CLIP : Mask CLIP

注意力: CLIP Surgery: V

NACLIP : k-k 相似度

clearCLIP: 最后一层移除 MLP 和两个残差连接 Q-Q

GEM: Q-Q, k-k, v-v 自注意力 全部层

SCLIP: 最后一个编码器关联自注意力 权重 Q-Q k-k 相加

Tag CLIP: ViT 初步分数 注意力掩码抑制 patch 交互 } 中间分割图生成掩码 → 再给 CLIP. → 两分数结合

中间层:

ITACLIP: 最后一层 (sCLIP 权重) 与中间层结合, 几何和像素变换 去最后一层 FNN

SC-CLIP: CLIP 倒数第2层, 有异常因子, 对它插值, 使用中间偏后注意力图

ResCLIP: 最终注意力 = 中间注意力权重 + 自注意力(最后一层) 共同
(最后一层 ← 第一次粗画)

外部掩码:

CaR: CLIP 修剪 Grad-CAM 掩码 掩码 → 迭代 再输入 CLIP

CLIPTrase: 最后一层 聚合 Q-Q, k-k, v-v 权重 DBSCAN → 掩码 迭代掩码

LaVG: ~~Kar~~ k-k 使用 Normalized Cut 使用 SCLIP 获得标签

其它: ReCO: 巨大参考库 patch 特征, 把图片分割获得 patch 特征相似度.

DIY-CLIP: 分割图片 1-1 2-2 3-3, 得到不同块类别上半评. FOUND 获得前一背景

结合 VFM-S

混合：Proxy CLIP：使用VFM的embeding计算自一自相似度 代替原CLIP最后一层权重

CASS：混合CLIP和DINO模型最后一层注意力权重
尝试匹配CLIP与DINO的个体注意力头

外部Mask：TAG：k-means对DINO特殊征聚类 对GEM对掩码进化 获得类

DBA-CLIP：SAM掩码 和 Mask CLIP的 得到类

Trident：图像分割 CLIP逐块embedding与VFM的注意力权重混合 拼接整个图像特征
SAM最后一层注意力权重

CorrCLIP：SAM掩码相似embedding掩码合并 DINO 自自注意力 应用于CLIP的
patch embedding 进行最终注意力操作

利用生成方法：

OVDiff：SD生成平均，由patch文本找到特征，把patch与特征比对来分割

DiffSegmentor：从SD得交叉注意力图（是什么）和自注意力图（分割）

EMERDIFF：利用SD获得语义相关latent embedding —k-means→ SD生成图片与原图比→mask
再使用MaskCLIP来语义

Mask Diffusion：SD的所有U-Net层输出，类别特定交叉注意力图进化得到原型，比较分割

Free Seg-Diff：SD内部特征图聚类→粗mask，分别给CLIP得到类

RIM：SD生成　SAM与DINO提取　前景保证：用U-Net交叉注意力给SAM
SAM→mask —DINO→ 得征，与内部数据库再对比

FOSSIL：SD生图和交叉注意力→视觉原型　Normalized Cut切前景
使用文本模型→标准视觉原型来切割。

FreeDA：SD和交叉注意力图对DINO区域进化→视觉原型
推理与CLS token相似性检索，分割CLIP的patch与类比较，DINO特征与原型比较

CLIPer：CLIP中间层注意力图平均替换最后一层权重，删最后一层MLP和残差连接
提取中间层embedding输入最后一层。