



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2024년02월27일
(11) 등록번호 10-2641283
(24) 등록일자 2024년02월22일

(51) 국제특허분류(Int. Cl.)
G06N 3/063 (2023.01) G06N 3/08 (2023.01)
G06N 5/04 (2023.01)
(52) CPC특허분류
G06N 3/063 (2013.01)
G06N 3/08 (2023.01)
(21) 출원번호 10-2022-7021145(분할)
(22) 출원일자(국제) 2016년04월29일
심사청구일자 2022년06월21일
(85) 번역문제출일자 2022년06월21일
(65) 공개번호 10-2022-0092642
(43) 공개일자 2022년07월01일
(62) 원출원 특허 10-2020-7011552
원출원일자(국제) 2016년04월29일
심사청구일자 2021년03월11일
(86) 국제출원번호 PCT/US2016/029965
(87) 국제공개번호 WO 2016/186810
국제공개일자 2016년11월24일
(30) 우선권주장
62/164,981 2015년05월21일 미국(US)
14/844,670 2015년09월03일 미국(US)

(73) 특허권자
구글 엘엘씨
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이
1600 (우:94043)
(72) 발명자
로쓰, 조나단
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이
1600 (우:94043)
(74) 대리인
특허법인 남앤남

(56) 선행기술조사문헌
EP00422348 A2*
(뒷면에 계속)

전체 청구항 수 : 총 23 항

심사관 : 서광훈

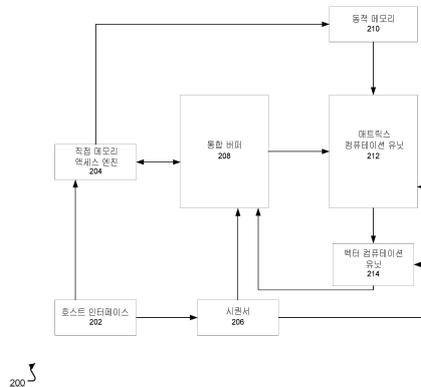
(54) 발명의 명칭 신경망 프로세서에서 사용하기 위한 가중치들의 프리페칭

(57) 요약

신경망에 대한 신경망 컴퓨테이션들을 수행하기 위한 회로가 개시되며, 회로는 복수의 셀들을 포함하는 시스틀릭 어레이; 가중치 페치 유닛 -가중치 페치 유닛은, 복수의 신경망 계층들의 각각에 대해, 복수의 가중치 입력들을 신경망 계층에 대한 시스틀릭 어레이의 제 1 차원을 따라 셀들에 전송하도록 구성됨-; 및 복수의 가중치 시퀀서

(뒷면에 계속)

대표도 - 도2



유닛들을 포함하고, 각각의 가중치 시퀀서 유닛은 시스틀릭 어레이의 제 1 차원을 따라 별개의 셀에 커플링되고, 복수의 가중치 시퀀서 유닛들은, 복수의 신경망 계층들 각각에 대해, 복수의 클록 사이클들을 통해 복수의 가중치 입력들을 신경망 계층에 대한 시스틀릭 어레이의 제 2 차원을 따라 셀들로 시프트하도록 구성되고, 각각의 셀은 곱셈 회로를 사용하여 활성화 입력 및 각각의 가중치 입력의 곱을 컴퓨팅하도록 구성된다.

(52) CPC특허분류
G06N 5/04 (2023.01)

(56) 선행기술조사문헌
KR100189195 B1
KR1020030082255 A
JP2004157756 A
JP07141454 A
US20040117710 A1
*는 심사관에 의하여 인용된 문헌

명세서

청구범위

청구항 1

복수의 계층들을 포함하는 신경망에 대한 신경망 컴퓨테이션들을 수행하기 위한 회로로서,

셀들의 2-차원 어레이를 위한 회로를 포함하는 하드웨어 매트릭스 컴퓨테이션 유닛 - 각각의 셀은 상기 셀의 외부에 있는 소스(source)로부터 수신되는 가중치 입력들을 저장하도록 구성됨 -;

복수의 신경망 계층들 각각에 대해, 신경망 계층에 대한 복수의 가중치 입력들을 상기 셀들의 2-차원 어레이에 전송하도록 구성되는 제1 하드웨어 회로;

상기 복수의 신경망 계층들 각각에 대해, 상기 셀들의 2-차원 어레이의 제1 차원을 따르는 셀들에 의해 수행되는 동작들을 제어하기 위해 상기 제1 차원을 따라 제어 신호들을 제공하도록 구성되는 제2 하드웨어 회로를 포함하고,

상기 동작들은 상기 셀들의 2-차원 어레이의 제2 차원을 따라 시프트되는 활성화 입력들 및 가중치 입력들을 사용하여 수행되며,

상기 제어 신호들은 각각의 셀이 상기 가중치 입력들 또는 상기 활성화 입력들을 인접한 셀들로 시프트 해야 하는지 여부를 결정하는,

회로.

청구항 2

제1항에 있어서,

상기 셀들의 2-차원 어레이의 각각의 셀은 가중치 레지스터를 포함하고,

상기 가중치 레지스터는, 상기 셀들의 2-차원 어레이의 상기 제2 차원을 따라 시프트되는 상기 가중치 입력들 중 하나 이상의 가중치 입력을 저장하는,

회로.

청구항 3

제2항에 있어서,

상기 셀들의 2-차원 어레이의 각각의 셀은, 상기 셀의 곱셈 회로를 사용하여, 상기 셀들의 2-차원 어레이의 상기 제2 차원을 따라 시프트되는 개개의 가중치 입력 및 활성화 입력의 곱을 계산하도록 구성되는,

회로.

청구항 4

제3항에 있어서,

상기 셀들의 2-차원 어레이의 각각의 셀은:

상기 셀들의 2-차원 어레이의 상기 제1 차원을 따라 제공되는 제어 신호들에 기초하여 제어 값을 수신하고; 그리고

상기 제어 값에 기초하여, 상기 셀들의 2-차원 어레이의 상기 제2 차원을 따라 시프트되는 개개의 가중치 입력을 상기 셀의 상기 가중치 레지스터에 저장할지를 결정하는,

회로.

청구항 5

제1항에 있어서,

상기 복수의 신경망 계층들 각각에 대해, 상기 회로는:

클럭 신호들을 생성하고;

상기 클럭 신호들의 타이밍에 기초하여, 상기 셀들의 2-차원 어레이의 상기 셀들 각각에 상기 제어 신호들을 전송하며;

상기 제어 신호들에 기초하여, 상이한 시간들에 그리고 상이한 셀들에서, 상기 동작들의 일부를 실행하는, 회로.

청구항 6

제5항에 있어서,

상기 제1 하드웨어 회로는:

상기 제어 신호들 및 상기 클럭 신호들의 타이밍에 기초하여, 상기 셀들의 2-차원 어레이의 상기 제2 차원을 따라 상이한 셀들에 상기 가중치 입력들 및 활성화 입력들을 시프트하도록 구성되는,

회로.

청구항 7

제1항에 있어서,

상기 회로는 프로세서를 포함하고,

상기 프로세서는:

상기 셀들의 2-차원 어레이의 상기 제1 차원을 따라 제공되는 상기 제어 신호들에 기초하여, 상기 매트릭스 컴퓨테이션 유닛의 대응하는 셀에 개개의 가중치 입력을 로딩시키도록 구성되고,

상기 개개의 가중치 입력은 상기 셀들의 2-차원 어레이의 상기 제2 차원을 따라 시프트되는 상기 가중치 입력들 중 하나의 가중치 입력인,

회로.

청구항 8

제7항에 있어서,

상기 프로세서는 추가로:

상기 제어 신호들에 기초하여, 복수의 클럭 사이클들에 걸쳐 상기 셀들의 2-차원 어레이의 각각의 셀에 개개의 가중치 입력을 로딩시키도록 구성되는,

회로.

청구항 9

제8항에 있어서,

상기 프로세서는 추가로:

상기 복수의 클럭 사이클들에 걸쳐, 상기 개개의 가중치 입력을 상기 셀에 로딩시키고 그리고 상기 셀의 대응하는 가중치 레지스터에 저장시키는,

회로.

청구항 10

제1항에 있어서,

상기 회로는:

상기 가중치 입력들을 프리페칭(prefetch)하고; 그리고

상기 가중치 입력들의 프리페칭에 기초하여, 상기 셀들의 2-차원 어레이의 상기 제2 차원을 따라 상이한 셀들 각각에 상기 가중치 입력들을 프리로딩(preload)하도록 구성되는,

회로.

청구항 11

제10항에 있어서,

상기 가중치 입력들은, 상기 활성화 입력들이 상기 셀들의 2-차원 어레이의 상기 제2 차원을 따라 시프트되기 전에 프리로딩되는,

회로.

청구항 12

제1항에 있어서,

상기 셀들의 2-차원 어레이의 상기 제1 차원은 상기 어레이의 컬럼에 대응하고, 그리고

상기 셀들의 2-차원 어레이의 상기 제2 차원은 상기 어레이의 로우에 대응하는,

회로.

청구항 13

매트릭스 컴퓨테이션 유닛을 포함하는 회로를 사용하여 복수의 계층들을 포함하는 신경망에 대한 신경망 컴퓨테이션들을 수행하기 위한 방법으로서,

복수의 신경망 계층들 중 적어도 하나의 계층에 대해:

셀들에 의해 수행되는 동작들을 제어하기 위해 상기 매트릭스 컴퓨테이션 유닛에 상기 셀들의 2-차원 어레이의 제1 차원을 따라 제어 신호들을 제공하는 단계;

상기 셀들의 2-차원 어레이의 제2 차원을 따라 셀들에 복수의 가중치 입력들을 전송하는 단계 - 상기 셀들 각각은 상기 셀의 외부에 있는 소스로부터 수신된 가중치 입력들을 저장하도록 구성됨 -;

상기 셀들의 2-차원 어레이의 제2 차원을 따라 활성화 입력들을 전송하는 단계; 및

상기 제어 신호들에 기초하여, 상기 셀들의 2-차원 어레이의 셀들 각각에 의해, 상기 셀들에 저장된 가중치 입력들 및 상기 활성화 입력들을 사용하여 동작들을 수행하는 단계를 포함하고,

상기 제어 신호들은 각각의 셀이 상기 가중치 입력들 또는 상기 활성화 입력들을 인접한 셀들로 시프트 해야 하는지 여부를 결정하는,

방법.

청구항 14

제13항에 있어서,

상기 셀들의 2-차원 어레이의 각각의 셀은 가중치 레지스터를 포함하고,

상기 방법은:

상기 셀들의 개개의 가중치 레지스터에, 상기 셀들의 2-차원 어레이의 상기 제2 차원을 따라 시프트되는 상기 가중치 입력들 중 하나 이상의 가중치 입력을 저장하는 단계를 더 포함하는,

방법.

청구항 15

제14항에 있어서,

상기 셀들의 2-차원 어레이의 각각의 셀은 개개의 곱셈 회로를 포함하고,

상기 방법은:

상기 셀들의 상기 개개의 곱셈 회로를 사용하여, 상기 셀들의 2-차원 어레이의 상기 제2 차원을 따라 시프트되는 개개의 가중치 입력 및 활성화 입력의 곱을 계산하는 단계를 더 포함하는,

방법.

