

## 1. はじめに

歩行者検出や標識認識において、高い性能を実現した Deep Convolutional Neural Network(CNN) をベースとした手法が高い性能を実現したことから注目されている。これらの手法は精度向上のために CNN の層を深くしたり、複数の識別器を併用したりする方法が多い。そのため、これらの手法の CNN は複雑な構造になる傾向がある。本研究では、1つの CNN で高性能な歩行者検出と標識認識を目的とし、Dropout[1] のアルゴリズムをベースとした Random Dropout と Ensemble Inference Networks(EIN) を提案する。提案手法では、1) Random Dropout により歩行者や標識の隠れや姿勢変動に頑健な CNN を構築し、2) EIN により異なる全結合層の構造を持つ CNN を生成することで、誤検出と誤認識を削減する。

## 2. 提案手法

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。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。本研究では、学習時に用いる Random Dropout と、識別時に用いる Ensemble Inference Networks(EIN) を提案する。以下に、2つの提案手法について述べる。

### 2.1 Random Dropout

Dropout が各層の応答値を 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 1 に Random Dropout のアルゴリズムを示す。Random Dropout は、更新 1 回目のとき各層のユニットの削減率を 25% と 33% としている。そして、更新 2 回目では、50% と 67% 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るように、パラメータが更新され、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。

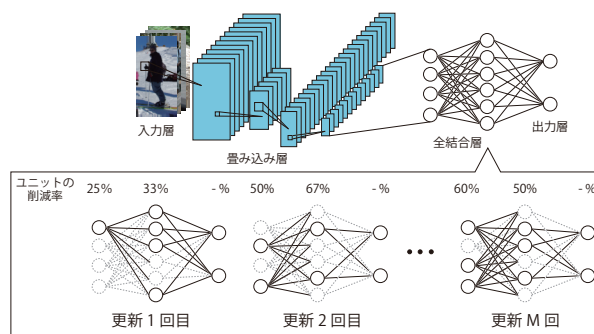


図 1: Random Dropout のアルゴリズム

### 2.2 Ensemble Inference Networks

EIN による識別処理は、図 2 のように 1) 畳み込み層と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。

#### 畳み込み層およびプーリング層

EIN の畳み込み層とプーリング層は、従来の CNN と同様の処理である。しかし、EIN は複数の異なる構成の全結合層に、畳み込み層とプーリング層で得られた特徴マップを入力する必要がある。そこで、畳み込み層とプーリング層を通した後の特徴マップを特徴ベクトルとして保持する。この保持した特徴ベクトルを生成した各全結合層に入力することで、畳み込み層とプーリング層の処理を繰り返し行う必要がなく、計算コストを削減できる。

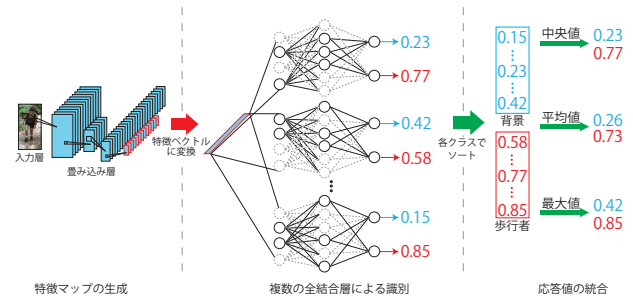


図 2: Ensemble Inference Networks のアルゴリズム

#### 全結合層の生成

従来の CNN による識別は、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出力する。EIN は、図 2 のように、学習したネットワークをもとにランダムに選択したユニットの応答値を 0 にした全結合層を  $N$  個生成する。そして、畳み込み層とプーリング層を通して得られた特徴ベクトルを、生成した全結合層に入力して各クラスのスコアを出力する。これらの処理により、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。

#### 最終出力の統合

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  $S_c$  とする。EIN の各クラスに対する最終的な出力  $S_c$  は、中央値  $S_c^{Median}$  や平均値  $S_c^{Mean}$ 、最大値  $S_c^{Max}$  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。

## 3. 評価実験

本研究では、Caltech Pedestrian Dataset と German Traffic Sign Recognition Benchmark (GTSRB) Dataset を用いて、提案手法の性能を評価する。Random Dropout 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、最も性能が良い結果を用いて精度を比較する。EIN の性能評価では、生成する全結合層のネットワークの数を 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比較することで最終的な性能を評価する。また、CNN をベースとした歩行者検出法である Switchable Deep Network(SDN)[2] とユニット数を比較することで、計算コストを評価する。

比較する手法は、各データセットのベンチマークを使用する。Caltech Pedestrian Dataset は 8,723 枚の評価サンプルを使用し、GTSRB Dataset では 12,631 枚の評価サンプルを使用する。各データセットに対して、畳み込み層 3 層と全結合層 3 層の CNN を使用する。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 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。

### 3.1 提案手法の導入による性能評価

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 3(a), (b) に示す。ここで、図 3(a) 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 の時の Miss rate を示している。図 3(a) より、Caltech Pedestrian Dataset において、Dropout は 70% の時に Miss rate が 40.45% である。提案手法である Random Dropout は削

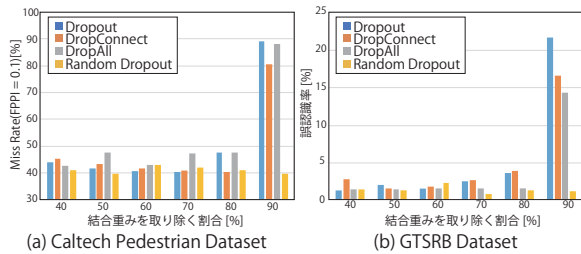


図 3 : Random Dropout の性能評価

減率の範囲が0%から70%の時に39.65%のMiss rateであり、最もMiss rateが低いことが確認できる。図3(b)のGTSRB Datasetも同様で、DropAllは削減率が50%の時Miss rateが1.39%であるのに対して、提案手法であるRandom Dropoutは削減率の範囲が0%から70%の時に0.81%の誤認識率であり、最も誤認識率が低い。

