情報処理学会論文誌 Vol.57 No.3 910–921 (Mar. 2016)

Random DropoutとEnsemble Inference Networksによる歩行者検出と標識認識

福井 宏[1,a)] 山下 隆義[1,b)] 山内 悠嗣[1,c)] 藤吉 弘亘[1,d)] 村瀬 洋[2,e)]

受付日 2015年6月30日, 採録日 2015年12月7日

概要：Advanced Driver Assistance System において，ドライバの運転支援と歩行者との接触事故防止を実現するために，画像による歩行者検出と標識認識が重要な技術となっている．歩行者検出と標識認識は，これまで手動で設計した特徴量を識別器に入力して学習する方法が一般的に用いられているが，近年，Deep Learning を用いたアプローチが，手動で設計した特徴量を用いたアプローチを大きく上回ったことで注目されている．しかし，Deep Learning を用いたアプローチは，精度向上のために層を深くしたり，他の識別器と併用したりするアプローチが多い．そのため，これらのアプローチはネットワークの構造が複雑になる傾向がある．本稿では，1 つのネットワークで高い性能の歩行者検出と標識認識を実現するため，Dropout のアルゴリズムをベースとした Random Dropout と Ensemble Inference Networks を提案する．従来の Dropout は，学習時の各更新処理において一定の割合で選択したユニットの応答値を 0 にしている．提案する Random Dropout は，ランダムに決定した割合で選択したユニットの応答値を 0 にすることで Dropout よりも汎化性能を向上させる．Ensemble Inference Networks は，評価時に構造の異なる全結合層を複数生成し，各全結合層の応答値を統合することで，最終的な出力を得る．本研究では，1) Random Dropout により，歩行者や標識の隠れや姿勢変動に頑健であるネットワークを構築する．2) Ensemble Inference Networks は，複数の全結合層により誤検出と誤認識を削減する．評価実験により，歩行者検出と標識認識において，提案手法は End-to-End で学習可能な単純な構造のネットワークであるにもかかわらず，各ベンチマークの高性能な Deep Learning ベースの手法と同等の精度であることを確認した．

キーワード：畳み込みニューラルネットワーク，Random Dropout，Ensemble Inference Networks，Dropout

Random Dropout and Ensemble Inference Networks for Pedestrian Detection and Traffic Sign Recognition

Hiroshi Fukui[1,a)] Takayoshi Yamashita[1,b)] Yuji Yamauchi[1,c)] Hironobu Fujiyoshi[1,d)] Hiroshi Murase[2,e)]

Received: June 30, 2015, Accepted: December 7, 2015

Abstract: To saving the driver and pedestrian from vehicle accident, vision based pedestrian detection and traffic sign recognition are important technology for Advanced Driver Assistance System. Pedestrian detection and traffic sign recognition based on machine learning and hand-craft feature became common approach. Deep Learning approaches are applied to challenging tasks and archives the state-of-the-art performance. However, network structure that becoming more deeper of using other classifier in order to improve the accuracy. In this paper, we propose Random Dropout and Ensemble Inference Networks to tackle these problems with same approach for both tasks. Random Dropout selects units at random with a flexible rate, instead of the fixed rate used in conventional Dropout. Ensemble Inference Networks generate multiple networks that have different structures in full connection layers. Our contributions are, 1) obtains better representation network at training process by Random Dropout, Here, the size of the our network is equivalent to the conventional CNN. 2) achieves the high performance for pedestrian detection and traffic sign recognition with Ensemble Inference Network. The proposed methods achieves comparable performance to high performance deep learning methods, even though the structure of the proposed methods are considerably simpler.

Keywords: deep convolutional neural network, Random Dropout, Ensemble Inference Networks, Dropout

1. はじめに

Advanced Driver Assistance System（ADAS）は，夜間など危険な状況下でのドライバの判断を支援するシステムである [1]．ドライバの判断を支援する機能として，歩行者検出や標識認識がある．これらの機能は，車載カメラ映像の単一フレームを用いる方法と複数フレームを用いる方法に分類される．複数フレームを用いる手法は，コンテキストや動き情報を利用できるため，高精度な検出や認識が可能となる．しかしながら，コンテキストや動き情報は，計算コストが非常に高い．それに対して，単一フレームによる処理では，実時間で検出が可能であるが，走行環境や天候の変化が生じても 1 枚の画像から高精度に検出，認識する必要がある．本稿では，車載カメラによる運転支援システムの普及を目的とするため，単一フレームによる方法で識別を行う．

単一フレームによる歩行者検出として，Dalal らが提案した Histogram of Oriented Gradients（HOG）特徴量と Support Vector Machine（SVM）を組み合わせた方法がある [2]．Dalal らの手法をベースとしたアプローチは，多数提案されている [3], [4], [5], [6]．その代表的な手法である Deformable Part Model（DPM）[3] は，歩行者の全身と頭や足などの部位を同時に検出することで，姿勢変化に頑健な歩行者検出を実現している．

標識認識においても，特徴量と機械学習を組み合わせたアプローチが数多く提案されている [7], [8], [9]．Zaklouta らは，HOG 特徴量と Canny のエッジ検出により得られたエッジを組み合わせて特徴量を設計し，Random Forest で学習することで高い認識性能を実現している [12]．

近年，畳み込みニューラルネットワーク（CNN）をベースとした手法が，一般物体認識や歩行者検出，標識認識のベンチマークにおいて高い性能を実現している [13], [14], [15]．

これらの手法は，高い性能を実現するために，汎化性能を向上させる Dropout を CNN の学習時に利用している．Dropout は，一定の割合でランダムにユニットを選択し，応答値を 0 にする．ランダムに選択するユニットは，各更新処理で異なる．我々は Dropout の以下の 2 点に着目する．1 つ目は，学習処理において，ユニットの応答値を 0 にする割合を固定している点，2 つ目は，学習時への導入にとどまっており，識別時に同様の処理は行われていない点である．

本稿では，Dropout をベースとした 2 つの手法を提案する．1 つ目は，学習時に用いる Random Dropout，2 つ目は，識別時に用いる Ensemble Inference Networks（EIN）である．Random Dropout は，各更新回数で応答値を 0 にする割合をランダムに指定する．これにより，歩行者や標識の隠れや照明変化に対して頑健なネットワークを学習し，汎化性能を向上させる．EIN は，学習した 1 つのネットワークからランダムに選択したユニットの応答値を 0 にする全結合層を複数生成し，最終的な出力は特定の統合方法を用いることで識別する．構造の異なる複数の全結合層のネットワークを用いて検出，認識をすることで，誤検出と誤認識を削減する．次章以降で，歩行者検出と標識認識の関連研究について述べた後，従来法である CNN と Dropout を説明し，提案手法である Random Dropout と EIN について説明する．そして，歩行者検出と標識認識の一般的な手法と性能を比較し，提案手法の有用性を評価する．

2. 関連研究

ADAS において，歩行者検出と標識認識は重要な技術の 1 つである．Dalal らが提案した HOG 特徴量は，輝度値の勾配を用いることで，歩行者の服装などの違いによる見えの変化に対して頑健な特徴量を獲得している．HOG 特徴量は，多くの歩行者検出法で用いられている [3], [4], [5], [6]．なかでも，DPM [3] は解像度の異なる HOG 特徴量から，歩行者の全身とパーツをとらえて歩行者検出するため，姿勢の変化に対して頑健な歩行者検出を実現している．標識認識では，手動で設計した特徴量を用いるアプローチとして，HOG 特徴量と機械学習を組み合わせた手法が提案されている．Zaklouta らは HOG 特徴量と Canny のエッジ検出を併用することで，標識のマークをとらえやすくしている [12]．これらの特徴量は，研究者の知識や経験に基づいて設計されている．

