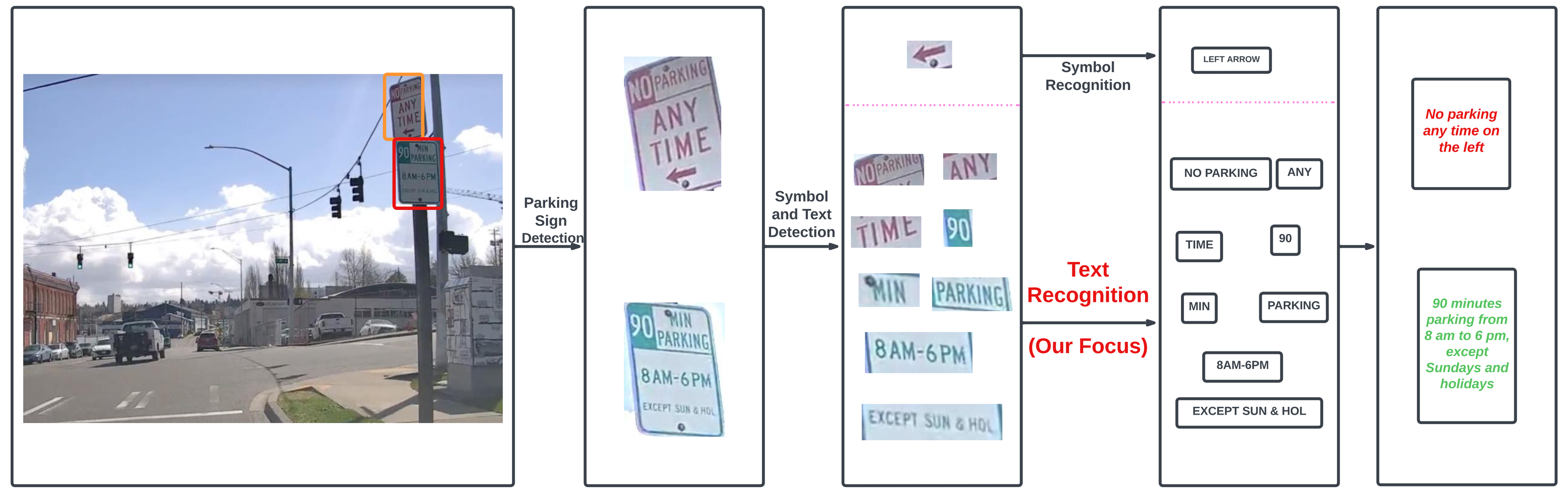
BMVC 2023

Dictionary-Guided Text Recognition for Smart Street Parking

Deyang Zhong, *Jiayu Li, Wei Cheng, and Juhua Hu*
School of Engineering and Technology, University of Washington, Tacoma, USA
{dyzhong, jiayu7, uwcheng, juhuah} @uw.edu

Motivation

- Current autonomous driving model
  - Relying on detecting and analyzing road conditions
  - In a too-simplified way, can result in parking tickets. For instance, 'Friday-Sunday' is totally different from 'Friday, Sunday'
  - Parnia [1] suggested a method for detecting and classifying parking signs. Jiang [2] proposed a comprehensive parking system. Both faced challenges with complex text including abbreviations and punctuation
- Our smart parking framework
  - The figure represents the smart street parking pipeline, from image capture and cropping to identifying and recognition.
  - This work focuses on improving text recognition accuracy
    - Making it capable of identifying complex text even with punctuation



