DenseRAN for Offline Handwritten Chinese Character Recognition

*Abstract*

In this study, we propose a novel radical analysis network with densely connected architecture (DenseRAN) to analyze Chinese character radicals and its two-dimensional structures simultaneously. DenseRAN first encodes input image to high- level visual features by employing DenseNet as an encoder. Then a decoder based on recurrent neural networks is employed, aiming at generating captions of Chinese characters by detecting radicals and two-dimensional structures through attention mechanism.

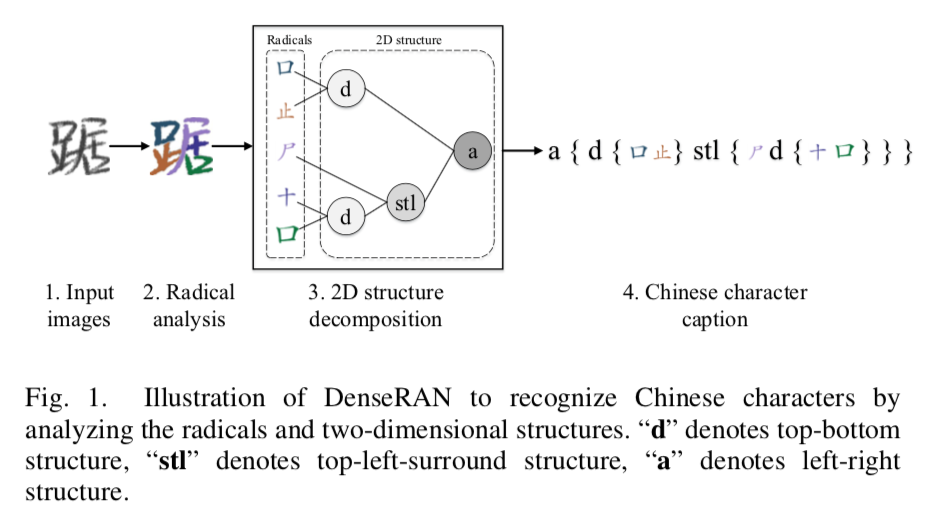
I. INTRODUCTION

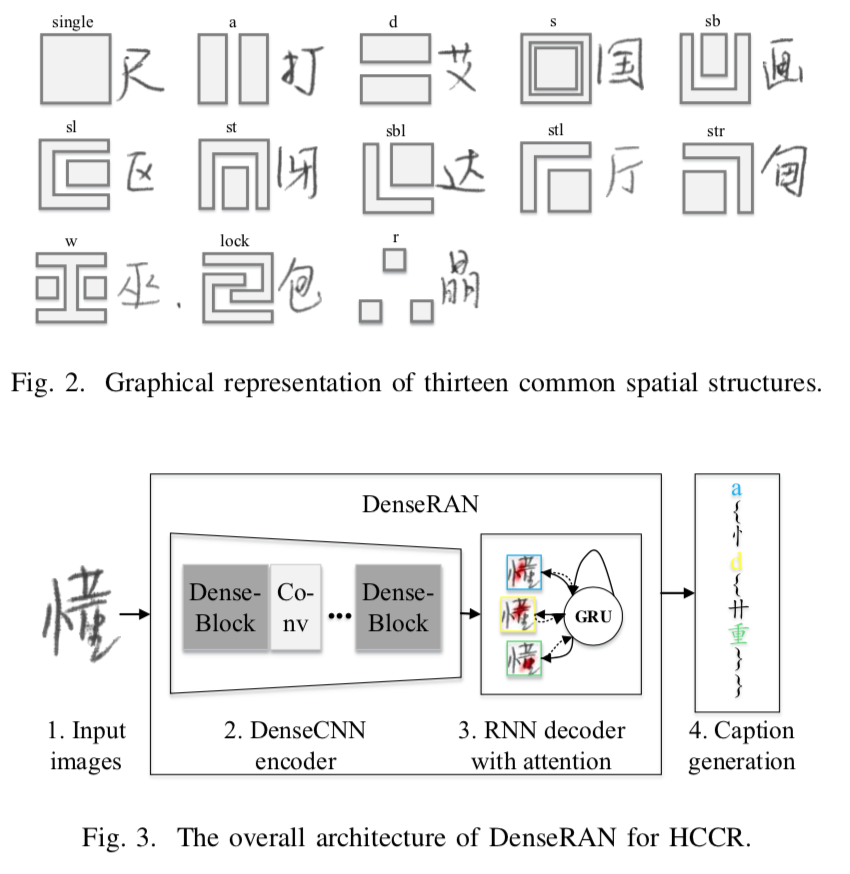
In traditional methods, the procedures for HCCR often include: image normalization, feature extraction, dimension reduction and classifier training. However, these algorithms can only recognize Chinese characters appeared in training set and have no ability to recognize unseen Chinese characters. Moreover, these algorithms treat each Chinese character as a whole without considering the similarites and sub-structures among Chinese characters.

