回帰型 Deep Convolutional Neural Network による 人検出と部位の位置推定

山下 隆義十 福井 宏十 村瀬 将之十 山内 悠嗣十 藤吉 弘亘†

†中部大学

E-mail: yamashita@cs.chubu.ac.jp

Abstract

運転支援システムにおける歩行者検出は、歩行者の 位置を検出するだけでなく、歩行者と車までの距離を 推定することが重要となる. 一般的な単眼カメラによ る距離推定は、歩行者を検出した後に、上端と下端を検 出する.そして、検出した上端と下端の位置情報とカメ ラの内部パラメータを用いて歩行者と車までの距離を 推定している。そのため、歩行者の検出、上端および下 端の検出,距離の推定を個別に処理する必要がある.そ こで、本研究では Deep Convolutional Neural Network (DCNN)を用いて、歩行者を検出すると同時に、頭部 と両足の位置を高精度に推定する回帰型 DCNN を提案 する.歩行者検出と頭部および両足の位置検出を同時 に行うことで、歩行者部位の位置情報を同時に与える ことができるため、従来の DCNN より歩行者検出に適 した特徴表現が可能となる.実験では、歩行者検出の みを行う手法との精度比較により、歩行者検出精度を 向上させることができ, 頭部および両足の位置検出精 度についても、位置検出のみの手法と比較して、精度 を向上させることができた.また,歩行者の位置検出 結果から、カメラの内部パラメータが既知の画像に対 して距離推定を行ったところ、約5%の誤差で距離推定 を行うことができた.

1 はじめに

多くの交通事故は,自動車と歩行者による事故が大 半を占めており,自動運転や高齢者の運転支援システ ムにおいて,歩行者を検知して回避または危険を知ら せるが重要となる.より高度な運転支援システムのた めには,歩行者検出だけでなく,歩行者までの距離を 計測する必要がある.歩行者までの距離を計測する方 法として,単眼カメラを用いる方法,ステレオカメラ を用いる方法,LIDARを用いる方法がある.LIDAR は照射したレーダーが物体に当たって戻ってくるまで の時間を計測する.照射するレーダーの本数に限界が あるため解像度が低い.また,LIDARは非常に高価で

あり、一般車への普及が進んでいない。一方、ステレ オカメラを用いた方法は、2つのカメラの視差から距 離を推定している。一般車への搭載が進みつつあるが、 装置の小型化に限界がある。単眼カメラを用いる方法 は、まず歩行者検出を行い、その後頭部および足下の 位置を推定する。そして、カメラパラメータと歩行者 の上端および下端の情報から距離を推定する。カメラ は安価であり、装置としての小型化が可能であるが、各 処理が個別に行われており、他の手法と比較して精度 が悪いため、歩行者検出精度および距離推定精度の向 上が求められている、そこで、本研究では、歩行者検 出と頭部および両足の位置推定を高精度に行うために, 歩行者を検出すると同時に,歩行者の上端と下端の位 置を同時に推定する回帰型 Deep Convolutional Neural Network(DCNN)を提案する. DCNN[1]は、歩行者検 出をはじめ、様々なベンチマークにおいてトップレベル の性能を達成している [2]. さらに、1 つのネットワー クで複数の認識を同時に行うことができる. Zhang ら は DCNN を用いることで顔器官点の検出に加え、顔の 向き、眼鏡の有無、笑顔の有無、性別を同時に認識し ている [3]. 我々は DCNN を用いることで歩行者検出 とともに、歩行者の頭部と両足の位置を回帰により検 出するネットワークを実現する. DCNN により得られ た歩行者の上端および下端の情報を用いて、カメラか ら歩行者までの距離推定行う.

2 関連研究

歩行者までの距離を計測する方法として、単眼カメラ を用いる方法、ステレオカメラを用いる方法、LIDAR を用いる方法がある.LIDAR を用いる方法は、レー ザを照射することで3次元点群を取得し、取得した3 次元点群の形状から歩行者などの物体を認識している. LIDAR は、レーザにより3次元点群を取得しているた め、天候の変化等に影響を受けづらいメリットがある. しかし、LIDAR は非常に高価なため、一般車への搭載 が大きな課題となっている.

