MMOCR: A Comprehensive Toolbox for Text Detection, Recognition and Understanding

Zhanghui Kuang
SenseTime Research
China

Hongbin Sun
SenseTime Research
China

Zhizhong Li
SenseTime Research
China

Xiaoyu Yue
Center for Perceptual and Interactive Intelligence
Hong Kong

Tsui Hin Lin
SenseTime Research
China

Jianyong Chen
South China University of Technology
China

Huaqiang Wei
SenseTime Research
China

Yiqin Zhu
South China University of Technology
China

Tong Gao
SenseTime Research
China

Wenwei Zhang
Nanyang Technological University
Singapore

Kai Chen
SenseTime Research,
Shanghai AI Laboratory
China

Wayne Zhang
SenseTime Research
China

Dahua Lin
The Chinese University of Hong Kong
Hong Kong

ABSTRACT

We present MMOCR—an open-source toolbox which provides a comprehensive pipeline for text detection and recognition, as well as their downstream tasks such as named entity recognition and key information extraction. MMOCR implements 14 state-of-the-art algorithms, which is significantly more than all the existing open-source OCR projects we are aware of to date. To facilitate future research and industrial applications of text recognition-related problems, we also provide a large number of trained models and detailed benchmarks to give insights into the performance of text detection, recognition and understanding. MMOCR is publicly released at https://github.com/open-mmlab/mmocr.

CCS CONCEPTS

• **Computing methodologies → Object recognition**.

Jianyong Chen and Yiqin Zhu are students from South China University of Technology China.

MM ’21, October 20–24, 2021, Virtual Event, China
© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-8651-7/21/10…$15.00
https://doi.org/10.1145/3474085.3478328

KEYWORDS

open source, text detection, text recognition, named entity recognition, key information extraction

ACM Reference Format:
Zhanghui Kuang, Hongbin Sun, Zhizhong Li, Xiaoyu Yue, Tsui Hin Lin, Jianyong Chen, Huaqiang Wei, Yiqin Zhu, Tong Gao, Wenwei Zhang, Kai Chen, Wayne Zhang, and Dahua Lin. 2021. MMOCR: A Comprehensive Toolbox for Text Detection, Recognition and Understanding. In *Proceedings of the 29th ACM International Conference on Multimedia (MM ’21), October 20–24, 2021, Virtual Event, China.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3474085.3478328

1 INTRODUCTION

In recent years, deep learning has achieved tremendous success in fundamental computer vision applications such as image recognition [8, 27, 30], object detection [6, 15, 20, 22] and image segmentation [7, 17]. In light of this, deep learning has also been applied to areas such as text detection [4, 34, 46] and text recognition [11, 13, 35, 37, 40], as well as their downstream tasks such as key information extraction [5, 29, 39] and named entity recognition [3, 36].

Different approaches utilize different training datasets, optimization strategies (*e.g.*, optimizers, learning rate schedules, epoch numbers, pre-trained weights, and data augmentation pipelines), and network designs (*e.g.* network architectures and losses). To encompass the diversity of components used in various models, we have proposed the MMOCR toolbox which covers recent popular text detection, recognition and understanding approaches in a unified framework. As of now, the toolbox implements seven text detection methods, five text recognition methods, one key information

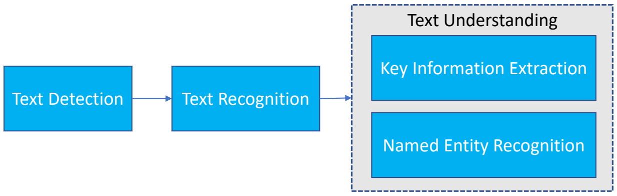

Figure 1: Overview of MMOCR. The supported text detection algorithms include DB [12], Mask R-CNN [7], PANet [34], PSENet [33], TextSnake [18], DRRG [43], and FCENet [46]. The supported text recognition algorithms are CRNN [24], NRTR [23], RobustScanner [40], SAR [11], and SegOCR [41]. The supported key information extraction algorithm is SDMG-R [29], and the supported named entity extraction algorithm is Bert-Softmax [36].

method and one named entity recognition method. Integrating various algorithms confers code reusability and therefore dramatically simplifies the implementation of algorithms. Moreover, the unified framework allows different approaches to be compared against each other fairly and that their key effective components can be easily investigated. To the best of our knowledge, MMOCR reimplements the largest number of deep learning-based text detection and recognition approaches amongst various open-source toolboxes, and we believe it will facilitate future research on text detection, recognition and understanding.