청구항 16

제15항에 있어서,

상기 셀들의 2-차원 어레이의 셀에 의해, 상기 셀들의 2-차원 어레이의 상기 제1 차원을 따라 제공되는 제어 신호들에 기초하여 제어 값을 수신하는 단계; 및

상기 제어 값에 기초하여, 상기 셀들의 2-차원 어레이의 상기 제2 차원을 따라 시프트되는 상기 개개의 가중치 입력을 상기 셀의 상기 개개의 가중치 레지스터에 저장할지를 결정하는,

방법.

청구항 17

제13항에 있어서,

상기 복수의 신경망 계층들 각각에 대해, 상기 방법은:

클록 신호들을 생성하는 단계;

상기 클록 신호들의 타이밍에 기초하여, 상기 셀들의 2-차원 어레이의 상기 셀들 각각에 상기 제어 신호들을 전송하는 단계; 및

상기 제어 신호들에 기초하여, 상이한 시간들에 그리고 상이한 셀들에서, 상기 제어 신호들의 일부를 실행하는 단계를 더 포함하는,

방법.

청구항 18

제17항에 있어서,

상기 제어 신호들 및 상기 클록 신호들의 타이밍에 기초하여, 상기 셀들의 2-차원 어레이의 상기 제2 차원을 따라 상이한 셀들에 상기 가중치 입력들 및 활성화 입력들을 시프트하는 단계를 더 포함하는,

방법.

청구항 19

제13항에 있어서,

상기 회로는 프로세서를 포함하고,

상기 방법은:

상기 셀들의 2-차원 어레이의 상기 제1 차원을 따라 제공되는 상기 제어 신호들에 기초하여, 상기 프로세서에 의해, 상기 매트릭스 컴퓨테이션 유닛의 대응하는 셀에 개개의 가중치 입력을 로딩시키는 단계를 더 포함하고,

상기 개개의 가중치 입력은 상기 셀들의 2-차원 어레이의 상기 제2 차원을 따라 시프트되는 상기 가중치 입력들 중 하나의 가중치 입력인,

방법.

청구항 20

제19항에 있어서,

상기 제어 신호들에 기초하여, 상기 프로세서에 의해, 복수의 클록 사이클들에 걸쳐 상기 셀들의 2-차원 어레이의 각각의 셀에 개개의 가중치 입력을 로딩시키는 단계를 더 포함하는,

방법.

청구항 21

제20항에 있어서,

상기 복수의 클록 사이클들에 걸쳐, 상기 프로세서에 의해, 상기 개개의 가중치 입력을 상기 셀에 로딩시키고 그리고 상기 셀의 대응하는 가중치 레지스터에 저장시키는,

방법.

청구항 22

제13항에 있어서,

상기 가중치 입력들을 프리페칭하는 단계; 및

상기 가중치 입력들의 프리페칭에 기초하여, 상기 셀들의 2-차원 어레이의 상기 제2 차원을 따라 상이한 셀들 각각에 상기 가중치 입력들을 프리로딩하는 단계를 더 포함하고,

상기 가중치 입력들은, 상기 활성화 입력들이 상기 셀들의 2-차원 어레이의 상기 제2 차원을 따라 시프트되기 전에 프리로딩되는,

방법.

청구항 23

매트릭스 컴퓨테이션 유닛을 포함하는 회로를 사용하여 복수의 신경망 계층들을 포함하는 신경망에 대한 신경망 컴퓨테이션들을 수행하기 위한 명령들이 저장된 비-일시적 기계-판독 저장 디바이스로서,

상기 명령들은 동작들을 수행시키기 위해 프로세싱 디바이스에 의해 실행가능하고, 상기 동작들은:

상기 복수의 신경망 계층 중 적어도 하나의 계층에 대해:

셀들에 의해 수행되는 동작들을 제어하기 위해 상기 매트릭스 컴퓨테이션 유닛에 상기 셀들의 2-차원 어레이의 제1 차원을 따라 제어 신호들을 제공하는 동작;

상기 셀들의 2-차원 어레이의 제2 차원을 따라 셀들에 복수의 가중치 입력들을 전송하는 동작 — 상기 셀들 각각은 상기 셀의 외부에 있는 소스로부터 수신된 가중치 입력들을 저장하도록 구성됨 —;

상기 셀들의 2-차원 어레이의 상기 제2 차원을 따라 활성화 입력들을 전송하는 동작; 및

상기 제어 신호들에 기초하여, 상기 셀들의 2-차원 어레이의 셀들 각각에 의해, 상기 셀들에 저장된 가중치 입력들 및 상기 활성화 입력들을 사용하여 동작들을 수행하는 동작을 포함하고,

상기 제어 신호들은 각각의 셀이 상기 가중치 입력들 또는 상기 활성화 입력들을 인접한 셀들로 시프트 해야 하는지 여부를 결정하는,

비-일시적 기계-판독 저장 디바이스.

발명의 설명

기술 분야

이 명세서는 하드웨어에서의 신경망 추론들을 컴퓨팅하는 것에 관한 것이다.

배경 기술

[0001]

[0002] 신경망들은 수신된 입력에 대한 출력(예를 들어, 분류)을 생성하기 위해 모델들의 하나 또는 그 초과 계층들을 활용하는 기계 학습 모델들이다. 일부 신경망들은 출력 계층 이외에도 하나 또는 그 초과 숨겨진 계층들을 포함한다. 각각의 숨겨진 계층의 출력은 네트워크의 다음 계층, 즉 네트워크의 다음 숨겨진 계층 또는 출력 계층에 대한 입력으로 사용된다. 네트워크의 각각의 계층은 파라미터들의 각각의 세트의 현재 값들에 따라, 수신된 입력으로부터 출력을 생성한다.

[0003] 일부 신경망들은 하나 또는 그 초과 컨벌루션 신경망 계층들을 포함한다. 각각의 컨벌루션 신경망 계층은 커널(kernel)들의 연관된 세트를 갖는다. 각각의 커널은 사용자에 의해 생성된 신경망 모델에 의해 확립된 값들을 포함한다. 일부 구현들에서, 커널들은 특정 이미지 컨투어들, 형상들 또는 컬러들을 식별한다. 커널들은 가중치 입력들의 매트릭스 구조로 나타내어질 수 있다. 각각의 컨벌루션 계층은 또한 활성화 입력들의 세트를 프로세싱할 수 있다. 활성화 입력들의 세트는 매트릭스 구조로도 나타내어질 수 있다.

[0004] 일부 기존 시스템들은 소프트웨어에서, 주어진 컨벌루션 계층에 대한 컴퓨테이션들을 수행한다. 예를 들어, 소프트웨어는 계층에 대한 각각의 커널을 활성화 입력들의 세트에 적용할 수 있다. 즉, 각각의 커널에 대해, 소프트웨어는 다차원적으로 표현될 수 있는 커널을, 다차원적으로 표현될 수 있는 활성화 입력들의 제 1 부분 위로 오버레이할 수 있다. 그런다음, 소프트웨어는 중첩된 엘리먼트(overlapped element)들로부터 도트 곱(dot product)을 컴퓨팅할 수 있다. 도트 곱은 단일 활성화 입력, 예를 들어, 중첩된 다차원 공간에서 상부-좌측 포지션을 갖는 활성화 입력 엘리먼트에 대응할 수 있다. 예를 들어, 슬라이딩 윈도우를 사용하여, 이후 소프트웨어는 커널을 시프트시켜 활성화 입력들의 제 2 부분을 오버레이하고 다른 활성화 입력에 대응하는 다른 도트 곱을 계산할 수 있다. 소프트웨어는, 각각의 활성화 입력이, 대응하는 도트 곱을 가질 때까지 이 프로세스를 반복적으로 수행할 수 있다. 일부 구현들에서, 도트 곱들은 활성화 값들을 생성하는 활성화 함수에 입력된다. 활성화 값들은, 신경망의 후속 계층으로 전송되기 전에 결합, 예를 들어, 풀링될 수 있다.

[0005] 컨벌루션 계산을 컴퓨팅하는 한 가지 방법은 큰 차원 공간에서 다수의 매트릭스 곱셈들을 요구한다. 프로세서는 억지 방법(brute force method)을 통해 매트릭스 곱셈들을 컴퓨팅할 수 있다. 예를 들어, 계산 집약적이고 시간 집약적이긴 하지만, 프로세서는 컨벌루션 계산을 위해 개개의 합산들 및 곱들을 반복적으로 계산할 수 있다. 프로세서가 계산들을 병렬화하는 정도는, 자신의 아키텍처로 인해 제한된다.

발명의 내용

[0006] 일반적으로, 본 명세서는 신경망 추론을 컴퓨팅하는 특수 목적 하드웨어 회로를 설명한다.

[0007] 일반적으로, 본 명세서에서 설명된 요지의 일 혁신적인 양상은 복수의 계층들을 포함하는 신경망에 대한 신경망 컴퓨테이션들을 수행하기 위한 회로에서 구현될 수 있으며, 회로는 복수의 셀들을 포함하는 시스틀릭 어레이; 가중치 페치 유닛 -가중치 페치 유닛은, 복수의 가중치 입력들을 신경망 계층에 대한 시스틀릭 어레이의 제 1 차원을 따라 셀들에 전송하도록, 복수의 신경망 계층들의 각각에 대해 구성됨-; 및 복수의 가중치 시퀀서 유닛들을 포함하고, 각각의 가중치 시퀀서 유닛은 시스틀릭 어레이의 제 1 차원을 따라 별개의 셀에 커플링되고, 복수의 가중치 시퀀서 유닛들은, 복수의 클럭 사이클들을 통해 복수의 가중치 입력들을 신경망 계층에 대한 시스틀릭 어레이의 제 2 차원을 따라 셀들로 시프트하도록, 복수의 신경망 계층들 각각에 대해 구성되고, 각각의 가중치 입력은 제 2 차원을 따라 각각의 셀 내부에 저장되고, 각각의 셀은 곱셈 회로를 사용하여 활성화 입력 및 각각의 가중치 입력의 곱을 컴퓨팅하도록 구성된다.

