the testing folder (assuming .jpg or .jpeg
format)
image_files = [f for f in os.listdir(testing_images_path) if
f.endswith('.jpg') or f.endswith('.jpeg')]

# Generate captions
results = []

for idx, img_file in enumerate(tqdm(image_files,
total=len(image_files))):
    img_path = os.path.join(testing_images_path, img_file)
    image = Image.open(img_path).convert("RGB")

    # Preprocess image and generate caption
    inputs = processor(images=image, return_tensors="pt").to(device)
    generated_ids = model.generate(**inputs, max_new_tokens=30)
    caption = processor.batch_decode(generated_ids,
skip_special_tokens=True)[0].strip()

    # Save results
    results.append({
        "Image_File": img_file,
        "Generated_Caption": caption
    })

# Save results as a CSV file
generated_results_df = pd.DataFrame(results)
generated_results_df.to_csv("/content/drive/MyDrive/generated_testing_
results.csv", index=False)

# Optionally, you can print the generated captions
for result in results:
    print(f"Image: {result['Image_File']}")
    print(f"Generated Caption: {result['Generated_Caption']}")
    print("-" * 50)
```

{"model_id":"01064db94761445cbaf00928666237b9","version_major":2,"version_minor":0}

100%|██████████| 30/30 [02:58<00:00, 5.93s/it]

```
Image: 000000023271.jpg
Generated Caption: two women in dresses walking in a field with an
umbrella
--------------------------------------------------
Image: 000000505458.jpg
Generated Caption: a bowl of vegetables with broccoli, carrots, and
zucchini
--------------------------------------------------
Image: 000000526327.jpg
Generated Caption: a person on a snowboard is riding down a ramp
--------------------------------------------------
Image: 000000488045.jpg
Generated Caption: a bench is sitting on a snowy path near a lake
--------------------------------------------------
Image: 000000471220.jpg
Generated Caption: a man and woman are playing a video game
--------------------------------------------------
Image: 000000526846.jpg
Generated Caption: a group of people walking around a display
--------------------------------------------------
Image: 000000509670.jpg
Generated Caption: a vase with a flower and a plant on top of a wooden
box
--------------------------------------------------
Image: 000000557330.jpg
Generated Caption: a man holding a tennis racket on a court
--------------------------------------------------
Image: 000000468714.jpg
Generated Caption: a city street with a lot of traffic and tall
buildings
--------------------------------------------------
Image: 000000326419.jpg
Generated Caption: a train is traveling down the tracks in the
mountains
--------------------------------------------------
Image: 000000379473.jpg
Generated Caption: a bird perched on a branch with leaves
--------------------------------------------------
Image: 000000304597.jpg
Generated Caption: a man riding a surfboard on a wave
--------------------------------------------------
Image: 000000114063.jpg
Generated Caption: a green bucket is sitting on the floor next to a
green hole
--------------------------------------------------
```

```
Image: 000000181543.jpg
Generated Caption: a person riding a wave on a surfboard
--------------------------------------------------
Image: 000000110820.jpg
Generated Caption: a man walking with an umbrella in the rain
--------------------------------------------------
Image: 000000130611.jpg
Generated Caption: a zebra standing in a field with trees behind it
--------------------------------------------------
Image: 000000248460.jpg
Generated Caption: a man kneeling down in the grass to grab a frisbee
--------------------------------------------------
Image: 000000241446.jpg
Generated Caption: two stuffed animals sitting on a railing next to a
train
--------------------------------------------------
Image: 000000297369.jpg
Generated Caption: a picnic table sits in a park with trees
--------------------------------------------------
Image: 000000122286.jpg
Generated Caption: a bedroom with a bed, desk, and a lamp
--------------------------------------------------
Image: 000000151122.jpg
Generated Caption: a plate with hot dogs and macaroni and cheese
--------------------------------------------------
Image: 000000098660.jpg
Generated Caption: a man sitting at a table looking at his phone
--------------------------------------------------
Image: 000000102209.jpg
Generated Caption: a man is throwing a frisbee on a dirt path
--------------------------------------------------
Image: 000000054498.jpg
Generated Caption: a man riding a motorcycle down a street with a
crowd
--------------------------------------------------
Image: 000000084457.jpg
Generated Caption: a man and a woman are playing a video game
--------------------------------------------------
Image: 000000006069.jpg
Generated Caption: a cat is looking at a bird feeder
--------------------------------------------------
Image: 000000040965.jpg
Generated Caption: a tray of food on a plane with a roll
--------------------------------------------------
Image: 000000026680.jpg
Generated Caption: two zebras walking in a field with mountains in the
background
--------------------------------------------------
Image: 000000004336.jpg
```

```
Generated Caption: a bed with a messy floor and a cat on top
--------------------------------------------------
Image: 000000047056.jpg
Generated Caption: a vase with flowers on a table with books
--------------------------------------------------
```

Blip 2 (2.7b model)

```
pip install torch transformers

Requirement already satisfied: torch in
/usr/local/lib/python3.11/dist-packages (2.6.0+cu124)
Requirement already satisfied: transformers in
/usr/local/lib/python3.11/dist-packages (4.51.3)
Requirement already satisfied: filelock in
/usr/local/lib/python3.11/dist-packages (from torch) (3.18.0)
Requirement already satisfied: typing-extensions>=4.10.0 in
/usr/local/lib/python3.11/dist-packages (from torch) (4.13.2)
Requirement already satisfied: networkx in
/usr/local/lib/python3.11/dist-packages (from torch) (3.4.2)
Requirement already satisfied: jinja2 in
/usr/local/lib/python3.11/dist-packages (from torch) (3.1.6)
Requirement already satisfied: fsspec in
/usr/local/lib/python3.11/dist-packages (from torch) (2025.3.2)
Collecting nvidia-cuda-nvrtc-cu12==12.4.127 (from torch)
  Downloading nvidia_cuda_nvrtc_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cuda-runtime-cu12==12.4.127 (from torch)
  Downloading nvidia_cuda_runtime_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cuda-cupti-cu12==12.4.127 (from torch)
  Downloading nvidia_cuda_cupti_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-cudnn-cu12==9.1.0.70 (from torch)
  Downloading nvidia_cudnn_cu12-9.1.0.70-py3-none-
manylinux2014_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-cublas-cu12==12.4.5.8 (from torch)
  Downloading nvidia_cublas_cu12-12.4.5.8-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cufft-cu12==11.2.1.3 (from torch)
  Downloading nvidia_cufft_cu12-11.2.1.3-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-curand-cu12==10.3.5.147 (from torch)
  Downloading nvidia_curand_cu12-10.3.5.147-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cusolver-cu12==11.6.1.9 (from torch)
```

```
  Downloading nvidia_cusolver_cu12-11.6.1.9-py3-none-
manylinux2014_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-cusparse-cu12==12.3.1.170 (from torch)
  Downloading nvidia_cusparse_cu12-12.3.1.170-py3-none-
manylinux2014_x86_64.whl.metadata (1.6 kB)
Requirement already satisfied: nvidia-cusparselt-cu12==0.6.2 in
/usr/local/lib/python3.11/dist-packages (from torch) (0.6.2)
Requirement already satisfied: nvidia-nccl-cu12==2.21.5 in
/usr/local/lib/python3.11/dist-packages (from torch) (2.21.5)
Requirement already satisfied: nvidia-nvtx-cu12==12.4.127 in
/usr/local/lib/python3.11/dist-packages (from torch) (12.4.127)
Collecting nvidia-nvjitlink-cu12==12.4.127 (from torch)
  Downloading nvidia_nvjitlink_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Requirement already satisfied: triton==3.2.0 in
/usr/local/lib/python3.11/dist-packages (from torch) (3.2.0)
Requirement already satisfied: sympy==1.13.1 in
/usr/local/lib/python3.11/dist-packages (from torch) (1.13.1)
Requirement already satisfied: mpmath<1.4,>=1.1.0 in
/usr/local/lib/python3.11/dist-packages (from sympy==1.13.1->torch)
(1.3.0)
Requirement already satisfied: huggingface-hub<1.0,>=0.30.0 in
/usr/local/lib/python3.11/dist-packages (from transformers) (0.30.2)
Requirement already satisfied: numpy>=1.17 in
/usr/local/lib/python3.11/dist-packages (from transformers) (2.0.2)
Requirement already satisfied: packaging>=20.0 in
/usr/local/lib/python3.11/dist-packages (from transformers) (24.2)
Requirement already satisfied: pyyaml>=5.1 in
/usr/local/lib/python3.11/dist-packages (from transformers) (6.0.2)
Requirement already satisfied: regex!=2019.12.17 in
/usr/local/lib/python3.11/dist-packages (from transformers)
(2024.11.6)
Requirement already satisfied: requests in
/usr/local/lib/python3.11/dist-packages (from transformers) (2.32.3)
Requirement already satisfied: tokenizers<0.22,>=0.21 in
/usr/local/lib/python3.11/dist-packages (from transformers) (0.21.1)
Requirement already satisfied: safetensors>=0.4.3 in
/usr/local/lib/python3.11/dist-packages (from transformers) (0.5.3)
Requirement already satisfied: tqdm>=4.27 in
/usr/local/lib/python3.11/dist-packages (from transformers) (4.67.1)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.11/dist-packages (from jinja2->torch) (3.0.2)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
(3.4.1)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
(3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
```

