(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0144374
(43) 공개일자 2022년10월26일

(51) 국제특허분류(Int. Cl.)
G16B 30/20 (2019.01) *G06N 3/04* (2006.01)
G06N 3/063 (2006.01) *G06N 3/08* (2006.01)
G06V 10/44 (2022.01) *G06V 10/762* (2022.01)
G06V 10/764 (2022.01) *G06V 10/77* (2022.01)
G06V 10/82 (2022.01) *G16B 40/10* (2019.01)
G16B 40/20 (2019.01)
(52) CPC특허분류
G16B 30/20 (2019.02)
G06N 3/0445 (2013.01)
(21) 출원번호 10-2022-7029024
(22) 출원일자(국제) 2021년02월19일
심사청구일자 없음
(85) 번역문제출일자 2022년08월22일
(86) 국제출원번호 PCT/US2021/018915
(87) 국제공개번호 WO 2021/168358
국제공개일자 2021년08월26일
(30) 우선권주장
62/979,399 2020년02월20일 미국(US)
(뒷면에 계속)

(71) 출원인
일루미나, 인코포레이티드
미국 캘리포니아 92122 샌디에고 일루미나 웨이 5200
(72) 발명자
두타 아닌디타
미국 캘리포니아주 92122 샌디에고 일루미나 웨이 5200
베세레 게리
미국 캘리포니아주 92122 샌디에고 일루미나 웨이 5200
(뒷면에 계속)
(74) 대리인
특허법인아주김장리

전체 청구항 수 : 총 20 항

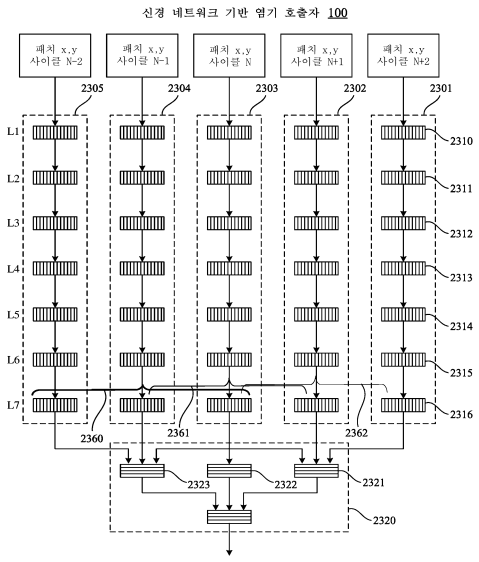
(54) 발명의 명칭 **인공 지능 기반 염기 호출자를 위한 분할 아키텍처**

(57) 요 약

개시된 기술은 공간 콘볼루션 네트워크 및 버스 네트워크를 포함하는 시스템에 관한 것이다. 공간 콘볼루션 네트워크는 각자의 공간 프로세싱 파이프라인들을 통해 각자의 사이클별 서열분석 이미지 세트들을 개별적으로 프로세싱함으로써 사이클별 서열분석 이미지 세트들의 윈도우를 사이클 단위로 프로세싱하여 각자의 서열분석 사이

(뒷면에 계속)

대 표 도 - 도23



클들 동안 각자의 사이클별 공간 특징부 맵 세트들을 생성하도록 구성된다. 버스 네트워크는 각자의 공간 프로세싱 파이프라인들 내의 공간 콘볼루션 층들 사이에 버스들을 형성하도록 구성된다. 버스들은, 특정 서열분석 사이클에 대한 공간 콘볼루션 층의 특성 시퀀스 내의 2개 이상의 공간 콘볼루션 층들에 의해 생성된 각자의 사이클별 공간 특징부 맵 세트들이 조합된 사이클별 공간 특징부 맵 세트로 조합되게 하도록, 그리고 조합된 사이클별 공간 특징부 맵 세트를 공간 콘볼루션 층의 특정 시퀀스 내의 다른 공간 콘볼루션 층에 대한 입력으로서 제공하도록 구성된다.

(52) CPC특허분류
G06N 3/0454 (2013.01)
G06N 3/063 (2013.01)
G06N 3/084 (2013.01)
G16B 40/10 (2019.02)
G16B 40/20 (2019.02)

(72) 발명자
카세프하기기 도르나
미국 캘리포니아주 92122 샌디에고 일루미나 웨이 5200
파르나비 개빈 데렉
미국 캘리포니아주 92122 샌디에고 일루미나 웨이 5200
자가나탄 키쇼르
미국 캘리포니아주 92122 샌디에고 일루미나 웨이 5200
키아 아미랄리
미국 캘리포니아주 94401 샌마테오 811 노스 엘 카미노 레알 아파트먼트 302

(30) 우선권주장
62/979,411 2020년02월20일 미국(US)
17/179,395 2021년02월18일 미국(US)
17/180,480 2021년02월19일 미국(US)
17/180,513 2021년02월19일 미국(US)

명 세 서

청구범위

청구항 1

시스템으로서,

공간 콘볼루션 층들의 각자의 시퀀스들을 통해 사이클별 서열분석 이미지 세트들의 윈도우 내의 각자의 사이클별 서열분석 이미지 세트들을 개별적으로 콘볼루션함으로써 서열분석 런의 일련의 서열분석 사이클들 동안 상기 사이클별 서열분석 이미지 세트들의 윈도우를 사이클 단위로 프로세싱하여 상기 일련의 서열분석 사이클들 중의 각자의 서열분석 사이클들 동안 각자의 사이클별 공간 특징부 맵 세트들을 생성하도록 구성된 공간 콘볼루션 네트워크

- 상기 공간 콘볼루션 층들의 각자의 시퀀스들은 공간 콘볼루션 필터 뱅크들의 각자의 시퀀스들을 갖고, 상기 공간 콘볼루션 필터 뱅크들의 각자의 시퀀스들의 공간 콘볼루션 필터 뱅크들 내의 공간 콘볼루션 필터들의 트레이닝된 계수들은 상기 공간 콘볼루션 층들의 각자의 시퀀스들 내의 공간 콘볼루션 층들의 시퀀스들 사이에서 변함 -; 및

제1 시간 콘볼루션 층의 각자의 시간 콘볼루션 필터 뱅크들을 사용하여 상기 사이클별 공간 특징부 맵 세트들 내의 사이클별 공간 특징부 맵 세트들의 각자의 중첩 그룹들을 콘볼루션함으로써 그룹별 단위로 상기 사이클별 공간 특징부 맵 세트들을 프로세싱하여 상기 사이클별 공간 특징부 맵 세트들의 각자의 중첩 그룹들에 대한 각자의 그룹별 시간 특징부 맵 세트들을 생성하도록 구성된 시간 콘볼루션 네트워크

- 상기 각자의 시간 콘볼루션 필터 뱅크들 내의 시간 콘볼루션 필터들의 트레이닝된 계수들은 상기 각자의 시간 콘볼루션 필터 뱅크들 중의 시간 콘볼루션 필터 뱅크들 사이에서 변함 - 를 포함하는, 시스템.

청구항 2

제1항에 있어서, 상기 공간 콘볼루션 필터들은 사이클내(intra-cycle) 분리형 콘볼루션들을 사용하는, 시스템.

청구항 3

제1항에 있어서, 상기 시간 콘볼루션 필터들은 사이클간(inter-cycle) 조합형 콘볼루션들을 사용하는, 시스템.

청구항 4

제1항에 있어서, 각자의 압축 콘볼루션 층들을 통해 상기 각자의 사이클별 공간 특징부 맵 세트들을 개별적으로 콘볼루션하여 상기 각자의 서열분석 사이클들 동안 각자의 사이클별 압축된 공간 특징부 맵 세트들을 생성하는 압축 네트워크를 포함하도록 추가로 구성되는, 시스템.

청구항 5

제4항에 있어서, 상기 각자의 압축 콘볼루션 층들 내의 압축 콘볼루션 필터들의 트레이닝된 계수들은 상기 각자의 압축 콘볼루션 층들 중의 압축 콘볼루션 층들 사이에서 변하는, 시스템.

청구항 6

제5항에 있어서, 상기 시간 콘볼루션 네트워크는 제2 시간 콘볼루션 층의 각자의 시간 콘볼루션 필터 뱅크들을 사용하여 상기 그룹별 시간 특징부 맵 세트들 중의 그룹별 시간 특징부 맵 세트들의 각자의 중첩 그룹들을 콘볼루션함으로써 상기 그룹별 시간 특징부 맵 세트들을 그룹별 단위로 프로세싱하여 상기 그룹별 시간 특징부 맵 세트들의 각자의 중첩 그룹들에 대한 각자의 추가 그룹별 시간 특징부 맵 세트들을 생성하도록 추가로 구성되는, 시스템.

청구항 7

제6항에 있어서, 최종 시간 콘볼루션 층에 의해 생성된 최종 시간 특징부 맵 세트를 프로세싱하여 최종 출력을

생성하는 출력 네트워크를 포함하도록 추가로 구성되는, 시스템.

청구항 8

제7항에 있어서, 상기 최종 출력에 기초하여 상기 일련의 서열분석 사이클들 중의 하나 이상의 서열분석 사이클들 동안 염기 호출 예측들을 생성하도록 추가로 구성되는, 시스템.

청구항 9

시스템으로서,

공간 콘볼루션 층들의 각자의 시퀀스들을 통해 사이클별 서열분석 이미지 세트들의 윈도우 내의 각자의 사이클별 서열분석 이미지 세트들을 개별적으로 콘볼루션함으로써 서열분석 런의 일련의 서열분석 사이클들 동안 상기 사이클별 서열분석 이미지 세트들의 윈도우를 사이클 단위로 프로세싱하여 상기 일련의 서열분석 사이클들 중의 각자의 서열분석 사이클들 동안 각자의 사이클별 공간 특징부 맵 세트들을 생성하도록 구성된 공간 콘볼루션 네트워크; 및

각자의 시간 콘볼루션 필터 뱅크들을 사용하여 상기 사이클별 공간 특징부 맵 세트들 내의 사이클별 공간 특징부 맵 세트들의 각자의 중첩 그룹들을 콘볼루션함으로써 그룹별 단위로 상기 사이클별 공간 특징부 맵 세트들을 프로세싱하여 상기 사이클별 공간 특징부 맵 세트들의 각자의 중첩 그룹들에 대한 각자의 그룹별 시간 특징부 맵 세트들을 생성하도록 구성된 시간 콘볼루션 네트워크를 포함하고,

상기 각자의 시간 콘볼루션 필터 뱅크들 내의 시간 콘볼루션 필터들의 트레이닝된 계수들은 상기 각자의 시간 콘볼루션 필터 뱅크들 중의 시간 콘볼루션 필터 뱅크들 사이에서 변하는, 시스템.

청구항 10

제9항에 있어서, 상기 공간 콘볼루션 층들의 각자의 시퀀스들은 공간 콘볼루션 필터 뱅크들의 각자의 시퀀스들을 갖고, 상기 공간 콘볼루션 필터 뱅크들의 각자의 시퀀스들의 공간 콘볼루션 필터 뱅크들 내의 공간 콘볼루션 필터들의 트레이닝된 계수들은 상기 공간 콘볼루션 층들의 각자의 시퀀스들 내의 공간 콘볼루션 층들의 시퀀스들 사이에서 공유되는, 시스템.

청구항 11

제9항에 있어서, 각자의 압축 콘볼루션 층들을 통해 상기 각자의 사이클별 공간 특징부 맵 세트들을 개별적으로 콘볼루션하여 상기 각자의 서열분석 사이클들 동안 각자의 사이클별 압축된 공간 특징부 맵 세트들을 생성하는 압축 네트워크를 포함하도록 추가로 구성되고, 상기 각자의 압축 콘볼루션 층들 내의 압축 콘볼루션 필터들의 트레이닝된 계수들은 상기 각자의 압축 콘볼루션 층들 중의 압축 콘볼루션 층들 사이에서 변하는, 시스템.

청구항 12

염기 호출의 인공 지능 기반 방법으로서,

공간 콘볼루션 네트워크를 통해, 공간 콘볼루션 층들의 각자의 시퀀스들을 통해 사이클별 서열분석 이미지 세트들의 윈도우 내의 각자의 사이클별 서열분석 이미지 세트들을 개별적으로 콘볼루션함으로써 서열분석 런의 일련의 서열분석 사이클들 동안 상기 사이클별 서열분석 이미지 세트들의 윈도우를 사이클 단위로 프로세싱하고 상기 일련의 서열분석 사이클들 중의 각자의 서열분석 사이클들 동안 각자의 사이클별 공간 특징부 맵 세트들을 생성하는 단계

- 상기 공간 콘볼루션 층들의 각자의 시퀀스들은 공간 콘볼루션 필터 뱅크들의 각자의 시퀀스들을 갖고, 상기 공간 콘볼루션 필터 뱅크들의 각자의 시퀀스들의 공간 콘볼루션 필터 뱅크들 내의 공간 콘볼루션 필터들의 트레이닝된 계수들은 상기 공간 콘볼루션 층들의 각자의 시퀀스들 내의 공간 콘볼루션 층들의 시퀀스들 사이에서 변함 -; 및

시간 콘볼루션 네트워크를 통해, 제1 시간 콘볼루션 층의 각자의 시간 콘볼루션 필터 뱅크들을 사용하여 상기 사이클별 공간 특징부 맵 세트들 내의 사이클별 공간 특징부 맵 세트들의 각자의 중첩 그룹들을 콘볼루션함으로써 그룹별 단위로 상기 사이클별 공간 특징부 맵 세트들을 프로세싱하고 상기 사이클별 공간 특징부 맵 세트들의 각자의 중첩 그룹들에 대한 각자의 그룹별 시간 특징부 맵 세트들을 생성하는 단계

- 상기 각자의 시간 콘볼루션 필터 뱅크들 내의 시간 콘볼루션 필터들의 트레이닝된 계수들은 상기 각자의 시간 콘볼루션 필터 뱅크들 중의 시간 콘볼루션 필터 뱅크들 사이에서 변함 - 를 포함하는, 인공 지능 기반 방법.

