CLIP Text Method Evaluation Framework

This repository provides code for evaluating CLIP-based models using GPT-generated text descriptions.

The `main.py` script runs the following steps:

1. Loads a CLIP model and a GPT model
2. Loads a dataset
3. Generates text descriptions for each image in the dataset using the GPT model
4. Computes the consistency score, silhouette score, and compound score of the generated text descriptions
5. Computes the zero-shot accuracy of the model and computes the correlation between the zero-shot accuracy and the consistency score, silhouette score, and compound score

The `generate_captions_alone.py` script generates descriptive captions for each class in the dataset only.

Features

- Support for multiple vision-language models:
  - OpenAI CLIP
  - Google SigLIP
  - Facebook FLAVA
- Integration with various CLIP Benchmark datasets
- Text generation using GPT models
- Consistency score computation

Installation

1. Install the required dependencies:

```
pip install -r requirements.txt
```

2. Set up your OpenAI API key in a `.env` file:

```
OPENAI_API_KEY=your_api_key_here
```

Usage

Basic Usage

Run the main evaluation script:

```
python main.py --config configs/config.yaml
```

Generate Captions Only

To only generate and cache captions without running evaluations:

```
python generate_captions_alone.py --config configs/config.yaml
```

Configuration

The framework is configured through `configs/config.yaml`. Key configuration options include:

- Dataset selection
- Model selection
- Experiment parameters
- Consistency scorer settings
- GPT model selection

Example configuration:

```yaml
datasets:
  - name: wds_cars
  - name: wds_fgvc_aircraft
  # Add more datasets as needed

models:
  - name: openai/clip-vit-base-patch32
  - name: google/siglip-large-patch16-384
  - name: facebook/flava-full

experiment_params:
  batch_size: 64
  num_captions: 35
  save_results: True
  enable_plot: False
```

Project Structure

- `main.py`: Main evaluation script
- `generate_captions_alone.py`: Standalone caption generation script
- `configs/`: Configuration files
- `utils/`: Utility functions and classes
  - `dataset.py`: Dataset loading and processing
  - `model_utils.py`: Model loading and inference
  - `evaluation_utils.py`: Evaluation metrics
  - `visualization_utils.py`: Plotting and visualization
  - `utils.py`: Caption generation and misc utilities