|  |
| --- |
| 论文名称： |
| ContourNet：Taking a Further Step toward Accurate Arbitrary-shaped Scene Text Detection |
| 论文信息： |
| Yuxin Wang, Hongtao Xie∗ , Zhengjun Zha, Mengting Xing, Zilong Fu and Yongdong Zhang  中科大 2020 |
| 摘要： |
| 场景文本检测近年来得到了迅速的发展。然而，仍存在两个主要的挑战：1)许多方法的文本表示形式都存在假阳性；2)场景文本的大规模方差使得网络难以学习样本。在本文中，我们提出了ContourNet，它有效地处理了这两个问题，为精确的任意形状的文本检测迈出了进一步的一步。  首先，提出了一种不敏感尺度的自适应区域提案网络(自适应RPN)，通过只关注预测边界框和GT边界框之间的交集(IoU)值来生成文本框。然后，一种新的局部正交纹理感知模块(LOTM)在两个正交方向上对提案特征的局部纹理信息进行建模，并用一组轮廓点表示文本区域。考虑到强单向或弱正交激活通常是由假阳性的单调纹理特征引起的。我们的方法只通过输出两种高响应值的预测来有效地抑制这些假阳性。这就提供了对文本区域的更准确的描述。 |
| 解决了什么问题： |
| 1. 在文字检测方法中，最终的检测结果中很容易出现误检，抑制误检会提升算法性能 2. 文本尺度变化较为大，使得模型在学习过程中不利于很好的学习 |
| 为什么会有这些问题： |
| 1. 文章认为，以往的直接分割边界的方法，由于FP的存在容易导致预测结果出现错误 2. 图像中文字大小不一致,回归的话存在困难 |
| 怎么解决： |
| 1. 文章将轮廓的分割分为水平方向以及竖直方向，对两张图进行re-scoring达到抑制FP的作用，利用两个方向的heatmap的融合得到较为精准的轮廓 2. 作者认为类似L2回归的方式对尺度变化比较敏感：不同尺度大小的proposals与真值的IOU在相同的情况下，L2或者smoothl1损失有可能不同。对于文字检测场景，smoothl1等损失函数不具备优势。选取IOULoss作为损失函数，重点优化IOU |
| 可借鉴的地方： |
| 1.Adaptive-RPN, 文章将anchor拆分至9个点或者5个点，预测与真值对应位置点的偏移 |
| 优点： |
| 1. 利用两个方向的 文本轮廓分割图，水平和垂直方向都参与投票，抑制FP 2. Adaptive-RPN, 文章将anchor拆分至9个点或者5个点 |
| 缺点： |
| 1. 速度好像很慢 |

|  |
| --- |
| 论文名称： |
| DBNet/ **Real-Time Scene Text Detection with Differentiable Binarization** |
| 论文信息： |
| **Minghui Liao,** **Zhaoyi Wan,** **Cong Yao,** **Kai Chen,** **白翔**  华中科大，旷视，上海交大 2020 |
| 摘要： |
| 近年来，基于分割的方法在场景文本检测中非常流行，因为分割结果可以更准确地描述各种形状的场景文本。然而，二值化的后处理对于基于分割的检测至关重要，它将分割方法生成的概率映射转换为文本的边界框/区域。在本文中，我们提出了一个名为可微分二值化(DB)的模块，它可以在分割网络中执行二值化过程。分割网络与DB模块一起优化，可以自适应地设置二值化的阈值，不仅简化了后处理，还提高了文本检测的性能。基于一个简单的分割网络，我们验证了DB在五个基准数据集上的性能改进，这在检测精度和速度方面都始终取得了最先进的结果。特别是，对于一个轻量级的骨干网，DB的性能改进是显著的，因此我们可以在检测精度和效率之间寻找一个理想的权衡。具体来说，通过ResNet-18的主干，我们的检测器在MSRA-TD500数据集上实现了82.8的F-measure，在62FPS的速度运行。 |
| 解决了什么问题： |
| 自适应地设置二值化的阈值，简化了后处理 |
| 为什么会有这些问题： |
| 基于分割的方法，二值化的后处理将分割方法生成的概率映射转换为文本的边界框/区域 |
| 怎么解决： |
| 提出 可微分二值化(DB)的模块，自适应地设置二值化的阈值，简化了后处理 |
| 可借鉴的地方： |
| 1. 可微分二值化(DB)的模块，自适应地设置二值化的阈值 |
| 优点： |
| 1. DB模块自适应地设置二值化的阈值，性能好，且速度快 |
| 缺点： |
| （有待深度理解，解读） |

