



R Talk | 旷视姚聪博士：深度学习时代的文字检测与识别技术

 旷视科技 

已认证的官方帐号

+

 关注他

73 人赞同了该文章

「R Talk」是「北京智源—旷视智能模型设计与图像感知联合实验室」推出的一个深度学习专栏，将通过不定期的推送，展示旷视研究院的学术分享及阶段性技术成果。「R Talk」旨在通过一场场精彩纷呈的深度学习分享，抛砖引玉，推陈出新，推动中国乃至全球领域深度学习技术的发展。

AI 研习社 · 大讲堂

Scene Text Detection and Recognition:
The Deep Learning Era

Dr. Cong Yao
Megvii (Face++) Researcher
yaocong@megvii.com



近日，旷视科技云服务事业部算法负责人姚聪博士应邀在雷锋网 AI 研习社做了一次主题为自然场景文字检测与识别的线上报告：《Scene Text Detection and Recognition: The Deep Learning Era》，其内容共分为 5 个部分：

3. Future Trends and Potential Directions
4. Typical Applications
5. Conclusion

第 1 部分给出了文字检测与识别的定义，为什么在视觉识别中如此重要，以及当下深度学习时代语境中，面临的 3 个显性挑战；第 2 部分把当下的文字检测与识别技术归纳为 5 个类别，并分别给出了代表性成果与算法实例，反映出该领域的最新进展；第 3 部分通过分析文字检测与识别技术发展趋势及潜在方向提出了 4 个未来机遇（挑战）；第 4 部分举出了一些典型的行业落地应用；第 5 部分则做了最后总结。

以下是主要内容

1. Background and Overview

视觉识别中，文字为什么如此重要？有两个原因，分别是作为载体的文字和作为线索的文字。

■ 文字作为载体

Text as a Carrier of High-Level Semantics

Text is a creature of humankind that

- carries rich and precise high level semantics
- conveys human thoughts and emotions



Face++ 旷视

首先，文字并非自然产生，而是人类的创造物，天生包含着丰富而准确的高层语义信息，传达人类的思想和情感，这是一般的图像或者视频信息很难直接描述的，但是文字可以做到。比如《史记》，通过它可以了解两千年前的某个历史事件相关的人物、时间和地点。

■ 文字作为线索

Text as a Cue in Visual Recognition

Text is **complementary** to other visual cues, such as contour, color and texture



Face++ 旷视

其次，文字是视觉识别的重要线索。如上所示，这是一张街景图像，有建筑物，有植物；如果进一步提问，图像拍摄地点是哪儿，光靠以上信息无从得知，而移除中间的灰色掩膜，会发现这是“中

的信息。

■ 定义

那么，从研究角度讲，如何界定文字检测与识别问题呢？

Problem Definition



Scene text detection is the process of predicting the presence of text and localizing each instance (if any), usually at word or line level, in natural scenes



7

简单来讲，文字检测是指通过算法计算判断自然场景中在单词或者文字行层面是否含有文字实例（如果有，则标出位置）的过程。

Problem Definition



Scene text recognition is the process of converting text regions into computer readable and editable symbols



8

文字识别则更进一步，它在文字检测基础之上把文字区域转化为计算机可读取和编辑的符号的过程。

■ 挑战

目前，文字检测与识别面临诸多挑战。首先，它与传统 OCR（光学字符识别）有所区别，因自然场景之中文字变化非常之多，如下所示，左边是一张典型的扫描文档图像，右边则是自然场景中采集的多张图像。

Challenges

Traditional OCR vs. Scene Text Detection and Recognition



- clean background vs. cluttered background
- regular font vs. various fonts
- plain layout vs. complex layouts
- monotone color vs. different colors

Face++ 旷视

9

通过对比可以发现，左图的背景非常干净，右图的背景非常杂乱；左边字体非常规整，右边则千变万化；左边布局较为平整统一，右边则多元复杂，缺乏规范；左边颜色单调，右边颜色种类繁多。

总体而言，文字检测与识别存在三个方面的重大挑战：

- 1) 场景文字的多样性。如文字的颜色、大小、方向、语言、字体等。

Challenges



Diversity of scene text:
different colors, scales, orientations, fonts, languages...

