

【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 0 6 N 3/063 (2006.01)

G 0 6 F 17/16 (2006.01)

【 F I 】

G 0 6 N 3/063

G 0 6 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開始する。ニューラルネットワークアクセラレータ 238 は RAM 226 に連結され、RAM 226 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、RAM 226 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。