|  |
| --- |
| 论文名称： |
| STKM/ Self-attention based Text Knowledge Mining for Text Detection |
| 论文信息： |
| Qi Wan , Haoqin Ji , 沈琳琳 2021  深圳大学  深圳市人工智能与社会机器人研究所  广东智能信息处理重点实验室 |
| 摘要： |
| 预训练模型在基于深度学习的文本检测器中发挥着重要作用。然而大多数方法忽略了自然图像和场景文本图像之间的差距，直接使用ImageNet预训练模型，另外一些方法首先使用大量的合成数据对模型进行预训练，然后在目标数据集上进行微调，这种方法是针对特定任务的，而且泛化能力有限。在本文中，作者们希望为文本检测提供通用的预训练模型。考虑到识别文本内容对文本检测的重要性，我们提出了STKM（基于自我注意的文本知识挖掘），它由一个CNN编码器和一个自我注意解码器组成，以从SynthText学习用于文本检测的一般先验知识。在只给定图像级别的文本标签的情况下，自我注意解码器直接将从CNN编码器中提取的特征解码为文本，而不需要检测，这引导CNN骨干明确地学习以前的方法所忽略的辨别性语义表征。之后，backbone学到的文本知识可以转移到各种文本检测器上，以显著提高它们的检测性能 |
| 解决了什么问题： |
| 预训练的模型在基于深度学习的文本检测器中发挥着重要作用。大多数方法忽略了自然图像和场景文本图像之间的差距，并直接应用ImageNet进行预训练。 |
| 为什么会有这些问题： |
| 自然图像和场景文本图像之间有差距 |
| 怎么解决： |
| 为文本检测器提供一般的预训练模型。考虑到探索文本内容对文本检测的重要性，提出了STKM（基于自注意力机制的文本知识挖掘），它由CNN编码器和自注意力解码器组成，以从合成文本中学习文本检测的一般先验知识。  只给定图像级的文本标签，自我注意力解码器直接将从CNN编码器中提取的特征解码为文本，而不需要检测，这引导CNN主干显式地学习以前方法忽略的判别性语义表示。 |
| 可借鉴的地方： |
| 预训练模型，文本识别的研究 |
| 优点： |
|  |
| 缺点： |
|  |

|  |
| --- |
| 论文名称： |
| PCRNet/ Progressive Contour Regression for Arbitrary-Shape Scene Text Detection |
| 论文信息： |
| Pengwen Dai, Sanyi Zhang, Hua Zhang, Xiaochun Cao  中科院；国科大；鹏城实验室；天津大学 2021 |
| 摘要： |
| 现有的场景文本检测方法通常从自底向上的角度对文本实例进行局部像素或局部分量的建模，因此对噪声非常敏感，并且依赖于复杂的启发式后处理，特别是对于任意形状的文本。为了解决这两个问题，我们建议以自上而下的方式将最初的文本方案逐步演化为任意形状的文本轮廓。初始水平文本方案是通过估计文本的中心和大小来生成的。为了缩小回归的范围，进化的第一阶段从最初的水平方向预测 定向文本框的角点。在第二阶段，有向文本提议框的轮廓被迭代回归为任意形状的轮廓。在这一阶段的最后一次迭代中，我们利用来自多个轮廓点的线索，而不是来自初始水平建议中心的单个线索（可能在任意形状的文本区域之外），重新评估最终局部文本的可信度。  此外，为了便于渐进的轮廓演化，我们设计了一个轮廓信息聚合机制来丰富文本轮廓的特征表示，同时考虑了循环拓扑和语义上下文。在CTW1500、Total Text、ArT和TD500上进行的实验表明，该方法特别适用于行级任意形状的文本。 |
| 解决了什么问题： |
| 目前挺多方法都是自下而上的建模方式，这种方式对噪声特别敏感，并且需要复杂的后处理，  提出采用自上而下的方式 逐步将初始文本框 演变为 任意形状的文本轮廓的网络模型 |
| 为什么会有这些问题： |
| 自下而上的建模方式具备的缺点 |
| 怎么解决： |
| 采用自上而下的方式 逐步将初始文本框 演变为 任意形状的文本轮廓 |
| 可借鉴的地方： |
| 逐步将初始文本框 演变为 任意形状的文本轮廓 |
| 优点： |
|  |
| 缺点： |
|  |

