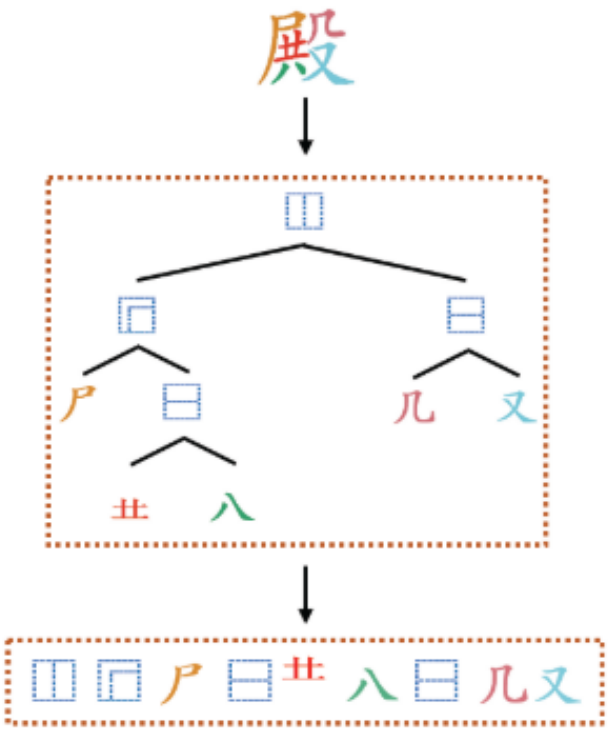
A Transformer-based Radical Analysis Network for Chinese Character Recognition

Chen Yang, Qing Wang, Jun Du, Jianshu Zhang, Changjie Wu, Jiaming Wang

NEL-SLIP Lab, University of Science and Technology of China,
Hefei, Anhui, P.R.China

Motivation

- Weakness of character-based methods
  - Don't consider the similarity and internal structures among different characters
  - Have difficulty in recognizing unseen Chinese characters

