Implementation Details

Environment Setup :

- Install Python 3.7 or later.

- Required libraries: PyTorch, torchvision, numpy, and OpenCV for image processing.

- Additional deep learning libraries: Install `DETR`, `Deformable DETR`, and other relevant transformer-based packages.

- Use `pip` or `conda` to install the packages

Data Preparation

Ensure that the infrared and visible image pairs are downloaded and preprocessed. This includes the **M3FD, FLIR, LLVIP, and VEDAI** datasets.

- Organize the datasets into proper train/test splits with annotations in COCO format or as specified in the dataset documentation.

Training Process

- Use pre-trained weights from the COCO dataset to initialize the CNN backbones.

- Train the model on infrared-visible image pair using the multispectral deformable cross-attention to handle complementary feature fusion.

- The loss function should include object detection loss (bounding boxes and class labels).

Evaluation

- After training, evaluate the model on the validation set of each dataset using mAP (mean Average Precision) at different thresholds (e.g., mAP50, mAP75).

Result on Dataset

1. M3FD Dataset:

- Mean Average Precision (mAP50): 80.2%
- Mean Average Precision (mAP75): 56.0%
- Overall mAP: 52.9%

2. FLIR Dataset:

- mAP50: 86.6%
- mAP75: 48.1%
- Overall mAP: 49.3%

3. LLVIP Dataset:

- mAP50: 97.9%
- mAP75: 79.1%
- Overall mAP: 69.6%

4. VEDAI Dataset (for small object detection):

- mAP50: 91.5%
- Overall mAP: 55.3%