[0008] 구현들은 다음 중 하나 또는 그 초과 것을 포함할 수 있다. 값 시퀀서 유닛은, 복수의 신경망 계층들 각각에 대해, 복수의 활성화 입력들을 신경망 계층에 대한 시스틀릭 어레이의 제 2 차원을 따라 셀들에 전송하도록 구성된다. 시스틀릭 어레이의 제 1 차원은 시스틀릭 어레이의 로우들에 대응하고, 시스틀릭 어레이의 제 2 차원은 시스틀릭 어레이의 컬럼들에 대응한다. 각각의 셀은 인접 셀에 가중치 제어 신호를 전달하도록 구성되고, 가중치 제어 신호는 인접 셀의 회로가 인접 셀에 대한 가중치 입력을 시프트하게 하거나 또는 로딩하게 한다. 각각의 셀은, 셀로 시프트된 가중치 입력을 저장하도록 구성된 가중치 경로 레지스터; 가중치 경로 레지스터에 커플링된 가중치 레지스터; 가중치 입력을 가중치 레지스터에 저장할지 여부를 결정하도록 구성된 가중치 제어 레지스터; 활성화 입력을 저장하도록 구성되고 그리고 활성화 입력을 제 1 차원을 따라 제 1 인접 셀의 다른 활성화 레지스터에 전송하도록 구성되는 활성화 레지스터; 가중치 레지스터 및 활성화 레지스터에 커플링되는 곱셈 회로 -곱셈 회로는 가중치 입력 및 활성화 입력의 곱을 출력하도록 구성됨-; 곱셈 회로에 커플링되고 제 2 차원을 따라 제 2 인접 셀로부터 제 1 부분 합 및 곱을 수신하도록 구성되는 합산 회로 -합산 회로는 제 1 부분 합 및 곱의 제 2 부분 합을 출력하도록 구성됨-; 및 합산 회로에 커플링되고 그리고 제 2 부분 합을 저장하

도록 구성되는 부분 합 레지스터를 포함하고, 부분 합 레지스터는 제 2 부분 합을 제 2 차원을 따라 제 3 인접 셀의 다른 합산 회로에 전송하도록 구성된다. 각각의 가중치 시퀀서 유닛은, 가중치 시퀀서 유닛에 커플링된 대응하는 셀 내의 가중치 제어 레지스터에 대응하는 일시정지 카운터; 및 감소 회로를 포함하고, 감소 회로는 감소된 출력을 생성하기 위해 가중치 시퀀서 유닛에 대한 입력을 감소시키고 감소된 출력을 일시정지 카운터에 전송하도록 구성된다. 각각의 일시정지 카운터의 값들이 동일하고, 각각의 가중치 시퀀서 유닛은 대응하는 가중치 입력을 시스틀릭 어레이의 대응하는 별개의 셀에 로딩하도록 구성되고, 로딩하는 것은 가중치 입력을 곱셈 회로에 전송하는 것을 포함한다. 각각의 일시정지 카운터의 값들이 미리결정된 값에 도달하면, 복수의 가중치 시퀀서 유닛들은 제 2 차원을 따르는 복수의 가중치 입력들의 시프트를 일시정지한다. 시스틀릭 어레이는, 복수의 신경망 계층들 각각에 대해, 각각의 곱으로부터 신경망 계층에 대한 누산 출력을 생성하도록 구성된다.

[0009] 본 명세서에 설명된 요지의 특정 실시예들은, 다음의 이점들 중 하나 또는 그 초과를 실현하기 위해서 구현될 수 있다. 가중치를 프리페칭하는 것은, 신경망 프로세서가 보다 효율적으로 컴퓨테이션들을 수행할 수 있게 한다. 프로세서는 가중치 페치 유닛 및 가중치 시퀀서 유닛을 사용하여 가중치 입력들을 시스틀릭 어레이에 로딩하는 것을 조정할 수 있음으로써, 외부 메모리 유닛을 시스틀릭 어레이 내의 각각의 셀에 커플링시키는 와이어의 필요성을 없앤다. 프로세서는 다수개의 컨벌루션 컴퓨테이션들의 수행을 동기화하기 위해서 가중치 입력들을 시프트시키는 것을 일시정지, 즉 "프리징(freeze)"시킬 수 있다.

[0010] 본 명세서의 요지의 하나 또는 그 초과를 실시예들의 상세들은 첨부된 도면들 및 아래의 설명에 제시된다. 요지의 다른 특징들, 양상들 및 이점들은 설명, 도면들 및 청구항들로부터 명백해질 것이다.

도면의 간단한 설명

[0011] 도 1은 신경망의 주어진 계층에 대한 컴퓨테이션을 수행하기 위한 예시적인 방법의 흐름도이다.

도 2는 예시적인 신경망 프로세싱 시스템을 도시한다.

도 3은 매트릭스 컴퓨테이션 유닛을 포함하는 예시적인 아키텍처를 도시한다.

도 4는 시스틀릭 어레이 내부의 셀의 예시적인 아키텍처를 도시한다.

도 5는 공간 차원들 및 특징 차원을 갖는 예시적인 매트릭스 구조를 도시한다.

도 6은 커널 매트릭스 구조가 시스틀릭 어레이에 전송되는 방법을 예시하는 예를 도시한다.

도 7은 3개의 클록 사이클들 이후 셀들 내부의 가중치 입력들을 예시하는 예를 도시한다.

도 8은 제어 신호들이 활성화 입력들을 시프트시키거나 또는 로딩시키는 방법을 예시하는 예이다.

다양한 도면들에서의 동일한 참조 번호들 및 표기들은 동일한 엘리먼트들을 표시한다.

발명을 실시하기 위한 구체적인 내용

[0012] 다수개의 계층들을 갖는 신경망은 추론들을 컴퓨팅하기 위해 사용될 수 있다. 예를 들어, 입력이 주어지면, 신경망은 입력에 대한 추론을 컴퓨팅할 수 있다. 신경망은 신경망의 계층들 각각을 통해 입력을 프로세싱함으로써 이 추론을 컴퓨팅한다. 특히, 신경망의 계층들은, 가중치들의 각각의 세트를 각각 가지고 시퀀스로 배열된다. 각각의 계층은 입력을 수신하고 계층에 대한 가중치들의 세트에 따라 입력을 프로세싱하여 출력을 생성한다.

[0013] 따라서, 수신된 입력으로부터 추론을 컴퓨팅하기 위해서, 신경망이 입력을 수신하고 시퀀스에서 신경망 계층들 각각을 통해 입력을 프로세싱하여 추론을 생성하며, 하나의 신경망 계층으로부터의 출력이 다음 신경망 계층에 대한 입력으로서 제공된다. 신경망 계층에 대한 데이터 입력들, 예를 들어, 신경망으로의 입력, 또는 시퀀스에서 계층 밑의 계층의, 신경망 계층으로의 출력들이 계층에 대한 활성화 입력들로 지칭될 수 있다.

[0014] 일부 구현들에서, 신경망의 계층들은 방향 그래프로 배열된다. 즉, 임의의 특정 계층은 다수개의 입력들, 다수개의 출력들, 또는 둘 모두를 수신할 수 있다. 신경망의 계층들은 또한, 계층의 출력이 이전 계층에 대한 입력으로서 다시 되돌려 보내질 수 있도록 배열될 수 있다.

[0015] 도 1은 특수 목적 하드웨어 회로를 사용하여 신경망의 주어진 계층에 대한 컴퓨테이션을 수행하기 위한 일 예시적인 프로세스(100)의 흐름도이다. 편의상, 방법(100)은, 방법(100)을 수행하는 하나 또는 그 초과회로들을 갖는 시스템과 관련하여 설명될 것이다. 방법(100)은, 수신된 입력으로부터 추론을 컴퓨팅하기 위해 신경망의

각각의 계층에 대해 수행될 수 있다.

- [0016] 시스템은 주어진 계층에 대한 가중치 입력들의 세트들을 수신하고(단계 102) 활성화 입력들의 세트들을 수신한다(단계 104). 가중치 입력들의 세트들 및 활성화 입력들의 세트들은, 특수 목적 하드웨어 회로의 동적 메모리 및 통합 버퍼로부터 각각 수신될 수 있다. 일부 구현들에서, 가중치 입력들의 세트들 및 활성화 입력들의 세트들 둘 모두가 통합 버퍼로부터 수신될 수 있다.
- [0017] 시스템은 특수 목적 하드웨어 회로의 매트릭스 곱셈 유닛을 사용하여 가중치 입력들 및 활성화 입력들로부터 누산 값들을 생성한다(단계 106). 일부 구현들에서, 누산 값들은 가중치 입력들의 세트들과 활성화 입력들의 세트들의 도트 곱들이다. 즉, 가중치들의 하나의 세트의 경우, 시스템은 각각의 가중치 입력을 각각의 활성화 입력과 곱셈하고 그 곱들을 함께 합산하여 누산 값을 형성할 수 있다. 그런 다음, 시스템은 가중치들의 다른 세트와 활성화 입력들의 다른 세트들과의 도트 곱을 컴퓨팅할 수 있다.
- [0018] 시스템은 특수 목적 하드웨어 회로의 벡터 계산 유닛을 사용하여 누산 값들로부터 계층 출력을 생성할 수 있다(단계 108). 일부 구현들에서, 벡터 컴퓨테이션 유닛은 누산 값들에 활성화 함수를 적용한다. 계층의 출력은 신경망의 후속 계층에 대한 입력으로 사용하기 위해 통합 버퍼에 저장될 수 있거나 또는 추론을 결정하는데 사용될 수 있다. 시스템은, 수신된 입력이 신경망의 각각의 계층을 통해 프로세싱되어 수신된 입력에 대한 추론이 생성되었을 경우 신경망의 프로세싱을 종료한다.
- [0019] 도 2는 신경망 컴퓨테이션들을 수행하기 위한 예시적인 특수 목적 통합 회로(200)를 도시한다. 시스템(200)은 호스트 인터페이스(202)를 포함한다. 호스트 인터페이스(202)는 신경망 컴퓨테이션에 대한 파라미터들을 포함하는 명령들을 수신할 수 있다. 파라미터들은 다음: 얼마나 많은 계층들이 프로세싱되어야 하는지, 계층의 각각의 계층에 대한 가중치 입력들의 대응하는 세트들, 활성화 입력들의 초기 세트, 즉, 추론이 컴퓨팅되는 신경망에 대한 입력, 각각의 계층의 대응하는 입력 및 출력 사이즈들, 신경망 컴퓨테이션을 위한 스트라이드 값, 및 프로세싱될 계층의 타입, 예를 들어, 컨벌루션 계층 또는 완전 연결된 계층 중 적어도 하나 또는 그 초과를 포함할 수 있다.
- [0020] 호스트 인터페이스(202)는, 명령들을 시퀀스(206)로 전송할 수 있으며, 시퀀스(206)는 명령들을, 신경망 컴퓨테이션들을 수행하도록 회로를 제어하는 저 레벨 제어 신호들로 변환한다. 일부 구현들에서, 제어 신호들은 회로에서의 데이터 흐름을 조절하는데, 예를 들어 가중치 입력들의 세트들 및 활성화 입력들의 세트들이 회로를 통해 흐르는 방법을 조절한다. 시퀀스(206)는 제어 신호들을 통합 버퍼(208), 매트릭스 컴퓨테이션 유닛(212), 및 벡터 컴퓨테이션 유닛(214)에 전송할 수 있다. 일부 구현에서, 시퀀스(206)는 또한 직접 메모리 액세스 엔진(204) 및 동적 메모리(210)에 제어 신호들을 전송한다. 일부 구현들에서, 시퀀스(206)는 클럭 신호들을 생성하는 프로세서이다. 시퀀스(206)는, 적절한 시간들에, 클럭 신호들을 회로(200)의 각각의 컴포넌트에 전송하기 위해 클럭 신호들의 타이밍을 사용할 수 있다. 일부 다른 구현들에서, 호스트 인터페이스(202)는 외부 프로세서로부터의 클럭 신호를 전달한다.
- [0021] 호스트 인터페이스(202)는 가중치 입력들의 세트들 및 활성화 입력들의 초기 세트들을 직접 메모리 액세스 엔진(204)에 전송할 수 있다. 직접 메모리 액세스 엔진(204)은 통합 버퍼(208)에 활성화 입력들의 세트들을 저장할 수 있다. 일부 구현들에서, 직접 메모리 액세스는 메모리 유닛일 수 있는 동적 메모리(210)에 가중치들의 세트들을 저장한다. 일부 구현들에서, 동적 메모리는 회로로부터 떨어져 위치된다.
- [0022] 통합 버퍼(208)는 메모리 버퍼이다. 이는, 직접 메모리 액세스 엔진(204)으로부터의 활성화 입력들의 세트 및 벡터 컴퓨테이션 유닛(214)의 출력들을 저장하는데 사용될 수 있다. 직접 메모리 액세스 엔진(204)은 또한 통합 버퍼(208)로부터 벡터 컴퓨테이션 유닛(214)의 출력들을 판독할 수 있다.
- [0023] 동적 메모리(210) 및 통합 버퍼(208)는, 가중치 입력들의 세트들 및 활성화 입력들의 세트들을 각각 매트릭스 컴퓨테이션 유닛(212)에 전송할 수 있다. 일부 구현들에서, 매트릭스 컴퓨테이션 유닛(212)은 2차원 시스템릭 어레이이다. 매트릭스 컴퓨테이션 유닛(212)은 또한, 1차원 시스템릭 어레이이거나 또는 수학적 연산들, 예를 들어, 곱셈 및 덧셈을 수행할 수 있는 다른 회로일 수 있다. 일부 구현들에서, 매트릭스 컴퓨테이션 유닛(212)은 범용 매트릭스 프로세서이다.
- [0024] 매트릭스 컴퓨테이션 유닛(212)은 가중치 입력들 및 활성화 입력들을 프로세싱하고 벡터 컴퓨테이션 유닛(214)에 출력들의 벡터를 제공할 수 있다. 몇몇 구현들에서, 매트릭스 컴퓨테이션 유닛은 출력들의 벡터를 통합 버퍼(208)에 전송하며, 통합 버퍼(208)는 출력들의 벡터를 벡터 컴퓨테이션 유닛(214)으로 전송한다. 벡터 컴퓨테이션 유닛은 출력들의 벡터를 프로세싱하고 프로세싱된 출력들의 벡터를 통합 버퍼(208)에 저장할 수 있다.

