```
!pip install diffusers --upgrade
!pip install invisible_watermark transformers accelerate safetensors
```

Requirement already satisfied: diffusers in
/usr/local/lib/python3.11/dist-packages (0.33.1)
Requirement already satisfied: importlib-metadata in
/usr/local/lib/python3.11/dist-packages (from diffusers) (8.7.0)
Requirement already satisfied: filelock in
/usr/local/lib/python3.11/dist-packages (from diffusers) (3.18.0)
Requirement already satisfied: huggingface-hub>=0.27.0 in
/usr/local/lib/python3.11/dist-packages (from diffusers) (0.30.2)
Requirement already satisfied: numpy in
/usr/local/lib/python3.11/dist-packages (from diffusers) (2.0.2)
Requirement already satisfied: regex!=2019.12.17 in
/usr/local/lib/python3.11/dist-packages (from diffusers) (2024.11.6)
Requirement already satisfied: requests in
/usr/local/lib/python3.11/dist-packages (from diffusers) (2.32.3)
Requirement already satisfied: safetensors>=0.3.1 in
/usr/local/lib/python3.11/dist-packages (from diffusers) (0.5.3)
Requirement already satisfied: Pillow in
/usr/local/lib/python3.11/dist-packages (from diffusers) (11.2.1)
Requirement already satisfied: fsspec>=2023.5.0 in
/usr/local/lib/python3.11/dist-packages (from huggingface-hub>=0.27.0-
>diffusers) (2025.3.2)
Requirement already satisfied: packaging>=20.9 in
/usr/local/lib/python3.11/dist-packages (from huggingface-hub>=0.27.0-
>diffusers) (24.2)
Requirement already satisfied: pyyaml>=5.1 in
/usr/local/lib/python3.11/dist-packages (from huggingface-hub>=0.27.0-
>diffusers) (6.0.2)
Requirement already satisfied: tqdm>=4.42.1 in
/usr/local/lib/python3.11/dist-packages (from huggingface-hub>=0.27.0-
>diffusers) (4.67.1)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
/usr/local/lib/python3.11/dist-packages (from huggingface-hub>=0.27.0-
>diffusers) (4.13.2)
Requirement already satisfied: zipp>=3.20 in
/usr/local/lib/python3.11/dist-packages (from importlib-metadata-
>diffusers) (3.21.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests->diffusers)
(3.4.1)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.11/dist-packages (from requests->diffusers)
(3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.11/dist-packages (from requests->diffusers)
(2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.11/dist-packages (from requests->diffusers)

```
(2025.4.26)
Collecting invisible_watermark
  Downloading invisible_watermark-0.2.0-py3-none-any.whl.metadata (8.2
kB)
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: safetensors in
/usr/local/lib/python3.11/dist-packages (0.5.3)
Requirement already satisfied: Pillow>=6.0.0 in
/usr/local/lib/python3.11/dist-packages (from invisible_watermark)
(11.2.1)
Collecting PyWavelets>=1.1.1 (from invisible_watermark)
  Downloading pywavelets-1.8.0-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (9.0 kB)
Requirement already satisfied: numpy>=1.17.0 in
/usr/local/lib/python3.11/dist-packages (from invisible_watermark)
(2.0.2)
Requirement already satisfied: opencv-python>=4.1.0.25 in
/usr/local/lib/python3.11/dist-packages (from invisible_watermark)
(4.11.0.86)
Requirement already satisfied: torch in
/usr/local/lib/python3.11/dist-packages (from invisible_watermark)
(2.6.0+cu124)
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: 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: 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: 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: networkx in
/usr/local/lib/python3.11/dist-packages (from torch-
>invisible_watermark) (3.4.2)
Requirement already satisfied: jinja2 in
/usr/local/lib/python3.11/dist-packages (from torch-
>invisible_watermark) (3.1.6)
Collecting nvidia-cuda-nvrtc-cu12==12.4.127 (from torch-
>invisible_watermark)
  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-
>invisible_watermark)
  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-
