

# 物体認識アルゴリズムを FPGA に実装するための基礎研究

吉元 裕真<sup>†</sup> 田向 権<sup>†</sup>

<sup>†</sup> 九州工業大学電子大学 大学院生命体工学研究科

## 1. はじめに

近年、一般物体認識の為の強力な手法として畳み込みニューラルネットワーク (Convolutional Neural Network; CNN) を用いた深層学習が注目を集めている。一般物体認識とは、実世界シーンの画像に含まれる物体を一般的な名称で認識するタスクであり、CNN を用いた高精度な物体認識アルゴリズムは家庭用ロボットや組み込みシステムへの導入が期待されている。しかしこれらの導入先では実時間での処理が求められるが、ソフトウェアのみで実時間処理を行うことは難しい。そこで専用のハードウェアアーキテクチャによる高速化が必要不可欠である。

本研究では、内部回路を書き換え可能なハードウェアである Field Programmable Gate Array (FPGA) に物体認識アルゴリズムを実装し、高精度な一般物体認識の実時間処理実現を目指す。本稿ではその前段階として、物体認識アルゴリズムを FPGA に実装する為のハードウェア指向化について検討する。

## 2. You only look once (YOLO)<sup>[1]</sup> とは

YOLO とは、CNN を用いた一般物体認識システムの 1 つであり、既存の手法よりも高速であることが報告されている。YOLO のシステムを図 1 に示す。システムは次のように動作する。

- 1) 画像を一定サイズにリサイズする。
- 2) リサイズ後の画像に CNN を適用し、出力として画像に含まれるオブジェクトのラベル、位置情報、サイズ情報、信頼度を得る。
- 3) 画像を  $S \times S$  のセルに分割し、CNN から得られた情報を元に、セル毎に物体検出とクラス分類を行う。
  - 3-a) 物体検出ではオブジェクトの位置情報、サイズ情報、信頼度情報から物体の位置を推測し、バウンディングボックスを設置する。
  - 3-b) クラス分類では CNN より得られるラベル情報から各セルがどの物体であるか推測し、分類を行う。
- 4) 3)の結果を組み合わせ、一般物体認識を行う。

YOLO ではクラス確率を用いているなど、一部ハードウェア実装に向かない点がある。本研究では、YOLO をハードウェア指向に再設計し、FPGA に実装する。

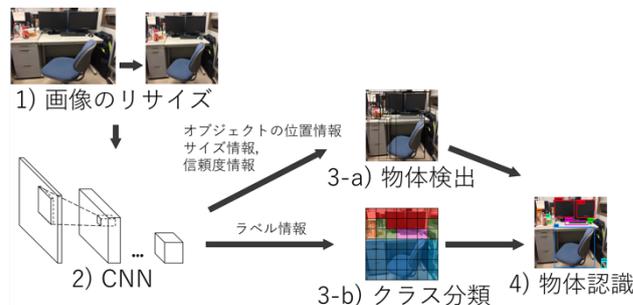


図 1 物体認識アルゴリズム YOLO

## 3. 実験

FPGA 実装の前段階として YOLO による物体認識の際、CNN において演算に時間のかかる層を調べた。結果を表 1 に示す。ここから Max Pooling 層前の Convolutional 層の演算量が多いと分かった。

表 1 YOLO の CNN における各層の演算時間比較

Layer	Filter	Time(sec)	Layer	Filter	Time(sec)	Layer	Filter	Time(sec)
1	conv	0.260	11	conv	0.358	21	conv	0.042
2	max	0.022	12	conv	0.041	22	conv	0.374
3	conv	<b>0.571</b>	13	conv	0.363	23	conv	0.042
4	max	0.008	14	conv	0.040	24	conv	0.366
5	conv	0.032	15	conv	0.360	25	conv	0.756
6	conv	0.367	16	conv	0.040	26	conv	0.180
7	conv	0.084	17	conv	0.364	27	conv	0.180
8	conv	<b>1.399</b>	18	conv	0.078	28	conv	0.181
9	max	0.005	19	conv	<b>1.419</b>	29	conv	0.339
10	conv	0.040	20	max	0.002	30	conv	0.007

## 4. まとめ

本稿では物体認識アルゴリズムである YOLO の FPGA 実装を提案した。しかし元アルゴリズムでは確率を用いているなど、一部ハードウェア実装に向かない点がある。また実験から Max Pooling 層前の Convolutional 層の演算量が多いことが分かった。今後は演算量の多い部分を短時間で処理できる工夫を加えつつ、まずはソフトウェアとしてハードウェア指向アルゴリズムを実現し、将来的に FPGA での実装を目指す。

## 参考文献

- [1] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, "You only look once: Unified, real-time object detection," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016, p. 779-788, 2016.