

ディープラーニングによるく ずし字認識組み込みシステム の開発

GAO HAN(コウ カン)

輪講-5.13

論文説明

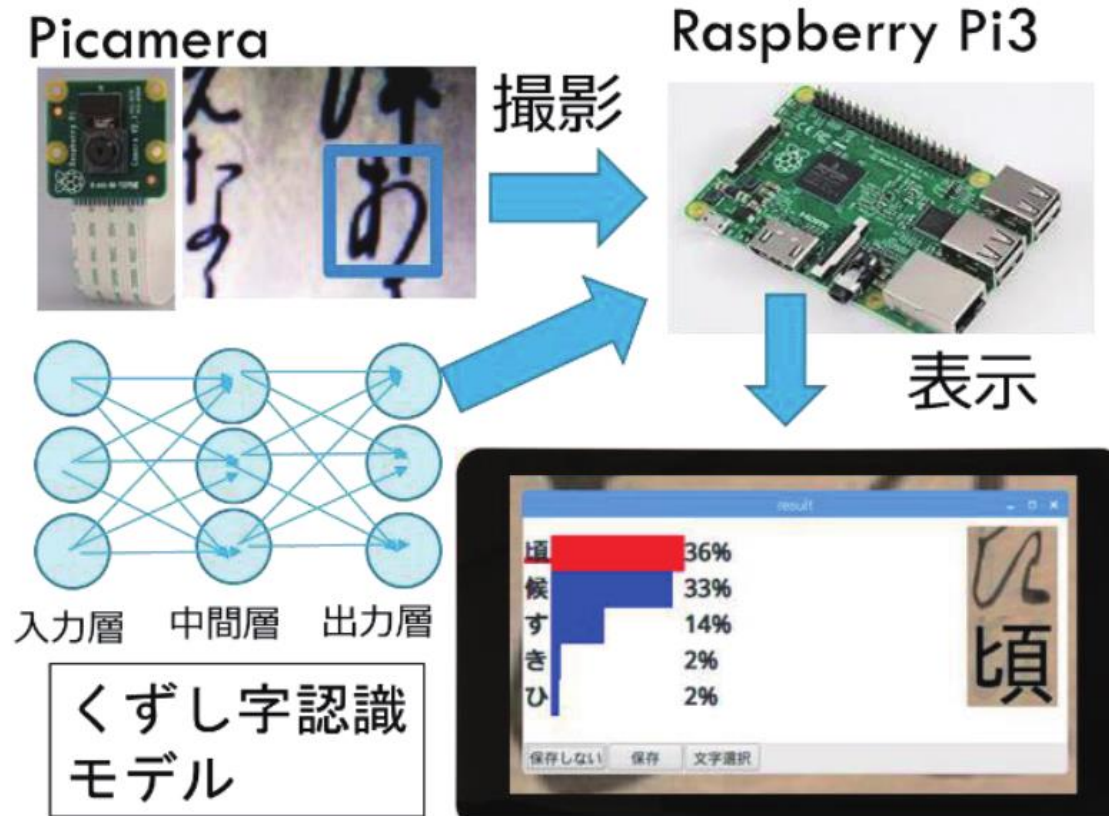
The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

ディープラーニングによるくずし字認識組み込みシステムの開発 Development of Embedded System for Recognizing *Kuzushiji* by Deep Learning

竹内正広^{*1} 早坂太一^{*1} 大野亙^{*1} 加藤弓枝^{*2} 山本和明^{*3} 石間衛^{*4} 石川徹也^{*4}
Masahiro Takeuchi, Taichi Hayasaka, Wataru Ohone, Yumie Kato, Kazuaki Yamamoto, Mamoru Ishima, and Tetsuya Ishikawa

^{*1} 豊田工業高等専門学校 ^{*2} 鶴見大学 ^{*3} 国文学研究資料館 ^{*4} TRC-ADEAC 株式会社
National Institute of Technology, Toyota College, Tsurumi University, National Institute of Japanese Literature, and TRC-ADEAC, Inc.

論文説明



- WWWアプリケーションとして実装されている学習モデルを応用し Raspberry Piを用いて開発されたくずし字認識組み込みシステムについて報告する。
- カメラジュールPicameraから古典籍画像を撮影し、用いる機械学習用ライブラリTensorFlow Liteを用いて、Python 3およびOpenCV 3を用いて認識する選択、画像処理などを行う。

