

Towards Scene Text Translation for Complex Writing Systems



Hour Kaing¹, Haiyue Song¹, Chenchen Ding¹, Jiannan Mao², Hideki Tanaka¹, Masao Utiyama¹

¹National Institute of Information and Communications Technology

²Gifu University

Task

Scene with Text in English



Translation
with our system

Scene with Text in Japanese



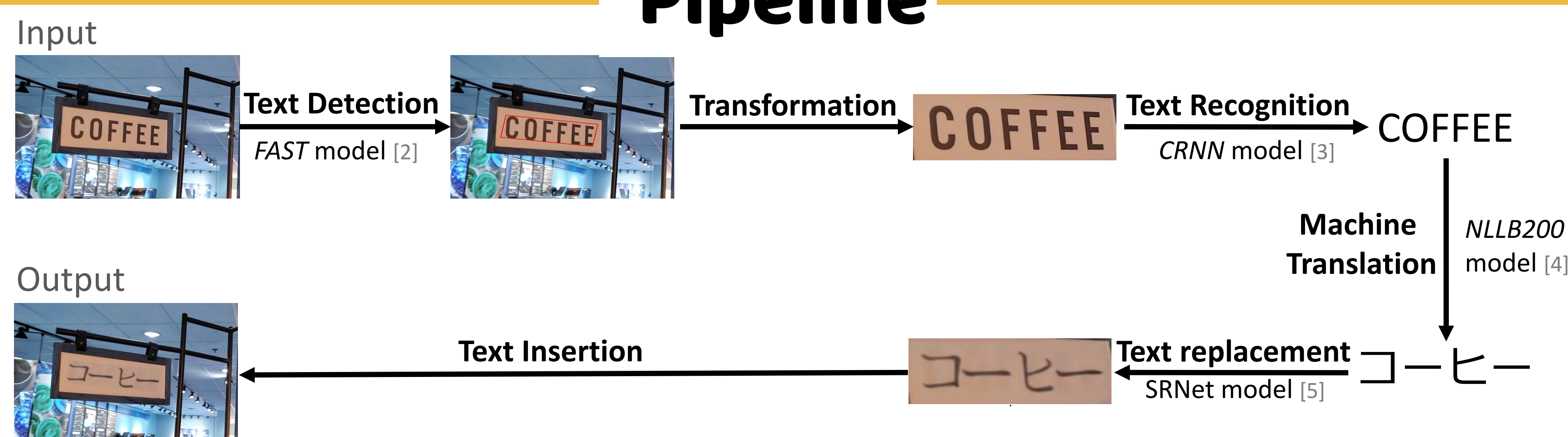
Challenges

- Visual features
 - location, distortion, font, background
- Japanese writing system
 - 漢字, ひらがな, カタカナ

Ideas

- Using existing pipeline [1], which is for simple writing systems
- Synthesize English-Japanese scenes to fine-tune the **text placement module** in it