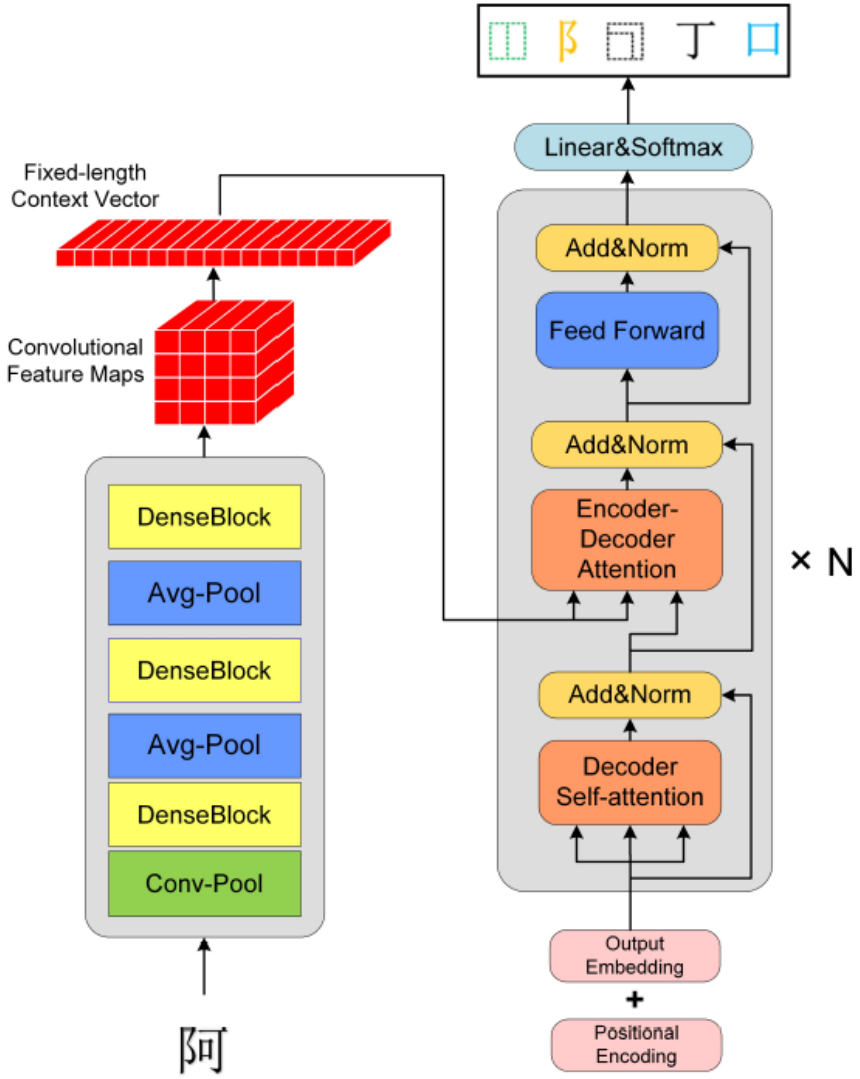


★ Hierarchical radical structure of an example Chinese character

Motivation

- Existing solutions
  - Radical-based methods decompose the Chinese character into a limited number of radicals
  - Have ability to learn a zero-shot problem
- Strengths of the Transformer architecture
  - Self-attention mechanism can capture long-range dependencies and the detailed internal pattern
  - The Transformer is composed of stacked blocks and aggregates the input context for each block, which naturally provides us with more hierarchical representations

Proposed Architecture



Transformer-based radical analysis network(**RTN**)

- Dense Encoder

$$\mathbf{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_L\}, \mathbf{a}_l \in \mathbb{R}^D$$

- Transformer Decoder

$$\mathbf{s}_i^n = \mathrm{MA}(\mathbf{c}_i^{n-1}, \mathbf{C}_{<i}^{n-1}, \mathbf{C}_{<i}^{n-1})$$

$$\mathbf{z}_i^n = \mathrm{MA}(\mathbf{s}_i^n, \mathbf{A}, \mathbf{A})$$

$$\mathbf{c}_i^n = \max\left(0, \mathbf{W}_1^n \mathbf{z}_i^n + \mathbf{b}_1^n\right) \mathbf{W}_2^n + \mathbf{b}_1^n$$

- Training Objective

$$\mathrm{P}(\mathbf{Y}|\mathbf{X}) = \prod_{i=1}^{|\mathbf{Y}|} \mathrm{P}\left(\mathbf{y}_i|\mathbf{Y}_{<i}, \mathbf{X}\right) = \prod_{i=1}^{|\mathbf{Y}|} \mathrm{P}\left(\mathbf{y}_i|\mathbf{c}_i^N\right)$$

Experiments

- Experiments on printed Chinese characters

TABLE I
COMPARISON OF ACCURACY RATE BEWTEEN RTN AND RAN WITH DIFFERENT CAPTION LENGTHS ON DIFFERENT FONT-STYLE UNSEEN CHINESE CHARACTERS RESPECTIVELY.

Font Style	RAN(%)			RTN(%)		
	ALL	≤ 6	> 6	ALL	≤ 6	> 6
Song	92.21	93.65	90.78	**94.54**	**94.93**	**94.16**
FangSong	91.04	91.98	90.11	**94.21**	**94.84**	**93.57**
Hei	90.41	91.34	89.50	**92.79**	**92.41**	**93.11**
Kaiti	88.57	90.59	86.58	**91.31**	**92.96**	**89.67**