論文説明

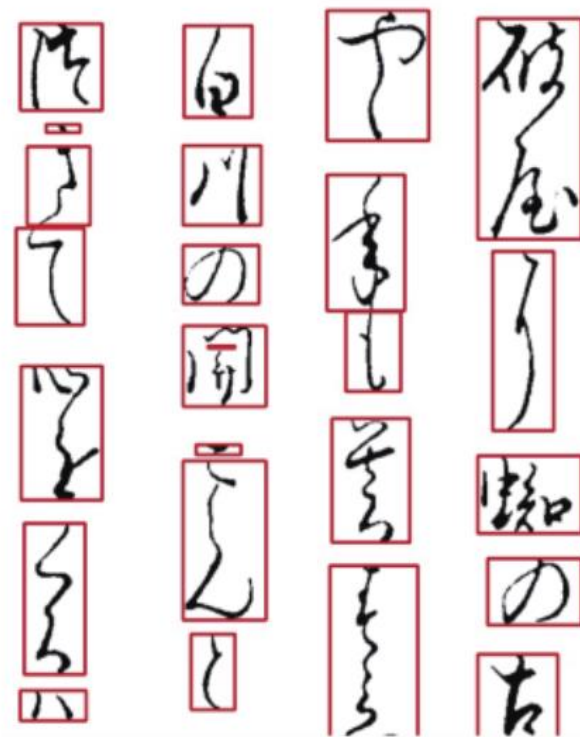
- タッチスクリーン指を触れてから離すまでのx-y座標を取得し、2値化を行った後に、学習モデルに入力する。
- 入力画像に対して、縦方向に対して膨張処理を行い、横方向に膨張処理を行い、1行の中から1文字の領域を検出する。
- 学習済みモデルの重み付けなどを浮動小数点形式から固定小数点形式に量子化し、モデルサイズを小さくすることで、能力がやや劣る機器で演算の高速化、および読み込みにかかる時間を短縮することができる。



(a)



(b)



(c)

画像の二値化

2値化処理とは、ある濃度値より小さいものは濃度値をいくらに変更し、ある濃度値より大きいものは濃度値をいくらに変更するといった濃度変換処理をする。

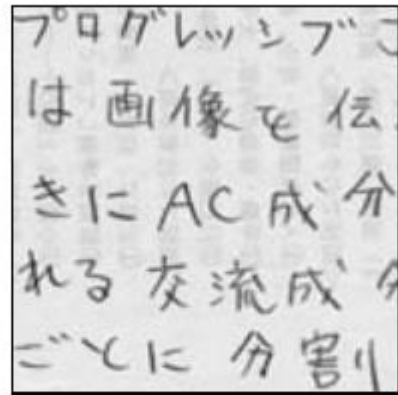


図.18 : 手書き文字画像
(128×128)

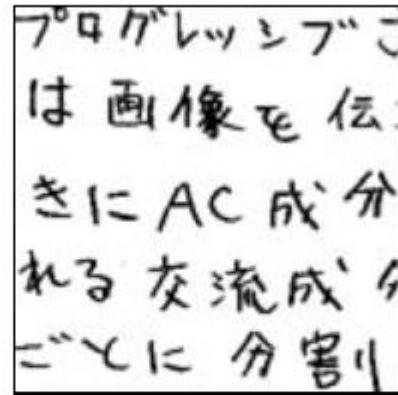


図.19 : 2値化画像
(128×128)

画像の二値化 Pタイル法

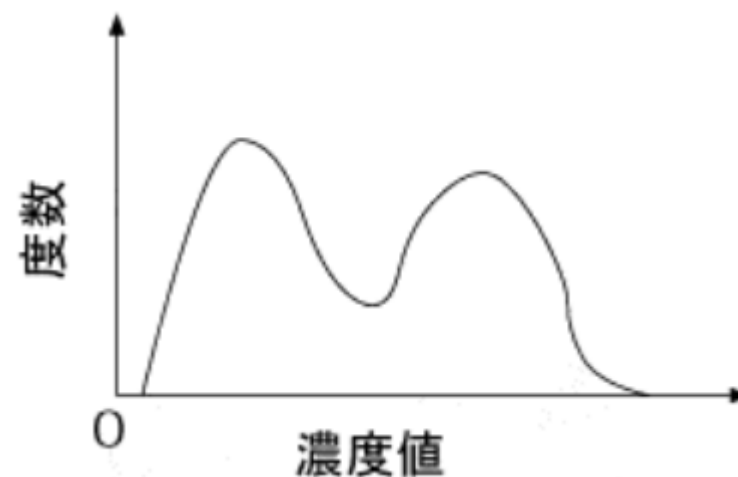
- 画像全体の面積を S 、対象図形的面積を S_0 とし画像全体に対する対象画像の面積の比率 p を求める。
- 画素数を N_0 とする。全体の画素数を N とするとときの濃度値をしきい値とする。