一方，一般物体認識において，CNN が従来の性能を大きく上回ったことで注目されている [13]．CNN は，学習過程において識別に適した特徴量を自動的に獲得することができ，家の番号認識 [16]，シーン認識 [17]，物体検出 [18] など様々なベンチマークで高い性能を達成している．Ouyang らは，CNN を用いた階層的な検出構造の歩行者検出法として，Joint Deep Learning を提案している [19]．Joint Deep Learning は，歩行者の部分的な領域から CNN を用いて特徴量を抽出し，制約付きボルツマンマシンに抽出した特徴量を与えて人か背景かを判定している．Joint Deep Learning は，歩行者の部分領域から特徴量を抽出することで，姿勢の変化に頑健な歩行者検出を実現している．また，Luo らは歩行者検出に対して有効な歩行者のパーツ領域を学習により選択する Switchable Deep Network を提案し，歩行者検出のベンチマークにおいて高い性能を実現している [15]．

標識認識においても，CNN を用いた手法が高い性能を

[1] 中部大学
Chubu University, Kasugai, Aichi 487–0027, Japan
[2] 名古屋大学
Nagoya University, Nagoya, Aichi 464–8601, Japan
a) fhiro@vision.cs.chubu.ac.jp
b) takayoshi@cs.chubu.ac.jp
c) yuu@vision.cs.chubu.ac.jp
d) hf@cs.chubu.ac.jp
e) murase@nagoya-u.jp

達成しており，人間の認識性能を上回る性能を実現している [20]．Ciresan らは，入力画像に対して解像度の変更やコントラスト正規化，ヒストグラム平坦化の処理を別々で施し，各変化を与えた入力画像を別々の CNN に入力することで識別する Multi-Column Deep Neural Network を提案している [14]．このように，CNN を用いた手法が様々なデータセットで高い性能を達成している．

しかし，これらの複数のネットワークを用いたアプローチなどは，End-to-End で学習できない構造となっているため，学習過程が複雑である．例として，Switchable Deep Network では，歩行者検出に有効な歩行者のパーツ領域を選択する際に使用している制限付きボルツマンマシンと，CNN を個々で学習する必要がある．Multi-Column Deep Neural Network においても，解像度の変更やコントラスト正規化など様々な変化を与えた各入力画像に対して別々の CNN で学習する必要がある．そこで，本研究では 1) End-to-End で学習できる単純なネットワーク構造を用いることで，学習コストを削減し，2) 汎化性能を向上させるための学習法を用いることで，性能が高いネットワークを構築する．3) そして，提案手法を歩行者検出と標識認識に適用し，認識性能の向上を実現する．

3. 畳み込みニューラルネットワーク

本章では，CNN の構造と学習方法，および汎化性能を向上させる手法である Dropout について説明する．

3.1 畳み込みニューラルネットワークの構造

図 1 のように，CNN は畳み込み層とプーリング層が積み重なった後に全結合層を配置する階層的な構造となっている．畳み込み層は，M 枚の $K_x \times K_y$ の重みフィルタ $\mathbf{V}$ を $l-1$ 層の特徴マップ $\mathbf{x}^{l-1}$ に対して畳み込む処理を行う．前層の特徴マップと重みフィルタの畳み込み処理を繰り返し行うことで，新たな特徴マップを得る．

$$\mathbf{h} = \phi\left(\mathbf{V}^T\mathbf{x}^{l-1} + \mathbf{b}\right) \tag{1}$$

ここで，$\mathbf{b}$ はバイアス項である．畳み込み処理により得られた値を活性化関数 ϕ に与えることで応答値を得る．活性化関数には，シグモイド関数や Rectified Linear Unit (ReLU)，Maxout [22] などがある．畳み込み層により特徴マップを得た後，プーリング層でサブサンプリングを行う．サブサンプリングは，局所領域の最大値でサンプリングする最大値プーリングが一般的に用いられる．最後のプーリング層の特徴マップは，次の層の全結合層に特徴ベクトルとして入力する．最終層である出力層は，式 (2) に示すソフトマックス関数を用いて各クラスに対するスコア O_c を出力する．算出したスコアをもとに最大となるスコアのクラスを識別結果として出力する．

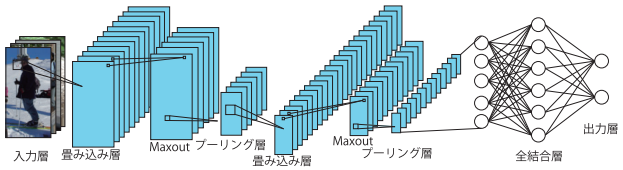


図 1 従来の CNN の構造
Fig. 1 Structure of convention CNN.

$$O_c = \frac{\exp\left(\mathbf{W}_j^L\mathbf{h}_j^L + b_j^L\right)}{\sum_{c=0}^{C}\exp\left(\mathbf{W}_c^L\mathbf{h}_c^L + b_c^L\right)} \tag{2}$$

3.2 畳み込みニューラルネットワークの学習

CNN の重みフィルタや結合重みの初期値は乱数で決定し，誤差逆伝播法により重みフィルタや結合重みを更新する [21]．誤差逆伝播法では，式 (3) のような誤差関数 E を最小化するように，重みフィルタや結合重みを式 (4) の確率的勾配降下法により更新する．

$$E = \frac{1}{2}\sum_{n=1}^{N} E_n \tag{3}$$

$$\mathbf{W}^{(l)} \leftarrow \mathbf{W}^{(l)} + \Delta\mathbf{W}^{(l)} = \mathbf{W}^{(l)} - \eta\frac{\partial E_n}{\partial \mathbf{W}^{(l)}} \tag{4}$$

ここで，$\{n|1,\ldots,N\}$ は学習サンプル，η は学習係数，$\mathbf{W}^{(l)}$ は l 層と $l+1$ 層の結合重みを示している．各サンプルの誤差 E_n は，各クラスに対する出力とラベルから算出する．認識問題の場合は，誤差関数にクロスエントロピーを用いる．

結合重みの更新量 $\Delta\mathbf{W}^{(l)}$ は，式 (5) のように求めることができる．

$$\Delta\mathbf{W}^{(l)} = -\eta\delta^{(l)}\mathbf{y}^{(l-1)} \tag{5}$$

$$\delta^{(l)} = \mathbf{e}\phi(\mathbf{A}^{(l)}) \tag{6}$$

$$\mathbf{A}^{(l)} = \mathbf{W}^{(l)} \cdot \mathbf{y}^{(l-1)} \tag{7}$$

$\mathbf{y}^{(l-1)}$ は $l-1$ 層の出力を示しており，$\mathbf{e}$ はユニットの誤差を示している．$\mathbf{A}^{(l)}$ は $l-1$ 層と l 層の結合重みと，$l-1$ 層の全ユニットの応答値の内積結果である．局所的な勾配 δ は，式 (6) から得ることができる．ネットワークの重みフィルタや結合重みの更新は，あらかじめ指定した更新回数または収束条件を満たすまで繰り返し行う．

誤差 E を算出するための学習サンプルの与え方として，バッチ学習，ミニバッチ学習，オンライン学習がある．バッチ学習は，すべての学習サンプルをネットワークに入力し，誤差 E を求め，パラメータを 1 回更新する学習法である．この方法は，誤差の勾配の変化が大きいことから，学習サンプルが増えるにつれて学習の収束が困難となる．オンライン学習は，1 つの学習サンプルをネットワークに入力し，誤差 E を求め，パラメータを 1 回更新する学習法である．この手法は，各学習サンプルで誤差の勾配を求めて更新を

しているため，学習サンプルが増加した場合でも最適解を獲得しやすいが，パラメータの更新回数が膨大となり計算コストが高くなる．ミニバッチ学習は，バッチ学習とオンライン学習の中間のアプローチであり，少量の学習サンプルをネットワークに入力し，誤差 E を求め，パラメータを 1 回更新する学習法である．少量の学習サンプルから誤差の勾配を求めることで，パラメータの更新回数を削減でき，学習サンプルが増加した場合でも最適解の獲得しやすい．そのため，CNN の一般的な学習方法として利用されている．

