

【公報種別】特許法第 17 条の 2 の規定による補正の掲載
 【部門区分】第 6 部門第 3 区分
 【発行日】令和 3 年 11 月 11 日 (2021.11.11)

【公開番号】特開 2020-57286 (P2020-57286A)
 【公開日】令和 2 年 4 月 9 日 (2020.4.9)
 【年通号数】公開・登録公報 2020-014
 【出願番号】特願 2018-188612 (P2018-188612)
 【国際特許分類】

G 0 6 F 17/10 (2006.01)

【F I】

G 0 6 F	17/10	S
G 0 6 F	17/10	D
G 0 6 F	17/10	A

【手続補正書】

【提出日】令和 3 年 10 月 4 日 (2021.10.4)

【手続補正 1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項 1】

複数の階層を有するネットワークにおけるそれぞれの階層について、該階層の複数のデータと複数のフィルタ係数との積和演算を実行する複数の積和演算手段と、

前記複数の積和演算手段による複数の積和演算結果を複数のシフトパラメータに基づいてシフトする複数のシフト演算手段と、

前記複数のシフト演算手段によりシフトされた複数の積和演算結果の総和を計算する加算手段と、

前記階層のビット幅に応じて前記シフトパラメータの値を切り替える制御手段とを備えることを特徴とする情報処理装置。

【請求項 2】

前記制御手段は、前記階層のビット幅に応じてフィルタ係数の転送回数を切り替えることを特徴とする請求項 1 に記載の情報処理装置。

【請求項 3】

前記シフト演算手段は複数のシフトパラメータに基づいて前記複数の積和演算結果をシフトすることを特徴とする請求項 1 または 2 に記載の情報処理装置。

【請求項 4】

複数の階層を有するネットワークにおけるそれぞれの階層について、該階層の複数のデータを複数のシフトパラメータに基づいてシフトする複数のシフト演算手段と、

前記複数のシフト演算手段によりシフトされた複数のデータと複数のフィルタ係数との積和演算を実行する複数の積和演算手段と、

前記複数の積和演算手段により計算された複数の積和演算結果の総和を計算する加算手段と、

前記階層のビット幅に応じて前記シフトパラメータの値を切り替える制御手段とを備えることを特徴とする情報処理装置。

【請求項 5】

前記複数の積和演算手段と前記複数のシフト演算手段とは並列に動作することを特徴とする請求項 1 乃至 4 の何れか 1 項に記載の情報処理装置。

【請求項 6】

前記加算手段は、ある階層に対する前記総和を該階層の次の階層のデータとしてメモリに格納することを特徴とする請求項 1 乃至 5 の何れか 1 項に記載の情報処理装置。

【請求項 7】

前記加算手段は、ある階層に対する前記総和に対して活性化処理、プーリング処理を行った結果を、該階層の次の階層のデータとしてメモリに格納することを特徴とする請求項 1 乃至 5 の何れか 1 項に記載の情報処理装置。

【請求項 8】

更に、

前記メモリに格納されたデータに基づいて、動画像における各フレームに対する画像処理および/または画像認識を行う手段を備えることを特徴とする請求項 6 または 7 に記載の情報処理装置。

【請求項 9】

前記ネットワークは、階層ごとにデータのビット幅が異なることを特徴とする請求項 1 乃至 8 の何れか 1 項に記載の情報処理装置。

【請求項 10】

情報処理装置が行う情報処理方法であって、

前記情報処理装置の複数の積和演算手段が、複数の階層を有するネットワークにおけるそれぞれの階層について、該階層の複数のデータと複数のフィルタ係数との積和演算を実行する工程と、

前記情報処理装置の複数のシフト演算手段が、前記複数の積和演算手段による複数の積和演算結果を複数のシフトパラメータに基づいてシフトする工程と、

前記情報処理装置の加算手段が、前記複数のシフト演算手段によりシフトされた複数の積和演算結果の総和を計算する工程と、

前記情報処理装置の制御手段が、前記階層のビット幅に応じて前記シフトパラメータの値を切り替える工程と

を備えることを特徴とする情報処理方法。

【請求項 11】

情報処理装置が行う情報処理方法であって、

前記情報処理装置の複数のシフト演算手段が、複数の階層を有するネットワークにおけるそれぞれの階層について、該階層の複数のデータを複数のシフトパラメータに基づいてシフトする工程と、