In this paper, we propose a novel radical-based approach to HCCR, namely radical analysis network with densely connected architecture (DenseRAN).

The main idea of DenseRAN is to decompose a Chinese character into a caption that describes its internal radicals and structures among radicals.

DenseRAN first encodes input image to high-level visual vectors using a densely connected convolutional networks (DenseNet). Then a RNN with gated recurrent units (GRU) decodes the high- level representations into output caption step by step. We adopt a coverage based spatial attention model built in the decoder to detect the radicals and internal two-dimensional structures simultaneously.





II. CHINESE CHARACTER DECOMPOSITION

Each Chinese character can be naturally decomposed into a caption of radicals and spatial structures.

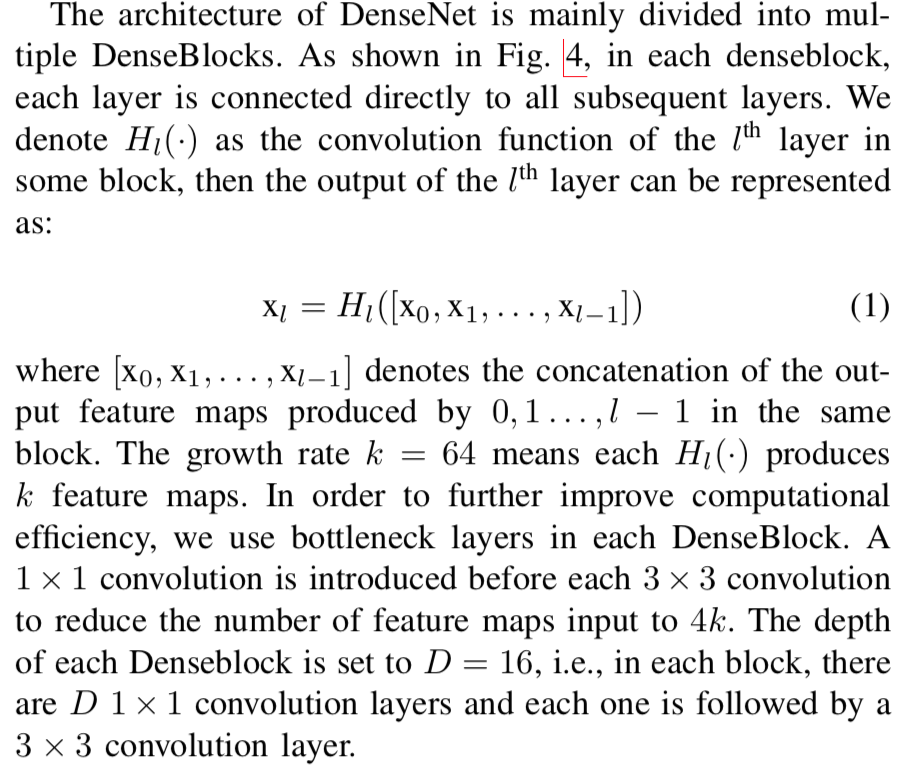
The character caption consists three key components: radicals, spatial structures and a pair of braces (e.g. “{” and “}”). One spatial structure with its radicals can be represented as: “structure {radical-1, radical-2}”.

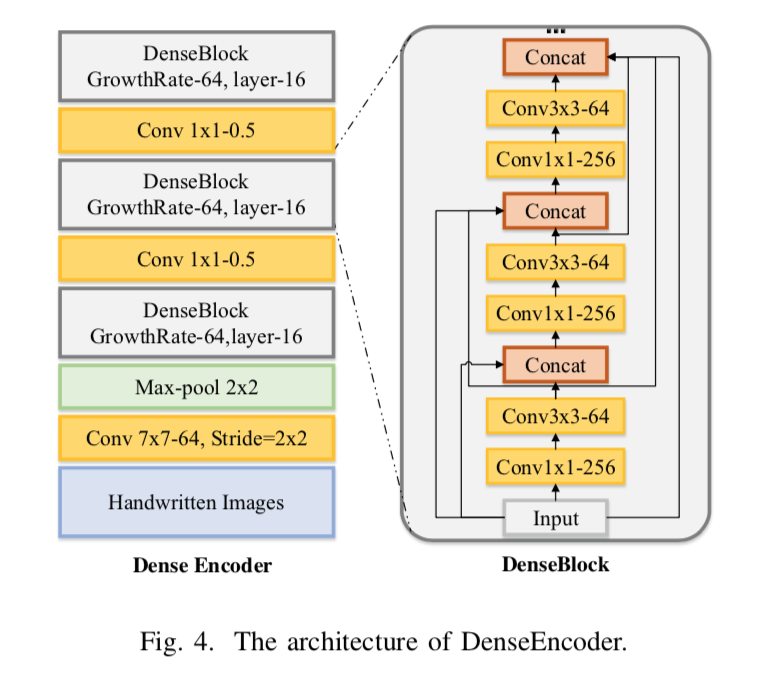
These thirteen structures are: single: some Chinese characters are radicals themselves. a: left-right structure. d: top-bottom structure. s: surround structure. sb: bottom-surround structure. sl: left- surround structure. st: top-surround structure. sbl: bottom-left-surround structure. stl: top-left-surround structure. str: top-right-surround structure. w: within structure. lock: lock structure. r: one radical repeated many times in a character.