3.3 Dropout

Dropout は，全結合層のユニットを指定した割合だけランダムに選択し，その応答値を 0 にして学習する手法である．式 (8) のように，ユニットの応答値を 0 にする場合は m_j^l を 0，ユニットの応答値を伝播する場合は m_j^l を 1 にすることで，l 層目のユニット j の応答値を制御している．

$$h_j^l = f\left(\mathbf{W}^l\mathbf{x} + \mathbf{b}^\mathbf{l}\right) \cdot m_j^l \tag{8}$$

Dropout は，各更新処理で異なるユニットを選択することで，異なる一部の結合が取り除かれた場合においても識別できるようにネットワーク全体のパラメータを学習するため，汎化性能を向上させることができる．識別時は，学習時に指定した応答値を 0 にする割合を全ユニットに対して乗算して識別する．

Dropout をもとに DropConnect [23]，DropAll [24]，Adaptive dropout [25] が提案されている．DropConnect は，結合重みをランダムに選択し，その結合重みを 0 にする手法である．DropAll は，ユニットと結合重みを両方選択し，選択したユニットにつながる結合重みと選択した結合重みの両方を 0 にする手法である．Adaptive dropout は，学習過程において応答値を 0 にする割合を求める方法である．Adaptive dropout は，応答値を 0 にする割合を求める際に，Dropout を行うネットワークと行わないネットワークの 2 つを学習する必要がある．そのため，学習コストが従来の Dropout と比べて 2 倍必要になる．

4. 提案手法

本稿では，Dropout をベースとした 2 つの手法を提案する．1 つ目は，学習時に用いる Random Dropout であり，2 つ目は，識別時に用いる EIN である．以下に 2 つの提案手法の詳細について述べる．

4.1 Random Dropout

従来の Dropout では，応答値を 0 にするユニットの割合は，各更新処理で一定である．提案する Random Dropout は，応答値を 0 にするユニットの割合を各更新処理でランダムに変化させる．**図 2** に Random Dropout のアルゴリズム例を示す．Random Dropout は，更新 1 回目のとき，各層のユニットの削減率を 60%と 30%としている．そして，更新 2 回目では，40%と 70%としており各更新処理で異なっている．これにより，ユニット数が異なっても同じような結果を得られるようにパラメータが更新されるため，汎化性能を向上させることができる．ここで，削減率の範囲は，あらかじめ指定するパラメータとし，全ユニットが削減されないようにする．

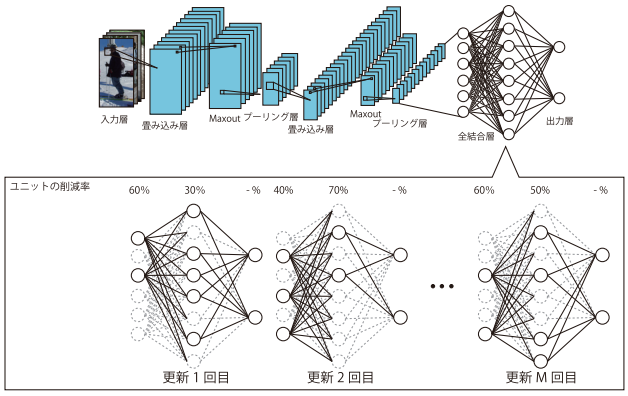


図 2 Random Dropout のアルゴリズム
Fig. 2 Algorithm of Random Dropout.

4.2 Ensemble Inference Networks

EIN は，学習した 1 つのネットワークから構成が異なる全結合層を複数生成し，それらの出力を統合することで，最終的な出力を得る．EIN による識別処理は，**図 3** のように，1) 畳み込み層およびプーリング層による特徴マップの生成，2) 複数の全結合層による識別，3) 応答値の統合の 3 つから構成されている．

4.2.1 畳み込み層およびプーリング層

EIN の畳み込み層およびプーリング層は，従来の CNN と同じ処理である．EIN は，異なる構成の全結合層を複数生成するが，畳み込み層とプーリング層は共通である．そして，畳み込み層とプーリング層を通した後の結果を特徴ベクトルとして保持する．これを各生成した全結合層に入力することで，畳み込み層およびプーリング層の処理を繰り返し行う必要がなく，計算コストを削減することができる．

4.2.2 全結合層の生成

従来の CNN による識別時は，図 3 (a) のように特徴ベクトルを全結合層に入力し，各クラスに対するスコアを取得する．このとき，生成する全結合層は 1 つである．EIN は，図 3 (b) のように学習したネットワークの全結合層をもとにランダムに選択したユニットの応答値を 0 にする全結合層を N 個用意する．そして，特徴ベクトルを生成した全結合層に入力して各クラスのスコアを求める．これを N 回行うことで，構造の異なる全結合層を通して得られた N 個の各クラスのスコアを求めることができる．

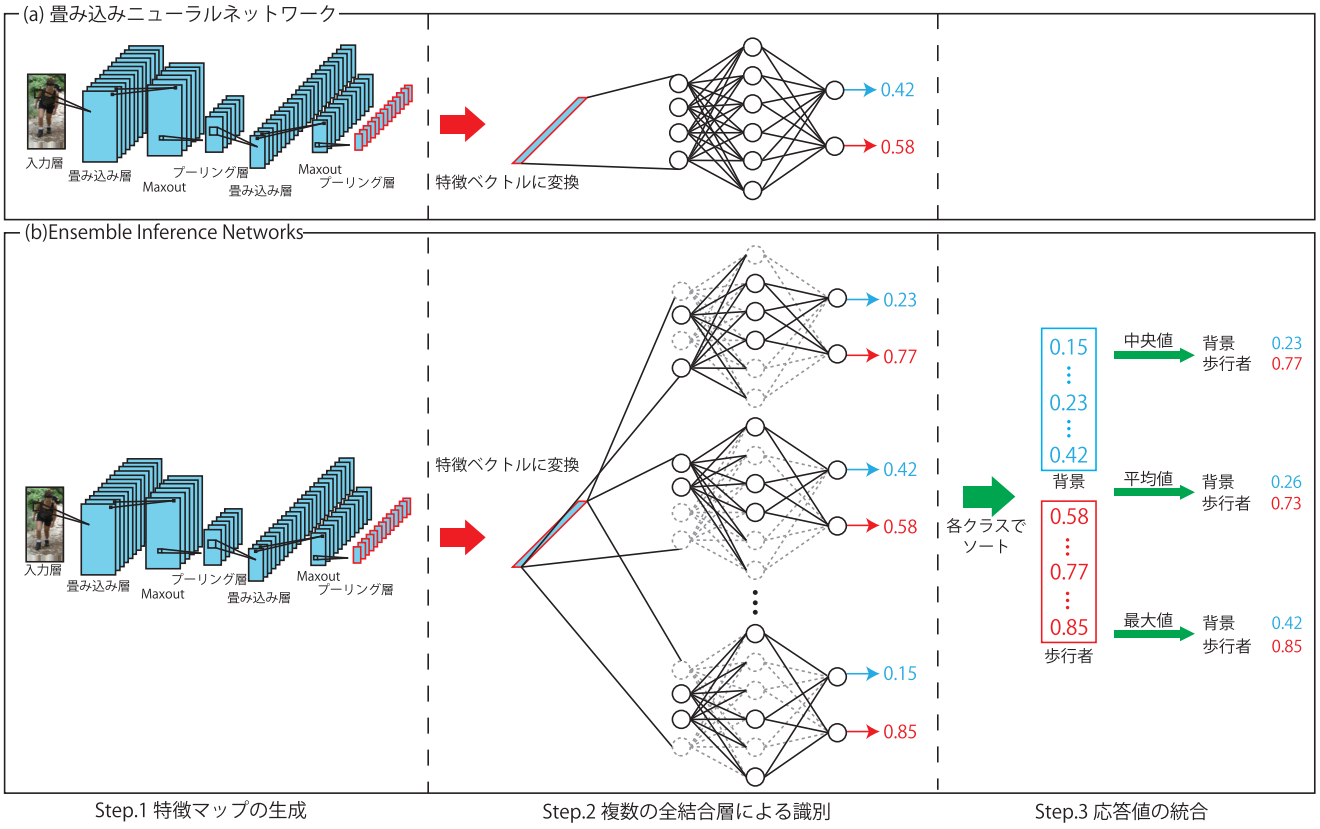


図 3 EIN のアルゴリズム
Fig. 3 Algorithm of EIN.