```
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
(2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
(2025.4.26)
Downloading nvidia_cublas_cu12-12.4.5.8-py3-none-
manylinux2014_x86_64.whl (363.4 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 363.4/363.4 MB 3.0 MB/s eta
0:00:00
anylinux2014_x86_64.whl (13.8 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 13.8/13.8 MB 130.5 MB/s eta
0:00:00
anylinux2014_x86_64.whl (24.6 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 24.6/24.6 MB 91.3 MB/s eta
0:00:00
e_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl (883 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 883.7/883.7 kB 51.4 MB/s eta
0:00:00
anylinux2014_x86_64.whl (664.8 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 664.8/664.8 MB 1.7 MB/s eta
0:00:00
anylinux2014_x86_64.whl (211.5 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 211.5/211.5 MB 12.1 MB/s eta
0:00:00
anylinux2014_x86_64.whl (56.3 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 56.3/56.3 MB 44.4 MB/s eta
0:00:00
anylinux2014_x86_64.whl (127.9 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 127.9/127.9 MB 19.9 MB/s eta
0:00:00
anylinux2014_x86_64.whl (207.5 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 207.5/207.5 MB 5.0 MB/s eta
0:00:00
anylinux2014_x86_64.whl (21.1 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 21.1/21.1 MB 107.1 MB/s eta
0:00:00
e-cu12, nvidia-cuda-nvrtc-cu12, nvidia-cuda-cupti-cu12, nvidia-cublas-
cu12, nvidia-cusparse-cu12, nvidia-cudnn-cu12, nvidia-cusolver-cu12
  Attempting uninstall: nvidia-nvjitlink-cu12
    Found existing installation: nvidia-nvjitlink-cu12 12.5.82
    Uninstalling nvidia-nvjitlink-cu12-12.5.82:
      Successfully uninstalled nvidia-nvjitlink-cu12-12.5.82
  Attempting uninstall: nvidia-curand-cu12
    Found existing installation: nvidia-curand-cu12 10.3.6.82
    Uninstalling nvidia-curand-cu12-10.3.6.82:
      Successfully uninstalled nvidia-curand-cu12-10.3.6.82
  Attempting uninstall: nvidia-cufft-cu12
    Found existing installation: nvidia-cufft-cu12 11.2.3.61
    Uninstalling nvidia-cufft-cu12-11.2.3.61:
```

```
      Successfully uninstalled nvidia-cufft-cu12-11.2.3.61
  Attempting uninstall: nvidia-cuda-runtime-cu12
    Found existing installation: nvidia-cuda-runtime-cu12 12.5.82
    Uninstalling nvidia-cuda-runtime-cu12-12.5.82:
      Successfully uninstalled nvidia-cuda-runtime-cu12-12.5.82
  Attempting uninstall: nvidia-cuda-nvrtc-cu12
    Found existing installation: nvidia-cuda-nvrtc-cu12 12.5.82
    Uninstalling nvidia-cuda-nvrtc-cu12-12.5.82:
      Successfully uninstalled nvidia-cuda-nvrtc-cu12-12.5.82
  Attempting uninstall: nvidia-cuda-cupti-cu12
    Found existing installation: nvidia-cuda-cupti-cu12 12.5.82
    Uninstalling nvidia-cuda-cupti-cu12-12.5.82:
      Successfully uninstalled nvidia-cuda-cupti-cu12-12.5.82
  Attempting uninstall: nvidia-cublas-cu12
    Found existing installation: nvidia-cublas-cu12 12.5.3.2
    Uninstalling nvidia-cublas-cu12-12.5.3.2:
      Successfully uninstalled nvidia-cublas-cu12-12.5.3.2
  Attempting uninstall: nvidia-cusparse-cu12
    Found existing installation: nvidia-cusparse-cu12 12.5.1.3
    Uninstalling nvidia-cusparse-cu12-12.5.1.3:
      Successfully uninstalled nvidia-cusparse-cu12-12.5.1.3
  Attempting uninstall: nvidia-cudnn-cu12
    Found existing installation: nvidia-cudnn-cu12 9.3.0.75
    Uninstalling nvidia-cudnn-cu12-9.3.0.75:
      Successfully uninstalled nvidia-cudnn-cu12-9.3.0.75
  Attempting uninstall: nvidia-cusolver-cu12
    Found existing installation: nvidia-cusolver-cu12 11.6.3.83
    Uninstalling nvidia-cusolver-cu12-11.6.3.83:
      Successfully uninstalled nvidia-cusolver-cu12-11.6.3.83
Successfully installed nvidia-cublas-cu12-12.4.5.8 nvidia-cuda-cupti-
cu12-12.4.127 nvidia-cuda-nvrtc-cu12-12.4.127 nvidia-cuda-runtime-
cu12-12.4.127 nvidia-cudnn-cu12-9.1.0.70 nvidia-cufft-cu12-11.2.1.3
nvidia-curand-cu12-10.3.5.147 nvidia-cusolver-cu12-11.6.1.9 nvidia-
cusparse-cu12-12.3.1.170 nvidia-nvjitlink-cu12-12.4.127
```

```
from transformers import Blip2Processor, Blip2ForConditionalGeneration

# Load BLIP-2 7B model and processor
import torch
processor = Blip2Processor.from_pretrained("Salesforce/blip2-opt-
2.7b")
model =
Blip2ForConditionalGeneration.from_pretrained("Salesforce/blip2-opt-
2.7b", device_map="auto", torch_dtype=torch.float16).eval()
```


```
{"model_id":"3023ee2275f04072bc2640c62723333c","version_major":2,"vers
ion_minor":0}
```

```
{"model_id":"705e28b83552464986da07194097db53","version_major":2,"vers
ion_minor":0}
```

{"model_id":"fc5ea901006c4e179e89983611f901b2","version_major":2,"version_minor":0}

{"model_id":"4080e93c536d4e809b3138fa58dfc5e3","version_major":2,"version_minor":0}

{"model_id":"391e2eea1ea24ab485109f762d200986","version_major":2,"version_minor":0}

{"model_id":"d1c170140281484abd978918674ef953","version_major":2,"version_minor":0}

{"model_id":"14804964d31145d184c17e6be49ecc73","version_major":2,"version_minor":0}

```
import json


device = "cuda" if torch.cuda.is_available() else "cpu"

images_path = "/content/drive/MyDrive/val2017/"
annotations_path =
"/content/drive/MyDrive/annotations/captions_val2017.json"

# Load annotations
# Load annotations
with open(annotations_path, "r") as f:
    coco_data = json.load(f)

image_id_to_filename = {img["id"]: img["file_name"] for img in
coco_data["images"]}
gt_captions = {}
for ann in coco_data["annotations"]:
    img_id = ann["image_id"]
    if img_id not in gt_captions:
        gt_captions[img_id] = []
    gt_captions[img_id].append(ann["caption"])

# Filter valid images
image_ids = list(gt_captions.keys())[:2000]  # limit to 2000 for your
dataset
image_files = [image_id_to_filename[i] for i in image_ids]


# Generate captions
from tqdm import tqdm  # Add this line to import tqdm
from PIL import Image
import os

results = []
gen_caps = []
```

```
ref_caps = []
ann_id = 0

for idx, (img_id, img_file) in enumerate(tqdm(zip(image_ids,
image_files), total=len(image_ids))):
    img_path = os.path.join(images_path, img_file)
    image = Image.open(img_path).convert("RGB")

    # Preprocess and generate captions
    inputs = processor(images=image, return_tensors="pt").to(device,
torch.float16)
    generated_ids = model.generate(**inputs, max_new_tokens=30)
    caption = processor.batch_decode(generated_ids,
skip_special_tokens=True)[0].strip()

    # Save results
    references = gt_captions[img_id]
    results.append({
        "Image_File": img_file,
        "Generated_Caption": caption,
        "Ground_Truth_Captions": references
    })
    gen_caps.append({"image_id": idx, "caption": caption})
    for ref in references:
        ref_caps.append({"image_id": idx, "id": ann_id, "caption":
ref})
        ann_id += 1



100%|██████████| 2000/2000 [35:40<00:00,  1.07s/it]

---------------------------------------------------------------------
-----
NameError                                 Traceback (most recent call
last)
<ipython-input-13-b63edec125bf> in <cell line: 0>()
     31
     32 # Save results to CSV
---> 33
pd.DataFrame(results).to_csv("/content/drive/MyDrive/generated_results
.csv", index=False)
     34
     35 # COCO-style Evaluation

NameError: name 'pd' is not defined

!git clone https://github.com/salaniz/pycocoevalcap.git
%cd pycocoevalcap
!pip install .
%cd ..
```