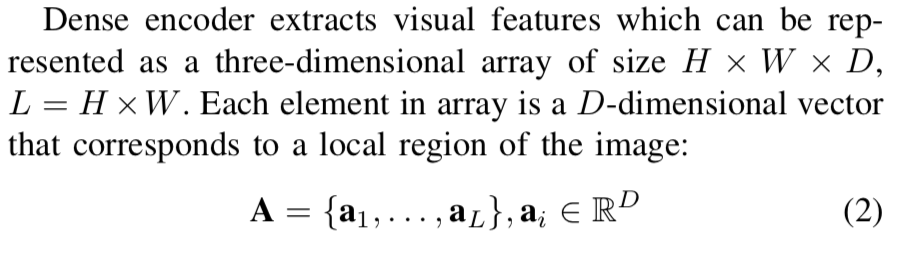
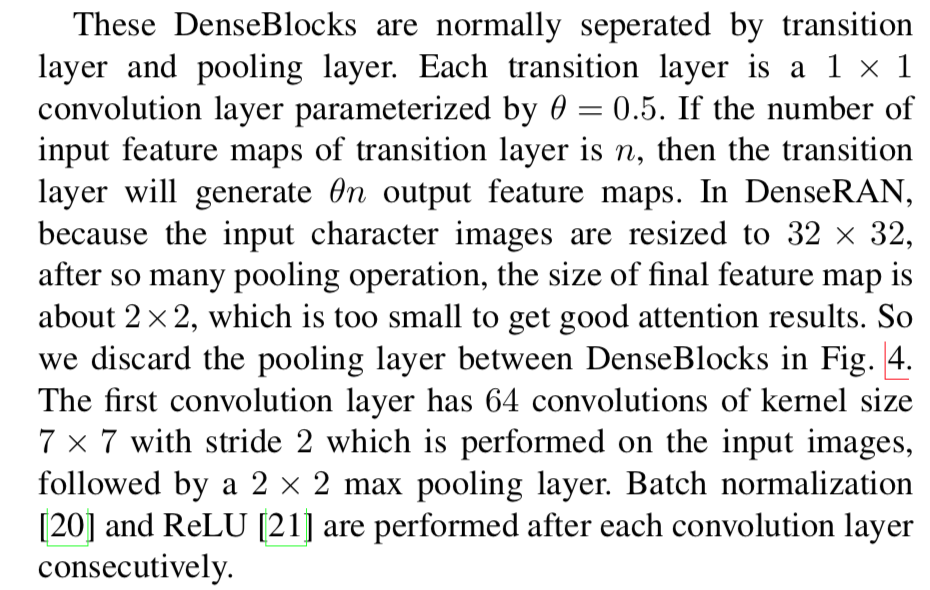
III. THE ARCHITECTURE OF DENSERAN

*A. Dense encoder*

We use DenseNet as the encoder to extract high- level visual features from images. Instead of extracting features after a fully connected layer, we discard fully connected layer and softmax layer in encoder, called fully convolutional neural networks. This allows the decoder to selectively pay attention to certain parts of an image by choosing specific portions from the extracted visual features.







*B. GRU decoder with attention model*

IV. EXPERIMENTS ON RECOGNIZING SEEN CHINESE CHARACTERS

*A. Dataset*

*B. Implementation details*

We normalize gray-scaled image to the size of 32×32 as the input.

We normalize gray-scaled image to the size of 32×32 as the input. The implementation details of Dense encoder has been introduced in Section III-A. The decoder is two unidirectional layers with 256 GRU units. The embedding dimension m and decoder state dimension n are set to 256. The convolution kernel of Q is set to 5 × 5 and the number of feature maps is set to 128. The model is trained with mini-batch size of 150 on one GPU. We utilize the adadelta with gradient clipping for optimization. The best model is determined in terms of word error rate (WER) of validation set. We use a weight decay of 10-4 and dropout after each convolution layer and set the dropout rate to 0.2.

In the decoding stage, we aim to generate a most likely caption string given the input character. The beam search algorithm is employed to find the optimal decoding path in the decoding process. The beam size is set to 10.

*C. Experiments results*

V. EXPERIMENTS ON RECOGNIZING UNSEEN CHINESE CHARACTERS

*A. Dataset*

*B. Experiments results*

VI. QUALITATIVE ANALYSIS

*A. Attention visualization*

*B. Error distribution of different two-dimensional structures*

VII. CONCLUSION AND FUTURE WORK