예를 들어, 벡터 컴퓨테이션 유닛(214)은 비선형 함수를 매트릭스 컴퓨테이션 유닛의 출력들, 예를 들어, 누산 값들의 벡터에 적용하여 활성화된 값들을 생성할 수 있다. 일부 구현들에서, 벡터 컴퓨테이션 유닛(214)은 정규화된 값들, 풀링된 값들, 또는 둘 모두를 생성한다. 프로세싱된 출력들의 벡터는, 예를 들어, 신경망의 후속 계층에서 사용하기 위해, 매트릭스 컴퓨테이션 유닛(212)에 대한 활성화 입력들로서 사용될 수 있다. 매트릭스 컴퓨테이션 유닛(212)은 도 3 및 도 4를 참조하여 아래에 더욱 상세히 설명될 것이다.

- [0025] 도 3은 매트릭스 컴퓨테이션 유닛을 포함하는 예시적인 아키텍처(300)를 도시한다. 매트릭스 컴퓨테이션 유닛은 2차원 시스틀릭 어레이(306)이다. 어레이(306)는 다수개의 셀들(304)을 포함한다. 일부 구현들에서, 시스틀릭 어레이(306)의 제 1 차원(320)은 셀들의 컬럼들에 대응하고 시스틀릭 어레이(306)의 제 2 차원(322)은 셀들의 로우들에 대응한다. 시스틀릭 어레이는 컬럼들보다 더 많은 로우들을 갖거나, 로우들보다 더 많은 컬럼들을 갖거나, 또는 같은 수의 컬럼들 및 로우들을 가질 수 있다.
- [0026] 예시된 예에서, 값 로더들(302)은 어레이(306)의 로우들에 활성화 입력들을 전송하고, 가중치 페처 인터페이스(308)는 가중치 입력들을 어레이(306)의 컬럼들에 전송한다. 그러나, 다른 구현들에서, 활성화 입력들이 컬럼들로 전달되고 가중치 입력들이 어레이(306)의 로우들로 전달된다.
- [0027] 값 로더들(302)은 통합 버퍼, 예컨대, 도 2의 통합 버퍼(208)로부터 활성화 입력들을 수신할 수 있다. 각각의 값 로더는 대응하는 활성화 입력을 어레이(306)의 별개의 제일 좌측 셀에 전송할 수 있다. 제일 좌측 셀은 어레이(306)의 제일 좌측 컬럼을 따라 있는 셀일 수 있다. 예를 들어, 값 로더(312)는 활성화 입력을 셀(314)에 전송할 수 있다. 또한, 값 로더는 인접한 값 로더에 활성화 입력을 전송할 수 있으며, 활성화 입력은 어레이(306)의 다른 제일 좌측 셀에서 사용될 수 있다. 이는, 활성화 입력들이 어레이(306)의 다른 특정 셀에서의 사용을 위해 시프트되도록 허용한다.
- [0028] 가중치 페처 인터페이스(308)는, 메모리 유닛, 예를 들어, 도 2의 동적 메모리(210)로부터 가중치 입력을 수신할 수 있다. 가중치 페처 인터페이스(308)는 대응하는 가중치 입력을 어레이(306)의 별개의 제일 상부 셀에 전송할 수 있다. 제일 상부 셀은 어레이(306)의 제일 상부 로우를 따라 있는 셀일 수 있다. 예를 들어, 가중치 페처 인터페이스(308)는 가중치 입력들을 셀들(314 및 316)에 전송할 수 있다.
- [0029] 일부 구현들에서, 호스트 인터페이스, 예를 들어, 도 2의 호스트 인터페이스(202)는 어레이(306) 전체에 걸친 활성화 입력들을 하나의 차원을 따라, 예를 들어, 우측으로 시프트하는 반면, 어레이(306) 전체에 걸친 가중치 입력들을 다른 차원을 따라, 예를 들어, 하부로 시프트한다. 예를 들어, 1 클록 사이클 동안, 셀(314)에서의 활성화 입력이, 셀(314)의 우측에 있는 셀(316)의 활성화 레지스터로 시프트할 수 있다. 유사하게, 셀(316)에서의 가중치 입력은 셀(314) 아래의 셀(318)의 가중치 레지스터로 시프트할 수 있다.
- [0030] 각각의 클록 사이클에서, 각각의 셀은 주어진 가중치 입력, 및 주어진 활성화 입력을 프로세싱하여 누산 출력을 생성할 수 있다. 누산 출력은 또한, 주어진 가중치 입력과 동일한 차원을 따라 인접 셀로 전달될 수 있다. 개개의 셀이 도 4를 참조하여 아래에서 추가로 설명된다.
- [0031] 일부 구현들에서, 가중치들 및 활성화들은, 하나의 컨벌루션 계산에서 다른 계산으로 전이하기 위해 주어진 클록 사이클 동안 1개 초과 셀만큼 시프트된다.
- [0032] 누산 출력이 가중치 입력과 동일한 컬럼을 따라 전달될 수 있는데, 예를 들어, 어레이(306)의 컬럼의 하부를 향하여 전달될 수 있다. 일부 구현들에서, 각각의 컬럼의 하부에서, 어레이(306)는, 컬럼들보다 더 많은 가중치 입력들을 갖는 계층들 또는 로우들보다 더 많은 활성화 입력들을 갖는 계층들과 함께 계산들을 수행할 경우 각각의 컬럼으로부터의 각각의 누산 출력을 저장하고 누산하는 누산기 유닛(310)을 포함할 수 있다. 일부 구현들에서, 각각의 누산기 유닛은 다수개의 병렬 누산들을 저장한다. 이는 도 6을 참조하여 아래에서 추가로 설명될 것이다. 누산기 유닛들(310)은 각각의 누산 출력을 누산하여 최종 누산 값을 생성할 수 있다. 최종 누산 값은 벡터 컴퓨테이션 유닛으로 전달될 수 있다. 일부 다른 구현들에서, 누산기 유닛들(310)은, 컬럼들보다 더 적은 수의 가중치 입력들을 갖는 계층들 또는 로우들보다 더 적은 수의 활성화 입력들을 갖는 계층들을 프로세싱할 경우 어떠한 누산들도 수행하지 않고 누산 값들을 벡터 계산 유닛으로 전달한다.
- [0033] 활성화 입력들 및 가중치 입력들이 회로를 통해 흐를 때, 회로는 누산 값들을 정확하게 계산하기 위해서 흐름으로부터 가중치 입력들의 세트를 "프리징"하거나 일시정지할 수 있다. 즉, 회로가 가중치 입력들의 세트를 일시정지할 수 있으므로 가중치 입력들의 특정 세트가 활성화 입력들의 특정 세트에 적용될 수 있다.
- [0034] 일부 구현들에서, 가중치 시퀀서들(324)은, 가중치 입력들이 인접 셀들로 시프트할지 여부를 구성한다. 가중치 시퀀서(326)는 호스트, 예를 들어, 도 2의 호스트 인터페이스(202)로부터, 또는 외부 프로세서로부터 제어 값을

수신할 수 있다. 각각의 체중 시퀀서는 어레이(306) 내의 대응하는 셀에 제어 값을 전달할 수 있다. 특히, 제어 값은 셀 내의 가중치 제어 레지스터, 예를 들어, 도 4의 가중치 제어 레지스터(414)에 저장될 수 있다. 제어 값은, 가중치 입력들이 어레이의 차원을 따라 시프트될지 또는 로딩될지 여부를 결정할 수 있으며, 이는 도 8과 관련하여 아래에 설명될 것이다. 또한, 가중치 시퀀서는 제어 값을 인접한 가중치 시퀀서에 전송할 수 있으며, 이 가중치 시퀀서는 대응하는 셀에 대한 대응하는 가중치 입력을 시프트시키는 것 또는 로딩하는 것을 조절할 수 있다.

[0035] 일부 구현들에서, 제어 값은 정수로 표현된다. 각각의 가중치 시퀀서는 정수를 저장하는 일시정지 카운터 레지스터를 포함할 수 있다. 또한, 가중치 시퀀서는 제어 값을 일시정지 카운터 레지스터에 저장하기 전에 정수를 감소시킬 수 있다. 일시정지 카운터 레지스터에 제어 값을 저장한 후, 가중치 시퀀서가 정수를 인접한 가중치 시퀀서로 전송하고 그 정수를 대응하는 셀에 전송할 수 있다. 예를 들어, 각각의 가중치 시퀀서는, 감소된 정수를 제어 값으로부터 생성하도록 구성된 감소 회로를 구비할 수 있다. 감소된 정수는 일시정지 카운터 레지스터에 저장될 수 있다. 저장된 제어 값들은 어레이의 전체 컬럼에 걸쳐 시프트의 동시 일시정지를 조정하기 위해 사용될 수 있으며, 이는 도 8과 관련하여 아래에 추가로 설명될 것이다.

[0036] 일부 구현들에서, 회로에서 가중치들을 일시정지하는 것은 개발자가 회로를 디버깅할 수 있게 한다.

[0037] 가중치들을 일시정지하는 다른 방법들이 가능하다. 예를 들어, 일시정지 카운터 레지스터들 내부의 값들을 인접한 일시정지 카운터 레지스터들로 전달하는 대신, 트리를 사용하여 제어 값이 전달될 수 있다. 즉, 주어진 셀에서, 신호는 모든 인접 셀들로 전달될 수 있고, 하나의 인접 셀에만 전달되지 않음으로써, 신호가 시스틀릭 어레이 전체에 걸쳐 신속하게 확산될 수 있게 한다.