Face++ 旷视

10

- 2) 图像背景的干扰。日常生活中随处可见的信号灯、指示标、栅栏、屋顶、窗户、砖块、花草等局部来看和文字有一定相似性，这为文字检测与识别过程带来很大干扰。

Challenges



Complexity of background:
elements like signs, fences, bricks, and grasses are virtually undistinguishable from true text

Face++ 旷视

11

Challenges



Various interference factors:

noise, blur, non-uniform illumination, low resolution, partial occlusion...

Face++ 旷视

12

2.Recent Advances and Representative Algorithms

正是由于存在上述多个挑战，研究者尝试从各种角度解决上述问题。这次探讨的主题是深度学习时代文字检测与识别领域的最新进展。这些进展被分成 5 个类别：1) 从语义分割和目标检测方法中汲取灵感，2) 更简化的 Pipeline，3) 处理任意形态文字，4) 使用 Attention，5) 使用合成数据。

■ 第一个分类：从语义分割和目标检测方法中汲取灵感

自然场景文字检测与识别技术从语义分割和目标检测方法中汲取灵感而产生的代表性工作主要有：

1) Holistic Multi-Channel Prediction, 2) TextBoxes, 3) Rotation Proposals, 4) Corner Localization and Region Segmentation。

深度学习方法相较于传统方法而言，会从通用物体语义分割和目标检测算法中汲取灵感。比如旷视科技 2015 年有一个工作，称之为 Holistic Multi-Channel Prediction。

Holistic, Multi-Channel Prediction



- holistic vs. local
- text detection is casted as a semantic segmentation problem
- conceptionally and functionally different from previous sliding-window or connected component based approaches

Face++ 旷视

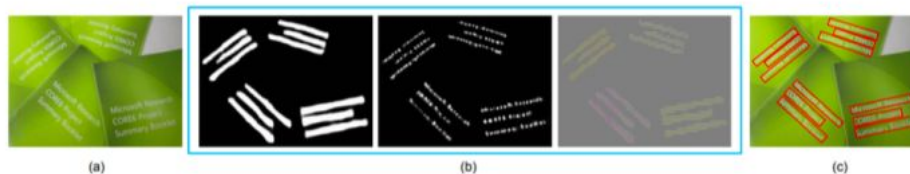
Yao et al., Scene Text Detection via Holistic, Multi-Channel Prediction. 2016. arXiv preprint arXiv:1606.09002

15

区别于传统文字检测方法，Holistic Multi-Channel Prediction 把文字检测问题转变为一个语义分割问题，通过对全图进行语义分割来获取文字区域，如上所示，左边为原图，右边为已预测的掩膜图，文字区域高亮，背景区域则被抑制。

Holistic Multi-Channel Prediction 的输入并非局部区域，而是整张图像；另外，无论从概念还是功能上来讲，该方法皆区别于滑窗或者连通分量方法。

Holistic, Multi-Channel Prediction



- **holistic, pixel-wise** predictions: text region map, character map and linking orientation map
- detections are formed using these three maps
- can **simultaneously** handle horizontal, multi-oriented and curved text in real-world natural images

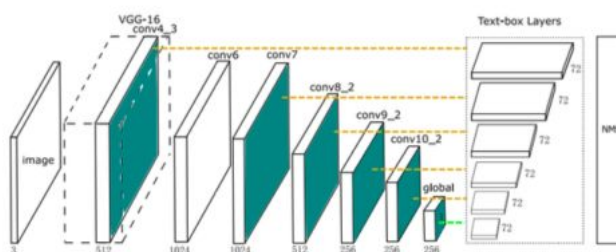


Yao et al., Scene Text Detection via Holistic, Multi-Channel Prediction, 2016, arXiv preprint arXiv:1606.09002

16

具体来讲，Holistic Multi-Channel Prediction 会输出全局的三种像素级预测，包括图像区域、字符位置和相邻字符间的连接方向。通过这三种信息，输出最右边的结果图，如红色矩形部分所示。这一方法的好处是可以同时处理水平、多方向、弯曲文字。

TextBoxes



- a text detection method inspired by **SSD**
- both high accuracy and efficiency

