

重み量子化時の識別精度低下を抑制する深層学習向け重み正規化手法の検討

1190016 井上 裕太 (集積システム研究室)

(指導教員 密山 幸男 准教授)

1. はじめに

深層学習は膨大な演算量を必要とする。そこで、演算速度と省電力化のためにハードウェア化し、回路規模削減と省メモリ化のためにパラメータを量子化する研究が行われているが、識別精度が低下する問題がある[1][2]。本研究では、量子化を行った値の演算とビットマスクを行った値の演算が同じ回路規模で実行できる演算器に注目し、画像識別処理を対象に、重み量子化時の識別精度低下を抑制する正規化手法を提案する。フレームワークに `caffe` を用いて提案手法を評価できる環境を構築し、量子化のみを行った場合との認識精度を比較、評価する。

2. 提案手法

提案手法の概要を図1に示す。提案手法では、畳み込み層の重み係数に対して行っており、想定する量子化ビット数より少し大きいビット数に量子化し、その差分をビットマスクする。これにより、ハードウェア実装時には単純量子化時のビット数と同じ回路規模であるとみなせる。しかし実際には、単純量子化時のビット幅より大きなビット数となっており、識別精度低下の抑制が期待される。丸め方法は、最近接丸めを採用している。

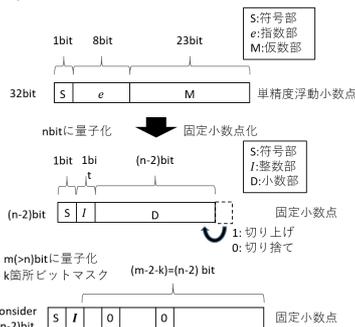


図1 提案手法の概要

3. 評価環境

`caffe` に付属している `LeNet_caffe(LeNet)` と `Cifar10_quick`、比較的大規模なネットワークである `VGG9` を用いて、提案手法の有効性を評価した。画像データセットは、`LeNet` の層構成では `Mnist` を用い、`Cifar10_quick` と `VGG9` の層構成では `Cifar10` を用いた。

4. 評価結果

4.1 LeNet_caffe

`LeNet_caffe` に提案手法と量子化のみを行った結果を表2、表3に示す。4bit のときは、量子化のみでは精度が得られないが、提案手法によりビットマスクをして精度をみなし 4bit にすると高い精度が得られた。5bit のときは、数組のビットマスクの組み合わせが量子化時より精度が向上した。また、4bit、5bit の両方でビットマスクを行った場合、より上位のビットがマスクされると精度が悪くなるのが分かった。

表2. 4bit 精度(LeNet)

量子化ビット数とビットマスク位置	認識精度	比較
4_0_0_0_0_0_0_0	xx.xx	0.1135
5_1_1_0_0_0_0_0	xx.x0x	0.9580 0.8445
5_1_2_0_0_0_0_0	xx.0xx	0.9656 0.8521
6_2_1_2_0_0_0_0	xx.x00x	0.9500 0.8365
6_2_1_3_0_0_0_0	xx.0x0x	0.9365 0.8230
6_2_2_3_0_0_0_0	xx.000x	0.9341 0.8206
7_3_1_2_3_0_0_0	xx.x000x	0.9312 0.8177
7_3_2_3_4_0_0_0	xx.000xx	0.7907 0.6772
9_5_1_2_3_4_5_0	xx.x0000x	0.9238 0.8103
9_5_2_3_4_5_6_0	xx.00000x	0.1135 0
10_6_1_2_3_4_5_6	xx.x00000x	0.9352 0.8217
10_6_2_3_4_5_6_7	xx.000000x	0.1135 0

表3. 5bit 精度(LeNet)

量子化ビット数とビットマスク位置	認識精度	比較
5_0_0_0_0_0_0_0	xx.xxxx	0.9528
6_1_1_0_0_0_0_0	xx.xx0x	0.9563 0.0035
6_1_2_0_0_0_0_0	xx.x0xx	0.9559 0.0031
6_1_3_0_0_0_0_0	xx.0xxx	0.9488 -0.0040
7_2_1_2_0_0_0_0	xx.xx00x	0.9585 0.0057
7_2_1_3_0_0_0_0	xx.x00x	0.9465 -0.0063
7_2_1_4_0_0_0_0	xx.000x	0.9365 -0.0163
7_2_2_3_0_0_0_0	xx.x000x	0.9425 -0.0103
8_3_1_2_3_0_0_0	xx.xx000x	0.9587 0.0059
8_3_1_3_4_0_0_0	xx.x000x	0.9390 -0.0138
9_4_1_2_3_4_0_0	xx.xx0000x	0.9533 0.0005
11_6_1_2_3_4_5_6	xx.xx000000x	0.9546 0.0018