|  |
| --- |
| 论文名称： |
| FCENet/ Fourier Contour Embedding for Arbitrary-Shaped Text Detection |
| 论文信息： |
|  |
| 摘要： |
| 任意形状的文本检测的主要挑战之一是设计一个良好的文本实例表示，允许网络学习不同的文本几何方差。大多数现有的方法通过笛卡尔坐标系或极坐标系中的掩模或轮廓点序列对图像空间域中的文本实例进行建模。然而，掩码表示可能会导致昂贵的后处理，而点序列对于建模具有高度弯曲形状的文本的能力可能很有限。为了解决这些问题，我们在傅里叶域中建模文本实例，并提出了一种新的傅里叶轮廓嵌入(FCE)方法，将任意形状的文本轮廓表示为紧凑的签名。我们进一步构造了具有主干、特征金字塔网络(FPN)和简单的后处理傅里叶逆变换(IFT)和非最大抑制(NMS)的FCENet。与之前的方法不同，FCENet首先预测文本实例的紧凑傅里叶特征，然后在测试期间通过IFT和NMS重构文本轮廓。大量的实验表明，即使是高曲线的形状，FCE在适应场景文本的轮廓方面也是准确的和鲁棒的，同时也验证了FCENet对任意形状文本检测的有效性和良好的泛化。此外，实验结果表明，我们的FCENet在CTW1500和Total-Text上优于最先进的(SOTA)方法，特别是在具有挑战性的高弯曲文本子集上。 |
| 解决了什么问题： |
| 更好地拟合 任意形状文本的 检测器 |
| 为什么会有这些问题： |
| 目前现有的任意形状文本检测器大多在图像的空间域中表示文本实例，基于空间域的表示方法大体上可以分为两种，即像素掩膜表示和轮廓点序列表示。其中，像素掩膜表示方法可能需要繁杂和耗时的后处理过程，同时对训练样本量的需求往往也会更大；而轮廓点序列表示方法对高度弯曲文本的表达能力有限。 |
| 怎么解决： |
| 傅里叶系数表示在理论上可以拟合任意的封闭曲线，并且文本轮廓更多集中在低频分量上，所以通过在傅立叶域对不规则场景文字实例进行表征能很好地解决上述问题，并且具有简单、紧凑、对复杂轮廓表达能力好的特点 |
| 可借鉴的地方： |
| 拟合任意的封闭曲线？？？？ |
| 优点： |
|  |
| 缺点： |
|  |