4.3 応答値の統合

前節で求めた各全結合層における各クラスのスコア O_{nc} を用いて最終的な出力を決める．まず，各全結合層のスコアをクラスごとに格納する．このとき，各クラスに対する応答値の集合を $\mathbf{S}_c$ とする．EIN の各クラスに対する最終的な出力 $\mathbf{S}_c$ は，中央値 $\mathbf{S}_c^{Median}$ や平均値 $\mathbf{S}_c^{Mean}$，最大値 $\mathbf{S}_c^{Max}$ から求める．ここでは，どの値の算出方法が適しているかは，問題設定ごとに決めることができるように一般化している．

5. 提案手法による歩行者検出

歩行者検出は，入力画像を網羅的にラスタスキャンし，各領域が歩行者かどうかを識別する．CNN の場合，畳み込み演算に多大な計算コストを要するため，ラスタスキャンにより入力画像 1 枚から多くの検出ウィンドウを対象とすると，非常に計算コストがかかる．そこで，本研究では，図 4 のように 2 段階の識別処理を行う．

1 段階目で HOG+SVM を用いて歩行者の候補領域の絞り込みを行う．そして，2 段階目で絞り込んだ領域に対して提案手法により最終的な識別を行う．

6. 評価実験

本稿では，Caltech Pedestrian Dataset [26]，Daimler

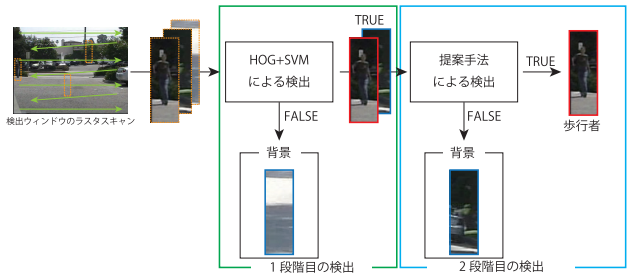


図 4 2 段階処理による歩行者検出
Fig. 4 Pedestrian detection by 2 stage process.

Mono Pedestrian Benchmark Dataset [30]，German Traffic Sign Recognition Benchmark (GTSRB) Dataset [31] を用いて，提案手法の性能を評価する．提案手法の評価では，Random Dropout と EIN の効果について検証する．また，性能を評価するために既存手法との比較を行う．

Random Dropout の性能評価では，Dropout の削減率および Random Dropout の削減率の範囲を変化させて比較する．Dropout または Dropout の発展形の手法の削減率は，40%から 90%まで変化させる．Random Dropout の削減率は，削減率の下限値を 0%とし，上限値を 40%から 90%まで変化させる．EIN の性能評価では，生成する全結合層の数を 1 から 33 まで変化させて性能を比較する．各比較実験の結果から，Random Dropout と EIN のパラメータを決定し，従来の歩行者検出法と標識認識法を

表 1 各データセットにおける CNN の構造
Table 1 Structure of CNN in each dataset.

データセット	入力層	1 層目			2 層目			3 層目			4 層目	5 層目	6 層目	出力層
		畳み込み	Maxout	最大値プーリング	畳み込み	Maxout	最大値プーリング	畳み込み	Maxout	最大値プーリング	ユニット数	ユニット数	ユニット数	ソフトマックス
Caltech	108x36x3	40,9x5	2	2x2	64,5x3	2	2x2	32,3x3	2	2x2	1,000	500	100	2
Daimler	96x48x1	100,5x3	2	2x2	80,5x4	2	2x2	70,5x4	2	2x2	1,000	500	100	2
GTSRB	64x64x3	60,5x5	2	2x2	80,5x5	2	2x2	120,4x4	2	2x2	2,000	1,000	300	43

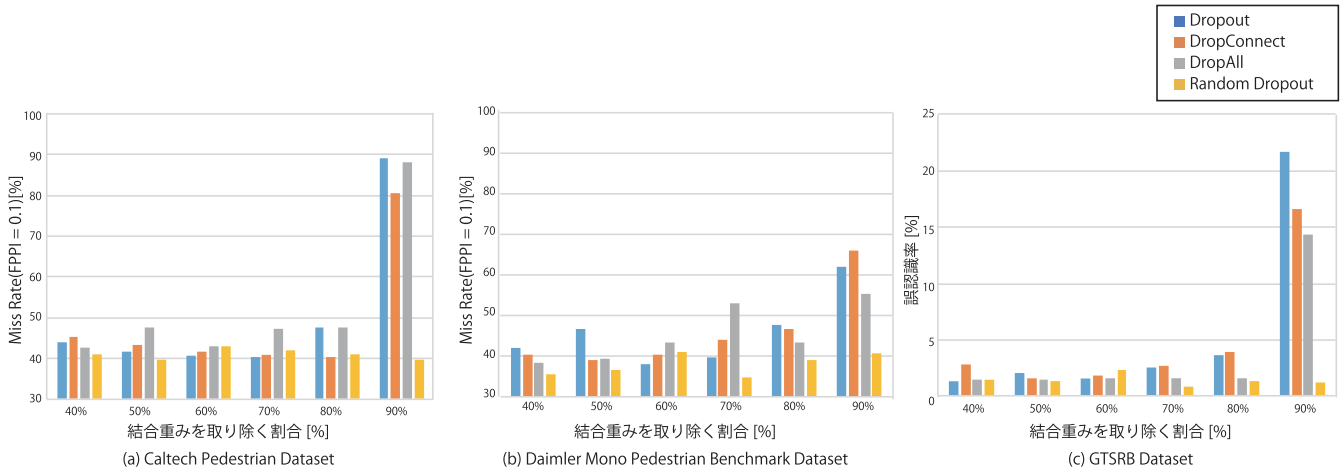


図 5 Dropout と関連する手法と Random Dropout の性能比較
Fig. 5 Performance comparison proposed method with various dropout methods.

比較することで最終的な性能を評価する．Caltech Pedestrian Dataset では，CNN，HOG [2]，HogLbp [4]，LatSvm-V2 [3]，VJ [32]，DBN-Isol [33]，ACF [34]，ACF-Caltech [34]，Pls [35]，FPDW [36]，ChuFtrs [37]，CrossTalk [38]，RandomForest [6]，MultiResC [39]，Roerei [40]，MOCO [41]，Joint Deep [19]，Switchable Deep Network [15] と比較する．Daimler Mono Pedestrian Benchmark Dataset では，CNN，HOG [2]，Shapelet [42]，HogLbp [4]，LatSvm-V2 [3]，VJ [32]，RandomForest [6]，MultiFtr [43]，MLS [5] と比較する．また，標識認識では，Random Forests，Human Performance [20]，Multi-Scale CNNs [28]，Multi-Column Deep Neural Network [14] と比較する．

各データセットに対して使用するネットワークの構造を**表 1** に示す．学習パラメータは，ネットワークの更新回数を 50 万回，学習係数は 0.01 とし，Caltech Pedestrian Dataset と Daimler Mono Pedestrian Benchmark Dataset のミニバッチサイズは 5，GTSRB Dataset のミニバッチサイズは 40 とする．各実験は，5 回試行している．

Caltech Pedestrian Dataset の評価は，学習データセットに対して HOG+SVM で検出した歩行者領域約 4,000 枚と誤検出領域およびランダムに切り出した背景サンプル約 20 万枚を学習に用い，評価画像 8,723 枚を評価した．歩行者サンプルは平行移動とスケーリングをランダムに加えて Data Augmentation したサンプル計 10 万枚を学習サンプルとして用いる．Daimler Mono Pedestrian Benchmark Dataset は，歩行者サンプル約 3 万枚に対して平行移動とスケーリングをランダムに加えて Data Augmentation したサンプル計 13 万枚と，学習データセットに対して HOG+SVM で検出した際に発生した誤検出領域およびランダムに切り出した背景サンプル約 25 万枚を学習に用い，評価画像 21,790 枚を評価した．