$$p = \frac{S_0}{S}$$

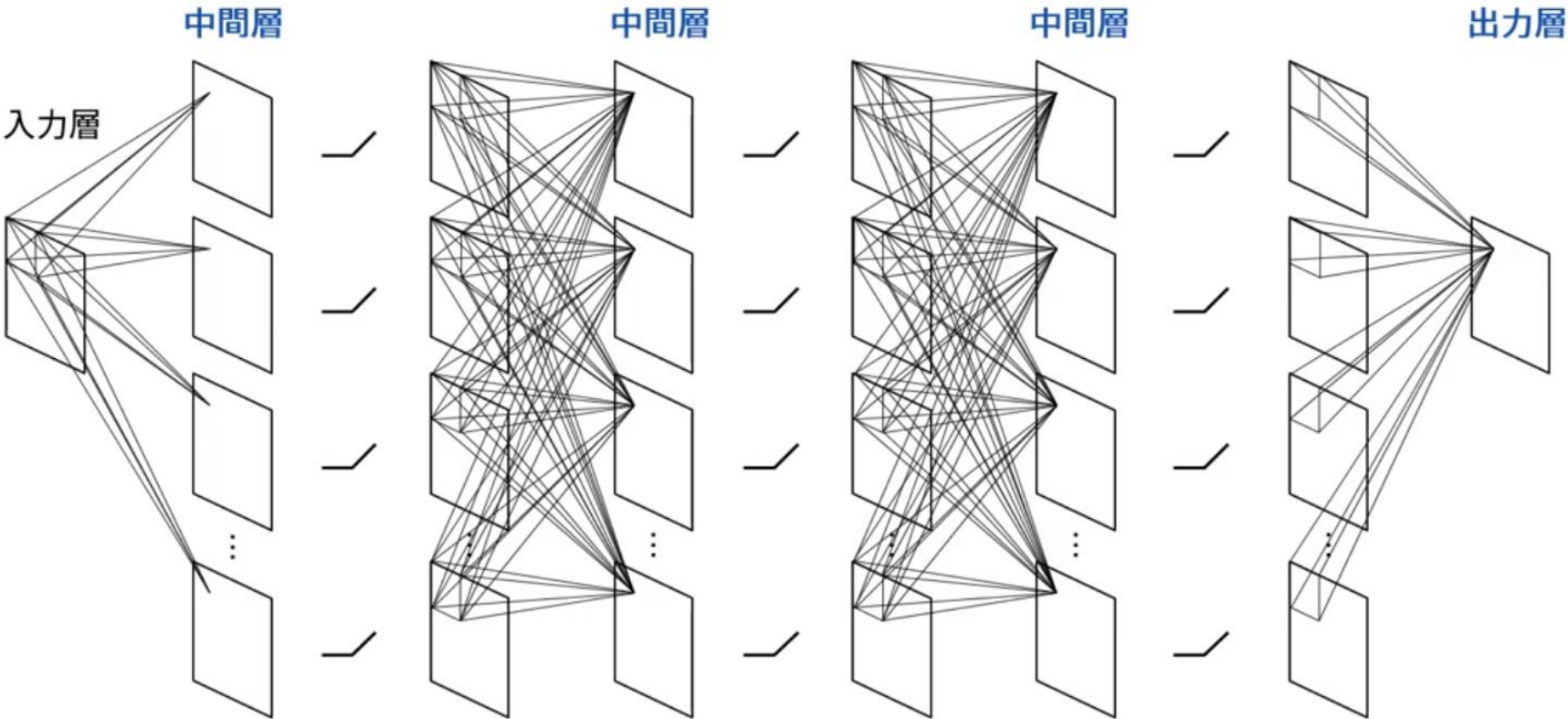
$$\frac{N_0}{N} = p$$

画像の二値化 モード法

- 画像の中の対象図形のおおよその面積がわからない場合でも、濃度ヒストグラムを作ってみると、多くの場合、双峰性のヒストグラムが得られる。このようなヒストグラムの場合は、その谷に当たる濃度値をしきい値とすれば良いのだ。



