

论文题目

Sequence-to-Sequence Domain Adaptation Network for Robust Text Image Recognition

论文作者

Yaping Zhang, Shuai Nie¹, Wenju Liu^{1*}, Xing Xu, Dongxiang Zhang, Heng Tao Shen
Chinese Academy of Sciences (CASIA)
University of Chinese Academy of Sciences (UCAS)
University of Electronic Science and Technology of China
Zhejiang University

论文时间

2019 年 6 月

论文的前提

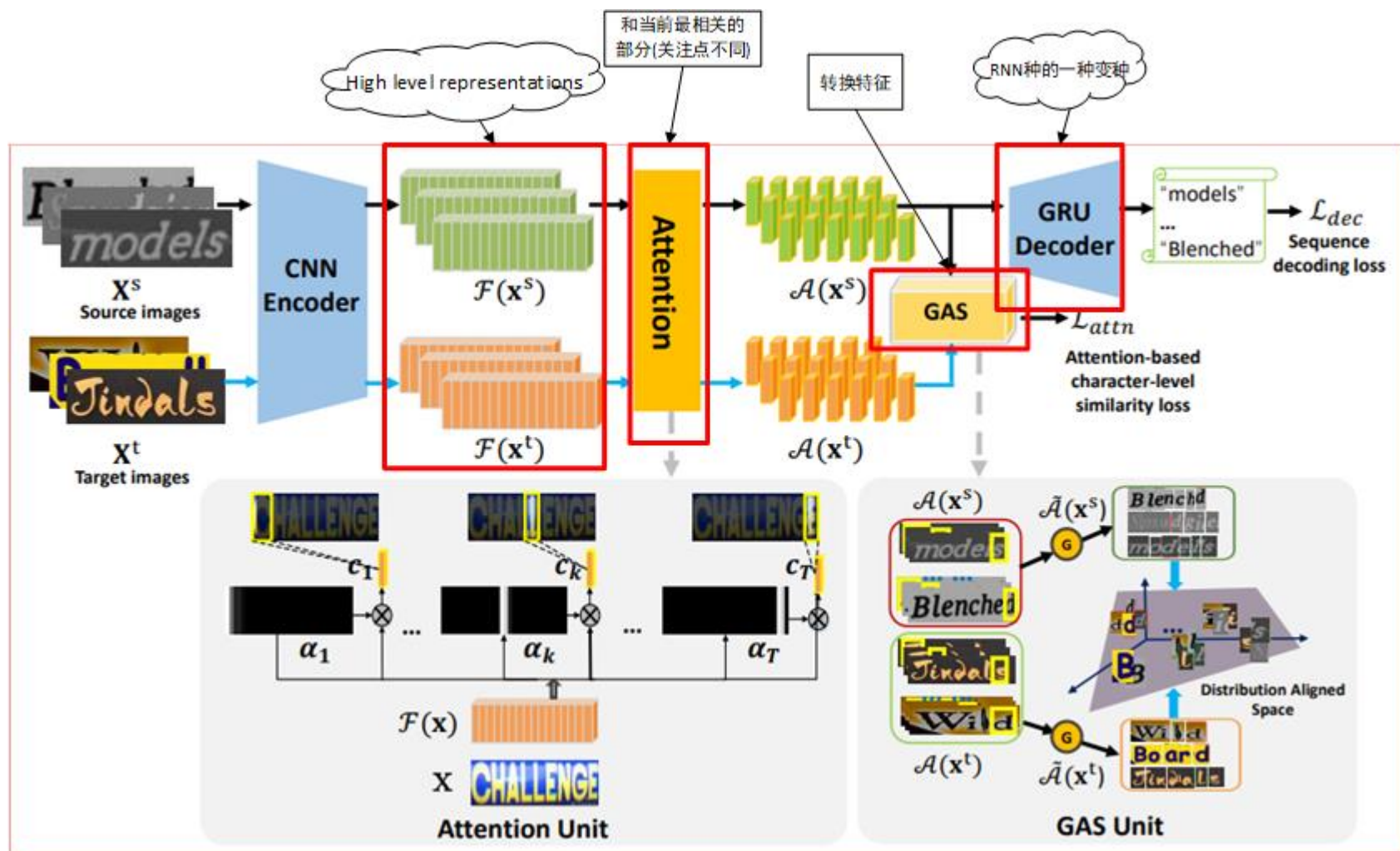
构建一个包括自然界的文字,手写文本,数学表达式的 robust 的 text recognizer
seq2seq 广泛用于语音处理的 encoder-decoder structural

	Scene text	Handwritten text	Mathematical expression text
Training (Source)			
Test (Target)			
Domain shift (Type)	Synthetic/Realistic	Writing style	Structure diversity

论文观点

使用 seq2seq 的方法实现 source domain 到 target domain 的自适应 shift

论文做法



1. encoder

image → CNN(backbone:DenseNet) → high level representations

high level representations → Attention(bi-LSTM hidden:256) → 时刻不同关注的 encoder 的特征不一样(和要 decode 的字符最相关的部分)

features → GAS → 转换变长成一定长度的 character features

2. decoder

character features → GRU(hidden:512) → 文字

最终效果

Model	IIIT5K	SVT	IC-03	IC-13
ANN [16]	—	71.7	89.6	81.8
STAR-Net [22]	83.3	83.6	89.9	89.1
R^2AM [20]	78.4	80.7	88.7	90.0
CRNN [31]	81.2	82.7	91.9	89.6
RARE [32]	81.9	81.9	90.1	88.6
Ghosh et al [12]	—	75.1	89.3	—
Gao et al [10]	81.8	82.7	89.2	88.0
ASTER [33]	83.2	87.6	92.4	89.7
Char-Net [21]	83.6	84.4	91.5	90.8
<i>SSDAN-base</i>	81.1	82.1	91.2	91.0
<i>SSDAN</i>	83.8	84.5	92.1	91.8

Method	WER	CER	Average
bluche2015deep [2]	24.7	7.3	16.00
bluche2016joint [3]	24.6	7.9	16.25
sueiras2018offline [34]	23.8	8.8	16.30
<i>SSDAN-base</i>	23.9	9.2	16.55
<i>SSDAN</i>	22.2	8.5	15.35

Method	ExpRate
I [28]	37.2
VI [28]	25.7
VII [28]	26.1
WYCIWYS [8]	28.7
Le et al [19]	35.2
IM2TEX [7]	38.7
<i>SSDAN-base</i>	39.9
<i>SSDAN</i>	41.6

结论

the fine-grained character-level knowledge transfer between the source and target sequence data is more effective