CNNを用いた商品札文字認識

PS2-51

成沢 淳史, 柳井 啓司
The University of Electro-Communications, Tokyo

背景 & 目的

情景文字認識 & 物体認識 による商品札認識

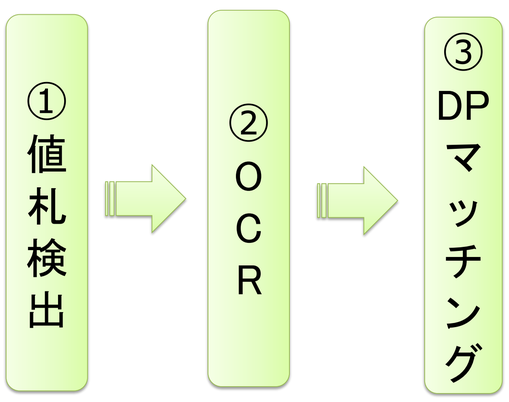
“みかん”

&

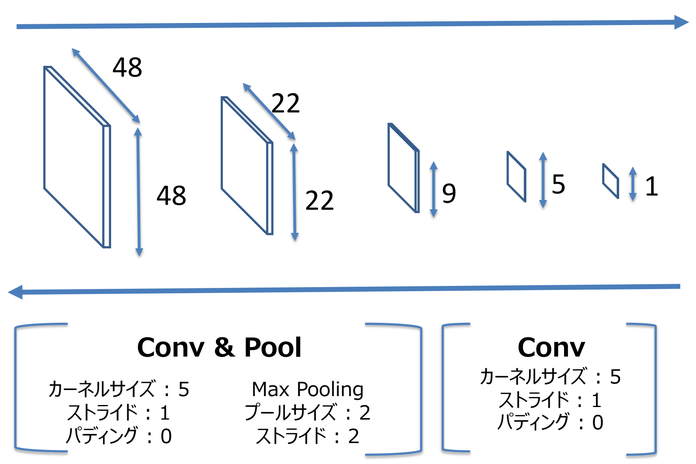
身近な価格情報を自動で読み取る システムの構築

- CNN によるテキストスポッティング
- 文字の CNN 特徴表現による動的マッチング
- シーン画像中から単語の認識まで

ワークフロー & ネットワーク



提案手法

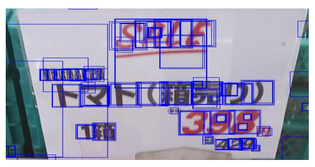


値札検出に使用する CNN

① 値札検出

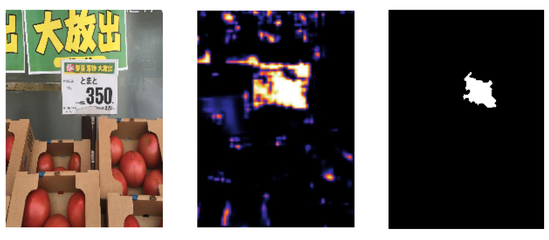
- 値札領域に対してバウンディングボックスの付与
  (値札の画像100 枚に対し人手による作業)
- 前景から MSER により文字画像パッチ, 背景からランダムに切り出しデータセット作成
- 学習：CNNを学習 (入力：48x48)
- 領域推定時：
  入力は 900x600 にリサイズ or 複数スケールの入力結果を統合
- ラベリングアルゴリズムを適用し最大の領域を取得する.

- 値札領域にアノテーション



- 値札領域にMSERアルゴリズムを適用

- CNN を使った 領域推定



- 左) 入力, 中央) CNNによる領域推定, 右) ラベリング後

② OCRによる読み取り

文字の読み取りにはオープンソースの OCR である pytesseract 使用
(与える画像の状態で認識結果が変化しやすい)

③ DP マッチング

- フォント画像を CNN で学習
- トレーニングデータの各文字ごとの softmax 出力平均を予め計算
- 予め計算した文字の特徴表現を利用してDP マッチング

さゆうり ->
きゅうり
にら
さけ

ほぅね ‘ん草 ->
ほうれん草
込格
なす

“税込価格” のように一部学習できなかった文字がある

経路コスト最小を探索

り				0.0
う			0.0	
ゆ	5.2	0.12		
さ	0.28	4.3		
	き	ゅ	う	り

辞書中の単語

実験

- 値札検出から文字の読み取りまで処理をテスト
  トレーニング 100 枚, テスト 20 枚を用意し値札領域を抽出
- 入力：
  900x600 のシングルスケール
  リサイズ後 0.3 , 0.5, 1.0 倍の複数結果を統合したマルチスケール
- 結果：
  テスト画像に対して **15** / 20 枚 (**75%**) の値札を検出

	値札検出	商品名読み取り	値段読み取り
20枚中	15 (75%)	11 (55%)	3 (15%)

- 上段) シングルスケール 下段) マルチスケール

値札検出失敗例

商品名読み取り失敗例

値札領域の取得は良好
OCRによる文字認識では
値段を全く読み取ることができなかった

まとめ & 今後の展開

- カメラ撮影ではノイズが多いので CNN で値札スポッティングを行った
- スーパーの商品札は活字のためOCRで読み取り
- CNN の文字特長表現によるDPマッチング
- 文字認識した単語を画像中の商品領域と結びつけ物体認識と組み合わせる