

## 多値化畳み込みニューラルネットワークによるリアルタイム物体認識

日高 憲太 指導教員：小林 邦和

## 1 はじめに

ロボカップは、自律移動型ロボットを用いたサッカー競技を通して、ロボット工学と人工知能の融合、発展を目指すランドマークプロジェクトである。その目標に向けて、サッカー標準プラットフォームリーグでは、年々ルールが変更されている。今までの大きな変更点として、ゴールの色が黄色から白色に変更、またボールの色が白黒ボールに変更、また試合の会場が屋内から屋外に変更された。このように多くのルール改正のある中、現在では自然光が入る照明環境下でボールやロボットなどフィールド上の物体を認識する必要がある。自然光が入ることで、物体への光の当たり方や影により、認識が難しくなる。

ロバスタな手法として、近年深層学習 (Deep Learning) が注目を集めている。通常、Deep Learning は膨大な計算コストが必要となるため、GPU コンピューティングの並列化処理が用いられている。しかしながら、一般的にロボットは GPU を搭載しておらず、CPU のみで物体認識を行わなければならないという制約がある。そのため Deep Learning をロボットへ実装することが困難という問題がある。

そこで本研究では、ロボットの CPU のみで物体認識を行うために、ロボカップ環境下で畳み込みニューラルネットワーク (Convolutional Neural Network) [1] の 1 種である You Only Look Once (YOLO) [2] モデルの改良、並びにラプラシアンピラミッド [3] を用いた入力画像のサイズ削減を提案する。

## 2 関連技術

CNN とは、畳み込み層と活性化関数、プーリング層が繰り返されるフィードフォワード型のニューラルネットワークである。提案手法では CNN とラプラシアンピラミッドを利用している。

以下に CNN のプーリング層とラプラシアンピラミッドについて説明する。

## 2.1 プーリング層

プーリング層は画像サイズを削減する役割を担う層である。プーリング層に入力した後は、情報が削減されるが次の層の処理速度向上させることができる。プーリングの一種である Average Pooling では、一定の領域内の画素値を平均することで、1 つの画素にする。Average Pooling は式 (1) で計算される。

$$y_{ij} = \frac{1}{|P_{ij}|} \sum_{(p,q) \in P_{ij}} x_{pq} \quad (1)$$

ここで  $x_{pq}$  は、画像中の座標  $(p, q)$  の画素値、 $P_{ij}$  は  $L \times L$  画素の正方領域  $\{i+i', j+j' | i' = 0, \dots, L-1, j' = 0, \dots, L-1\}$  である。

## 2.2 ラプラシアンピラミッド

ラプラシアンピラミッドとは、多重解像度表現に基づいた階層表現を取得する手法である。縦幅  $h$ 、横幅  $w$  の入力画像  $G_i$  に対して、式 (2) を適用することでラプラシアンピラミッド  $L_i$  を作成していく。

$$\begin{aligned} L_i &= G_i - G'_i \\ G'_i &= \text{UpSampling}(G_{i-1}) \\ G_{i-1} &= \text{DownSampling}(G_i) \end{aligned} \quad (2)$$

ここで、 $G_{i-1}$  は  $G_i$  を Down Sampling を行った画像、 $G'_i$  は画

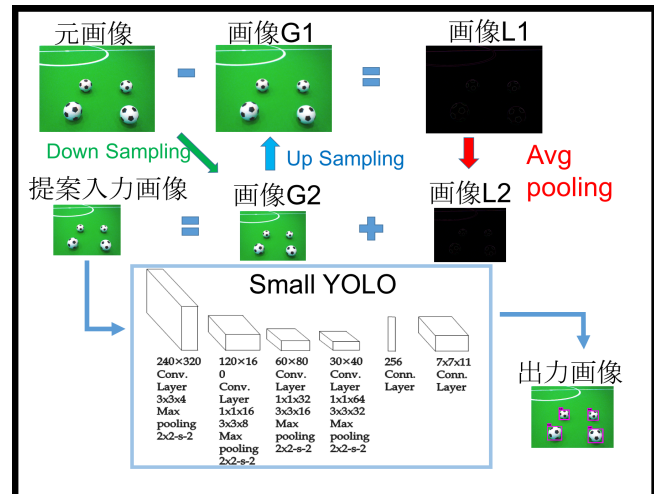


図1 提案手法の概要図

像  $G_{i-1}$  を Up Sampling を行った画像である。Down Sampling は、正の整数  $N$  を利用した  $N \times N$  のガウシアンカーネルで画像を畳み込み、偶数行、偶数列を新しい画像とする処理である。Down Sampling では画像サイズを  $h \times w$  から  $h/2 \times w/2$  に圧縮することができる。

