ISS-SP-206

CNN を用いたテロップの外接矩形検出

Detecting bounding rectangles of captions using convolutional neural networks

柳瀬直人 [1]　島田裕 [1]　谷口行信 [1]
Naoto Yanase　Yutaka Shimada　Yukinobu Taniguchi

東京理科大学 工学部第一部 経営工学科 [1]
Faculty of Engineering, Tokyo University of Science

1　はじめに

映像データの効率的な検索のために，映像中のテロップを解析しインデックスとして利用するアプローチがある [1]．Minetto ら [2] は文字の形状特徴と HOG 特徴量に基づき SVM を用いてテキスト領域を検出した．また，Huang ら [3] は，畳込みニューラルネットワーク (Convolutional Neural Network，以下 CNN) を用いたテキスト検出法を提案した．しかし，この 2 つの手法はいずれも風景中の文字列（シーンテキスト）を検出するもので，映像に重畳されるテロップの検出に特化したものではない．本稿では，CNN を用いてテロップの外接矩形を検出する手法を提案する．具体的には，局所的なテロップらしさを CNN により評価した上で，テロップ文字列に特有の性質—水平垂直方向に整列し，外接矩形が長方形—をポストフィルタにより評価することでシーンテキストの誤検出を抑制する．

2　提案手法

提案手法の概要を説明する．あらかじめ，テロップらしさを表すスコア関数を CNN により学習する．CNN の構造は，物体認識のタスクでよく用いられる AlexNet と同様とした．20 × 20 画素 ×3 チャンネルの画像パッチを入力とし，テロップを含む場合は 1，それ以外は 0 を理想出力とする．検出の際には，入力画像の各画素に対して，CNN を用いてテロップらしさを表すスコアを算出し (図 1:(b))，行（あるいは列）ごとのスコアの総和を用いてスコアを更新する (図 1:(c))．その後，大津の二値化によって得られた閾値よりスコアが大きい画素を白，小さい画素を黒にすることにより二値化を行い，ラベリング処理によって白の連結領域を求める (図 1:(d))．最後に，1 行ごとに矩形が抽出できるように，各領域の外接矩形に対して矩形分割を行う (図 1:(e))．以下では，スコア計算と矩形分割についてさらに説明する．

スコア計算 サイズ $w \times h$ の入力画像の画素 (x, y) に対して，テロップらしさを表すスコアを $s(x, y)$，初期値を $s(x, y) = 0$ とする．20 × 20 画素の走査窓を，水平および垂直方向に 4 画素ずつ移動させながらフレーム全体を走査する．その際，走査窓内の画像を CNN へ入力して，得られた出力値 p を，走査窓内の全画素に対して $s(x, y) \leftarrow s(x, y) + p$ と加算する．その後，第 i 列目 $(i = 0, 1, \cdots, w-1)$ の画素のスコアの総和 $V_i = \sum_{k=0}^{h-1} s(i, k)$ と，第 j 行目 $(j = 0, 1, \cdots, h-1)$ の画素のスコアの総和 $H_j = \sum_{k=0}^{w-1} s(k, j)$ を用いて，スコアを $\tilde{s}(x, y) \leftarrow \max(V_x, H_y) \cdot s(x, y)$ と更新する．テロップ領域が長方形に近い形状をしていることから，テロップが存在する行（あるいは列）における $\max(V_x, H_y)$ は大きな値を持つことが多い．したがって，テロップ領域内の画素における $\tilde{s}(x, y)$ の値は相対的に大きくなる．

矩形分割 各連結領域の外接矩形に対して，白い画素を 1，黒い画素を 0 とした行列を定義し，この行列において隣接する行の行ベクトル同士のコサイン類似度を計算する．コサイン類似度が閾値 θ 以下の行ベクトル間を境界線とし，矩形分割を行う．図 2 に矩形分割の例を示す．

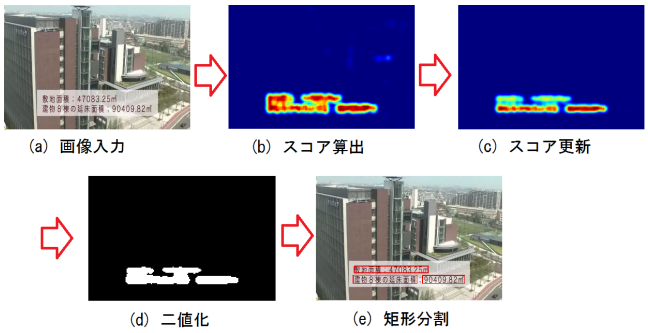

図 1　提案手法のフロー

表 1　テロップ外接矩形の検出精度

	再現率	適合率	F 値
提案手法	0.85	0.75	**0.78**
Snooper-Text	0.64	0.83	0.71

3　実験・考察

提案手法の有効性を示すため，テロップ外接矩形の検出精度を従来手法と比較した．実験には，CNN のフレームワークである Caffe[4] を利用した．テロップを含むフレーム画像を 20×20 の矩形に切り出したものを学習画像とし，画像内にテロップの一部が含まれるものを正例，それ以外のものを負例とした．ここで，正例の個数は 11,286 個，負例の個数は 25,053 個である．評価には，google および YouTube から無作為に取得した 50 枚の画像を用いた．これらの画像に含まれるテロップ外接矩形の総数は 290 個である．比較対象として，Snooper-Text[2] と呼ばれる SVM ベースの手法を用いた．実験の結果を表 1 に示す．比較手法と比べて，提案手法の F 値が 7% 上回っていることが確認できる．誤検出が多かった画像は，テロップ以外の CG やイラスト，幾何学模様である．このような誤検出が生じる原因として，エッジ強度などの空間的性質がテロップと似ているため，CNN による識別がうまくいかなかったためと考えられる．また，1 つのフレーム内に異なる長さのテロップが存在する場合に，短い方のテロップが検出されないケースが見られた．

4　おわりに

本稿では，テロップ外接矩形を検出する手法を提案し，従来手法を上回る精度を確認した．今後はネットワークのパラメータを最適化して，さらなる精度向上を目指す．

参考文献

[1] 佐藤ほか, 信学論, **J81-D-II**, (8), 1847–1855, 1998.
[2] R. Minetto et al., CVIU, **122**, 92–104, 2014.
[3] W. Huang et al., ECCV 2014, pp. 497–511, 2014.
[4] `http://caffe.berkeleyvision.org/`

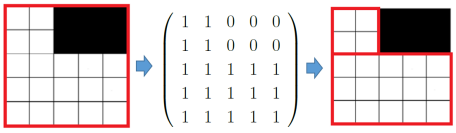

図 2　矩形分割のイメージ