

## 高精度画像検索法の検討

王 ショウ 木本 伊彦  
東洋大学大学院理工学研究科

## 1. はじめに

技術の進歩と共に、インターネットだけではなく、人間の日常生活にも大量な類似画像が徐々に出現している。混雑した背景、そして複雑な要求に対して、従来の画像検索法の実現は難しくなる。

近年、ディープラーニングによる画像検索は、画像の高レベルの特徴量まで抽出できるため、ますます注意を集めている。正解率、識別性などの性能から見ると、ディープラーニングによる画像検索法は従来のアルゴリズムを超えているが、検索効率にはまだ改善する余地がある。

本稿では、Faster RCNN ネットに用いられる Non-Maximum-Suppression(NMS)方法を中心に、高密度のターゲットが存在する画像に対して、モデルの精度を上げる方法を検討する。

## 2. 従来の Faster RCNN 法による画像検索

## 2.1 Faster RCNN の構成

ディープラーニングを用いて、初めて目標検索領域に End-To-End やり方を実現した代表的な検索方法として、Faster RCNN[1]がある。

Faster RCNN とは、入力された画像データから、VGG-16 などの Convolutional Neural Networks(CNN) ネットを使用し、特徴マップを先に生成する。注意力メカニズムネットワーク Region Proposal Network(RPN)を使い、特徴マップにターゲットが存在する可能性のある区域(前景)を推定する。最後に、特徴マップに載せた前景ボックスを CNN ネットに送り、各ボックスの所属種類と位置関係を予測できるネットワークである。図 1 に Faster RCNN の構成を示す。

Faster RCNN は、画像検索の分野において、検索スピードと検索精度を著しく向上させた。

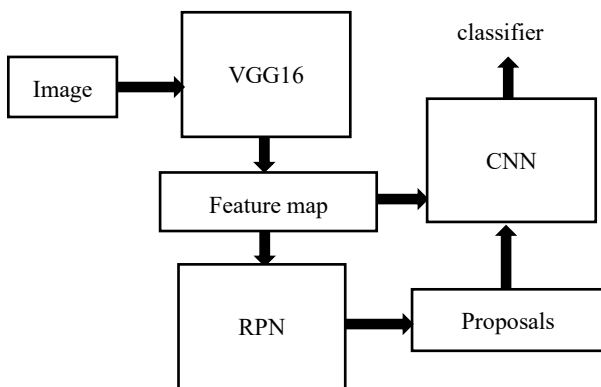


図 1 Faster RCNN の構成

## 2.2 画像検索の原理

画像のターゲットには、位置以外、サイズと形状も相違する無数の可能性がある。ワンパターンの矩形ボックスを通じて、すべての目標を適切に獲得することはできない。この問題を解決するために、Faster RCNN に初めて Anchor という概念を導入した。

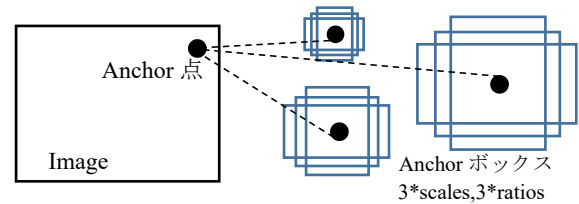


図 2 Anchor

Anchor とは、画像の枠内で、事前に固定される位置に展開されるボックスのことである。図 2 に示す。区別するために、固定する位置を Anchor 点、ボックスのことを Anchor ボックスとする。目標の形状とサイズに合わせるため、一つの Anchor 点を中心に、9 個の Anchor ボックスを生成する。目標位置の問題を考慮することから、全ての Anchor 点を挙げて、各点の Anchor ボックスがターゲットになれるかという問題に変換する。

したがって、大量の重複度がある Anchor ボックスの中で、必要なボックスだけを保存することは重要である。その不可欠な役目を果たすものは NMS 方法である。

## (1) IOU

NMS に重要なパラメータとして、Intersection over Union (IOU)がある。二つのボックスの共通部分が全体に対する割合と定義される。図 3.(a)で示す。