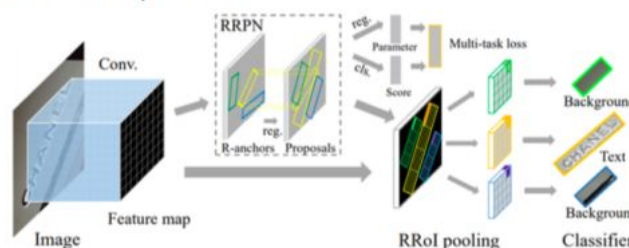


Liao et al., TextBoxes: A Fast Text Detector with a Single Deep Neural Network, AAAI, 2017.

17

另一种方法是 TextBoxes，它受到单步的通用物体检测器 SSD 启发，其本质是把文字作为一种特殊的目标，通过 SSD 对其进行建模。这里的基础模型是 VGG-16，卷积层输出目标文字是否存在以及长宽、大小、方向等信息，同时取得了很高的精度及效率。

Rotation Proposals



- a multi-oriented text detection method based on **Faster RCNN**
- propose several modifications to better detect scene text

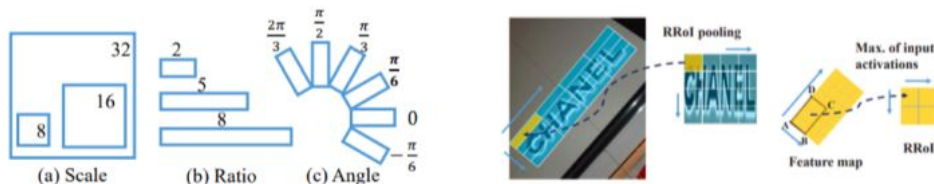


Ma et al., Arbitrary-Oriented Scene Text Detection via Rotation Proposals, arxiv, 2017.

18

目标检测领域还存在两步的检测器，比如 Faster R-CNN。第三种文字检测方法称之为 Rotation Proposals，它属于两步检测的范畴，主要借鉴了 Faster R-CNN，在框架保持一致之时，又有所修

Rotation Proposals



- anchor strategy: add orientations, change aspect ratios
- RRoI pooling layer: to handle **rotated** boxes

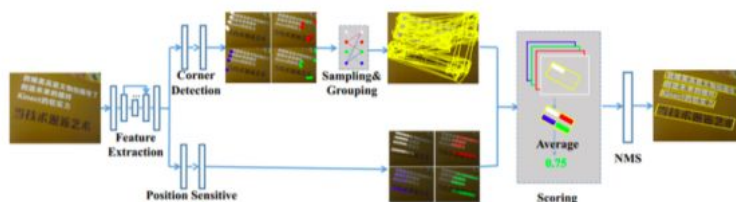


Ma et al., Arbitrary-Oriented Scene Text Detection via Rotation Proposals. arxiv, 2017.

19

Rotation Proposals 主要做两件事：第一是拓展锚点，增加方向，改变长宽比变化；第二是 RRoI 池化层，处理各种旋转情况，分离出规整的特征图，方便后续操作。

Corner Localization and Region Segmentation (A Megvii work in CVPR 2018)



- a compound text detection method: **corner localization** and **region segmentation**



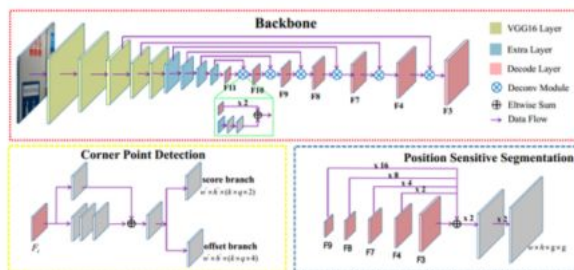
Lyu et al., Multi-Oriented Scene Text Detection via Corner Localization and Region Segmentation. CVPR, 2018.