前記情報処理装置の複数の積和演算手段が、前記複数のシフト演算手段によりシフトされた複数のデータと複数のフィルタ係数との積和演算を実行する工程と、

前記情報処理装置の加算手段が、前記複数の積和演算手段により計算された複数の積和演算結果の総和を計算する工程と、

前記情報処理装置の制御手段が、前記階層のビット幅に応じて前記シフトパラメータの値を切り替える工程と

を備えることを特徴とする情報処理方法。

【請求項 12】

コンピュータを、請求項 1 乃至 9 の何れか 1 項に記載の情報処理装置の各手段として機能させるためのコンピュータプログラム。

【手続補正 2】

【補正対象書類名】明細書

【補正対象項目名】0010

【補正方法】変更

【補正の内容】

【0010】

非特許文献 2 に記載の手法では、複数種類のビット幅の特徴データを処理可能な SIMD (シングルインストラクションマルチプルデータ) 構成の RNN (Recurrent

Neural Network) 専用ハードウェアが提案されている。同じハードウェアで 8 ビット、16 ビット、32 ビットのデータを処理可能であるが、並行して出力されたデータの総和を計算する際に、一旦メモリに保持してからもう一回 SIMD のコマンドを実行する必要がある、処理時間が長くなる。本発明では、複数の階層を有するネットワークの中に複数種類のビット幅の特徴データがあっても効率的な処理を実現するための技術を提供する。

【手続補正 3】

【補正対象書類名】明細書

【補正対象項目名】0044

【補正方法】変更

【補正の内容】

【0044】

ステップ S113 では、制御部 405 は、出力特徴画像のループの終了判定をする。全ての出力特徴画像の処理が終わった場合にはステップ S114 に進み、未処理の出力特徴画像が残っている場合には、ステップ S105 に戻り、未処理の出力特徴画像について処理を開始する。

【手続補正 4】

【補正対象書類名】明細書

【補正対象項目名】0053

【補正方法】変更

【補正の内容】

【0053】

フィルタサイズが 1×1 よりも大きい場合、積和演算部 402 はフィルタ係数と入力特徴データとの畳み込み演算結果を計算するが、フィルタサイズが 1×1 の場合、積和演算部 402 は $I(m)$ と $C(m, n)$ の積を計算する。

【手続補正 5】

【補正対象書類名】明細書

【補正対象項目名】0054

【補正方法】変更

【補正の内容】

【0054】

処理対象の特徴データはビット幅が ビット とビット幅が ビット の 2 種類であるとする。図 4 に示した積和演算部 402 は積和演算結果を計算する ビット のデータの積和演算ユニットが P 個あり、シフト演算部 403 はシフト演算結果を計算する ビット のデータのシフト演算ユニットが P 個ある。 、 、 P は以下の条件を満たしている。

【手続補正 6】

【補正対象書類名】明細書

【補正対象項目名】0089

【補正方法】変更

【補正の内容】

【0089】

< ビット幅が異なる場合の詳細説明 >

本実施形態では、シフト演算部 701 はシフト演算結果を計算する ビット のデータのシフト演算ユニットが P 個あり、積和演算部 702 は積和演算結果を計算する ビット のデータの積和演算ユニットが P 個ある。積和演算部 702 の出力は (式 15) で表され、(式 5) に示したシフト演算部 403 の出力と等価である。

【手続補正 7】

【補正対象書類名】明細書

【補正対象項目名】0097

【補正方法】変更

【補正の内容】

【0097】

$O(n)$ は n 枚目の出力特徴画像の積和演算結果であり、 $I(\quad), p(m)$ は ビットのデータの積和演算ユニットの入力データであり、 $C_p(m, n)$ はフィルタ係数であり、 $S(p)$ はシフトパラメータである。変数 m は ビットの入力特徴画像グループ (1 グループ = P 枚) の番号 (積和演算部 402 の処理番号) である。積和演算ユニットとシフト演算ユニットはそれぞれ P / P' セットに分かれており、変数 q は積和演算ユニットのセットの番号である。変数 p はセット内の積和演算ユニット、シフト演算ユニットの番号であり、変数 n は出力特徴画像の番号である。シフト演算は 2 のべき乗の処理で表現される。