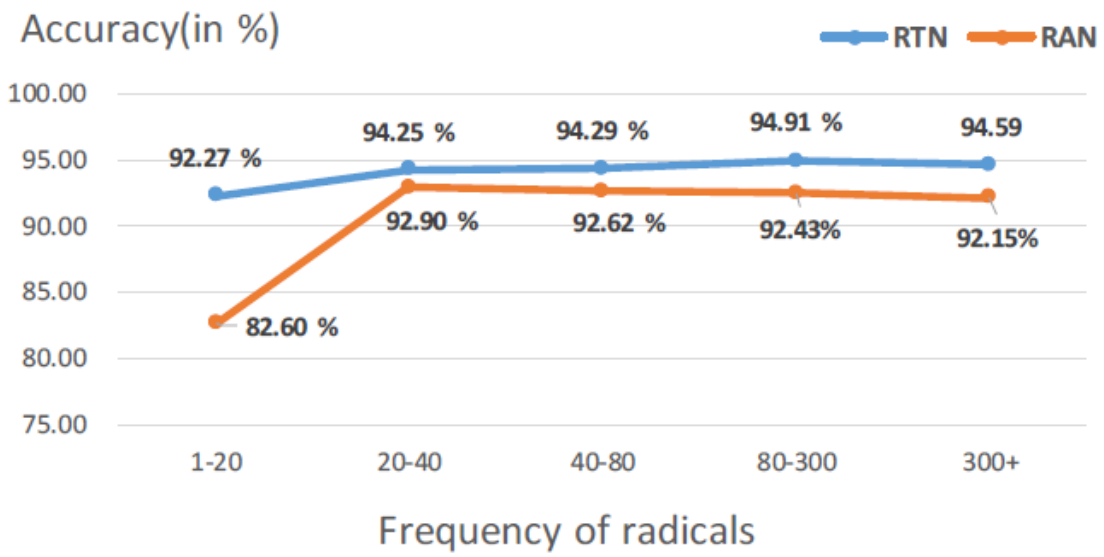


Fig. 2. Comparison of character-level accuracy between RTN and RAN with respect to the frequency of radicals. Approximately 100 radical categories are included in each range.

[1] Zhang, Jianshu, et al. "Radical analysis network for zero-shot learning in printed Chinese character recognition." *2018 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2018.

Experiments

- Experiments on natural scene Chinese characters

TABLE II
COMPARISON OF THE RECOGNITION PERFORMANCE OF RTN AND RAN WITH DIFFERENT CAPTION LENGTHS ON THE CTW VALID DATABASE

Model	Caption length		
	ALL	≤ 4	> 4
RAN	85.95%	89.33%	82.31%
RTN	**87.51%**	**90.02%**	**84.80%**
Accuracy↑	**1.56%**	0.69%	**2.49%**

TABLE III
COMPARISON OF THE PERFORMANCE OF RAN AND RTN WITH THE DIFFERENT APPEARANCE FREQUENCY OF CHARACTER-LEVEL CATEGORIES.

Frequency	≤ 20	≤ 50	≤ 100	HF	ALL
Categories	398	511	335	1044	2015
Samples	1128	1229	1663	48745	52765
RAN	25.88%	47.92%	65.12%	89.01%	85.95%
RTN	**41.84%**	**61.51%**	**71.67%**	**89.76%**	**87.51%**

TABLE IV
COMPARISON OF THE RECOGNITION PERFORMANCE OF RAN AND RTN WITH RESPECT TO 6 ATTRIBUTES ON THE CTW TEST DATASET; ALL INCLUDES ALL CHARACTERS ON THE TEST DATABASE.