>invisible_watermark)
  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-
>invisible_watermark)
  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-
>invisible_watermark)
  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-
>invisible_watermark)
  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-
>invisible_watermark)
  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-
>invisible_watermark)
  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-
>invisible_watermark)
  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-
>invisible_watermark) (0.6.2)
Requirement already satisfied: nvidia-nccl-cu12==2.21.5 in
/usr/local/lib/python3.11/dist-packages (from torch-
>invisible_watermark) (2.21.5)

Requirement already satisfied: nvidia-nvtx-cu12==12.4.127 in
/usr/local/lib/python3.11/dist-packages (from torch-
>invisible_watermark) (12.4.127)
Collecting nvidia-nvjitlink-cu12==12.4.127 (from torch-
>invisible_watermark)
  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-
>invisible_watermark) (3.2.0)
Requirement already satisfied: sympy==1.13.1 in
/usr/local/lib/python3.11/dist-packages (from torch-
>invisible_watermark) (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-
>invisible_watermark) (1.3.0)
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: MarkupSafe>=2.0 in
/usr/local/lib/python3.11/dist-packages (from jinja2->torch-
>invisible_watermark) (3.0.2)
Downloading invisible_watermark-0.2.0-py3-none-any.whl (1.6 MB)
━━━ 1.6/1.6 MB 25.0 MB/s eta
0:00:00
anylinux_2_17_x86_64.manylinux2014_x86_64.whl (4.5 MB)
━━━ 4.5/4.5 MB 31.2 MB/s eta
0:00:00
anylinux2014_x86_64.whl (363.4 MB)
━━━ 363.4/363.4 MB 4.2 MB/s eta
0:00:00
anylinux2014_x86_64.whl (13.8 MB)
━━━ 13.8/13.8 MB 46.8 MB/s eta
0:00:00
anylinux2014_x86_64.whl (24.6 MB)
━━━ 24.6/24.6 MB 29.2 MB/s eta
0:00:00
e_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl (883 kB)
━━━ 883.7/883.7 kB 39.9 MB/s eta
0:00:00

anylinux2014_x86_64.whl (664.8 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 664.8/664.8 MB 2.9 MB/s eta
0:00:00
anylinux2014_x86_64.whl (211.5 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 211.5/211.5 MB 5.4 MB/s eta
0:00:00
anylinux2014_x86_64.whl (56.3 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 56.3/56.3 MB 14.9 MB/s eta
0:00:00
anylinux2014_x86_64.whl (127.9 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 127.9/127.9 MB 7.0 MB/s eta
0:00:00
anylinux2014_x86_64.whl (207.5 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 207.5/207.5 MB 6.2 MB/s eta
0:00:00
anylinux2014_x86_64.whl (21.1 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 21.1/21.1 MB 77.1 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,
invisible_watermark
  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 PyWavelets-1.8.0 invisible_watermark-0.2.0
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
```

```python
#----------------------------IMPORT
LIBRARIES--------------------------------
import torch
from diffusers import DiffusionPipeline

import cv2
from PIL import Image
import matplotlib.pyplot as plt
from transformers import VisionEncoderDecoderModel, ViTImageProcessor,
AutoTokenizer

import tensorflow as tf
import numpy as np
from tensorflow.keras.datasets import cifar10

from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
import numpy as np

#------------------------------------DEFINE
MODEL-----------------------------------------------

model = VisionEncoderDecoderModel.from_pretrained("nlpconnect/vit-
gpt2-image-captioning") #load pre-trained model from hugging face ;
Vision Transformer (ViT) to understand the image and a GPT-2 model to
generate the caption
feature_extractor = ViTImageProcessor.from_pretrained("nlpconnect/vit-
gpt2-image-captioning") # Get Image Feature Extractor
tokenizer = AutoTokenizer.from_pretrained("nlpconnect/vit-gpt2-image-
captioning") # Get Tokenizer Mode
```

