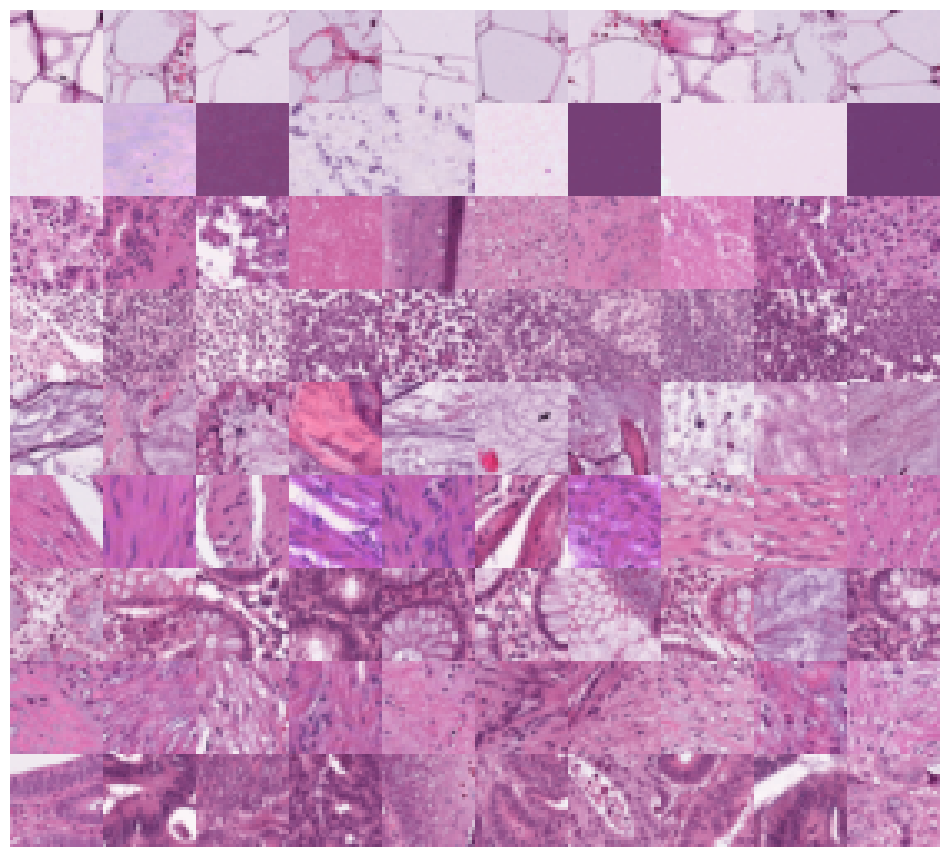
# Introduction:

# Data exploration:

原始样本包含training 89996,validation 10004, test 7180, 由学校提供的版本仅包含training 32000条和test 8000条。

所有的样本被分为9种类别，class0-class8.

