

上半身検出手法とその評価

浅利広織 中島克人

東京電機大学未来科学研究科

1. はじめに

カメラを用いた人検出は、車載カメラによる自動車停止システムや、監視カメラによる挙動解析等、様々な応用で注目されている。

人検出器として、主に顔検出器や全身検出器が挙げられる。しかし、顔検出器では横顔や後ろ姿の検出が困難である。また、全身検出器ではオクルージョンに弱く、混雑した状況での検出が困難となる。そこで、これら問題を軽減するためには人の上半身を検出することを考える。

人検出器等に多用される HOG[3]特徴と SVM による上半身検出器は藤井ら[1]が既に提案している。

一方、識別器を構成する手法とし、近年、ニューラルネットワーク(NN)の一種である多層パーセプトロン(MLP)の派生形である Deep Learning が注目されている。画像識別では、Deep Learning の考え方に基づく Convolutional Neural Network (CNN)[2]が主流であり、これを用いた高精度な識別器が数多く報告されている。

今回、CNN による上半身検出器を構築・評価し、藤井らによる HOG 特徴量と線形 SVM による上半身検出器と比較したので報告する。

2. HOG 特徴による上半身検出器

藤井らによる上半身検出器の概要は以下の通りである。

検出対象の上半身画像は「人物の頭頂部が画像の上辺から縦幅の 1/8 の位置にあり、かつ、頭部が画像の縦幅の約 3/5、横幅の約 1/2 の領域に収まっている」と定義する。

識別に用いる特徴量として、HOG に加えて CSS[4]を使用する。HOG により人のシルエットを捉え、CSS により領域ごとの色の類似度を評価することで、上半身の形状・色の関係を捉えている。

識別には線形 SVM を使用する。線形 SVM は、正例・負例を分類する超平面を、両データ間のマージンが最大になるように決定する識別器である。

3. CNN による上半身検出器

3.1 CNN の概要

CNN は、畳み込み処理を取り入れることで画像の識別精度を向上させる NN の一種である。一般的な CNN では、畳み込み層が 2 から 3 層続いた後に、隠れ層が数層続き、ロジスティック回帰層で結果を出力する。

○各層の説明

畳み込み層では、入力画像に対する畳み込みが行われた後に、画像の部分サンプリングとマックスプーリング処理が行われる。

畳み込みはニューロンの重み値 w を用いて、ニューロンの数だけ行われる。

部分サンプリングでは、畳み込みの結果画像を、 $N \times N$ 個の小領域に分割する。その後、 $N \times N$ 枚の画像の中で最も応答値の高いものを次の層へ渡す(マックスプーリング)。これにより、ネットワークの複雑さを軽減しつつ、検出対象の画像の位置に対する不変性を得ることが出来る。

SVM による識別器構築では HOG や CSS 等の特徴量を認識対象依存で設計する必要があるが、CNN ではこのように特徴量が内部で自動的に設計されるため、識別対象に関する事前知識を必要としない。

隠れ層では、それまで画像として扱ってきた入力を数値列として扱い、パラメータの学習を行う。

ロジスティック回帰層では、入力を用いてクラスごとに尤度を計算し、最も高いクラスを出力とする。

○モデルパラメータの更新アルゴリズム

NN のパラメータ更新には確率的勾配降下法を用いる。具体的には、学習データのサブセットを用い、学習率に基づいてパラメータ更新を行う。

○学習の手順と停止条件

学習セットの一部を途中評価用セットとして取り除き、残りを多数の学習単位に分割する。学習はこの学習単位ごとに繰り返し行う。全ての単位の学習(学習の 1 世代)が終わると途中評価用セットで誤識別率を求め、以下の 2 つの停止条件のどちらかを満たさなければ、再度同じ学習セットを用いて次の世代の学習を行う。

1. 規定の NN パラメータ更新回数を超えた場合。ただし、一定以上の誤識別率の改善が見られる場合、更新回数を一定数増加させる(Early Stopping)。
2. 最大世代数に達した場合。

3.2 上半身検出のための CNN の構成

今回の CNN の構成とパラメータ、学習時の学習率や停止条件等のパラメータを以下に示す。

表 1 に NN の各層の種類と使用する活性化関数を示す。表内の Softmax 関数は式 1 で定義される。ここで i は i 番目のニューロンを示すクラスインデックス、 W , W_i , W_j はニューロンの重み値、 b_i , b_j はニューロンのバイアス値である。