EINの最適な全結合層数、最終出力決定方法を各評価において比較する。図4(a)にCaltech Pedestrian Datasetにおける全結合層数に対するMiss rateの変化、図4(b)にGTSRB Datasetにおける全結合層数に対する誤認識率の変化を示す。なお、図4(a)はFPPが0.1の時の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4(a), (b)の縦軸上の各点に相当する。図4(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、全結合層の生成数が33のときMiss rateが37.77%で最も精度が良い。また、従来のDropoutを用いたCNNと比較したとき、Miss Rateが最大で7.73%減少していることがわかる。標識認識のベンチマークGTSRB Datasetでの結果を図4(b)に示す。図4(b)より、Random Dropoutを導入し、EINの統合方法に最大値を使用して、全結合層の生成数が17のとき認識率が99.17%で最も良い。また、EINを導入することで認識性能が0.1%向上していることが確認できる。図4(b)より、多クラス認識では、EINの最終出力の方法に最大値を用いることで、認識性能が向上している。

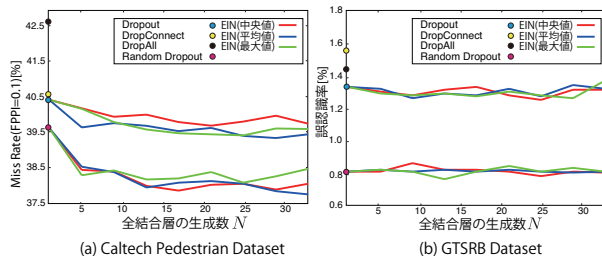


図 4 : EIN の性能評価

### 3.2 従来法との比較

図5(a)にCaltech Pedestrian Dataset, 表1にGTSRB Datasetの比較結果を示す。図5にCaltech Pedestrian Datasetの比較結果を示す。Caltech Pedestrian Datasetでは、FPPが0.1のときに従来のCNNよりMiss Rateを8.54%改善している。CNNをベースとした歩行者検出方法であるSwitchable Deep Network(SDN)のMiss rateは37.87%であり、同等以上の性能を実現していることが確認できる。

表1より、Random DropoutとEINを導入した際の認識率が99.22%であり、人の認識性能(Human Performance)より高性能に認識でき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。

表2に提案手法とSDNのユニット数の比較を示す。表2より、CNNのユニット数は約22万個、EINによる生成数を33にした場合のユニット数は約23万個である。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

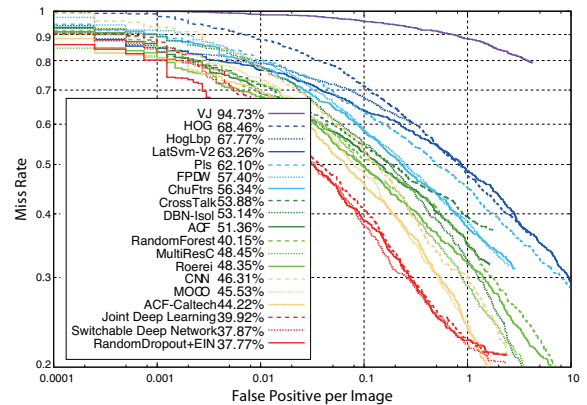


図 5 : 従来法との比較 (Caltech Pedestrian Dataset)

表 1 : 従来法との比較 (GTSRB Dataset)

手法	認識率 [%]
Multi-Column Deep Neural Network	99.46
Random Dropout + EIN	<b>99.22</b>
Human Performance	98.84
Multi-Scale CNN	98.31
Random Forests	96.14

リング層の共有化と全結合ユニットの削減により、ユニット数の増加を大幅に抑えることができる。また、SDNのユニット数と比較した場合、ネットワークの規模が約3分の1であることが確認できる。

表 2 : 提案手法とSDNのユニット数の比較

手法	ネットワークの構造	ユニット数
CNN	Conv(40 × 9 × 5, 64 × 5 × 3, 32 × 3 × 3)	226,002
	Full(1,000, 500, 100, 2)	
提案手法	Conv(40 × 9 × 5, 64 × 5 × 3, 32 × 3 × 3)	232,402
	Full(1,000, 500, 100, 2) × 33	
SDN	Conv(64 × 9 × 9)	604,402
	Full(2,100)	
	Switchable層 Full(2,000, 2)	

## 4. おわりに

本研究では、CNNの汎化性能の向上を目的とした2つの手法を提案した。提案手法の導入により、歩行者検出と標識認識において高精度な識別を実現した。今後は、歩行者を検出する際に歩行者属性情報を用いることで、より高精度な歩行者検出を実現する。

## 参考文献

- [1] G.E.Hinton, *et al.*, "Improving neural networks by preventing co-adaptation of feature detectors", CoRR abs/1207.0580, 2012.
- [2] P. Luo, *et al.*, "Switchable Deep Network for Pedestrian Detection", CVPR, 2014.

## 研究業績

- [1] H. Fukui, T. Yamashita, Y. Yamauchi, H. Fujiyoshi, H. Murase, "Pedestrian Detection Based on Deep Convolutional Neural Network with Ensemble Inference Network", IEEE IV Symposium, 2015.

(他 学会口頭発表 3件)

## 受賞

- [1] PRMU アルゴリズムコンテスト 2015 最優秀賞
- [2] SSII 2015 インタラクティブセッション賞