修士学位論文要約（平成29年３月）

畳み込みニューラルネットワークを用いた文字境界の形状評価に基づく単語認識に関する研究

柄澤 匡哉

指導教員：田中 和之，　学位論文指導教員：片岡 駿

A Study on the Word Recognition based on Evaluation of Boundary between Characters Using Convolutional Neural Network

Masaya KARASAWA

Supervisor: Kazuyuki TANAKA, Research Advisor: Shun KATAOKA

Scene text recognition is a practical application but still a challenging task due to complex backgrounds, font variations and so on. To improve the recognition accuracy, we propose a method based on evaluation of the shape of a character boundary. Although there are many studies on text recognition in natural images, there's no system considering the boundary shape between characters. By detecting a boundary between two characters based on its shape, contiguous characters can be segmented, false detection in non-character area can be reduced and estimating character area can revise position gap for recognition. We use Convolutional Neural Network (CNN) to build boundary classifier. We show some numerical results of our method using ICDAR 2003 datasets and compared with the existing methods.

1.　序論

近年のカメラ付きスマートフォンやウェアラブル端末の普及により，日常生活中の文字情報の撮影はより身近なものとなってきている．文字の撮影によって，カメラによる自動翻訳や商品情報の読み取り，ウェブ上の画像の整理や検索といった応用が可能なことから，実生活情景中の画像を対象とした文字認識技術の研究は広く行われている．

単語中の文字認識においては，画素値の射影によって個別文字を切り出して認識を行うものや [1]，スライディングウインドウによって画像上の個別文字を検出する手法が広く行われる [2)3)]．しかし情景中の文字にはフォントの多様性や背景ノイズ等の変動要因が多く，切り出しの失敗や，文字ではない領域における文字の誤検出が起こりうる．このような先行研究においては，単語画像中の文字間の形状を考慮した文字認識は行われていない．本研究においては，風景中の文字間の境界の形状の認識結果を利用した単語認識を提案し，精度向上を期待する．

2.　単語認識手法

はじめに単語認識における個別文字分類器の構築を行う．C を 62 クラスの文字の集合を示すものとする．この 62 クラスは 26 種の大文字アルファベット，26 種の小文字アルファベット，10 種の数字の合計からなる．

個別文字を示す画像を x として，画像が示す文字種 c に分類するために，x を d 次元のベクトル表現に変換する射影関数 $\boldsymbol{f}(x)$ を考える．分類器 ϕ は，文字種の事後確率分布 $p(c|x) = \phi(\boldsymbol{f}(x))$ を計算することで訓練がなされる．

分類される字種は，事後確率を最大化するもの

$$\tilde{c} = \operatorname{argmax} p(c|x) \tag{1}$$

として認識される．この $\boldsymbol{f}(x)$ として畳み込みニューラルネットワーク (CNN) の一種である Network in Network を用いる．

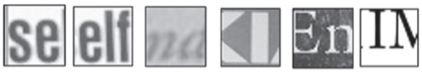
続いて，訓練された個別文字分類器を用いて単語画像から単語の推論を行う．単語画像上を個別文字分類器で走査することで，画像の各位置における各字種らしさを表す行列 $\mathbf{M}(36 \times T)$ を生成する．この際に同字種の大文字・小文字の区別をなくし，36 は字種のクラス数，T は入力画像の横幅に対応する．

この行列 $\mathbf{M}$ の各列 $t(0 \leq t \leq T)$ 中の最大値と 2 番目に大きい値との差を $R(t)$ として，式 (2) の Non Maximum Supression(NMS) を適用することで検出の重複を抑制する．式 (2) 中の δ は抑制の幅を制御するパラメータとなる．この処理によって，$\tilde{R(t)} > 0$ となる t の最大値のクラスを指す文字を用いて，認識結果となる文字列 z を出力する．

$$\tilde{R(t)} = \begin{cases} R(t) & \text{if } R(t) \geq R(t'), \forall t' |t - t'| < \delta \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

ここで得られた文字列 z との編集距離の小さい辞書（単語画像と共に与えられる単語集合）中の単語を候補単語 w とする．行列 $\mathbf{M}$ に基づいた候補単語の妥当性の検証を式 (3) によって行う．

$$S(w, \mathbf{M}) = \frac{1}{L} \Sigma_{i=1}^{L} s(c_i, p_i, \mathbf{M}). \tag{3}$$



