

【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25日(2021.11.25)

【公表番号】特表2020-537785(P2020-537785A)

【公表日】令和2年12月24日(2020.12.24)

【年通号数】公開・登録公報2020-052

【出願番号】特願2020-521412(P2020-521412)

【国際特許分類】

G 06 N 3/063 (2006.01)

G 06 F 17/16 (2006.01)

【F I】

G 06 N 3/063

G 06 F 17/16 S

【手続補正書】

【提出日】令和3年10月14日(2021.10.14)

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】0036

【補正方法】変更

【補正の内容】

【0036】

図4は、一実装形態による、例示的なニューラルネットワーク処理システム400を示す。システムは、ニューラルネットワークアクセラレータ238に通信可能に接続されたホストコンピュータシステム402を含む。ホストコンピュータシステム402は、図2に示されているように、処理システム210と、システムメモリ216と、ストレージ218とを含むことができる。ホストコンピュータシステム402は、詳細には、機械学習(ML)フレームワーク410とニューラルネットワークアクセラレータ(KA)インターフェース412とによってプログラムされる。図1のアプリケーション112に対応するMLフレームワークプログラムは、特定のニューラルネットワークアプリケーション、たとえば、画像または音声処理を指定し、図1の場合のようにアクセラレーションライブラリとして実装され得るKAインターフェースは、MLフレームワークからのニューラルネットワーク処理についての要求に応答して、ニューラルネットワークアクセラレータ上でニューラルネットワーク動作を始動する。ニューラルネットワークアクセラレータ238はRAM226に連結され、RAM226を通して、ホストとニューラルネットワークアクセラレータとが通信する。ニューラルネットワークアクセラレータは構成レジスタ408のセットを有する。構成レジスタは、RAM226中のメモリバッファのアドレスと、一般的な行列乗算(GEMM)のための行列次元、および畳み込みのためのストライド/ウィンドウなど、ニューラルネットワーク動作のための構成パラメータとを記憶するために、KAインターフェース412にとってアクセス可能である。