```
Cloning into 'pycocoevalcap'...
remote: Enumerating objects: 821, done.ote: Counting objects: 100%
(12/12), done.ote: Compressing objects: 100% (9/9), done.ote: Total
821 (delta 4), reused 3 (delta 3), pack-reused 809 (from 2)etadata
(setup.py) ... ent already satisfied: pycocotools>=2.0.2 in
/usr/local/lib/python3.11/dist-packages (from pycocoevalcap==1.2)
(2.0.8)
Requirement already satisfied: matplotlib>=2.1.0 in
/usr/local/lib/python3.11/dist-packages (from pycocotools>=2.0.2-
>pycocoevalcap==1.2) (3.10.0)
Requirement already satisfied: numpy in
/usr/local/lib/python3.11/dist-packages (from pycocotools>=2.0.2-
>pycocoevalcap==1.2) (2.0.2)
Requirement already satisfied: contourpy>=1.0.1 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools>=2.0.2->pycocoevalcap==1.2) (1.3.2)
Requirement already satisfied: cycler>=0.10 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools>=2.0.2->pycocoevalcap==1.2) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools>=2.0.2->pycocoevalcap==1.2) (4.57.0)
Requirement already satisfied: kiwisolver>=1.3.1 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools>=2.0.2->pycocoevalcap==1.2) (1.4.8)
Requirement already satisfied: packaging>=20.0 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools>=2.0.2->pycocoevalcap==1.2) (24.2)
Requirement already satisfied: pillow>=8 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools>=2.0.2->pycocoevalcap==1.2) (11.2.1)
Requirement already satisfied: pyparsing>=2.3.1 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools>=2.0.2->pycocoevalcap==1.2) (3.2.3)
Requirement already satisfied: python-dateutil>=2.7 in
/usr/local/lib/python3.11/dist-packages (from matplotlib>=2.1.0-
>pycocotools>=2.0.2->pycocoevalcap==1.2) (2.9.0.post0)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.7-
>matplotlib>=2.1.0->pycocotools>=2.0.2->pycocoevalcap==1.2) (1.17.0)
Building wheels for collected packages: pycocoevalcap
  Building wheel for pycocoevalcap (setup.py) ... e=pycocoevalcap-1.2-
py3-none-any.whl size=104312245
sha256=2d051328d71f9b1c7b5fc52d1175bcb3c0cad91f1632db8526a164259f517c4
9
  Stored in directory:
/tmp/pip-ephem-wheel-cache-t8g9cjze/wheels/0e/98/9f/b6578f2310a0adf702
387edf950a2ba69dbf680c0b6830b312
Successfully built pycocoevalcap
Installing collected packages: pycocoevalcap
```

```
Successfully installed pycocoevalcap-1.2
/content

import pandas as pd
import tempfile
from pycocotools.coco import COCO
from pycocoevalcap.eval import COCOEvalCap

# Save results to CSV
pd.DataFrame(results).to_csv("/content/drive/MyDrive/generated_results
.csv", index=False)

# COCO-style Evaluation
pred_file = tempfile.NamedTemporaryFile(delete=False,
suffix='.json').name
ref_file = tempfile.NamedTemporaryFile(delete=False,
suffix='.json').name

with open(pred_file, "w") as f:
    json.dump(gen_caps, f)

image_list = [{"id": i} for i in range(len(image_ids))]
with open(ref_file, "w") as f:
    json.dump({"annotations": ref_caps, "images": image_list}, f)

coco = COCO(ref_file)
res = coco.loadRes(pred_file)
cocoEval = COCOEvalCap(coco, res)
cocoEval.evaluate()

# Print evaluation metrics
metrics = {k: round(v, 4) for k, v in cocoEval.eval.items()}
print(metrics)

loading annotations into memory...
Done (t=0.01s)
creating index...
index created!
Loading and preparing results...
DONE (t=0.00s)
creating index...
index created!
tokenization...
setting up scorers...
Downloading stanford-corenlp-3.6.0 for SPICE ...
Progress: 384.5M / 384.5M (100.0%)
Extracting stanford-corenlp-3.6.0 ...
Done.
computing Bleu score...
{'testlen': 16550, 'reflen': 18242, 'guess': [16550, 14550, 12550,
```

```
10550], 'correct': [13705, 8112, 4155, 2052]}
ratio: 0.9072470123889427
Bleu_1: 0.748
Bleu_2: 0.613
Bleu_3: 0.483
Bleu_4: 0.375
computing METEOR score...
METEOR: 0.275
computing Rouge score...
ROUGE_L: 0.580
computing CIDEr score...
CIDEr: 1.251
computing SPICE score...
SPICE: 0.217
{'Bleu_1': 0.7476, 'Bleu_2': 0.6134, 'Bleu_3': 0.4827, 'Bleu_4':
0.3749, 'METEOR': 0.2747, 'ROUGE_L': np.float64(0.5805), 'CIDEr':
np.float64(1.2514), 'SPICE': np.float64(0.2173)}
```

```
import torch
from transformers import Blip2Processor, Blip2ForConditionalGeneration
from PIL import Image
import os
import pandas as pd
from tqdm import tqdm

# Set device to CPU
device = "cpu"

# Load BLIP-2 FLAN-T5-XL model
processor = Blip2Processor.from_pretrained("Salesforce/blip2-opt-
2.7b")
model =
Blip2ForConditionalGeneration.from_pretrained("Salesforce/blip2-opt-
2.7b").to(device).eval()

# Path to the testing folder containing your images
testing_images_path = "/content/drive/MyDrive/test2017"

# Get all image files in the testing folder (assuming .jpg or .jpeg
format)
image_files = [f for f in os.listdir(testing_images_path) if
f.endswith('.jpg') or f.endswith('.jpeg')]

# Generate captions
results = []

for idx, img_file in enumerate(tqdm(image_files,
total=len(image_files))):
    img_path = os.path.join(testing_images_path, img_file)
    image = Image.open(img_path).convert("RGB")
```

```
    # Preprocess image and generate caption
    inputs = processor(images=image, return_tensors="pt").to(device)
    generated_ids = model.generate(**inputs, max_new_tokens=30)
    caption = processor.batch_decode(generated_ids,
skip_special_tokens=True)[0].strip()

    # Save results
    results.append({
        "Image_File": img_file,
        "Generated_Caption": caption
    })

# Save results as a CSV file
generated_results_df = pd.DataFrame(results)
generated_results_df.to_csv("/content/drive/MyDrive/generated_testing_
results_blip2base.csv", index=False)

# Optionally, you can print the generated captions
for result in results:
    print(f"Image: {result['Image_File']}")
    print(f"Generated Caption: {result['Generated_Caption']}")
    print("-" * 50)
```

{"model_id":"2064e9a464954099a66a64e256e7442f","version_major":2,"vers
ion_minor":0}

```
100%|██████████| 30/30 [02:27<00:00,  4.93s/it]

Image: 000000047056.jpg
Generated Caption: a vase with flowers in it
--------------------------------------------------
Image: 000000054498.jpg
Generated Caption: a man riding a motorcycle in a parade
--------------------------------------------------
Image: 000000102209.jpg
Generated Caption: a man throwing a frisbee in a park
--------------------------------------------------
Image: 000000040965.jpg
Generated Caption: a tray of food on a plane
--------------------------------------------------
Image: 000000026680.jpg
Generated Caption: two zebras walking in a field with mountains in the
background
--------------------------------------------------
Image: 000000181543.jpg
Generated Caption: a person riding a wave on a surfboard
--------------------------------------------------
Image: 000000241446.jpg
Generated Caption: a train is passing by
```

```
--------------------------------------------------
Image: 000000023271.jpg
Generated Caption: two women walking in the grass with an umbrella
--------------------------------------------------
Image: 000000488045.jpg
Generated Caption: a bench is covered in snow next to a lake
--------------------------------------------------
Image: 000000114063.jpg
Generated Caption: a small green toilet in a bathroom
--------------------------------------------------
Image: 000000151122.jpg
Generated Caption: a white plate with two hot dogs and a bowl of
macaroni and cheese
--------------------------------------------------
Image: 000000526846.jpg
Generated Caption: a group of people standing around a display of a
computer
--------------------------------------------------
Image: 000000505458.jpg
Generated Caption: a pan full of vegetables
--------------------------------------------------
Image: 000000379473.jpg
Generated Caption: a bird is perched on a branch in a black and white
photo
--------------------------------------------------
Image: 000000110820.jpg
Generated Caption: a man walking with an umbrella
--------------------------------------------------
Image: 000000526327.jpg
Generated Caption: a person on a snowboard doing a trick on a ramp
--------------------------------------------------
Image: 000000304597.jpg
Generated Caption: a man riding a surfboard on a wave
--------------------------------------------------
Image: 000000130611.jpg
Generated Caption: a zebra standing in the sun
--------------------------------------------------
Image: 000000471220.jpg
Generated Caption: a woman holding a wii remote
--------------------------------------------------
Image: 000000297369.jpg
Generated Caption: a picnic table in a park with a tree in the
background
--------------------------------------------------
Image: 000000468714.jpg
Generated Caption: a busy city street with traffic lights and signs
--------------------------------------------------
Image: 000000122286.jpg
Generated Caption: a bedroom with a bed, desk, and chair
```

```
--------------------------------------------------
Image: 000000557330.jpg
Generated Caption: a man holding a tennis racket on a tennis court
--------------------------------------------------
Image: 000000004336.jpg
Generated Caption: a cat is sitting on a bed
--------------------------------------------------
Image: 000000084457.jpg
Generated Caption: a man and a woman playing a video game
--------------------------------------------------
Image: 000000509670.jpg
Generated Caption: a wooden table with a vase, a vase with flowers,
and a vase with a plant
--------------------------------------------------
Image: 000000006069.jpg
Generated Caption: a bird is sitting on a bird feeder
--------------------------------------------------
Image: 000000098660.jpg
Generated Caption: a man sitting at a table with a phone and a cup
--------------------------------------------------
Image: 000000326419.jpg
Generated Caption: a train on the tracks
--------------------------------------------------
Image: 000000248460.jpg
Generated Caption: a man kneeling on the grass
--------------------------------------------------
```