Attributes	Training Samples	RAN(%)	RTN(%)
All	760107	85.56	87.31
occluded	101393	71.55	73.94
background	218560	82.84	84.57
distorted	192481	71.55	83.60
3D raised	199066	76.17	78.06
wordart	65983	87.11	84.25
handwritten	6661	63.58	66.70

Experiments

- Attention visualization

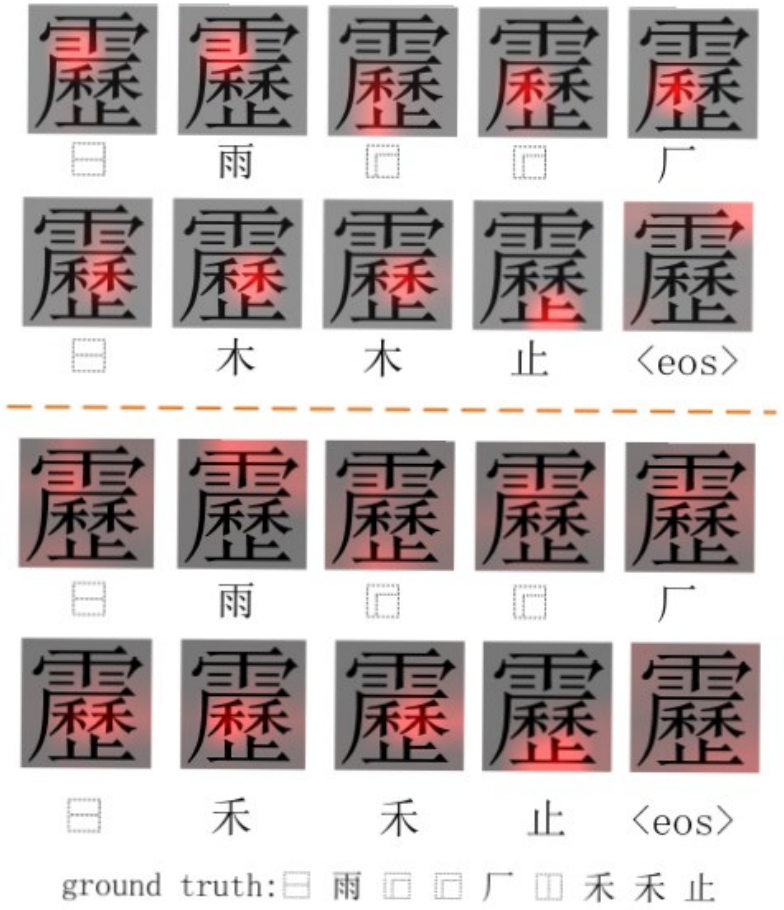


Fig. 4. Attention visualization of recognizing a complicated Chinese character step by step; the above is the process of RAN and the below is the process of RTN; symbols below the images are the predicted radicals or structures.

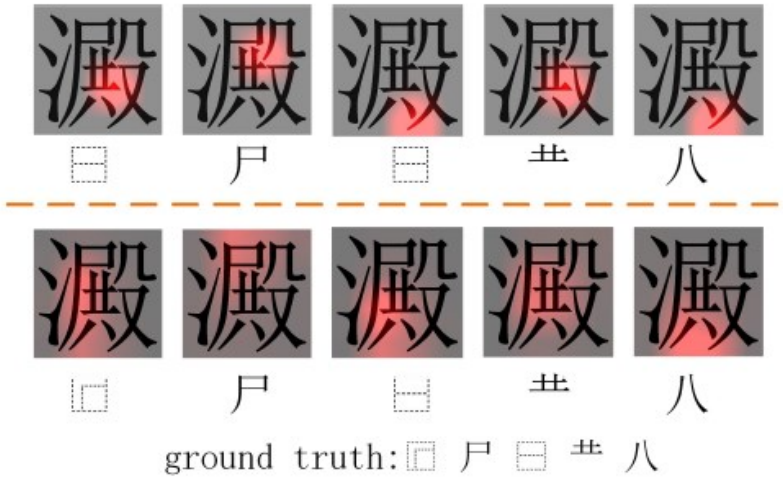


Fig. 5. Comparison of RAN and RTN on aligning the internal spatial structures; the above is the partial process of RAN and the below is the partial process of RTN.

Future Work

- Evaluate on zero-shot scene or handwritten character recognition
- Apply to Chinese text line recognition
- Design more task-specific transformer-based decoder

Thank you!