モデルの量子化



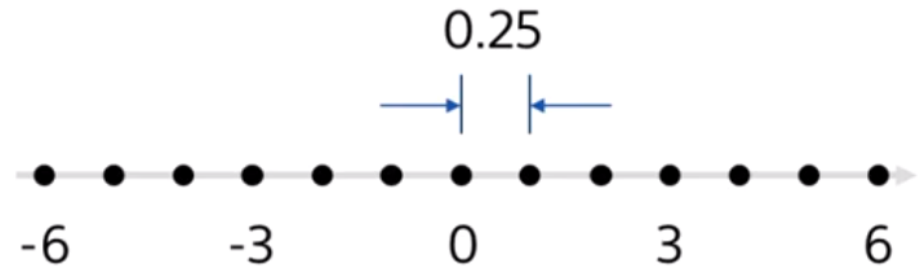
モデルの量子化

連続量



浮動小数点数

離散量



整数

モデルの量子化

32bit 浮動小数点数



8bit 整数

回路規模

大

小

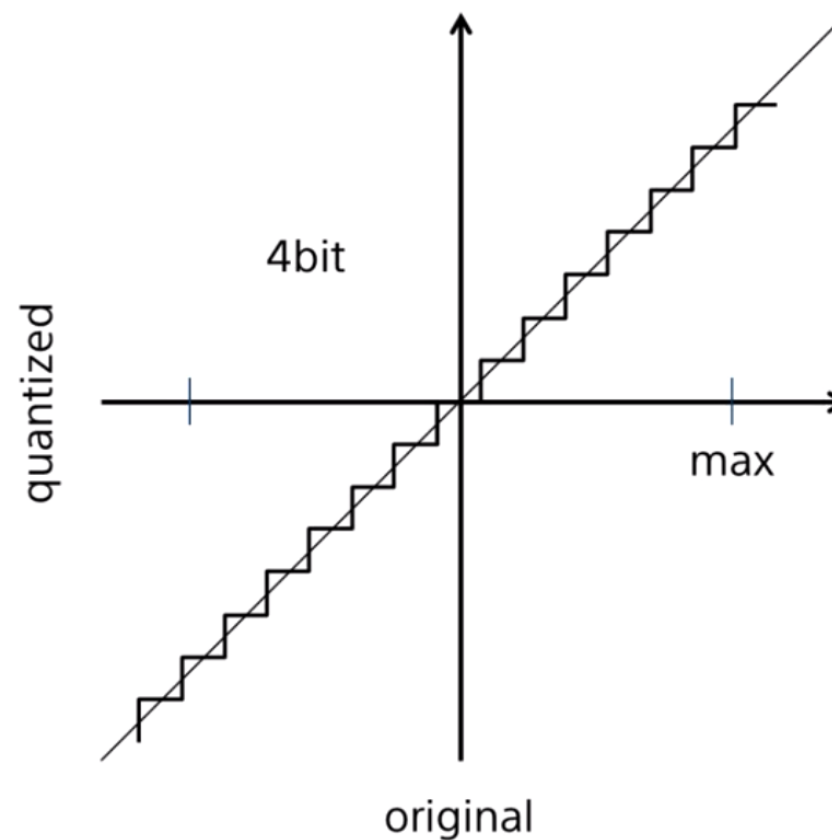
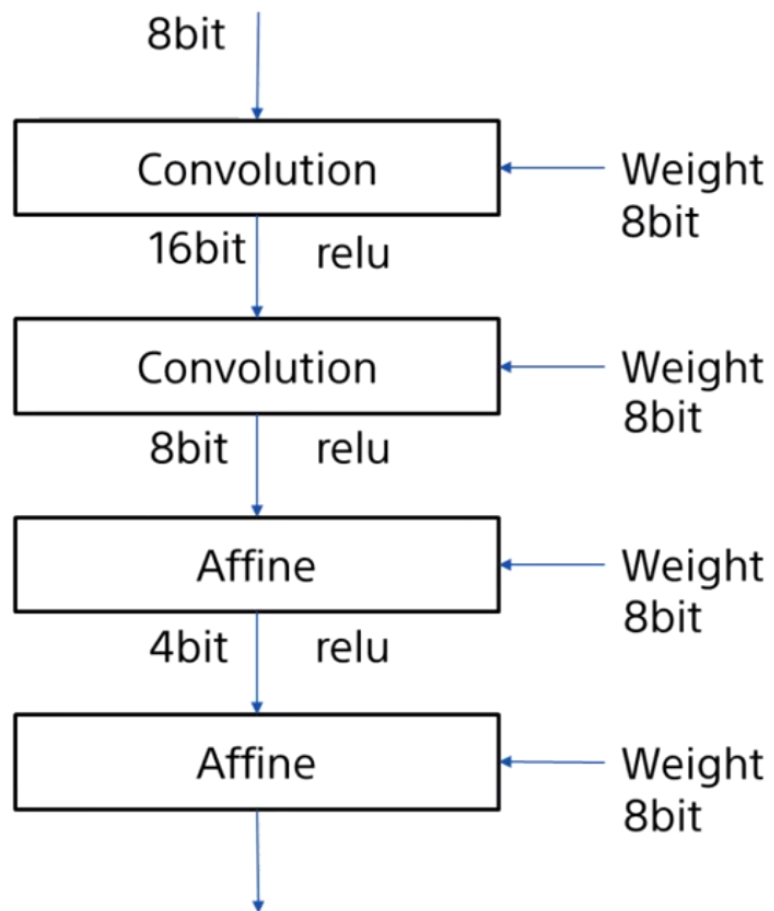
必要メモリ