청구항 13

제12항에 있어서, 압축 네트워크의 각자의 압축 콘볼루션 층들을 통해 상기 각자의 사이클별 공간 특징부 맵 세트들을 개별적으로 콘볼루션하고, 상기 각자의 서열분석 사이클들 동안 각자의 사이클별 압축된 공간 특징부 맵 세트들을 생성하는 단계를 추가로 포함하는, 인공 지능 기반 방법.

청구항 14

제13항에 있어서, 상기 각자의 압축 콘볼루션 층들 내의 압축 콘볼루션 필터들의 트레이닝된 계수들은 상기 각자의 압축 콘볼루션 층들 중의 압축 콘볼루션 층들 사이에서 변하는, 인공 지능 기반 방법.

청구항 15

제14항에 있어서, 상기 시간 콘볼루션 네트워크를 통해, 제2 시간 콘볼루션 층의 각자의 시간 콘볼루션 필터 뱅크들을 사용하여 상기 그룹별 시간 특징부 맵 세트들 중의 그룹별 시간 특징부 맵 세트들의 각자의 중첩 그룹들을 콘볼루션함으로써 상기 그룹별 시간 특징부 맵 세트들을 그룹별 단위로 프로세싱하고 상기 그룹별 시간 특징부 맵 세트들의 각자의 중첩 그룹들에 대한 각자의 추가 그룹별 시간 특징부 맵 세트들을 생성하는 단계를 추가로 포함하는, 인공 지능 기반 방법.

청구항 16

제15항에 있어서, 출력 네트워크를 통해, 최종 시간 콘볼루션 층에 의해 생성된 최종 시간 특징부 맵 세트를 프로세싱하고 최종 출력을 생성하는 단계를 추가로 포함하는, 인공 지능 기반 방법.

청구항 17

제16항에 있어서, 상기 최종 출력에 기초하여 상기 일련의 서열분석 사이클들 중의 하나 이상의 서열분석 사이클들 동안 염기 호출 예측들을 생성하는 단계를 추가로 포함하는, 인공 지능 기반 방법.

청구항 18

시스템으로서,

사이클별 서열분석 이미지들의 윈도우에서 공간 콘볼루션 층들의 각자의 시퀀스들을 각자의 사이클별 서열분석 이미지들에 적용하도록 구성된 공간 콘볼루션 네트워크를 포함하고,

상기 공간 콘볼루션 층들의 각자의 시퀀스들은 공간 콘볼루션 층들의 시퀀스마다 상이한 공간 콘볼루션 필터 뱅크들의 각자의 시퀀스들을 갖는, 시스템.

청구항 19

시스템으로서,

시간 콘볼루션 필터들의 각자의 세트들을 공간 특징부 맵들의 각자의 슬라이딩 윈도우들에 적용하도록 구성된 제1 시간 콘볼루션 층으로 구성된 시간 콘볼루션 네트워크를 포함하고,

상기 제1 시간 콘볼루션 층 내의 상기 시간 콘볼루션 필터들의 각자의 세트들은 시간 콘볼루션 필터들의 세트마다 상이한 시간 콘볼루션 필터들을 갖는, 시스템.

청구항 20

제19항에 있어서, 상기 시간 콘볼루션 네트워크는 상기 제1 시간 콘볼루션 층에 후행하는 제2 시간 콘볼루션 층으로 구성되고, 상기 제2 콘볼루션 층은 시간 콘볼루션 필터들의 각자의 세트들을 시간 특징부 맵들의 각자의 슬라이딩 윈도우들에 적용하도록 구성되고, 상기 제2 시간 콘볼루션 층 내의 상기 시간 콘볼루션 필터들의 각자의 세트들은 시간 콘볼루션 필터들의 세트마다 상이한 시간 콘볼루션 필터들을 갖는, 시스템.

발명의 설명

기 술 분 야

[0001] 개시된 기술은 지능의 에뮬레이션을 위한; 그리고 불확실성 추리 시스템들(예컨대, 퍼지 논리 시스템들), 적응적 시스템들, 기계 학습 시스템들, 및 인공 신경 네트워크들을 포함하는, 인공 지능 유형 컴퓨터들 및 디지털 데이터 프로세싱 시스템들, 및 대응하는 데이터 프로세싱 방법들 및 제품들(즉, 지식 기반 시스템들, 추리 시스템들, 및 지식 획득 시스템들)에 관한 것이다. 특히, 개시된 기술은 데이터를 분석하기 위한 심층 콘볼루션(deep convolutional) 신경 네트워크들과 같은 심층 신경 네트워크들을 사용하는 것에 관한 것이다.

[0002] 우선권 출원

[0003] 본 PCT 출원은 발명의 명칭이 "DATA COMPRESSION FOR ARTIFICIAL INTELLIGENCE-BASED BASE CALLING"이고 2020년 2월 20일자로 출원된 미국 가특허 출원 제62/979,411호(대리인 문서 번호 ILLM 1029-1/IP-1964-PRV) 및 발명의 명칭이 "DATA COMPRESSION FOR ARTIFICIAL INTELLIGENCE-BASED BASE CALLING"이고 2021년 2월 18일자로 출원된 미국 특허 출원 제17/179,395호(대리인 문서 번호 ILLM 1029-2/IP-1964-US)에 대한 우선권 및 그의 이익을 주장한다. 우선권 출원들은 이로써 모든 목적들을 위해 본 명세서에 완전히 설명된 것처럼 참고로 포함된다.

[0004] 본 PCT 출원은 발명의 명칭이 "SQUEEZING LAYER FOR ARTIFICIAL INTELLIGENCE-BASED BASE CALLING"이고 2020년 2월 20일자로 출원된 미국 가특허 출원 제62/979,399호(대리인 문서 번호 ILLM 1030-1/IP-1982-PRV) 및 발명의 명칭이 "SPLIT ARCHITECTURE FOR ARTIFICIAL INTELLIGENCE-BASED BASE CALLER"이고 2021년 2월 19일자로 출원된 미국 특허 출원 제17/180,480호(대리인 문서 번호 ILLM 1030-2/IP-1982-US)에 대한 우선권 및 그의 이익을 주장한다. 우선권 출원들은 이로써 모든 목적들을 위해 본 명세서에 완전히 설명된 것처럼 참고로 포함된다.

[0005] 본 PCT 출원은 발명의 명칭이 "NETWORK FOR ARTIFICIAL INTELLIGENCE-BASED BASE CALLER"이고 2021년 2월 19일자로 출원된 미국 특허 출원 제17/180,513호(대리인 문서 번호 1031-2/IP-1965-US)에 대한 우선권 및 그의 이익을 주장한다. 우선권 출원은 이로써 모든 목적들을 위해 본 명세서에 완전히 설명된 것처럼 참고로 포함된다.

[0006] **참조 문헌들**

[0007] 하기는 본 명세서에 완전히 설명된 것처럼 참고로 포함된다:

[0008] 발명의 명칭이 "ARTIFICIAL INTELLIGENCE-BASED BASE CALLING OF INDEX SEQUENCES"이고 2020년 2월 20일자로 출원된 미국 가특허 출원 제62/979,384호(대리인 문서 번호 ILLM 1015-1/IP-1857-PRV);

[0009] 발명의 명칭이 "ARTIFICIAL INTELLIGENCE-BASED MANY-TO-MANY BASE CALLING"이고 2020년 2월 20일자로 출원된 미국 가특허 출원 제62/979,414호(대리인 문서 번호 ILLM 1016-1/IP-1858-PRV);

[0010] 발명의 명칭이 "KNOWLEDGE DISTILLATION-BASED COMPRESSION OF ARTIFICIAL INTELLIGENCE-BASED BASE CALLER"이고 2020년 2월 20일자로 출원된 미국 가특허 출원 제62/979,385호(대리인 문서 번호 ILLM 1017-1/IP-1859-PRV);

[0011] 발명의 명칭이 "DETECTING AND FILTERING CLUSTERS BASED ON ARTIFICIAL INTELLIGENCE-PREDICTED BASE CALLS"이고 2020년 8월 28일자로 출원된 미국 가특허 출원 제63/072,032호(대리인 문서 번호 ILLM 1018-1/IP-1860-PRV);

[0012] 발명의 명칭이 "MULTI-CYCLE CLUSTER BASED REAL TIME ANALYSIS SYSTEM"이고 2020년 2월 20일자로 출원된 미국 가특허 출원 제62/979,412호(대리인 문서 번호 ILLM 1020-1/IP-1866-PRV);

[0013] 발명의 명칭이 "TRAINING DATA GENERATION FOR ARTIFICIAL INTELLIGENCE-BASED SEQUENCING"이고 2020년 3월 20일자로 출원된 미국 가특허 출원 제16/825,987호(대리인 문서 번호 ILLM 1008-16/IP-1693-US);

[0014] 발명의 명칭이 "ARTIFICIAL INTELLIGENCE-BASED GENERATION OF SEQUENCING METADATA"이고 2020년 3월 20일자로 출원된 미국 특허 출원 제16/825,991호(대리인 문서 번호 ILLM 1008-17/IP-1741-US);

[0015] 발명의 명칭이 "ARTIFICIAL INTELLIGENCE-BASED BASE CALLING"이고 2020년 3월 20일자로 출원된 미국 특허 출원 제16/826,126호(대리인 문서 번호 ILLM 1008-18/IP-1744-US);

[0016] 발명의 명칭이 "ARTIFICIAL INTELLIGENCE-BASED QUALITY SCORING"이고 2020년 3월 20일자로 출원된 미국 특허 출원 제16/826,134호(대리인 문서 번호 ILLM 1008-19/IP-1747-US); 및

[0017] 발명의 명칭이 "ARTIFICIAL INTELLIGENCE-BASED SEQUENCING"이고 2020년 3월 21일자로 출원된 미국 특허 출원 제16/826,168호(대리인 문서 번호 ILLM 1008-20/IP-1752-PRV-US).

배 경 기 술

[0018] 이 섹션에서 논의되는 주제는 단지 이 섹션 내에서의 그의 언급의 결과로서 종래기술이라고 가정되어서는 안 된다. 유사하게, 이 섹션에서 언급되거나 배경기술로서 제공되는 주제와 연관된 문제는 종래기술에서 이전에 인식되었다고 가정되어서는 안 된다. 이 섹션에서의 주제는 단지 상이한 접근법들을 표현할 뿐이며, 그 접근법들 자체는 청구되는 기술의 구현들에 또한 대응할 수 있다.

[0019] 계산 능력에서의 신속한 개선은 유의하게 개선된 정확도를 갖는 많은 컴퓨터 비전 태스크들에 대해 최근 수년간 심층 콘볼루션 신경 네트워크(convolution neural network, CNN)들을 큰 성공으로 만들었다. 추론 페이즈 동안, 많은 애플리케이션들은 그래픽 프로세싱 유닛(graphics processing unit, GPU) 및 다른 범용 플랫폼의 효율을 감소시키는 엄격한 전력 소비 요건을 갖는 하나의 이미지의 낮은 레이턴시 프로세싱을 요구하여, 심층 학습 알고리즘 추론에 대해 특정적인 디지털 회로를 커스텀화함으로써 특정 가속화 하드웨어, 예컨대 필드 프로그래밍가능 게이트 어레이(field programmable gate array, FPGA)에 기회들을 가져온다. 그러나, 휴대용 및 임베디드 시스템들에 CNN들을 배치하는 것은 큰 데이터 볼륨, 집중적인 연산, 다양한 알고리즘 구조들, 및 빈번한 메모리 액세스들로 인해 여전히 도전적이다.

[0020] 콘볼루션이 CNN들에서 대부분의 연산들에 기여함에 따라, 콘볼루션 가속화 스킴은 하드웨어 CNN 가속기의 효율 및 성능에 유의하게 영향을 미친다. 콘볼루션은 커널 및 특징부 맵(feature map)들을 따라 슬라이딩하는 4개의 레벨들의 루프들과의 승산 누적(multiply and accumulate, MAC) 연산들을 수반한다. 제1 루프 레벨은 커널 윈도우 내의 픽셀들의 MAC를 연산한다. 제2 루프 레벨은 상이한 입력 특징부 맵들에 걸쳐 MAC의 곱셈들의 합을 누산한다. 제1 및 제2 루프 레벨들을 완료한 후, 바이어스를 추가함으로써 최종 출력 픽셀이 얻어진다. 제3 루프 레벨은 입력 특징부 맵 내에서 커널 윈도우를 슬라이딩한다. 제4 루프 레벨은 상이한 출력 특징부 맵들을 생성한다.

[0021] FPGA들은 그들의 (1) 고도의 재구성가능성, (2) CNN들의 빠른 진화를 따라잡기 위한, 주문형 반도체(application specific integrated circuit, ASIC)들에 비해 더 빠른 개발 시간, (3) 양호한 성능, 및 (4) GPU들에 비해 우수한 에너지 효율로 인해, 특히 추론 태스크들을 가속화하기 위해 증가하는 관심 및 인기를 얻어왔다. FPGA의 높은 성능 및 효율은 특정 연산을 위해 커스텀화되는 회로를 합성하여 커스텀화된 메모리 시스템들로 수십억 회의 연산들을 직접 프로세싱함으로써 실현될 수 있다. 예를 들어, 최신 FPGA들 상의 수백 내지 수천 개의 디지털 신호 프로세싱(digital signal processing, DSP) 블록들은 높은 병렬성과 함께 코어 콘볼루션 연산, 예컨대, 곱셈 및 덧셈을 지원한다. 외부 온-칩 메모리와 온-칩 프로세싱 엔진(processing engine, PE)들 사이의 전용 데이터 버퍼들은, FPGA 칩 상에 수십 Mbyte 온-칩 블록 랜덤 액세스 메모리(block random access memory, BRAM)들을 구성함으로써 바람직한 데이터 흐름을 실현하도록 설계될 수 있다.

