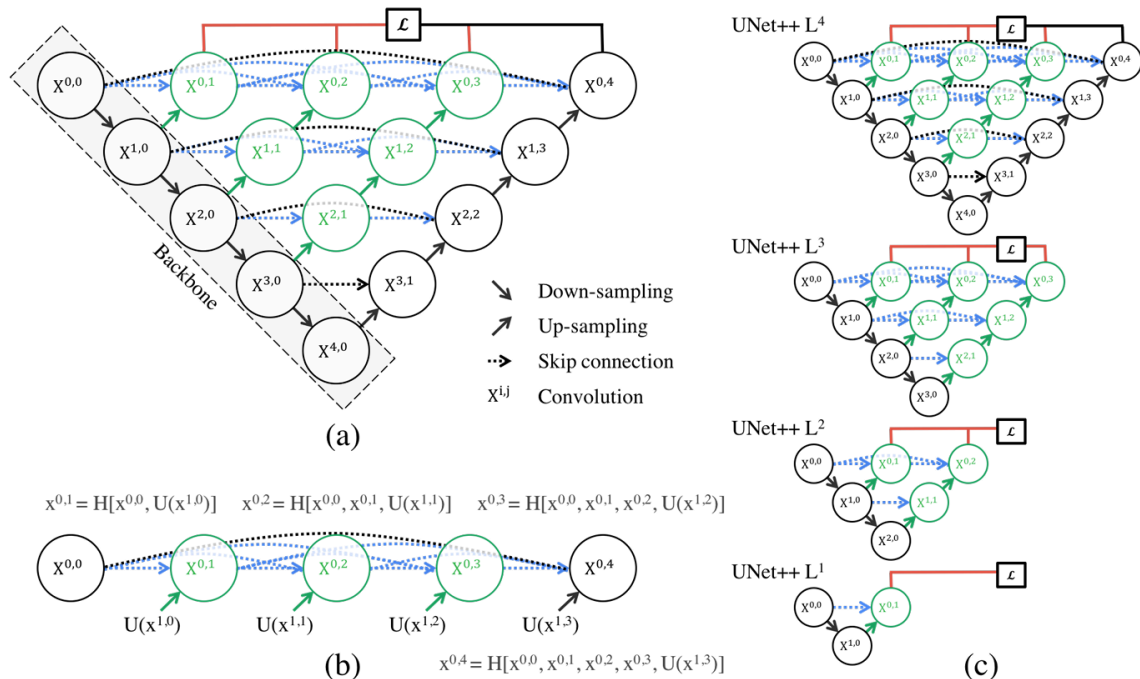


UNet++和MNIST-CLIP

UNet和UNet++框架概述

UNet++结构图（其实就是UNet+残差网络，非常简单的结构）



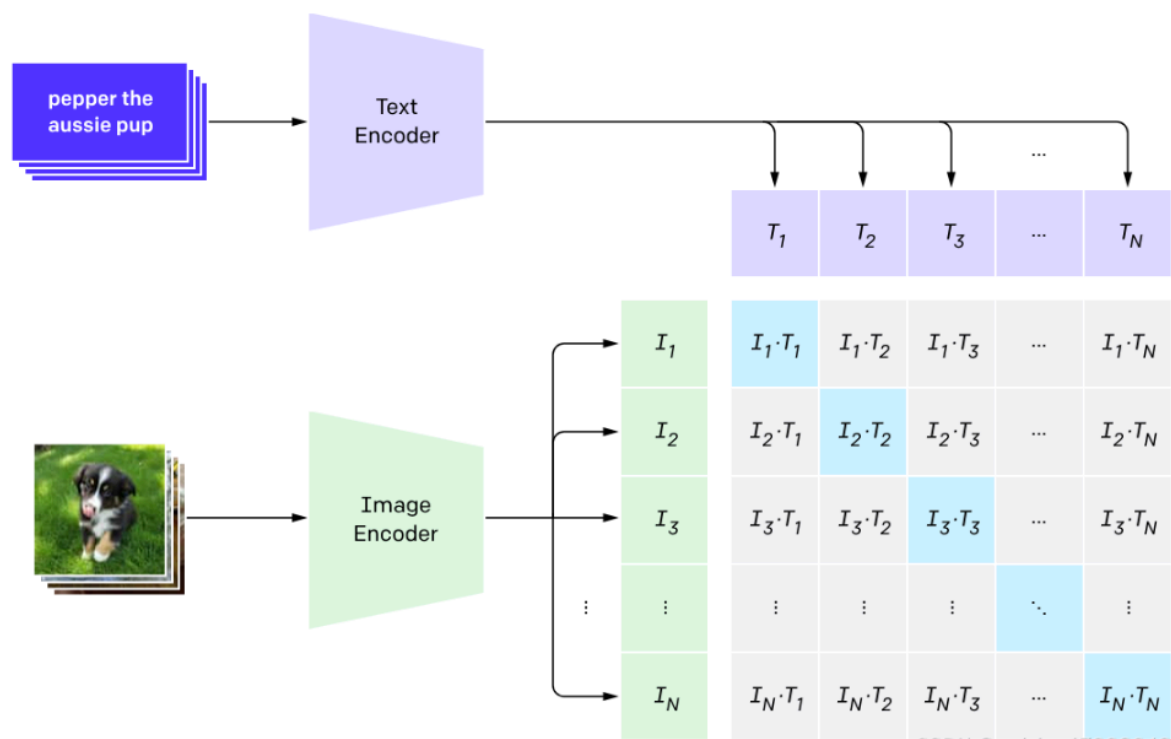
关于UNet网络的具体讲解可以参考这篇文章：[Unet网络架构讲解（从零到一，逐行编写并重点讲解数据维度变化）-CSDN博客](#)

UNet++代码配置

我这里做了一些小改动，因为他这个代码比较老旧，存在一些旧的函数已经无了的情况，我这里改了之后应该正常按照他的安装教程来就行，代码就用我发的压缩包来就行，配置教程依旧可以按照原Github项目的readme来

项目Github：[4uiiurz1/pytorch-nested-unet: PyTorch implementation of UNet++ \(Nested U-Net\). \(github.com\)](#)

CLIP框架概述



CLIP模型结构解析可以参考这篇文章：[【小白】一文读懂CLIP图文多模态模型 clip模型-CSDN博客](#)

主要思想其实是对比学习，让文本的特征向量和图像的特征向量对应起来，使得匹配的文本-图像对的特征向量尽可能接近，不匹配的尽量远离。

大家感兴趣的话可以直接搜索CLIP的论文，看看原文是怎么讲的，微软也公布了源码，大家也可以跑一下源码，但是想要训练一个性能比较好的CLIP，在一般的电脑是没法训练的，因为对显存的要求很高

MNIST-CLIP

个人手写简化版，因为原来的CLIP训练起来要求比较高，所以这里用手写数字数据集MNIST写了一个简化版，就是把上面的Image Encoder简化了一下（用了比较简单的CNN，原来的CLIP里面是用了vision-transformer）。然后Text Encoder这里就是用了最简单的编码（因为只有数字，没有文本，都不需要获得词向量，原来的CLIP用了Bert模型）。

具体的操作教程可以看项目里面的readme。