表 2, 表 3 に NN への入力画像サイズ、層ごとの

畳み込みフィルタの数とそのサイズ，部分サンプリングによる縦・横方向への分割数を示す。

表4に学習率，一度の更新で使用する画像の枚数，使用パラメータの更新回数（初期値），更新回数を増加させる倍率，最大世代数を示す。

表5と表6に学習と性能テストに用いた画像の枚数と，CNNの学習の経過をそれぞれ示す。

表6において，17世代で誤識別率が最小，F値が最大となるが，最大適合率は16世代，最大再現率は8世代である。適合率，もしくは再現率に応じてパラメータ更新回数の増加を図ることで，どちらかに特化した識別器を構成できると考えられる。

なお，今回はPythonで実装された機械学習用ライブラリTheano[5]を用いてCNNの構築を行った。

表1 NN構成

段数	種類	活性化関数
1	畳み込み層	Tanh
2	畳み込み層	Tanh
3	隠れ層	Tanh
4	ロジスティック回帰層	Softmax

$$Softmax_i(Wx + b) = \frac{e^{W_i x + b_i}}{\sum_j e^{W_j x + b_j}}$$

式1 Softmax関数

表2 畳み込み層パラメータ1

層	入力画像サイズ(pixel)	フィルタ数
1	64×64	20
2	30×30	50

表3 畳み込み層パラメータ2

層	フィルタサイズ(pixel)	部分サンプリング分割数
1	5×5	2×2
2	5×5	2×2

表4 学習パラメータ

学習率	0.1
1回の更新に用いる画像枚数	100
NNパラメータの更新回数（初期値）	10,000
パラメータ更新回数増加倍率	2
最大世代数	200

4. 識別器の性能比較

HOG特徴量と線形SVMによる上半身検出器と，CNNによる上半身検出器を，識別性能指標の1つである適合率・再現率・F値について比較を行った。線形SVMの学習と評価には，CNNと同様に表5の画像データを用いた。

表7にテストセットを用いて識別した結果を元に計算した性能指標の比較を示す。

適合率はSVM識別器が約0.03だけ上回るが，再現率はCNN識別器が約0.2上回り，総合的な指標のF値について0.1以上良い結果となった。

表5 画像データ枚数

	正例	負例	総数
学習画像	24,000	76,000	100,000
テストセット	4,000	36,000	40,000

表6 CNN学習経過

世代	誤識別率	適合率	再現率	F値
1	14.96	0.818	0.220	0.346
2	8.01	0.679	0.594	0.634
3	6.59	0.545	0.789	0.645
4	5.27	0.610	0.826	0.702
5	4.06	0.666	0.839	0.742
6	3.78	0.663	0.854	0.747
7	3.50	0.682	0.863	0.762
8	3.39	0.701	0.872	0.777
9	2.73	0.757	0.849	0.800
10	2.87			
11	2.53	0.814	0.834	0.824
12	2.4	0.841	0.822	0.832
13	2.53			
14	2.17	0.874	0.817	0.845
15	2.21			
16	2.14	0.892	0.803	0.845
17	1.98	0.871	0.831	0.851
33	2.00			

Early Stoppingにより停止

表7 識別性能

	HOG・SVM	CNN
適合率	0.903	0.872
再現率	0.638	0.831
F値	0.748	0.851

5. 今後の課題

今後の課題として，以下が挙げられる。

○交差検定による再評価

今回データセットの一部を分離したテストセットにより評価を行ったが，それらを学習データに含め，交差検定での性能向上を検証する必要がある。

○識別速度評価

最終的には最終的に実時間で上半身検出を行うことを目標としているため，それぞれの検出器の速度評価と，必要ならば高速化の検討が必要である。

参考文献

- [1] 藤井龍也，他，“HOGとSVMによる上半身検出器の特徴の抽出位置に関する考察”，第10回情報科学技術フォーラム(FIT2011)，H-001，2011.
- [2] Lecun.Y, Bottou.L, Bengio.Y, Haffner.P, “Gradient-based learning applied to document recognition”, Proceedings of the IEEE, Volume:86 Issue:11, Page:2278-2324, 1998.
- [3] N.Dalal, et al., “Histograms of Oriented Gradients for Human Detection”, Proc.CVPR, vol.1, pp.886-893, 2005.
- [4] S. Walk, N. Majer, K. Schindler, and B. Schiele, “New features and insights for pedestrian detection”, pp.1030-1037, In CVPR, 2010.
- [5] Theano
<http://deeplearning.net/software/theano/>