Up Sampling は、画像  $G'_{i-1}$  の各行、各列の後に 0 の値を追加し、Down Sampling のときと同じガウシアンカーネルのそれぞれのピクセルの値を 4 倍したもので、偶数行、列にその結果を畳み込むことで新しい画像とする処理である。Up Sampling では画像サイズを  $h/2 \times w/2$  から  $h \times w$  に次元を拡張することができる。

## 3 提案手法

本研究で提案する手法は、入力画像のサイズを削減し、その画像に元画像から作成した強調画像を加える手法である。画像サイズを削減する際に情報が削減されてしまい、高周波成分を含むエッジは弱まってしまう。そのため、ラプラシアンピラミッドを用いることにより、Down Sampling する前の画像を利用することで、高周波成分が含まれる領域を強調する。また本研究では、処理速度の速い Fast YOLO モデルの層数、フィルタサイズ、フィルタ枚数を削減し、CNN(以下、small YOLO と呼ぶ)を作成した。

提案手法のアルゴリズムについて説明していく。

## 3.1 アルゴリズム

図 1 に提案手法の概要図を示す。同図では画像サイズを 1 段階削減する概要図である。また small YOLO は図 1 のモデルである。

提案手法のアルゴリズムを以下に示す。ここで、 $n$  は階層数、Average Pooling を行った画像は添字の数字を 1 つ増やし、'(プライム記号)を増やすものとする。例えば  $L'_1$  を Average Pooling した場合、 $L''_2$  となる。 $n$  段階行うことで画像サイズは、画像サイズは  $w \times h$  から  $w/(2n) \times h/(2n)$  に削減される。

1. 元画像を  $n$  段階削減する。
2. 元画像  $G_1$  に対してガウシアンピラミッド  $G_1 \sim G_n$  を作成する。
3. 元画像  $G_1$  に対してラプラシアンピラミッド  $L_1 \sim L_n$  を作成する。

表 1 処理速度

モデル	処理速度 (fps)
Model448	1.10
Model320	15.0
Model320L	13.9
Model160	48.5
Model160L	42.7

表 2 評価指標 (閾値 0.5)

モデル	適合率	再現率
Model448	80.4	52.6
Model320	88.7	23.3
Model320L	89.5	34.0
Model160	71.7	9.43
Model160L	92.7	23.3

表 3 評価指標 (閾値 0.4)

モデル	適合率	再現率
Model448	63.0	65.8
Model320	79.1	45.9
Model320L	71.9	51.4
Model160	71.6	25.1
Model160L	79.2	44.4

表 4 評価指標 (閾値 0.3)

モデル	適合率	再現率
Model448	46.9	76.4
Model320	56.7	63.3
Model320L	52.4	62.5
Model160	61.0	50.6
Model160L	62.2	62.5

- $G_n$  の画像サイズと同サイズの強調画像を作成するために、ラプラシアンピラミッドの全ての画像を Average Pooling を用いて、強調画像の集合  $L = \{ L_n''', \dots, L_n'', L_n' \}$  を作成する。
- $G_n$  に  $L$  の集合の画像をすべて加え、CNN の入力画像とする。

#### 4 計算機シミュレーション

本研究の計算機シミュレーションについて、シミュレーション環境、本研究で扱う学習モデル、パラメータ設定、評価方法、実験結果の順で説明していく。

##### 4.1 シミュレーション環境

本研究の問題設定はロボカップサッカー標準プラットフォームリーグ環境でボールとロボットとそれ以外の 3 クラスを認識する。今回使用する CPU はロボットの CPU ではない Intel®Xeon(R) CPU E5-2690 v4 を使用した。また YOLO モデルを実装するにあたり、Darknet というフレームワークを用いた。

また学習データは  $640 \times 480$  のサイズでサッカー標準プラットフォームリーグ環境で撮影したボールのみが存在する画像を 500 枚、ロボットのみが存在する画像を 400 枚のデータを用い、各モデルに対して 120000 回学習する。テストデータは、ボールとロボットそれぞれ 200 枚ずつで評価する。本研究では、学習時と認識時に画像の輝度値に影響されないように学習、テスト画像の前処理として、減算正規化を行った。