大

小

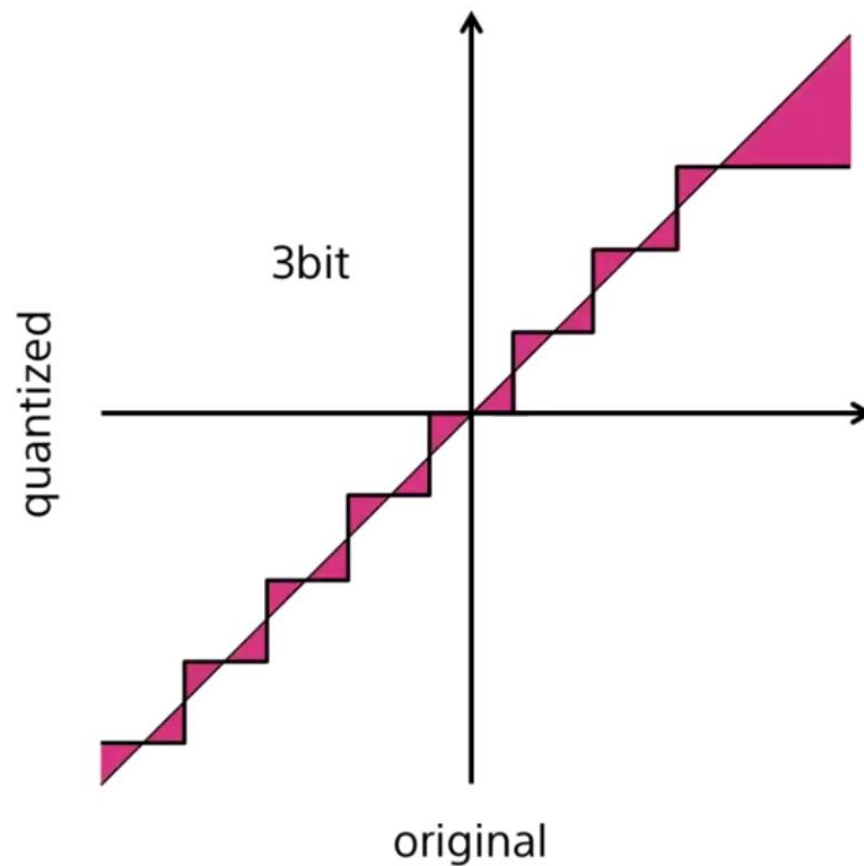
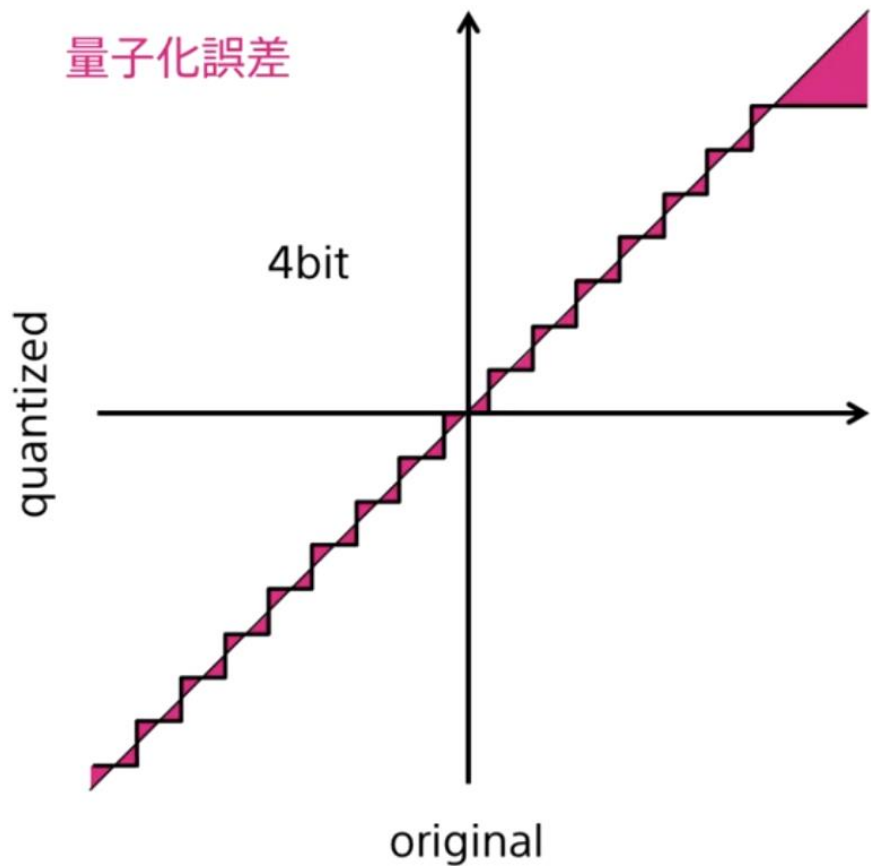
モデルの量子化