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

/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":"d5df331f3aee4943a140954b11cdbb4b","version_major":2,"version_minor":0}

Xet Storage is enabled for this repo, but the 'hf_xet' package is not
installed. Falling back to regular HTTP download. For better
performance, install the package with: `pip install
huggingface_hub[hf_xet]` or `pip install hf_xet`
WARNING:huggingface_hub.file_download:Xet Storage is enabled for this
repo, but the 'hf_xet' package is not installed. Falling back to
regular HTTP download. For better performance, install the package
with: `pip install huggingface_hub[hf_xet]` or `pip install hf_xet`

{"model_id":"01a7966552304a31937e80a2c8d60166","version_major":2,"version_minor":0}

Xet Storage is enabled for this repo, but the 'hf_xet' package is not
installed. Falling back to regular HTTP download. For better
performance, install the package with: `pip install
huggingface_hub[hf_xet]` or `pip install hf_xet`
WARNING:huggingface_hub.file_download:Xet Storage is enabled for this
repo, but the 'hf_xet' package is not installed. Falling back to
regular HTTP download. For better performance, install the package
with: `pip install huggingface_hub[hf_xet]` or `pip install hf_xet`

{"model_id":"013e4b6559f24d6bb86169d7650f08bd","version_major":2,"version_minor":0}

Config of the encoder: <class
'transformers.models.vit.modeling_vit.ViTModel'> is overwritten by
shared encoder config: ViTConfig {
  "architectures": [
    "ViTModel"
  ],
  "attention_probs_dropout_prob": 0.0,
  "encoder_stride": 16,
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.0,

```
  "hidden_size": 768,
  "image_size": 224,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-12,
  "model_type": "vit",
  "num_attention_heads": 12,
  "num_channels": 3,
  "num_hidden_layers": 12,
  "patch_size": 16,
  "pooler_act": "tanh",
  "pooler_output_size": 768,
  "qkv_bias": true,
  "torch_dtype": "float32",
  "transformers_version": "4.51.3"
}

Config of the decoder: <class
'transformers.models.gpt2.modeling_gpt2.GPT2LMHeadModel'> is
overwritten by shared decoder config: GPT2Config {
  "activation_function": "gelu_new",
  "add_cross_attention": true,
  "architectures": [
    "GPT2LMHeadModel"
  ],
  "attn_pdrop": 0.1,
  "bos_token_id": 50256,
  "decoder_start_token_id": 50256,
  "embd_pdrop": 0.1,
  "eos_token_id": 50256,
  "initializer_range": 0.02,
  "is_decoder": true,
  "layer_norm_epsilon": 1e-05,
  "model_type": "gpt2",
  "n_ctx": 1024,
  "n_embd": 768,
  "n_head": 12,
  "n_inner": null,
  "n_layer": 12,
  "n_positions": 1024,
  "pad_token_id": 50256,
  "reorder_and_upcast_attn": false,
  "resid_pdrop": 0.1,
  "scale_attn_by_inverse_layer_idx": false,
  "scale_attn_weights": true,
  "summary_activation": null,
  "summary_first_dropout": 0.1,
  "summary_proj_to_labels": true,
  "summary_type": "cls_index",
```
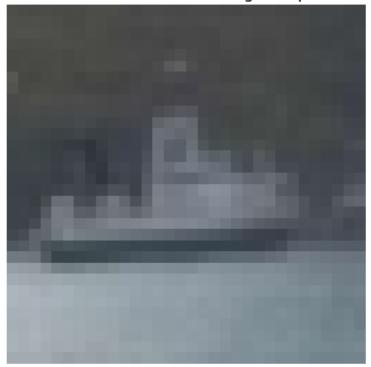
```
  "summary_use_proj": true,
  "task_specific_params": {
    "text-generation": {
      "do_sample": true,
      "max_length": 50
    }
  },
  "torch_dtype": "float32",
  "transformers_version": "4.51.3",
  "use_cache": true,
  "vocab_size": 50257
}
```

