歩行者検出における DCNN のネットワーク規模の縮小による計算コスト削減

EP13063 小芝駿王　　指導教授：藤吉弘亘

1. はじめに

認識問題に用いられる Deep Convolutional Neural Network(DCNN) は，多層化され，高い認識精度を達成している．その反面，学習及び評価時の計算コストが高く，処理時間を要する．歩行者検出は 2 クラス分類であり，多層のネットワーク構造を必要とせずコンパクト化できる可能性がある．そこで，本研究では精度を維持したまま，どこまで DCNN のネットワーク規模を小さくできるか検討する．

2. AlexNet

AlexNet[1] は 1000 クラスの物体認識を目的とした DCNN であり，ImageNet Large Scale Visual Recognition Challenge 2012 でトップの精度を達成した．ネットワーク構造は図 1 に示すように畳み込み層 5 層，全結合層 3 層である．歩行者検出のような 2 クラスの分類問題では，AlexNet はネットワーク規模が大きく，計算コストや処理時間がかかる．

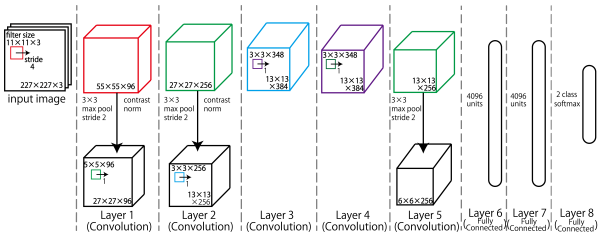

図 1：AlexNet の構造

3. ネットワーク規模を縮小した歩行者検出

本研究では図 2 に示すように Locally Decorrelated Channel Features(LDCF) とネットワーク規模を縮小した DCNN を組み合わせて 2 段階の歩行者検出を行う．はじめに，LDCF で歩行者候補領域を検出する．ここで，DCNN での処理において領域の端付近の特徴が消滅する問題を防ぐために歩行者候補領域の端にオフセットをつける．次に歩行者候補領域を DCNN に入力して歩行者か否か判定する．

図 2：2 段階の歩行者検出の流れ

1 つの歩行者領域に対して，AlexNet は識別に 217.4ms 要するため，本研究では計算コスト削減を目的とし DCNN のネットワーク規模を縮小を検討する．縮小させる観点として以下がある．

- 層数の削減
- フィルタ数の削減
- フィルタサイズの縮小
- 全結合層のユニット数の削減

これらを組合わせてネットワークサイズを縮小し，処理時間の短縮とメモリ使用量削減を実現する．

4. 評価実験

歩行者検出の精度を維持しつつ，DCNN のネットワーク規模をどこまで縮小可能か実験を行う．

4.1. 実験概要

本実験では，データセットに Caltech Pedestrian Dataset を用いる．DCNN の学習には LDCF で検出した歩行者 121,266 枚，背景 355,487 枚の画像を使用する．また，バッチサイズは 128，学習更新回数 500,000 回とする．本実験では，メモリ使用量，1 枚当たりの DCNN の処理時間を比較するとともに検出精度の推移を確認する．

4.2. 実験結果

図 3 に LDCF と AlexNet，縮小をした DCNN の検出精度を示す．AlexNet および縮小した DCNN の検出精度は，LDCF と比べて向上しており，後段の DCNN が有用であることがわかる．

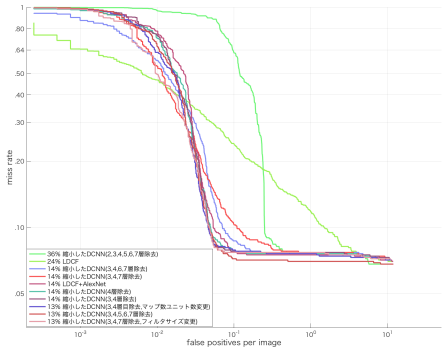

図 3：歩行者検出精度の比較

表 1 に計算コストとメモリ使用量，検出精度の比較を示す．表 1 の畳み込み層を削除したモデル 1，2，7，8 では使用メモリ量，処理時間の削減を実現した．また，結合層を削除したモデル 4，6 でも使用メモリ量，処理時間の削減を実現した．更に，フィルタ数を削減した縮小モデル 3，フィルタサイズを変更したモデル 5 でも使用メモリ量，処理時間の削減を実現した．しかし，畳み込み層 1 層，全結合層 1 層のモデル 8 は，検出精度が大きく低下した．

表 1：計算コストの比較

	層数		フィルタ	フィルタ	メモリ	処理時間	精度
	畳込	全結合	数	サイズ	[MB]	[ms]	[%]
AlexNet	5	3	96〜384	11×11〜3×3	3500	217.4	85.75
モデル 1	4	3	96〜384	11×11〜3×3	2400	40.3	86.15
モデル 2	3	3	96，256	11×11〜3×3	1900	16.0	86.18
モデル 3	3	3	16，32	11×11〜3×3	1800	14.0	86.78
モデル 4	3	2	16，32	11×11〜3×3	1500	10.1	85.54
モデル 5	3	2	16，32	5×5，3×3	900	4.2	87.31
モデル 6	3	1	16，32	5×5，3×3	500	3.6	85.52
モデル 7	2	1	16，32	5×5，3×3	200	2.1	87.10
モデル 8	1	1	16	5×5	119	1.6	64.32

精度を維持した最小の DCNN の構造を図 4 に示す．入力サイズは 57×57，畳み込み層は 2 層，全結合層は 1 層である．AlexNet と比べ，処理時間を約 215.3ms，メモリ使用量を 3300MB 削減した．

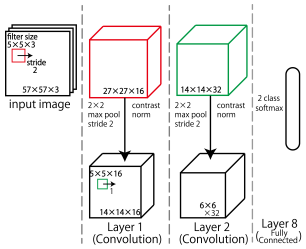

図 4：精度を維持した最小の DCNN の構造

5. おわりに

AlexNet のネットワーク規模を縮小することにより，精度を維持したまま計算コストを削減する縮小方法を実現した．今後は，ネットワークの縮小を続けるとともに検出精度の向上を目標とする．

参考文献

[1] J. Hosang, *et al.*, “Taking a Deeper Look at Pedestrians”, CVPR, 2016.