Extracting structured information such as "shop name", "shop address" and "total payment" in receipt images, and "name" and "organization name" in document images plays an important role in many practical scenarios. For example, in the case of office automation, such structured information is useful for efficient archiving or compliance checking. To provide a comprehensive pipeline for practical applications, MMOCR reimplements not only text detection and text recognition approaches, but also their downstream tasks such as key information extraction and named entity recognition as illustrated in Figure 1. In this way, MMOCR can meet the document image processing requirements in a one-stop-shopping manner.

MMOCR is publicly released at https://github.com/open-mmlab/mmocr under the Apache-2.0 License. The repository contains all the source code and detailed documentation including installation instructions, dataset preparation scripts, API documentation, model zoo, tutorials and user manual. MMOCR re-implements more than ten state-of-the-art text detection, recognition, and understanding algorithms, and provides extensive benchmarks and models trained on popular academic datasets. To support multilingual OCR tasks, MMOCR also releases Chinese text recognition models trained on industrial datasets [1]. In addition to (distributed) training and testing scripts, MMOCR offers a rich set of utility tools covering visualization, demonstration and deployment. The models provided by MMOCR are easily converted to onnx [2] which is widely supported by deployment frameworks and hardware devices. Therefore, it is useful for both academic researchers and industrial developers.

2 RELATED WORK

Text detection. Text detection aims to localize the bounding boxes of text instances [4, 7, 14, 31, 42, 46]. Recent research focus has shifted to challenging arbitrary-shaped text detection [4, 46]. While Mask R-CNN [7, 14] can be used to detect texts, it might fail to detect curved and dense texts due to the rectangle-based ROI proposals. On the other hand, TextSnake [18] describes text instances with a series of ordered, overlapping disks. PSENet [33] proposes a progressive scale expansion network which enables the differentiation of curved text instances that are located close together. DB [12] simplifies the post-processing of binarization for scene-text segmentation by proposing a differentiable binarization function to a segmentation network, where the threshold value at every point of the probability map of an image can be adaptively predicted.

Text recognition. Text recognition has gained increasing attention due to its ability to extract rich semantic information from text images. Convolutional Recurrent Neural Network (CRNN) [24] uses an end-to-end trainable neural network which consists of a Deep Convolutional Neural Networks (DCNN) for the feature extraction, a Recurrent Neural Networks (RNN) for the sequential prediction and a transcription layer to produce a label sequence. RobustScanner [40] is capable of recognizing contextless texts by using a novel position enhancement branch and a dynamic fusion module which mitigate the misrecognition issue of random text images. Efforts have been made to rectify irregular texts input into regular ones which are compatible with typical text recognizers. For instance, Thin-Plate-Spline (TPS) transformation is employed in a deep neural network that combines a Spatial Transformer Network (STN) and a Sequence Recognition Network (SRN) to rectify curved and perspective texts in STN before they are fed into SRN [26].

Key information extraction. Key Information Extraction (KIE) for unstructured document images, such as receipts or credit notes, is most notably used for office automation tasks including efficient archiving and compliance checking. Conventional approaches, such as template matching, fail to generalize well on documents of unseen templates. Several models are proposed to resolve the generalization problem. For example, CloudScan [19] employs NER to analyze the concatenated one-dimensional text sequence for the entire invoice. Chargrid [5] encodes each document page as a two-dimensional grid of characters to conduct semantic segmentation, but it cannot make full use of the non-local, distant spatial relation between text regions since it covers two-dimensional spatial layout information with small neighborhood only. Recently, an end-to-end Spatial Dual Modality Graph Reasoning (SDMG-R) model [29] has been developed which is particularly robust against text recognition errors. It models unstructured document images as spatial dual-modality graphs with graph nodes as detected text boxes and graph edges as spatial relations between nodes.

