

RARE(Robust text recognizer with Automatic Rectification)模型介绍

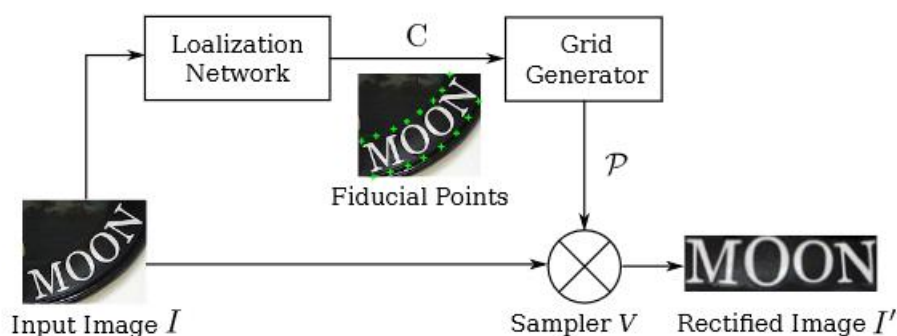
(华中科大)

Spatial Transformer Network (STN):

在 STN 内部通过 TPS 变换, 将输入的不正常文本转换成适合作为后序 SRN 输入的正常文本。而 TPS 变换是由一组基准点配置的, 这些基准点的坐标由 CNN 回归得到。

Thin-Plate-Spline (TPS) transformation

具体来说, localization network(定位网络)预测一组基准点 C , 接着在 grid generator(网格生成器)里, 从基准点 C 计算 TPS 参数, 在输入图像 I 上生成采样网格 P , 最后 sampler(采样器)用输入图像 I 和网格 P 生成校正过的图像 I' 。



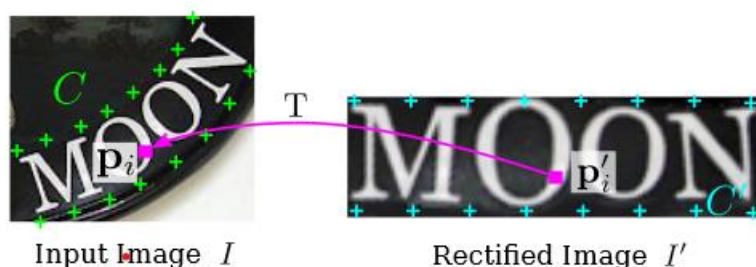
Localization Network: (CNN)

基准点数量 K 是偶数, 用归一化坐标系来定位每个基准点 c_i , x 、 y 的值在 $[-1,1]$ 之间。使用 CNN 网络, 包括卷积层、池化层和全连接层解决 (x, y) 的回归问题。

Output layer 是全连接层, 输出维度为 $2K$, 激活函数为 \tanh 。最后 output vector 被 reshape 为 c 的形式。

Grid Generator:

首先定义基本基准点 C' , 由于 K 是常数且为归一化坐标系, 则 C' 为常数。



接着通过 TPS 变换 T , 找到 p'_i 对应在原图 I 中的 p_i , 矩阵 T 中存储 TPS 参数。

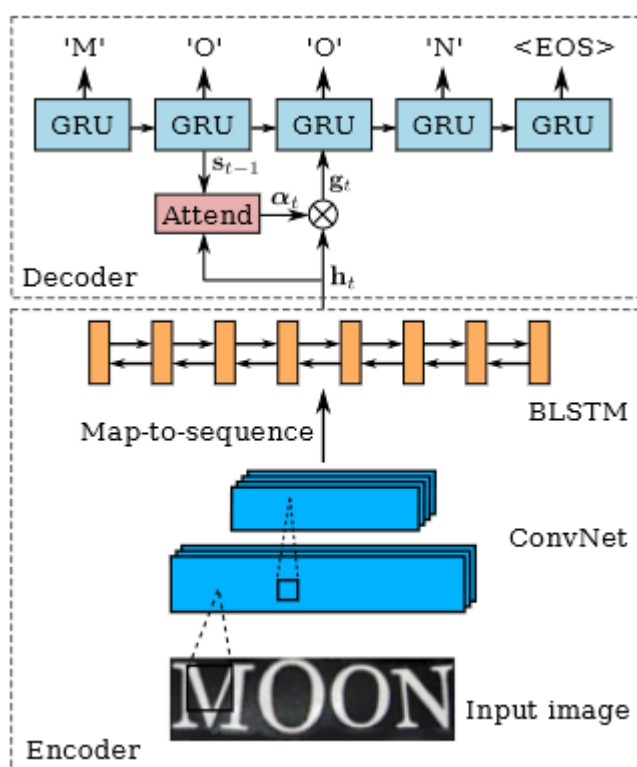
$$\mathbf{T} = \left(\Delta_{\mathbf{C}'}^{-1} \begin{bmatrix} \mathbf{C}^\top \\ \mathbf{0}_{3 \times 2} \end{bmatrix} \right)^\top$$

通过矩阵乘法，将网格像素点 p_i' 转换到原图 I 上的 p_i ，由于是矩阵乘法，自然而然可以传播梯度。

Sampler:

通过对原始图像 I 中 p_i 周围的像素颜色双线性插值，得到 p_i 的像素颜色。双线性插值同样可微。

Sequence Recognition Network (SRN): attention-based model



Encoder: (CRNN)

对输入图像生成一个序列特征表示（特征向量序列）(CNN+BLSTM)。

在底部的几个卷积层之后产生 shape 为 (D,H,W) 的 feature maps，将 feature maps 降维处理，得到 W 个 $(DH,)$ 的向量（或理解为 shape 为 (W,DH) 的 feature maps）。接着应用两层的 BLSTM，得到长度为 L (W) 的输出序列。

$$\mathbf{h} = (\mathbf{h}_1, \dots, \mathbf{h}_L),$$

Decoder: (GRU + attention)

每一步解码由 attention 机制决定, 循环生成一个基于输入序列的字符序列。Decode 过程有 T 步, 在第 t 步 decoder 产生一个 attention 权重向量, 长度为 L , 方法如下:

$$\alpha_t = \text{Attend}(s_{t-1}, \alpha_{t-1}, \mathbf{h}),$$

其中 s_{t-1} 是 $t-1$ 步中 GRU 单元的状态, 特别地, $t=1$ 时 s_0 和 α_0 都是 0 向量, 接着 glimpse g_t 通过线性方法计算:

$$\mathbf{g}_t = \sum_{i=1}^L \alpha_{ti} \mathbf{h}_i.$$

α 中的所有值之和为 1。

接着更新 s_t :

$$\mathbf{s}_t = \text{GRU}(l_{t-1}, \mathbf{g}_t, \mathbf{s}_{t-1}),$$

训练中, l_{t-1} 是 $t-1$ 个真实标签, 测试中, 是 $t-1$ 步中预测的标签。

接着计算概率分布:

$$\hat{\mathbf{y}}_t = \text{softmax}(\mathbf{W}^T \mathbf{s}_t).$$

通过取最大值得到 t 步的预测标签 l_t 。

###STN 和 SRN 的训练是一体的, 即 STN 没有对基准点的 label, STN 接收从 SRN backprop 回来的梯度, 对基准点的位置进行调整。

