



- home
- papers
- study
- life
- About me

CLIP

CLIP is a self-supervised learning framework that enables zero-shot transfer learning across a wide variety of tasks. It consists of a `TextTransformer` and a `VisionTransformer` that share a common embedding space. The model is trained on a dataset of image-text pairs, where the goal is to maximize the similarity between the embeddings of the image and text. The model is trained using a cross-entropy loss function. The model is trained on a dataset of image-text pairs, where the goal is to maximize the similarity between the embeddings of the image and text. The model is trained using a cross-entropy loss function. The model is trained on a dataset of image-text pairs, where the goal is to maximize the similarity between the embeddings of the image and text. The model is trained using a cross-entropy loss function.

[Newer](#)

[Older](#)

8th May 2025

LLaVA

4th December 2024

BNLN