```python
import torch
from transformers import Blip2Processor, Blip2ForConditionalGeneration
from PIL import Image
import os
import pandas as pd
from tqdm import tqdm

# Set device to CPU
device = "cpu"

# Load BLIP-2 FLAN-T5-XL model
processor = Blip2Processor.from_pretrained("Salesforce/blip2-opt-
2.7b")
model =
Blip2ForConditionalGeneration.from_pretrained("Salesforce/blip2-opt-
2.7b").to(device).eval()

# Path to the testing folder containing your images
testing_images_path = "/content/drive/MyDrive/testing"

# Get all image files in the testing folder (assuming .jpg or .jpeg
```

```
format)
image_files = [f for f in os.listdir(testing_images_path) if
f.endswith('.jpg') or f.endswith('.jpeg')]

# Generate captions
results = []

for idx, img_file in enumerate(tqdm(image_files,
total=len(image_files))):
    img_path = os.path.join(testing_images_path, img_file)
    image = Image.open(img_path).convert("RGB")

    # Preprocess image and generate caption
    inputs = processor(images=image, return_tensors="pt").to(device)
    generated_ids = model.generate(**inputs, max_new_tokens=30)
    caption = processor.batch_decode(generated_ids,
skip_special_tokens=True)[0].strip()

    # Save results
    results.append({
        "Image_File": img_file,
        "Generated_Caption": caption
    })

# Save results as a CSV file
generated_results_df = pd.DataFrame(results)
generated_results_df.to_csv("/content/drive/MyDrive/generated_testing_
results_blip2base_2.7b.csv", index=False)

# Optionally, you can print the generated captions
for result in results:
    print(f"Image: {result['Image_File']}")
    print(f"Generated Caption: {result['Generated_Caption']}")
    print("-" * 50)
```


```
{"model_id":"317cd35ebb1b417cbba61c6283d698f8","version_major":2,"vers
ion_minor":0}

{"model_id":"5f2e9619b1a54253bbde6f241ae167c8","version_major":2,"vers
ion_minor":0}

{"model_id":"5dcec06c21c64c7fbf5b3ae4f31f1675","version_major":2,"vers
ion_minor":0}
```


```
100%|██████████| 24/24 [01:36<00:00,  4.02s/it]

Image: 000000581886.jpg
Generated Caption: two people sitting on the beach with surfboards
--------------------------------------------------
Image: 000000581402.jpg
Generated Caption: a cat sleeping in a bowl
```

```
--------------------------------------------------
Image: 000000580975.jpg
Generated Caption: a row of motorcycles parked on the side of the road
--------------------------------------------------
Image: 000000581873.jpg
Generated Caption: a dog standing on a checkered floor
--------------------------------------------------
Image: 000000581422.jpg
Generated Caption: a man sitting on a bench next to a cow
--------------------------------------------------
Image: 000000581396.jpg
Generated Caption: a man taking a picture of himself in the mirror
--------------------------------------------------
Image: 000000581686.jpg
Generated Caption: two pigeons are sitting on a ledge
--------------------------------------------------
Image: 000000581702.jpg
Generated Caption: a bird with a red head and black body
--------------------------------------------------
Image: 000000581036.jpg
Generated Caption: a piece of cake on a plate
--------------------------------------------------
Image: 000000580951.jpg
Generated Caption: an elephant is standing in the grass near water
--------------------------------------------------
Image: 000000581738.jpg
Generated Caption: two red buses parked in a parking lot
--------------------------------------------------
Image: 000000581719.jpg
Generated Caption: a young boy wearing a baseball uniform
--------------------------------------------------
Image: 000000581582.jpg
Generated Caption: a train is pulling into a station with a building
in the background
--------------------------------------------------
Image: 000000581827.jpg
Generated Caption: a person playing tennis on a court
--------------------------------------------------
Image: 000000581249.jpg
Generated Caption: a cat on top of a suitcase
--------------------------------------------------
Image: 000000581542.jpg
Generated Caption: a man holding a cell phone
--------------------------------------------------
Image: 000000581593.jpg
Generated Caption: a plate of food with bread and a glass of wine
--------------------------------------------------
Image: 000000581704.jpg
Generated Caption: a black dog standing next to a suitcase
```

```
--------------------------------------------------
Image: 000000581089.jpg
Generated Caption: a dog laying on a bed with pillows and blankets
--------------------------------------------------
Image: 000000581326.jpg
Generated Caption: a woman is standing next to two horses on a street
--------------------------------------------------
Image: 000000581009.jpg
Generated Caption: a young man and a young woman riding a skateboard
down a street
--------------------------------------------------
Image: 000000581711.jpg
Generated Caption: a plate with a sandwich and a plate of tomatoes and
beets
--------------------------------------------------
Image: 000000581821.jpg
Generated Caption: a living room with a couch, a bookcase, and a table
--------------------------------------------------
Image: 000000581657.jpg
Generated Caption: a man using a cell phone
--------------------------------------------------
```

CLIP Score calculation

```
!pip install transformers torch torchvision datasets sentencepiece
clip-score
!pip install git+https://github.com/openai/CLIP.git

Requirement already satisfied: transformers in
/usr/local/lib/python3.11/dist-packages (4.51.3)
Requirement already satisfied: torch in
/usr/local/lib/python3.11/dist-packages (2.6.0+cu124)
Requirement already satisfied: torchvision in
/usr/local/lib/python3.11/dist-packages (0.21.0+cu124)
Collecting datasets
  Downloading datasets-3.6.0-py3-none-any.whl.metadata (19 kB)
Requirement already satisfied: sentencepiece in
/usr/local/lib/python3.11/dist-packages (0.2.0)
Collecting clip-score
  Downloading clip_score-0.2.1-py2.py3-none-any.whl.metadata (6.4 kB)
Requirement already satisfied: filelock in
/usr/local/lib/python3.11/dist-packages (from transformers) (3.18.0)
Requirement already satisfied: huggingface-hub<1.0,>=0.30.0 in
/usr/local/lib/python3.11/dist-packages (from transformers) (0.30.2)
Requirement already satisfied: numpy>=1.17 in
/usr/local/lib/python3.11/dist-packages (from transformers) (2.0.2)
```

```
Requirement already satisfied: packaging>=20.0 in
/usr/local/lib/python3.11/dist-packages (from transformers) (24.2)
Requirement already satisfied: pyyaml>=5.1 in
/usr/local/lib/python3.11/dist-packages (from transformers) (6.0.2)
Requirement already satisfied: regex!=2019.12.17 in
/usr/local/lib/python3.11/dist-packages (from transformers)
(2024.11.6)
Requirement already satisfied: requests in
/usr/local/lib/python3.11/dist-packages (from transformers) (2.32.3)
Requirement already satisfied: tokenizers<0.22,>=0.21 in
/usr/local/lib/python3.11/dist-packages (from transformers) (0.21.1)
Requirement already satisfied: safetensors>=0.4.3 in
/usr/local/lib/python3.11/dist-packages (from transformers) (0.5.3)
Requirement already satisfied: tqdm>=4.27 in
/usr/local/lib/python3.11/dist-packages (from transformers) (4.67.1)
Requirement already satisfied: typing-extensions>=4.10.0 in
/usr/local/lib/python3.11/dist-packages (from torch) (4.13.2)
Requirement already satisfied: networkx in
/usr/local/lib/python3.11/dist-packages (from torch) (3.4.2)
Requirement already satisfied: jinja2 in
/usr/local/lib/python3.11/dist-packages (from torch) (3.1.6)
Requirement already satisfied: fsspec in
/usr/local/lib/python3.11/dist-packages (from torch) (2025.3.2)
Collecting nvidia-cuda-nvrtc-cu12==12.4.127 (from torch)
  Downloading nvidia_cuda_nvrtc_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cuda-runtime-cu12==12.4.127 (from torch)
  Downloading nvidia_cuda_runtime_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cuda-cupti-cu12==12.4.127 (from torch)
  Downloading nvidia_cuda_cupti_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-cudnn-cu12==9.1.0.70 (from torch)
  Downloading nvidia_cudnn_cu12-9.1.0.70-py3-none-
manylinux2014_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-cublas-cu12==12.4.5.8 (from torch)
  Downloading nvidia_cublas_cu12-12.4.5.8-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cufft-cu12==11.2.1.3 (from torch)
  Downloading nvidia_cufft_cu12-11.2.1.3-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-curand-cu12==10.3.5.147 (from torch)
  Downloading nvidia_curand_cu12-10.3.5.147-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cusolver-cu12==11.6.1.9 (from torch)
  Downloading nvidia_cusolver_cu12-11.6.1.9-py3-none-
manylinux2014_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-cusparse-cu12==12.3.1.170 (from torch)
  Downloading nvidia_cusparse_cu12-12.3.1.170-py3-none-
```

