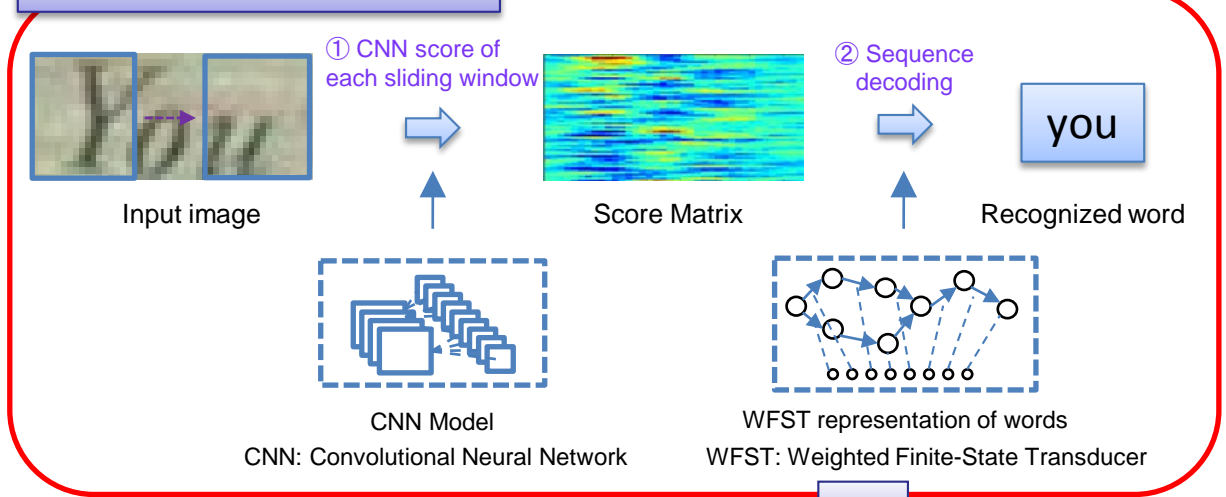


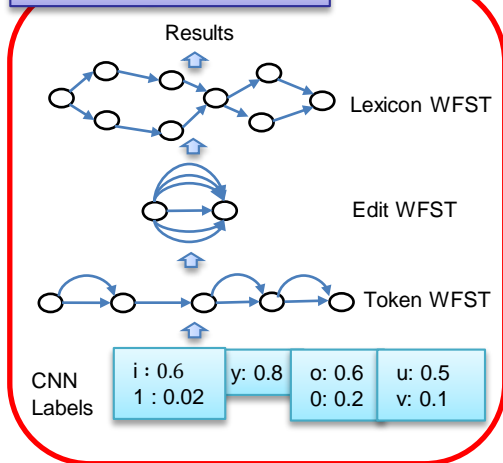
Abstract

Text in natural scene images usually contains a lot of semantic value and recognizing the texts is an important step for understanding the scene. Unlike the printed documents, text in a natural scene is more difficult due to large variations in geographical placement, backgrounds, textures, fonts, and illumination conditions. In this work, we propose a method which first detects and recognizes characters by utilizing the Convolutional Neural Network (CNN), and then decodes a series of recognized characters into words with a Weight Finite State Transducer (WFST). WFST has been successfully utilized in the speech recognition field, where it is shown that it can efficiently incorporate lexicon or high order language model in the word labelling tasks. In the experiments, we have shown that the proposed algorithm can robustly recognize words in the scene images from the public datasets such ICDAR 2003, and SVT-WORD.

Flowchart of the word recognition



Decoding with WFST



Results

Recognition accuracy on ICDAR and SVT-WORD dataset

Method \ Dataset (lexicon size)	ICDAR2003 (50)	ICDAR2003 (860)	SVT-WORD (50)
Jaderberg et al. [ECCV 2014]	96.2	91.5	86.1
Proposed	96.9	91.8	93.4

ICDAR: Intl. Conf. on Document Analysis and Recognition Dataset (860 images)

SVT-WORD: Street View Text Dataset (647 images)

[Reference]

[1] Xinhao Liu Takahito Kawanishi Xiaomeng Wu and Kunio Kashino, "Scene Text Recognition With High Performance CNN Classifier and Efficient Word Inference," in Proc. of International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2016.

[Contact]

Xinhao LIU Media Information Laboratory, Recognition Research Group
E-mail : liu.xinhao(at)lab.ntt.co.jp