GTSRB Dataset では，学習サンプルに 39,000 枚の標識画像と，評価サンプルに 12,631 枚の標識画像を用いる．GTSRB Dataset では，Data Augmentation による画像の生成は行わない．

6.1 Random Dropout の性能評価

Random Dropout の効果について，Dropout，DropConnect，DropAll と比較する．各データセットに対して，各手法で学習した CNN の Miss rate および誤認識率を**図 5** (a)，(b)，(c) に示す．ここで，図 5 (a)，(b) の Miss rate は，Receiver Operating Characteristic（ROC）カーブを用いて評価した際に，False Positive per Image（FPPI）が 0.1 のときの Miss rate を示している．

図 5 (a) より，Caltech Pedestrian Dataset において，Dropout は削減率が 70%のとき Miss rate が 40.45%である．一方，DropConnect は削減率が 80%のときに Miss rate が 40.52%となっている．提案手法である Random Dropout は削減率の範囲が 0%から 90%のときに 39.65%の Miss rate であり，提案手法の Miss rate が最も低いことが分かる．図 5 (b) の Daimler Mono Pedestrian Benchmark Dataset では，Dropout は削減率が 70%のとき Miss rate が 39.81%である．一方，DropAll は削減率が 40%のときに 38.22%となっている．提案手法である Random Dropout は削減率

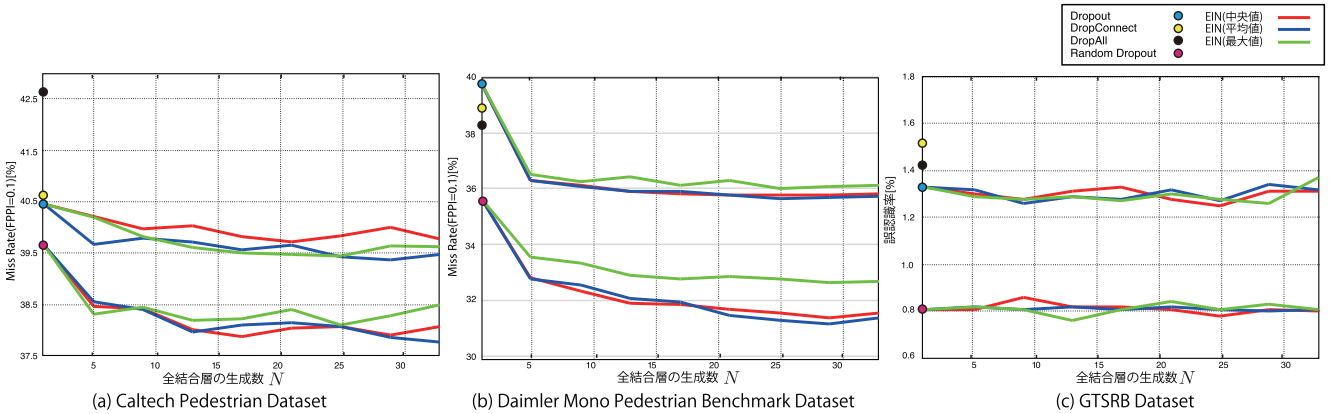


図 6 生成する全結合層数の変化による精度の比較

Fig. 6 Performance comparison with the number of generated fully-connection networks.

の範囲が 0%から 70%のときに Miss rate が 35.79%であり，提案手法が最も Miss rate が低いことが分かる．

図 5 (c) の GTSRB Dataset では，Dropout は削減率が 40%のとき 1.33%の誤認識率である．一方，DropAll は削減率が 50%のときに 1.39%となっている．提案手法である Random Dropout は削減率の範囲が 0%から 70%のときに 0.81%の誤認識率であり，提案手法の誤認識率が最も低いことが分かる．

6.2 Ensemble Inference Networks の性能評価

EIN の全結合層の生成数，最終出力の決定方法を各評価において比較する．**図 6** (a)，(b) に歩行者検出データセットにおける，全結合層の生成数に対する Miss rate の変化，図 6 (c) に標識認識データセットにおける全結合層の生成数に対する誤認識率の変化を示す．なお，図 6 (a)，(b) は FPPI が 0.1 のときの Miss rate である．ここで，EIN の出力値の統合方法を中央値・平均値・最大値の 3 パターンで比較する．全結合層の生成数が 1 の場合は，EIN を用いていない場合であり，図 6 (a)，(b)，(c) の縦軸上の各点に相当する．前節での実験結果と同様に，Random Dropout が最も良い性能となっている．

次に EIN を用いた場合，図 6 (a) から，Caltech Pedestrian Dataset では Random Dropout を導入し，EIN の統合方法に平均値を使用して，生成数が 33 のとき Miss rate が 37.77%で最も精度が良い．EIN を用いていないときの Random Dropout より最大で Miss Rate が 1.88%減少していることが分かる．図 6 (b) から，Daimler Mono Pedestrian Benchmark Dataset では Random Dropout を導入し，EIN の統合方法に中央値を使用して，生成数が 33 のとき Miss Rate が 31.34%で最も未検出が低い．同じように，EIN を用いていないときの Random Dropout より最大で Miss Rate が 4.45%減少していることが分かる．図 6 (a)，(b) より，歩行者検出において，EIN の最終出力の方法に中央値か平均値を用いることで，検出性能を向上させることができる．

標識認識のベンチマークである，GTSRB Dataset の結果を図 6 (c) に示す．図 6 (c) より，Random Dropout を導入し，EIN の統合方法に最大値を使用して，生成数が 17 のとき Miss Rate が 0.76%で最も良い．また，EIN を導入することで 0.1%認識性能が向上していることが確認できる．図 6 (c) より，多クラス認識では最大値を最終出力として用いるのが良い．

6.3 従来法の性能比較

各データセットにおける従来法と提案手法の比較を行う．**図 7** (a) に Caltech Pedestrian Dataset，図 7 (b) に Daimler Mono Pedestrian Benchmark Dataset の比較結果を示す．Caltech Pedestrian Dataset では，FPPI が 0.1 のとき従来の CNN に対して提案手法は，Miss rate を 8.54%改善することができている．Deep Learning による手法において最も高精度な歩行者検出方法である Switchable Deep Learning の Miss rate は 37.87%であり，提案手法の Miss rate は 37.77%であることから，Deep Learning による手法で最も性能が良いことが分かる．同様に，図 7 (b) の Daimler Mono Pedestrian Benchmark Dataset においても，FPPI が 0.1 のとき従来の CNN と提案手法を比べて Miss rate が 1.16%減少していることが確認できる．

表 2 に GTSRB Dataset の比較結果を示す．人間の認識性能が 98.84%であるのに対して，提案手法では 99.24%であり，人間の認識性能を上回っていることが確認できる．CNN ベースの手法である Multi-Scale CNN と Multi-Column Deep Neural Network の性能を比較したとき，それぞれ 98.31%と 99.46%であることから，提案手法はこれらの手法と同等の性能を実現している．

図 8 に Caltech Pedestrian Dataset と Daimler Mono Pedestrian Benchmark Detection の歩行者検出例を示す．1，4 列目の歩行者検出例は従来の歩行者検出法である HOG+SVM の検出例であり，2，5 列目の歩行者検出例は