Named entity recognition. Named entity recognition (NER) [3, 10, 36, 44] aims to locate and classify named entities into pre-defined categories such as the name of a person or organization. They are based on either bidirectional LSTMs or conditional random fields.

Open source OCR toolbox. Several open-source OCR toolboxes have been developed over the years to meet the increasing demand from both academia and industry. Tesseract[3] is the pioneer of open-source OCR toolbox. It was publicly released in 2005, and provides CLI tools to extract printed font texts from images. It initially followed a traditional, step-by-step pipeline comprising the connected component analysis, text line finding, baseline fitting,

[1] https://github.com/chineseocr/chineseocr
[2] https://github.com/onnx/onnx
[3] https://github.com/tesseract-ocr/tesseract

Table 1: Comparison between different open-source OCR toolboxes.

Toolbox	tesseract	chineseocr	chineseocr_lite	EasyOCR	PaddleOCR	MMOCR
DL library	—	PyTorch	PyTorch	PyTorch	PaddlePaddle	PyTorch
Inference engine	—	OpenCV DNN	NCNN TNN onnx runtime	PyTorch	Paddle inference Paddle lite	PyTorch onnx runtime TensorRT
OS	— Linux Android IOS	Windows Linux — —	Windows Linux Android IOS	Windows Linux — —	Windows Linux Android IOS	Windows Linux — —
Language#	100+	2	2	80+	80+	2
Detection	convention	YOLOV3 [21]	DB [12]	CRAFT [1]	EAST [45], DB [12], SAST [31]	MaskRCNN [7], PAN [34], PSENet [33] DB [12], TextSnake [18], DRRG [43], FCENet [46]
Recognition	convention LSTM	CRNN [24]	DB [12]	CRNN [24]	CRNN [24], Rosetta [2], SRN [38] Star-Net [16], RARE [25]	CRNN [24], RobustScanner [40], SAR [11] SegOCR [41], Transformer [11]
Downstream tasks						KIE, NER
Support training	Yes	Yes	No	No	Yes	Yes

Table 2: The effects of backbones. All models are pre-trained on ImageNet, and trained on ICDAR2015 training set and evaluated on its test set.

		PSENet			PAN			DB		
Backbone	FLOPs	Recall	Precision	H-mean	Recall	Precision	H-mean	Recall	Precision	H-mean
ResNet18	37.1G	73.5	83.8	78.3	73.4	85.6	79.1	73.1	87.1	79.5
ResNet50	78.9G	78.4	83.1	80.7	73.2	85.5	78.9	77.8	82.1	79.9
ddrnet23-slim [9]	16.7G	75.2	80.1	77.6	72.3	83.4	77.5	76.7	78.5	77.6

Table 3: The effects of necks. All models are pre-trained on ImageNet, and trained on ICDAR2015 training set and evaluated on its test set.

		PSENet			PAN			DB		
Necks	FLOPs	Recall	Precision	H-mean	Recall	Precision	H-mean	Recall	Precision	H-mean
FPNF [33]	208.6G	78.4	83.1	80.7	72.4	86.4	78.8	77.5	82.3	79.8
PFNC [12]	22.4G	75.6	80.0	77.7	70.9	83.3	76.6	73.1	87.1	79.5
FPEM_FFM [34]	7.79G	71.7	82.0	76.5	73.4	85.6	79.1	71.8	86.7	78.6

fixed pitch detection and chopping, proportional word finding, and word recognition [28]. It now supports an LSTM-based OCR engine and supports more than 100 languages. Deep learning-based open-source OCR toolbox EasyOCR [4] has been released recently. It provides simple APIs for industrial users and supports more than 80 languages. It implemented the CRAFT [1] detector and CRNN [24] recognizer. However, it is for inference only and does not support model training. ChineseOCR [5] is another popular open-source OCR toolbox. It uses YOLO-v3 [21] and CRNN [24] for text detection and recognition respectively, and uses OpenCV DNN for deep models inference. By contrast, ChineseOCR_lite [6] releases a lightweight Chinese detection and recognition toolbox that uses DB [12] to detect texts and CRNN [24] to recognize texts. It provides forward inference based on NCNN [7] and TNN [8], and can be deployed easily on multiple platforms such as Windows, Linux and Android. PaddleOCR [9] is a practical open-source OCR toolbox based on PaddlePaddle and can be deployed on multiple platforms such as Linux, Windows and MacOS. It currently supports more than 80 languages and implements three text detection methods (EAST [45], DB [12], and SAST [31]), five recognition methods (CRNN [24], Rosetta [2], STAR-Net [16], RARE [25] and SRN [38]), and one end-to-end text spotting method (PGNet) [32]. Comprehensive comparisons among these open-source toolboxes are given in Table 1.