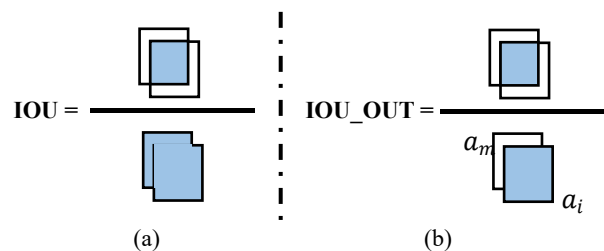


図 3 IOU と IOU\_OUT

## (2) NMS

NMS とは、同じターゲットに所属するボックスの集合から、代表ボックスを選択し、他の重複ボックスを削除するアルゴリズムである。

代表ボックスを選択する根拠は、ボックスの点数が高ければ、ターゲットになる可能性も高くなることである。そして、二つのボックスの IOU 値が設定された閾値  $N_t$  を超えることにより、二つのボックスが同じターゲットに所属すると判断され、低いスコアのボックスを削除する。図 4 には、NMS[2]の流れを示す。

†Graduate School of Science and Engineering, Toyo University

†Xiao Wang †Tadahiko Kimoto

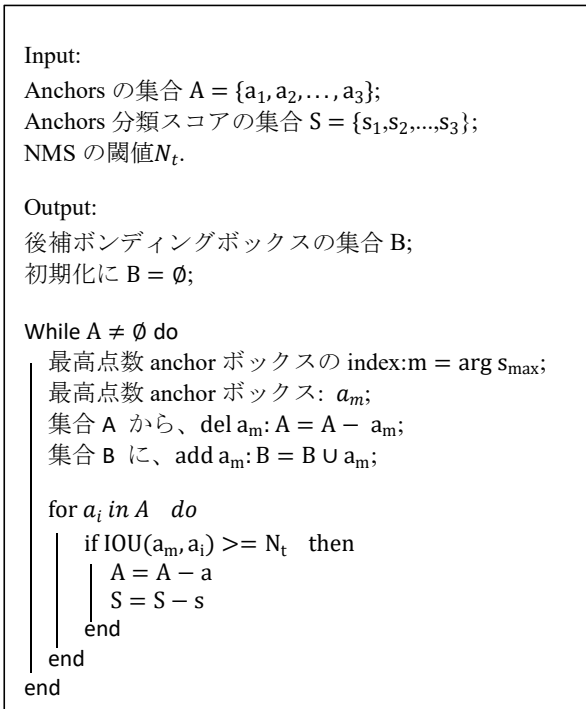


図4 NMS の流れ

### 2.3 画像検索における問題点

高密度のターゲットが存在する画像に対して、Faster-RCNN を応用すると、図5に示すように、(a)右側にいる人は検出されないこと及び真ん中にある男の予測範囲に両辺にいる人の部分を含める現象も多発する。(b)左側にいる男の予測範囲に隣の女の大部分を含むことと右側に一つのボックスに二人が入ってる状況になっている。どちらにしても、検出結果には望ましいことではない。

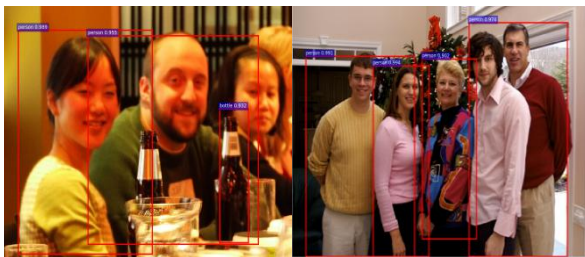


図5 良くない例

本稿の目的は次の二つである。一つは、図5のような誤検出と未検出の形成原因を分析する。二つ目は検索効率を上げる方法を提案する。

### 2.4 誤検出と未検出の原因の分析

図4より、NMSのやり方は明白である。まず、最高点数のボックスを探す。そして、そのボックスとのIOU値がNMSの閾値を越えるボックスを削除する。同じ操作を繰り返し、最後に残ったボックスの集合は別のターゲットの集合である。