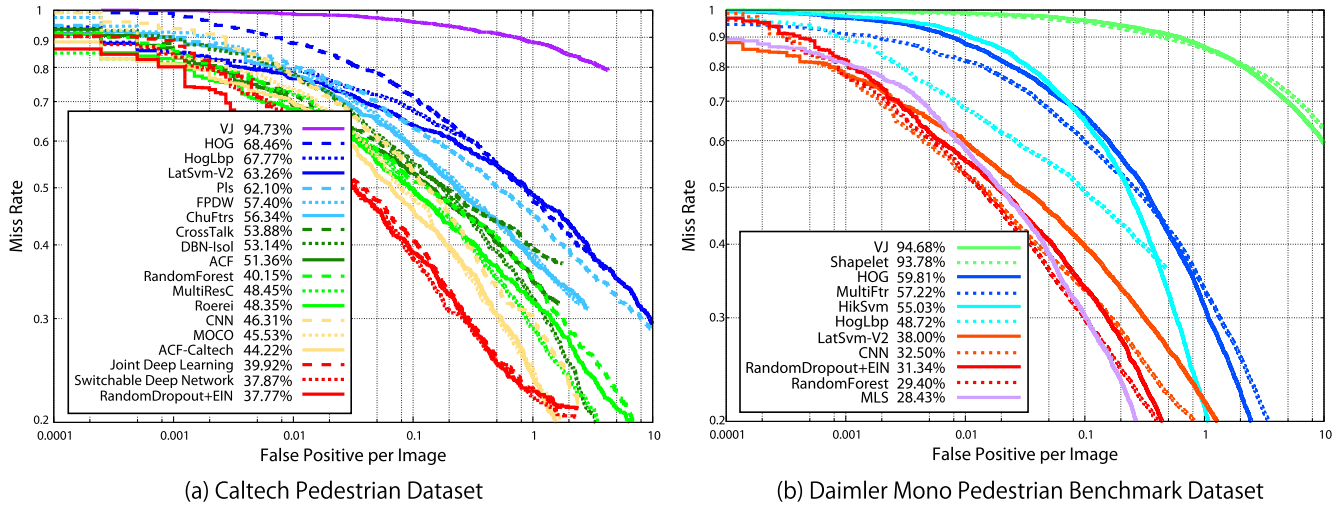


図 7 歩行者検出における提案手法と従来法の比較

Fig. 7 Performance comparison proposed methods with conventional methods.

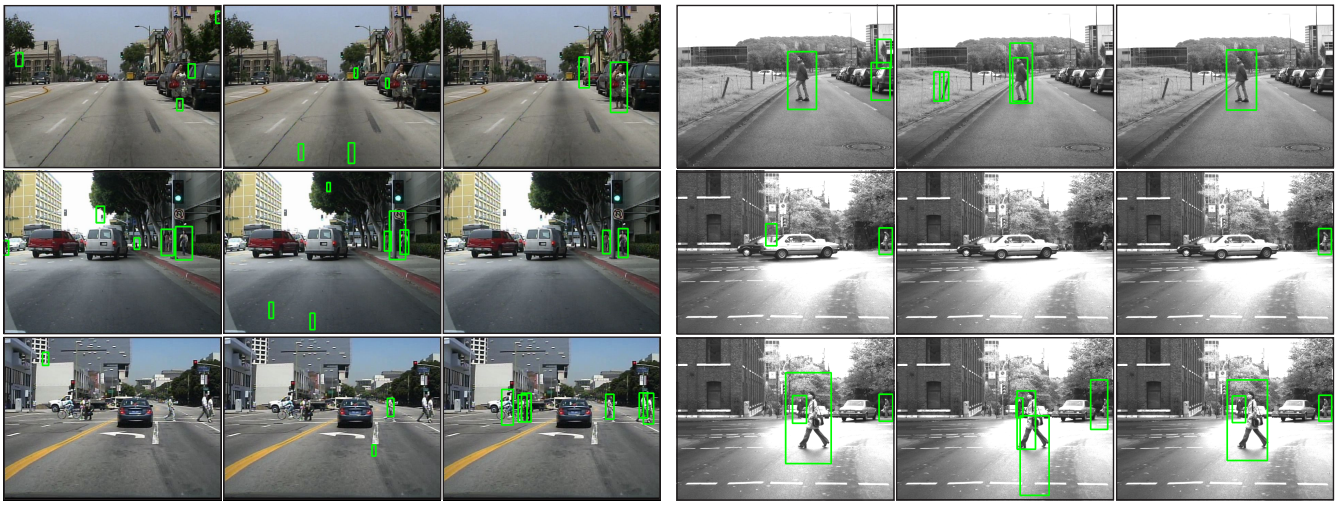


図 8 歩行者検出結果の比較

Fig. 8 Detection results in each evaluation dataset.

表 2 GTSRB Dataset における提案手法と従来法の比較

Table 2 Compare proposed methods and conventional methods for GTSRB Dataset.

手法	認識率 [%]
Random Dropout + EIN	99.24
Multi-Column Deep Neural Network	99.46
Human Performance	98.84
Multi-Scale CNN	98.31
Random Forests	96.14
LDA on HOG	95.68

DPM の検出例，3，6 列目の歩行者検出例は提案手法の検出例である．従来の歩行者検出法では，遠方に存在する歩行者や体型の変化が大きい歩行者に対して検出が困難であり，多くの誤検出が発生している．それに対して，提案手法を導入することで従来の歩行者検出法で検出できなかった歩行者を検出し，誤検出を削減できていることが確認できる．

図 9 に，標識認識データセットにおける Dropout を導入した CNN と提案手法の認識結果を示す．ここで，各グラフは上位 3 位までの結果である．また，赤いグラフはターゲットクラスの尤度を示しており，青いグラフは他のクラスの尤度を示している．図 9 より，従来の CNN では照明変動や標識の向き，解像度の低下，オクルージョンの発生に対して誤認識が発生していることが分かる．それに対して，提案手法では照明変動や標識の向き，解像度の低下，オクルージョンの発生が生じた場合においても認識が可能であることが確認できる．

6.4 考察

学習サンプル数やネットワーク構成による性能を比較す

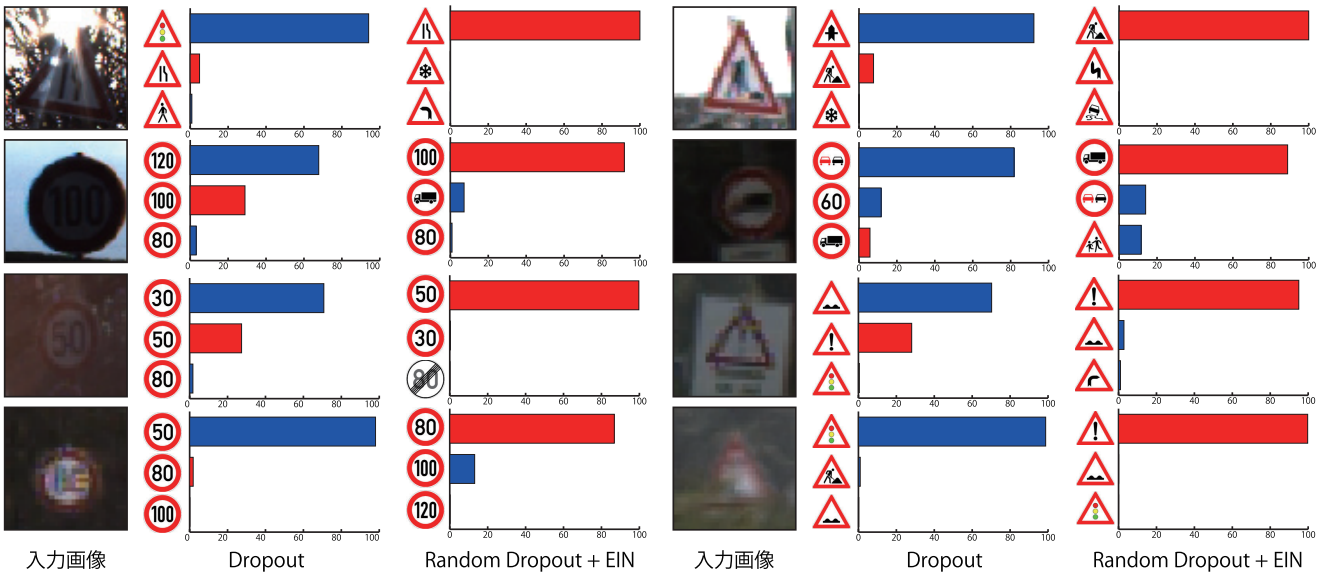


図 9 GTSRB Dataset における提案手法と従来の CNN の上位 3 つのスコアの比較
Fig. 9 Top three scores of conventional Dropout and proposed methods in GTSRB Dataset.

表 3 学習サンプル数に対する Miss rate の推移
Table 3 Compare of Miss rate for number of training samples.