20

旷视科技 CVPR 2018 收录论文《Multi-Oriented Scene Text Detection via Corner Localization and Region Segmentation》提出了一种复合的文字检测方法——Corner Localization and Region Segmentation（角点检测和区域分割），它最大亮点是综合利用目标分割与语义分割两种方法。

为什么提出这种方法？这源于实际场景文字实例有时相邻较近甚至粘连，分割和检测变得无济于事。这时通过两个分支进行处理，一个分支定位文字区域的角点，另一个分支进行文字区域分割，结合这两种信息文字输出文字位置和置信度。

Corner Localization and Region Segmentation



- corner localization: corner detection with SSD
- region segmentation: position-sensitive segmentation with R-FCN



Lyu et al., Multi-Oriented Scene Text Detection via Corner Localization and Region Segmentation. CVPR, 2018.

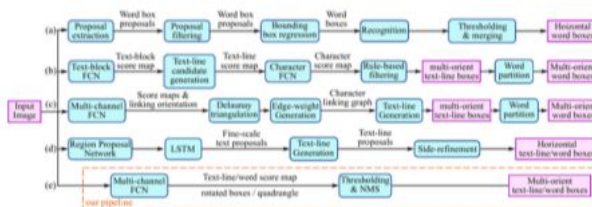
21

如上所示，Corner Localization and Region Segmentation 基础模型是 VGG-16，其上添加大量的卷积层，以提取特征，再往上是两个分支，1) 角点检测分支通过 SSD 定位角点，通过网络提取角点，最终得到角点位置；2) 文字区域分割分支则利用基于 R-FCN 的位置敏感分割，生成不同相对位置的分割图，得到更准确的文字检测结果。

第二个分类：更简化的 Pipeline

深度学习时代，几乎所有的文字检测与识别方法都会使用更加简化、更加高效的 Pipelines。这是要分享的第二个类方法，其代表性工作是 EAST。

EAST (A Megvii work in CVPR 2017)



- highly simplified pipeline

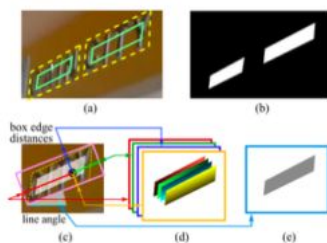


Zhou et al., EAST: An Efficient and Accurate Scene Text Detector. CVPR, 2017.

23

旷视科技在 CVPR 2017 收录论文《EAST: An Efficient and Accurate Scene Text Detector》提出一种高度简化的 Pipeline 结构。如上图所示，最左侧是输入图像，最右侧是算法输出结果，中间则是处理步骤，EAST（最下面）把 Pipeline 精简为中间两步，其中一步是通过多通道 FCN 进行几何信息预测以及前景、背景预测；另外一步是 NMS，最终得到多方向文字检测结果。

EAST



- main idea: predict location, scale and orientation of text with **a single model and multiple loss functions (multi-task training)**
- advantages:
 - (a). accuracy: allow for end-to-end training and optimization
 - (b). efficiency: remove redundant stages and processings



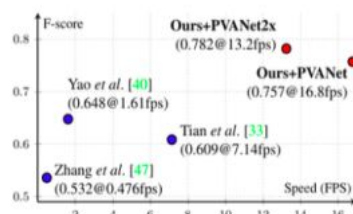
Zhou et al., EAST: An Efficient and Accurate Scene Text Detector. CVPR, 2017.

24

那么如何实现 EAST 呢？很简单，通过单一网络刻画并建模位置、尺度、方向等文字的主要信息，同时加上多个损失函数，即所谓的多任务训练。如上所示，再借助对几何信息的预测即可得到文字检测的结果。

这种方法的好处主要体现在两个方面：1) 精度方面，允许端到端的训练和优化，2) 效率方面，剔除了中间冗余的处理步骤。

EAST



- strike a good balance between accuracy and speed
- code available at: <https://github.com/argman/EAST> (reimplemented by a student outside Megvii (Face++), credit goes to @argman)
- now an official module in OpenCV
(https://github.com/opencv/opencv/blob/63564039640c81256ccfed77899e37746352d45/samples/dnn/text_detection.cpp)



Zhou et al., EAST: An Efficient and Accurate Scene Text Detector. CVPR, 2017.