[0022] 높은 성능을 달성하기 위해 리소스 활용을 최대화하면서 데이터 통신을 최소화하기 위해, CNN 가속화의 효율적인 데이터 흐름 및 하드웨어 아키텍처가 요구된다. 높은 성능, 효율 및 유연성을 갖는 가속화 하드웨어 상에서 다양한 CNN 알고리즘들의 추론 프로세스를 가속화하기 위한 방법론 및 프레임워크를 설계할 기회가 발생한다.

도면의 간단한 설명

[0023] 도면에서, 유사한 도면 부호는, 대체로, 상이한 도면들 전체에 걸쳐서 유사한 부분들을 지칭한다. 또한, 도면은 반드시 축척대로인 것은 아니며, 그 대신, 대체적으로, 개시된 기술의 원리들을 예시할 시에 강조된다. 하기의 설명에서, 개시된 기술의 다양한 구현예들이 하기의 도면을 참조하여 기술된다.

도 1a는 염기 호출의 제1 반복 동안 압축된 공간 맵 세트들을 생성하는 개시된 압축 로직의 하나의 구현예를 도시한다.

도 1b는 개시된 시간 로직을 통해 압축된 공간 맵 세트들을 프로세싱하여 시간 맵 세트들을 생성하는 하나의 구현예를 도시한다.

도 1c는 개시된 출력 로직을 통해 시간 맵 세트들을 프로세싱하여 염기 호출 분류 데이터를 생성하는 하나의 구현예를 도시한다.

도 1d는 대상 서열분석 사이클 동안 사이클별 이미지 패치(patch)들을 프로세싱한 것에 응답하여 공간 콘볼루션

층들의 캐스케이드에 의해 연속적으로 생성된 특징부 맵 볼륨들의 시퀀스의 일례를 도시한다.

도 1e는 1x1 콘볼루션들이 특징부 맵들을 어떻게 압축하는지를 예시하는 일례를 도시한다.

도 1f는 개시된 압축 로직에 의해 달성되는 압축비가 개시된 압축 로직에 의해 적용된 압축 필터들의 개수의 함수임을 도시한다.

도 1g는 예시적인 소프트맥스(softmax) 함수를 도시한다.

도 1h는 개시된 기술에 의해 생성된 예시적인 클러스터별, 사이클별 확률 쿼드러플들을 도시한다.

도 2a는 염기 호출의 제2 반복 동안, 공간 맵들 및 대응하는 압축된 공간 맵들이 비중첩 서열분석 사이클 6 동안만 생성되는 것을 도시한다.

도 2b는 염기 호출의 제1 반복 동안 생성된 압축된 공간 맵 세트들이 염기 호출의 제2 반복 동안 생성된 압축된 공간 맵 세트와 함께 사용되어 중앙 서열분석 사이클 4 동안 염기 호출들을 생성하는 것을 도시한다.

도 2c는 출력 층이 염기 호출의 제2 반복 동안 생성된 최종 시간 맵 세트를 프로세싱하고 중앙 서열분석 사이클 4 동안 염기 호출들을 생성하는 것을 도시한다.

도 3a는 염기 호출의 제3 반복 동안, 공간 맵들 및 대응하는 압축된 공간 맵들이 비중첩 서열분석 사이클 7 동안만 생성되는 것을 도시한다.

도 3b는 염기 호출의 제1 반복 및 제2 반복 동안 생성된 압축된 공간 맵 세트들이 염기 호출의 제3 반복 동안 생성된 압축된 공간 맵 세트와 함께 사용되어 중앙 서열분석 사이클 5 동안 염기 호출들을 생성하는 것을 도시한다.

도 3c는 출력 층이 염기 호출의 제3 반복 동안 생성된 최종 시간 맵 세트를 프로세싱하고 중앙 서열분석 사이클 5 동안 염기 호출들을 생성하는 것을 도시한다.

도 4a는 중앙 서열분석 사이클 16을 염기 호출하기 위한 염기 호출의 제14 반복을 도시한다.

도 4b는 서열분석 사이클들 1 내지 29 동안 이전에 생성된 압축된 공간 맵들이 중앙 서열분석 사이클 16을 염기 호출하기 위한 최종 시간 맵 세트를 생성하는 데 사용된다는 것을 도시한다.

도 4c는 출력 층이 염기 호출의 제14 반복 동안 생성된 최종 시간 맵 세트를 프로세싱하고 중앙 서열분석 사이클 16 동안 염기 호출들을 생성하는 것을 도시한다.

도 5a는 필터링 로직을 사용하여 각자의 서열분석 사이클들 1, 2, 3, 4 및 5 동안 압축된 공간 맵 세트들을 필터링하여, 염기 호출의 제1 반복 동안 각자의 압축된, 필터링된 공간 맵들을 생성하는 하나의 구현예를 도시한다.

도 5b는 출력 층이 염기 호출의 제1 반복 동안 생성된 최종 필터링된 시간 맵 세트를 프로세싱하고 중앙 서열분석 사이클 3 동안 염기 호출들을 생성하는 것을 도시한다.

도 6a는 필터링 로직을 사용하여 각자의 서열분석 사이클들 2, 3, 4, 5 및 6 동안 압축된 공간 맵 세트들을 필터링하여, 염기 호출의 제2 반복 동안 각자의 압축된, 필터링된 공간 맵들을 생성하는 하나의 구현예를 도시한다.

도 6b는 출력 층이 염기 호출의 제2 반복 동안 생성된 최종 필터링된 시간 맵 세트를 프로세싱하고 중앙 서열분석 사이클 4 동안 염기 호출들을 생성하는 것을 도시한다.

도 7a는 필터링 로직을 사용하여 각자의 서열분석 사이클들 3, 4, 5, 6 및 7 동안 압축된 공간 맵 세트들을 필터링하여, 염기 호출의 제3 반복 동안 각자의 압축된, 필터링된 공간 맵들을 생성하는 하나의 구현예를 도시한다.

도 7b는 출력 층이 염기 호출의 제3 반복 동안 생성된 최종 필터링된 시간 맵 세트를 프로세싱하고 중앙 서열분석 사이클 5 동안 염기 호출들을 생성하는 것을 도시한다.

도 8a는 압축 로직을 통해 염기 호출의 제1 반복 동안 생성된 시간 특징부 맵들의 세트들을 프로세싱하여, 압축된 시간 특징부 맵들의 각자의 세트들을 생성하는 하나의 구현예를 도시한다.

도 8b는 출력 층이 염기 호출의 제1 반복 동안 생성된 최종 압축된 시간 맵 세트를 프로세싱하고 중앙 서열분석

사이클 3 동안 염기 호출들을 생성하는 것을 도시한다.

도 9a는 염기 호출의 제1 반복 동안 생성된 압축된 시간 맵 세트들이 염기 호출의 제2 반복 동안 생성된 압축된 시간 맵 세트와 함께 사용되어 중앙 서열분석 사이클 4 동안 염기 호출들을 생성하는 것을 도시한다.

도 9b는 출력 층이 염기 호출의 제2 반복 동안 생성된 최종 압축된 시간 맵 세트를 프로세싱하고 중앙 서열분석 사이클 4 동안 염기 호출들을 생성하는 것을 도시한다.

도 10a는 염기 호출의 제1 반복 및 제2 반복 동안 생성된 압축된 시간 맵 세트들이 염기 호출의 제3 반복 동안 생성된 압축된 시간 맵 세트와 함께 사용되어 중앙 서열분석 사이클 5 동안 염기 호출들을 생성하는 것을 도시한다.

도 10b는 출력 층이 염기 호출의 제3 반복 동안 생성된 최종 압축된 시간 맵 세트를 프로세싱하고 중앙 서열분석 사이클 5 동안 염기 호출들을 생성하는 것을 도시한다.

도 11a는 압축 로직을 통해 염기 호출의 제1 반복 동안 생성된 필터링된 시간 특징부 맵들의 세트들을 프로세싱하여, 압축된, 필터링된 시간 특징부 맵들의 각자의 세트들을 생성하는 하나의 구현예를 도시한다.

도 11b는 출력 층이 염기 호출의 제1 반복 동안 생성된 최종 압축된, 필터링된 시간 맵 세트를 프로세싱하고 중앙 서열분석 사이클 3 동안 염기 호출들을 생성하는 것을 도시한다.

도 12a는 염기 호출의 제1 반복 동안 생성된 압축된, 필터링된 시간 맵 세트들이 염기 호출의 제2 반복 동안 생성된 압축된, 필터링된 시간 맵 세트와 함께 사용되어 중앙 서열분석 사이클 4 동안 염기 호출들을 생성하는 것을 도시한다.

도 12b는 출력 층이 염기 호출의 제2 반복 동안 생성된 최종 압축된, 필터링된 시간 맵 세트를 프로세싱하고 중앙 서열분석 사이클 4 동안 염기 호출들을 생성하는 것을 도시한다.

도 13a는 염기 호출의 제1 반복 및 제2 반복 동안 생성된 압축된, 필터링된 시간 맵 세트들이 염기 호출의 제3 반복 동안 생성된 압축된, 필터링된 시간 맵 세트와 함께 사용되어 중앙 서열분석 사이클 5 동안 염기 호출들을 생성하는 것을 도시한다.

도 13b는 출력 층이 염기 호출의 제3 반복 동안 생성된 최종 압축된, 필터링된 시간 맵 세트를 프로세싱하고 중앙 서열분석 사이클 5 동안 염기 호출들을 생성하는 것을 도시한다.

도 14는 본 명세서에 개시된 신경 네트워크 기반 염기 호출자의 제1 예시적인 아키텍처를 도시한다.

도 15는 본 명세서에 개시된 신경 네트워크 기반 염기 호출자의 제2 예시적인 아키텍처를 도시한다.

도 16은 본 명세서에 개시된 신경 네트워크 기반 염기 호출자의 제3 예시적인 아키텍처를 도시한다.

도 17은 본 명세서에 개시된 신경 네트워크 기반 염기 호출자의 제4 예시적인 아키텍처를 도시한다.

도 18은 입력 데이터 내의 채널들의 개수에 따라 압축 층에 콘볼루션 필터들의 카운트(또는 수)를 구성하는 필터 구성 로직의 하나의 구현예를 도시한다.

도 19a 및 도 19b는 서열분석 시스템의 하나의 구현예를 도시한다. 서열분석 시스템은 구성가능한 프로세서를 포함한다.

도 19c는 염기 호출 센서 출력들과 같은, 서열분석 시스템으로부터의 센서 데이터의 분석을 위한 시스템의 단순화된 블록도이다.

도 20a는 호스트 프로세서에 의해 실행되는 런타임 프로그램의 기능들을 포함한, 염기 호출 동작의 태양들을 보여주는 단순화된 도면이다.

도 20b는 구성가능한 프로세서의 구성의 단순화된 도면이다.

도 21은 제1 염기 호출 반복 동안 생성된 압축된 공간 맵들을 오프-칩 메모리(예컨대, 오프-칩 DRAM, 호스트 RAM, 호스트 고대역폭 메모리(high bandwidth memory, HBM))로부터 제2 염기 호출 반복 동안 이용가능하게 하는 개시된 데이터 흐름 로직의 다른 구현예를 도시한다.

도 22는 제1 염기 호출 반복 동안 생성된 압축된 공간 맵들을 온-칩 메모리(예컨대, 온-칩 DRAM, 온-칩 SRAM, 온-칩 BRAM, 상호접속부를 통해 프로세서에 부착되는 DRAM)로부터 제2 염기 호출 반복 동안 이용가능하게 하는

개시된 데이터 흐름 로직의 하나의 구현예를 도시한다.

도 23은 개시된 신경 네트워크 기반 염기 호출자의 소위 분할 아키텍처의 하나의 구현예를 도시한다.

도 24a는 특징부 맵 추가를 통해 이전 정보를 다운스트림으로 재주입하는 잔차 접속을 도시한다.

도 24b는 잔차 블록들 및 스킵 접속들의 하나의 구현예를 도시한다.

도 24c는 공간 콘볼루션 층들이 스킵 접속들을 갖는 잔차 블록들로 그룹화되는 신경 네트워크 기반 염기 호출자의 잔차 아키텍처를 도시한다.

도 25a는 본 명세서에 기술된 신경 네트워크 기반 염기 호출자의 개시된 버스 네트워크의 세부사항들을 도시한다.

도 25b는 개시된 버스 네트워크의 예시적인 동작을 도시한다.

도 25c는 개시된 버스 네트워크의 디멘션 호환성 로직의 하나의 구현예를 도시한다.

도 26은 개시된 버스 네트워크의 다른 예를 도시한다.

도 27은 개시된 버스 네트워크의 또 다른 예를 도시한다.

도 28은 개시된 버스 네트워크의 스케일링 로직의 하나의 구현예를 도시한다.

도 29는 시간 네트워크의 시간 콘볼루션 층들 사이의 스킵 접속들의 하나의 구현예를 도시한다.

도 30은 압축 로직으로 구성된 네트워크 네트워크 기반 염기 호출자(sqz2 염기 호출자)에 의한 염기 호출 성능을, 압축 로직이 없는 네트워크 네트워크 기반 염기 호출자에 대해 그리고 Illumina의 비-신경 네트워크 기반 염기 호출자 실시간 분석(Real-Time Analysis, RTA) 소프트웨어에 대해 비교한다.

도 31은 개시된 압축 로직의 사용에 의해 RAM 및 DRAM 사용량의 절약이 유발되었음을 도시한다.

도 32는 분할 및 스킵 아키텍처들로 구성된 네트워크 네트워크 기반 염기 호출자(split_res)에 의한 염기 호출 성능을, RTA 염기 호출자 및 분할 및 스킵 아키텍처들이 없는 네트워크 네트워크 기반 염기 호출자(distilled)의 다른 버전에 대해 비교한다.

도 33은 개시된 기술을 구현하는데 사용될 수 있는 컴퓨터 시스템이다.