##### 4.2 学習モデル

処理速度と精度を比較するため、学習するモデルは 5 つで評価する。1 つ目は入力が  $448 \times 448$  の Fast YOLO である (以下、Model448 と呼ぶ)。2~5 つ目は small YOLO を利用する。それぞれのネットワークは入力サイズが  $320 \times 240$ 、 $160 \times 120$  の Model320, Model160, また small YOLO にラプラシアンピラミッドを用い強調したモデルを Model320L, Model160L とする。

##### 4.3 パラメータ設定

本研究で利用したパラメータは、ガウシアンフィルタの縦横幅  $N$  を 5、またそれぞれの YOLO のモデルの Pooling の正方形領域  $L$  を 2、学習係数を 0.00001、モーメント係数は 0.9、decay 係数は 0.0005、YOLO モデルのグリッド分割数を  $8 \times 8$ 、各グリッドの矩形領域の推定数を 5 とする。

##### 4.4 評価方法

本研究の評価方法として、矩形領域の信頼度の閾値として 0.3 ~ 0.5 の間を 0.1 きざみで設定し、処理速度とボールとロボットの平均の再現率と適合率で評価する。ただし、矩形領域の信頼度とは、矩形領域に物体 (本研究では、ボールとロボット) が存在する確率とする。

##### 4.5 実験結果

処理速度の結果を表 1、適合率、再現率の評価指標の結果を表 2~4 に示す。

処理速度は、提案手法の方が Fast YOLO より速い。また同

サイズのモデルにおいて提案手法を利用することでラプラシアンピラミッドの処理が余分にかかるが入力サイズが  $320 \times 240$ 、 $160 \times 120$  の場合では small YOLO に比べ、12% 以内に悪化率が収まっている。

再現率は入力サイズが  $160 \times 120$  の場合、提案手法の有無によって、改善できていることが確認出来る。だが入力サイズが  $320 \times 240$  の場合、提案手法の方が改悪する場合がある。

適合率は入力サイズが  $160 \times 120$  の場合、再現率と同じように提案手法の有無によって、改善できていることが確認出来る。だが入力サイズが  $320 \times 240$  の場合、提案手法の方が改悪される場合がある。

実験結果に対する考察として、提案手法を用いる場合において、1 段階画像サイズを縮小することで、再現率と適合率、改悪される場合があることを確認した。だが画像サイズを 2 段階、縮小することにより、再現率と適合率が改善されていることを確認した。これらより、画像サイズを 1 段階縮小した場合と 2 段階縮小した場合の CNN の畳み込み層の出力を可視化することで特徴抽出の違いを確かめる必要がある。

ロボカップサッカー標準プラットフォームリーグの環境において、画像サイズが  $160 \times 120$ 、矩形領域の閾値を 0.3, 0.4 にすることで適合率と再現率が small YOLO より提案手法の方が良いと考えられる。

#### 5 おわりに

本研究では、ロボットに CNN を実装するために、small YOLO モデルにラプラシアンピラミッドを導入した。その結果、提案手法では Fast YOLO と比べて処理速度は向上し、small YOLO において提案手法を利用することで、利用していないモデルに比べ、画像の入力サイズによって F 値が改善することを確認できた。本研究では、ロボカップサッカー標準プラットフォームリーグの環境を想定しているため、ボールとロボットの認識を行った。

今後の課題として、画像サイズによって F 値が改悪する場合があるためその原因の追究、またロボットの低スペック CPU ではリアルタイム処理が困難と思われるため、画像サイズをさらに削減するなどの工夫が必要となる。

#### 参考文献

- [1] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel 『Backpropagation applied to handwritten zip code recognition』, Neural computation, 1 巻-4 号, 541/551(1989)
- [2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi 『You only look once: Unified, real-time object detection』, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 779/788(2016)
- [3] P. Burt, and A. Edward 『The Laplacian pyramid as a compact image code』, IEEE Transactions on communications, 31 巻-4 号, 532/540(1983)
- [4] 日高 憲太, 鈴木 拓夫, 小林 邦和 『ラプラシアンピラミッドを導入した CNN によるロボットの物体認識』, システム・情報部門学術講演会ポスターセッション, SS09-14(2017)