ステレオカメラを用いる方法は,複数のカメラから 得られる視差マップを用いて物体の位置や立体的な形 状を認識する.車載としてステレオを用いる場合,2つ のカメラを車体前方に設置し,2つのカメラの視差から 歩行者を検出している.Zhaoらは,単眼カメラから取 得した画像とステレオカメラから取得した視差マップ を用いて,歩行者のエッジを取得し,取得したエッジ を特徴量として多層パーセプトロンに入力することで 高精度かつリアルタイムでの歩行者検出を実現してい る[4].しかし,ステレオで用いている視差は,それぞ れの画像上の共通する物体の位置の違いから生じてお り,天候の変化等で物体の見えが大きく変化した場合 に認識が困難になる.

一方,単眼カメラを用いて歩行者までの距離を推定す るためには、まず歩行者の検出を行い、その後歩行者の 上端および下端の位置を検出する必要がある。その後、 カメラパラメータと上端および下端の情報から距離を 推定する。歩行者検出は、Dalal らが提案した HOG 特 徴量により精度が大きく向上し [5]、その後、HOG 特 徴量を改良した手法が提案されている [6][7][8]. 人の姿 勢変化に対応する方法として、歩行者の全身とパーツ を同時に推定することができる DPM が提案されてい る [9].

DCNN[1]は、学習処理において識別処理に適した特 徴量を自動獲得できる点が注目され、歩行者検出をは じめ、様々なベンチマークにおいてトップの性能を出 している [2]. Sermanet らは, Sparse Auto Encoder を DCNN の学習処理に対して導入し、サブサンプリング した特徴マップを識別部の全結合層に入力することで, 高精度な歩行者検出を実現している [10]. また,同年に 歩行者の各パーツのスコアを DCNN で求めた後に,各 パーツのスコアと特徴マップを用いて歩行者を認識す る階層型ニューラルネットワークを用いた Joint Deep Learning が提案されている. Joint Deep Learning は歩 行者の各パーツに対して考慮することで、様々な姿勢に 対して頑健な検出が可能となっている。さらに、DCNN は、出力層のユニットをそれぞれの認識で振り分ける ことで、複数の認識を同時に行うことができる. Zhang らは DCNN を用いることで顔器官点の検出に加え,顔 の向き、眼鏡の有無、笑顔の有無、性別を同時に認識 している [3]. そこで, 我々は DCNN を用いることで 歩行者の検出をはじめ、歩行者の頭部と両足を回帰に より検出・推定を同時に行う.

単眼カメラで取得した画像から,カメラと歩行者の 距離を推定するには歩行者の上端と下端を推定する必 要がある.岸野らは,歩行者を検出した後に,投影法 を用いることでカメラと歩行者の距離を推定している [11].我々は,DCNNにより得られた歩行者の上端お よび下端情報を用いて,投影法によりカメラから歩行 者までの距離を推定する.

3 提案手法

我々は、歩行者検出と同時に頭部および両足の位置を 推定するために、検出と位置推定を全て回帰により求め る回帰型 DCNNを提案する。回帰問題に用いる DCNN の構造および学習について、以下に述べる。

3.1 回帰問題に用いる DCNN

図1に示すように、DCNN は畳み込み層とプーリン グ層が階層的になっており,その後に全結合層が上位 層に続いている。入力としては、RGB 画像やグレース ケール画像、勾配画像、または正規化処理を施した画 像が与えられる。与えられた画像に対して、畳み込み 層では、サイズ $k_x \times k_u$ のフィルタを畳み込む. 畳み 込んで得られた値xは活性化関数f(x)を通した後,特 徴マップに格納される。 畳み込み層のフィルタは M 個 あり、各フィルタからそれぞれ特徴マップを作成する. 活性化関数には、シグモイド関数や Rectified Linear Unit(ReLU), Maxout が用いられる. シグモイド関数 はxが大きくなると出力 f(x)が一定の値に留まるため, 学習する際に勾配が得られないという問題があった. 一 方, Krizhevsky らが提案した活性化関数 ReLU は,式 (1)のように、*x*が0以下であれば0とし、0以上であ ればその値である f(x) を出力している.

$$f(x) = \max(0, x) \tag{1}$$

これにより, xが大きな値となった場合に, その値を そのまま出力することができ, 勾配が得られないとい う問題を解決している. Maxout は, 複数のフィルタで 畳み込んだ値 h_k の最大値 h' を次の層の特徴マップに 格納する活性化関数である [12]. フィルタが M 個ある 場合, それらを式 (2) に示すように, 特徴マップを Kごとに分けて, それぞれで最大値 h' を選択する.

