

第7月 進捗確認会報告資料

変形ARマーカ認識のための手法調査

安井 理

平成32年7月10日

1 はじめに

変形ARマーカの読み取りを最終目標とし、今回までは、前任者の鈴木さんがベースとしていたfaster-rcnnの調査とSSD-6D論文の手法の理解を進め、SSD-6Dでは実装が可能かどうか調べた。結果として、トレーニングコードのコードが乗っていない実装することは、現実的ではなかったため、論文の理解を一通りするところまで行い、次の調査、実装に向けて新しい論文調査をする方針にした。

2 進捗報告

2.1 調査内容

今回調査した手法及び論文は以下である。

- ・ **Faster-RCNN**
- ・ **SSD-6D** Making RGB-based 3D detection and 6D pose estimation great again

2.2 Faster-RCNN

2.2.1 Faster-RCNN を用いた先行研究

先行研究でFaster-RCNNを用いて鈴木さんが行った研究内容として、主に3つである。次のステップとしてARマーカの四隅の座標、回転方向の推定を計算できるようにすることである。

- ・ **Faster-RCNNの正常性の確認** 環境画像に張り付けたARマーカの読み取り
- ・ **Faster-RCNNの応用** アーチ状、旗状、円状、に変化させたARマーカの読み取り
- ・ **研究目的の読み取り** 仮想空間で円柱に張り付けたARマーカの検出

2.2.2 Faster-RCNN とは

大きく特徴マップを抽出する Convolutional Layer と物体領域を抽出する RPN に加え分類, 回帰の結果を出力するネットワークで構成されている. 学習のポイントは, ある矩形が背景なのか物体なのか, 検出されたものが物体だった場合にそれが何かの点である. 特徴的なのが背景か物体かの学習を行う際に RPN (Region Proposal Network) と呼ばれる CNN の構造である.

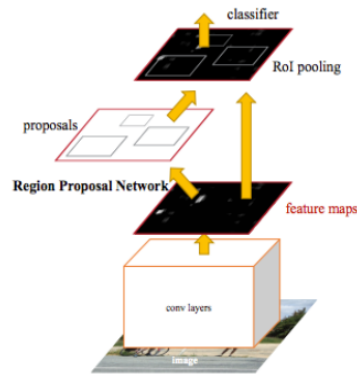


図 1: 二次元姿勢推定.

2.2.3 RPN

Region Proposal Network は画像内から物体領域の候補となる箇所を抽出する. 元々は「Fast RCNN」ではこの物体領域抽出に Selective Search と呼ばれる手法を使っていたため, 物体候補領域の抽出に時間がかかっていたが, その箇所も End-to-End による実装で従来より高速・精度改善される. 主な処理は下記の 3 つである.

- ・ 物体の候補領域を見つける **Anchor** を作成
- ・ (学習時) 誤差を計算
- ・ 候補の中から NMS などを実施して、**Region Proposal Network** での物体の候補領域を抽出

2.2.4 Anchor

Anchor と呼ばれる領域を生成→物体の領域を示している矩形. これらを画像内に一定間隔で生成する. これは学習時に使われる Loss Function で使われる.

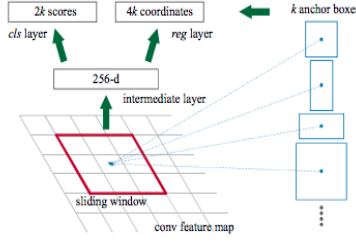


図 2: Anchor.

2.2.5 Loss Function

RPN 自身を更新するための誤差を計算し, ニューラルネットワークの全体の誤差と併せて更新される. 以下の式で表される.

$$L(p_i, t_i) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i L_{reg}(t_i, t_i^*) \quad (1)$$

RPN の誤差は基本的には 2 つの要素で成り立っている, その領域は物体か背景であるかどこに領域があるかの領域の位置 (詳細) を計算. のバウンディングボックスの決定は Anchor と物体の IoU が 0.7 以上の場合物体であることを示す. IoU が 0.3 以下では物体が存在しないことを示す.

2.3 SSD-6D

2.3.1 SSD-6D とは

ディープラーニングの分野で画像の分類と検出は大幅に進歩しているが 6 次元の姿勢推定はまだ伸びしろのある分野である, 座標認識と物体の回転座標の認識は自分の研究テーマにおける変形 AR マーカの推定に必要なものであり, 今回 2017 年の 6 次元物体認識のコンテスト (ECCV 17) で Best Paper Award をとった論文の手法を調べていくことにした. 6 次元の姿勢推定とは 3 次元空間の X,Y,Z,3 座標にそれぞれ回転軸を加えたものであり, 6 自由度の方向推定である. 6 次元姿勢推定はすべて合成モデルデータのみでトレーニングを行い RGB 画像のみで推定をすることができる, SSD を拡張したもので, 複雑なデータを扱う方法に匹敵するかそれ以上の精度の認識を行えるもの.