```
manylinux2014_x86_64.whl.metadata (1.6 kB)
Requirement already satisfied: nvidia-cusparselt-cu12==0.6.2 in
/usr/local/lib/python3.11/dist-packages (from torch) (0.6.2)
Requirement already satisfied: nvidia-nccl-cu12==2.21.5 in
/usr/local/lib/python3.11/dist-packages (from torch) (2.21.5)
Requirement already satisfied: nvidia-nvtx-cu12==12.4.127 in
/usr/local/lib/python3.11/dist-packages (from torch) (12.4.127)
Collecting nvidia-nvjitlink-cu12==12.4.127 (from torch)
  Downloading nvidia_nvjitlink_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Requirement already satisfied: triton==3.2.0 in
/usr/local/lib/python3.11/dist-packages (from torch) (3.2.0)
Requirement already satisfied: sympy==1.13.1 in
/usr/local/lib/python3.11/dist-packages (from torch) (1.13.1)
Requirement already satisfied: mpmath<1.4,>=1.1.0 in
/usr/local/lib/python3.11/dist-packages (from sympy==1.13.1->torch)
(1.3.0)
Requirement already satisfied: pillow!=8.3.*,>=5.3.0 in
/usr/local/lib/python3.11/dist-packages (from torchvision) (11.2.1)
Requirement already satisfied: pyarrow>=15.0.0 in
/usr/local/lib/python3.11/dist-packages (from datasets) (18.1.0)
Collecting dill<0.3.9,>=0.3.0 (from datasets)
  Downloading dill-0.3.8-py3-none-any.whl.metadata (10 kB)
Requirement already satisfied: pandas in
/usr/local/lib/python3.11/dist-packages (from datasets) (2.2.2)
Collecting xxhash (from datasets)
  Downloading xxhash-3.5.0-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (12 kB)
Collecting multiprocess<0.70.17 (from datasets)
  Downloading multiprocess-0.70.16-py311-none-any.whl.metadata (7.2
kB)
Collecting fsspec (from torch)
  Downloading fsspec-2025.3.0-py3-none-any.whl.metadata (11 kB)
Collecting ftfy (from clip-score)
  Downloading ftfy-6.3.1-py3-none-any.whl.metadata (7.3 kB)
Requirement already satisfied: aiohttp!=4.0.0a0,!=4.0.0a1 in
/usr/local/lib/python3.11/dist-packages (from
fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (3.11.15)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
(3.4.1)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
(3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
(2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
```

```
(2025.4.26)
Requirement already satisfied: wcwidth in
/usr/local/lib/python3.11/dist-packages (from ftfy->clip-score)
(0.2.13)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.11/dist-packages (from jinja2->torch) (3.0.2)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.11/dist-packages (from pandas->datasets)
(2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in
/usr/local/lib/python3.11/dist-packages (from pandas->datasets)
(2025.2)
Requirement already satisfied: tzdata>=2022.7 in
/usr/local/lib/python3.11/dist-packages (from pandas->datasets)
(2025.2)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (2.6.1)
Requirement already satisfied: aiosignal>=1.1.2 in
/usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (1.3.2)
Requirement already satisfied: attrs>=17.3.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (25.3.0)
Requirement already satisfied: frozenlist>=1.1.1 in
/usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (1.6.0)
Requirement already satisfied: multidict<7.0,>=4.5 in
/usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (6.4.3)
Requirement already satisfied: propcache>=0.2.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (0.3.1)
Requirement already satisfied: yarl<2.0,>=1.17.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (1.20.0)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2-
>pandas->datasets) (1.17.0)
Downloading nvidia_cublas_cu12-12.4.5.8-py3-none-
manylinux2014_x86_64.whl (363.4 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 363.4/363.4 MB 3.2 MB/s eta
0:00:00
anylinux2014_x86_64.whl (13.8 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 13.8/13.8 MB 72.6 MB/s eta
0:00:00
anylinux2014_x86_64.whl (24.6 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 24.6/24.6 MB 68.2 MB/s eta
0:00:00
```

```
e_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl (883 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 883.7/883.7 kB 50.1 MB/s eta
0:00:00
anylinux2014_x86_64.whl (664.8 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 664.8/664.8 MB 2.1 MB/s eta
0:00:00
anylinux2014_x86_64.whl (211.5 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 211.5/211.5 MB 11.3 MB/s eta
0:00:00
anylinux2014_x86_64.whl (56.3 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 56.3/56.3 MB 35.8 MB/s eta
0:00:00
anylinux2014_x86_64.whl (127.9 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 127.9/127.9 MB 14.9 MB/s eta
0:00:00
anylinux2014_x86_64.whl (207.5 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 207.5/207.5 MB 3.4 MB/s eta
0:00:00
anylinux2014_x86_64.whl (21.1 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 21.1/21.1 MB 73.9 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 491.5/491.5 kB 40.1 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 116.3/116.3 kB 11.0 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 193.6/193.6 kB 18.8 MB/s eta
0:00:00
ultiprocess-0.70.16-py311-none-any.whl (143 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 143.5/143.5 kB 14.5 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 44.8/44.8 kB 4.0 MB/s eta
0:00:00
anylinux_2_17_x86_64.manylinux2014_x86_64.whl (194 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 194.8/194.8 kB 20.7 MB/s eta
0:00:00
e-cu12, nvidia-cuda-nvrtc-cu12, nvidia-cuda-cupti-cu12, nvidia-cublas-
cu12, ftfy, fsspec, dill, nvidia-cusparse-cu12, nvidia-cudnn-cu12,
multiprocess, nvidia-cusolver-cu12, datasets, clip-score
  Attempting uninstall: nvidia-nvjitlink-cu12
    Found existing installation: nvidia-nvjitlink-cu12 12.5.82
    Uninstalling nvidia-nvjitlink-cu12-12.5.82:
      Successfully uninstalled nvidia-nvjitlink-cu12-12.5.82
  Attempting uninstall: nvidia-curand-cu12
    Found existing installation: nvidia-curand-cu12 10.3.6.82
    Uninstalling nvidia-curand-cu12-10.3.6.82:
      Successfully uninstalled nvidia-curand-cu12-10.3.6.82
  Attempting uninstall: nvidia-cufft-cu12
    Found existing installation: nvidia-cufft-cu12 11.2.3.61
    Uninstalling nvidia-cufft-cu12-11.2.3.61:
```

```
      Successfully uninstalled nvidia-cufft-cu12-11.2.3.61
  Attempting uninstall: nvidia-cuda-runtime-cu12
    Found existing installation: nvidia-cuda-runtime-cu12 12.5.82
    Uninstalling nvidia-cuda-runtime-cu12-12.5.82:
      Successfully uninstalled nvidia-cuda-runtime-cu12-12.5.82
  Attempting uninstall: nvidia-cuda-nvrtc-cu12
    Found existing installation: nvidia-cuda-nvrtc-cu12 12.5.82
    Uninstalling nvidia-cuda-nvrtc-cu12-12.5.82:
      Successfully uninstalled nvidia-cuda-nvrtc-cu12-12.5.82
  Attempting uninstall: nvidia-cuda-cupti-cu12
    Found existing installation: nvidia-cuda-cupti-cu12 12.5.82
    Uninstalling nvidia-cuda-cupti-cu12-12.5.82:
      Successfully uninstalled nvidia-cuda-cupti-cu12-12.5.82
  Attempting uninstall: nvidia-cublas-cu12
    Found existing installation: nvidia-cublas-cu12 12.5.3.2
    Uninstalling nvidia-cublas-cu12-12.5.3.2:
      Successfully uninstalled nvidia-cublas-cu12-12.5.3.2
  Attempting uninstall: fsspec
    Found existing installation: fsspec 2025.3.2
    Uninstalling fsspec-2025.3.2:
      Successfully uninstalled fsspec-2025.3.2
  Attempting uninstall: nvidia-cusparse-cu12
    Found existing installation: nvidia-cusparse-cu12 12.5.1.3
    Uninstalling nvidia-cusparse-cu12-12.5.1.3:
      Successfully uninstalled nvidia-cusparse-cu12-12.5.1.3
  Attempting uninstall: nvidia-cudnn-cu12
    Found existing installation: nvidia-cudnn-cu12 9.3.0.75
    Uninstalling nvidia-cudnn-cu12-9.3.0.75:
      Successfully uninstalled nvidia-cudnn-cu12-9.3.0.75
  Attempting uninstall: nvidia-cusolver-cu12
    Found existing installation: nvidia-cusolver-cu12 11.6.3.83
    Uninstalling nvidia-cusolver-cu12-11.6.3.83:
      Successfully uninstalled nvidia-cusolver-cu12-11.6.3.83
ERROR: pip's dependency resolver does not currently take into account
all the packages that are installed. This behaviour is the source of
the following dependency conflicts.
gcsfs 2025.3.2 requires fsspec==2025.3.2, but you have fsspec 2025.3.0
which is incompatible.
Successfully installed clip-score-0.2.1 datasets-3.6.0 dill-0.3.8
fsspec-2025.3.0 ftfy-6.3.1 multiprocess-0.70.16 nvidia-cublas-cu12-
12.4.5.8 nvidia-cuda-cupti-cu12-12.4.127 nvidia-cuda-nvrtc-cu12-
12.4.127 nvidia-cuda-runtime-cu12-12.4.127 nvidia-cudnn-cu12-9.1.0.70
nvidia-cufft-cu12-11.2.1.3 nvidia-curand-cu12-10.3.5.147 nvidia-
cusolver-cu12-11.6.1.9 nvidia-cusparse-cu12-12.3.1.170 nvidia-
nvjitlink-cu12-12.4.127 xxhash-3.5.0
Collecting git+https://github.com/openai/CLIP.git
  Cloning https://github.com/openai/CLIP.git to /tmp/pip-req-build-
be4o4anr
  Running command git clone --filter=blob:none --quiet
```