[0038] 도 4는 시스틀릭 어레이, 예를 들어, 도 3의 시스틀릭 어레이(306) 내부의 셀의 예시적인 아키텍처(400)를 도시한다.

[0039] 셀은 활성화 입력을 저장하는 활성화 레지스터(406)를 포함할 수 있다. 활성화 레지스터는, 시스틀릭 어레이 내의 셀의 포지션에 따라, 좌측 인접 셀, 즉, 주어진 셀의 좌측에 위치되는 인접 셀로부터, 또는 통합 버퍼로부터 활성화 입력을 수신할 수 있다. 셀은 가중치 입력을 저장하는 가중치 레지스터(402)를 포함할 수 있다. 가중치 입력은, 시스틀릭 어레이 내의 셀의 포지션에 따라, 상부 인접 셀로부터 또는 가중치 패치 인터페이스로부터 전달될 수 있다. 셀은 또한 레지스터의 합(404)을 포함할 수 있다. 레지스터의 합(404)은 상부 인접 셀로부터의 누산 값을 저장할 수 있다. 곱셈 회로(408)는 가중치 레지스터(402)로부터의 가중치 입력을 활성화 레지스터(406)로부터의 활성화 입력과 곱하기 위해 사용될 수 있다. 곱셈 회로(408)는 곱을 합산 회로(410)로 출력할 수 있다.

[0040] 합산 회로는 레지스터의 합(404)으로부터의 누산 값을 곱과 합산하여 새로운 누산 값을 생성할 수 있다. 그런 다음, 합산 회로(410)는 새로운 누산 값을 하부 인접 셀에 위치한 레지스터 내의 다른 합으로 전송할 수 있다. 새로운 누산 값은 하부 인접 셀에서의 합산을 위한 피연산자로서 사용될 수 있다.

[0041] 일부 구현들에서, 셀은 또한 일반 제어 레지스터를 포함한다. 제어 레지스터는, 셀이 가중치 입력 또는 활성화 입력을 인접 셀들로 시프트해야 하는지 여부를 결정하는 제어 신호를 저장할 수 있다. 일부 구현들에서, 가중치 입력 또는 활성화 입력을 시프트하는 것은 2 이상의 클록 사이클이 걸린다. 제어 신호는 또한, 활성화 입력 또는 가중치 입력들이 곱셈 회로(408)로 전달되는지 여부를 결정할 수 있거나, 또는 곱셈 회로(408)가 활성화 입력 및 가중치 입력에 대해 연산하는지를 여부를 결정할 수 있다. 제어 신호는 또한, 예를 들어, 와이어를 사용하여, 하나 또는 그 초과인 인접 셀들로 전달될 수 있다.

[0042] 일부 구현들에서, 가중치들은 가중치 경로 레지스터(412)로 프리-시프트된다. 가중치 경로 레지스터(412)는, 예를 들어, 상부 인접 셀로부터 가중치 입력을 수신할 수 있고, 제어 신호에 기초하여 가중치 입력을 가중치 레지스터(402)에 전달할 수 있다. 가중치 레지스터(402)는, 활성화 입력들이, 예를 들어, 활성화 레지스터(406)를 통해 다수개의 클록 사이클들 동안 셀로 전달될 때, 가중치 입력이 셀 내에 유지되고 인접 셀로 전달되지 않도록 가중치 입력을 정적으로 저장할 수 있다. 따라서, 가중치 입력은, 예를 들어, 곱셈 회로(408)를 사용하여 다수개의 활성화 입력들에 적용될 수 있고, 각각의 누산 값들이 인접 셀로 전달될 수 있다.

[0043] 일부 구현들에서, 가중치 제어 레지스터(414)는, 가중치 입력이 가중치 레지스터(402)에 저장되는지 여부를 제어한다. 예를 들어, 가중치 제어 레지스터(414)가 0의 제어 값을 저장하는 경우, 가중치 레지스터(402)는 가중치 경로 레지스터(412)에 의해 전송된 가중치 입력을 저장할 수 있다. 일부 구현들에서, 가중치 입력을 가중치 레지스터(402)에 저장하는 것은 가중치 입력을 로딩하는 것으로 지칭된다. 일단 가중치 입력이 로딩되면, 가중

치 입력은 프로세싱을 위해 곱셈 회로(408)로 전송될 수 있다. 가중치 제어 레지스터(414)가 0-제어 값을 저장하는 경우, 가중치 레지스터(402)는 가중치 경로 레지스터(412)에 의해 전송된 가중치 입력을 무시할 수 있다. 가중치 제어 레지스터(414)에 저장된 제어 값은, 하나 또는 그 초과인 인접 셀들로 전달될 수 있고, 예를 들어, 주어진 셀의 경우, 제어 값이 그 주어진 셀의 우측에 위치한 셀의 가중치 제어 레지스터로 전송될 수 있다.

[0044] 셀은 또한, 가중치 입력 및 활성화 입력을 인접 셀들로 시프트시킬 수 있다. 예를 들어, 가중치 경로 레지스터(412)는 가중치 입력을 하부 인접 셀 내의 다른 가중치 경로 레지스터에 전송할 수 있다. 활성화 레지스터(406)는 활성화 입력을 우측 인접 셀 내의 다른 활성화 레지스터에 전송할 수 있다. 따라서, 가중치 입력 및 활성화 입력 둘 모두는, 후속 클럭 사이클에서 어레이 내의 다른 셀들에 의해 재사용될 수 있다.

[0045] 도 5는 공간 차원들 및 특징 차원을 갖는 예시적인 매트릭스 구조(500)를 도시한다. 매트릭스 구조(500)는 활성화 입력들의 세트 또는 가중치 입력들의 세트를 나타낼 수 있다. 활성화 입력들의 세트에 대한 매트릭스 구조는 본 명세서에서 활성화 매트릭스 구조로 지칭될 것이고, 가중치 입력들의 세트에 대한 매트릭스 구조는 본 명세서에서 커널 매트릭스 구조로 지칭될 것이다. 매트릭스 구조(500)는 3개 차원들: 2개의 공간 차원들 및 1개의 특징 차원을 갖는다.

[0046] 일부 구현들에서, 공간 차원들은 활성화 입력들의 세트의 공간 또는 포지션에 대응한다. 예를 들어, 신경망이 2개의 차원들을 갖는 이미지를 프로세싱하고 있는 경우, 매트릭스 구조들은 이미지의 공간 좌표들, 즉 XY 좌표들에 대응하는 2개의 공간 차원들을 가질 수 있다.

[0047] 특징 차원은 활성화 입력으로부터의 특징들에 대응한다. 각각의 특징 차원은 깊이 레벨들을 가질 수 있다: 예를 들어, 매트릭스 구조(500)는 깊이 레벨들(502, 504 및 506)을 갖는다. 예시로서, 매트릭스 구조(500)가 활성화 입력들의 세트로서 제 1 계층에 전송된 $3 \times 3 \times 3$ 이미지를 나타내는 경우, 이미지(3×3)의 X 및 Y 차원들은 공간 차원들이 될 수 있고, Z 차원(3)은 R, G 및 B 값들에 대응하는 특징 차원이 될 수 있다. 즉, 깊이 레벨(502)은, 9개의 '1' 활성화 입력들의 특징, 예를 들어, 적색 값들에 대응할 수 있고, 깊이 레벨(504)은 9개의 '2' 활성화 입력들의 특징, 예를 들어, 녹색 값들에 대응할 수 있고, 깊이 레벨(506)은 9개의 '3' 활성화 입력들의 특징, 예를 들어 청색 값들에 대응할 수 있다.

[0048] 도 5의 예에서 특징 차원에 대해 3개의 깊이 레벨들만이 도시되어 있지만, 주어진 특징 차원은 많은 수, 예를 들어, 수백 개의 특징 차원들을 가질 수 있다. 유사하게, 하나의 특징 차원만이 예시되어 있지만, 주어진 매트릭스 구조는 다수개의 특징 차원들을 가질 수 있다.

[0049] 매트릭스 구조(500)를 사용하여, 컨벌루션 계층에 대한 컴퓨테이션을 수행하기 위해, 시스템은 컨벌루션 컴퓨테이션을 2-차원 매트릭스 곱셈으로 변환해야 한다.

[0050] 도 6은 도 5의 매트릭스 구조(500)가 주어진 컨벌루션 계층에서 시스틀릭 어레이(606)에 의해 프로세싱되는 방법을 예시하는 예를 도시한다. 매트릭스 구조(600)는 활성화 입력들의 세트일 수 있다. 일반적으로, 신경망 프로세서는 활성화 입력들, 예를 들어, 매트릭스 구조(600) 내의 엘리먼트들, 및 가중치 입력들, 예를 들어, 커널들 A-D(610)를, 각각 어레이의 로우들 및 컬럼들에 전송할 수 있다. 활성화 및 가중치 입력들은 시스틀릭 어레이의 우측 및 하부로 각각 시프트될 수 있고, 특정 포지션, 예를 들어, 특정 셀의 특정 레지스터에 도달해야 한다. 예를 들어, 제어 신호를 검사함으로써 입력들이 제자리에 있는 것으로 결정되면, 프로세서는 셀들 내에 저장된 입력들을 사용하여 계산들을 수행하여 주어진 계층의 출력을 생성할 수 있다.

[0051] 신경망 프로세서는, 상술된 바와 같이, 구조(600)의 부분들을 시스틀릭 어레이의 로우들로 전송하기 전에 매트릭스 구조(600)를 "평탄화(flatten)"한다. 즉, 신경망 프로세서는 매트릭스 구조(600)의 깊이 계층들(602), 예를 들어, 도 6의 깊이 계층들(602, 604 및 606)을 분할하고, 각각의 깊이 계층을 별개의 셀로 전송할 수 있다. 일부 구현들에서, 각각의 깊이 계층은 시스틀릭 어레이(606)의 상이한 로우 상의 셀로 전송된다. 예를 들어, 프로세서는 제 1 깊이 계층으로부터의 활성화 입력들, 예를 들어, 9개의 '1' 활성화 입력들의 매트릭스를 시스틀릭 어레이(606)의 제 1 로우에 있는 제일 좌측 셀로, 제 2 깊이 계층으로부터의 활성화 입력들, 예를 들어, 9개의 '2' 활성화 입력들의 매트릭스를 제 2 로우에 있는 제일 좌측 셀로, 제 3 깊이 계층으로부터의 활성화 입력들, 예를 들어, 9개의 '3' 활성화 입력들의 매트릭스를 제 3 로우에 있는 제일 좌측 셀로 전송하는 식이다.

[0052] 주어진 계층은 다수의 커널들, 예를 들어, 커널들 A-D(610)를 가질 수 있다. 커널들 A-D(610)는 차원 $3 \times 3 \times 10$ 의 매트릭스 구조들을 가질 수 있다. 프로세서는 각각의 커널 매트릭스 구조를 시스틀릭 어레이(606)의 별개의 컬럼에 있는 셀에 전송할 수 있다. 예를 들어, 커널 A는 제 1 컬럼의 상부 셀로 전송될 수 있고, 커널

B는 제 2 컬럼의 상부 셀로 전송될 수 있는 식이다.

