"**Document Layout Analysis** is performed to determine physical structure of a document, that is, to determine document components. These document components can consist of single connected components-regions of pixels that are adjacent to form single regions, or group of text lines. A text line is a group of characters, symbols, and words that are adjacent, “relatively close” to each other and through which a straight line can be drawn (usually with horizontal or vertical orientation)."

Dataset reprocess：

数据预处理阶段，我过滤掉原图中的背景部分，并将Mask中的黄色（文本框）部分去掉。之后我将使用高斯核或者其他方法对原图片进行去噪。由于图片过大（960x640），我将对其进行resize，使其变为原来的二分之一（480x320）。

Dataset and Dataloader：

我按照8：1：1的比例将数据集划分为训练集，验证集和测试集。其中，测试集用于应用matric评价指标来衡量模型是否表现良好。

对于JPG格式的文档图片，我将其读取为三通道的图片数据。需要注意的是，对于GIF格式的Mask图片，不能将其颜色作为标签（如红色的标签是[1, 0, 0]），而是应该按照一共有多少种颜色将其进行编码。这里既可以采用仅仅是一个单通道的图像，其中每个像素的整数值直接表示类别标签。例如，0代表背景，1代表第一个类别，2代表第二个类别等；也可以采用One-hot编码的Mask：在这种情况下，每个像素的标签会被编码成一个向量，向量的长度等于可能的标签数。比如，如果有三个类别（背景，类别1，类别2），那么每个像素的标签将会是一个三维的向量（我们的任务中，除开黄色的文本框，一共有六个类别）。

Model:

这里我目前使用的是基础的UNet。我在实验中注意到了在其下采样过程中，特征图出现的“漏墨”问题，并利用残差连接缓解了这一问题。

对于不同的模型，如果有时间的话，我会做消融实验。但模型的性能好坏似乎并不是这次任务的主要目的，所以这里我不再展开细说。

Training process：

训练过程中，我同时使用像素级别的交叉熵损失，以及分割图像边缘级别的boundary loss，当然，具体情况还需要根据训练结果而定。

我在训练集上累计梯度，更新模型，并在验证集上判断结果的好坏，根据结果判断是否要保存模型，进而得到验证集上表现最好的模型。

Matrics：

这是这次任务最容易让人迷糊的部分：我们该怎样知道我的模型表现良好？也就是说，图像分割任务上最常见的指标是什么？

Pixel Accuracy 指的是图像中被正确分类的像素的百分比。虽然它看起来很不错，毕竟我们也是用它作为优化目标来进行训练的，但它并不是一个好的衡量标准。就比如说，模型给出的分割图像大体上和Mask非常相似，计算得到Pixel Accuracy高达95%以上。但实际上有一些需要被分割出来的关键词或者小插画根本就没有被分割出来——这种现象被称为class imbalance，说明高像素精度并不总是意味着卓越的分割能力。这就要求我们去寻找其他衡量指标。

Intersection-Over-Union (IoU)，是语义分割中最常用的指标之一，计算方式是预测分割和真实情况之间的重叠面积除以预测分割和真实情况之间的并集面积。对于二值（两类）或多类分割，图像的平均 IoU 是通过对每个类的 IoU 取平均值来计算的（如我们这个任务）。这将会给予那些面积不是很大但需要被分割的图像块更多的关注。

Dice Coefficient (F1 Score)指的是2 \* 重叠面积除以两个图像中的像素总数。Dice 系数与 IoU 非常相似。它们是正相关的，范围都是从 0 到 1，其中 1 表示预测与事实之间的相似度最大。但它们并不是一回事：一般来说，IoU 指标在数量上倾向于比 F 得分更多地惩罚不良分类的单个实例，即使它们都同意这个实例是不好的。与 L2 比 L1 更能惩罚最大错误类似，IoU 指标往往会对相对于 F 分数的错误产生“平方”效应。因此，F 分数往往衡量更接近平均性能的东西，而 IoU 分数衡量更接近最坏情况性能的东西。

IoU和F1 Score都有一个问题：它们过分强调小错误的重要性，尤其是在真实正样本很少的情况下。在平均多个样本的评分时，这会使得含有少量正样本的图像的错误过度影响整体评价结果。为了避免这种问题，我们可以使用一个更简单的度量标准，即总错误数（FN + FP，即假阴性和假阳性的总和）。这种方法简单地统计分类错误的像素数量，而不是计算比例或相对大小。如果在具体应用中某类错误更为严重，可以通过赋予假阳性（FP）和假阴性（FN）不同的权重（c0和c1）来调整这种度量方式。

所以，我将使用IoU、F1 Score和FN + FP来作为总的评价指标，衡量模型的好坏。

Output：

Model的输出取决于一开始怎么处理Mask。我的想法是：如果Mask是单通道的图像，那么我的输出就是使用softmax处理之后生成的单通道的Mask；如果每个像素的标签是一个六维的向量，那么模型的输出就是经过Sigmoid处理得到的六通道的二分类向量组成的Mask：[batchsize, 6, 480, 320]。总之，Mask的标签不能是颜色而是离散的label，而且模型的输出要与Mask对应，方便损失函数的计算。