

1. はじめに

顔画像には、目や口などの器官点、年齢や性別などの属性を含め、多くの情報が含まれている。従来、これらの情報を得るためには、顔検出、器官点検出、属性推定と、順に各タスクを処理する。そのため処理時間がタスク数に応じて増えるという問題がある。そこで、本研究では、高速な物体検出手法である Single Shot MultiBox Detector(SSD) をマルチタスク化し、これらのタスクを同時に行う手法を提案する。また、従来の SSD では傾いた矩形を出力できないため、傾いた顔に対する認識精度が低下する恐れがある。そのため、顔角度推定タスクの追加を検討する。

2. Single Shot MultiBox Detector

Single Shot MultiBox Detector(SSD)[1] は物体候補領域と物体クラスを同時に推定する手法である。複数解像度の特徴マップを用いて、様々なスケールの物体を検出することができる。予測レイヤでは、デフォルトボックスと呼ぶ矩形に対するオフセットを特徴マップ (Localization maps) として出力する。同様に予測レイヤで得られる物体のクラス尤度を表す特徴マップ (Confidence maps) を用いて、物体のクラスを識別する。その後、Non-maximum Suppression を行い最終的な検出を行う。

3. 提案手法

本研究では、顔検出に加えて、顔器官点検出と顔属性推定の学習および推定を同時に行う。

3.1. 提案するネットワーク構造

提案手法のネットワーク構造を図 1 に示す。顔検出と顔器官点検出は異なるタスクであるが、共に位置情報を出力するタスクである。そこで、本研究ではこれらを共通な Localization map(図 1 の青色マップ) として出力する。また、各デフォルトボックスが顔であるか否かのクラス尤度を Confidence maps(図 1 の緑色マップ)、各顔属性を Attribute maps(図 1 の赤色マップ) として同時に出力する。また、顔角度推定を行うため Localization map を拡張し、角度推定タスクを可能とする。

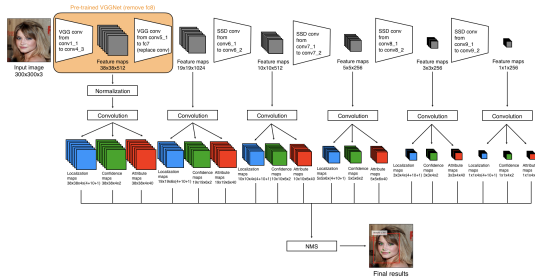


図 1：顔器官点検出と顔属性推定を含んだ SSD

3.2. ネットワークの学習

提案手法のネットワーク全体の学習誤差関数 L_{all} を式 (1) に示す。ここで、 L_{bbox} 、 L_{land} 、 L_{conf} 、 L_{attr} 、 L_{angle} はそれぞれ、顔検出、顔器官点検出、クラス識別、顔属性推定、角度推定の学習誤差関数を表す。 N は教師信号と一致した顔領域の数を示している。また、 $L_{loc} = L_{bbox} + L_{land}$ である。

$$L_{all} = \frac{1}{N}(L_{loc} + L_{conf} + L_{attr} + L_{angle}) \quad (1)$$

L_{bbox} と L_{land} 、 L_{attr} 、 L_{angle} は、学習誤差関数に式 (2) より Smooth L_1 誤差を使用する。

$$L_* = \sum_i \sum_m x_{ij}^{smooth} (y_i^m - \hat{t}_i^m) \quad (2)$$

ここで、 y を出力結果、 t を教師信号とする。 m は中心 cx, cy 、高さ h 、幅 w のパラメータであり、 K はカテゴリ数である。 $x = \{0, 1\}$ は、顔領域として推定された領域と教師信号の一致を示している。また、 L_{conf} はクロスエントロピー誤差を用いる。

4. 評価実験

本実験では、CelebA Dataset[2] を使用し、顔検出・両目・鼻・口角の 5 点の顔器官点検出と、男性・黒髪などの 40 種類の顔属性推定、顔角度推定のタスク対象とする。

4.1. 実験概要

学習サンプルに 162,770 枚、評価サンプルに 19,962 枚使用する。入力画像サイズは 300×300 、学習回数は 120,000 回とする。最適化手法はモーメント付の SGD とする。また、顔の角度推定を行うにあたり、CelebA Dataset の顔画像を回転させるデータオーグメンテーションを行う。比較手法として、顔検出のみを学習した SSD を用いる。

4.2. 評価結果

SSD を用いた顔検出、顔器官点検出、顔属性推定、顔角度推定の推定結果を表 1 に示す。表 1 より、顔角度を推定しない提案手法は従来手法よりも顔検出、顔器官点認識、顔属性推定の精度の向上が確認できる。一方で、角度を推定する場合、各タスクの精度が低下している。これは角度誤差が 11.0 度と大きいため、他のタスクに悪影響を及ぼしていると考えられる。また、処理時間の比較結果を表 2 に示す。表 2 より、顔検出のみの処理時間が処理時間が 0.023[sec] となった。器官点検出を追加した場合、0.03[sec] の処理速度の増加でタスクを増やすことができた。また、属性推定、角度推定を追加した場合、0.02[sec] の処理速度の増加でタスクを増やすことができた。

表 1：SSD の認識結果

タスク	従来手法	提案手法	
		角度なし	角度あり
顔検出 [%]	90.3	90.5	86.2
平均顔器官点検出誤差 [pix]	—	4.2	5.4
顔属性認識 [%]	—	86.2	85.5
平均角度誤差 [度]	—	—	11.0

表 2：処理時間の比較結果

	顔検出	顔検出 + 顔器官点検出	顔検出 + 顔器官点検出 + 顔属性推定	顔検出 + 顔器官点検出 + 顔属性推定 + 顔角度推定
処理時間 [sec]	0.023	0.026	0.026	0.028

従来手法と提案手法の推定結果を図 2 に示す。ここで、顔検出は青枠で表示してあり、顔器官点検出を緑点で表示してある。図 2 から、従来手法と角度なし提案手法を比べると顔検出はほぼ同等の精度を出力している。また、提案手法を比較すると顔検出、顔器官点検出共に精度が低下していることがわかる。

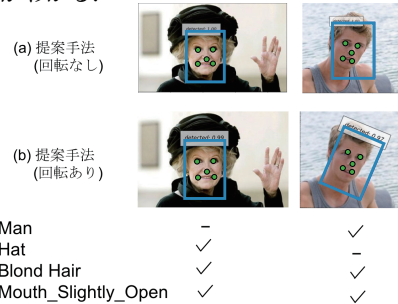


図 2：従来手法と提案手法の比較

5. おわりに

本研究では、SSD を用いて顔検出、顔器官点検出、顔属性認識、顔角度推定を行った。顔角度を推定しない場合、従来手法と比べて顔検出が 0.2% 向上している。顔角度を推定した場合よりも各タスクにおいて同等以上の精度を出している。また、顔検出のみの場合が最も処理時間が短かったが、SSD をマルチタスク化したことによって他タスクを追加した時の処理時間の増加が抑えられている。今後は、顔角度推定を含んだ SSD の精度の向上に取り組む。

参考文献

- [1] W. Liu, et al., “SSD : Single Shot MultiBox Detector”, ECCV, 2016.
- [2] Z. Liu, et al., “Deep Learning Face Attributes in the Wild”, ICCV, 2015.