- [0053] 매트릭스 구조가 셀로 전송될 경우, 매트릭스의 제 1 엘리먼트는 1 클록 사이클 동안 셀에 저장될 수 있다. 다음 클록 사이클에서, 다음 엘리먼트가 셀에 저장될 수 있다. 저장된 제 1 엘리먼트는, 도 4를 참조하여 상술된 바와 같이 인접 셀로 시프트될 수 있다. 입력들을 시프트하는 것은 매트릭스 구조의 모든 엘리먼트들이 시스틀릭 어레이(606)에 저장될 때까지 계속될 수 있다. 활성화 입력들 및 가중치 입력들 둘 모두는 하나 또는 그 초과 클록 사이클들 이후에 각각의 셀 전체에 걸쳐 시프트될 수 있다. 시스틀릭 어레이 내부의 입력들을 시프트시키는 것이 도 7과 관련하여 아래에서 추가로 설명될 것이다.
- [0054] 도 7은 3개 클록 사이클들 이후의 예시적인 3×3 시스틀릭 어레이의 셀들 내부의 가중치 입력들을 예시하는 예(700)를 도시한다. 각각의 셀은, 도 5를 참조하여 상술된 바와 같이 가중치 입력 및 활성화 입력을 저장할 수 있다. 가중치 입력들은, 도 7과 관련하여 상술된 바와 같이, 컨벌루션 계산을 위해 시스틀릭 어레이의 별개의 컬럼들에 있는 셀들로 전송될 수 있다. 예로서, 시스템은 1, 2 및 4의 가중치 입력들을 갖는 제 1 커널 매트릭스 구조를 시스틀릭 어레이의 제 1 컬럼에 전송한다. 시스템은 3, 5 및 7의 가중치 입력들을 갖는 제 2 커널 구조를 제 2 컬럼으로 전송한다. 시스템은 6, 8 및 10의 가중치들을 갖는 제 3 커널 구조를 제 3 컬럼으로 전송한다. 모든 각각의 클록 사이클 이후, 가중치 입력들이 1 차원으로 시프트될 수 있는데, 예를 들어, 상부에서 하부로 시프트될 수 있는 반면, 활성화 입력들은 다른 차원으로 시프트될 수 있으며, 예를 들어 좌측에서 우측으로 시프트될 수 있다(예시안됨).
- [0055] 가중치 입력들이 스테거식 방식으로 셀들 내에 저장될 수 있다. 즉, 제 1 클록 사이클(702) 이후의 시스틀릭 어레이의 상태는 상부-좌측 셀 내부에 '1'을 나타낸다. '1'은 셀에 저장된 '1'의 가중치 입력을 나타낸다. 다음 클록 사이클(704)에서, '1'은 상부-좌측 셀 아래의 셀으로 시프트되고, 커널로부터의 다른 가중치 입력인 '2'가 상부-좌측 셀에 저장될뿐만 아니라, '3'의 가중치 입력이 제 2 컬럼에 있는 제일 상부 셀에 저장된다.
- [0056] 제 3 클록 사이클(706)에서, 각각의 가중치가 다시 시프트된다. 제 1 컬럼에서, 제일 하부 셀은 '1'의 가중치 입력을 저장하고, '2'의 가중치 입력이, 이전 사이클에서 '1'의 가중치 입력이 저장되었던 위치에 저장되며, '4'의 가중치 입력이 상부-제일 좌측 셀에 저장된다. 유사하게, 제 2 컬럼에서, '3'이 아래로 시프트되고 '5' 가중치 입력이 상부-중간 셀에 저장된다. 제 3 컬럼에서, '6' 가중치 입력이 상부-제일 우측 셀에 저장된다.
- [0057] 일부 구현들에서, 가중치 입력들이 시프트되어야 하는지 여부를 결정하는 가중치 입력들에 대한 제어 신호가 또한 가중치 입력들과 함께 시프트된다.
- [0058] 활성화 입력들이 유사한 방식으로 다른 차원에서 시프트될 수 있는데, 예를 들어, 좌측에서 우측으로 시프트될 수 있다.
- [0059] 일단 활성화 입력들 및 가중치 입력들이 제자리에 있으면, 프로세서는, 예를 들어, 셀들 내부의 곱셈 및 합산 회로들을 사용함으로써 컨벌루션 계산을 수행하여 벡터 컴퓨테이션 유닛에서 사용될 누산 값의 세트를 생성할 수 있다.
- [0060] 이 시스템은 가중치 입력들이 어레이의 컬럼들로 전송되고 활성화 입력들이 어레이의 로우들로 전송되는 것으로 설명되었지만, 일부 구현들에서는, 가중치 입력들이 어레이의 로우들로 전송되고 활성화 입력들이 어레이의 컬럼들로 전송된다.
- [0061] 도 8은 제어 값들이 가중치 입력들을 시프트시키거나 또는 로딩시키는 방법을 예시하는 예이다. 제어 값들(806)은, 도 3과 관련하여 상술된 바와 같이, 호스트에 의해 전송될 수 있고 가중치 시퀀서들에 의해 저장될 수 있다. 그래프 내의 값들은, 시스틀릭 어레이의 로우들 1 내지 4(804)에 각각 대응하는 가중치 시퀀서들(808-814)에 저장된 제어 값들을, 클록 사이클 단위(802)로 나타낸다.
- [0062] 일부 구현들에서, 주어진 가중치 시퀀서의 제어 값이 널-제로인 경우, 시스틀릭 어레이의 대응하는 셀에서의 가중치 입력이 인접 셀로 시프트될 것이다. 주어진 가중치 시퀀서의 제어 값이 제로인 경우, 가중치 입력은, 대응하는 셀로 로딩되고 셀의 활성화 입력과의 곱을 컴퓨팅하기 위해 사용될 수 있다.
- [0063] 예시로서, 호스트는, 4개의 가중치 입력들이 로딩되기 전에 이들이 시프트되어야 한다고 결정할 수 있다. 클록 사이클 0에서, 호스트는 5의 제어 값을 가중치 시퀀서(808), 즉, 로우 1에 대응하는 가중치 시퀀서에 전송할 수 있다. 가중치 시퀀서(808)는 감소 회로를 포함하며, 감소 회로는 1 클록 사이클을 소요하여 5의 제어 값에 기초하여 4의 제어 값을 출력한다. 따라서, 4의 제어 값이, 후속 클록 사이클, 즉 클록 사이클 1에서 가중치 시퀀서(808)에 저장된다.

- [0064] 클록 사이클 1에서, 호스트는 4의 제어 값을 가중치 시퀀서(808)로 전송한다. 따라서, 클록 사이클 2에서, 가중치 시퀀서(808)는, 예를 들어, 감소 회로를 사용하여 3의 제어 값을 저장한다. 클록 사이클 1에서, 가중치 시퀀서(808)는 4의 제어 값을 가중치 시퀀서(810)에 전송할 수 있다. 따라서, 클록 사이클 2에서, 4의 제어 값이 가중치 시퀀서(810)의 감소 회로에 의해 프로세싱된 후, 가중치 시퀀서(810)는 3의 제어 값을 저장할 수 있다.
- [0065] 유사하게, 호스트는 3, 2 및 1의 제어 값들을, 각각 클록 사이클들 2, 3 및 4에서 전송할 수 있다. 각각의 가중치 시퀀서(808-814) 내의 감소 회로는, 제어 값들이 감소할 경우, 지연을 야기하기 때문에, 각각의 클록 사이클에서 제어 값들(806)을 감소시키는 것은 결국, 각각의 가중치 시퀀서로 하여금, 동일한 제어 값, 즉 클록 사이클 4에서 1의 제어 값들을 그리고 클록 사이클 5에서 0의 제어 값들을 저장하게 할 수 있다.
- [0066] 일부 구현들에서, 각각의 가중치 시퀀서가 0의 제어 값을 출력할 경우, 시스톨릭 어레이는 가중치 입력들의 시프트를 일시정지하고 각각의 셀에 가중치 입력들을 로딩한다. 즉, 가중치 입력들을 로딩함으로써, 시스톨릭 어레이는, 가중치 입력들이 도트 곱 컴퓨테이션들에서 피연산자로서 사용될 수 있게 함으로써, 신경망의 계층을 프로세싱하기 시작한다.
- [0067] 일부 구현들에서, 컴퓨테이션이 완료된 후, 가중치들을 시프트시키는 것을 다시 시작하기 위해서, 호스트는 제어 값들을 난-제로의 수로 변경하는데, 예를 들어, 클록 사이클 7 동안 5의 제어 값을 전송한다. 시프트 프로세스는 클록 사이클 0과 관련하여 상술된 바와 같이 반복될 수 있다.
- [0068] 일부 구현들에서, 제어 값들이 다른 오프셋, 예를 들어, 1에서 시작한다.
- [0069] 본 명세서에서 설명한 기능적 동작들 및 요지의 실시예들은 디지털 전자 회로에서, 유형적으로 구현된 컴퓨터 소프트웨어 또는 펌웨어에서, 본 명세서에 개시된 구조들 및 이들의 구조적 등가물들을 포함하는 컴퓨터 하드웨어에서, 또는 이들 중 하나 또는 그 초과에 대한 결합들에서 구현될 수 있다. 본 명세서에서 설명한 요지의 실시예들은 하나 또는 그 초과에 대한 컴퓨터 프로그램들, 즉 데이터 프로세싱 장치에 의한 실행을 위해 또는 데이터 프로세싱 장치의 동작을 제어하기 위해 유형의 비일시적 프로그램 캐리어 상에 인코딩되는 컴퓨터 프로그램 명령들의 하나 또는 그 초과에 대한 모듈들로서 구현될 수 있다. 대안으로 또는 추가로, 프로그램 명령들은 데이터 프로세싱 장치에 의한 실행을 위해 적절한 수신기 장치로의 송신을 위한 정보를 인코딩하기 위해 발생하는 인공적으로 발생한 전파 신호, 예를 들어 기계 발생 전기, 광학 또는 전자기 신호에 대해 인코딩될 수 있다. 컴퓨터 저장 매체는 기계 판독 가능 저장 디바이스, 기계 판독 가능 저장 기관, 랜덤 또는 직렬 액세스 메모리 디바이스, 또는 이들 중 하나 또는 그 초과에 대한 결합일 수 있다.
- [0070] "데이터 프로세싱 장치"라는 용어는 예로서 프로그래밍 가능 프로세서, 컴퓨터 또는 다수개의 프로세서들이나 컴퓨터들을 포함하여, 데이터를 프로세싱하기 위한 모든 종류의 장치, 디바이스들 및 기계들을 포괄한다. 장치는 특수 목적 로직 회로, 예를 들어, FPGA(field programmable gate array) 또는 ASIC(application specific integrated circuit)을 포함할 수 있다. 장치는 또한 하드웨어뿐만 아니라, 해당 컴퓨터 프로그램에 대한 실행 환경을 생성하는 코드, 예를 들어 프로세서 펌웨어, 프로토콜 스택, 데이터베이스 관리 시스템, 운영 시스템, 또는 이들 중 하나 또는 그 초과에 대한 결합을 구성하는 코드를 포함할 수 있다.
- [0071] (프로그램, 소프트웨어, 소프트웨어 애플리케이션, 모듈, 소프트웨어 모듈, 스크립트 또는 코드로 또한 지칭되거나 이로서 설명될 수 있는) 컴퓨터 프로그램은 컴파일링된 또는 해석된 언어들, 또는 순수-기능적 또는 서술적 또는 절차적 언어들, 또는 임의의 형태의 프로그래밍 언어로 기록될 수 있고, 이는 독립형 프로그램으로서 또는 모듈, 컴포넌트, 서브루틴, 또는 컴퓨팅 환경에 사용하기에 적절한 다른 유닛으로서의 형태를 포함하는 임의의 형태로 전개될 수 있다. 컴퓨터 프로그램은 파일 시스템 내의 파일에 대응할 수도 있지만 반드시 그런 것은 아니다. 프로그램은 다른 프로그램들 또는 데이터, 예를 들어 마크업 언어 문서에 저장된 하나 또는 그 초과에 대한 스크립트들을 보유하는 파일의 일부에, 해당 프로그램에 전용된 단일 파일에, 또는 다수개의 조정된 파일들, 예를 들어 하나 또는 그 초과에 대한 모듈들, 하위 프로그램들, 또는 코드의 부분들을 저장하는 파일들에 저장될 수 있다. 컴퓨터 프로그램은 하나의 컴퓨터 상에서 또는 한 사이트에 로케이팅되거나 다수개의 사이트들에 걸쳐 분포되어 통신 네트워크에 의해 상호 접속되는 다수개의 컴퓨터들 상에서 실행되도록 전개될 수 있다.
- [0072] 본 명세서에서 설명한 프로세스들 및 로직 흐름들은 입력 데이터에 대해 동작하여 출력을 발생시킴으로써 기능들을 수행하기 위해 하나 또는 그 초과에 대한 컴퓨터 프로그램들을 실행하는 하나 또는 그 초과에 대한 프로그래밍 가능 컴퓨터들에 의해 수행될 수 있다. 프로세스들 및 로직 흐름들은 또한 특수 목적 로직 회로, 예를 들어 FPGA(field programmable gate array) 또는 ASIC(application specific integrated circuit)에 의해 수행될