발명을 실시하기 위한 구체적인 내용

[0024] 아래의 논의는 어느 당업자라도 개시된 기술을 제조하고 사용할 수 있게 하도록 제시되며, 특정의 응용 및 그의 요건과 관련하여 제공된다. 개시된 구현예들에 대한 다양한 변형들은 당업자들에게 용이하게 명백할 것이며, 본 명세서에서 정의된 일반적인 원리들은 개시된 기술의 사상 및 범주로부터 벗어남이 없이 다른 구현예들 및 응용예들에 적용될 수 있다. 따라서, 개시된 기술은 도시된 구현예들로 제한되도록 의도된 것이 아니라, 본 명세서에 개시된 원리들 및 특징들과 일치하는 가장 넓은 범주에 부합되어야 한다.

[0025] 서열분석 이미지들

[0026] 염기 호출은 서열의 뉴클레오티드 조성을 결정하는 프로세스이다. 염기 호출은 Illumina의 iSeq, HiSeqX, HiSeq 3000, HiSeq 4000, HiSeq 2500, NovaSeq 6000, NextSeq 550, NextSeq 1000, NextSeq 2000, NextSeqDx, MiSeq, 및 MiSeqDx와 같은 서열분석 기구에 의해 수행되는 서열분석 런(또는 서열분석 반응) 동안 생성된 이미지 데이터, 즉 서열분석 이미지들을 분석하는 것을 수반한다.

[0027] 하기의 논의는 하나의 구현예에 따라, 서열분석 이미지들이 어떻게 생성되고 그들이 무엇을 묘사하는지의 개요를 서술한다.

[0028] 염기 호출은 서열분석 이미지들에 인코딩된 세기 데이터를 뉴클레오티드 서열들로 디코딩한다. 하나의 구현예에서, Illumina 서열분석 플랫폼들은 염기 호출을 위한 순환 가역적 종결(cyclic reversible termination, CRT) 화학을 채용한다. 프로세스는 각각의 새롭게 추가된 뉴클레오티드의 방출된 신호를 추적하면서, 형광-라벨링된 뉴클레오티드들을 갖는 템플릿 가닥들에 상보적인 발생기 가닥들을 성장시키는 것에 의존한다. 형광-라벨링된 뉴클레오티드들은 뉴클레오티드 유형의 형광단 신호를 앵커링하는 3' 제거가능한 블록을 갖는다.

[0029] 서열분석은 반복 사이클들에서 발생하는데, 각각은 3개의 단계들을 포함한다: (a) 형광-라벨링된 뉴클레오티드

를 추가하는 것에 의한 발생기 가닥의 연장; (b) 서열분석 기구의 광학 시스템의 하나 이상의 레이저들을 사용한 형광단의 여기(excitation)하고 광학 시스템의 상이한 필터들을 통해 이미징하여, 서열분석 이미지들을 산출; (c) 다음 서열분석 사이클을 대비한 형광단의 절단(cleavage) 및 3' 블록의 제거. 혼입 및 이미징 사이클들은 지정된 수의 서열분석 사이클들까지 반복되어, 판독물 길이를 정의한다. 이러한 접근법을 사용하여, 각각의 사이클이 템플릿 가닥들을 따라 새로운 포지션을 심문한다.

[0030] Illumina 서열분석기들의 엄청난 힘은 CRT 반응들을 겪는 수백만 또는 심지어 수십억 개의 클러스터들("피분석물들"로도 칭해짐)을 동시에 실행하고 감지하는 그들의 능력으로부터 유래한다. 클러스터는 템블릿 가닥의 대략 1000개의 동일한 복제물들을 포함하지만, 클러스터들은 크기 및 형상이 다르다. 클러스터들은, 서열분석 런 전에, 입력 라이브러리의 브리지 증폭(bridge amplification) 또는 배제 증폭(exclusion amplification)에 의해 템플릿 가닥으로부터 성장된다. 증폭 및 클러스터 성장의 목적은 방출된 신호의 세기를 증가시키는 것인데, 이는 이미징 디바이스가 단일 가닥의 형광단 신호를 신뢰성 있게 감지할 수 없기 때문이다. 그러나, 클러스터 내의 가닥들의 물리적 거리는 작고, 따라서, 이미징 디바이스는 가닥들의 클러스터를 단일 스폿으로 인지한다.

[0031] 서열분석은 플로우 셀(또는 바이오센서) - 입력 가닥들을 보유하는 작은 유리 슬라이드 - 에서 발생한다. 플로우 셀은 현미경 이미징, 여기 레이저들, 및 형광 필터들을 포함하는 광학 시스템에 접속된다. 플로우 셀은 레인들로 칭해지는 다수의 챔버들을 포함한다. 레인들은 서로 물리적으로 분리되어 있고, 샘플 교차 오염 없이 구별가능한 상이한 태깅된 서열분석 라이브러리들을 포함할 수 있다. 일부 구현예들에서, 플로우 셀은 패턴화된 표면을 포함한다. "패턴화된 표면"은 고체 지지체의 노출된 층 내의 또는 그 상에서의 상이한 영역들의 배열을 지칭한다.

[0032] 서열분석 기구의 이미징 디바이스(예컨대, 전하 결합 소자(charge-coupled device, CCD) 또는 상보성 금속 산화물 반도체(complementary metal-oxide-semiconductor, CMOS) 센서와 같은 솔리드 스테이트 이미저)는 타일들로 칭해지는 일련의 비-중첩 영역들 내의 레인들을 따르는 다수의 위치들에서 스냅숏(snapshot)들을 촬영한다. 예를 들어, 레인당 64개 또는 96개의 타일들이 있을 수 있다. 타일은 수십만 내지 수백만 개의 클러스터들을 보유한다.

[0033] 서열분석 런의 출력은 서열분석 이미지들이다. 서열분석 이미지들은 픽셀화된 단위들(예컨대, 픽셀들, 슈퍼픽셀들, 서브픽셀들)의 그리드(또는 어레이)를 사용하여 클러스터들 및 그들의 주변 배경의 세기 방출물들을 묘사한다. 세기 방출물들은 픽셀화된 단위들의 세기 값들로서 저장된다. 서열분석 이미지들은 픽셀화된 단위들의 그리드의 디멘션들 w x h를 가지며, 여기서 w(폭) 및 h(높이)는 1 내지 100,000 범위의 임의의 수들(예컨대, 115 x 115, 200 x 200, 1800 x 2000, 2200 x 25000, 2800 x 3600, 4000 x 400)이다 . 일부 구현예들에서, w 및 h는 동일하다. 다른 구현예들에서, w 및 h는 상이하다. 서열분석 이미지들은 서열분석 런 동안 뉴클레오티드 서열들 내의 뉴클레오티드 혼입의 결과로서 생성되는 세기 방출물들을 묘사한다. 세기 방출물들은 연관된 클러스터들 및 그들의 주변 배경으로부터의 것이다.

[0034] 신경 네트워크 기반 염기 호출

[0035] 하기의 논의는 본 명세서에 기술된 신경 네트워크 기반 염기 호출자(100)에 초점을 맞춘다. 먼저, 하나의 구현예에 따르면, 신경 네트워크 기반 염기 호출자(100)에 대한 입력이 기술된다. 이어서, 신경 네트워크 기반 염기 호출자(100)의 구조 및 형태의 예들이 제공된다. 마지막으로, 하나의 구현예에 따르면, 신경 네트워크 기반 염기 호출자(100)의 출력이 기술된다.

[0036] 데이터 흐름 로직은 염기 호출을 위해 신경 네트워크 기반 염기 호출자(100)에 서열분석 이미지들을 제공한다. 신경 네트워크 기반 염기 호출자(100)는 패치 단위(또는 타일 단위)로 서열분석 이미지들에 액세스한다. 패치들 각각은 서열분석 이미지들을 형성하는 픽셀화된 단위들의 그리드 내의 픽셀화된 단위들의 서브그리드(또는 서브어레이)이다. 패치들은 픽셀화된 단위들의 서브그리드의 디멘션들 q x r을 가지며, 여기서 q(폭) 및 r(높이)은 1 내지 10,000 범위의 임의의 수들(예컨대, 3 x 3, 5 x 5, 7 x 7, 10 x 10, 15 x 15, 25 x 25, 64 x 64, 78 x 78, 115 x 115)이다 . 일부 구현예들에서, q 및 r은 동일하다. 다른 구현예들에서, q 및 r은 상이하다. 일부 구현예들에서, 서열분석 이미지로부터 추출된 패치들은 동일한 크기의 것들이다. 다른 구현예들에서, 패치들은 상이한 크기들의 것들이다. 일부 구현예들에서, 패치들은 (예컨대, 에지들 상에서) 중첩되는 픽셀화된 단위들을 가질 수 있다.

[0037] 서열분석은 대응하는 m개의 이미지 채널들에 대해 서열분석 사이클당 m개의 서열분석 이미지들을 생성한다. 즉, 서열분석 이미지들 각각은 하나 이상의 이미지(또는 세기) 채널들(색상 이미지의 적색, 녹색, 청색(red,

green, blue, RGB) 채널들과 유사함)을 갖는다. 하나의 구현예에서, 각각의 이미지 채널은 복수의 필터 파장 대역들 중 하나에 대응한다. 다른 구현예에서, 각각의 이미지 채널은 서열분석 사이클에서의 복수의 이미징 이벤트들 중 하나에 대응한다. 또 다른 구현예에서, 각각의 이미지 채널은 특정 레이저를 사용하는 조명과 특정 광학 필터를 통한 이미징의 조합에 대응한다. 이미지 패치들은 특정 서열분석 사이클 동안 m개의 이미지 채널들 각각으로부터 타일링(또는 액세스)된다. 4-, 2-, 및 1-채널 화학들과 같은 상이한 구현예들에서, m은 4 또는 2이다. 다른 구현예들에서, m은 1, 3, 또는 4 초과이다.

[0038] 예를 들어, 서열분석 런이 2개의 상이한 이미지 채널들, 즉 청색 채널 및 녹색 채널을 사용하여 구현된다는 것을 고려한다. 이어서, 각각의 서열분석 사이클에서, 서열분석 런은 청색 이미지 및 녹색 이미지를 생성한다. 이러한 방식으로, 서열분석 런의 일련의 k개의 서열분석 사이클들 동안, 청색 및 녹색 이미지들의 k개의 쌍들의 서열이 출력으로서 생성되고, 서열분석 이미지들로서 저장된다. 따라서, 신경 네트워크 기반 염기 호출자(100)에 의한 패치 레벨 프로세싱에 대해, 청색 및 녹색 이미지 패치들의 k개의 쌍들의 서열이 생성된다.

[0039] 염기 호출의 단일 반복(또는 단일 순방향 순회(traversal) 또는 순방향 패스의 단일 인스턴스)을 위한 신경 네트워크 기반 염기 호출자(100)에 대한 입력 이미지 데이터는 다수의 서열분석 사이클들의 슬라이딩 윈도우에 대한 데이터를 포함한다. 슬라이딩 윈도우는 예를 들어, 현재 서열분석 사이클, 하나 이상의 선행 서열분석 사이클들, 및 하나 이상의 연속 서열분석 사이클들을 포함할 수 있다.

[0040] 하나의 구현예에서, 입력 이미지 데이터는 3개의 서열분석 사이클들 동안의 데이터를 포함하여서, 염기 호출될 현재(시간 t) 서열분석 사이클 동안의 데이터에 (i) 좌측 플랭킹/콘텍스트/이전/선행/우선(시간 t-1) 서열분석 사이클 동안의 데이터, 및 (ii) 우측 플랭킹/콘텍스트/다음/연속/후속(시간 t+1) 서열분석 사이클 동안의 데이터가 동반되도록 한다.

[0041] 다른 구현예에서, 입력 이미지 데이터는 5개의 서열분석 사이클들 동안의 데이터를 포함하여서, 염기 호출될 현재(시간 t) 서열분석 사이클 동안의 데이터에 (i) 제1 좌측 플랭킹/콘텍스트/이전/선행/우선(시간 t-1) 서열분석 사이클 동안의 데이터, (ii) 제2 좌측 플랭킹/콘텍스트/이전/선행/우선(시간 t-2) 서열분석 사이클 동안의 데이터, (iii) 제1 우측 플랭킹/콘텍스트/다음/연속/후속(시간 t+1)에 대한 데이터, 및 (iv) 제2 우측 플랭킹/콘텍스트/다음/연속/후속(시간 t+2) 서열분석 사이클 동안의 데이터가 동반되도록 한다.

[0042] 또 다른 구현예에서, 입력 이미지 데이터는 7개의 서열분석 사이클들 동안의 데이터를 포함하여서, 염기 호출될 현재(시간 t) 서열분석 사이클 동안의 데이터에 (i) 제1 좌측 플랭킹/콘텍스트/이전/선행/우선(시간 t-1) 서열분석 사이클 동안의 데이터, (ii) 제2 좌측 플랭킹/콘텍스트/이전/선행/우선(시간 t-2) 서열분석 사이클 동안의 데이터, (iii) 제3 좌측 플랭킹/콘텍스트/이전/선행/우선(시간 t-3) 서열분석 사이클 동안의 데이터, (iv) 제1 우측 플랭킹/콘텍스트/다음/연속/후속(시간 t+1)에 대한 데이터, (v) 제2 우측 플랭킹/콘텍스트/다음/연속/후속(시간 t+2) 서열분석 사이클 동안의 데이터, 및 (vi) 제3 우측 플랭킹/콘텍스트/다음/연속/후속(시간 t+3) 서열분석 사이클 동안의 데이터가 동반되도록 한다. 다른 구현예들에서, 입력 이미지 데이터는 단일 서열분석 사이클 동안의 데이터를 포함한다. 또 다른 구현예들에서, 입력 이미지 데이터는 10개, 15개, 20개, 30개, 58개, 75개, 92개, 130개, 168개, 175개, 209개, 225개, 230개, 275개, 318개, 325개, 330개, 525개, 또는 625개의 서열분석 사이클들 동안의 데이터를 포함한다.