25

上图是多种方法的应用对比，横轴是速度，纵轴是精度，红点是 EAST 的两个变体，可以看到其在精度和速度上都优于传统方法，且在精度和速度质检取得了极佳的平衡。EAST 已成为为行业标准方法，且代码开源，有趣的是，这是由热心网友 @argman 完成的（旷视贡献了部分代码）有兴趣的童鞋可以尝试；如果只想使用不想看源代码，目前 EAST 也作为官方模块，集成到最新版 OpenCV 之中。

第三个分类：处理任意形态文字

要处理现实世界的文字还面临着一个挑战：文字形态的多变性。文字检测与识别算法要如何应对呢？旷视科技为此给出了两个代表性方案：1) TextSnake, 2) Mask TextSpotter。

知乎

首发于
旷视Text组

关注

TextSnake (A Megvii work in ECCV 2018)



- a novel and flexible representation
- able to effectively and precisely describe the geometric properties, such as location, scale, and bending of curved text, while the other representations (axis-aligned rectangle, rotated rectangle or quadrangle) struggle

Face++ 旷视

Long et al., TextSnake: A Flexible Representation for Detecting Text of Arbitrary Shapes, ECCV, 2018.

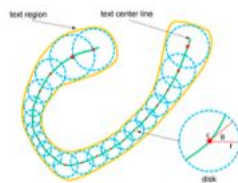
27

旷视科技 ECCV 2018 收录论文《TextSnake: A Flexible Representation for Detecting Text of Arbitrary Shapes》提出一种全新而灵活的表征，称之为 TextSnake。

如上所示，对于弯曲的文字区域，图 a 使用了传统的坐标对齐的矩形，包含了大量不需要的背景内容；图 b 将矩形框旋转，背景内容减少，适应力也随之变强；图 c 使用了不规则四边形进行包围，效果依然不理想。由此可以得出结论，四边形无法很好地包围弯曲的文字区域。

为了更精确地处理这种情况，图 d 使用了 TextSnake 方法，用一系列圆盘覆盖文字区域，更好地适应文字的变化，包括尺度、方向、形变等等。

TextSnake (A Megvii work in ECCV 2018)



$$S(t) = \{D_0, D_1, \dots, D_i, \dots, D_n\}$$

$$D = (c, r, \theta)$$

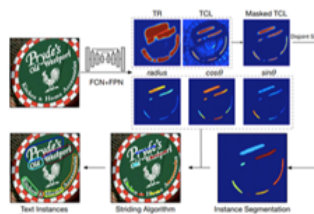
- a text instance is described as a sequence of ordered, overlapping disks centered at symmetric axes, each of which is associated with potentially variable radius and orientation

Face++ 旷视

Long et al., TextSnake: A Flexible Representation for Detecting Text of Arbitrary Shapes, ECCV, 2018.

28

TextSnake 原理示意图如上，黄色区域代表不规则的弯曲文字区域，绿线代表文字区域的中心线，红点代表覆盖文字区域的圆盘中心点，用 c 表示，圆盘半径用 r 表示，方向信息用 θ 表示。由此，文字实例可建模为一系列以对称轴为中心的有序重叠的圆盘。由于文字的尺度和方向是变化的，圆盘也有着可变的半径和方向，随文字变化而变化。这种灵活的表示方式可以精确适应各种形态的文字，比如多方向、弯曲文字。



- such geometry attributes of text instances are estimated via FCN

有了 TextSnake 之后，如何设计一个计算模型完成文字检测呢？其实很简单，即几何属性定义完成，通过 FCN 输出多个通道的预测结果，并进行整合，分开前景、背景，找出中心线，根据其半径和方向即可恢复出文字区域。

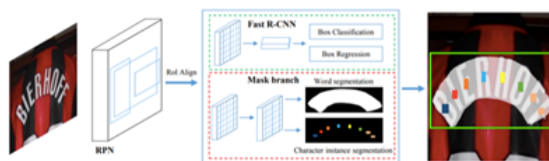
Mask TextSpotter (A Megvii work in ECCV 2018)



- an end-to-end system for both text detection and recognition
- inspired by Mask R-CNN

旷视科技 ECCV 2018 收录论文《Mask TextSpotter: An End-to-End Trainable Neural Network for Spotting Text with Arbitrary Shapes》完成了另外一项工作，在 Mask R-CNN 启发之下提出一种新模型 Mask TextSpotter，通过端到端的方式同时实现了文字检测和识别。Mask TextSpotter 整体框架基于 Mask R-CNN 并进行改造，同样也把文字当作一种特殊目标处理。

