



Self-Training for Domain Adaptive Scene Text Detection

Yudi Chen^{*1,2}, Wei Wang^{*1,2}, Yu Zhou¹, Fei Yang³, Dongbao Yang¹, Weiping Wang¹ ¹ Institute of Information Engineering, Chinese Academy of Sciences ²University of Chinese Academy of Sciences ³TAL Education Group

*Equally contributed

Wei Wang Date: 15/01/2021

Domain Trouble on Text Detection



- Background
- Language
- Font
- Shape

Related Work

- Weak/semi-supervised methods
 - Still need label
- Data generation methods
 - Not realistic enough
 - Not complicated enough
- Domain adaption methods
 - GA-DAN: Need box annotation[Zhan et al. ICCV'19]
 - Domain adaptation for general object detection
 Not work well on text





SynthText [Gupta et al. CVPR'16] SynthText3D [Liao et al. SCIS]

Self-Training for Domain Adaptive



Detector

Self-Training for Domain Adaptive



Self-Training for Domain Adaptive



Target Domain





Results from previous frame

Results for next frame

Text Mining Module

Detection Results



Text Mining Module

Detection Results



When No Video Available



Ablation study

| Train set | Test set | Video | TMM | Balance loss | Precision | Recall | F-measure |
|-----------|----------|-------|-----------------------|-----------------------|-----------|--------|-----------|
| VISDT | IC15 | × | X | × | 54.8 | 57.7 | 56.2 |
| | | ~ | × | × | 67.7 | 52.1 | 58.9 |
| | | ~ | ✓ | × | 62.3 | 61.6 | 61.9 |
| | | ~ | ~ | ~ | 64.3 | 61.7 | 63.0 |
| IC15 | IC15 | × | X | × | 83.0 | 80.4 | 81.7 |
| | | ~ | × | × | 83.9 | 81.1 | 82.5 |
| | | ~ | ✓ | × | 85.4 | 81.3 | 83.3 |
| | | ~ | ~ | ~ | 87.7 | 80.3 | 83.8 |
| VISDT | 15VID | × | X | × | 49.8 | 48.0 | 48.9 |
| | | ~ | × | × | 55.7 | 49.3 | 52.3 |
| | | ~ | ✓ | × | 56.3 | 54.0 | 55.1 |
| | | ~ | ~ | ~ | 60.9 | 53.6 | 57.0 |
| IC15 | 15VID | × | X | × | 63.5 | 60.2 | 61.8 |
| | | ~ | × | × | 64.1 | 60.7 | 62.4 |
| | | ~ | ✓ | × | 65.4 | 60.5 | 62.9 |
| | | ✓ | ✓ | ✓ | 65.4 | 61.9 | 63.6 |

Baseline detector: Mask RCNN

TMM

works

Comparison with data augmentation

| Method | Precision | Recall | F-measure |
|---------------------|-------------|--------|-----------|
| Mask R-CNN | 83.0 | 80.4 | 81.7 |
| Mask R-CNN+Ours | 87.7 | 80.3 | 83.8 |
| Mask R-CNN+Aug | 86.1 | 84.1 | 85.1 |
| Mask R-CNN+Aug+Ours | 89.8 | 82.5 | 86.0 |

Comparison with other video-lacking strategies

| Method | Precision | Recall | F-measure |
|--------------|-----------|--------|-----------|
| None | 54.8 | 57.7 | 56.2 |
| Base | 62.7 | 58.9 | 60.7 |
| Base-Trans | 64.1 | 59.1 | 61.5 |
| Gen-Straight | 64.4 | 60.0 | 62.1 |
| Gen-Loop | 66.9 | 64.7 | 65.8 |

Comparison with cross domain and data generation methods

| Method | Р | R | F |
|--------------------|------|------|------|
| Synthtext3d [19] | 64.5 | 56.7 | 60.3 |
| GA-DAN [22][AD] | 69.9 | 59.6 | 64.4 |
| GA-DAN [22][10-AD] | 67.3 | 71.6 | 69.4 |
| VISDT | 54.8 | 57.7 | 56.2 |
| VISDT+15VID | 64.3 | 61.7 | 63.0 |
| VISDT+15VID-2 | 70.5 | 64.7 | 67.5 |
| Synthtext3d [19] | 86.6 | 79.2 | 82.7 |
| GA-DAN [22][AD] | 83.7 | 79.2 | 81.4 |
| GA-DAN [22][10-AD] | 85.6 | 81.6 | 83.5 |
| Target | 80.3 | 81.7 | 81.0 |
| VISDT→Target | 83.0 | 82.2 | 82.6 |
| 15VID→Target | 86.9 | 81.7 | 84.2 |

Comparison with SoTA on ICDAR2015

| Method | Ext. | Р | R | F |
|--------------------------------|-----------------------|------|------|------|
| PAN [32] | - | 82.9 | 77.8 | 80.3 |
| PSENet [11] | - | 81.5 | 79.7 | 80.6 |
| Synthtext3d [19] | - | 86.6 | 79.2 | 82.7 |
| GA-DAN [22] | - | 85.6 | 81.6 | 83.5 |
| Ours[IC15] | - | 80.3 | 81.7 | 81.0 |
| Ours[15VID+IC15] | - | 85.4 | 81.3 | 83.3 |
| Ours[15VID→IC15] | - | 86.9 | 81.7 | 84.2 |
| MSR [33] | ~ | 86.6 | 78.4 | 82.3 |
| Mask TextSpotter [34] | ~ | 85.8 | 81.2 | 83.4 |
| FOTS [35] | ✓ | 88.8 | 82.0 | 85.3 |
| PSENet-1s [11] | ~ | 86.9 | 84.5 | 85.7 |
| BDN [36] | ~ | 89.4 | 83.8 | 86.5 |
| SPCNet [37] | ~ | 88.7 | 85.8 | 87.2 |
| LOMO [38] | ~ | 91.3 | 83.5 | 87.2 |
| GNNets [39] | ✓ | 90.4 | 86.7 | 88.5 |
| Ours[IC15] | ✓ | 88.0 | 83.9 | 85.9 |
| $Ours[15VID \rightarrow IC15]$ | ✓ | 88.3 | 85.7 | 87.0 |
| Ours[15VID→IC15]+AUG | ✓ | 91.2 | 85.4 | 88.2 |

13

Before self-training

After self-training







THANKS