{"model_id":"d0d88e9cb376447381c0b1680506bc91","version_major":2,"version_minor":0}

{"model_id":"767a2d4cffdf46a1940e5cdf793a9a95","version_major":2,"version_minor":0}

{"model_id":"760ed27fca8c42d29c024da25862acee","version_major":2,"version_minor":0}

{"model_id":"1f8da688107c48c6b28d05369fdc12d1","version_major":2,"version_minor":0}

{"model_id":"6a7b17d171a64baca7299a9e15bdc521","version_major":2,"version_minor":0}

{"model_id":"1dae4e0b562f409c9be2ae2d1e116141","version_major":2,"version_minor":0}

```
VisionEncoderDecoderModel(
  (encoder): ViTModel(
    (embeddings): ViTEmbeddings(
      (patch_embeddings): ViTPatchEmbeddings(
        (projection): Conv2d(3, 768, kernel_size=(16, 16), stride=(16, 16))
      )
      (dropout): Dropout(p=0.0, inplace=False)
    )
    (encoder): ViTEncoder(
      (layer): ModuleList(
        (0-11): 12 x ViTLayer(
          (attention): ViTAttention(
            (attention): ViTSelfAttention(
              (query): Linear(in_features=768, out_features=768, bias=True)
              (key): Linear(in_features=768, out_features=768, bias=True)
              (value): Linear(in_features=768, out_features=768,
```

```
bias=True)
            )
            (output): ViTSelfOutput(
              (dense): Linear(in_features=768, out_features=768,
bias=True)
              (dropout): Dropout(p=0.0, inplace=False)
            )
          )
          (intermediate): ViTIntermediate(
            (dense): Linear(in_features=768, out_features=3072,
bias=True)
            (intermediate_act_fn): GELUActivation()
          )
          (output): ViTOutput(
            (dense): Linear(in_features=3072, out_features=768,
bias=True)
            (dropout): Dropout(p=0.0, inplace=False)
          )
          (layernorm_before): LayerNorm((768,), eps=1e-12,
elementwise_affine=True)
          (layernorm_after): LayerNorm((768,), eps=1e-12,
elementwise_affine=True)
        )
      )
    )
    (layernorm): LayerNorm((768,), eps=1e-12, elementwise_affine=True)
    (pooler): ViTPooler(
      (dense): Linear(in_features=768, out_features=768, bias=True)
      (activation): Tanh()
    )
  )
  (decoder): GPT2LMHeadModel(
    (transformer): GPT2Model(
      (wte): Embedding(50257, 768)
      (wpe): Embedding(1024, 768)
      (drop): Dropout(p=0.1, inplace=False)
      (h): ModuleList(
        (0-11): 12 x GPT2Block(
          (ln_1): LayerNorm((768,), eps=1e-05,
elementwise_affine=True)
          (attn): GPT2Attention(
            (c_attn): Conv1D(nf=2304, nx=768)
            (c_proj): Conv1D(nf=768, nx=768)
            (attn_dropout): Dropout(p=0.1, inplace=False)
            (resid_dropout): Dropout(p=0.1, inplace=False)
          )
          (ln_2): LayerNorm((768,), eps=1e-05,
elementwise_affine=True)
          (crossattention): GPT2Attention(
```

```
            (c_attn): Conv1D(nf=1536, nx=768)
            (q_attn): Conv1D(nf=768, nx=768)
            (c_proj): Conv1D(nf=768, nx=768)
            (attn_dropout): Dropout(p=0.1, inplace=False)
            (resid_dropout): Dropout(p=0.1, inplace=False)
          )
          (ln_cross_attn): LayerNorm((768,), eps=1e-05,
elementwise_affine=True)
          (mlp): GPT2MLP(
            (c_fc): Conv1D(nf=3072, nx=768)
            (c_proj): Conv1D(nf=768, nx=3072)
            (act): NewGELUActivation()
            (dropout): Dropout(p=0.1, inplace=False)
          )
        )
      )
      (ln_f): LayerNorm((768,), eps=1e-05, elementwise_affine=True)
    )
    (lm_head): Linear(in_features=768, out_features=50257, bias=False)
  )
)

#--------------------------------PREDICT STEP FOR
IMAGES----------------------------------------------------------

max_length = 16  # The maximum length the generated tokens can have
num_beams = 4 #  Number of beams for beam search
gen_kwargs = {"max_length": max_length, "num_beams": num_beams} #
Generation Config

def predict_step(images):  # predict list of  images
    pil_images = []
    for img_array in images:
        img = Image.fromarray(img_array.astype('uint8'), 'RGB')
        pil_images.append(img)

    pixel_values = feature_extractor(images=pil_images,
return_tensors="pt").pixel_values  # Feature extractor
    pixel_values = pixel_values.to(device)

    output_ids = model.generate(pixel_values, **gen_kwargs)  # Apply
model

    preds = tokenizer.batch_decode(output_ids,
skip_special_tokens=True) # Get text tokens description of image
    preds = [pred.strip() for pred in preds]
    return preds

#-----------------------LOAD DATASET AND TEST
MODEL--------------------------
```

```python
(x_train, y_train), (x_test, y_test) = cifar10.load_data() # Load
CIFAR-10
class_names = ['airplane', 'automobile', 'bird', 'cat', 'deer','dog',
'frog', 'horse', 'ship', 'truck'] #Labels

idx = np.random.randint(0, len(x_train)) # Pick a random index
img_array = x_train[idx]  # Input shape of dataset (32, 32, 3)- input
shape of dataset
label = class_names[y_train[idx][0]]

img = Image.fromarray(img_array) # Convert to high-quality image

plt.imshow(img) # Display the image
plt.title(f"Random CIFAR-10 Image: {label}")
plt.axis('off')
plt.show()
print(predict_step([x_train[idx]]))
```

```
Downloading data from https://www.cs.toronto.edu/~kriz/cifar-10-
python.tar.gz
170498071/170498071 ━━━━━━━━━━━━━━━━━ 4s 0us/step
```