しかし、高いIOU値を持つことは必ず同じターゲットに所属することではない。距離の近いターゲットは常に高いIOU値を持つが、違う分類に判断しかねない。また、二つのボックスの共通面積以外の部分が多かったら、同じ

ターゲットに所属するべきが、違う分類されることも発生する。同様に、低いIOU値を持つことは必ず違うターゲットに分類するべきことではない。

実験結果から見ると、Anchorボックスのスコアの高低は、ボックス内部に存在する特徴量の割合に関係がある。高密度のターゲットの場合、ターゲットとのエッジを判断することが難しくなる。よって、図5に示されるように、予測される目標の範囲に誤差がある。また、誤差のあるボックスを使い、NMS方法を行うことで、他のターゲットの判断にも影響を与えて、未検出問題の原因にもなる。

また、NMSに使われる閾値の大小も、検索結果を左右する要因となる。

## 3. 改善策の提案

Faster RCNNを再現した結果からすると、高密度の目標に対して、部分ターゲットを検出できないことと一つのAnchorボックスに複数の物体が存在する現象が多い。それらの原因は、NMSに使用される閾値の不変性、及び基準と選択されるAnchorボックスの不適切な位置に関係がある。

また、IOU値は1に近ければ、近いほど二つのボックスが同じターゲットになる可能性も高まる。逆の場合は成立しない。つまり、IOU値が低くなると、二つのボックスの位置関係を判断できないと考える。

### 3.1 IOU\_OUT

IOU値が閾値以下の場合、従来のNMSの手法は直接同じターゲットに所属しないと判断され、次の判断集合に加わる。その代わりに、IOU\_OUTというパラメータを提出する。図3.(b)で示すように、最高点数anchorボックス $a_m$ との共通部分が自身に対する割合をIOU\_OUTと定義する。このパラメータを用いて、他のターゲットのボックスによって、ボックス $a_m$ の位置を修正することができる。

### 3.2 ボックスの修正

最高点数anchorボックス $a_m$ とのIOU値が閾値 $N_t$ 以下のボックスの集合を集合Cとする。集合Cのボックス $a_i$ と $a_m$ のIOU( $a_m, a_i$ )は閾値 $N_t$ を超えていないが、IOU\_OUTが高い数値を持っている。ここで、何もしないと、従来法に従って、閾値 $N_t$ 以下だから、ボックス $a_i$ が他のターゲットに所属すると知りながら、他のターゲットのIOU\_OUT割の特徴量を含める状態でボックス $a_m$ を結果にして保存する。それは一つのAnchorボックスに複数の物体が存在する結果をもたらす原因となる。

スコアの差が小さければ、まず修正に使うボックス $a_i$ の信頼度を保証できる。そしてボックス $a_m$ に対する共通部分の重要性が低いことと考えられる。逆に、ボックス $a_i$ に対する共通部分の重要性が高くなる。つまり、共通部分はボックス $a_i$ に所属するべきと考慮する。

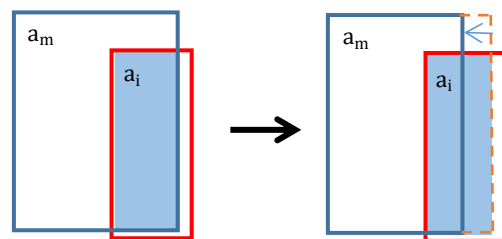


図6 ボックスの修正

$\forall a_i \in C, N_h = 0.8:$

$$IOU(a_m, a_i) < N_t \quad (1.1)$$

$$abs(s_m - s_i) < 0.1 \quad (1.2)$$

$$IOU\_OUT(a_m, a_i) > N_h \quad (1.3)$$

そのため、ボックス  $a_i$  が上の式を満たす時に、図 6 のようにボックス  $a_m$  のエッジを調整することは望ましい。しかし、スコアを用いて、調整する根拠になるボックス  $a_i$  の信頼度を保証しても、図 7 のように、ボックス  $a_m$  とボックス  $a_i$  の形状差別が大きい場合、右辺は左方向に修正されるが、ボックス  $a_m$  の特徴量減少が削除される共通面積よりも大きすぎる。