学習済みモデルの事後的量子化



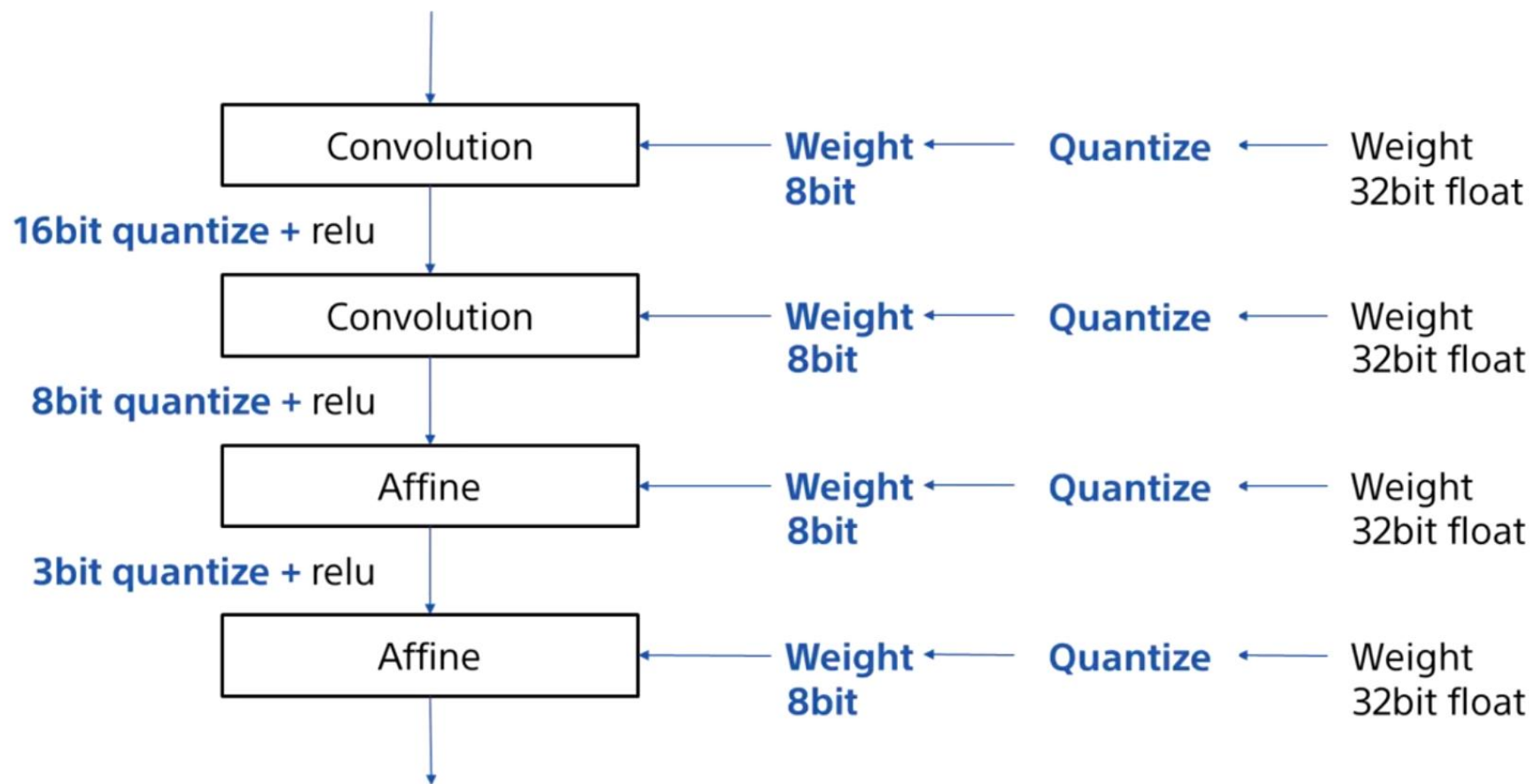
モデルの量子化

学習済みモデルの事後的量子化