3 TEXT DETECTION STUDIES

Many important factors can affect the performance of deep learning-based models. In this section, we investigate the backbones and necks of network architectures. We exchange the above components between different segmentation-based text detection approaches to measure the performance and computational complexity effects.

Backbone. ResNet18 [8] and ResNet50 [8] are commonly used in text detection approaches. For practical applications, we also introduce a GPU-friendly lightweight backbone ddrnet23-slim [9]. Table 2 compares ResNet18, ResNet50 and ddrnet23-slim in terms of FLOPs and H-mean by plugging them in PSENet, PAN and DB. It has been shown that ddrnet23-slim performs slightly worse than ResNet18 and ResNet50, as it only has 45% and 21% FLOPs of ResNet18 and ResNet50 respectively.

Neck. PSENet, PAN and DB propose different FPN-like necks to fuse multi-scale features. Our experimental results in Table 3 show that the FPNF proposed in PSENet [33] can achieve the best H-mean in PSENet and DB [12]. However, its FLOPs are substantially higher than those of PFNC proposed in DB [12] and FPEM_FFM proposed

[4] https://github.com/JaidedAI/EasyOCR
[5] https://github.com/chineseocr/chineseocr
[6] https://github.com/DayBreak-u/chineseocr_lite
[7] https://github.com/Tencent/ncnn
[8] https://github.com/Tencent/TNN
[9] https://github.com/PaddlePaddle/PaddleOCR

in PAN [34]. By contrast, FPEM_FFM has the lowest FLOPs and achieves the best H-mean in PAN [34].

4 CONCLUSIONS

We have publicly released MMOCR, which is a comprehensive toolbox for text detection, recognition and understanding. MMOCR has implemented 14 state-of-the-art algorithms, which is more than all the existing open-source OCR projects. Moreover, it has offered a wide range of trained models, benchmarks, detailed documents, and utility tools. In this report, we have extensively compared MMOCR with other open-source OCR projects. Besides, we have introduced a GPU-friendly lightweight backbone-ddrnet23-slim, and carefully studied the effects of backbones and necks in terms of detection performance and computational complexity which can guide industrial applications.

5 ACKNOWLEDGEMENT

This work was supported by the Shanghai Committee of Science and Technology, China (Grant No. 20DZ1100800).

REFERENCES

[1] Youngmin Baek, Bado Lee, Dongyoon Han, Sangdoo Yun, and Hwalsuk Lee. 2019. Character region awareness for text detection. In *CVPR*. 9365–9374.

[2] Fedor Borisyuk, Albert Gordo, and Viswanath Sivakumar. 2019. Rosetta: Large scale system for text detection and recognition in images. *ACM SIGKDD* (2019), 71–79.

[3] Jason P.C. Chiu and Eric Nichols. 2016. Named Entity Recognition with Bidirectional LSTM-CNNs. *Transactions of the Association for Computational Linguistics* 4 (2016), 357–370.

[4] Jiaqi Duan, Youjiang Xu, Zhanghui Kuang, Xiaoyu Yue, Hongbin Sun, Yue Guan, and Wayne Zhang. 2019. Geometry normalization networks for accurate scene text detection. In *ICCV*. 9136–9145.

[5] Anoop Raveendra Katti Faddoul, Christian Reisswig Cordula Guder, Sebastian Brarda, Steffen Bickel, Johannes Höhne, and Jean Baptiste. 2018. Chargrid: Towards Understanding 2D Documents. In *EMNLP*. 4459–4469.

[6] Ross Girshick. 2015. Fast R-CNN. In *ICCV*. 1440–1448.

[7] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. Mask R-CNN. In *ICCV*. 2961–2969.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *CVPR*. 770–778.

