ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021
"putting Artificial Intelligence to work on patterns"

Cross-Lingual Text Image Recognition via Multi-Task Sequence to Sequence Learning

Zhuo Chen[1,2], Fei Yin[1,2], Xu-Yao Zhang[1,2], Qing Yang[1,2], Cheng-Lin Liu[1,2,3]

[1]National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences
[3]CAS Center for Excellence in Brain Science and Intelligence Technology

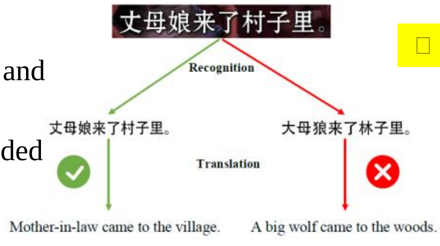
Introduction

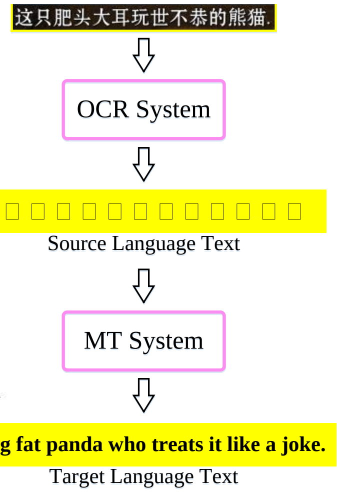
Background

People traveling to or living in foreign countries often have trouble in recognizing foreign texts in natural scene. Thus, transforming text images to understandable information automatically has become an intense demand.

Cross-Lingual Text Image Recognition

- CLTIR: Recognizing texts in a source language and translating into a target language
- All schemes for handling this problem are cascaded
- Apart system is potentially error prone





Contribution

- We raise a new problem called CLTIR
- We propose a novel end-to-end multi-task system with two different sequence to sequence learning methods
- The proposed framework achieves promising results on the dataset of movie subtitle images

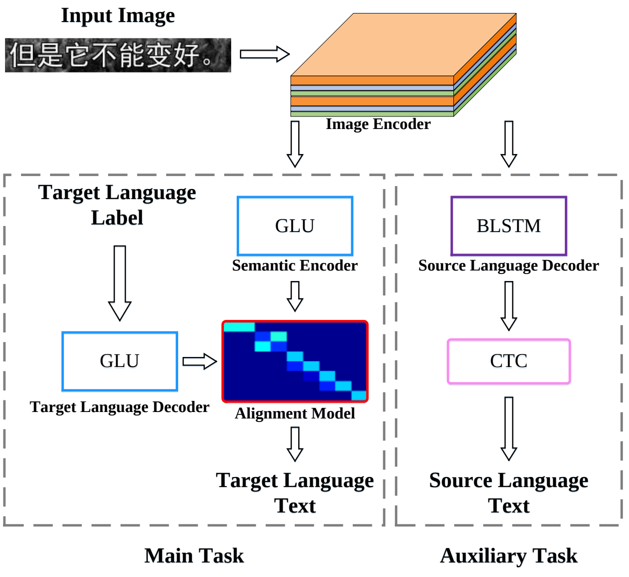
Methods

Multi-task learning

- Main task: cross-lingual text image recognition
- Auxiliary task: mono-lingual text image recognition

Sequence to sequence learning

- Attention based
- BLSTM + CTC



Experiments

Dataset

There was no existing text recognition dataset with label in other language. We finally get English-Chinese bilingual subtitles from 50 English animated films.

Sets	Training	Validation	Total
#Movies	45	5	50
#Corpus	51,614	3,666	55,280
#En Vocab	13,608	1,679	13,823
#Zh Vocab	3,488	2,382	3,501
#Samples	450,000	50,000	500,000

Results on the CLTIR dataset

Compared with cascade system and single-task system, the multi-task system is better on both source and target language recognition task.

System	Accuracy	BLEU
Recognition Model	98.05	—
Translation Model	—	43.73
Cascade System	—	41.84
Single-Task	—	38.64
Multi-Task	**99.13**	**42.91** ↑ 1.07

Ablation study of encoder architecture

Architecture	BLEU
VGG-19 [21]	42.91
ResNet-18 [22]	43.00
DenseNet-29 [23]	43.42