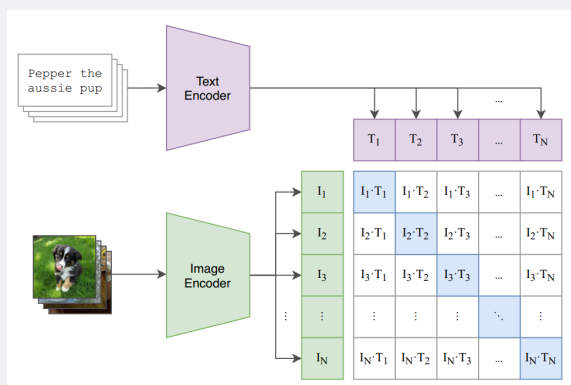


CLIP

新范式：数据不用标注类别，给出文字描述即可

400M数据集

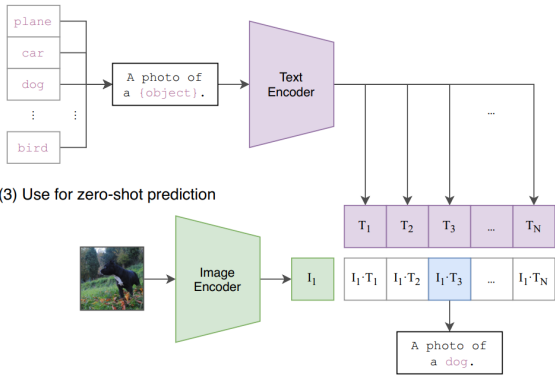


学习方式：对比学习

temperature可学习，不好调参

学习过程见论文伪代码

(2) Create dataset classifier from label text



分类方式：zero-shot

每一个类别组成句子送入text encoder

prompt engineering: 怎么构造句子，避免歧义等

prompt ensemble：同一图片设计多个句子分类然后ensemble

zero-shot

本质是提取vision & language两个模态的“共同信息”，比如其所指是同一个实体