Mask TextSpotter (A Megvii work in ECCV 2018)



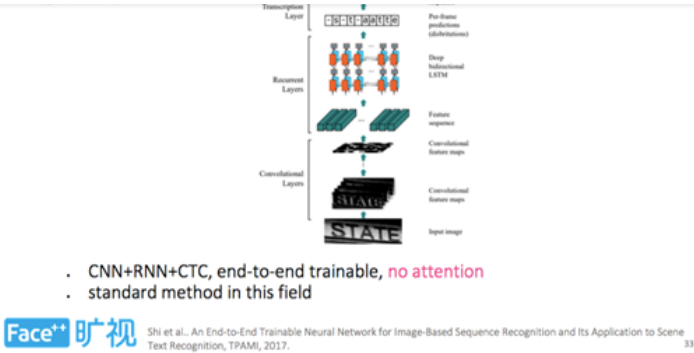
- RPN for text proposal generation
- Fast R-CNN for proposal classification and regression
- mask branch for character segmentation and recognition

具体细节如上所示，左边是输入图像，右边是输出结果。RPN 做文字候选区域的产生；Fast R-CNN 对候选区域打分并回归其坐标；另外还有一个掩膜分支，同时对字符做像素级的分割和识别。

这有别于传统方法的识别模块基于 RNN 序列进行建模，或者借助 CTC loss 进行计算，Mask TextSpotter 则直接做像素级分割，对每一像素进行分类，最后连接所有结果完成识别。

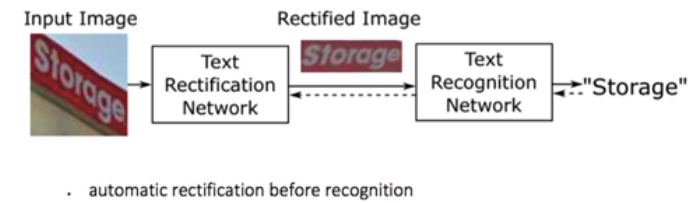
第四个分类：借鉴 Attention

由于 NLP 领域兴起的 Attention 模型的重大影响，其也进入了文字检测与识别的视野，激发出一些新想法，代表性成果有：1) CRNN, 2) ASTER, 3) FAN。

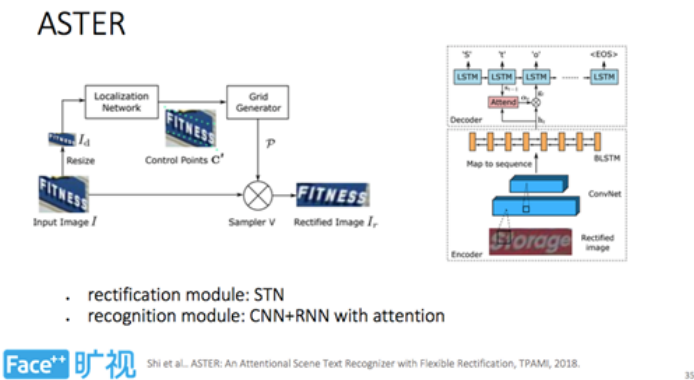


在讲 Attention 之前，首先讲一下旷视科技 TPAMI 2017 的一个工作，称之为 CRNN，其底层用 CNN 提取特征，中层用 RNN 进行序列建模，上层用 CTC loss 对目标进行优化。它是一个端到端可训练的文字识别结构，但并未使用 Attention。目前，CRNN 已成长为该领域的一个标准方法，在 GitHub 上已开源。

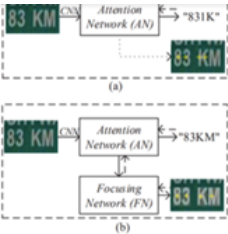
ASTER



随后，旷视科技在 TPAMI 2018 提出一个称之为 ASTER 的解决方案。由于文字存在倾斜、弯曲等问题，在识别阶段，检测也不一定是最理想的，这时需要分两步做识别。第一步是给定一张输入图像，把其中的文字矫正到一个有利于识别的状态；第二步则是进行识别。这里需强调的是矫正过程是网络自动学习的，并不存在标注。