**Data Modality:** Colon Pathology，

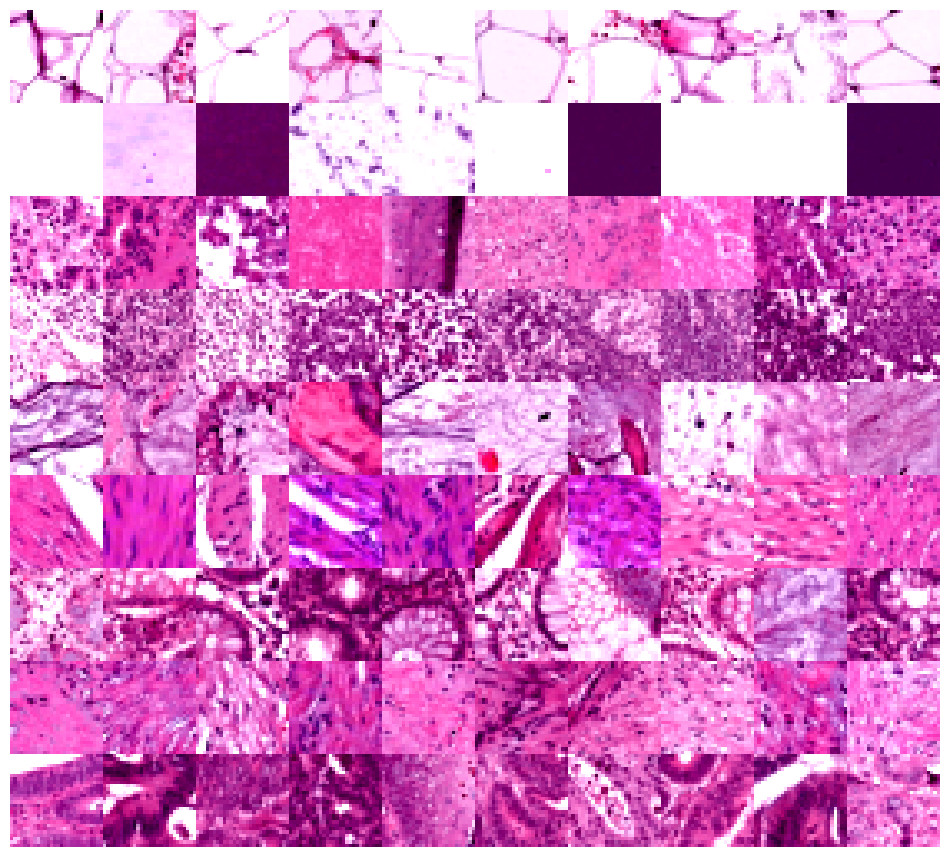


图为试输出的数据图片，从上到下分别为Label0-label8, label0 为adipose, label1为background,label2 为debris, label3 为lymphocytes， label4为mucus, label 5为 smooth muscle， label6为normal colon mucosa，label7 为 cancer-associated stroma，label8为 colorectal adenocarcinoma epithelium。

通过观察输出图片了解到，background类别内存在显著的颜色差异，可能造成捕捉特征的困难; smooth muscle 类别有些样本的颜色特别的鲜艳; normal colon mucosa、cancer-associated stroma和colorectal adenocarcinoma epithelium类别之间，有些样本无法通过人眼的观察分辨出其区别，这些样本的特征并不明显，可能对模型的训练造成阻碍。