4.2 Cifar10_quick

`Cifar10_quick` に提案手法と量子化のみを行った結果を表4、表5に示す。12bit のときは、表2の結果と同様に、量子化のみでは精度が得られないが、ビットマスクをして精度をみなし 12bit とすると高い精度が得られた。ビットマスクを行う箇所による精度の影響も `LeNet_caffe` と同様に、より上位のビットがマスクされると、精度が悪くなった。

表4. 12bit 精度(Cifar10)

量子化ビット数とビットマスク位置	認識精度	比較
12_0_0_0_0_0_0_0_0	xxxxxxxxxxxx	0.1
13_1_1_0_0_0_0_0_0	xxxxxxxxxxxx0x	0.7528 0.6528
13_1_2_0_0_0_0_0_0	xxxxxxxxxxxx0xx	0.7431 0.6431
13_1_3_0_0_0_0_0_0	xxxxxxxxxxxx0xxx	0.6692 0.5692
13_1_8_0_0_0_0_0_0	xxxxxxxxxxxx0xxxxxx	0.2334 0.1334
13_1_9_0_0_0_0_0_0	xxxxxxxxxxxx0xxxxxx	0.2059 0.1059
13_1_10_0_0_0_0_0_0	xxxxxxxxxxxx0xxxxxx	0.1000 0
14_2_1_2_0_0_0_0_0_0	xxxxxxxxxxxx00xx	0.7435 0.6435
14_2_2_3_0_0_0_0_0_0	xxxxxxxxxxxx00xxx	0.6990 0.5990
14_2_10_11_0_0_0_0_0_0	xxxxxxxxxxxx000xxxxxx	0.2208 0.1208
15_3_1_2_3_0_0_0_0_0_0	xxxxxxxxxxxx000xxx	0.7342 0.6342

表5. 13bit 精度(Cifar10)

量子化ビット数とビットマスク位置	認識精度	比較
13_0_0_0_0_0_0_0_0_0	xxxxxxxxxxxx	0.7607
14_1_1_0_0_0_0_0_0_0	xxxxxxxxxxxx0x	0.7606 -0.0001
14_1_2_0_0_0_0_0_0_0	xxxxxxxxxxxx0xx	0.7591 -0.0016
14_1_3_0_0_0_0_0_0_0	xxxxxxxxxxxx0xxx	0.7323 -0.0284
14_1_9_0_0_0_0_0_0_0	xxxxxxxxxxxx0xxxxxx	0.2230 -0.5377
14_1_10_0_0_0_0_0_0_0	xxxxxxxxxxxx0xxxxxx	0.2147 -0.5460
14_1_11_0_0_0_0_0_0_0	xxxxxxxxxxxx0xxxxxx	0.2045 -0.5562
15_2_1_2_0_0_0_0_0_0_0	xxxxxxxxxxxx00xx	0.7537 -0.0070
15_2_2_12_0_0_0_0_0_0_0	xxxxxxxxxxxx00xxx	0.2114 -0.5493
15_2_7_8_0_0_0_0_0_0_0	xxxxxxxxxxxx000xxxxxx	0.2223 -0.5384
15_2_11_12_0_0_0_0_0_0_0	xxxxxxxxxxxx000xxxxxx	0.2186 -0.5421

4.3 VGG9

`VGG9` に提案手法と量子化のみを行った結果を表6、表7に示す。6bit のときは表2、表4と同様にビットマスクをしてみなし 6bit とすると精度が大幅に向上した。一方で、7bit のときは量子化のみを行ったときと比較して精度が大幅に低下した。

表6. 6bit 精度(VGG9)

