

## 多値化畳み込みニューラルネットワークを用いた画像分割による物体の認識性能の向上

情報科学科 山田 裕史

指導教員：小林 邦和, 鈴木 拓央

## 1 はじめに

深層学習を用いた物体検出の技術は年々進化している。しかし、未だに深層学習は基本的に多大な計算コストがかかるため Graphics Processing Unit(GPU) を使用して処理をすることが一般的である。しかし、You Only Look Once(YOLO) モデルは、従来の物体領域候補を複数抽出した後、それぞれの候補領域に対してクラス分類する R-CNN といったモデルと違い、1つの CNN で物体の検出からクラス分類を行うことができるモデルである。その結果、CPU のみでリアルタイム物体検出ができるようになってきている。しかし、ロボカップサッカー標準プラットフォーム NAO に搭載されている CPU では多大な計算コストを処理するのが難しいのが現状である。そこで本研究では、NAO 搭載 CPU のみで物体検出行えるモデルの作成を目指し、計算コストの削減を目的として、画像分割のアイデアを導入する。

## 2 提案手法

## 2.1 ネットワークモデル

本研究では YOLO モデルの計算コストの削減を目指すため YOLOv3-tiny モデルに XNOR-Net というバイナリネットを組み込んだ YOLOv3-tiny-XNOR モデルを使用し、そのフィルター数を削減した。図 1 に CNN 内部の処理内容を示す。この図では畳み込み層とプーリング層のフィルターサイズとスライド数を示している。

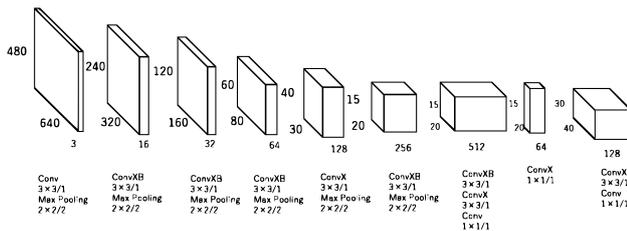


図 1 ネットワークモデル

## 2.2 画像分割による入力サイズの削減

本研究で提案する手法は、入力画像を分割して処理することによって、より小さな画像サイズで物体認識をおこない物体検出の計算コストを削減するものである。分割方法は図 2 に示す。縦横それぞれ入力サイズの  $\frac{2}{3}$  の大ききで分割した。

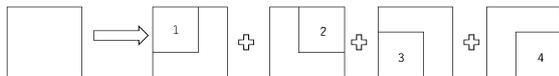


図 2 分割方法 (4 分割)

## 3 計算機シミュレーション

## 3.1 環境

現在、NAOv5 に搭載されている CPU では計算能力が足りず実験できないため、今後 SPL に導入予定の NAOv6 に実装を目指す。したがって、シミュレーションに使用する CPU は、NAO に搭載されていない CPU で Intel(R) Xeon(R) CPU E5-2690 v4 を使用した。また YOLO モデルを使用するにあたり Darknet というフレームワークを使用した。

本研究の問題設定としては、SPL 環境で使用されるボールの 1 クラスを認識する。また、NAO の CPU 学習データには  $640 \times 480$  のボールが複数個写っている画像 128 枚を用いた。また、評価には SPL で使用されるボールが写っている画像 64 枚

で評価した。比較対象の分割した手法では、元の学習データと同じ画像を使用し、オーバーラップありの  $640 \times 480$  の画像を 4 分割した画像で 512 枚を学習させた。また評価データも同様に 4 分割した画像 256 枚を使用して評価した。

## 3.2 シミュレーション方法

学習するモデルは入力サイズが異なっている 3 種類とその画像を 4 分割したものを使用する。それぞれの入力サイズとモデル名は表 1 のようにする。例えばモデル A を分割させたモデルがモデル A' となる。

表 1 モデル名の定義

モデル名	分割なし	モデル名	分割あり
モデル A	$640 \times 480$	モデル A'	$448 \times 320$
モデル B	$384 \times 288$	モデル B'	$256 \times 192$
モデル C	$128 \times 96$	モデル C'	$96 \times 64$

## 3.3 評価方法

本研究の評価方法としては、処理速度、再現率、適合率、F 値、IOU(Intersect Over Union) の 5 項目で評価する。

## 3.4 実験結果

実験結果を表 2 に示す。分割することによって、処理速度はモデル A' は A と比較して +126%, モデル B' は B と比較して +140%, モデル C' は C と比べて +15% それぞれ改善した。また、再現率、適合率、F 値、IOU はそれぞれモデル A は 5% 以内、モデル B は 10% 以内に悪化率が収まっている。しかし、モデル C は、30% 以上の悪化率になっていた。

実験結果に対する考察として、画像分割することによって処理速度が改善することが確認できた。しかし、再現率、適合率、F 値、IOU などの精度は入力サイズが小さくなるにつれて悪化率が大きくなることが確かめられた。処理速度と適合率などの精度から、SPL 環境下ではモデル B の入力サイズが適切だという結果となった。

表 2 実験結果

モデル名	処理速度 (fps)	適合率 (%)	再現率 (%)	F 値 (%)	IOU (%)
モデル A	2.94	96	93	95	69
モデル A'	6.66	92	89	91	67
モデル B	8.33	90	89	90	64
モデル B'	20.0	81	86	84	62
モデル C	83.0	58	55	56	35
モデル C'	96.0	40	37	38	23

## 4 おわりに

本研究では、CNN を用いたリアルタイム物体検出を NAO に搭載されている CPU のみで処理を可能にするために、画像分割という手法をも用いて、入力サイズを削減した。画像分割することによって計算コストの削減に成功した。しかし、入力サイズによっては精度が著しく下がることが確認できた。

## 参考文献

- [1] REDMON, Joseph; FARHADI, Ali. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767, 2018.
- [2] RASTEGARI, Mohammad, et al. Xnor-net: Imagenet classification using binary convolutional neural networks. In: European Conference on Computer Vision. Springer, Cham, 2016. p. 525-542