[9] Yuanduo Hong, Huihui Pan, Weichao Sun, and Yisong Jia. 2021. Deep Dual-resolution Networks for Real-time and Accurate Semantic Segmentation of Road Scenes. *CoRR* abs/2101.06085 (2021).

[10] Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. Neural architectures for named entity recognition. *arXiv preprint arXiv:1603.01360* (2016).

[11] Hui Li, Peng Wang, Chunhua Shen, and Guyu Zhang. 2019. Show, Attend and Read: A Simple and Strong Baseline for Irregular Text Recognition. *AAAI* (2019), 8610–8617.

[12] Minghui Liao, Zhaoyi Wan, Cong Yao, Kai Chen, and Xiang Bai. 2020. Real-Time Scene Text Detection with Differentiable Binarization. In *AAAI*. 11474–11481.

[13] Minghui Liao, Jian Zhang, Zhaoyi Wan, Fengming Xie, Jiajun Liang, Pengyuan Lyu, Cong Yao, and Xiang Bai. 2019. Scene text recognition from two-dimensional perspective. *AAAI* (2019), 8714–8721.

[14] Jingchao Liu, Xuebo Liu, Jie Sheng, Ding Liang, Xin Li, and Qingjie Liu. 2019. Pyramid Mask Text Detector. *CoRR* (2019).

[15] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng Yang Fu, and Alexander C. Berg. 2016. SSD: single shot multibox detector. In *ECCV*. 21–37.

[16] Wei Liu, Chaofeng Chen, Kwan-Yee K Wong, Zhizhong Su, and Junyu Han. 2016. STAR-Net: A SpaTial Attention Residue Network for Scene Text Recognition.. In *BMVC*.

[17] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. In *CVPR*. 3431–3440.

[18] Shangbang Long, Jiaqiang Ruan, Wenjie Zhang, Xin He, Wenhao Wu, and Cong Yao. 2018. TextSnake: A Flexible Representation for Detecting Text of Arbitrary Shapes. In *ECCV*. 19–35.

[19] Rasmus Berg Palm, Ole Winther, and Florian Laws. 2017. CloudScan - A configuration-free invoice analysis system using recurrent neural networks. In *ICDAR*. 406–413.

[20] Joseph Redmon and Ali Farhadi. 2017. YOLO9000: Better, Faster, Stronger. In *CVPR*. 6517–6525.

[21] Joseph Redmon and Ali Farhadi. 2018. YOLOv3: An Incremental Improvement. *CoRR* abs/1804.02767 (2018).

[22] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. In *NIPS*. 91–99.

[23] Fenfen Sheng, Zhineng Chen, and Bo Xu. 2019. NRTR: A no-recurrence sequence-to-sequence model for scene text recognition. In *ICDAR*. 781–786.

[24] Baoguang Shi, Xiang Bai, and Cong Yao. 2016. An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition. *PAMI* 39, 11 (2016), 2298–2304.

[25] Baoguang Shi, Xinggang Wang, Pengyuan Lyu, Cong Yao, and Xiang Bai. 2016. Robust Scene Text Recognition with Automatic Rectification. In *CVPR*. 4168–4176.

[26] Baoguang Shi, Mingkun Yang, Xinggang Wang, Pengyuan Lyu, Cong Yao, and Xiang Bai. 2018. ASTER : An Attentional Scene Text Recognizer with Flexible Rectification. *PAMI* 41, 9 (2018), 2035–2048.

[27] Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *ICLR*.

[28] Ray Smith. 2007. An Overview of the Tesseract OCR Engine. In *ICDAR*. 629–633.

[29] Hongbin Sun, Zhanghui Kuang, Xiaoyu Yue, Chenhao Lin, and Wayne Zhang. 2021. Spatial Dual-Modality Graph Reasoning for Key Information Extraction. *arXiv preprint* (2021).

[30] C Szegedy, W Liu, Y Jia, and P Sermanet. 2015. Going deeper with convolutions. In *CVPR*. 1–9.

[31] Pengfei Wang, Chengquan Zhang, Fei Qi, Zuming Huang, Mengyi En, Junyu Han, Jingtuo Liu, Errui Ding, and Guangming Shi. 2019. A Single-Shot Arbitrarily-Shaped Text Detector based on Context Attended Multi-Task Learning. In *ACM MM*. 1277–1285.