[0043] 신경 네트워크 기반 염기 호출자(100)는 하나의 구현예에 따르면, 이미지 패치들을 그의 콘볼루션 층들을 통해 프로세싱하고 대안적인 표현을 생성한다. 이어서, 대안적인 표현은, 바로 현재(시간 t) 서열분석 사이클 또는 서열분석 사이클들 각각, 즉 현재(시간 t) 서열분석 사이클, 제1 및 제2 선행(시간 t-1, 시간 t-2) 서열분석 사이클들, 및 제1 및 제2 후행(시간 t+1, 시간 t+2) 서열분석 사이클들 중 어느 하나의 서열분석 사이클 동안 염기 호출을 생성하기 위한 출력 층(예컨대, 소프트맥스 층)에 의해 사용된다. 생성된 염기 호출들은 서열분석 판독물들을 형성한다.

[0044] 하나의 구현예에서, 신경 네트워크 기반 염기 호출자(100)는 특정 서열분석 사이클 동안 단일 표적 클러스터에 대한 염기 호출을 출력한다. 다른 구현예에서, 신경 네트워크 기반 염기 호출자(100)는 특정 서열분석 사이클 동안 복수의 표적 클러스터들 내의 각각의 표적 클러스터에 대한 염기 호출을 출력한다. 또 다른 구현예에서, 신경 네트워크 기반 염기 호출자(100)는 복수의 서열분석 사이클들 중의 각각의 서열분석 사이클 동안 복수의 표적 클러스터들 내의 각각의 표적 클러스터에 대한 염기 호출을 출력하여, 이에 의해, 각각의 표적 클러스터에 대한 염기 호출 서열을 생성한다.

[0045] 하나의 구현예에서, 신경 네트워크 기반 염기 호출자(100)는 다층 퍼셉트론(multilayer perceptron, MLP)이다.

다른 구현예에서, 신경 네트워크 기반 염기 호출자(100)는 피드포워드 신경 네트워크이다. 또 다른 구현예에서, 신경 네트워크 기반 염기 호출자(100)는 완전 접속 신경 네트워크이다. 추가 구현예에서, 신경 네트워크 기반 염기 호출자(100)는 완전 콘볼루션 신경 네트워크이다. 다른 추가 구현예에서, 신경 네트워크 기반 염기 호출자(100)는 시맨틱 세그먼트화 신경 네트워크이다. 또 다른 추가 구현예에서, 신경 네트워크 기반 염기 호출자(100)는 생성적 대립쌍 네트워크(generative adversarial network, GAN)이다.

[0046] 하나의 구현예에서, 신경 네트워크 기반 염기 호출자(100)는 복수의 콘볼루션 층들을 갖는 콘볼루션 신경 네트워크(CNN)이다. 다른 구현예에서, 신경 네트워크 기반 염기 호출자(100)는 장단기 메모리 네트워크(short-term memory network, LSTM), 양방향 LSTM(bi-directional LSTM, Bi-LSTM), 또는 게이트형 순환 유닛(gated recurrent unit, GRU)과 같은 순환 신경 네트워크(recurrent neural network, RNN)이다. 또 다른 구현예에서, 신경 네트워크 기반 염기 호출자(100)는 CNN 및 RNN 둘 모두를 포함한다.

[0047] 또 다른 구현예들에서, 신경 네트워크 기반 염기 호출자(100)는 1D 콘볼루션, 2D 콘볼루션, 3D 콘볼루션, 4D 콘볼루션, 5D 콘볼루션, 확장형 또는 아트로스 콘볼루션, 전치 콘볼루션, 깊이별 분리가능 콘볼루션, 포인트별 콘볼루션, 1 x 1 콘볼루션, 그룹 콘볼루션, 편평형 콘볼루션, 공간 및 교차 채널 콘볼루션, 셔플 그룹형 콘볼루션, 공간 분리가능 콘볼루션, 및 디콘볼루션을 사용할 수 있다. 신경 네트워크 기반 염기 호출자(100)는 하나 이상의 손실 함수들, 예컨대 로지스틱 회귀(logistic regression)/로그(log) 손실, 다중클래스 교차-엔트로피(multi-class cross-entropy)/소프트맥스 손실, 이진 교차-엔트로피(binary cross-entropy) 손실, 평균 제곱 에러(mean-squared error) 손실, L1 손실, L2 손실, 평활한(smooth) L1 손실, 및 Huber 손실을 사용할 수 있다. 신경 네트워크 기반 염기 호출자(100)는 임의의 병렬성(parallelism), 효율성, 및 압축 스킴들, 예컨대 TFRecords, 압축 인코딩(예컨대, PNG), 샤딩(sharding), 맵 변환을 위한 병렬 호출, 배칭(batching), 프리페칭(prefetching), 모델 병렬성, 데이터 병렬성, 및 동기식/비동기식 확률적 기울기 하강(stochastic gradient descent, SGD)을 사용할 수 있다. 신경 네트워크 기반 염기 호출자(100)는 업샘플링 층, 다운샘플링 층, 순환 접속, 게이트 및 게이트형 메모리 유닛(예컨대, LSTM 또는 GRU), 잔차 블록, 잔차 접속, 하이웨이 접속, 스킵 접속, 핍홀(peephole) 접속, 활성화 함수(예컨대, ReLU(rectifying linear unit), 리키 ReLU(leaky ReLU), ELU(exponential liner unit), 시그모이드 및 tanh(hyperbolic tangent)와 같은 비선형 변환 함수), 배치 정규화 층, 규칙화 층, 드롭아웃, 풀링 층(예컨대, 최대 또는 평균 풀링), 글로벌 평균 풀링 층, 및 감쇠 메커니즘을 포함할 수 있다.

[0048] 신경 네트워크 기반 염기 호출자(100)는 역전파 기반 기울기 업데이트 기법들을 사용하여 트레이닝된다. 신경 네트워크 기반 염기 호출자(100)를 트레이닝시키기 위해 사용될 수 있는 예시적인 기울기 하강 기법들은 확률적 기울기 하강, 배치 기울기 하강, 및 미니 배치 기울기 하강을 포함한다. 신경 네트워크 기반 염기 호출자(100)를 트레이닝시키는 데 사용될 수 있는 기울기 하강 최적화 알고리즘들의 일부 예들은 Momentum, Nesterov 가속화된 기울기, Adagrad, Adadelta, RMSprop, Adam, AdaMax, Nadam, 및 AMSGrad이다.

[0049] 하나의 구현예에서, 신경 네트워크 기반 염기 호출자(100)는 특화된 아키텍처를 사용하여, 상이한 서열분석 사이클들 동안의 데이터의 프로세싱을 분리(segregate)한다. 특화된 아키텍처를 사용하기 위한 동기가 먼저 기술된다. 상기에서 논의된 바와 같이, 신경 네트워크 기반 염기 호출자(100)는 현재 서열분석 사이클, 하나 이상의 선행 서열분석 사이클들, 및 하나 이상의 연속 서열분석 사이클들 동안 이미지 패치들을 프로세싱한다. 추가적인 서열분석 사이클들 동안 데이터는 서열 특정 콘텍스트를 제공한다. 신경 네트워크 기반 염기 호출자(100)는 트레이닝 동안 서열 특정 콘텍스트를 학습하고, 그들을 염기 호출한다. 더욱이, 프리 및 포스트 서열분석 사이클들 동안의 데이터는 현재 서열분석 사이클 동안 프리-페이징 및 페이징 신호들의 2차 기여를 제공한다.

[0050] 그러나, 상이한 서열분석 사이클들에서 그리고 상이한 이미지 채널들에서 캡처된 이미지들은 오정렬되어 있고, 서로에 대해 잔차 정합 에러를 갖는다. 이러한 오정렬을 처리하기 위해, 특화된 아키텍처는 서열분석 사이클들 사이의 정보를 혼합하지 않고 단지 서열분석 사이클 내에서의 정보만을 혼합하는 공간 콘볼루션 층을 포함한다.

[0051] 공간 콘볼루션 층들(또는 공간 로직)은, 콘볼루션들의 "전용 비공유" 시퀀스를 통해 복수의 서열분석 사이클들 각각 동안 데이터를 독립적으로 프로세싱함으로써 분리를 조작할 수 있게 하는 소위 "분리형 콘볼루션들"을 사용한다. 분리형 콘볼루션들은, 임의의 다른 서열분석 사이클의 데이터 및 생성된 특징부 맵들에 걸쳐서 콘볼루션하지 않고서, 단지 주어진 서열분석 사이클, 즉, 사이클내(intra-cycle) 전용의 데이터 및 생성된 특징부 맵들에 걸쳐서 콘볼루션한다.

[0052] 예를 들어, 입력 이미지 데이터는 (i) 염기 호출될 현재(시간 t) 서열분석 사이클 동안의 현재 이미지 패치,

(ii) 이전(시간 t-1) 서열분석 사이클 동안의 이전 이미지 패치, 및 (iii) 다음(시간 t+1) 서열분석 사이클 동안의 다음 이미지 패치를 포함한다는 것을 고려한다. 이어서, 특화된 아키텍처는 3개의 별개의 콘볼루션 파이프라인들, 즉, 현재 콘볼루션 파이프라인, 이전 콘볼루션 파이프라인, 및 다음 콘볼루션 파이프라인을 개시한다. 현재 데이터 프로세싱 파이프라인은 현재(시간 t) 서열분석 사이클 동안 현재 이미지 패치를 입력으로서 수신하고, 이를 복수의 공간 콘볼루션 층들을 통해 독립적으로 프로세싱하여 최종 공간 콘볼루션 층의 출력으로서 소위 "현재 공간적으로 콘볼루션된 표현"을 생성한다. 이전 콘볼루션 파이프라인은 이전(시간 t-1) 서열분석 사이클 동안 이전 이미지 패치를 입력으로서 수신하고, 이를 복수의 공간 콘볼루션 층들을 통해 독립적으로 프로세싱하여 최종 공간 콘볼루션 층의 출력으로서 소위 "이전에 공간적으로 콘볼루션된 표현"을 생성한다. 다음 콘볼루션 파이프라인은 다음(시간 t+1) 서열분석 사이클 동안 다음 이미지 패치를 입력으로서 수신하고, 이를 복수의 공간 콘볼루션 층들을 통해 독립적으로 프로세싱하여 최종 공간 콘볼루션 층의 출력으로서 소위 "다음에 공간적으로 콘볼루션된 표현"을 생성한다.

[0053] 일부 구현예들에서, 현재, 이전, 및 다음 콘볼루션 파이프라인들은 병렬로 실행된다. 일부 구현예들에서, 공간 콘볼루션 층들은 특화된 아키텍처 내의 공간 콘볼루션 네트워크(또는 서브네트워크)의 일부이다.

[0054] 신경 네트워크 기반 염기 호출자(*100*)는 서열분석 사이클들 사이의, 즉, 사이클간(inter-cycle)의 정보를 혼합하는 시간 콘볼루션 층들(또는 시간 로직)을 추가로 포함한다. 시간 콘볼루션 층들은 공간 콘볼루션 네트워크로부터 그들의 입력들을 수신하고, 각자의 데이터 프로세싱 파이프라인들에 대해 최종 공간 콘볼루션 층에 의해 생성되는 공간적으로 콘볼루션된 표현들 상에서 동작한다.

[0055] 시간 콘볼루션 층들의 사이클간 동작가능성 자유는, 공간 콘볼루션 네트워크에 대한 입력으로서 공급되는 이미지 데이터 내에 존재하는 오정렬 속성이 공간 콘볼루션 층들의 시퀀스에 의해 수행되는 분리형 콘볼루션들의 스택 또는 캐스케이드에 의해, 공간적으로 콘볼루션된 표현들로부터 제거된다(purged out)는 사실로부터 나온다.

[0056] 시간 콘볼루션 층들은 슬라이딩 윈도우 단위로 연속 입력들에서 입력 채널들에 걸쳐서 그룹별로 콘볼루션하는 소위 "조합형 콘볼루션들"을 사용한다. 하나의 구현예에서, 연속 입력들은 이전 공간 콘볼루션 층 또는 이전 시간 콘볼루션 층에 의해 생성되는 연속 출력들이다.

[0057] 일부 구현예들에서, 시간 콘볼루션 층들은 특화된 아키텍처 내의 시간 콘볼루션 네트워크(또는 서브네트워크)의 일부이다. 시간 콘볼루션 네트워크는 그의 입력들을 공간 콘볼루션 네트워크로부터 수신한다. 하나의 구현예에서, 시간 콘볼루션 네트워크의 제1 시간 콘볼루션 층은 서열분석 사이클들 사이의 공간적으로 콘볼루션된 표현들을 그룹별로 조합한다. 다른 구현예에서, 시간 콘볼루션 네트워크의 후속 시간 콘볼루션 층들은 이전 시간 콘볼루션 층들의 연속 출력들을 조합한다. 최종 시간 콘볼루션 층의 출력은 출력을 생성하는 출력 층에 공급된다. 출력은 하나 이상의 서열분석 사이클들에서 하나 이상의 클러스터들을 염기 호출하는 데 사용된다.

[0058] 신경 네트워크 기반 염기 호출자(100)에 관한 추가적인 세부사항들은 발명의 명칭이 "ARTIFICIAL INTELLIGENCE-BASED SEQUENCING"이고 2019년 3월 21일자로 출원된 미국 가특허 출원 제62/821,766호(대리인 문서 번호 ILLM 1008-9/IP-1752-PRV)에서 찾을 수 있으며, 이는 본 명세서에 참고로 포함된다.

