manuscript ID:FCS-XXXXX

Label Distribution Learning for Scene Text Detection

Haoyu MA, Ningning LU, Junjun MEI, Tao GUAN, Yu ZHANG, Xin GENG

Frontiers of Computer Science, DOI: 10.1007/s11704-022-1446-5

Problems & Ideas

- Problems of conventional approaches for scene text detection:
  - Traditional regression-based approaches can only detect the regular text instances.
  - The post-processing for existing Segmentation-based model is complicated, which restricts the speed for inference.
- Ideas: We propose a novel scene text detection algorithm, which combines label distribution learning and label shrinkage via multi-task learning.

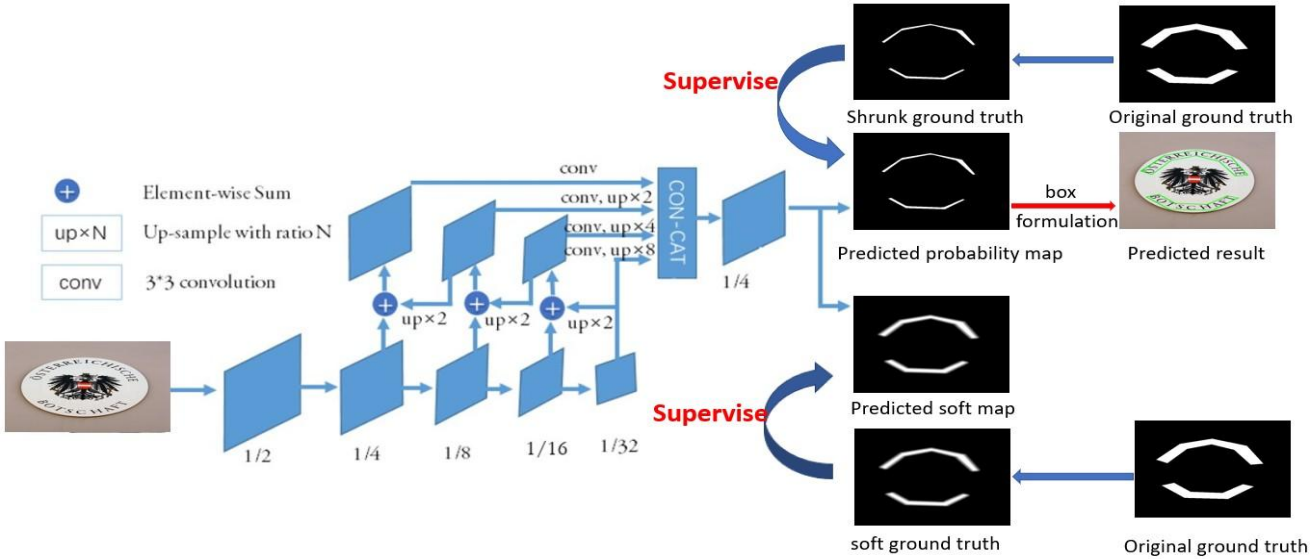

Illustration of our pipeline. Soft ground truth is the core component for our model, with each pixel value representing distribution of 0-1 instead of discrete label representing inside text region or no

Main Contributions

- Contributions:
  - We proposed to use label distribution learning for scene text detection.
  - Our proposed method can be easily combined with existing segmentation-based algorithms to improve their performance.
  - Compared with traditional methods, our proposed method can achieve comparable or better performance on three benchmark datasets.

Table 6 Detection results on the MSRA-TD500 dataset.

Method	Precision	Recall	F-measure
He et al. [44]	71	61	69
DeepReg [45]	77	70	74
RRPN [46]	82	68	74
RRD [42]	87	73	79
MCN [43]	88	79	83
PixelLink [11]	83	73.2	77.8
Corner [41]	87.6	76.2	81.5
TextSnake [12]	83.2	73.9	78.3
Xue et al. [47]	83.0	77.4	80.1
MSR [48]	87.4	76.7	81.7
CRAFT [36]	88.2	78.2	82.9
SAE [38]	84.2	81.7	82.9
DB-ResNet50(736) [14]	91.5	79.2	84.9
Ours-ResNet50(736, so)	91.4	82.4	86.7
Ours-ResNet50(736)	**95.4**	**86.2**	**90.6**

Table 7 Comparison over inference speed on the MSRA-TD500

Method	FPS
DeepReg [45]	1.1
RRD [42]	10
PixelLink [11]	3
Corner [41]	5.7
TextSnake [12]	1.1
CRAFT [36]	8.6
DB-ResNet18(736) [14]	62
DB-ResNet50(736) [14]	32
Ours-ResNet18(736)	60
Ours-ResNet50(736)	32

Table 6 and 7 show the performance and speed of our model in conventional dataset MSRA TD500. From these two tables, we can find that our model can achieve both better performance and faster speed compared to previous approaches.