量子化ビット数とビットマスク位置	認識精度	比較
6_0_0_0_0_0_0_0_0_0	xxxxxx	0.1997
7_1_1_0_0_0_0_0_0_0	xxxx0xx	0.5607 0.3610
7_1_2_0_0_0_0_0_0_0	xxxx0xxx	0.3450 0.1453
7_1_3_0_0_0_0_0_0_0	xxxx0xxx	0.2904 0.0907
7_1_4_0_0_0_0_0_0_0	xxxx0xxx	0.2922 0.0925
8_2_1_3_0_0_0_0_0_0	xxxx000xx	0.2924 0.0927
8_2_1_4_0_0_0_0_0_0	xxxx000xx	0.2872 0.0875
8_2_1_5_0_0_0_0_0_0	xxxx000xx	0.2858 0.0861
8_2_2_3_0_0_0_0_0_0	xxxx000xx	0.2891 0.0894
8_2_2_4_0_0_0_0_0_0	xxxx000xx	0.2868 0.0871
8_2_2_5_0_0_0_0_0_0	xxxx000xx	0.2857 0.0860
8_2_3_4_0_0_0_0_0_0	xxxx000xx	0.2877 0.0880
8_2_3_5_0_0_0_0_0_0	xxxx000xx	0.2865 0.0868
8_2_4_5_0_0_0_0_0_0	xxxx000xx	0.2826 0.0829

表7. 7bit 精度(VGG9)

量子化ビット数とビットマスク位置	認識精度	比較
7_0_0_0_0_0_0_0_0_0	xxxxxxx	0.8462 0
8_1_1_0_0_0_0_0_0_0	xxxxxxx0x	0.2896 -0.5566
8_1_2_0_0_0_0_0_0_0	xxxxxxx0xx	0.2848 -0.5614
8_1_3_0_0_0_0_0_0_0	xxxxxxx0xxx	0.2861 -0.5601
8_1_4_0_0_0_0_0_0_0	xxxxxxx0xxx	0.2870 -0.5592
8_1_5_0_0_0_0_0_0_0	xxxxxxx0xxx	0.2789 -0.5673
8_2_1_2_0_0_0_0_0_0_0	xxxxxxx00xx	0.2833 -0.5629
8_2_1_3_0_0_0_0_0_0_0	xxxxxxx00xxx	0.2903 -0.5559
8_2_1_4_0_0_0_0_0_0_0	xxxxxxx00xxx	0.2903 -0.5559
8_2_1_5_0_0_0_0_0_0_0	xxxxxxx00xxx	0.2824 -0.5638
8_2_2_6_0_0_0_0_0_0_0	xxxxxxx00xxx	0.2766 -0.5696
8_2_2_3_0_0_0_0_0_0_0	xxxxxxx00xxx	0.2896 -0.5566
8_2_2_4_0_0_0_0_0_0_0	xxxxxxx00xxx	0.2865 -0.5597
8_2_2_5_0_0_0_0_0_0_0	xxxxxxx00xxx	0.2835 -0.5627
8_2_2_6_0_0_0_0_0_0_0	xxxxxxx00xxx	0.2793 -0.5669

5. 考察

量子化時では精度が得られない場合でも提案手法によりビットマスクをしてみなしビットとすることで高い精度が得られた。これは、量子化によって切り捨てられる値が、全体のビット幅が増えることによって表現できるようになったためと考えられる。また、ビットマスクがより上位のビットに適応されると精度が悪くなったことから、より上位のビットが精度向上に重要であると考えられる。

`VGG9` で提案手法を適用した場合、認識精度の向上が確認できたが、一部の条件で大幅な精度低下が起こった。その原因として、学習パラメータ設定の影響などが考えられる。

6. まとめ

量子化のみでは認識精度が低いビット数の場合でも、提案手法によって認識精度の向上が可能であることを示した。量子化のみを行ったときより少ないビット数で高い精度の学習を行えるため、ハードウェア実装時での回路規模と演算量を削減が可能となる。一方、`VGG9` の評価実験では、一部の条件において認識精度が大幅に低下することが確認されたため、提案手法の改良が必要であると考えられる。

参考文献

[1] S. Zhou, Z. Ni, X. Zhou, H. Wen, Y. Wu, and Y. Zou, "DoReFa-Net: Training low bitwidth convolutional neural networks with low bitwidth gradients", arXiv preprint arXiv: 1606.06160, Jun. 2016.
 [2] S. Gupta, A. Agrawal, K. Gopalakrishnan, "Deep Learning with Limited Numerical Precision", arXiv preprint arXiv:1502.02551v1, Feb. 2015.