[0059] <u>압축 네트워크</u>

[0060] 상기에서 논의된 바와 같이, 신경 네트워크 기반 염기 호출자(100)의 특화된 아키텍처는 대응하는 서열분석 사이클들 동안 이미지 패치들의 슬라이딩 윈도우들을 프로세싱한다. 연속 슬라이딩 윈도우들의 서열분석 사이클들 사이에 중첩이 존재한다. 이는 신경 네트워크 기반 염기 호출자(100)가 중첩 서열분석 사이클들 동안 이미지 패치들을 리던던트로 프로세싱하게 한다. 이는 이어서, 연산 리소스들의 낭비를 초래한다. 예를 들어, 하나의 구현예에서, 신경 네트워크 기반 염기 호출자(100)의 각각의 공간 콘볼루션 층은 거의 1억 개의 곱셈 연산들을 갖는다. 이어서, 5개의 서열분석 사이클들의 윈도우 및 7개의 공간 콘볼루션 층들의 캐스케이드(또는 시퀀스)에 대해, 공간 콘볼루션 신경 네트워크는 약 6억 2천만 개의 곱셈 연산들을 실행한다. 더욱이, 시간 콘볼루션 신경 네트워크는 약 1천만 개의 곱셈 연산들을 실행한다.

[0061] 현재 슬라이딩 윈도우(또는 염기 호출의 현재 반복) 내의 사이클 N-1 동안의 이미지 데이터가 이전 슬라이딩 윈도우(또는 염기 호출의 이전 반복) 내의 사이클 N으로서 프로세싱되므로, 현재 슬라이딩 윈도우에서 행해진 프로세싱의 중간 결과들 및 후속 슬라이딩 윈도우들에서의 중간 결과들을 저장하고, 이에 의해, 연속 슬라이딩 윈도우들 사이의 중첩 서열분석 사이클들 동안의 입력 이미지 데이터의 리던던트 프로세싱(또는 재프로세싱)을 우회(또는 배제)시키는 기회가 발생한다.

[0062] 그러나, 중간 결과들은 비현실적인 양의 저장소를 요구하는 수 테라바이트의 데이터이다. 이러한 기술적 문제

를 극복하기 위해, 개시된 기술은, 중간 결과들이 신경 네트워크 기반 염기 호출자(100)에 의해 처음 생성될 때 중간 결과들을 압축하고 후속 슬라이딩 윈도우들에서 압축된 중간 결과들을 용도변경하여 리던던트 연산을 회피시키는 것, 및 이에 의해, 중간 결과들을 재생성하지 않는 것(또는 그들을 1회만 생성하는 것)을 제안한다. 일부 구현예들에서, 개시된 기술은 신경 네트워크 기반 염기 호출자(100)의 공간 네트워크에서의 콘볼루션들의 약 80%를 절약한다. 하나의 구현예에서, 후속 서열분석 사이클들에서의 압축된 특징부 맵들의 용도변경 및 압축 로직이 5개의 서열분석 사이클들(예컨대, 사이클 N, 사이클 N+1, 사이클 N-1, 사이클 N+2, 및 사이클 N-2)의 입력 윈도우에 사용될 때 공간 콘볼루션들에서 80% 절약이 관찰된다. 다른 구현예에서, 후속 서열분석 사이클들에서의 압축된 특징부 맵들의 용도변경 및 압축 로직이 10개의 서열분석 사이클들(예컨대, 사이클 N, 사이클 N+1, 사이클 N-1, 사이클 N+2, 사이클 N-2, 사이클 N+3, 및 사이클 N-3)의 입력 윈도우에 사용될 때 공간 콘볼루션들에서 90% 절약이 관찰된다. 즉, 윈도우 크기가 클수록, 압축된 특징부 맵들의 용도변경 및 압축 로직의 사용으로부터의 절약이 더 클수록, 그리고 윈도우 크기가 클수록, 추가적인 플랭킹 사이클들로부터의 더 큰 콘텍스트의 혼입으로 인해 염기 호출 성능이 더 양호하다. 따라서, 더 큰 윈도우들에 대한 더 큰 절약은 주어진 계산 능력에 대한 전체 성능을 개선한다.

[0063] 압축 로직에 의해 유발된 계산 효율 및 컴팩트한 계산 풋프린트는, 중앙 프로세싱 유닛(Central Processing Unit, CPU), 그래픽 프로세싱 유닛(GPU), 필드 프로그래밍가능 게이트 어레이(FPGA), 코어스-그레인드 재구성가능 아키텍처(Coarse-Grained Reconfigurable Architecture, CGRA), 주문형 반도체(ASIC), 주문형 명령어 세트 프로세서(Application Specific Instruction-set Processor, ASIP), 및 디지털 신호 프로세서(DSP)와 같은 리소스 제약 프로세서들 상에서 신경 네트워크 기반 염기 호출자(100)의 하드웨어 구현을 용이하게 한다.

[0064] 압축 로직에 의해 저장된 계산은 신경 네트워크 기반 염기 호출자(100)에 더 많은 콘볼루션 연산자들을 혼입하는 것을 허용한다. 예들은 공간 및 시간 콘볼루션 층들에 더 많은 콘볼루션 필터들을 추가하는 것, 콘볼루션 필터들의 크기를 증가시키는 것, 및 공간 및 시간 콘볼루션 층들의 수를 증가시키는 것을 포함한다. 추가적인 콘볼루션 연산들은 신경 네트워크 기반 염기 호출자(100)의 세기 패턴 검출 및 전체 염기 호출 정확도를 개선한다.

[0065] 압축 로직에 의해 저장된 계산은 또한, 대상 슬라이딩 윈도우가 증가된 수의 서열분석 사이클들을 포함하기 위해 입력 이미지 데이터를 확장하는 것을 허용한다. 확장된 슬라이딩 윈도우들은 추가적인 플랭킹 서열분석 사이클들로부터 과잉 이미지 패치들을 가져옴으로써 염기 호출 콘텍스트를 넓힌다.

[0066] 더욱이, 오리지널 중간 결과들과는 대조적으로, 압축된 중간 결과들의 사용으로 인해 발생할 수 있는 정확성의 임의의 감소는 추가적인 콘볼루션 연산자의 혼입 및 슬라이딩 윈도우들의 확장에 의해 보상된다.

[0067] 도 1a는 염기 호출의 제1 반복 동안 압축된 공간 맵 세트들을 생성하는 개시된 압축 로직의 하나의 구현예를 도시한다. 도시된 예에서, 서열분석 사이클들의 제1 윈도우는 서열분석 사이클들 1, 2, 3, 4, 및 5를 포함한다. 각자의 서열분석 사이클들 1, 2, 3, 4, 및 5 동안의 각자의 이미지 패치들(102, 112, 122, 132, 142)(또는 사이클별 분석물 채널 세트들)은 공간 로직(104)(또는 공간 네트워크 또는 공간 서브네트워크 또는 공간 콘볼루션 신경 네트워크)을 통해 개별적으로 프로세싱되어, 각자의 서열분석 사이클들 1, 2, 3, 4, 및 5에 대한 각자의 공간 맵들(106, 116, 126, 136, 146)(또는 중간 결과들 또는 공간 출력 세트들 또는 공간 특징부 맵 세트들)을 생성한다. 공간 콘볼루션 네트워크(104)는 1D, 2D, 또는 3D 콘볼루션들을 사용할 수 있다.

[0068] 공간 로직(104)은 공간 콘볼루션 층들의 시퀀스(또는 캐스케이드)를 포함한다. 각각의 공간 콘볼루션 층은 분리형 콘볼루션들을 구현하는 복수의 공간 콘볼루션 필터들을 갖는 필터 뱅크를 갖는다. 따라서, 각각의 공간 콘볼루션 층은 복수의 공간 특징부 맵들을 출력으로서 생성한다. 대상 공간 콘볼루션 층에 의해 생성된 공간 특징부 맵들의 개수는 대상 공간 콘볼루션 층에 구성된 공간 콘볼루션 필터들의 개수의 함수이다. 예를 들어, 대상 공간 콘볼루션 층이 14개의 공간 콘볼루션 필터들을 갖는 경우, 대상 공간 콘볼루션 층은 14개의 공간 특징부 맵들을 생성한다. 집적(aggregate) 관점에서, 14개의 공간 특징부 맵들은 14개의 채널들(또는 깊이 디멘션 = 14)을 갖는 공간 특징부 맵 볼륨(또는 텐서)으로 간주될 수 있다.

[0069] 더욱이, 대상 공간 콘볼루션 층 뒤에 이어지는 다음 공간 콘볼루션 층도 또한, 14개의 공간 콘볼루션 필터들로 구성될 수 있다. 그러한 경우에, 다음 공간 콘볼루션 층은 대상 공간 콘볼루션 층에 의해 생성된 14개의 공간 특징부 맵들을 입력으로서 프로세싱하고 그 자체로 14개의 새로운 공간 특징부 맵들을 출력으로서 생성한다. 도 1a는 각자의 서열분석 사이클들 1, 2, 3, 4, 및 5 동안 공간 네트워크(104)의 최종 공간 콘볼루션 층에 의해 생성된 5개의 공간 특징부 맵 세트들(106, 116, 126, 136, 146)을 도시한다. 도시된 예에서, 5개의 공간 특징부 맵 세트들(106, 116, 126, 136, 146) 각각은 14개의 특징부 맵들을 갖는다.

[0070] 도 1d는 공간 네트워크(104)의 7개의 공간 콘볼루션 층들의 캐스케이드에 의해 생성된 7개의 공간 특징부 맵 세트들(196a, 196b, 196c, 196d, 196e, 196f, 196g)의 시퀀스를 도시한다. 대상 서열분석 사이클 i 동안의 사이클별 입력 패치 데이터(194)는 (오리지널 서열분석 이미지들 내의 2개의 이미지 채널들로 인해) 115 x 115의 공간 디멘션수(spatial dimensionality) 및 2의 깊이 디멘션수를 갖는다. 하나의 구현예에서, 7개의 공간 콘볼루션 층들 각각은 연속 공간 특징부 맵 볼륨들의 공간 디멘션수를 2만큼 감소시키는, 예를 들어, 10 x 10으로부터 8 x 8로 감소시키는 3 x 3 콘볼루션들을 사용한다.

[0071] 제1 공간 특징부 맵 볼륨(196a)은 공간 디멘션들 113 x 113(즉, 115 x 115로부터 제1 공간 콘볼루션 층의 3 x 3 콘볼루션들만큼 감소됨) 및 14의 깊이 디멘션(즉, 제1 공간 콘볼루션 층 내의 14개의 공간 콘볼루션 필터들로 인한 14개의 특징부 맵들 또는 14개의 채널들)을 갖는다. 제2 공간 특징부 맵 볼륨(196b)은 공간 디멘션들 111 x 111(즉, 113 x 113로부터 제2 공간 콘볼루션 층의 3 x 3 콘볼루션들만큼 감소됨) 및 14의 깊이 디멘션(즉, 제2 공간 콘볼루션 층 내의 14개의 공간 콘볼루션 필터들로 인한 14개의 특징부 맵들 또는 14개의 채널들)을 갖는다. 제3 공간 특징부 맵 볼륨(196c)은 공간 디멘션들 109 x 109(즉, 111 x 111로부터 제3 공간 콘볼루션 층의 3 x 3 콘볼루션들만큼 감소됨) 및 14의 깊이 디멘션(즉, 제3 공간 콘볼루션 층 내의 14개의 공간 콘볼루션 필터들로 인한 14개의 특징부 맵들 또는 14개의 채널들)을 갖는다. 제4 공간 특징부 맵 볼륨(196d)은 공간 디멘션들 107 x 107(즉, 109 x 109로부터 제4 공간 콘볼루션 층의 3 x 3 콘볼루션들만큼 감소됨) 및 14의 깊이 디멘션(즉, 제4 공간 콘볼루션 층 내의 14개의 공간 콘볼루션 필터들로 인한 14개의 특징부 맵들 또는 14개의 채널들)을 갖는다. 제5 공간 특징부 맵 볼륨(196e)은 공간 디멘션들 105 x 105(즉, 107 x 107로부터 제5 공간 콘볼루션 층의 3 x 3 콘볼루션들만큼 감소됨) 및 14의 깊이 디멘션(즉, 제5 공간 콘볼루션 층 내의 14개의 공간 콘볼루션 필터들로 인한 14개의 특징부 맵들 또는 14개의 채널들)을 갖는다. 제6 공간 특징부 맵 볼륨(196f)은 공간 디멘션들 103 x 103(즉, 105 x 105로부터 제6 공간 콘볼루션 층의 3 x 3 콘볼루션들만큼 감소됨) 및 14의 깊이 디멘션(즉, 제6 공간 콘볼루션 층 내의 14개의 공간 콘볼루션 필터들로 인한 14개의 특징부 맵들 또는 14개의 채널들)을 갖는다. 제7 공간 특징부 맵 볼륨(196g)은 공간 디멘션들 101 x 101(즉, 103 x 103로부터 제7 공간 콘볼루션 층의 3 x 3 콘볼루션들만큼 감소됨) 및 14의 깊이 디멘션(즉, 제7 공간 콘볼루션 층 내의 14개의 공간 콘볼루션 필터들로 인한 14개의 특징부 맵들 또는 14개의 채널들)을 갖는다.

[0072] 도 1a에 도시된 다중-사이클 예와 유사하게, 5개의 서열분석 사이클들 1, 2, 3, 4, 및 5와 5개의 사이클별 이미지 패치들(102, 112, 122, 132, 142)에 대해, 공간 로직(104)은 7개의 공간 특징부 맵 볼륨들(196a, 196b, 196c, 196d, 196e, 196f, 196g)의 5개의 각자의 시퀀스들을 개별적으로 생성하며, 이때 도 1a의 공간 맵들(106, 116, 126, 136, 146)은 도 1d의 최종 공간 특징부 맵 볼륨(196g)의 5개의 개별 인스턴스들과 동등하다.

[0073] 압축 로직(108)(또는 압축 네트워크 또는 압축 서브네트워크 또는 압축 층 또는 스퀴즈 층)은 공간 로직(104)의 출력들을 프로세싱하고 출력들의 압축된 표현을 생성한다. 하나의 구현예에서, 압축 네트워크(108)는 공간 네트워크(104)에 의해 생성된 특징부 맵들의 깊이 디멘션수를 감소시키는 압축 콘볼루션 층을 포함한다.