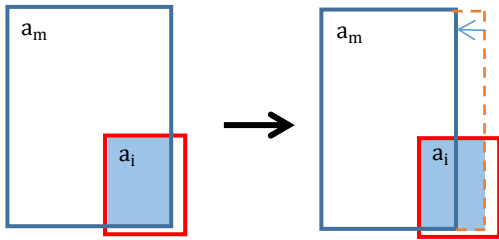


図7 ボックスの過修正

二つのボックスの共通部分はボックス  $a_i$  の高い比率を占有することより、両者の形状には一貫性を持つと考えられる。そこで、制限条件を追加する。共通部分の横と縦とボックス  $a_m$  の横と縦のどちらかの比率は高い閾値  $N_h(0.8)$  を超えると、修正可能である。

$$\frac{W_{inter}}{W_m} > N_h \quad (1.4)$$

$$\frac{H_{inter}}{H_m} > N_h \quad (1.5)$$

修正量は、 $IOU\_OUT(a_m, a_i)$  を低い閾値  $N_{low}(0.3)$  まで減少させるため、共通部分からボックス  $a_m$  だけを修正する面積である。式 1.4 を満たす時に、ボックス  $a_m$  の高さを修正する。式 1.5 を満足するときに、ボックス  $a_m$  の長さを修正する。また、二つの条件同時に成立する場合は、ボックス  $a_m$  の内部にあるボックス  $a_i$  との位置関係が決まらないので、修正できない。

### 3.3 閾値の提案

従来の NMS には、単一固定の閾値を使って、閾値を超える部分を直接削除する。閾値が大きければ、重複検出される結果になる。小さければ、部分ターゲットの未検出問題もある。

本稿では、三つの閾値を使用し、区間に分けて IOU 値を処理する。

① IOU 値が  $N_h(0.8)$  を超えると、直接削除する。  $N_h$  は 3.4 の再検索のデータ量を減少するために固定される。

② IOU 値が  $N_t(0.7 \text{ or } 0.3)$  と  $N_h(0.8)$  の間に、再検索に使う。

③ IOU 値が  $N_t(0.7 \text{ or } 0.3)$  以下の場合、集合 A に残る。

③  $N_{low} = 0.5 * (N_t + 0.3)$ 、再検索に使う。

### 3.4 再検索

3.2 より、最高点数 anchor ボックス  $a_m$  のエッジを修正したので、ボックス  $a_m$  は前と比べて、単一のターゲットを含む可能性が増加する。そして、高密度の対象が未検出になる問題を防ぐために 3.3 の②を満たす全てのボックスとの IOU 値を再判断する必要がある。再判断される IOU 値が  $N_{low}$  より小さくなると、このボックスも集合 B の列に増加する。

ここまでは従来法と一つ同じ問題を持っている。それは判断は必ずスコアの減少する順番で行う。簡単に説明すると、スコアの低いボックスにスコアの低いボックスを含んではいけないが、スコアの低いボックスにスコアの高いボックスを含むことは許す。この問題を解決するために、毎回集合 B 中のボックスを集合 A の末に添付する必要がある。これにより、逆修正もできる。図 8 には新 NMS の流れを示す。

Input:

Anchor の集合  $A = \{a_1, a_2, \dots, a_n, \dots\}$ ;

Anchor 分類スコアの集合  $S = \{s_1, s_2, \dots, s_n, \dots\}$ ;

NMS の閾値  $N_t, N_h = 0.8, N_{low} = 0.5 * (N_t + 0.3)$ .