那么，如上所示，ASTER 主要有矫正和识别两个模块。矫正模块在 STN 的基础上做了优化，使得控制点的预测更精确；识别模块则是一个经典的 CNN+RNN 同时带有 Attention 的结构，可以对序列进行预测。



- propose the concept of attention drift
- force deviated attentions to focus on target areas (character centers)

Face++ 旷视

Cheng et al., Focusing Attention: Towards Accurate Text Recognition in Natural Images, ICCV, 2017.

36

当然，Attention 本身也存在一些问题，比如 ICCV 2017 收录论文《Focusing Attention: Towards Accurate Text Recognition in Natural Images》提出了 FAN 这一工作。

某些情况下，Attention 预测不一定准确，如图 a 中 Attention 把后两个点定位在“K”上，造成 Attention 点发生重合，进而导致定位不准确与漂移。FAN 方法通过约束 Attention 点位置，将其锁定为目标文字中心，这样就避免了重合、漂移等情况，Attention 区域变得更整齐，最终识别效果也得到了提升。

■ 第五个分类：使用合成数据

深度学习时代，对数据的需求量大增，大量数据有利于训练出优秀模型。因此，深度学习时代的文字检测和识别方法几乎都会采用合成数据，代表性数据集有 SynthText。

SynthText



- a synthetically generated scene text image dataset
- widely used for scene text detection and recognition

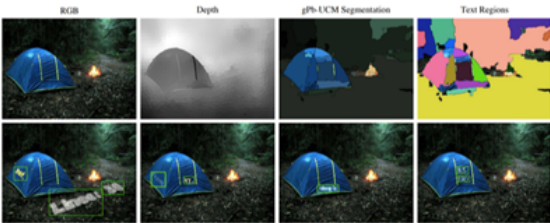
Face++ 旷视

Gupta et al., Synthetic Data for Text Localisation in Natural Images, CVPR, 2016.

38

SynthText 是一个经典的通过合成方式产生的场景文字数据集，已在文字检测和识别领域获得广泛应用。其主要思路是先采集几千张真实场景图像，然后往上贴字，如上图所示。

SynthText



- segmentation and geometry estimation
- local colour/texture sensitive placement

Face++ 旷视

Gupta et al., Synthetic Data for Text Localisation in Natural Images, CVPR, 2016.

39

具体而言，给定一些自然场景图像之后，SynthText 使用基础算法估计几何信息以及区域分割的结果，比如估计深度，通过底层算法分割连通的区域，同时找到一些平滑的区域放置文字，最后生成包含文字的图像数据集。

3.Future Trends and Potential Directions

别，2) 读取任意形态的文字，3) 文字图像合成，4) 模型鲁棒性。

■ 多语言文字检测与识别

Multi-Lingual Text Detection and Recognition



Can we find unified approaches for handling different scripts?

Face++ 旷视 Images from ICDAR 2017 MLT competition. Homepage: <http://rrc.cvc.uab.es/?ch=8>

42

针对较为整齐的文字区域，目前文字检测技术在使用上问题不大，但是一旦涉及文字识别，超过两百种语言文字，不同的书写方式，千差万别的结构、顺序，带来了重重阻碍，而针对每一种文字都训练一种模型显然很不合理。因此，是否可以找到一种通用方法处理不同的文字类型呢？这是未来挑战之一。

■ 读取任意形态的文字

Reading Text of Arbitrary Shapes



Can we find unified approaches for text instances of arbitrary shapes (e.g., horizontal, oriented, curved)?

Face++ 旷视 Images from Total-Text and IIT-5K Word, respectively

43

当前文字的形态多种多样，面对不同的颜色、字体、组合带来的不同程度的困难，文字检测与识别技术目前的表现并不是令人很满意。那么，是否存在一种通用模型胜任任意形态的文字呢？这是未来挑战之二。

■ 文字图像合成

Text Image Synthesis



How to synthesize more realistic and diverse text images?