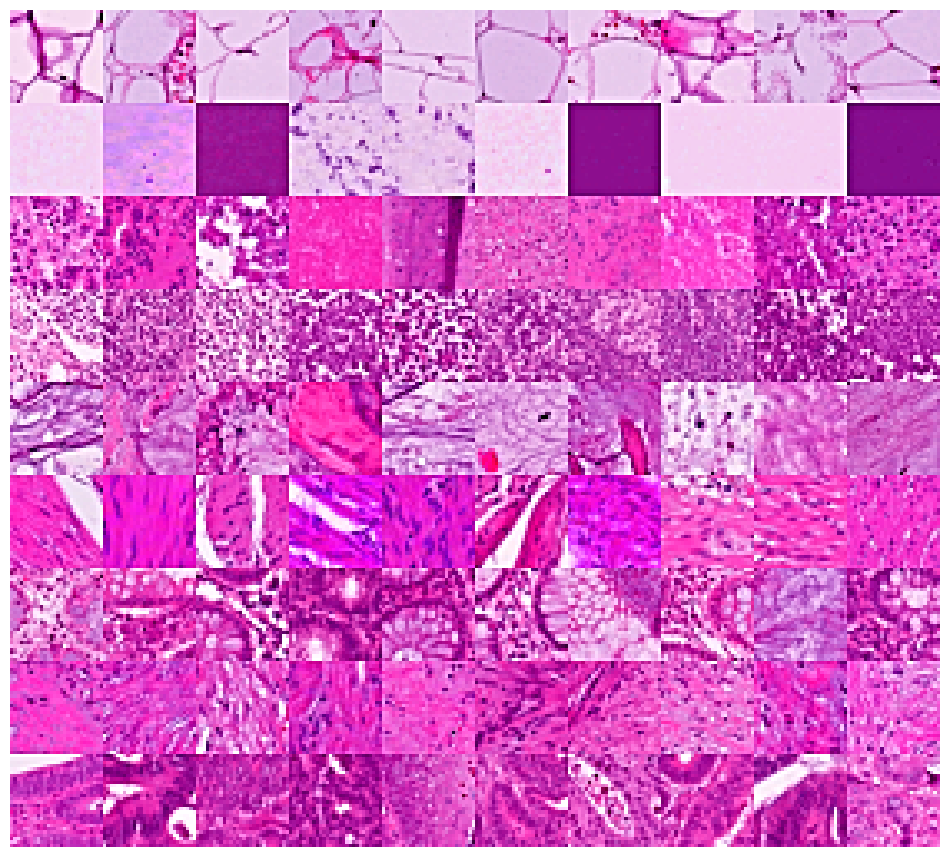
# Preprocessing

1. 首先考虑到样本图片的主要色调差异不大，考虑通过强化对比度来强化其特征。

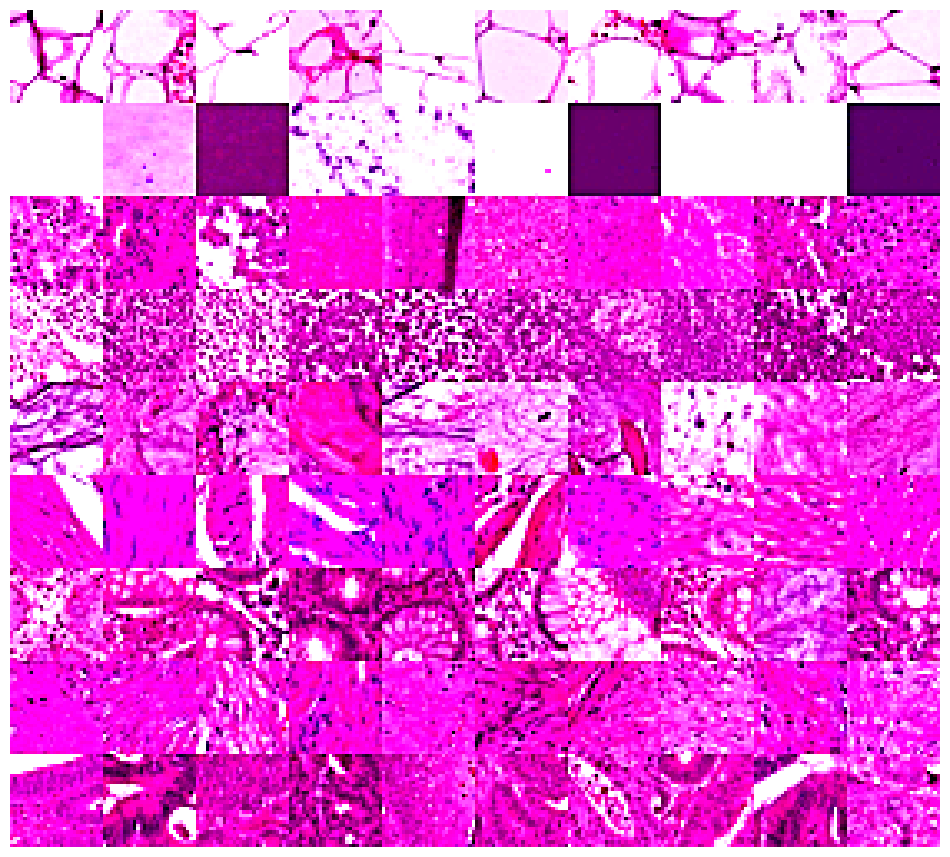


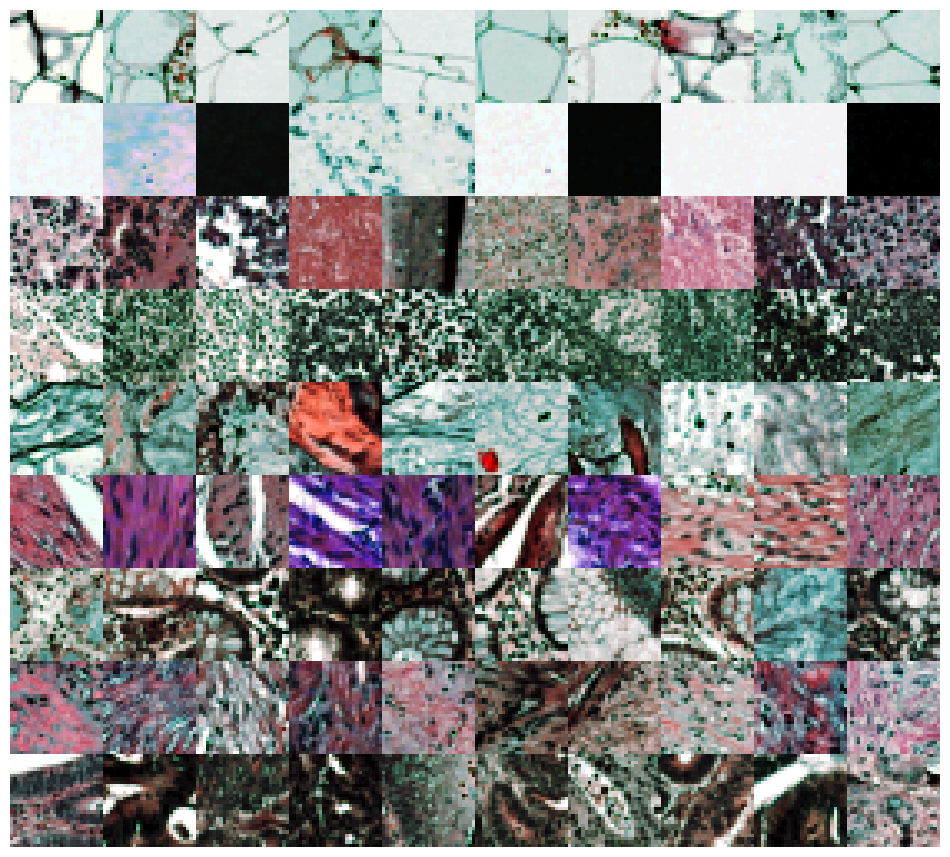
可以看到，在放大对比度之后，样本显示出了更加显著的特征。

下面这张是锐化的样本图，锐化之后纹理的分界变得更加的清晰，可以帮助模型捕捉特征。



下面这张是先锐化在增强了对比度的组织图，



下图为自适应对比度+局部拉伸

同时也尝试了其他的方法来降噪，但是结果非常的模糊，可能是因为样本图片很小，降噪模糊掉了细节，这份输出结果也会尝试用于训练模型。