|  |
| --- |
| 论文名称： |
| **MSR: Multi-Scale Shape Regression for Scene Text Detection** |
| 论文信息： |
|  |
| 摘要： |
| 最先进的场景文本检测技术可以预测在处理场景中不同方向和长度的直线或曲线文本行时容易出现定位错误的四边形框。本文提出了一种新颖的多尺度形状回归网络 (MSR)，它能够定位场景中不同长度、形状和曲率的文本行。所提出的 MSR 通过预测密集文本边界点来检测场景文本，与使用提议或分割的现有技术相比，这些边界点本质上准确地捕获了文本行的位置和形状，并且对文本行长度的变化也更加宽容。此外，多尺度网络提取和融合不同尺度的特征，显示出对文本尺度变化的极好容忍度。 |
| 解决了什么问题： |
| 自然场景的文本检测存在如下两个问题：   (1)不精确的定位：缘于特殊的文本形状、细长的文字、变化的长度、不同的弯曲方式，文本检测的定位精度不够良好。为此，基于文本提议框的方法通过调整适当的横纵比锚框来处理长度变化的文本行;当长文本线中间附近的像素距四边形回归框顶点很远时基于分割方法可能会出现较大的回归错误。因此，曲折文本的检测框可能会包含大量的非文本像素;   (2)不可靠的检测：相对于传统的通用目标检测，场景文本通常存在大规模的尺寸波动，最大与最小的文本实例的横纵比可能相差230倍，因此最小的文本实例可能漏检测而最大的文本实例可能会部分检测或断裂; |
| 为什么会有这些问题： |
|  |
| 怎么解决： |
| 1. 回归文本中心区域，中心区域的点到距离最近边框点的XY距离 2. 提出了Multi-scale NetWork,既可以预测不同层级的细节，又可以提取图片中不同尺度的文本特征. |
| 可借鉴的地方： |
| 考虑到多尺度的问题，提出了一种适合于多尺度的网络结构 |
| 优点： |
|  |
| 缺点： |
|  |

|  |
| --- |
| 论文名称： |
| **Sequential Deformation for Accurate Scene Text Detection** |
| 论文信息： |
| ECCV2020 清华；现代汽车集团 |
| 摘要： |
|  |
| 解决了什么问题： |
| 由于文字尺寸、方向、形状和比例的多种多样，CNN对于这种几何变换能力有限；本文中，作者提出了一个序列变形方法对 场景文本的线条形状 进行建模。同时进一步引入了一种辅助字符计数监督来指导序列偏移预测，整个训练网络可以多任务端对端。 |
| 为什么会有这些问题： |
| 1. 标准卷积的固定感受野和几何变换的有限能力削弱了它们的性能和泛化能力。   包含许多不需要的背景信息，无法很好地提取出具有不同尺度和高宽比的实例的表示 |
| 怎么解决： |
| 1. 提出了一种端到端可训练的顺序变形模块（SDM），用于精确的场景文本检测，该方法通过对特征级构件进行分组，有效地扩展了文本几何配置的建模能力，并沿文本线学习信息化实例级语义表示。 |
| 可借鉴的地方： |
|  |
| 优点： |
|  |
| 缺点： |
|  |

|  |
| --- |
| 论文名称： |
| **pan** |
| 论文信息： |
| ICCV2019 |
| 摘要： |
| 近年来，有人提出了一些处理任意形状文本检测的方法，但很少考虑到整个管道的速度，这在实际应用中可能不足。  在本文中，我们提出了一种高效而准确的任意形状的文本检测器，称为像素聚合网络(PAN)，它具有低计算成本的分割头和可学习的后处理。  分割头由特征金字塔增强模块(FPEM)和特征融合模块(FFM)组成。  FFPEM是一个可级联的u型模块，它可以引入多层次的信息来指导更好的分割。FFM可以将不同深度的fpem给出的特征收集成最终的特征进行分割。  可学习的后处理是通过像素聚合(PA)实现的，它可以通过预测的相似性向量精确地聚合文本像素。 |
| 解决了什么问题： |
| 第一个问题是速度和准确性之间的权衡。  第二种方法是对任意形状的文本实例进行建模。  保证网络的高效率，我们就需要采取措施来降低这两步的耗时：采用的****更轻量级的backbone****和****更简单的后处理****。 |
| 为什么会有这些问题： |
| 1. 标准卷积的固定感受野和几何变换的有限能力削弱了它们的性能和泛化能力。   包含许多不需要的背景信息，无法很好地提取出具有不同尺度和高宽比的实例的表示 |
| 怎么解决： |
| ****本文使用将 ResNet-18 作为 PAN 的默认 Backbone，并提出了低计算量的Segmentation Head(FPFE + FFM)**** 以解决因为使用 ResNet-18 而导致的特征提取能力较弱，特征感受野较小且表征能力不足的缺点 |
| 可借鉴的地方： |
|  |
| 优点： |
|  |
| 缺点： |
|  |