```
https://github.com/openai/CLIP.git /tmp/pip-req-build-be4o4anr
  Resolved https://github.com/openai/CLIP.git to commit
dcba3cb2e2827b402d2701e7e1c7d9fed8a20ef1
  Preparing metadata (setup.py) ... ent already satisfied: ftfy in
/usr/local/lib/python3.11/dist-packages (from clip==1.0) (6.3.1)
Requirement already satisfied: packaging in
/usr/local/lib/python3.11/dist-packages (from clip==1.0) (24.2)
Requirement already satisfied: regex in
/usr/local/lib/python3.11/dist-packages (from clip==1.0) (2024.11.6)
Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-
packages (from clip==1.0) (4.67.1)
Requirement already satisfied: torch in
/usr/local/lib/python3.11/dist-packages (from clip==1.0) (2.6.0+cu124)
Requirement already satisfied: torchvision in
/usr/local/lib/python3.11/dist-packages (from clip==1.0)
(0.21.0+cu124)
Requirement already satisfied: wcwidth in
/usr/local/lib/python3.11/dist-packages (from ftfy->clip==1.0)
(0.2.13)
Requirement already satisfied: filelock in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(3.18.0)
Requirement already satisfied: typing-extensions>=4.10.0 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(4.13.2)
Requirement already satisfied: networkx in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(3.4.2)
Requirement already satisfied: jinja2 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(3.1.6)
Requirement already satisfied: fsspec in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(2025.3.0)
Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.4.127 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(12.4.127)
Requirement already satisfied: nvidia-cuda-runtime-cu12==12.4.127
in /usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(12.4.127)
Requirement already satisfied: nvidia-cuda-cupti-cu12==12.4.127 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(12.4.127)
Requirement already satisfied: nvidia-cudnn-cu12==9.1.0.70 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(9.1.0.70)
Requirement already satisfied: nvidia-cublas-cu12==12.4.5.8 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(12.4.5.8)
```

```
Requirement already satisfied: nvidia-cufft-cu12==11.2.1.3 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(11.2.1.3)
Requirement already satisfied: nvidia-curand-cu12==10.3.5.147 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(10.3.5.147)
Requirement already satisfied: nvidia-cusolver-cu12==11.6.1.9 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(11.6.1.9)
Requirement already satisfied: nvidia-cusparse-cu12==12.3.1.170 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(12.3.1.170)
Requirement already satisfied: nvidia-cusparselt-cu12==0.6.2 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(0.6.2)
Requirement already satisfied: nvidia-nccl-cu12==2.21.5 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(2.21.5)
Requirement already satisfied: nvidia-nvtx-cu12==12.4.127 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(12.4.127)
Requirement already satisfied: nvidia-nvjitlink-cu12==12.4.127 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(12.4.127)
Requirement already satisfied: triton==3.2.0 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(3.2.0)
Requirement already satisfied: sympy==1.13.1 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(1.13.1)
Requirement already satisfied: mpmath<1.4,>=1.1.0 in
/usr/local/lib/python3.11/dist-packages (from sympy==1.13.1->torch-
>clip==1.0) (1.3.0)
Requirement already satisfied: numpy in
/usr/local/lib/python3.11/dist-packages (from torchvision->clip==1.0)
(2.0.2)
Requirement already satisfied: pillow!=8.3.*,>=5.3.0 in
/usr/local/lib/python3.11/dist-packages (from torchvision->clip==1.0)
(11.2.1)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.11/dist-packages (from jinja2->torch-
>clip==1.0) (3.0.2)
Building wheels for collected packages: clip
  Building wheel for clip (setup.py) ... e=clip-1.0-py3-none-any.whl
size=1369490
sha256=e08b57eb52bf1d6047a21c1aee074038073f7a762858681b10e36d66aefa81d
2
  Stored in directory:
/tmp/pip-ephem-wheel-cache-sx0izlpc/wheels/3f/7c/a4/9b490845988bf7a4db
33674d52f709f088f64392063872eb9a
```

```
Successfully built clip
Installing collected packages: clip
Successfully installed clip-1.0



from torch.utils.data import Dataset
from PIL import Image
import json
import os

class CocoEvalDataset(Dataset):
    def __init__(self, image_dir, annotation_path, num_samples=500):
        with open(annotation_path) as f:
            data = json.load(f)

        # Get first 2000 images
        self.image_ids = [img['id'] for img in data['images']
[:num_samples]]
        self.image_files = {img['id']: img['file_name'] for img in
data['images']}

        # Map image_id to captions
        self.gt_captions = {}
        for ann in data['annotations']:
            if ann['image_id'] in self.image_ids:
                if ann['image_id'] not in self.gt_captions:
                    self.gt_captions[ann['image_id']] = []

self.gt_captions[ann['image_id']].append(ann['caption'])

        self.image_dir = image_dir

    def __len__(self):
        return len(self.image_ids)

    def __getitem__(self, idx):
        img_id = self.image_ids[idx]
        img_path = os.path.join(self.image_dir,
self.image_files[img_id])
        image = Image.open(img_path).convert("RGB")
        captions = self.gt_captions[img_id]
        return image, captions, str(img_id)

# Initialize dataset
eval_dataset = CocoEvalDataset(
    image_dir="/content/drive/MyDrive/val2017",

annotation_path="/content/drive/MyDrive/annotations/captions_val2017.j
son"
)
```

```
!pip install git+https://github.com/openai/CLIP.git
!pip install clip-score==0.1.0  # Specific working version
```

```
Collecting git+https://github.com/openai/CLIP.git
  Cloning https://github.com/openai/CLIP.git to /tmp/pip-req-build-
_ovhuzit
  Running command git clone --filter=blob:none --quiet
https://github.com/openai/CLIP.git /tmp/pip-req-build-_ovhuzit
  Resolved https://github.com/openai/CLIP.git to commit
dcba3cb2e2827b402d2701e7e1c7d9fed8a20ef1
  Preparing metadata (setup.py) ... ent already satisfied: ftfy in
/usr/local/lib/python3.11/dist-packages (from clip==1.0) (6.3.1)
Requirement already satisfied: packaging in
/usr/local/lib/python3.11/dist-packages (from clip==1.0) (24.2)
Requirement already satisfied: regex in
/usr/local/lib/python3.11/dist-packages (from clip==1.0) (2024.11.6)
Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-
packages (from clip==1.0) (4.67.1)
Requirement already satisfied: torch in
/usr/local/lib/python3.11/dist-packages (from clip==1.0) (2.6.0+cu124)
Requirement already satisfied: torchvision in
/usr/local/lib/python3.11/dist-packages (from clip==1.0)
(0.21.0+cu124)
Requirement already satisfied: wcwidth in
/usr/local/lib/python3.11/dist-packages (from ftfy->clip==1.0)
(0.2.13)
Requirement already satisfied: filelock in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(3.18.0)
Requirement already satisfied: typing-extensions>=4.10.0 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(4.13.2)
Requirement already satisfied: networkx in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(3.4.2)
Requirement already satisfied: jinja2 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(3.1.6)
Requirement already satisfied: fsspec in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(2025.3.0)
Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.4.127 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(12.4.127)
Requirement already satisfied: nvidia-cuda-runtime-cu12==12.4.127
in /usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(12.4.127)
Requirement already satisfied: nvidia-cuda-cupti-cu12==12.4.127 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(12.4.127)
```

```
Requirement already satisfied: nvidia-cudnn-cu12==9.1.0.70 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(9.1.0.70)
Requirement already satisfied: nvidia-cublas-cu12==12.4.5.8 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(12.4.5.8)
Requirement already satisfied: nvidia-cufft-cu12==11.2.1.3 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(11.2.1.3)
Requirement already satisfied: nvidia-curand-cu12==10.3.5.147 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(10.3.5.147)
Requirement already satisfied: nvidia-cusolver-cu12==11.6.1.9 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(11.6.1.9)
Requirement already satisfied: nvidia-cusparse-cu12==12.3.1.170 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(12.3.1.170)
Requirement already satisfied: nvidia-cusparselt-cu12==0.6.2 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(0.6.2)
Requirement already satisfied: nvidia-nccl-cu12==2.21.5 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(2.21.5)
Requirement already satisfied: nvidia-nvtx-cu12==12.4.127 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(12.4.127)
Requirement already satisfied: nvidia-nvjitlink-cu12==12.4.127 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(12.4.127)
Requirement already satisfied: triton==3.2.0 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(3.2.0)
Requirement already satisfied: sympy==1.13.1 in
/usr/local/lib/python3.11/dist-packages (from torch->clip==1.0)
(1.13.1)
Requirement already satisfied: mpmath<1.4,>=1.1.0 in
/usr/local/lib/python3.11/dist-packages (from sympy==1.13.1->torch-
>clip==1.0) (1.3.0)
Requirement already satisfied: numpy in
/usr/local/lib/python3.11/dist-packages (from torchvision->clip==1.0)
(2.0.2)
Requirement already satisfied: pillow!=8.3.*,>=5.3.0 in
/usr/local/lib/python3.11/dist-packages (from torchvision->clip==1.0)
(11.2.1)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.11/dist-packages (from jinja2->torch-
>clip==1.0) (3.0.2)
Collecting clip-score==0.1.0
```