Output:

後補ボンディングボックスの集合 B;

初期化に  $B = \emptyset$ ;

While  $A \neq \emptyset$  do

    最高点数 anchor ボックスの  $index:m = \arg s_{max}$ ;

    最高点数 anchor ボックス:  $a_m$ ;

$\forall a_i \in A, N_t < IOU(a_i, a_m) < N_h$  を満たす  $a_i$  の集合を集合 C にする。  $C = \{c_1, c_2, \dots, c_n, \dots\}$ ;

    集合 A から、 $del a_m: A = A - a_m$ ;

    集合 B に、 $add a_m: B = B \cup a_m$ ;

    for  $a_i$  in A do

        if  $IOU(a_m, a_i) < N_h$  then

            if  $IOU\_OUT(a_m, a_i) > N_h$  then

$a_m$  のエッジを修正する

                for  $c_k$  in C do

                    if  $IOU(a_m, c_k) < N_{low}$  then

$B = B \cup c_k$

                        Break

                Break

        if  $IOU(a_m, a_i) \geq N_h$  then

$A = A - a$

$S = S - s$

    end

$A = A \cup B$

    if  $a_1 == b_1$ :

        break

end

図8 新 NMS の流れ

4. 実験

4.1 実験環境

ハードウェア:

CPU: Intel(R) Core(TM) i7-8700K CPU@3.70GHz\*12;

GPU: GeForce GTX 1080Ti.

ソフトウェア:

システム: 64x Windows 10 20H2;

言語: Python 3.6.2;

ディープラーニング:

Tensorflow 1.14.0; Cuda: 10.0. Cudnn 7.6.5.

4.2 実験手法

本稿では、Pascal VOC 2007 データ集[3]を用いて、Pascal VOC 挑戦の20種類の検索分類任務で改善されるNMS方法の有効性を検証する。提案手法の評価にあたり、画像検索に一般的に使われる Average Precision(AP)及び Mean Average Precision(MAP)を使用する。

人を予測する問題を考えて、結果は人と人ではない二種類に分ける。正解率(P)は、人に予測される結果の中で、実際の答えも人の比率である。リーコール(Recall)率は実際の答えは人の集合の中で、正確に予測される結果の比率である。リコール率が増加すると、正解率は実際の人の比率まで減少する。P-R 曲線の面積は AP となる。MAP とは、各種類の AP を用いて、平均値を計算することである。

4.3 実験結果

(1) 画像

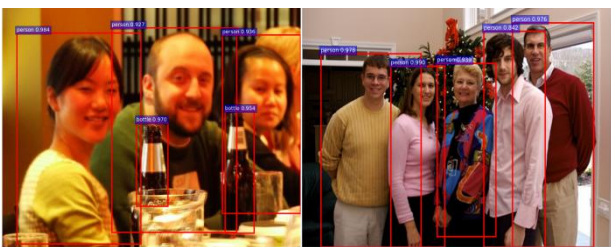


図9 新NMS法の結果

図9に新NMS法で検索する結果を示す。図5の画像と比較して、(a)には、右側にいる人と左にあるボトルを検出できるようになっている。真ん中の人ボックスが左方向に修正され、適切になっている。(b)には、左側にいる人のボックスを左方向に修正された。右側にいる人のボックスを右側に修正することより、前の未検出の男のボックスが検出されている。

(2) AP と MAP

図10より、VOC 2007 データ集の test 集を使い、AP を計算する結果を示す。従来法と比べて、新 NMS 法は各種

類の平均正解率(AP)を向上させた。また、MAP が 7.18% 増加した。



MAP = 0.5661

(a) 従来法

MAP = 0.6379

(b) 新NMS法

図10 AP と MAP

5. まとめ

本稿では、高密度のターゲットが存在する画像に対する新NMS方法を提案した。従来の方と比較して、ターゲットの検出率、正確率と所属範囲の正確性を向上させた。今後は、ネットワークの訓練速度の向上を検討する。

参考文献

[1] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in Neural Information Processing Systems 28 (NIPS 2015)

[2] Bicheng Zhang. A research of target detection and recognition algorithm with region based convolutional neural network[D]. University of Electronic Science and Technology of China(2020).

[3] Pascal VOC. [https://lib-arts.hatenablog.com/entry/dataset\\_ml1\[EB/OL\]\(2019\)](https://lib-arts.hatenablog.com/entry/dataset_ml1[EB/OL](2019)