[32] Pengfei Wang, Chengquan Zhang, Fei Qi, Shanshan Liu, Xiaoqiang Zhang, Pengyuan Lyu, Junyu Han, Jingtuo Liu, Errui Ding, and Guangming Shi. 2021. PGNet: Real-time Arbitrarily-Shaped Text Spotting with Point Gathering Network. In *AAAI*. 2782–2790.

[33] Wenhai Wang, Enze Xie, Xiang Li, Wenbo Hou, Tong Lu, Gang Yu, and Shuai Shao. 2019. Shape robust text detection with progressive scale expansion network. In *CVPR*. 9336–9345.

[34] Wenhai Wang, Enze Xie, Xiaoge Song, Yuhang Zang, Wenjia Wang, Tong Lu, Gang Yu, and Chunhua Shen. 2019. Efficient and Accurate Arbitrary-Shaped Text Detection with Pixel Aggregation Network. In *ICCV*. 8439–8448.

[35] Zecheng Xie, Yaoxiong Huang, Yuanzhi Zhu, Lianwen Jin, Yuliang Liu, and Lele Xie. 2019. Aggregation cross-entropy for sequence recognition. In *CVPR*. 6538–6547.

[36] Liang Xu, Yu Tong, Qianqian Dong, Yixuan Liao, Cong Yu, Yin Tian, Weitang Liu, Lu Li, Caiquan Liu, and Xuanwei Zhang. 2020. CLUENER2020: Fine-grained Named Entity Recognition Dataset and Benchmark for Chinese. *arXiv preprint* (2020).

[37] Mingkun Yang, Yushuo Guan, Minghui Liao, Xin He, Kaigui Bian, Song Bai, Cong Yao, and Xiang Bai. 2019. Symmetry-constrained rectification network for scene text recognition. In *ICCV*. 9146–9155.

[38] Deli Yu, Xuan Li, Chengquan Zhang, Tao Liu, Junyu Han, Jingtuo Liu, and Errui Ding. 2020. Towards Accurate Scene Text Recognition With Semantic Reasoning Networks. In *CVPR*. 12110–12119.

[39] Wenwen Yu, Ning Lu, Xianbiao Qi, Ping Gong, and Rong Xiao. 2020. PICK: Processing key information extraction from documents using improved graph learning-convolutional networks. In *ICPR*. 4363–4370.

[40] Xiaoyu Yue, Zhanghui Kuang, Chenhao Lin, Hongbin Sun, and Wayne Zhang. 2020. RobustScanner: Dynamically Enhancing Positional Clues for Robust Text Recognition. In *ECCV*. 135–151.

[41] Xiaoyu Yue, Zhanghui Kuang, and Wayne Zhang. 2021. SegOCR: Simple Baseline. In *Unpublished Manuscript*.

[42] Xiaoyu Yue, Zhanghui Kuang, Zhaoyang Zhang, Zhenfang Chen, Pan He, Yu Qiao, and Wei Zhang. 2018. Boosting up Scene Text Detectors with Guided CNN. In *BMVC*.

[43] Shi-Xue Zhang, Xiaobin Zhu, Jie-Bo Hou, Chang Liu, Chun Yang, Hongfa Wang, and Xu-Cheng Yin. 2020. Deep Relational Reasoning Graph Network for Arbitrary Shape Text Detection. In *CVPR*. 9696–9705.

[44] Suncong Zheng, Feng Wang, Hongyun Bao, Yuexing Hao, Peng Zhou, and Bo Xu. 2017. Joint Extraction of Entities and Relations Based on a Novel Tagging Scheme. In *ACL*. 1227–1236.

[45] Xinyu Zhou, Cong Yao, He Wen, Yuzhi Wang, Shuchang Zhou, Weiran He, and Jiajun Liang. 2017. EAST: An Efficient and Accurate Scene Text Detector. *CVPR* (2017), 2642–2651.

[46] Yiqin Zhu, Jianyong Chen, Lingyu Liang, Zhuanghui Kuang, Lianwen Jin, and Wayne Zhang. 2021. Fourier Contour Embedding for Arbitrary-Shaped Text Detection. In *CVPR*.