

## VC-MRAM での実装を目指した 量子化 YOLO モデル開発と書き込みエラー耐性評価

Development of quantization YOLO model and WER tolerance evaluation for VC-MRAM  
implementation

九工大生命体工<sup>1</sup>, 産総研<sup>2</sup> ○ヨー ヨンジェ<sup>1</sup>, 田向 権<sup>1</sup>,

野村 修<sup>1</sup>, 荒井 礼子<sup>2</sup>, 今村 裕志<sup>2</sup>, 森江 隆<sup>1</sup>

LSSE, Kyushu Inst. Tech<sup>1</sup>, AIST<sup>2</sup> ○Yoengjye Yeoh<sup>1</sup>, Hakaru Tamukoh<sup>1</sup>,

Osamu Nomura<sup>1</sup>, Hiroko Arai<sup>2</sup>, Hiroshi Imamura<sup>2</sup>, Takashi Morie<sup>1</sup>

E-mail: yeoh.yoeng-jye542@mail.kyutech.jp

近年、人工知能の発展は著しく、ロボットなど様々な分野に広く適用されている。ロボットでの応用のひとつとして、物体検出と物体認識が挙げられる。特に YOLO<sup>[1]</sup>は、カメラで捉えた物体の画像内の位置と種類を同時に取得することができる。ここで、物体の位置は座標情報であるため連続値を出力する回帰問題、物体の種類は 0, 1 の二値ラベルを出力する分類問題である。YOLO は同一ネットワークで回帰と分類問題を解くことができる。しかし YOLO は、計算コストが高く、ロボットへの実装を直接行うことは難しい。YOLO のような高速・高精度な物体検出能力を実装し、ロボットの性能を向上させるためには、計算コストや消費電力を下げることも必要となる。

電圧駆動型の磁気ランダムアクセスメモリ (VC-MRAM) は、印加電圧によって磁性層の磁気異方性を制御することにより、高速で低消費電力な書き込みを行うことができるため、ロボットへの応用も期待できる。現状では書き込みエラー率 (WER) は  $10^{-3} \sim 10^{-4}$  程度、最良の場合でも  $10^{-6}$  程度<sup>[2]</sup>であるため、DRAM 置き換え用途としては WER のさらなる改善が求められる。一方、人工知能・機械学習用途においては WER への要請は緩和できると考えられる。

そこで本研究では VC-MRAM での実装を目指し、量子化 YOLO モデル開発と書き込みエラー耐性を評価した。まず、従来の浮動小数点演算ベースの計算コストが膨大なため、計算や記憶に必要なビット数を量子化し削減する。本研究では、YOLO の特徴抽出層を二値化する一方で、連続値の出力が必要な回帰分類層を 3~8 ビットで量子化し、固定小数点演算へと変換した上で演算精度を評価した。ロボット応用を見越し、ホームロボットの国際競技会 World Robot Summit で使われているデータセットを用いて YOLO を構築した。また、VC-MRAM の WER 耐性評価のため、最悪の場合を仮定し YOLO の各層での演算結果を VC-MRAM に書き込む際に、確率  $10^{-3}$  のノイズを入れ評価した。

実験の結果、回帰分類層の整数部を 4 ビット量子化した際、評価指標である mAP (mean Average Precision) がノイズ無しの場合は 70.71%, WER =  $10^{-3}$  となるノイズを導入した場合には 69.08% となり、1 ポイント程度の劣化で量子化 YOLO への VC-MRAM の適用が可能であることを見出した。

謝辞：この成果は、国立研究開発法人新エネルギー・産業技術総合開発機構(NEDO)の委託業務 (JPNP16007) の結果得られたものです。

[1] J. Redmon, et al., (2016). IEEE-CVPR (pp. 779-788).

[2] T. Yamamoto et al., J. Phys. D: Appl. Phys. 52 (2019) 164001.