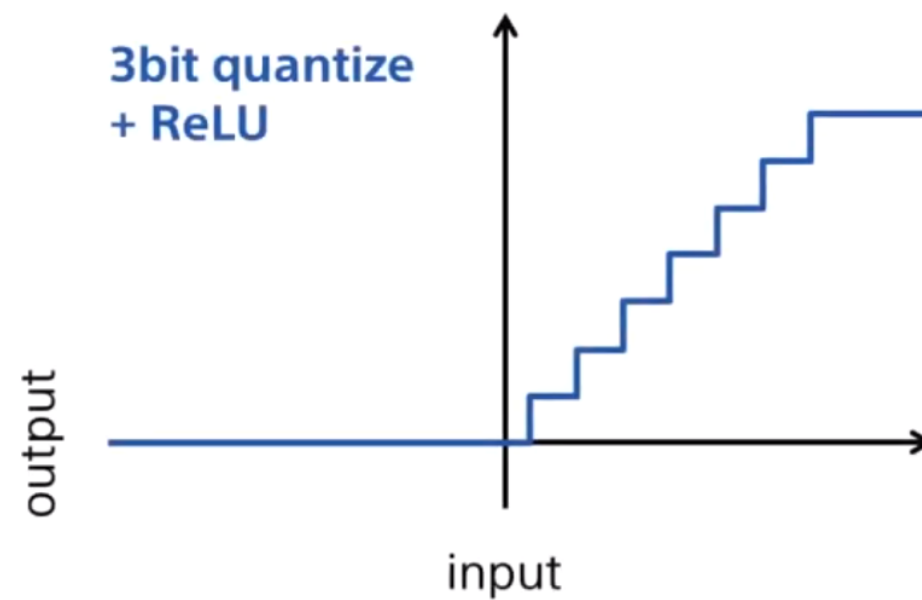
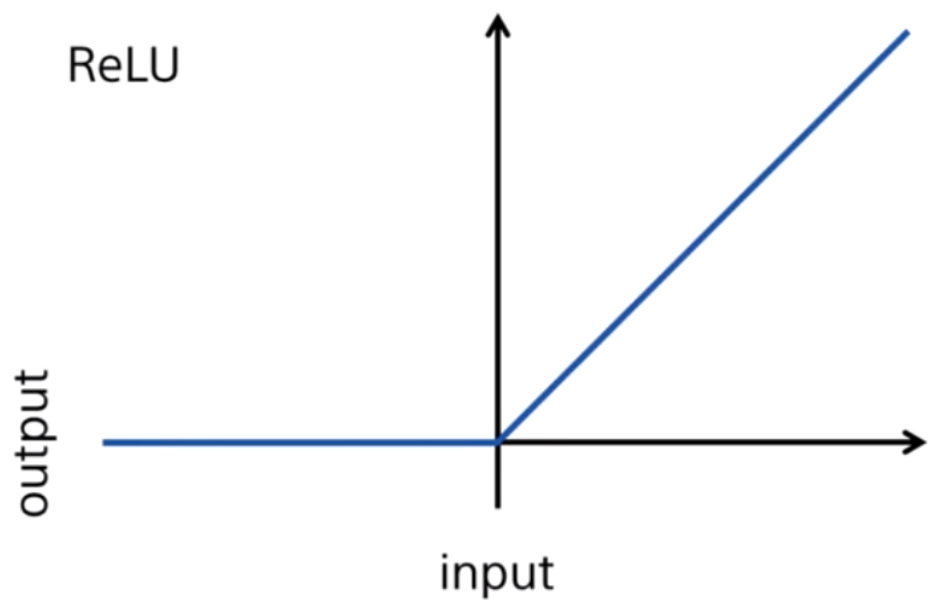
モデルの量子化

