Query-Based Object Detection

Overview

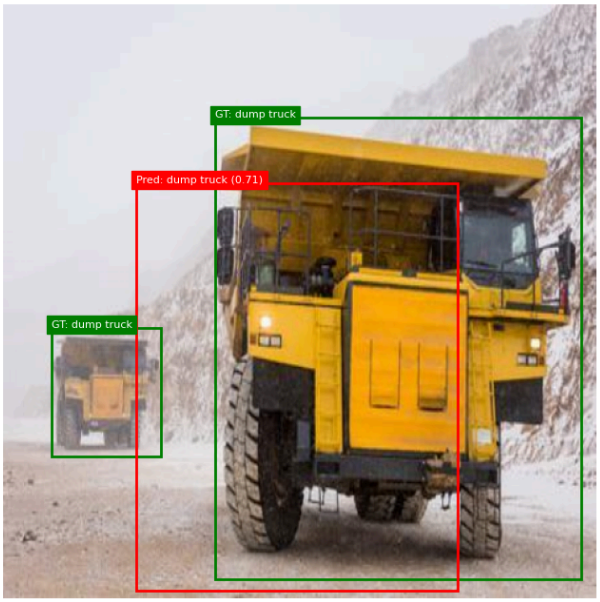
This code provides a concise implementation of a DETR-style (Detection Transformer) object detection model. It uses a **pre-trained ResNet-18** as a feature-extracting backbone, followed by a Transformer Decoder. The core of the model lies in its use of learnable object , which act as initial "slots" for potential objects. The decoder refines these queries by attending to the image features, ultimately populating each slot with object-specific information. Finally, two prediction heads interpret the output from the decoder to produce a class label and a bounding box for each query, allowing the model to perform end-to-end object detection.

Results

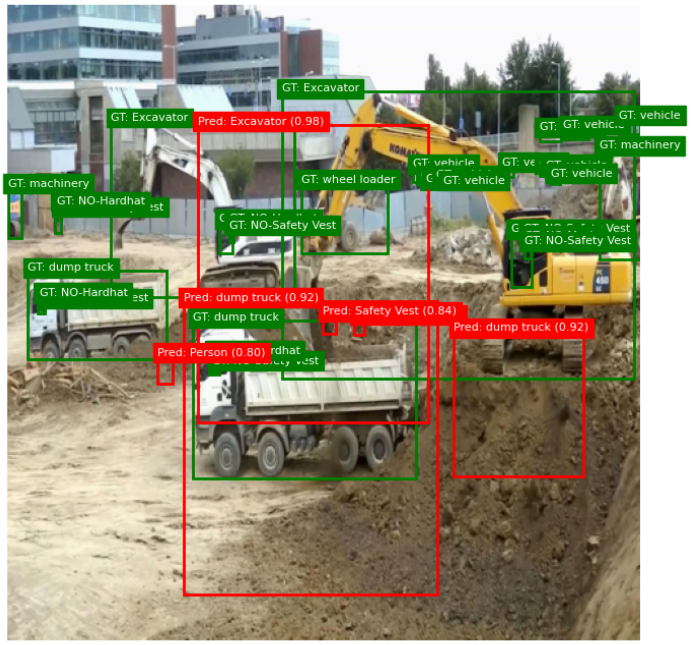
The training results, as shown in the loss graph, indicate that the model is learning successfully. The training loss shows a **strong, consistent downward trend**, which confirms that the model is effectively optimizing its weights on the training data.

The **validation loss** begins to plateau and flatten out after approximately seven epochs. This growing gap between the steadily decreasing training loss and the stagnant validation loss is a classic sign of **overfitting**,

Outputs



Detection 1



Detection 2