

## 飾り文字でも読み取れる

～深層学習と系列デコーディングによる文字列認識～

### どんな研究

スマートフォンで撮影した街中の風景など、**実環境の情景中に存在する文字を読み取る**ことは、その有用性が期待されながら、今なお困難な課題です。本研究は、デザインされた飾り文字も含めて、情景中に存在する**様々な書体の文字列**を正確に読み取ることを目指しています。

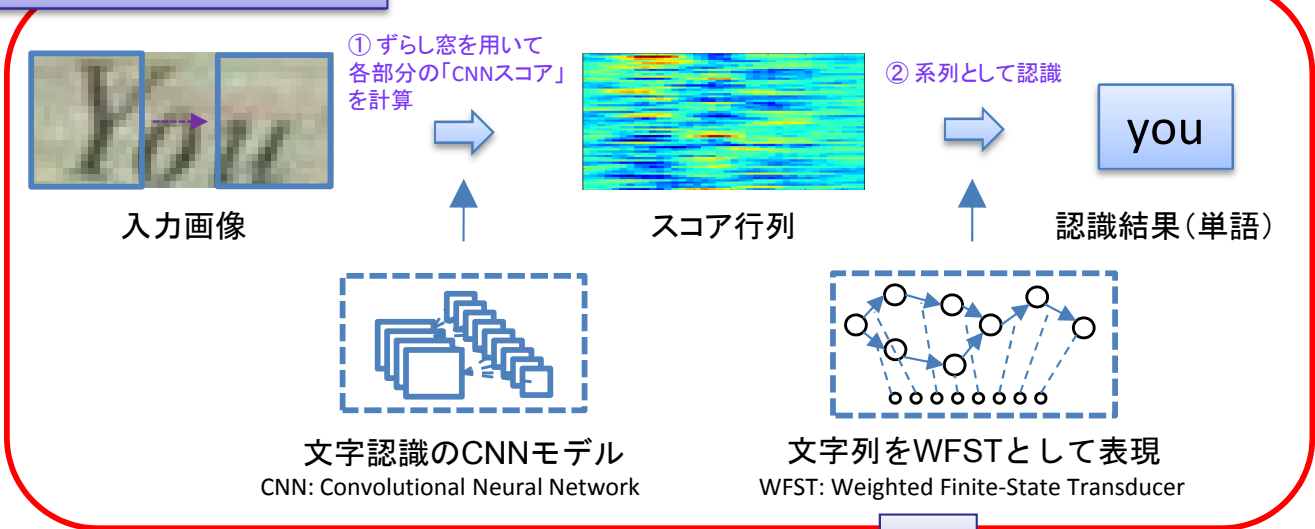
### どこが凄い

同じ文字でも書体や撮影条件などにより画像には大きな違いが生まれ、認識が困難になります。これに対し、**深層学習**と、音声認識で有効性が示されている**系列の認識手法であるWFST（重み付き有限状態トランスデューサ）**とを組み合わせる方法を提案し**高い認識精度を実現**しました。

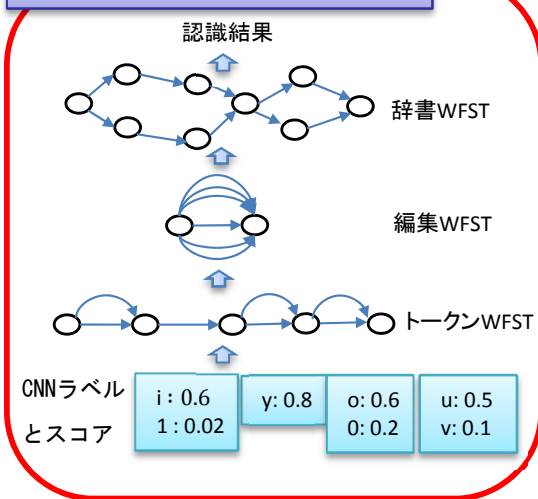
### 目指す未来

文字を読み取ることで、コンピュータが効率的に情報を取得することや、ユーザにとって必要な情報を的確に提示することが可能になります。効率的な画像探索や駅・店頭・市街地での**歩行者用ナビゲーション**、**車の自動運転**、**画面の自動翻訳**など、様々な応用の可能性が広がります。

### 文字列認識の流れ



### WFSTによる系列デコーディング



### 実験結果

画像データセットを用いた単語認識精度の評価

Dataset (辞書の単語数)	ICDAR2003 (50)	ICDAR2003 (860)	SVT-WORD (50)
Jaderberg et al. [ECCV 2014]	96.2	91.5	86.1
Proposed	<b>96.9</b>	<b>91.8</b>	<b>93.4</b>

ICDAR: Intl. Conf. on Document Analysis and Recognition Dataset (画像860枚)  
SVT-WORD: Street View Text Dataset (画像647枚)

### 【関連文献】

[1] X. Liu, T. Kawahishi, X. Wu, K. Kashino, "Scene text recognition with high performance CNN classifier and efficient word inference," in Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2016.

### 【連絡先】

劉新豪 (Xinhao LIU)   メディア情報研究部   メディア認識研究グループ  
E-mail : liu.xinhao(at)lab.ntt.co.jp