Pipeline



Synthetic Data

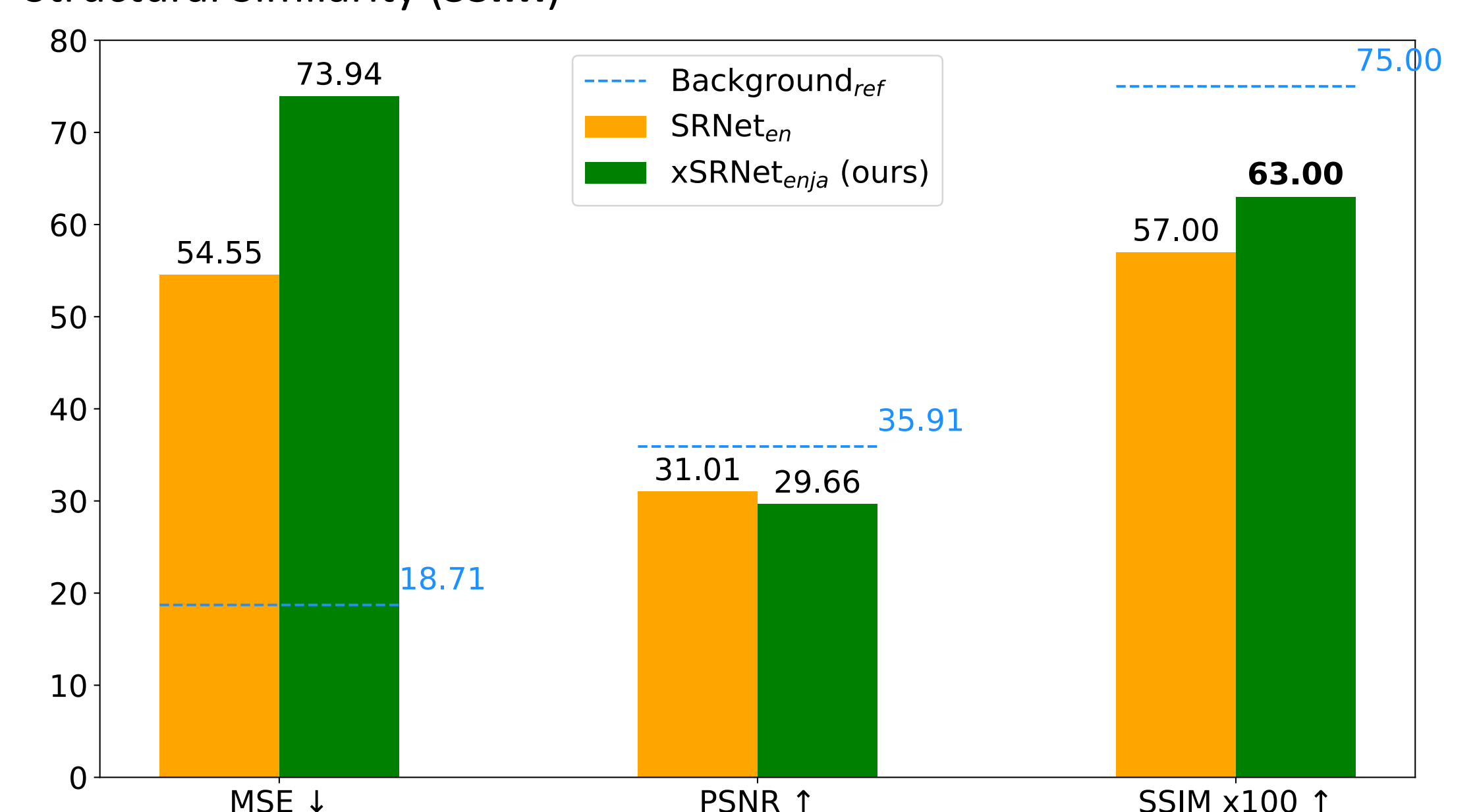
1. Random text
 - Ja/En word (Edictionary), visual features
2. Random background

■ Background		
■ Text skeleton		
■ Foreground text		
■ Target image (Ja)		
■ Source image (En)		

Results

Metrics

- Mean Squared Error (MSE) on pixels
- Peak Signal-to-Noise Ratio (PSNR)
- Structural Similarity (SSIM)



Samples

Outputs of the text replacement module

Input Image					
Ground Truth					
SRNet _{en}					
xSRNet _{enja}					

Outputs of the whole pipeline

Input Image					
Translated Image					

Future Directions

- Scene-context-aware translation
- Video-level pipeline
- Text-aware evaluation

References

- [1] Show me the world in my language: Establishing the first baseline for scene-text to scene-text translation. 2024.
- [2] Fast: Faster arbitrarily-shaped text detector with minimalist kernel representation. 2021.
- [3] An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. 2016.
- [4] No language left behind: Scaling human-centered machine translation. 2022.
- [5] Editing text in the wild. 2019.