学習サンプル数	Dropout	提案手法
3 万枚	47.73%	46.17%
6 万枚	43.13%	43.12%
10 万枚	40.45%	**37.77%**

表 4 全結合層の層数に対する Miss rate の推移
Table 4 Compare of Miss rate for number of fully-connection layers.

層数	Dropout	提案手法
1 層	45.47%	51.05%
2 層	37.96%	45.52%
3 層	40.45%	**37.77%**

表 5 全結合層のユニット数に対する Miss rate の推移
Table 5 Compare of Miss rate for number of units in fully-connection layers.

各層のユニット数	Dropout	提案手法
500, 250, 50	40.45%	42.66%
1,000, 500, 100	40.49%	**37.77%**
1,500, 750, 150	40.12%	38.33%

表 6 学習時の処理時間
Table 6 Training processing time.

	Dropout	提案手法
学習時間	2 時間 53 分	3 時間 9 分

る．また，EIN による全結合層の生成数に対する計算コストを比較する．本実験では Caltech Pedestrian Dataset を使用し，Random Dropout の範囲を 0%から 90%，EIN による生成数を 33 とする．また，応答値の統合方法に平均値を使用する．

表 3 に学習サンプルに対する提案手法の Miss rate を示す．学習サンプルは，3 万枚から 10 万枚まで変化させている．これより，学習サンプル数を増やすことで Miss rate が低下しており，Data Augmentation により学習サンプル数を増やすことが重要であることが分かる．また，各学習サンプル数において，Dropout を用いて学習した場合よりも Miss rate を低下させることができている．

次に，ネットワーク構成による比較実験を行う．ベースとなるネットワーク構成は，表 1 で示した構成であり，Random Dropout と EIN の処理に関する全結合層の層数とユニット数を変化させて比較する．**表 4** は，全結合層の層の数による性能比較，**表 5** はユニット数による性能比較の結果である．表 4 より，全結合層の層数を 3 層とした場合が最も性能が良いことが分かる．一方，表 5 よりユニット数を少なくすると性能が低下することが分かる．ユニット数を各層それぞれ 1,000，500，100 とした場合が最も性能が良い．

表 6 に学習時の処理時間を示す．学習は，GPU: GeForce GTX 980 で行っている．表 6 より，提案手法の学習時に用いるネットワーク構造は Dropout とほぼ同じであるため，学習の処理時間に大きな差がないことが確認できる．また，**図 10** に全結合層の生成数による識別時の処理時間を示す．識別処理は，CPU: Intel(R) Core(TM) i7-4790K CPU @ 4.00 GHz で行っている．図 10 より，EIN を用いない場合（N=1），処理時間は 11 ms，EIN による生成数を 33 にした場合の処理時間は 37 ms である．全結合層の生成数は 33 倍になっているが，畳み込み層とプーリング層の共有化と全結合ユニットの削減により 3 倍程度となっている．これより，学習サンプルおよびネットワーク構成

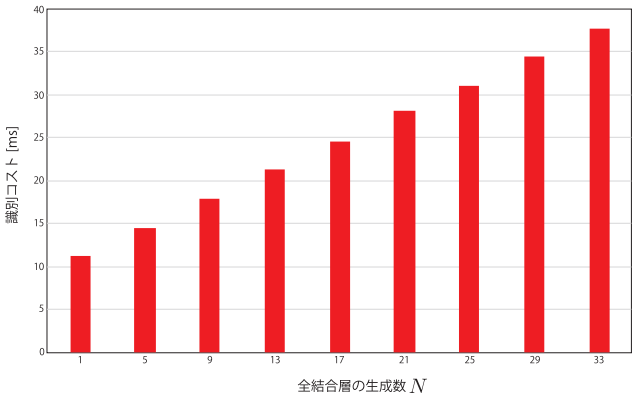


図 10 全結合層の生成数による識別時の処理時間

Fig. 10 Evaluation processing time with the number of fully-connection networks.

を変えた場合でも提案手法が有効であることが分かる．

7. おわりに

本稿では，汎化性能の向上を目的とした 2 つの手法を提案した．Dropout の割合を学習の更新回数ごとにランダムで決定する Random Dropout を学習に導入することで，汎化性能を向上させることができた．EIN による識別処理では，識別過程でランダムに選択した応答値を 0 にし，複数の全結合層を生成する．そして，生成した全結合層の応答値から最終的な出力を算出することで高精度な識別を実現した．今後の課題として，リアルタイムで歩行者検出を実現するために CNN の高速化があげられる．

謝辞 本研究の一部は，独立行政法人科学技術振興機構（JST）の研究成果展開事業「センター・オブ・イノベーション（COI）プログラム」の支援により行われた．

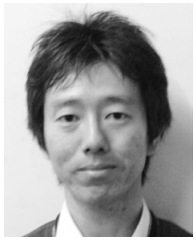
参考文献

[1] Geronimo, D., Lopez, A.M. and Sappa, A.D.: Survey of Pedestrian Detection for Advanced Driver Assistance Systems, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.32, No.7, pp.1239–1258 (2010).
[2] Dalal, N. and Triggs, B.: Histograms of oriented gradients for human detection, *Computer Vision and Pattern Recognition* (2005).
[3] Felzenszwalb, P., McAllester, D. and Ramaman, D.: A Discriminatively Trained, Multi scale, Deformable Part Model, *Computer Vision and Pattern Recognition* (2008).
[4] Wang, X., Han, T.X. and Yan, S.: An HOG-LBP Human Detection with Partial Occlusion, *International Conference on Computer Vision* (2009).
[5] Nam, W., Han, B. and Han, J.H.: Improving Object Localization Using Macrofeature Layout Selection, *International Conference on Computer Vision Workshop on Visual Surveillance* (2011).
[6] Marin, J., Vazquez, D., Lopez, A., Amores, J. and Leibe, B.: Random Forests of Local Experts for Pedestrian Detection, *International Conference on Computer Vision* (2012).
[7] Miura, J., Kanda, T. and Shirai, Y.: An Active Vision System for Real-time Traffic Sign Recognition, *Intelligent Transportation Systems* (2000).
[8] De la Escalera, A., Armingol, J. and Mata, M.: Traffic Sign Recognition and Analysis for Intelligent Vehicles, *Image and Vision Computing*, Vol.3, pp.247–258 (2003).
[9] Gavrila, D.: Traffic Sign Recognition Revisited, *Musterkennung*, pp.86–93 (1999).
[10] Bahlmann, C., Zhu, Y., Ramesh, V., Pellkofer, M. and Koehler, T.: A System for Traffic Sign Detection, Tracking, and Recognition Using Color, Shape, and Motion Information, *Intelligent Vehicles* (2005).
[11] Bascon, S.M., Arroyo, S.L., Jimenez, P.G. and Moreno, H.G.: Road-sign Detection and Recognition based on Support Vector Machines, *IEEE Trans. Intelligent Transportation Systems*, Vol.8, No.2 (2007).
[12] Zaklouta, F., Stanciulescu, B. and Hamdoun, O.: Traffic Sign Classification Using K-d Trees and Random Forests, *International Joint Conference on Neural Network* (2011).
[13] Krizhevsky, A., Ilva, S. and Hinton, G.E.: ImageNet Classification with Deep Convolutional Neural Network, *Advances in Neural Information Processing System 25*, pp.1097–1105 (2012).
[14] Ciresan, D., Meier, U., Masci, J. and Schmidhuber, J.: Multi-Column Deep Neural Network for Traffic Sign Classification, *Neural Network*, Vol.32, pp.333–338 (2012).
[15] Luo, P., Tian, Y., Wang, X. and Tang, X.: Switchable Deep Network for Pedestrian Detection, *Computer Vision and Pattern Recognition* (2014).
[16] Goodfellow, I.J., Bulatov, Y., Ibarz, J., Arnold, S. and Shet, V.: Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks, *CoRR*, Vol.abs/1312.6082 (2013).
[17] Farabet, C., Couprie, C., Najman, L. and LeCun, Y.: Learning Hierarchical Features for Scene Labeling, *IEEE Transacations on Pattern Analysis and Machine Intelligence* (2012).
[18] Girshick, R., Donahue, J., Darrell, T. and Malik, J.: Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, *Computer Vision and Pattern Recognition* (2014).
[19] Ouyang, W. and Wang, X.: Joint Deep Learning for Pedestrian Detection, *Computer Vision and Pattern Recognition* (2013).
[20] Stallkamp, J., Schlipsing, M., Salmen, J. and Igel, C.: Man vs. Computer: Benchmarking Machine Learning Algorithms for Traffic Sign Recognition, *Neural Network*, Vol.32, pp.323–332 (2012).
[21] Rumelhart, D.E., Hinton, G.E. and Williams, R.J.: Learning representations by back-propagating errors, *Neurocomputing*, pp.696–699 (1988).
[22] Goodfellow, I., Warde-Farley, D., Mirza, M., Couville, A.C. and Bengio, Y.: Maxout Network, *International Conference on Machine Learning*, pp.1319–1327 (2013).
[23] Wan, L., Zeiler, M., Zhang, S., LeCun, Y. and Fergus, R.: Regularization of Neural Networks using DropConnect, *International Conference on Machine Learning* (2013).
[24] Frazao, X. and Alexandre, L.A.: DropAll: Generalization of Two Convolutional Neural Network Regularization Methods, *International Conf. on Image Analysis and Recognition*, LNCS.8814, pp.282–289 (2014).
[25] Ba, L.J. and Frey, B.: Adaptive Dropout for Training Deep Neural Networks, *Neural Information Processing Systems*, pp.3084–3092 (2013).