[0074] 예를 들어, 도 1a에서, 공간 맵들(106, 116, 126, 136, 146)의 깊이 디멘션수는 14(즉, 공간 출력당 14개의 특징부 맵들 또는 14개의 채널들)이다. 압축 네트워크(108)는 각자의 서열분석 사이클들 1, 2, 3, 4 및 5 동안 공간 맵들(106, 116, 126, 136, 146)을 각자의 압축된 공간 맵 세트들(110, 120, 130, 140, 150)로 감쇠시킨다. 압축된 공간 맵 세트들(110, 120, 130, 140, 150) 각각은 2의 깊이 디멘션수(즉, 압축된 공간 출력당 2개의 특징부 맵들 또는 2개의 채널들)를 갖는다. 다른 구현예들에서, 압축된 공간 맵 세트들(110, 120, 130, 140, 150)은 3 또는 4의 깊이 디멘션수(즉, 압축된 공간 출력당 3개 또는 4개의 특징부 맵들 또는 3개 또는 4개의 채널들)를 가질 수 있다. 또 다른 구현예들에서, 압축된 공간 맵 세트들(110, 120, 130, 140, 150)은 1의 깊이 디멘션수(즉, 압축된 공간 출력당 1개의 특징부 맵 또는 1개의 채널)를 가질 수 있다. 하나의 구현예에서, 압축 층(108)은 ReLU와 같은 활성화 함수를 포함하지 않는다. 다른 구현예들에서, 이는 활성화 함수를 포함할 수 있다. 다른 구현예들에서, 압축 로직(108)은 대응하는 압축된 공간 맵 세트들을, 4개 초과의 특징부 맵들을 각각 갖도록 구성할 수 있다.

[0075] 이제, 압축 로직(108)이 압축된 출력들을 어떻게 생성하는지에 대해 논의된다.

[0076] 하나의 구현예에서, 압축 로직(108)은 1 x 1 콘볼루션들을 사용하여, 비선형성을 도입하면서 특징부 맵들의 개수(즉, 깊이 디멘션 또는 채널들의 개수)를 감소시킨다. 1x1 콘볼루션들은 1의 커널 크기를 갖는다. 1x1 콘볼루션들은 공간 디멘션들을 변경하지 않으면서 볼륨 깊이를 다른 스퀴즈된 또는 확장된 표현으로 변환할 수 있다. 1x1 콘볼루션은 입력 채널들에 걸쳐 완전 접속된 선형 층처럼 동작한다. 이것은 더 많은 채널들을 갖는 특징부 맵들로부터 더 적은 특징부 맵들로 맵핑시키는 데 유용하다. 도 1e에서, 2개의 특징부 맵들을 갖는 입

력 텐서에 단일 1x1 콘볼루션이 적용된다. 1x1 콘볼루션은 2-채널 입력을 단일 채널 출력으로 압축한다.

[0077] 압축 층(108)에 의해 생성된 압축된 출력들(또는 압축된 특징부 맵들 또는 압축된 공간 맵들 또는 압축된 시간 맵들)의 개수는 압축 층(108)에 구성된 1x1 콘볼루션 필터들(또는 압축 콘볼루션 필터들 또는 압축 필터들)의 개수의 함수이다. 도 1f에서, 압축 층(108)은 2개의 1x1 콘볼루션 필터들(198a, 198b)을 갖는다. 제1 1x1 콘볼루션 필터(198a)는 14개의 특징부 맵들을 갖는 공간 특징부 볼륨(196g)을 프로세싱하고, 101 x 101의 공간 디멘션수를 보존하면서 제1 특징부 맵(199a)을 생성한다. 제2 1x1 콘볼루션 필터(198b)는 또한, 14개의 특징부 맵들을 갖는 공간 특징부 볼륨(196g)을 프로세싱하고, 101 x 101의 공간 디멘션수를 보존하면서 제2 특징부 맵(199b)을 생성한다. 따라서, 압축 층(108)은 14개의 특징부 맵들을 갖는 공간 특징부 볼륨(196g)을, 2개의 공간 특징부 맵들(199a, 199b)을 갖는 압축된 출력으로 감소시킨다(즉, 압축비 = 7).

[0078] 시계열 관점에서, 서열분석 사이클 5는 중앙 서열분석 사이클(N)이고, 서열분석 사이클들 1 및 2는 좌측 플랭킹 서열분석 사이클들(N-1, n-2)이고, 서열분석 사이클들 4 및 5는 좌측 플랭킹 서열분석 사이클들(N+1, N+2)이다. 따라서, 중심의 압축된 출력(130)은 중앙 서열분석 사이클(N) 동안 생성되고, 좌측의 플랭킹 압축된 출력(120)은 좌측의 플랭킹 서열분석 사이클(N-1) 동안 생성되고, 더 좌측의 플랭킹 압축된 출력(110)은 더 좌측의 플랭킹 서열분석 사이클(N-2) 동안 생성되고, 우측의 플랭킹 압축된 출력(140)은 우측의 플랭킹 서열분석 사이클(N+1) 동안 생성되고, 더 우측의 플랭킹 압축된 출력(150)은 더 우측의 플랭킹 서열분석 사이클(N+2) 동안 생성된다.

[0079] 파이프라인 관점으로부터, 신경 네트워크 기반 염기 호출자(100)는 공간 로직(104) 및 압축 로직(108)을 통해(예컨대, 데이터 병렬성에 기초한 다중 스레드 실행 또는 다중 클러스터 실행으로서) 이미지 패치들(102, 112, 122, 132, 142)을 각각 프로세싱하는 5개의 병렬 및 독립적 파이프라인들을 실행한다. 따라서, 5개의 압축된 출력들(110, 120, 130, 140, 150)은 신경 네트워크 기반 염기 호출자(100)에 의해 개별적으로, 동시에, 그리고 독립적으로 생성된다.

[0080] 일부 구현예들에서, 압축 층(108)은 공간 네트워크(104)의 최종 공간 콘볼루션 층으로 간주될 수 있다. 다른 구현예들에서, 압축 네트워크(108)는 별개의 네트워크를, 신경 네트워크 기반 염기 호출자(100)의 특화된 아키텍처 내부에 또는 외부에 있는 것으로 간주될 수 있다.

[0081] 도 1b는 신경 네트워크 기반 염기 호출자(100)의 시간 로직(160)(또는 시간 네트워크 또는 시간 서브네트워크 또는 시간 콘볼루션 신경 네트워크)을 통해 압축된 공간 맵 세트들(110, 120, 130, 140, 150)을 프로세싱하는 하나의 구현예를 도시한다. 시간 로직(160)은 슬라이딩 윈도우 단위로 연속적인 압축된 공간 맵 세트들의 그룹들을 프로세싱한다. 예를 들어, 도 1b에서, 시간 로직(160)은 각자의 서열분석 사이클들 1, 2, 및 3 동안 압축된 공간 맵 세트들(110, 120, 130)의 제1 그룹/윈도우를 프로세싱하고 시간 맵들(172)(또는 시간 맵 세트들 또는 시간 특징부 맵들 또는 시간 특징부 맵 세트들)을 출력으로서 생성한다. 시간 로직(160)은 각자의 서열분석 사이클들 2, 3, 및 4 동안 압축된 공간 맵 세트들(120, 130, 140)의 제2 그룹/윈도우를 프로세싱하고 시간 맵들(174)을 출력으로서 생성한다. 시간 로직(160)은 각자의 서열분석 사이클들 3, 4, 및 5 동안 압축된 공간 맵 세트들(130, 140, 150)의 제3 그룹/윈도우를 프로세싱하고 시간 맵들(176)을 출력으로서 생성한다. 시간 콘볼루션 네트워크(160)는 1D, 2D, 또는 3D 콘볼루션들을 사용할 수 있다.

[0082] 도 1b에 도시된 시간 로직(160)의 3개의 인스턴스들은 시간 네트워크(160)의 제1 시간 콘볼루션 층의 3개의 필터 뱅크들을 표현한다. 제1 필터 뱅크는 압축된 공간 맵들(110, 120, 130)의 제1 그룹 상에 시간 콘볼루션 필터들의 제1 세트를 적용하고, 시간 맵들(172)의 제1 세트를 생성한다. 제2 필터 뱅크는 압축된 공간 맵들(120, 130, 140)의 제2 그룹 상에 시간 콘볼루션 필터들의 제2 세트를 적용하고, 시간 맵들(174)의 제2 세트를 생성한다. 제3 필터 뱅크는 압축된 공간 맵들(130, 140, 150)의 제3 그룹 상에 시간 콘볼루션 필터들의 제3 세트를 적용하고, 시간 맵들(176)의 제3 세트를 생성한다.

[0083] 시간 맵들(172, 174, 176)의 제1, 제2 및 제3 세트들은 시간 로직(160)에 의해 그룹으로서 프로세싱되어 시간 맵들(182)을 생성한다. 도 1b에 도시된 시간 로직(160)의 제4 인스턴스는 도 1a의 신경 네트워크 기반 염기 호출자(100)에 대한 입력으로서 사이클별 이미지 패치 쌍들이 공급된 모든 서열분석 사이클들 1, 2, 3, 4, 및 5 동안 출력을 생성하는 시간 네트워크(160)의 제2 시간 콘볼루션 층을 표현한다.

[0084] 시간 네트워크(160)는 시간 콘볼루션 층들의 캐스케이드를 갖는다(예컨대, 2개, 3개, 4개, 5개 또는 그 이상의 시간 콘볼루션 층들이 시퀀스로 배열됨). 시간 콘볼루션 층들의 캐스케이드는 상이한 레벨들의 그룹화를 갖는 계층구조 형태로 데이터를 프로세싱한다. 즉, 주어진 레벨에서, 슬라이딩 윈도우 접근법은 주어진 레벨에서의

입력들을 그룹별로 프로세싱하여, 후속적으로 슬라이딩 윈도우 방식으로 다음 레벨에서 그룹별로 프로세싱되는 출력들을 생성한다.

[0085] 시간 콘볼루션 층들은 조합형 콘볼루션들을 구현하는 시간 콘볼루션 필터들로 구성된다. 조합형 콘볼루션들은 다수의 서열분석 사이클들에 걸쳐있는 특징부 맵들 사이의 정보를 혼합한다. 조합형 콘볼루션들은 시간 네트워크(160) 내의 현재 레벨에서의 대상 그룹/윈도우 내의 연속 서열분석 사이클들 사이의 데이터를 조합한다. 예를 들어, 제1 시간 콘볼루션 층은 서열분석 사이클들 1, 2, 및 3의 제1 그룹 동안 압축된 공간 맵들(110, 120, 130)의 제1 그룹을 조합하여 시간 맵들(172)의 제1 세트를 생성하고; 서열분석 사이클들 2, 3, 및 4의 제2 그룹 동안 압축된 공간 맵들(120, 130, 140)의 제2 그룹을 조합하여 시간 맵들(174)의 제2 세트를 생성하고; 서열분석 사이클들 3, 4, 및 5의 제3 그룹 동안 압축된 공간 맵들(130, 140, 150)의 제3 그룹을 조합하여 시간 맵들(176)의 제3 세트를 생성한다.

[0086] 조합형 콘볼루션들은 또한, 시간 네트워크(160) 내의 현재 레벨에서의 대상 그룹/윈도우 내의 서열분석 사이클들의 연속 그룹들 사이의 데이터를 조합한다. 예를 들어, 제2 시간 콘볼루션 층은 시간 맵들(172, 174, 176)의 제1, 제2 및 제3 세트들을 시간 맵들(182)의 최종 세트로 조합한다. 레벨 2에서는, 레벨 1로부터의 서열분석 사이클들의 제1, 제2, 및 제3 그룹들/윈도우들이 서열분석 사이클들 1, 2, 3, 4, 및 5의 제1 그룹/윈도우 내에 그룹화된다.

[0087] 조합형 콘볼루션들은 조합될 입력들의 개수만큼 많은 커널들로 구성된다(즉, 시간 콘볼루션 필터들의 깊이 컬럼 또는 파이버(fibre)는 현재 레벨에서의 대상 그룹/윈도우 내의 입력들의 개수와 매칭됨). 예를 들어, 시간 콘볼루션 층이 3개의 압축된 공간 맵들을 조합할 때, 그것은 시간 콘볼루션 필터들을 사용하는데, 이들은 각각, 3개의 압축된 공간 맵들의 깊이 전체에 걸쳐 요소별 곱셈 및 합산을 수행하는 3개의 커널들을 갖는다.

[0088] 시간 맵들(182)의 최종 세트는 시간 네트워크(160)의 최종(또는 마지막) 시간 콘볼루션 층에 의해 생성된다. 도 1c는 개시된 출력 로직(190)(또는 출력 층 또는 출력 네트워크 또는 출력 서브네트워크)을 통해 최종 시간 맵 세트들(182)을 프로세싱하여 염기 호출 분류 데이터를 생성하는 하나의 구현예를 도시한다. 하나의 구현예에서, 복수의 클러스터들은 하나 이상의 서열분석 사이클들 동안 동시에 염기 호출된다. 도 1c에 도시된 예에서, 염기 호출들(192)은 중앙 서열분석 사이클 3동안만 많은 클러스터들에 대해 생성된다. 다른 구현예들에서, 개시된 기술은, 하나의 구현예에 따라, 입력의 주어진 윈도우에 대해, 출력 로직(190)이 중앙 서열분석 사이클 동안뿐만 아니라 플랭킹 서열분석 사이클들(선택적 파선들로 표시된 바와 같음) 동안에도 염기 호출들을 생성하게 한다. 예를 들어, 하나의 구현예에서, 개시된 기술은 주어진 입력 윈도우 동안 사이클 N, 사이클 N+1, 사이클 N-1, 사이클 N+2, 사이클 N-2 등 동안 동시에 염기 호출들을 생성한다. 즉, 신경 네트워크 기반 염기 호출자(102)의 단일 순방향 전파/순회/염기 호출 반복은 서열분석 사이클들의 입력 윈도우 내의 다수의 서열분석 사이클들 동안 염기 호출들을 생성하는데, 이는 본 명세서에서 "다-대-다(many-to-many) 염기 호출"로 지칭된다.