수 있으며, 장치가 또한 이로서 구현될 수 있다.

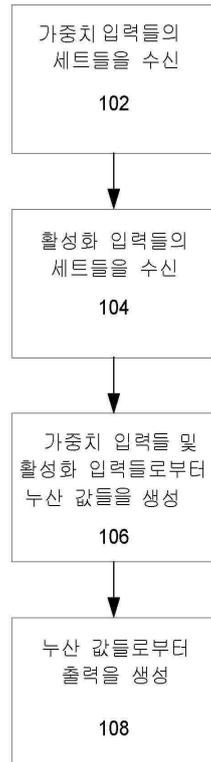
- [0073] 컴퓨터 프로그램의 실행에 적합한 컴퓨터들은 범용 또는 특수 목적 마이크로프로세서들 또는 이 둘 모두, 또는 임의의 다른 종류의 중앙 프로세싱 유닛을 포함하며, 예로서 이에 기반할 수 있다. 일반적으로, 중앙 프로세싱 유닛은 판독 전용 메모리 또는 랜덤 액세스 메모리 또는 이 둘 모두로부터 명령들 및 데이터를 수신할 것이다. 컴퓨터의 필수 엘리먼트들은 명령들을 수행 또는 실행하기 위한 중앙 프로세싱 유닛 그리고 명령들 및 데이터를 저장하기 위한 하나 또는 그 초과 메모리 디바이스들이다. 일반적으로, 컴퓨터는 또한 데이터를 저장하기 위한 하나 또는 그 초과 대용량 저장 디바이스들, 예를 들어, 자기, 마그네토 광 디스크들, 또는 광 디스크들을 포함하거나, 이들로부터 데이터를 수신하고 또는 이들에 데이터를 전송하도록, 또는 이 둘 모두를 위해 동작 가능하게 연결될 것이다. 그러나 컴퓨터가 이러한 디바이스들을 가질 필요는 없다. 더욱이, 컴퓨터는 다른 디바이스, 몇 가지만 예로 들자면, 예를 들어 모바일 전화, PDA(personal digital assistant), 모바일 오디오 또는 비디오 플레이어, 게임 콘솔, GPS(Global Positioning System) 수신기, 또는 휴대용 저장 디바이스, 예를 들어 USB(universal serial bus) 플래시 드라이브에 내장될 수 있다.
- [0074] 컴퓨터 프로그램 명령들 및 데이터를 저장하기에 적합한 컴퓨터 판독 가능 매체들은, 예로서 반도체 메모리 디바이스들, 예를 들어 EPROM, EEPROM, 및 플래시 메모리 디바이스들; 자기 디스크들, 예를 들어 내부 하드 디스크들 또는 착탈식 디스크들; 마그네토 광 디스크들; 그리고 CD ROM 및 DVD-ROM 디스크들을 포함하는 모든 형태들의 비휘발성 메모리, 매체들 및 메모리 디바이스들을 포함한다. 프로세서 및 메모리는 특수 목적 로직 회로에 의해 보완되거나 특수 목적 로직 회로에 포함될 수 있다.
- [0075] 사용자와의 상호 작용을 전송하기 위해, 본 명세서에서 설명한 요지의 실시예들은 사용자에게 정보를 디스플레이하기 위한 디스플레이 디바이스, 예를 들어 CRT(cathode ray tube) 또는 LCD(liquid crystal display) 모니터, 및 사용자가 컴퓨터에 입력을 전송할 수 있게 하는 키보드와 포인팅 디바이스, 예를 들어 마우스 또는 트랙볼을 갖는 컴퓨터 상에 구현될 수 있다. 다른 종류들의 디바이스들이 사용자와의 상호 작용을 전송하기 위해 또한 사용될 수 있는데; 예를 들어, 사용자에게 제공되는 피드백은 임의의 형태의 감각 피드백, 예를 들어 시각 피드백, 청각 피드백 또는 촉각 피드백일 수 있고; 사용자로부터의 입력은 음향, 음성 또는 촉각 입력을 포함하는 임의의 형태로 수신될 수 있다. 추가로, 컴퓨터는 사용자에게 의해 사용되는 디바이스에 문서들을 전송하고 이러한 디바이스로부터 문서들을 수신함으로써; 예를 들어, 웹 브라우저로부터 수신된 요청들에 대한 응답으로 사용자의 클라이언트 디바이스 상의 웹 브라우저에 웹 페이지들을 전송함으로써 사용자와 상호 작용할 수 있다.
- [0076] 본 명세서에서 설명한 요지의 실시예들은 예를 들어, 데이터 서버로서 백엔드 컴포넌트를 포함하는, 또는 미들웨어 컴포넌트, 예를 들어, 애플리케이션 서버를 포함하는, 또는 프론트엔드 컴포넌트, 예를 들어, 본 명세서에서 설명한 요지의 구현과 사용자가 상호 작용할 수 있게 하는 그래픽 사용자 인터페이스 또는 웹 브라우저를 갖는 클라이언트 컴퓨터를 포함하는 컴퓨팅 시스템으로, 또는 이러한 하나 또는 그 초과 백엔드, 미들웨어 또는 프론트엔드 컴포넌트들의 임의의 결합으로 구현될 수 있다. 시스템의 컴포넌트들은 임의의 형태 또는 매체의 디지털 데이터 통신, 예를 들어 통신 네트워크에 의해 상호 접속될 수 있다. 통신 네트워크들의 예들은 "LAN"(local area network) 및 "WAN"(wide area network), 예를 들어 인터넷을 포함한다.
- [0077] 컴퓨팅 시스템은 클라이언트들 및 서버들을 포함할 수 있다. 클라이언트 및 서버는 일반적으로 서로로부터 원거리이며 일반적으로 통신 네트워크를 통해 상호 작용한다. 클라이언트와 서버의 관계는 각각의 컴퓨터들 상에서 실행되며 서로 클라이언트-서버 관계를 갖는 컴퓨터 프로그램들에 의해 발생한다.
- [0078] 본 명세서는 많은 특정 구현 세부사항들을 포함하지만, 이들은 청구될 수 있는 것의 또는 임의의 발명의 범위에 대한 한정들로서가 아니라, 그보다는 특정 발명들의 특정 실시예들에 특정할 수 있는 특징들의 설명으로서 해석되어야 한다. 개별 실시예들과 관련하여 본 명세서에 설명되는 특정 특징들은 또한 단일 실시예로 결합하여 구현될 수 있다. 반대로, 단일 실시예와 관련하여 설명되는 다양한 특징들은 또한 다수개의 실시예들로 개별적으로 또는 임의의 적절한 하위 결합으로 구현될 수 있다. 더욱이, 특징들이 특정한 결합들로 작용하는 것으로 앞서 설명되고 심지어 초기에 이와 같이 청구될 수 있다 하더라도, 어떤 경우들에는 청구된 결합으로부터의 하나 또는 그 초과 특징들이 그 결합으로부터 삭제될 수 있고, 청구된 결합은 하위 결합 또는 하위 결합의 변형에 관련될 수 있다.
- [0079] 유사하게, 동작들이 특정 순서로 도면들에 도시되지만, 이는 바람직한 결과들을 달성하기 위해 이러한 동작들이 도시된 특정 순서로 또는 순차적인 순서로 수행될 것을, 또는 예시된 모든 동작들이 수행될 것을 요구하는 것으로 이해되지 않아야 한다. 특정 상황들에서는, 다중 작업 및 병렬 프로세싱이 유리할 수도 있다. 더욱이, 앞서 설명한 실시예들에서 다양한 시스템 모듈들 및 컴포넌트들의 분리는 모든 실시예들에서 이러한 분리를 필요

로 하는 것으로 이해되지 않아야 하며, 설명한 프로그램 컴포넌트들 및 시스템들은 일반적으로 단일 소프트웨어 제품으로 함께 통합되거나 다수개의 소프트웨어 제품들로 패키징될 수 있다는 것이 이해되어야 한다.

[0080] 요지의 특정 실시예들이 설명되었다. 다른 실시예들이 다음의 청구항들의 범위 내에 있다. 예를 들어, 청구항들에서 언급되는 동작들은 다른 순서로 수행되며 여전히 바람직한 결과들을 달성할 수 있다. 일례로, 첨부 도면들에 도시된 프로세스들은 바람직한 결과들을 달성하기 위해 반드시 도시된 특정 순서 또는 순차적인 순서를 필요로 하는 것은 아니다. 특정 구현들에서는, 다중 작업 및 병렬 프로세싱이 유리할 수 있다.