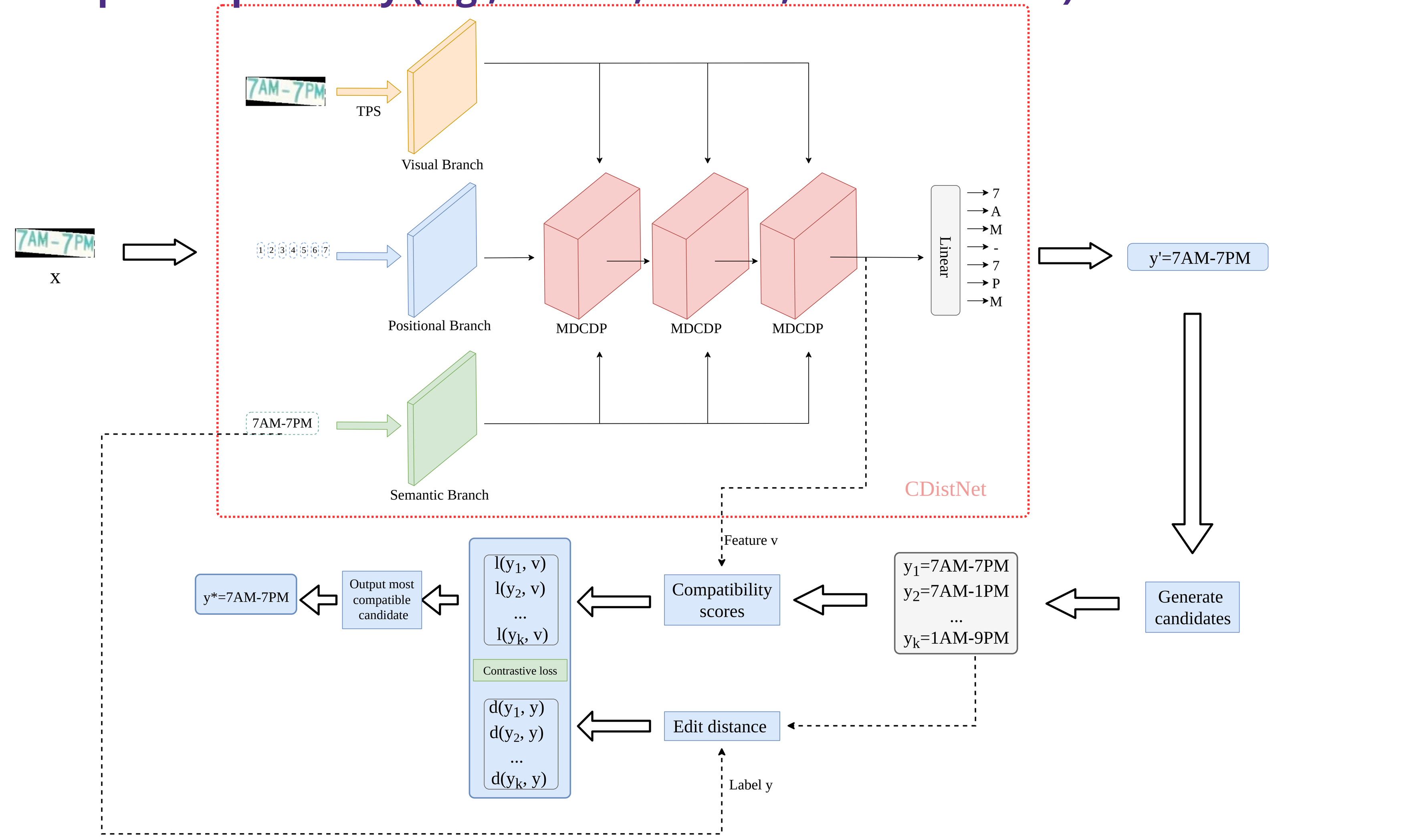
Smart Curbside Parking System

Text Recog Background

- Scene text recognition
  - Traditional methods:
    - Manually construct text features by observations/statistical models
    - Histograms of oriented gradient used by Wang et al.[3]
    - Multiscale representation proposed by Yao[4].
    - Problems especially concerning accuracy and automation
  - Deep learning methods:
    - Segmentation-based approaches
      - Create a character segmentation, identify each character separately, and gather them into one text line
      - Ask for high requirements of the character detection model
    - The segmentation-free approaches
      - Regard cropped word images as a whole
      - The attention-based method is popular in the prediction stage
      - One new method CDistNet[5] achieves great performance
- Parking text recognition
  - Irshad [6] introduced a novel framework for parking sign recognition
  - Li [7] directly deployed CRNN model in the smart parking system
  - Both failed to process compound signs and achieve great performance

Our Method

- Introducing a dictionary method [8] to increase the probability of success based on the CDistNet structure
- Adding special parking-related words and punctuations into the dictionary to improve specificity (e.g., "9 AM", "Mon.", "9AM-4:30PM")



- Step 1: Input a text image x to CDistNet and obtain preliminary result y', then generate k candidates according to y' from the dictionary
- Step 2: Compute compatibility scores $l(y_i, v) = \sum_{j=1}^{len(y_i)} \log(P[y_i^j, j])$ and edit distances $d(y_i, y')$. P is the probability matrix of size $s \times m$, where m is the max length of a word and s is the size of alphabet. y_i^j represents the index of the jth character of y_i: $y_i^j \in \{1, 2, \dots, s\}$.
- Step 3: Compute KL-divergence between two probability distributions D and L: $KL(D||L) \propto -\sum_{i=1}^{k} D_i \log(L_i)$, $L_i = \frac{exp(-l(y_i,v))}{\sum_{j=1}^{k} exp(-l(y_j,v))}$, $D_i = \frac{exp(-d(y_i,y)/T)}{\sum_{j=1}^{k} exp(-d(y_j,y)/T)}$.
- Step 4: Compute training loss is $\mathcal{L}(x, y) = l(y', v) + \lambda KL(D||L)$, and output most compatible candidate y^*

Experiments

- Datasets: SynthText [9] is a synthetic text image dataset where the word image contains punctuations. The second dataset is from our curbside parking sign images. The sizes of these two datasets are separately 7000k and 9k. The training set is SynthText plus 80% of our parking set, and the test set is the remaining 20% parking set.
- Baselines:
  - CRNN [10]: an established deep neural network that has demonstrated competitive- ness in recognizing sequence-like images
  - CDistNet: a recently proposed state-of-the-art architecture specifically designed for text recognition tasks
  - E-CDistNet: an edited model based on CDistNet as a variant baseline to address the limitation to handle punctuation marks
  - TrOCR [11]: a recently proposed OCR method that has gained recognition as an exceptional model for text recognition

Table 1: Parking Sign Text Recognition Performance Comparison (ER: Error Rate).

