Multi-Task DSSD による広告画像からの文字領域検出

EP16010 伊藤咲季　　指導教授：山下隆義

1. はじめに

広告画像には，期間を設けずに「タイムセール」や「期間限定」などと表記する不当表示が存在する．広告画像が不当表示かを判断するには，複雑な背景から文字領域の検出を正確に行う必要がある．本研究では，広告画像の文字領域の検出を目的とし，文字領域の検出とセグメンテーションを Multi-Task DSSD により同時に行うことで高精度化を実現する．

2. Multi-Task DSSD

Multi-Task Deconvolutional Single Shot Detector (M-T-DSSD)[1] は，Single Shot Multibox Detector (SSD) に Deconvolution 層を追加した Deconvolutional Single Shot Detector (DSSD) をベースとし，セグメンテーションタスクを追加したマルチタスクネットワークである．MT-DSSD のネットワーク構造を図 1 に示す．MT-DSSD は，デフォルトボックスと呼ばれるテンプレートとのオフセットを回帰することで物体の位置を予測する．

予測レイヤは，3 つの branch で構成されている．Confidence branch で物体のクラス識別を行い，Localization branch で物体検出を行う．そして，Segmentation branch で各物体クラスのセグメンテーションを行う．Segmentation branch で得たセグメンテーションマップを全て結合することで物体の大きさに関わらないセグメンテーション結果を得る．Non-Maximum Suppression (NMS) を行う際に，セグメンテーション結果を反映させることで物体検出の位置精度を向上させている．

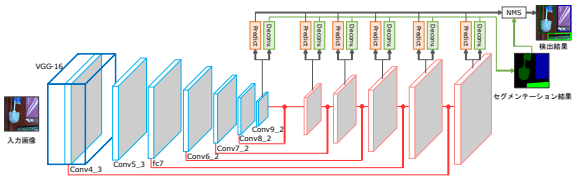

図 1：MT-DSSD のネットワーク構造

MT-DSSD の損失関数を式 (1) に示す．ここで，y は予測結果，t は教師データであり，添字の c はクラス確率，l は予測したバウンディングボックス，s はセグメンテーションマップである．L_{conf} は Confidence loss，L_{seg} は Segmentation loss であり，Softmax 関数と交差エントロピーで求める．L_{loc} は Localization loss であり，平均二乗誤差で求める．

$$L(y,t) = \frac{1}{3}(L_{conf}(y_c,t_c) + L_{loc}(y_l,t_l) + L_{seg}(y_s,t_s)) \tag{1}$$

3. 提案手法

文字領域は，一般物体よりも横に長い形状であり，サイズも小さい．そのため，一般物体検出を目的とした MT-DSSD では，文字領域の検出は困難である．本研究では，デフォルトボックスのアスペクト比を文字領域検出に適したアスペクト比に変更する．また，セグメンテーションをより重点的に学習するために，損失関数の重みを変更する．

3.1. デフォルトボックスのアスペクト比の変更

MT-DSSD のデフォルトボックスのアスペクト比は [1:1, 2:1, 3:1, 1:2, 1:3] である．提案手法では，広告画像における文字領域のアスペクト比の分布より，横文字中心の文字領域の検出に適した [1:1, 2:1, 3:1, 7:1, 10:1] とする．

3.2. 損失関数に重み係数を追加

MT-DSSD では，損失関数の L_{conf}，L_{loc}，L_{seg} の重みを全て 1 としている．本手法では，L_{conf}，L_{loc}，L_{seg} の重みをそれぞれ 0.75，0.75，1.5 に変更し，セグメンテーションをより重点的に学習する．

4. 評価実験

アスペクト比の変更と損失関数の重み係数追加の有効性の調査のため，作成したデータセットを用いて MT-DSSD と比較を行う．なお，本実験では，セグメンテーション結果を反映せずに NMS を行う．

4.1. 学習用データセットの生成

画像上に文字を生成する Text Image Generator を用いて学習用データセットを生成する．楽天市場に掲載されている商品画像をベースとし，生成した文字を合成する．文字は 8 種類の単語のフォントをランダムに変更し，1,000 枚の画像を生成する．フォントの種類は 13 種類である．Text Image Generator による生成例を図 2 に示す．

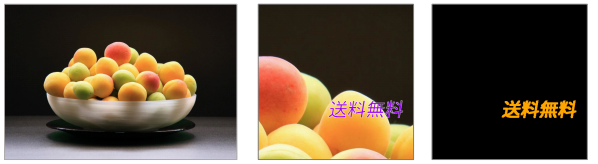

図 2：画像の生成例

4.2. 実験結果

検出精度の比較を表 1 に示す．なお，表 1 の P は Precision，R は Recall，F は F-measure を表す．MT-DSSD とアスペクト比を変更した MT-DSSD の F-measure を比較すると，アスペクト比を変更した場合，変更前よりも 3.5 高い精度であった．また，MT-DSSD と損失関数に重み係数を追加した MT-DSSD の F-measure を比較すると，損失関数の重み係数を追加した場合，追加前よりも 9.5pt 高い精度であった．

表 1：検出精度の比較

手法名	P	R	F
MT-DSSD	45.5	14.7	22.2
MT-DSSD(アスペクト比変更)	47.4	17.6	25.7
MT-DSSD(重み係数追加)	**53.5**	**22.6**	**31.7**

図 3 に各手法の検出例を示す．図 3 より，MT-DSSD では検出できていない文字領域を重み係数を追加した MT-DSSD では検出できており，重み係数追加により検出数が増加していることがわかる．

図 3：評価実験による検出結果

5. おわりに

本研究では，MT-DSSD のデフォルトボックスのアスペクト比の変更と損失関数の重みの変更の有効性の調査を行った．評価実験の結果，デフォルトボックスのアスペクト比の変更，損失関数の重み係数追加による検出精度の向上を確認した．また，重み係数の追加による検出数の増加を確認した．今後は，データセットの見直しやハイパーパラメータの変更による精度向上を目指す．

参考文献

[1] 荒木諒介 等,“マルチタスク学習を導入した Deconvolutional Single Shot Detector による物体検出とセグメンテーションの高精度化”, MIRU, 2018.