$$h' = \max_{k \in [1,K]} h_k \tag{2}$$

Maxout は、複数のフィルタの畳み込み値から値を 選択することで、区間凸関数として表現することがで き、他の活性化関数より表現力が高くなっている。我々 は、DCNNの汎化性能を向上させるために Maxout を 用いる。

次に、プーリング層において特徴マップを縮小する 処理を行う. Pooling には、Max Pooling や Average Pooling などの種類がある. Max pooling は、 $2 \times 2 \sigma$ ようなあらかじめ決めた領域における最大値により間 引きを行う方法である. 一方、Average pooling は、あ らかじめ決めた領域における平均値で縮小を行う方法 である. このように、Pooling には様々な手法が提案さ れているが、中でも Max Pooling が最も性能が良いと されている.



図 1 Convolutional Neural Network の構造

畳み込み層とプーリング層は階層的になっており、こ れらを繰り返すことで深いネットワーク構造を形成し ている.これらの階層的な処理の後に従来のニューラ ルネットワークと同様の全結合層が続いている.全結 合層は、1層前の特徴マップを1次元にして入力層と している.入力された値は式(3)のように、重み付きの 全結合を行う.そして、畳み込み層と同様に活性化関 数f(.)から得られた値を出力値 $h_i(x)$ としている.

$$h_i(\boldsymbol{x}) = f\left(\sum_{j=1}^N w_{ij}x_j + b_i\right)$$
(3)

認識問題に用いる DCNN の場合,出力層のユニット 数は認識対象のクラス数となり,特定のユニットは1 に近くなり,そのほかのユニットは0に近い出力値を クラス確率として出力する.一方,回帰を行う DCNN の場合,出力層のユニット数は,人らしさと背景らしさ に相当するユニット,頭部および両足のx座標とy座 標に対応するユニットの計8つとなる.各ユニットは 0から1の値を出力し,それらを画像サイズで乗算す ることで,画像上の座標値を得ることができる.

認識問題に用いる DCNN との違いは,出力層のユニットの出力値の算出方法である.認識問題の場合は,入力画像 x に対する各クラスの確率 p(x) をソフトマックス関数を通して得る.

3.2 DCNN の学習

DCNN を構成する要素は畳み込み層のフィルタと全結合層の結合重みおよびバイアスである。多層の構成にした場合、これらの要素の数は非常に膨大となり、最適なパラメータ群を一意に決めることは困難である。そのため、逆誤差伝播法を用いて、これらの要素を繰り返し更新しながら最適な値を求める。その際、初期値は乱数で初期化されている。逆誤差伝播法は、まず式(4)のように各学習サンプルの誤差 E_m を累積して全体の誤差 Eを算出する。

$$E = \frac{1}{2} \sum_{m=1}^{M} E_m$$
 (4)

 ${m|1,...,M}$ は学習サンプルである. 誤差の算出に は,式(5)のような二乗誤差を用いる. 二乗誤差は,各 クラスkの教師信号 t_k と出力値 y_k から誤差を算出し, 全ての出力ノードに対する誤差を求めている.

$$E_m = \sum_{k=1}^{C} (t_k - y_k)^2$$
 (5)

ここで, C は出力層のノード数を示している。逆誤 差伝播法は,勾配法により全体の誤差 E を最小とする ようにフィルタの値や重み等のパラメータを更新する。 更新された各パラメータは式 (6) に示すように,誤差 E の偏微分から求める。

$$w_{ji}^{(l)} \leftarrow w_{ji}^{(l)} + \Delta w_{ji}^{(l)} = w_{ji}^{(l)} - \lambda \frac{\partial E}{\partial w_{ji}^{(l)}} \tag{6}$$

 λ は学習における更新率, $w_{ji}^{(l)}$ は l 番目の層のノート i と次層のノード j との結合重みである。更新量および勾配は,式(7) と式(8) から求めることができる。

$$\Delta w_{ji}^{(l)} = -\lambda \delta_k^{(l)} y_j^{(l-1)} \tag{7}$$

$$\delta_k^{(l)} = e_k f(V_k^{(l)}) \tag{8}$$

$$V_k^{(l)} = \sum_j w_{kj}^{(l)} * y_j^{(l-1)}$$
(9)