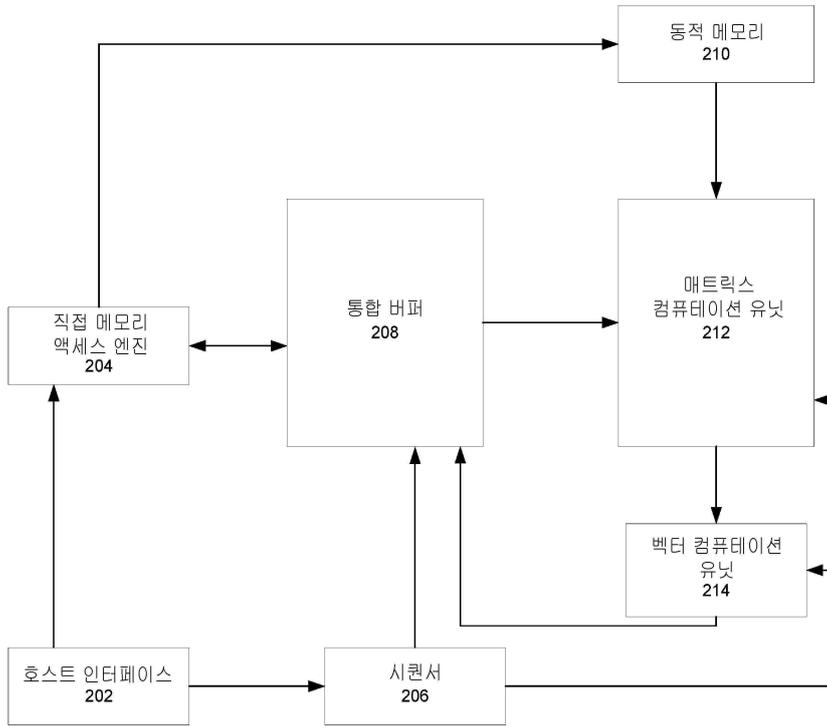
도면

도면1



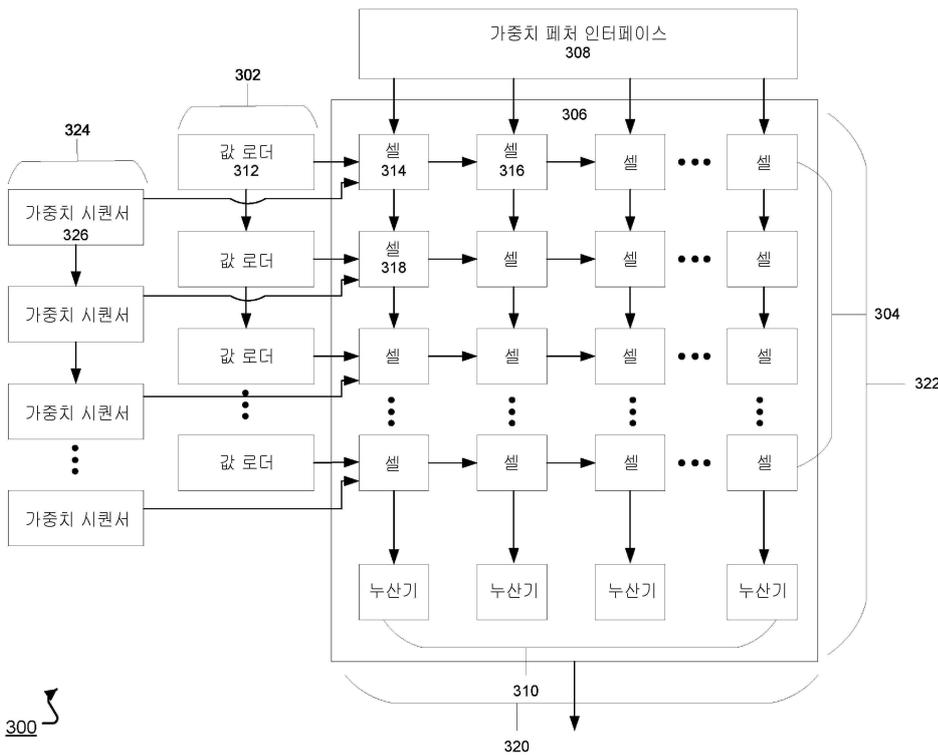
100

도면2



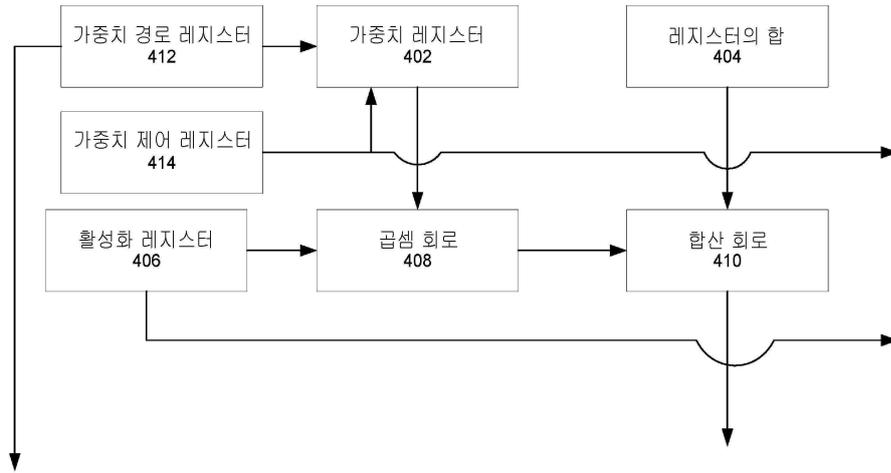
200

도면3



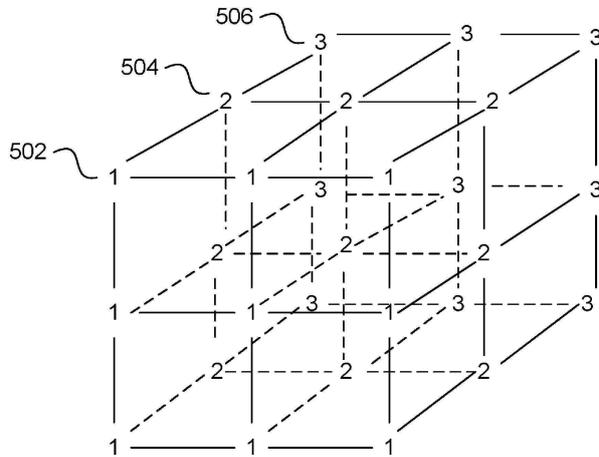
300

도면4



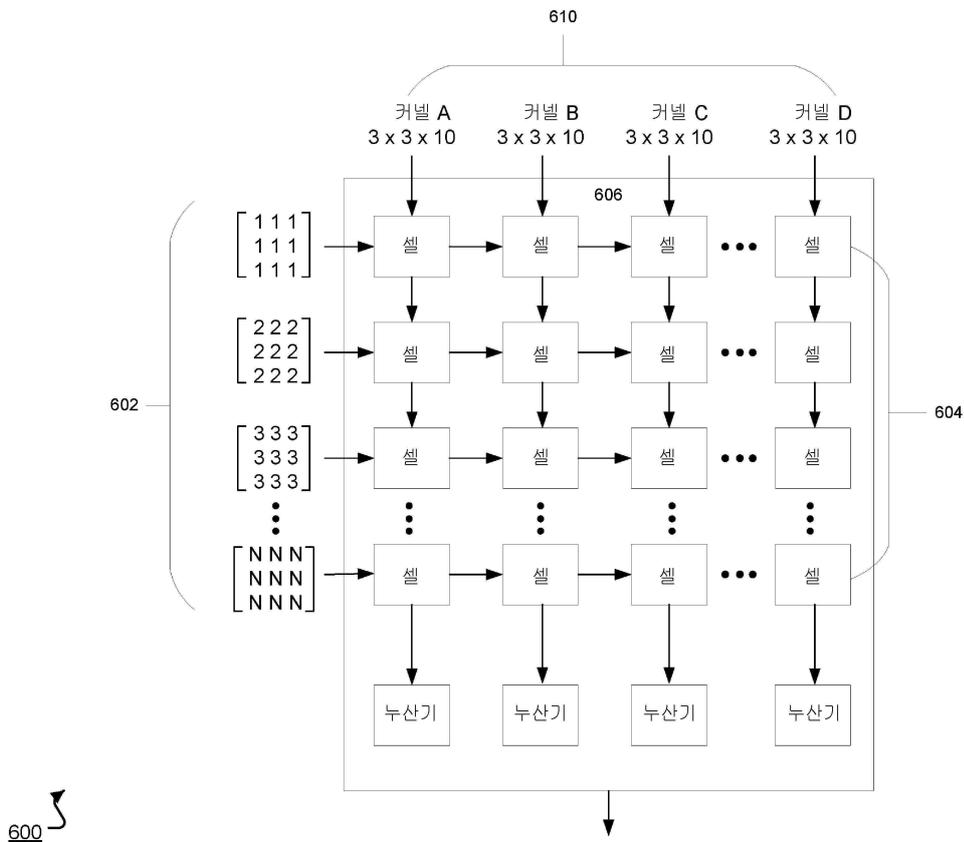
400 ↗

도면5

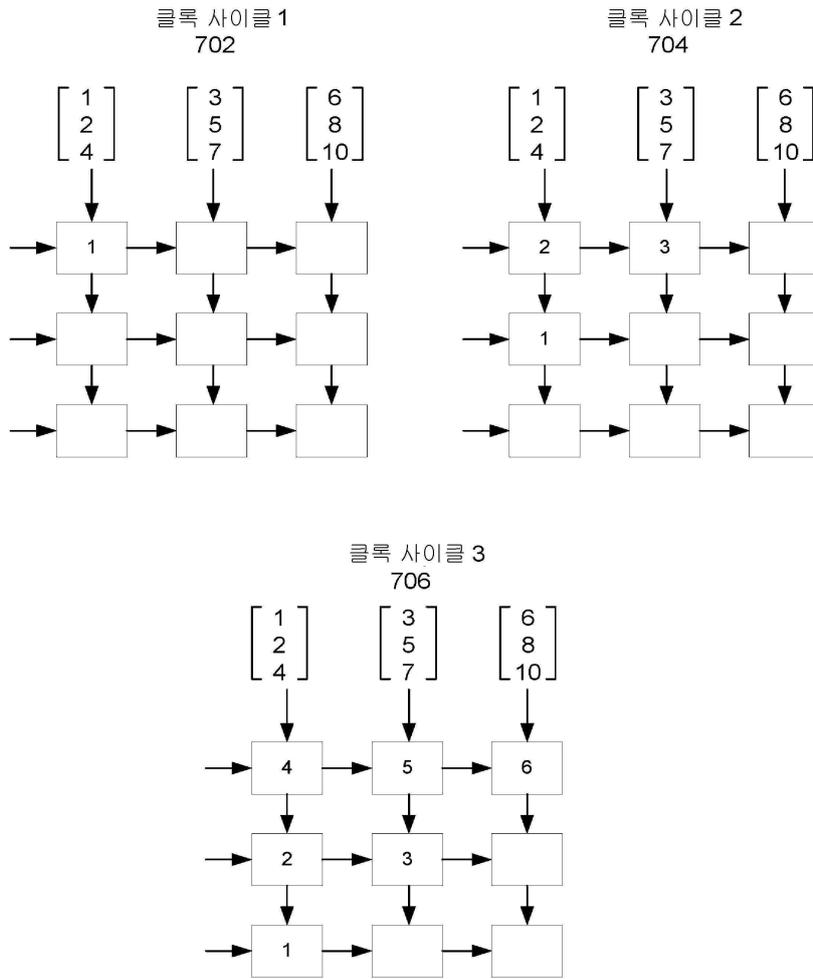


500 ↗

도면6



도면7



700 ↷

도면8

제어 806	5	4	3	2	1	1	1	5	4	
시퀀서 808	0	4	3	2	1	0	0	0	4	문우 1
시퀀서 810	0	0	3	2	1	0	0	0	0	문우 2
시퀀서 812	0	0	0	2	1	0	0	0	0	문우 3
시퀀서 814	0	0	0	0	1	0	0	0	0	문우 4
	0	1	2	3	4	5	6	7	8	9

클럭 사이클들
802

문우들
804

800 ↷