Model	CRNN	CDistNet	E-CDistNet	TrOCR	Our Model
ER (%)	13.3	16.7	10.0	7.4	**4.3**
Repairable ER (%)	4.8	5.3	3.7	5.4	**3.2**
Irreparable ER (%)	8.5	11.4	6.3	2.0	**1.1**
CER (%)	10.1	13.2	7.5	3.8	**2.2**

Table 2: Recognition Examples.

Images					
CRNN	STOPPING	700AM	9AM-4PM	7AM;8:3OPM	SAT-SUN:HOL
CDistNet	STOPPING	700AM	9AM4PM	7AN830PM	SATSUNHOL
E-CDistNet	STOPPING	7:00AM	9AM-4PM	7AN-8:30PM	SAT-SUN.-HOL
TrOCR	STOPPING	7:00AM	9AM-4PM	7AM-8:30PM	SAT-SUN.-HOL
Our Model	STOPPING	7:00AM	9AM-4PM	7AM-8:30PM	SAT.-SUN.-HOL.

From Table 1 and Table 2, we find that:

- The original CDistNet model yields the highest error rate among the evaluated models due to its inability to handle punctuation
- The CRNN model has the capability of punctuation but still exhibits poor performance
- E-CDistNet model shows significant improvement and surpasses the CRNN model
- The TrOCR model, despite not being specifically designed for this task, exhibits remarkable performance while evaluating our parking text test data.
- Our model, which builds a parking-specific dictionary as an additional text recognition guidance, demonstrates significantly better performance across all metrics.

- Ablation study
  - To show the effect of the number of candidates to consider, we change k in {3,5,8,10,15} and show the results in Table 3

Table 3: Effect of k.

k	3	5	8	10	15
ER(%)	7.8	4.9	4.5	4.3	4.3
Repairable ER(%)	4.9	3.6	3.4	3.2	3.2
Irreparable ER(%)	2.9	1.3	1.1	1.1	1.1
CER	4.5	2.5	**2.2**	**2.2**	**2.2**

Conclusions

- We propose an innovative model for recognizing parking text that outperforms other methods, particularly in the following areas:
  - Collecting a dataset for street parking sign text recognition
  - Overcoming punctuation recognition challenges through well-designed strategies and building a parking-specific dictionary
  - Significantly increasing accuracy for text recognition, especially for parking-specific words, like abbreviations of time words

References

[1] Parnia Haji Faraji, Hamid Reza Tohidypour, Yixiao Wang, Panos Nasiopoulos, Simon Ren, Arash Rizvi, Cloris Feng, Mahsa T Pourazad, and Victor CM Leung. Deep learn- ing based street parking sign detection and classification for smart cities. In *Proceed- ings of the Conference on Information Technology for Social Good*, pages 254–258, 2021.
[2] Zhongyu Jiang. Street parking sign detection, recognition and trust system. Master's thesis, University of Washington, 2019.
[3] Kai Wang and Serge Belongie. Word spotting in the wild. In *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part I 11*, pages 591–604. Springer, 2010.
[4] Cong Yao, Xiang Bai, Baoguang Shi,and WenyuLiu. Strokelets: A learned multi-scale representation for scene text recognition. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 4042–4049, 2014. doi: 10.1109/CVPR.2014.515.
[5] Tianlun Zheng, Zhineng Chen, Shancheng Fang, Hongtao Xie, and Yu-Gang Jiang. Cdistnet: Perceiving multi-domain character distance for robust text recognition. *International Journal of Computer Vision*, pages 1–19, 2023.
[6] Humayun Irshad, Qazaleh Mirsharif, and Jennifer Prendki. Crowd sourcing based active learning approach for parking sign recognition. *arXiv preprint arXiv:1812.01081*, 2018.
[7] Jiayu Li. An algorithm for street parking sign rule generation. Master's thesis, University of Washington, 2020.
[8] Nguyen Nguyen, Thu Nguyen, Vinh Tran, Minh-Triet Tran, Thanh Duc Ngo, Thien Huu Nguyen, and Minh Hoai. Dictionary-guided scene text recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7383–7392, 2021.
[9] Ankush Gupta, Andrea Vedaldi, and Andrew Zisserman. Synthetic data for text localization in natural images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2315–2324, 2016.
[10] Baoguang Shi, Xiang Bai, and Cong Yao. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE transactions on pattern analysis and machine intelligence*, 39(11):2298–2304, 2016.
[11] Minghao Li, Tengchao Lv, Jingye Chen, Lei Cui, Yijuan Lu, Dinei Florencio, Cha Zhang, Zhoujun Li, and Furu Wei. Trocr: Transformer-based optical character recognition with pre-trained models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 13094–13102, 2023.