MENTOR: Multilingual tExt detectioN TOward leaRning by analogy

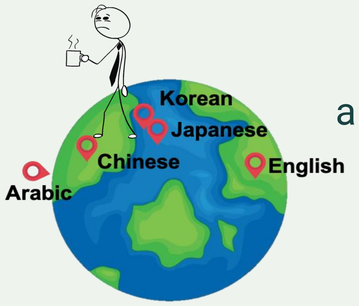
Hsin-Ju Lin, Tsu-Chun Chung, Ching-Chun Hsiao, Pin-Yu Chen, Wei-Chen Chiu, and Ching-Chun Huang*

IBM research, National Yang Ming Chiao Tung University*

OCTOBER 1 - 5, 2023

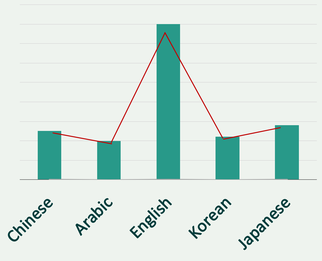
IEEE/RSJ International Conference on Intelligent Robots and Systems

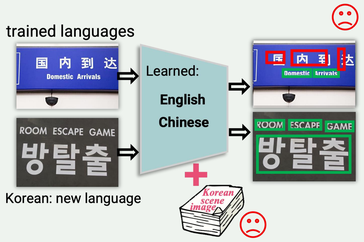
Challenges



a. Collecting universal text dataset is time-consuming and resource-intensive

b. Text training datasets that are imbalanced between languages





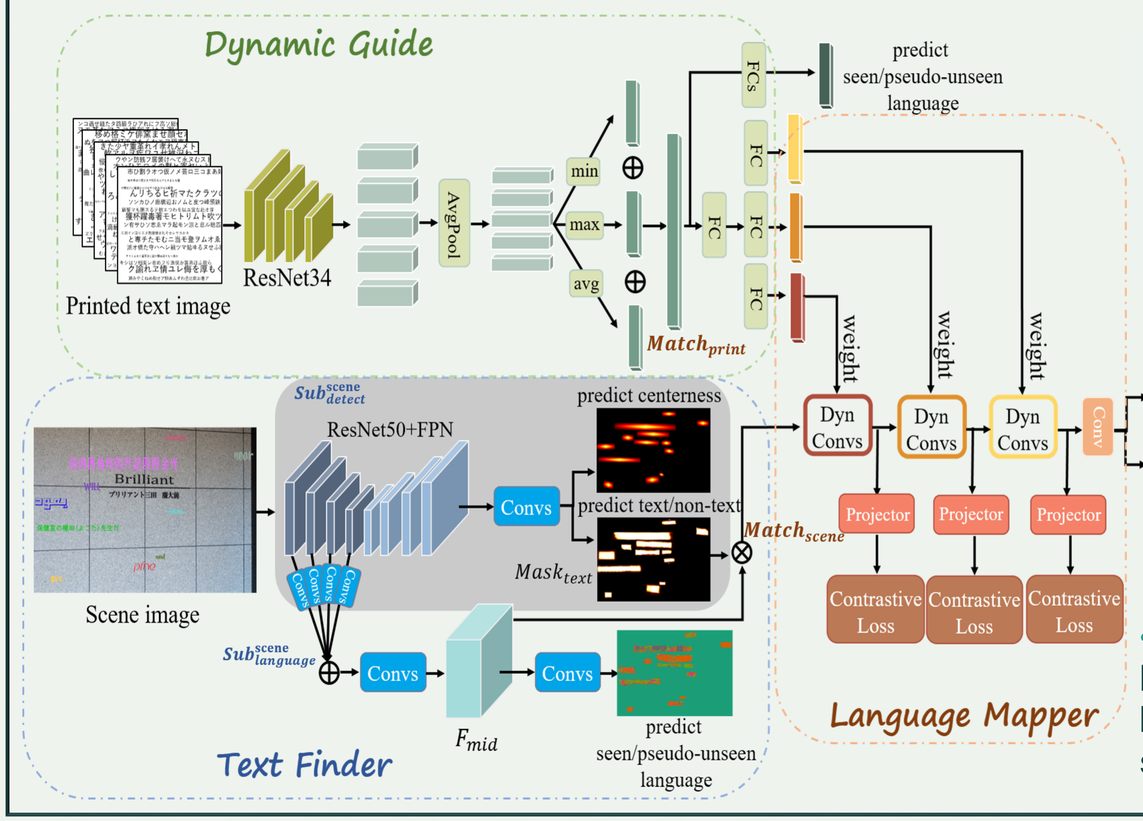
c. The detector must be re-trained or fine-tune to detect new languages

New Problem Setting

A generalizable multilingual text detector

- **SHOULD** detect both seen and unseen language regions
- **NO NEED** to collect supervised training data for unseen language
- **NO NEED** model re-training for unseen language

Proposed Method



- **Dynamic Guide :** generate language representational attributes from printed text images
- **Text Finder :** obtain differentiable features to identify text regions
- **Language Mapper :** learn the corresponding relationship between side information and the scene text feature

Experimental Result

Evaluating F1-score on MLT17 and Malayalam in IIIT-ILST dataset

Method	English	Arabic	Bangla	Chinese	Japanese	Korean	Malayalam
E2E-MLT	55.43	55.431	3.027	50.594	12.9	32.715	x
MultiplexedOCR	83.284	80.074	78.104	56.251	70.986	67.862	11.932
(a) Ours	84.031	80.952	81.76	83.585	76.896	72.479	65.837
(b) Ours	82.527	82.092	80.909	69.046	83.51	84.013	55.895
(c) Ours	81.778	54.646	82.134	82.554	80.737	83.143	43.992

Arabic (seen)

Chinese (unseen)

SCAN ME for more detail