```
  Downloading clip_score-0.1.0-py3-none-any.whl.metadata (3.3 kB)
Requirement already satisfied: numpy in
/usr/local/lib/python3.11/dist-packages (from clip-score==0.1.0)
(2.0.2)
Requirement already satisfied: pillow in
/usr/local/lib/python3.11/dist-packages (from clip-score==0.1.0)
(11.2.1)
Requirement already satisfied: torch>=1.7.1 in
/usr/local/lib/python3.11/dist-packages (from clip-score==0.1.0)
(2.6.0+cu124)
Requirement already satisfied: torchvision>=0.8.2 in
/usr/local/lib/python3.11/dist-packages (from clip-score==0.1.0)
(0.21.0+cu124)
Requirement already satisfied: ftfy in /usr/local/lib/python3.11/dist-
packages (from clip-score==0.1.0) (6.3.1)
Requirement already satisfied: regex in
/usr/local/lib/python3.11/dist-packages (from clip-score==0.1.0)
(2024.11.6)
Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-
packages (from clip-score==0.1.0) (4.67.1)
Requirement already satisfied: filelock in
/usr/local/lib/python3.11/dist-packages (from torch>=1.7.1->clip-
score==0.1.0) (3.18.0)
Requirement already satisfied: typing-extensions>=4.10.0 in
/usr/local/lib/python3.11/dist-packages (from torch>=1.7.1->clip-
score==0.1.0) (4.13.2)
Requirement already satisfied: networkx in
/usr/local/lib/python3.11/dist-packages (from torch>=1.7.1->clip-
score==0.1.0) (3.4.2)
Requirement already satisfied: jinja2 in
/usr/local/lib/python3.11/dist-packages (from torch>=1.7.1->clip-
score==0.1.0) (3.1.6)
Requirement already satisfied: fsspec in
/usr/local/lib/python3.11/dist-packages (from torch>=1.7.1->clip-
score==0.1.0) (2025.3.0)
Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.4.127 in
/usr/local/lib/python3.11/dist-packages (from torch>=1.7.1->clip-
score==0.1.0) (12.4.127)
Requirement already satisfied: nvidia-cuda-runtime-cu12==12.4.127
in /usr/local/lib/python3.11/dist-packages (from torch>=1.7.1->clip-
score==0.1.0) (12.4.127)
Requirement already satisfied: nvidia-cuda-cupti-cu12==12.4.127 in
/usr/local/lib/python3.11/dist-packages (from torch>=1.7.1->clip-
score==0.1.0) (12.4.127)
Requirement already satisfied: nvidia-cudnn-cu12==9.1.0.70 in
/usr/local/lib/python3.11/dist-packages (from torch>=1.7.1->clip-
score==0.1.0) (9.1.0.70)
Requirement already satisfied: nvidia-cublas-cu12==12.4.5.8 in
/usr/local/lib/python3.11/dist-packages (from torch>=1.7.1->clip-
```

```
score==0.1.0) (12.4.5.8)
Requirement already satisfied: nvidia-cufft-cu12==11.2.1.3 in
/usr/local/lib/python3.11/dist-packages (from torch>=1.7.1->clip-
score==0.1.0) (11.2.1.3)
Requirement already satisfied: nvidia-curand-cu12==10.3.5.147 in
/usr/local/lib/python3.11/dist-packages (from torch>=1.7.1->clip-
score==0.1.0) (10.3.5.147)
Requirement already satisfied: nvidia-cusolver-cu12==11.6.1.9 in
/usr/local/lib/python3.11/dist-packages (from torch>=1.7.1->clip-
score==0.1.0) (11.6.1.9)
Requirement already satisfied: nvidia-cusparse-cu12==12.3.1.170 in
/usr/local/lib/python3.11/dist-packages (from torch>=1.7.1->clip-
score==0.1.0) (12.3.1.170)
Requirement already satisfied: nvidia-cusparselt-cu12==0.6.2 in
/usr/local/lib/python3.11/dist-packages (from torch>=1.7.1->clip-
score==0.1.0) (0.6.2)
Requirement already satisfied: nvidia-nccl-cu12==2.21.5 in
/usr/local/lib/python3.11/dist-packages (from torch>=1.7.1->clip-
score==0.1.0) (2.21.5)
Requirement already satisfied: nvidia-nvtx-cu12==12.4.127 in
/usr/local/lib/python3.11/dist-packages (from torch>=1.7.1->clip-
score==0.1.0) (12.4.127)
Requirement already satisfied: nvidia-nvjitlink-cu12==12.4.127 in
/usr/local/lib/python3.11/dist-packages (from torch>=1.7.1->clip-
score==0.1.0) (12.4.127)
Requirement already satisfied: triton==3.2.0 in
/usr/local/lib/python3.11/dist-packages (from torch>=1.7.1->clip-
score==0.1.0) (3.2.0)
Requirement already satisfied: sympy==1.13.1 in
/usr/local/lib/python3.11/dist-packages (from torch>=1.7.1->clip-
score==0.1.0) (1.13.1)
Requirement already satisfied: mpmath<1.4,>=1.1.0 in
/usr/local/lib/python3.11/dist-packages (from sympy==1.13.1-
>torch>=1.7.1->clip-score==0.1.0) (1.3.0)
Requirement already satisfied: wcwidth in
/usr/local/lib/python3.11/dist-packages (from ftfy->clip-score==0.1.0)
(0.2.13)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.11/dist-packages (from jinja2->torch>=1.7.1-
>clip-score==0.1.0) (3.0.2)
Downloading clip_score-0.1.0-py3-none-any.whl (10 kB)
Installing collected packages: clip-score
  Attempting uninstall: clip-score
    Found existing installation: clip-score 0.2.1
    Uninstalling clip-score-0.2.1:
      Successfully uninstalled clip-score-0.2.1
Successfully installed clip-score-0.1.0
```

{"id":"39ec3151ef544f94a612973932109001","pip_warning":{"packages":
["clip_score"]}}

```
!pip install git+https://github.com/openai/CLIP.git
!pip install torch torchvision



import torch
import clip
from PIL import Image
import numpy as np

device = "cuda" if torch.cuda.is_available() else "cpu"

# Load CLIP model
clip_model, clip_preprocess = clip.load("ViT-B/32", device=device)

def calculate_clip_score(image, text):
    """Calculate CLIPScore between image and text"""
    try:
        # Preprocess image
        image_input = clip_preprocess(image).unsqueeze(0).to(device)

        # Tokenize text
        text_input = clip.tokenize([text]).to(device)

        # Extract features
        with torch.no_grad():
            image_features = clip_model.encode_image(image_input)
            text_features = clip_model.encode_text(text_input)

        # Normalize features
        image_features /= image_features.norm(dim=-1, keepdim=True)
        text_features /= text_features.norm(dim=-1, keepdim=True)

        # Calculate similarity (original CLIPScore uses
2.5*max(0,cos_sim))
        similarity = (image_features @ text_features.T).item()
        return 2.5 * max(0, similarity)

    except Exception as e:
        print(f"Error calculating CLIPScore: {e}")
        return 0.0

def evaluate_clipscore(model, processor, dataset, num_samples=None):
    """Evaluate model on dataset using CLIPScore"""
    results = []
    model.eval()

    if num_samples:
        dataset = list(dataset)[:num_samples]  # Limit samples if
specified
```

```
    for idx, (image, gt_captions, img_id) in enumerate(dataset):
        try:
            # Generate caption
            inputs = processor(images=image,
return_tensors="pt").to(device)
            with torch.no_grad():
                outputs = model.generate(**inputs)
            pred_caption = processor.decode(outputs[0],
skip_special_tokens=True)

            # Calculate CLIPScore
            score = calculate_clip_score(image, pred_caption)

            results.append({
                "image_id": img_id,
                "predicted_caption": pred_caption,
                "clip_score": score,
                "reference_captions": gt_captions
            })

            if (idx+1) % 50 == 0:
                print(f"Processed {idx+1} samples...")

        except Exception as e:
            print(f"Error processing sample {img_id}: {e}")
            continue

    avg_score = np.mean([r['clip_score'] for r in results]) if results
else 0.0
    return avg_score, results

from transformers import (
    BlipForConditionalGeneration,
    Blip2ForConditionalGeneration,
    AutoProcessor
)
import pandas as pd

models_to_evaluate = {
    "blip-base": "Salesforce/blip-image-captioning-base",
    "blip2-flan-t5-xl": "Salesforce/blip2-flan-t5-xl-coco",
    "blip2-opt-2.7b": "Salesforce/blip2-opt-2.7b"
}

results = []

for model_name, model_path in models_to_evaluate.items():
    print(f"\nEvaluating {model_name}...")