 $y_j^{(l-1)}$ は (l-1)番目の層のノード j であり, e_k は ノード k の誤差, $V_k^{(l)}$ は前層の全てのノードからの重 み付き累積の値である。各要素の値は、あらかじめ決 められた回数または収束条件を満足するまで繰り返し 更新される。誤差 E を算出するための学習サンプルの 与え方として, full-batch, online, and mini-batch があ



図 2 Dropout による全結合層

る. Full-batchは、全ての学習サンプルを用いて1回の 更新を行う方法である.この方法は、勾配の変化が大き いため収束しにくい. Online は学習サンプルを1枚ず つ与え逐次更新する方法である.この方法は、勾配の変 化が小さいため最適解に得やすいが、非常に多くの更 新回数を要するため処理時間がかかる. Mini-batchは、 これらの中間的なアプローチであり、少量のサンプル を用いて1回の更新を行う.この方法は、更新に要す る処理時間が比較的短く、繰り返し行えるため、十分 な勾配の変化を得ることがでる.そのため、DCNNの 学習において、Mini-batchは一般的な方法として利用 されている.

DCNNの学習は、逆誤差伝播法により学習誤差が小 さくなるようにネットワーク全体のパラメータを更新 することで、最適なネットワークを学習する.DCNN の汎化性能を向上させるための方法として Dropout が ある.Dropout は、全結合層のユニットの一部の応答 値を0にして、ネットワークの結合を取り除いて学習 を行う.図2に示すように、t層目の応答値の一部が0 となるが、学習の繰り返し回数ごとで0となるノード が変わる.あらかじめ決めた割合のノード数だけ0と するノードをランダムに選択する.割合は50%とする ことが多い.Dropout は、多くのノードとの結合を除 くため、誤差が下位層まで伝播しやすくなる.そして、 一部の結合が無くても認識できるようにネットワーク 全体のパラメータが更新されることで汎化性能が向上 している.

3.3 歩行者までの距離算出

単眼カメラで撮影した画像から距離を推定する方法 として,対象の大きさを利用する方法と位置を利用す る方法がある.大きさを利用する方法は,歩行者の身長 によって画像上での大きさに個人差が生じるため,安 定した値を得にくい.本稿では,位置を利用する方法 により距離を推定する.その際,足の下端点は地面に 接地しているとする.カメラの高さ*h*を固定し,地面



とカメラは平行であると仮定した場合,カメラから歩 行者までの距離を *d* とすると,画像上の y 座標は,以 下のようになる.

$$y = \frac{fh}{d} + \frac{H}{2} \tag{10}$$

ここで f は焦点距離, H は画像の縦サイズである.こ れより,距離は式(11)となる.

$$d = fh\frac{2}{2y - H} \tag{11}$$

4 評価実験

提案する回帰型 DCNN の有効性を確認するために評 価実験を行う.本章では、歩行者検出の精度と歩行者部 位検出の精度,実シーンにおける身長推定精度の評価に 分けて実験する.歩行者検出と歩行者部位検出の評価に は、図3(a)の車載カメラから撮影されたデータセットで ある Daimler Mono Pedestrian Dataset(以下 Daimler データセット)と、図3(b)の歩行姿勢のバリエーション が多い INRIA Person Dataset(以下 INRIA データセッ ト)を用いる. Daimler データセットは、学習データとし て、ポジティブサンプル 31,320 枚とネガティブサンプル 254,356枚,評価データとして 21,790枚のサンプルが含 まれている. DCNN の学習には, Data Augmentation によりポジティブサンプルを 250.560 枚に拡張したも のを使用する. INRIA データセットは、学習データと して、ポジティブサンプル 2,100 枚とネガティブサン プル 50,000 枚あるサンプルを, Daimler データセット と同様にポジティブサンプルを Data Augmentation に より 50,000 に拡張したものを学習に使用する。評価に は、ポジティブサンプル 1,000 枚とネガティブサンプル 9,000 枚を使用する.