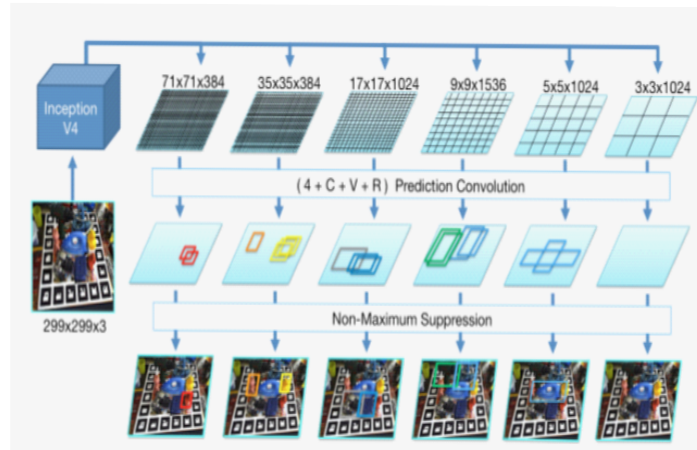


図 3: SSD-6D 全体.

2.3.2 ネットワーク

299*299のカラー画像を入力とし, 事前にトレーニングされた InceptionV4 にかへ, その後複数の縮尺のフィーチャーマップを用意し, $[71*71*84 \Rightarrow \text{Inception} \cdot A]$ $[\text{Inception} \cdot A * 35*35*384 \Rightarrow \text{Inception} \cdot B]$ $[\text{Inception} \cdot B * 17*17*1024 \Rightarrow \text{Inception} \cdot C]$ $[\text{Inception} \cdot C * 9*9*1536]$ と小さい特徴から順に大きい物の特徴へと特徴を抽出していく。より大きな特徴を抽出するために $5*5*1024$ のフィルタと $3*3*1024$ のフィルタにさらに通していく。

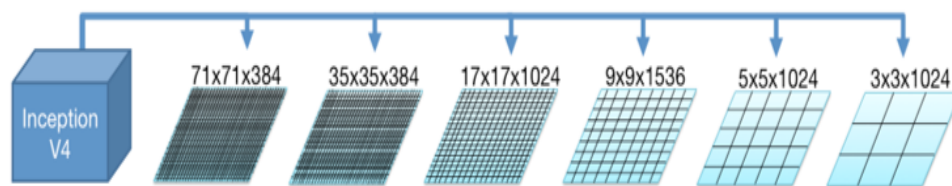


図 4: SSD-6D のネットワーク.

2.3.3 検出

検出特徴を得たらそれぞれの特徴マップをトレーニング済みの予測カーネル $C \cdot 4 \cdot V \cdot R$ (オブジェクトクラス・バウンディングボックス・視点・画内回転) に通す. 全結合層の部分に値する場

所である。特徴マップからまず、位置の推定予想が行われる、想定するのは物体の高さ、幅、チャンネルの深さの3つを求めバウンディングボックスを推定、そこから物体のID(分類)、物体の視点（画内回転）、ボックスの一致度を評価、バウンディングボックスは各、ピクセルに高さ幅3*3の9つの大きさのboxを用意し最も一致度の高いバウンディングボックスを選択する。

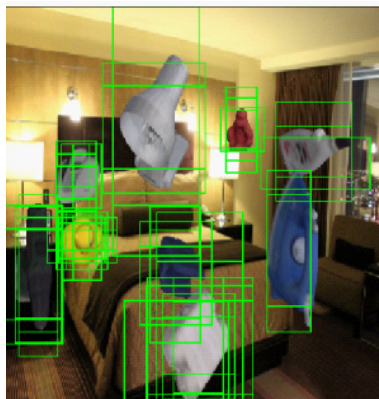


図 5: ボックス検出.

2.3.4 推定

ネットワークで出力されたボックスの特定のしきい値を超えた検出を収集してからハードネガティブマイニングを行う、これにより適切なバウンディングボックスが選択される。これらの選ばれたboxから分類・画内回転のそれぞれのIDをもとに6次元の推定を行う。事前に計算した物体の視点ごとのバウンディングボックスの対角線を l_r 、ネットワークで予想されたバウンディングボックスの対角線を l_s として物体との距離を $z_r=0.5$ mで計算をする。 $z_s = l_r/l_s * z_r$ と表現でき用意されたカメラでインライン関数を用いて投影の重心位置を推定し逆投影することが可能になる

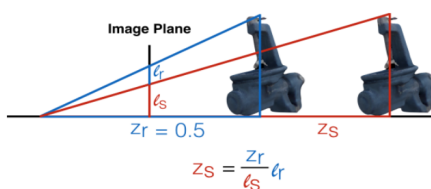


図 6: SSD-6D 推定.

2.4 終わりに

今後は実装していけるように現実的に使える論文を読んで理解し進めていく。ペースが少し遅いのでスピードアップもしていきたい。次は RGB データを用いた 6 次元の物体推定の論文が使えるか調べてから次に、何をするのか考え進めていく。

参考文献

- [1] SSD-6D, <https://openaccess.thecvf.com/>
- [2] Faster-RCN, <https://medium.com/lsc-psd/faster-r-cnn/>
- [3] Faster-RCN, <https://nonbiri-tereka.hatenablog.com/>