    # Load model
```

```
    if "blip2" in model_name:
        processor = AutoProcessor.from_pretrained(model_path)
        model = Blip2ForConditionalGeneration.from_pretrained(
            model_path,
            torch_dtype=torch.float16,
            device_map="auto"
        )
    else:
        processor = AutoProcessor.from_pretrained(model_path)
        model =
BlipForConditionalGeneration.from_pretrained(model_path).to(device)

    # Evaluate
    avg_score, detailed_results = evaluate_clipscore(model, processor,
eval_dataset)
    results.append({
        "model": model_name,
        "avg_clip_score": avg_score
    })

    # Save detailed results

pd.DataFrame(detailed_results).to_csv(f"{model_name}_clipscore_results
.csv", index=False)

    # Clean up
    del model, processor
    torch.cuda.empty_cache()

# Print summary
print("\nEvaluation Summary:")
print(pd.DataFrame(results))
```

```
Evaluating blip-base...
Processed 50 samples...
Processed 100 samples...
Processed 150 samples...
Processed 200 samples...
Processed 250 samples...
Processed 300 samples...
Processed 350 samples...
Processed 400 samples...
Processed 450 samples...
Processed 500 samples...

Evaluating blip2-flan-t5-xl...
```


```
{"model_id":"2fbb373752734f26b9dc2db717fceaf7","version_major":2,"vers
ion_minor":0}
```

{"model_id":"19a2cc9e3474454f9208e30ff67b80c1","version_major":2,"version_minor":0}

{"model_id":"67d502fcd7dd430984cdcf2a8b1c48f1","version_major":2,"version_minor":0}

{"model_id":"816be02403a04aa3bf3a337561cd2061","version_major":2,"version_minor":0}

{"model_id":"84d58c4c7c6340c19019b24e50c768ef","version_major":2,"version_minor":0}

{"model_id":"f4a0d75f36094cf5a30eb4c87367923a","version_major":2,"version_minor":0}

{"model_id":"1eb570112adf40618ce318be9c72c1e0","version_major":2,"version_minor":0}

{"model_id":"27ff07836b8f4278a3335ee9ed102b97","version_major":2,"version_minor":0}

{"model_id":"c0e8339580e34a19917b3c1f0a1d679f","version_major":2,"version_minor":0}

{"model_id":"1f35a1583abc46699c3744a8f8bfe430","version_major":2,"version_minor":0}

{"model_id":"476a52ccd6ee4fb0a69f1ab2d64541e2","version_major":2,"version_minor":0}

```
Expanding inputs for image tokens in BLIP-2 should be done in
processing. Please follow instruction here
(https://gist.github.com/zucchini-nlp/e9f20b054fa322f84ac9311d9ab67042
) to update your BLIP-2 model. Using processors without these
attributes in the config is deprecated and will throw an error in
v4.50.

Processed 50 samples...
Processed 100 samples...
Processed 150 samples...
Processed 200 samples...
Processed 250 samples...
Processed 300 samples...
Processed 350 samples...
Processed 400 samples...
Processed 450 samples...
Processed 500 samples...

Evaluating blip2-opt-2.7b...
```

{"model_id":"5a96507106404b958074ecbe2114a533","version_major":2,"version_minor":0}

{"model_id":"aa9799b14f074e29b6e6109bcf0cd4bc","version_major":2,"version_minor":0}

{"model_id":"73d73031d9c7431fbe5637bf40d5b59e","version_major":2,"version_minor":0}

{"model_id":"8f9bdab37140423bb008311216ba90d4","version_major":2,"version_minor":0}

{"model_id":"de39ab04131e4a319141a48ff788199d","version_major":2,"version_minor":0}

{"model_id":"d941cdecf71b4bf8b3992a3fbf5c74c7","version_major":2,"version_minor":0}

{"model_id":"b056f7fe5901415cbbcf2586e884024d","version_major":2,"version_minor":0}

{"model_id":"458634122bea4490b584161202e60b8d","version_major":2,"version_minor":0}

{"model_id":"ea140bfc93e344dfa8624c08a3897724","version_major":2,"version_minor":0}

{"model_id":"402b7448eb2a4c1fbc9c35acbd3ff791","version_major":2,"version_minor":0}

{"model_id":"b7e0e9c184e546019e888a41a196f3c4","version_major":2,"version_minor":0}

{"model_id":"17a0282270834766ba582e99d037e547","version_major":2,"version_minor":0}

{"model_id":"e9ea3fbf64e84eb48e9754ec1dca47cc","version_major":2,"version_minor":0}

{"model_id":"32d19a0c69114a6d9d47731389091452","version_major":2,"version_minor":0}

{"model_id":"c6837654699e4b20b7d2dc125565cfc7","version_major":2,"version_minor":0}

```
Processed 50 samples...
Processed 100 samples...
Processed 150 samples...
Processed 200 samples...
Processed 250 samples...
Processed 300 samples...
Processed 350 samples...
Processed 400 samples...
Processed 450 samples...
Processed 500 samples...
```

```
Evaluation Summary:
              model  avg_clip_score
0         blip-base        0.730167
1  blip2-flan-t5-xl        0.749786
2    blip2-opt-2.7b        0.759352
```

Clip Score calculation for fine-tuned model

```python
from transformers import BlipForConditionalGeneration, AutoProcessor
import torch
from PIL import Image
import clip
import pandas as pd

# 1. Load your fine-tuned model
output_dir = "/content/drive/MyDrive/finetuned/final_model"
device = "cuda" if torch.cuda.is_available() else "cpu"

# Load processor and model
processor = AutoProcessor.from_pretrained(output_dir)
model =
BlipForConditionalGeneration.from_pretrained(output_dir).to(device)
model.eval()

# 2. Load CLIP model for scoring
clip_model, clip_preprocess = clip.load("ViT-B/32", device=device)

def calculate_clip_score(image, text):
    """Calculate CLIPScore between image and text"""
    try:
        image_input = clip_preprocess(image).unsqueeze(0).to(device)
        text_input = clip.tokenize([text]).to(device)

        with torch.no_grad():
            image_features = clip_model.encode_image(image_input)
            text_features = clip_model.encode_text(text_input)

        # Normalize features and calculate score
        image_features /= image_features.norm(dim=-1, keepdim=True)
        text_features /= text_features.norm(dim=-1, keepdim=True)
        similarity = (image_features @ text_features.T).item()
        return 2.5 * max(0, similarity)
    except Exception as e:
        print(f"Error calculating CLIPScore: {e}")
        return 0.0

# 3. Evaluation function for your fine-tuned model
```

```
def evaluate_finetuned_model(dataset, num_samples=None):
    results = []
    dataset = list(dataset)[:num_samples] if num_samples else dataset

    for idx, (image, gt_captions, img_id) in enumerate(dataset):
        try:
            # Generate caption
            inputs = processor(images=image,
return_tensors="pt").to(device)
            with torch.no_grad():
                outputs = model.generate(**inputs)
            pred_caption = processor.decode(outputs[0],
skip_special_tokens=True)

            # Calculate CLIPScore
            score = calculate_clip_score(image, pred_caption)

            results.append({
                "image_id": img_id,
                "predicted_caption": pred_caption,
                "clip_score": score,
                "reference_captions": gt_captions
            })

            if (idx+1) % 20 == 0:
                print(f"Processed {idx+1}/{len(dataset)} samples...")

        except Exception as e:
            print(f"Error processing {img_id}: {e}")
            continue

    return results

# 4. Run evaluation on your validation dataset
eval_results = evaluate_finetuned_model(eval_dataset, num_samples=500)
# Adjust sample size

# 5. Analyze results
avg_score = sum(r['clip_score'] for r in eval_results) /
len(eval_results)
print(f"\nAverage CLIPScore: {avg_score:.4f}")

# Save detailed results
results_df = pd.DataFrame(eval_results)
results_df.to_csv("/content/drive/MyDrive/finetuned/clipscore_results.
csv", index=False)
'''
# Optional: Compare with original model
print("\nComparing with original model...")
original_model = BlipForConditionalGeneration.from_pretrained(
```

```
    "Salesforce/blip-image-captioning-base"
).to(device)
original_results = evaluate_finetuned_model(eval_dataset,
num_samples=200)  # Same function works
original_avg = sum(r['clip_score'] for r in original_results) /
len(original_results)
'''
print(f"\nFine-tuned model score: {avg_score:.4f}")
#print(f"Original model score: {original_avg:.4f}")
#print(f"Improvement: {avg_score - original_avg:+.4f}")
```

```
Processed 20/500 samples...
Processed 40/500 samples...
Processed 60/500 samples...
Processed 80/500 samples...
Processed 100/500 samples...
Processed 120/500 samples...
Processed 140/500 samples...
Processed 160/500 samples...
Processed 180/500 samples...
Processed 200/500 samples...
Processed 220/500 samples...
Processed 240/500 samples...
Processed 260/500 samples...
Processed 280/500 samples...
Processed 300/500 samples...
Processed 320/500 samples...
Processed 340/500 samples...
Processed 360/500 samples...
Processed 380/500 samples...
Processed 400/500 samples...
Processed 420/500 samples...
Processed 440/500 samples...
Processed 460/500 samples...
Processed 480/500 samples...
Processed 500/500 samples...

Average CLIPScore: 0.7715

Fine-tuned model score: 0.7715
```