[26] Dolla, P., Christian, C., Schiele, B. and Perona, P.: Pedestrian Detection: An Evaluation of the State of the Art, *Pattern Analysis and Machine Intelligence*, Vol.34 (2012).
[27] LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P.: Gradient-Based Learning Applied to Document Recognition, *Proc. IEEE* (1998).
[28] Sermanet, P. and LeCun, Y.: Traffic Sign Recognition with Multi-scale Convolutional Networks, *International Joint Conference on Neural Network* (2011).
[29] Hinton, G.E., Srivastava, N., Krizhevsky, A., Ilya, S. and Salakhutdinov, R.: *Improving neural networks by preventing co-adaptation of feature detectors*, *Clinical Orthopaedics and Related Research*, Vol.abs/1207.0 (2012).
[30] Enzweiler, M. and Gavrila, D.M.: Monocular Pedestrian Detection: Survey and Experiments, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.31, No.12, pp.2179–2195 (2009).
[31] Houben, S., Stalkamp, J., Schlipsing, M., Salmen, J. and Igel, C.: The German Traffic Sign Detection Benchmark, *International Joint Conference on Neural Networks* (2013).
[32] Viola, P. and Jones, M.: Robust Real-Time Face Detection, *Computer Vision and Pattern Recognition* (2004).
[33] Ouyang, W. and Wang, X.: A Discriminative Deep Model for Pedestrian Detection with Occlusion Handling, *Computer Vision and Pattern Recognition* (2012).
[34] Dallar, P., Appel, R., Belongic, S. and Perona, P.: Fast Feature Pyramids for Object Detection, *Pattern Analysis and Machine Intelligence* (2014).
[35] Schwartz, W.R., Kembhavi, A., Harwood, D. and Davis, L.S.: Human Detection Using Partial Least Squares Analysis, *International Conference on Computer Vision* (2009).
[36] Dollar, P., Belongie, S. and Perona, P.: The Fastest Pedestrian Detection, *British Machine Vision Conference* (2010).
[37] Dollar, P., Tu, Z., Perona, P. and Belongie, S.: Integral Channel Feature, *British Machine Vision Conference* (2009).
[38] Dollar, P., Appel, R. and Kienzle, W.: Crosstalk Cascades for Frame-Rate Pedestrian Detection, *European Conference on Computer Vision* (2012).
[39] Park, D., Ramanan, D. and Fowlkes, C.: Multi Resolution Models for Object Detection, *European Conference on Computer Vision* (2010).
[40] Benenson, R., Mathias, M., Tuytelaars, T. and Gool, L.V.: Seeking the Strongest Rigid Detector, *Computer Vision and Pattern Recognition* (2013).
[41] Chen, G., Ding, Y., Xiao, J. and Han, T.: Detection Evolution with Multi-order Contextural Co-occurrence, *Computer Vision and Pattern Recognition* (2013).
[42] Sabzmeydani, P. and Mori, G.: Detecting Pedestrians by Learning Shapelet Features, *Computer Vision and Pattern Recognition* (2007).
[43] Wojek, C. and Schiele, B.: A Performance Evaluation of Single and Multi-Feature People Detection, *German Association for Pattern Recognition* (2009).
[44] Sermanet, P., Kavukcuoglu, K., Chintala, S. and LeCun, Y.: Pedestrian Detection with Unsupervised Multi-stage Feature Learning, *Computer Vision and Pattern Recognition*, pp.3626–3633 (2013).

福井 宏

2014 年中部大学工学部情報工学科卒業．現在，同大学大学院工学研究科情報工学専攻博士前期課程在学中．画像を用いた物体認識の研究に従事．



山下 隆義 （正会員）

2002 年奈良先端科学技術大学大学院大学博士前期課程修了，2002 年オムロン株式会社入社，2009 年中部大学大学院博士後期課程修了（社会人ドクター），2014 年中部大学講師，人の理解に向けた動画像処理，パターン認識・機械学習の研究に従事，2009 年画像センシングシンポジウム高木賞，2013 年電子情報通信学会情報・システムソサエティ賞，2013 年電子情報通信学会 PRMU 研究会研究推奨賞，2014・2015 年画像センシングシンポジウムオーディエンス賞．

山内 悠嗣 （正会員）

2012 年中部大学大学院博士後期課程修了，2010 年独立行政法人日本学術振興会特別研究員 DC，2012 年中部大学院博士研究員，2014 年中部大学助手．コンピュータビジョン・パターン認識の研究に従事．

藤吉 弘亘 （正会員）

1997 年中部大学大学院博士後期課程修了．1997～2000 年米カーネギーメロン大学ロボット工学研究所 Postdoctoral Fellow，2000 年中部大学講師，2004 年より同大学教授，2005～2006 年米カーネギーメロン大学ロボット工学研究所客員研究員．計算機視覚，動画像処理，パターン認識・理解の研究に従事．2005 年ロボカップ研究賞，2009 年情報処理学会論文誌コンピュータビジョンとイメージメディア優秀論文賞，2009 年山下記念研究賞，2010・2013 年画像センシングシンポジウム優秀学術賞，2013 年電子情報通信学会情報・システムソサエティ論文賞．

村瀬 洋 （正会員）

1978 年名古屋大学工学部電気電子工学科卒業，1980 年名古屋大学大学院電気電子工学専攻修士課程修了，1980 年 NTT（当時，日本電信電話公社）入社，1987 年名古屋大学大学院情報工学専攻工学博士取得（論文博士）．文字・図形認識，コンピュータビジョン，マルチメディア認識の研究に従事．1992 年米国コロンビア大学コンピュータ科学部客員研究員，2003 年名古屋大学大学院情報科学研究科メディア専攻教授．1985 年篠原学術奨励賞，1992 年テレコムシステム技術賞，1994 年 IEEE Best Paper Award：CVPR，1995 年山下記念研究賞，1996 年 IEEE Best Video Award：ICRA，2001 年高柳記念奨励賞，システムソサエティ論文賞，2002 年電子情報通信学会業績賞，2003 年文部科学大臣賞，2004 年 IEEE 論文賞 Trans. on Multimedia，画像認識理解シンポジウム MIRU2004 優秀論文賞，2005 年テレコムシステム技術奨励賞，2006 年 IEEE フェロー，論文賞（Best Industry Related Paper Award），2007 年 FIT2007 論文賞，Most Influential Paper over the Decade Award，電子情報通信学会フェロー称号授与，2009 年論文賞，2010 年前島賞，2012 年紫綬褒章．