歩行者検出の評価実験では、歩行者検出のみを行う DCNN と比較を行う. この DCNN は2つの出力ユニッ トがあり,認識問題と同様にそれぞれ歩行者と歩行者以 外の確率を出力する.最終結果は確率の高いクラスとな る.また、歩行者部位検出の評価実験では、部位検出の みを行う DCNN と比較を行う.この DCNN は6つの 出力ユニットがあり、各部位のx 座標およびy 座標を出 力する.提案手法および比較手法の DCNN の構成を表 1に示す.出力ユニット数以外の構成は提案手法と比較

(a)Daimler データセット														
手法	Input		Layer1			Layer2		Layer3		Layer4	Layer5	Layer6	Output	
		Conv	Max Pooling	Maxout	Conv	Max Pooling	Maxout	Conv	Max Pooling	Maxout	# of unit	# of unit	# of unit	
検出用	96x48x1	8,5x3	2x2	2	16,5x4	2x2	2	32,5x4	2x2	2	1,000	500	100	Softmax 2
回帰用	96x48x1	8,5x3	2x2	2	16,5x4	2x2	2	32,5x4	2x2	2	1,000	500	100	Sigmoid 6
提案手法	96x48x1	8,5x3	2x2	2	16,5x4	2x2	2	32,5x4	2x2	2	1,000	500	100	Sigmoid 8
(a)INRIA データセット														
手法	Input	Layer1			Layer2			Layer3			Layer4	Layer5	Layer6	Output
		Conv	Max Pooling	Maxout	Conv	Max Pooling	Maxout	Conv	Max Pooling	Maxout	# of unit	# of unit	# of unit	
検出用	64x128x3	20,9x5	2x2	2	64,5x3	2x2	2	32,3x3	2x2	2	1,000	500	100	Softmax 2
回帰用	128x64x3	20,9x5	2x2	2	64,5x3	2x2	2	32,3x3	2x2	2	1,000	500	100	Sigmoid 6
提案手法	128x64x3	20,9x5	2x2	2	64,5x3	2x2	2	32,3x3	2x2	2	1,000	500	100	Sigmoid 8

表1 各データセットで用いる DCNN の構造



手法で同じで,3つの畳み込み層およびプーリング層, その後に3つの全結合層が続いている.Daimler デー タセットに用いる DCNN は,入力画像サイズが96×48 のグレースケール画像であり,INRIA データセットは, 入力画像サイズが128×64のカラー画像である.

実シーンにおける身長推定をするには、カメラパラ メータおよび歩行者までの距離の真値が必要であるが、 Daimler データセットおよび INRIA データセットには、 これらの情報が含まれていない。そのため、身長推定 は我々で収集した画像を用いて評価する。

4.1 歩行者検出精度の比較

DCNN により歩行者検出を行う場合,処理時間と誤 検出を軽減するために,前段処理として別の検出器を 用いる.ここでは,HOGとSVMによる歩行者検出器 を用いて歩行者候補領域を検出する.Daimler データ セットにおける歩行者検出精度について,図4に示す. これより,FPPI (False Positive Per Image)が0.1の



時,検出型の DCNN は未検出率が 38%であるが,回帰型の DCNN は 32%となっている.回帰型の DCNN は, 頭部および両足の座標を教師信号として与えているため,歩行者領域を捉える特徴が選ばれやすくなってい ると考えられる.

次に INRIA データセットにおける歩行者検出精度に ついて,図5に示す. INRIA データセットは Daimler データセットに比べて人の姿勢のバリエーションが多 い.前段処理の HOG 特徴と SVM では検出できない場 合が多いため,評価データセットから人領域を切り出 した画像と人を含まない背景領域をランダムに切り出 した画像を用いて評価する.図5より,FPPW が0.01 の時,提案手法は未検出率が14%と検出型の DCNN の 39%を大幅に上回ることが出来ている.

4.2 歩行者部位検出の精度

Daimler データセットおよび INRIA データセットに おける各部位の位置精度を表 2 に示す. Daimler データ

表 2 部位の位置ずれ精度 [pixel] (a)Daimler データセット

壬注		亚均			
Ъ!Д	頭部	左足	右足	十均	
部位回帰のみ	6.1	5.9	10.7	7.6	
検出と部位回帰	4.2	5.3	9.4	6.3	

(b)INRIA データセット

毛辻		パーツ		亚齿
丁亿	頭部	左足	右足	十均
部位回帰のみ	8.1	8.7	9.9	8.9
検出と部位回帰	6.5	8.0	9.2	7.9



図 6 INRIA データセットにおける部位検出結 果例

セットにおいて,提案手法は,部位の回帰のみを行う場 合と比較して,誤差が小さくなっている.Daimler デー タセットの画像サイズは,96×48であり,誤差をサイ ズで正規化すると約6%の位置ずれとなっている.一方, INRIA データセットにおける位置ずれ精度についても 提案手法の方が誤差が小さい.INRIA データセットの 画像サイズは,128×64であり,誤差をサイズで正規 化すると5.5%の位置ずれとなっている.INRIA データ セットにおける部位検出結果例を図6に示す.これよ り,足の開き方や姿勢が異なる場合でも頭部および足 の位置を推定することができている.

