Documentation of the Custom Siamese Architecture Used in the Model

Introduction

The Siamese architecture is used to train a model through contrastive learning. This approach utilizes a twin encoder network to extract latent features from images. These latent features are compared for similarity using algorithms such as KL Divergence or Euclidean distance. During training, the encoder learns to identify specific features within the images, generating embeddings that are used in a regression model to calculate the percentage similarity.

Model Inputs

The model operates in two distinct phases:

1. Training Phase:
   - Faulty Image
   - Reference Image
   - Label (% similarity)
2. Test Phase:
   - Test Image

Model Predictions

The model predicts the posture accuracy of individuals performing gym activities, outputting a percentage score indicating the correctness of the posture. While the current focus is on posture detection, the architecture can be adapted for other applications as well.

Model Scope

This architecture can be extended by integrating it with an LSTM or Transformer model to generate textual feedback on posture. Furthermore, it is scalable and can be adapted to solve a variety of other tasks.