# CLIP模型

1. 添加新编码模型
   1. 残差结构
      1. CLIP-Adapter: Better Vision-Language Models with Feature Adapters
   2. transformer解码器
      1. Frozen CLIP Models are Efficient Video Learners
2. 对CLIP编码器进行修改：
   1. 添加时间、位置和注意力信息
      1. Frozen CLIP Models are Efficient Video Learners
3. 图像特征与文本特征关系
   1. 图像特征与文本特征相等
      1. CLIP-Forge: Towards Zero-Shot Text-to-Shape Generation
   2. 图像特征与文本特征属于同一种特征
      1. Multimodality Helps Unimodality: Cross-Modal Few-Shot Learning With Multimodal Models
   3. 利用图像文本之间的自信息
      1. ZegCLIP: Towards Adapting CLIP for Zero-shot Semantic Segmentation
4. 提示词
   1. Learning to Prompt for Vision-Language Models
5. 设置不同的损失函数
   1. focal loss
      1. ZegCLIP: Towards Adapting CLIP for Zero-shot Semantic Segmentation
      2. Align and Attend: Multimodal Summarization With Dual Contrastive Losses
   2. 类间对比损失
      1. Align and Attend: Multimodal Summarization With Dual Contrastive Losses
   3. 类内对比损失
      1. Align and Attend: Multimodal Summarization With Dual Contrastive Losses
6. 输入阶段
   1. 注意力掩膜机制
      1. Align and Attend: Multimodal Summarization With Dual Contrastive Losses