Random CIFAR-10 Image: ship


```
The attention mask is not set and cannot be inferred from input
because pad token is same as eos token. As a consequence, you may
```

```
observe unexpected behavior. Please pass your input's `attention_mask`
to obtain reliable results.
We strongly recommend passing in an `attention_mask` since your
input_ids may be padded. See
https://huggingface.co/docs/transformers/troubleshooting#incorrect-
output-when-padding-tokens-arent-masked.
You may ignore this warning if your `pad_token_id` (50256) is
identical to the `bos_token_id` (50256), `eos_token_id` (50256), or
the `sep_token_id` (None), and your input is not padded.

['a boat is docked in a body of water']

#------------------------------
VALIDATION---------------------------------------------------

num_samples = 5
x_sample = x_test[:num_samples]
y_true = y_test[:num_samples].flatten()
y_pred = predict_step(x_sample)
true_labels = [class_names[i] for i in y_true]


fig, axes = plt.subplots(num_samples, 2, figsize=(8, num_samples * 2))
for i in range(num_samples):
    img = x_sample[i]

    # True label
    axes[i, 0].imshow(img)
    axes[i, 0].axis('off')
    axes[i, 0].set_title(f"True: {true_labels[i]}", color='green')

    # Predicted label
    axes[i, 1].imshow(img)
    axes[i, 1].axis('off')
    axes[i, 1].set_title(f"Predicted: {y_pred[i]}", color='red')

plt.show()
```

True: cat   Predicted: a cat laying on top of a pile of clothes

True: ship   Predicted: a row of boats lined up in a row

True: ship   Predicted: a white and blue boat on the water

True: airplane   Predicted: a cartoon of a bird flying through the air

True: frog Predicted: a small black and white bird standing on top of a green field