4.3 実シーンにおける身長推定精度の評価

各部位の位置推定を利用して,歩行者までの距離精 度を評価する.距離を推定するためのデータセットは, カメラパラメータと距離があらかじめ分かるように我々 で撮影したサンプルを使用する.撮影距離は5m,10m および15m であり,各距離について7枚ずつ撮影し, 平均の距離精度を算出している.距離推定結果を表3 および評価画像例を図7に示す.距離が5mの時,推定 精度は4.89m となっており,2.2%の誤差であった.距 離が10m および15m と遠くなるに従って,推定精度は 低下するが,10m 以上離れた場合でも5%程度の誤差で 距離推定できていることがわかる.

5 **まとめ**

本稿では、歩行者検出と同時に頭部および両足の位 置を検出する回帰型 DCNN を提案した. DCNN は出

表 3 距離推定精度

距離	推定距離	誤差 [%]
5m	4.89m	2.2
10m	$9.26\mathrm{m}$	5.3
15m	14.12m	5.8

カ層が歩行者と歩行者以外の確率,頭部および両足の x 座標と y 座標を出力する.2つの検出を同時に行う ことで互いに有効な特徴を学習することができ,歩行 者検出精度を向上させるとともに,各部位の検出精度 も向上させることができた.また,各部位の位置情報 とカメラ内部パラメータから歩行者までの距離推定を 行った結果,約5%の誤差にすることができた.今後は, DCNN による各部位の位置精度を向上させ,距離推定 精度の向上を図る.

参考文献

- Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, " Gradient-Based Learning Applied to Document Recognition", Proceedings of the IEEE, 1998.
- [2] W. Ouyang, X. Wang, " Joint Deep Learning for Pedestrian Detection", Computer Vision and Pattern Recognition, 2013.
- [3] Z. Zhang, P. Luo, C. Change, T. Xiaoou, "Facial Landmark Detection by Deep Multi-task Learning", European Conference on Computer Vision, 2014.
- [4] L. Zhao, C. E. Thorpe, "Stereo- and Neural Network-Based Pedestrian Detection", IEEE Transactions on Intelligent Transportation Systems, Vol.1, No.3, pp.148-154, 2000.
- [5] N.Dalal, B. Triggs, "Histograms of oriented gradients for human detection", Computer Vision and Pattern Recognition, 2005.
- [6] X. Wang, T. X. Han, S. Yan, "An HOG-LBP Human Detection with Partial Occlusion", International Conference on Computer Vision, 2009.
- [7] J. Marin, D. Vazquez, A. Lopez, J. Amores, B. Leibe, "Random Forests of Local Experts for Pedestrian Detection", International Conference on Computer Vision, 2012.
- [8] W. Nam, B. Han, J. H. Han, "Improving Object Localization Using Macrofeature Layout Selection", International Conference on Computer Vision Workshop on Visual Surveillance, 2011.
- [9] P. Felzenszwalb, D. McAllester, D. Ramaman,"A Discriminatively Trained, Multiscale, Deformable



(a) 5m

(b) 10m

(c) 15m

図7 距離推定データ例

Part Model ", Computer Vision and Pattern Recognition 2008.

- [10] P. Sermanet, K. Kavukcuoglu, S. Chintala, Y. LeCun, "Pedestrian Detection with Unsupervised Multi-stage Feature Learning" Computer Vision and Pattern Recognition, pp.3626-3633, 2013.
- [11] 岸野嵩久, R. Micheletto, "視覚障害者歩行支援の ための単眼カメラを用いた歩行者検出システム", 感覚代行シンポジウム, Vol. 38, pp.13-16, 2012.
- [12] I. Goodfellow, D. Warde-Farley, M. Mirza, A. C. Couville, Y. Bengio, "Maxout Network", International Conference on Machine Learning, pp.1319-1327, 2013.
- [13] G. E. Hinton, N. Srivastava, A. Krizhevsky, S. Ilya, R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors", Clinical Orthopaedics and Related Research, vol. abs/1207.0, 2012.