(a) 境界画像：文字境界が画像中心に位置

(b) 非境界画像：文字境界が画像中心以外に位置

図 1: モデル訓練用画像

$s(c_i, p_i, \mathbf{M})$ は候補単語 w 中の i 番目の文字 c_i の認識スコアであり，p_i は c_i が中心となる位置，L は単語の文字数である．スコアの参照位置 p_i は，文字 c_i が画像中に均等に分布すると想定されることより，単語画像の幅から候補単語の文字数で均等に割ることで求められる．この S を最大化する単語を最終的な認識結果として出力する．

3.　境界の形状評価に基づく認識

前節にて述べた単語認識手法においては，画像中の走査の際に文字の境界周辺で不適切な反応が生じうる，単語認識時に不適切な位置で字種らしさを評価しうる，といった問題点がある．この改良のために，文字間の境界の形状を学習し，その認識結果を利用した単語認識手法を提案する．

はじめに学習モデルの訓練用の画像を境界画像，非境界画像として定義する．図 1 に例を示す．これらを CNN モデル（Network in Network）に入力し，入力が単語中の文字領域を示すか，境界領域を示すかの 2 クラス分類を行う．個別文字分類と同様に単語画像上を走査することで，単語中の文字らしさを表すベクトル $\boldsymbol{b_0} = \{b_{01}, b_{02}, ..., b_{0T}\}$ を生成する．

この学習結果 $\boldsymbol{b_0}$ を利用した二つの提案を行う．第一に，個別文字分類時の行列 $\mathbf{M}$ の各行の要素と $\boldsymbol{b_0}$ の要素との積をとることで，行列 $\mathbf{M'}$ として定義する．これにより境界位置における文字検出の出力を抑制し，誤認識を削減する．本手法を “スコア補正” と定義する．第二に，スコア $\boldsymbol{b_0}$ と二つのしきい値を指すパラメータを用いてスコアの参照位置を決定する．これにより，不適切な位置における文字の評価を削減する．本手法を “位置決定” と定義する．

4.　数値実験

ICDAR2003 の単語認識用試験画像 861 枚を対象とした単語認識実験を行い，境界情報の利用に関する各手法の比較を行った．評価辞書は ICDAR2003 中の全考慮単語とする．各手法における認識精度の一例を表 1 に示す．表中の “境界無し” が既存手法を示す．

本結果より，提案手法による境界位置での誤った文字のスコア抑制や，均等割りにおける適切な位置からずれが生じていた文字のスコア向上といった働きが確認できた．認識結果の一例を表 2 に示す．

表 1: 各手法における単語認識精度 [%]

	認識精度
境界無し	83.97
スコア補正	85.71
位置決定	85.13
スコア補正*位置決定	**86.41**

表 2: 認識結果の例

入力画像と境界値	境界無し	スコア補正*位置決定
	CHIPS	**HILLS**
	ROOM	**FROM**

5.　結論

本論文では，風景画像中の単語認識における文字の誤識別を抑制するために，文字境界の形状評価を利用した単語認識を提案した．これにより諸条件において境界情報を用いた手法の認識精度が既存手法に対し上回る傾向を確認した．スコアの補正や位置決定によって，不適切な候補単語に対し重みをかける，単語画像に対してより有効な位置の補正を加える，といった作用が有効に働き，単語認識精度の向上に繋がったと結論付けた．

今後は，境界情報の単語認識への導入手法の改良，辞書の評価の改良，不鮮明画像や角度変化を伴う画像を追加しての学習等により，更なる結果の改善が期待できる．

文献

1) Huang, X., et al. “Text Detection and Recognition in Natural Scene Images” in 2015 International Conference on Estimation, Detection and Information Fusion (ICEDlF 2015) (2015)

2) Mishra, A., et al. “Top-Down and Bottom-up Cues for Scene Text Recognition” in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2012, pp. 2687-2694. (2012)

3) Liu, X., et al. “Scene Text Recognition With High Performance CNN Classifier And Efficient Word Inference” Acoustics, Speech and Signal Processing (ICASSP), 2016 International Conference on. IEEE, 1322-1326 (2016)