

## 论文题目

ESIR: End-to-end Scene Text Recognition via Iterative Image Rectification

## 论文作者

Fangneng Zhan, Shijian Lu  
Nanyang Technological University

## 论文时间

2018 年 12 月

## 论文的前提

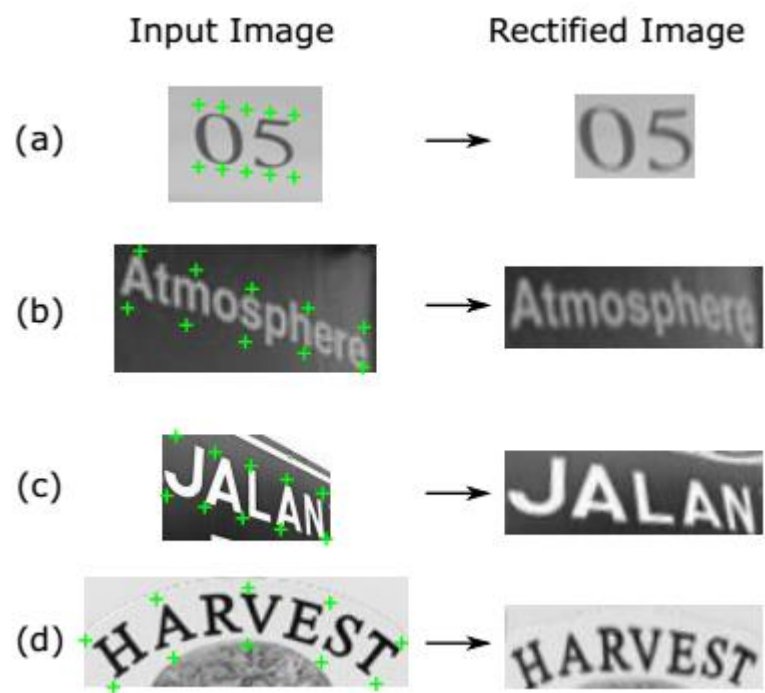
用于解决不规则排列文字的文字识别

是基于 ASTER, ESIR 有些提高

都由 Rectification network(矫正网络)和 recognition network(识别网络)组成

将不规则排列的文字矫正成正常排列的文字在进行识别(end-to-end)

TPS(Thin-Plate-Spline):对形变图像进行矫正,通过对 control points 进行定位和映射来得到矫正过后的图片





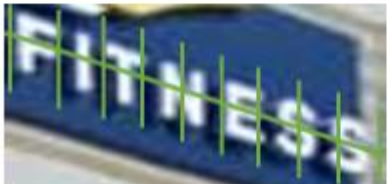



## 论文观点

本文改进地方主要在矫正网络部分,改变了拟合形变的方式,并增加了迭代的矫正流程

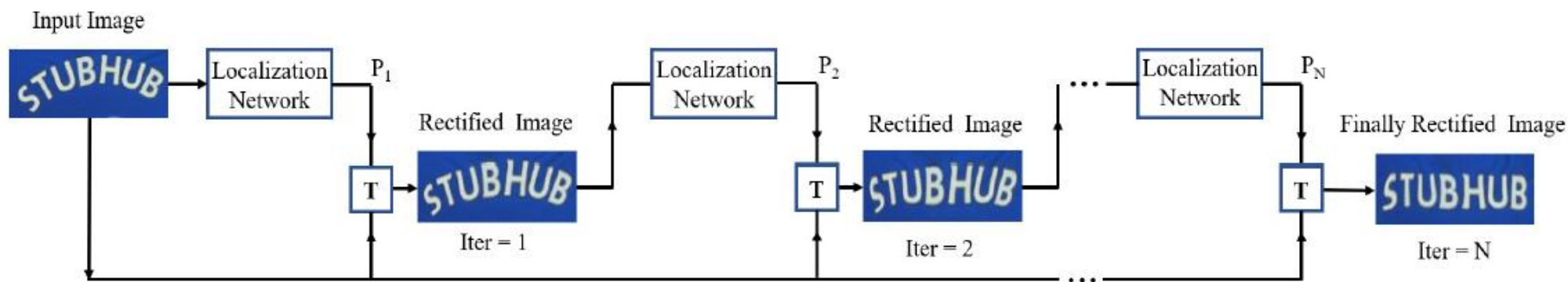
## 论文做法

使用直线和线段帮助矫正形变,与 ASTER 使用 K 个控制点辅助矫正不同,ESIR 使用 1 条水平中线(拟合文字走向)和 L 条线段(估计竖直方向和边界)来辅助矫正

Input Images		Rectified Images
	⇒	
	⇒	
	⇒	

$$y = a_K * x^K + a_{K-1} * x^{K-1} + \dots + a_1 * x + a_0 \quad y = b_{1,l} * x + b_{0,l} \mid r_l, \quad l = 1, 2, \dots, L$$

iterative rectification framework.多次(迭代)矫正,图片变形时,ASTER 学习到 TPS 变换后仅做一次变换,本算法循环 TPS 变换 N 次,ASTER 只能生成一张矫正图,ESIR 则在这点做了改进,通过迭代可以生成多张矫正图



recognition network: 预测网络本文选的是 ResNet+BiLSTM+Attention

## 最终效果

Methods	ICDAR2013	ICDAR2015	IIIT5K			SVT		SVTP	CUTE
	None	None	50	1k	None	50	None	None	None
Wang [44] [-]	-	-	-	-	-	70.0	-	-	-
Bissacco [4] [-]	87.6	-	-	-	-	-	-	-	-
Yao [47] [-]	-	-	80.2	69.3	-	75.9	-	-	-
Almazan [1] [-]	-	-	91.2	82.1	-	89.2	-	-	-
Gordo [10] [-]	-	-	93.3	86.6	-	91.8	-	-	-
Jaderberg [16] [VGG, SK]	81.8	-	95.5	89.6	-	93.2	71.7	-	-
Jaderberg [17] [VGG, SK]	90.8	-	97.1	92.7	-	95.4	80.7	-	-
Shi [37] [VGG, SK]	88.6	-	96.2	93.8	81.9	95.5	81.9	71.8	59.2
Yang [46] [VGG, Private]	-	-	97.8	96.1	-	95.2	-	75.8	69.3
Cheng [7] [ResNet, SK+ST]	<b>93.3</b>	70.6	99.3	97.5	87.4	97.1	85.9	71.5	63.9
Cheng [8] [VGG, SK+ST]	-	68.2	99.6	98.1	87.0	96.0	82.8	73.0	76.8
Shi [38] [ResNet, SK+ST]	91.8	76.1	<b>99.6</b>	<b>98.8</b>	<b>93.4</b>	97.4	89.5	78.5	79.5
ESIR [VGG, SK]	87.4	68.4	95.8	92.9	81.3	96.7	84.5	73.8	68.4
ESIR [ResNet, SK]	89.1	70.1	97.8	96.1	82.9	97.1	85.9	75.8	72.1
ESIR [ResNet, SK+ST]	91.3	<b>76.9</b>	<b>99.6</b>	<b>98.8</b>	93.3	<b>97.4</b>	<b>90.2</b>	<b>79.6</b>	<b>83.3</b>

## 缺点

如果不加控制的多次直接迭代,则会造成 boundary effect 问题,即每次迭代都会使一部分像素点在采样区域外面,这样就会忽视掉一些 text 的像素点

矫正网络对参数初始化十分敏感,ASTER 中也提到完全随机的初始化参数会导致收敛问题,其产生的高度扭曲的图片会影响识别网络的效果,进而影响矫正网络