量子化を考慮した学習



モデルの量子化

量子化を考慮した学習

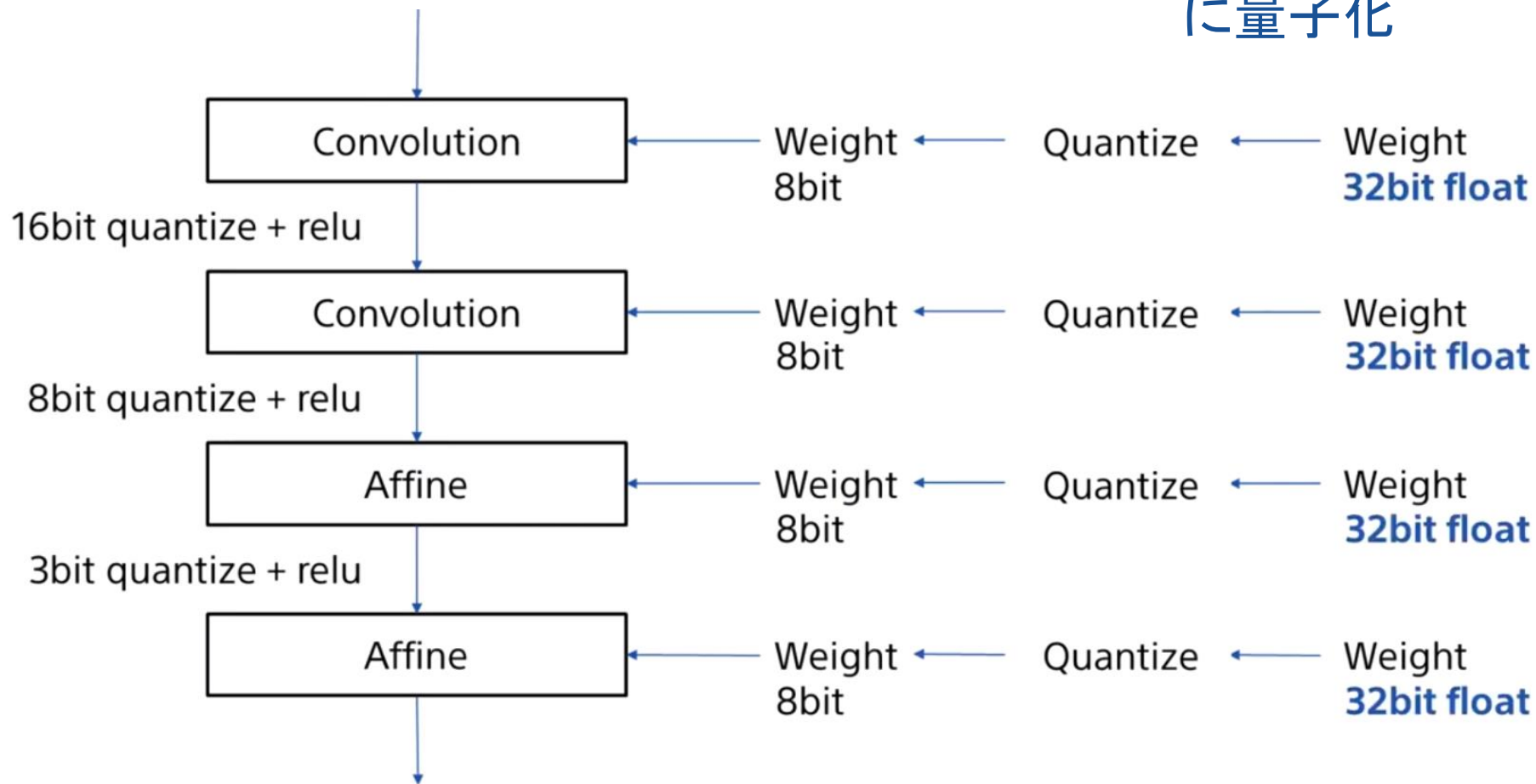


モデルの量子化

量子化を考慮した学習

学習時

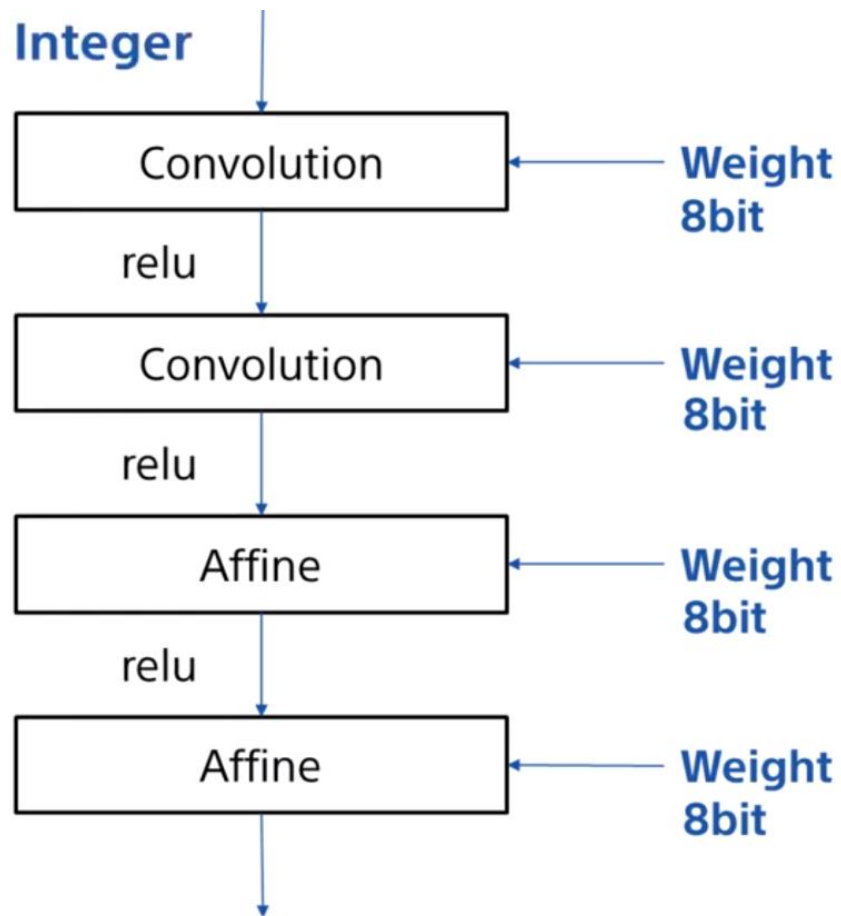
浮動小数点数のパラメータ・データパスを動的に量子化



モデルの量子化

量子化を考慮した学習

推論時



量子化済みのパラ
メータのみ必要

モデルの量子化

その他のコンパクト化テクニック

- Binary Neural Networks (2値化ネットワーク)
- Pruning (枝刈り)
 - 0付近のパラメータの演算を省くことでパラメータ量、演算量を削減
- アーキテクチャの工夫
 - CNNの工夫
 - 重み係数の共有
- その他、より高度な量子化
 - 事後的な量子化においても精度劣化を抑えながらよりbit数を削減
 - 量子化単位やbit数の自動調整