Face++ 旷视 Images from SynthText

44

虽然合成场景文字数据集非常有利于模型训练，但是由于该技术目前尚不成熟，生成图像不够真实，图像多样性欠缺，最终导致数据集质量不高，训练提升有限。那么，如何才能合成更加真实丰

■ 模型鲁棒性

Model Robustness



Existing recognition models seem to be very sensitive to subtle disturbances (e.g., small shift, rotation, distortion). How to make them more robust?



Liao et al., Scene Text Recognition from Two-Dimensional Perspective, AAAI, 2019

45

模型鲁棒性是一个基础问题，也是解决问题的关键。旷视科技过去曾针对输入图像的边界进行像素扰动，结果发现输出发生较大偏移，甚至完全错误。这也意味着现有识别模型对微小扰动（如小偏移、旋转、形变）非常敏感。那么，如何增强模型鲁棒性呢？这是未来挑战之四。

4. Typical Applications

在深度学习助力之下，文字检测与识别技术获得跃进式发展，在相关场景和行业获得广泛应用，比如旷视科技人工智能开放平台 Face++ 提供卡片、证件文字识别 API 调用服务，给出卡片、证件、单据的通用框架解决方案 TemplateOCR。

■ 卡片、证件文字识别

Text Information Extraction from Cards/Certificates



Available at: <https://www.faceplusplus.com.cn/bank-card-recognition/#demo>
<https://www.faceplusplus.com.cn/license-plate-recognition/>

48

卡片、证件的文字识别一直是热门的行业需求，旷视科技 Face++ 官网提供了 API 可供用户调用，用户只需上传图片，系统便会实时生成结果。目前系统支持身份证、驾驶证、行驶证、银行卡、车牌号等文字识别。

■ TemplateOCR



What if we need to recognize hundreds of types of cards/certificates/receipts?



49

力、物力、财力，显然并不划算。

Solution: TemplateOCR



- Upload a clean image -> Specify fields to be recognized -> Go! (API ready instantly)
- No training, no waiting!

Face++ 旷视 Will be available soon at: <https://www.faceplusplus.com.cn/>

51

TemplateOCR 是旷视给出的解决方案。这是一种用于识别结构化卡片、证件、单据的通用框架，只需三步操作。第一步是上传清晰的图片作为模板，第二步是指定需要识别的文字区域(不需要指定参考区域)，然后即可开始识别(API 即时生成)。该方法不仅省去了针对不同框架的文字进行模型训练的麻烦，也使识别过程变得更加快速和准确。

5.Conclusion

目前来看，深度学习时代之下的场景文字检测与识别技术依然存在巨大挑战，其主要体现在以下 3 个方面：1) 文字的差异性，存在着语言、字体、方向、排列等各种各样的形态；2) 背景的复杂性，比如几乎无法区分的元素（标志、篱笆、墙砖、草地等）；3) 干扰的多样性，诸如噪声、模糊、失真、低分辨率、光照不均匀、部分遮挡等情况。

相应而言，针对上述挑战，文字检测与识别具体研究有着以下 4 个技术趋势：1) 设计更强大的模型；2) 识别多方向、弯曲文字；3) 识别多语言文字；4) 合成更丰富逼真的数据集。

编辑于 2019-10-25

计算机视觉 神经网络 文字识别（技术）

赞同 73 2 条评论 分享 喜欢 收藏 ...

文章被以下专栏收录

旷视Text组

关注专栏

推荐阅读

基于深度学习的图像目标检测(下)

在“基于深度学习的图像目标检测(上)”里面，我们详细介绍了R-CNN走到端到端模型的Faster R-CNN的进化流程。这里，我们介绍，后续如何变得更快、更强！天下武功唯快不破！所以在如何...

史博 发表于聊聊-深度...



深度学习在计算机视觉领域（包括图像，视频，3-D点云，深...

黄浴



第5弹！CS231n 2017中文字幕视频：卷积神经网络

AI研习社 发表于AI研习社...



视频换脸一起动，

量子位

知乎




首发于
旷视Text组

关注

2 条评论


⇌ 切换为时间排序

- 写下你的评论...

😊
-  lijinchao2007

好文

8 个月前

👍 赞
-  lijinchao2007

CRNN厉害

8 个月前

👍 赞