[0089] 출력 층(190)의 예들은 소프트맥스 함수, 로그-소프트맥스 함수, 앙상블 출력 평균 함수, 다층 퍼셉트론 불확실성 함수, 베이즈 가우시안(Bayes Gaussian) 분포 함수, 및 클러스터 세기 함수를 포함한다. 하나의 구현예에서, 출력 층(190)은 각각의 클러스터에 대해 그리고 각각의 서열분석 사이클 동안 클러스터별, 사이클별 확률 쿼드러플(quadruple)을 생성한다.

[0090] 하기의 논의는 일례로서 소프트맥스 함수를 사용한 클러스터별, 사이클별 확률 쿼드러플들에 중점을 둔다. 먼저 소프트맥스 함수를 설명하고, 이어서, 클러스터별, 사이클별 확률 쿼드러플들을 설명한다.

[0091] 소프트맥스 함수는 다중 클래스 분류를 위한 바람직한 함수이다. 소프트맥스 함수는 모든 가능한 표적 클래스들에 걸쳐 각각의 표적 클래스의 확률들을 계산한다. 소프트맥스 함수의 출력 범위는 0과 1 사이이고, 모든 확률들의 합은 1과 동일하다. 소프트맥스 함수는 주어진 입력 값의 지수 및 모든 입력 값들의 지수 값들의 합을 계산한다. 입력 값의 지수와 지수 값들의 합의 비는 소프트맥스 함수의 출력이고, 이는 본 명세서에서 "지수 정규화(exponential normalization)"로 지칭된다.

[0092] 공식적으로, 소위 소프트맥스 분류기를 트레이닝시키는 것은 진정한 분류기라기보다는 클래스 확률로의 회귀이며, 이는 그것이 클래스를 반환하는 것이 아니라 오히려 각각의 클래스의 확률의 신뢰도 예측이기 때문이다. 소프트맥스 함수는 값들의 클래스를 취하고, 이들을 1로 합산하는 확률들로 변환한다. 소프트맥스 함수는 0 내지 1의 범위 내의 실제 값들의 n-차원 벡터에 대한 임의의 실제 값들의 n-차원 벡터를 스퀴싱한다. 따라서, 소프트맥스 함수를 사용하는 것은, 출력이 유효한 지수적으로 정규화된 확률 질량 함수(비-음성이고, 1로 합산함)인 것을 보장한다.

[0093] 직관적으로, 소프트맥스 함수는 최대 함수의 "소프트" 버전이다. 용어 "소프트"는 소프트맥스 함수가 연속적이고 구별가능하다는 사실로부터 유래한다. 하나의 최대 요소를 선택하는 대신, 그것은 벡터를 전체의 부분들로 분해하며, 이때 최대 입력 요소는 비례적으로 더 큰 값을 얻고 다른 것은 더 적은 비율의 값을 얻게 된다. 확률 분포를 출력하는 속성은 소프트맥스 함수를 분류 태스크들에서 확률적 해석에 적합하게 만든다.

[0094] z를 소프트맥스 층에 대한 입력들의 벡터로서 고려한다. 소프트맥스 층 유닛들은 소프트맥스 층의 노드들의 개수이고, 따라서, z 벡터의 길이는 소프트맥스 층 내의 유닛들의 개수이다(10개의 출력 유닛들을 갖는 경우, 10개의 z 요소들이 있음).

[0095] n-차원 벡터 $Z=[z_1,z_2,...z_n]$의 경우, 소프트맥스 함수는 지수 정규화(exp)를 사용하여, 정규화된 값들이 범위 [0, 1] 내에 있고 1을 가산하는 다른 n-차원 벡터 $p(Z)$ 를 생성한다:

[0096] $$Z=\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \text{ 및 } p(Z) \rightarrow \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix}$$

[0097] $$p_j=\frac{\exp^{z_j}}{\sum_{k=1}^{n}\exp^{z_k}} \forall\ j \in 1,2,...,n$$

[0098] 도 1g는 예시적인 소프트맥스 함수를 도시한다. 소프트맥스 함수는 $z \mapsto \text{소프트맥스}([z;\frac{z}{10};-2z])$ 로서 3개의 클래스들에 적용된다. 3개의 출력들이 항상 1로 합산된다는 것에 유의한다. 따라서, 이들은 이산 확률 질량 함수를 한정한다.

[0099] 특정 클러스터별, 사이클별 확률 쿼드러플은 특정 서열분석 사이클에서 특정 클러스터에 혼입된 염기가 A, C, T, 및 G일 확률들을 식별한다. 신경 네트워크 기반 염기 호출자(100)의 출력 층이 소프트맥스 함수를 사용할 때, 클러스터별, 사이클별 확률 쿼드러플에서의 확률들은 1을 합산하는 지수적으로 정규화된 분류 스코어들이다. 도 1h는 클러스터 1(121, 갈색으로 도시됨)에 대해 그리고 서열분석 사이클들 1 내지 S(122) 동안 각각 소프트맥스 함수에 의해 생성된 예시적인 클러스터별, 사이클별 확률 쿼드러플들(123)을 도시한다. 다시 말해서, 서열분석 사이클들의 제1 서브세트는 S개의 서열분석 사이클들을 포함한다.

[0100] 신뢰성없는 클러스터 식별자(125)는 클러스터별, 사이클별 확률 쿼드러플로부터 필터 값들을 생성한 것에 기초하여 신뢰성없는 클러스터들을 식별한다. 본 출원에서, 클러스터별, 사이클별 확률 쿼드러플들은 또한, 염기 호출 분류 스코어들 또는 정규화된 염기 호출 분류 스코어들 또는 초기 염기 호출 분류 스코어들 또는 정규화된 초기 염기 호출 분류 스코어들 또는 초기 염기 호출들로 지칭된다.

[0101] 필터 계산기(127)는 그것이 식별하는 확률들에 기초하여 각각의 클러스터별, 사이클별 확률 쿼드러플에 대한 필터 값을 결정하여, 이에 의해, 각각의 클러스터에 대한 필터 값들의 시퀀스를 생성한다. 필터 값들의 시퀀스는 필터 값들(124)로서 저장된다.

[0102] 클러스터별, 사이클별 확률 쿼드러플에 대한 필터 값은 확률들 중 하나 이상을 수반하는 계산에 기초하여 결정된다. 하나의 구현예에서, 필터 계산기(127)에 의해 사용되는 계산은 뺄셈이다. 예를 들어, 도 1h에 도시된 구현예에서, 클러스터별, 사이클별 확률 쿼드러플에 대한 필터 값은 확률들 중 가장 높은 확률(자홍색으로 도시됨)로부터 확률들 중 두 번째로 가장 높은 확률(청색으로 도시됨)을 감산함으로써 결정된다.

[0103] 다른 구현예에서, 필터 계산기(116)에 의해 사용되는 계산은 나눗셈이다. 예를 들어, 클러스터별, 사이클별 확률 쿼드러플에 대한 필터 값은 확률들 중 가장 높은 확률(자홍색으로 도시됨) 대 확률들 중 두 번째로 가장 높은 확률(청색으로 도시됨)의 비로서 결정된다. 또 다른 구현예에서, 필터 계산기(127)에 의해 사용되는 계산은 덧셈이다. 또 다른 추가 구현예에서, 필터 계산기(127)에 의해 사용되는 계산은 곱셈이다.

[0104] 하나의 구현예에서, 필터 계산기(127)는 필터링 함수를 사용하여 필터 값들(124)을 생성한다. 하나의 예에서,

필터링 함수는 가장 밝은 염기 세기를 가장 밝은 염기 세기 및 두 번째로 가장 밝은 염기 세기의 합으로 나눈 비로서 순도(chastity)를 정의하는 순도 필터이다. 다른 예에서, 필터링 함수는 최대 로그 확률 함수, 최소 제곱 오차 함수, 평균 신호대잡음비(signal-to-noise ratio, SNR), 및 최소 절대 오차 함수 중 적어도 하나이다.

[0105] 신뢰성없는 클러스터 식별자(125)는 필터 값들(124)을 사용하여, 복수의 클러스터들 중의 일부 클러스터들을 신뢰성없는 클러스터들(128)로서 식별한다. 신뢰성없는 클러스터들(128)을 식별하는 데이터가 컴퓨터 판독가능 포맷 또는 매체에 있을 수 있다. 신뢰성없는 클러스터들은 기구 ID, 기구 상의 실행 횟수, 플로우 셀 ID, 레인 번호, 타일 번호, 클러스터의 X 좌표, 클러스터의 Y 좌표, 및 고유 분자 식별자(unique molecular identifier, UMI)들에 의해 식별될 수 있다. 신뢰성없는 클러스터 식별자(125)는 복수의 클러스터들 내의 그 클러스터들을 신뢰성없는 클러스터들로서 식별하고, 여기서 그들의 필터 값들의 시퀀스들은 임계치 "H" 미만의 "G"개의 필터 값들을 포함한다. 하나의 구현예에서, "G"는 1 내지 5의 범위이다. 다른 구현예에서, "H"는 0.5 내지 0.99의 범위이다. 하나의 구현예에서, 신뢰성없는 클러스터들(128)은 신뢰성없는 클러스터들에 대응하는(즉, 이들의 세기 방출들을 묘사하는) 그 픽셀들을 식별한다. 그러한 픽셀들은 이후에 본 출원에서 설명되는 바와 같이, 필터링 로직(502)에 의해 필터링된다.

[0106] 신뢰성없는 클러스터들은 백그라운드 신호와 비교하여 미미한 일정 양의 원하는 신호를 방출하는 저품질 클러스터들이다. 신뢰성없는 클러스터들에 대한 신호대잡음비는 실질적으로 낮은데, 예를 들어 1 미만이다. 일부 구현예들에서, 신뢰성없는 클러스터들은 어떠한 양의 원하는 신호도 생성하지 않을 수 있다. 다른 구현예들에서, 신뢰성없는 클러스터들은 배경에 비해 매우 적은 양의 신호를 생성할 수 있다. 하나의 구현예에서, 신호는 광학 신호이고, 예를 들어 형광, 발광, 산란, 또는 흡수 신호들을 포함하도록 의도된다. 신호 레벨은 원하는 또는 미리정의된 특성을 갖는 검출된 에너지 또는 코딩된 정보의 양 또는 분량을 지칭한다. 예를 들어, 광학 신호는 세기, 파장, 에너지, 주파수, 전력 휘도 등 중 하나 이상에 의해 정량화될 수 있다. 다른 신호들은 전압, 전류, 전기장 강도, 자기장 강도, 주파수, 전력, 온도 등과 같은 특성들에 따라 정량화될 수 있다. 신뢰성없는 클러스터들 내에서의 신호의 부재는 0의 신호 레벨 또는 잡음과 유의하게 구별되지 않는 신호 레벨인 것으로 이해된다.

[0107] 신뢰성없는 클러스터들의 열악한 품질 신호들에 대한 많은 잠재적인 이유들이 있다. 신뢰성없는 클러스터 내의 대략 1000개의 분자들 중 상당한 비율이 소정 위치에서 상이한 염기를 함유하도록 콜로니 증폭 시에 폴리머라제 사슬 반응(polymerase chain reaction, PCR) 에러가 있었다면, 2개의 염기들에 대한 신호를 관찰할 수 있다 - 이것은 불량한 품질의 표시로서 해석되고, 페이즈 에러(phase error)로 지칭될 수 있음. 페이즈 에러는, 신뢰성없는 클러스터 내의 개별 분자들이 (예컨대, 페이징(phasing)으로 명명된, 3' 종결자들의 불완전한 제거 때문에) 일부 사이클에서 뉴클레오티드를 혼입하지 않고 이어서 다른 분자들보다 뒤떨어질 때, 또는 개별 분자가 (프리-페이징(prephasing)으로 명명된, 유효 3'-블로킹이 없는 뉴클레오티드들의 혼입 때문에) 단일 사이클에서 1개 초과의 뉴클레오티드를 혼입할 때 발생한다. 이는 서열 복사본들의 판독 시에 동시성(synchrony)의 손실을 초래한다. 페이징 및 프리-페이징에 의해 영향을 받는 신뢰성없는 클러스터들 내의 서열들의 비율은 사이클 개수에 따라 증가하는데, 이는 높은 사이클 개수들에서 판독물들의 품질이 저하하는 경향이 있는 주요 이유이다.

[0108] 신뢰성없는 클러스터들은 또한 페이딩(fading)으로부터 기인한다. 페이딩은 사이클 개수의 함수로서 신뢰성없는 클러스터들의 신호 세기의 지수 감쇠이다. 서열분석 런이 진행됨에 따라, 신뢰성없는 클러스터들 내의 가닥들은 과도하게 세척되고, 반응성 종들을 생성하는 레이저 방출들에 노출되고, 가혹한 환경 조건들의 대상이 된다. 이들 모두는 신뢰성없는 클러스터들에서 단편들의 점진적인 손실을 초래하여, 그들의 신호 세기를 감소시킨다.

[0109] 신뢰성없는 클러스터들은 또한, 미개발 콜로니들, 즉, 패턴화된 플로우 셀 상의 비어 있는 또는 부분적으로 충전된 웰(well)들을 생성하는 작은 클러스터 크기들의 신뢰성없는 클러스터들로부터 기인한다. 즉, 일부 구현예들에서, 신뢰성없는 클러스터들은 패턴화된 플로우 셀 상의 비어 있는 웰, 다클론성(polyclonal) 웰, 및 디밍 웰을 나타낸다. 신뢰성없는 클러스터들은 또한, 비배타적인 증폭에 의해 야기된 중첩 콜로니들로부터 기인한다. 신뢰성없는 클러스터들은 또한, 예를 들어, 플로우 셀의 에지들 상에 위치된 것으로 인한, 불충분한 조명(under-illumination) 또는 불균일한 조명으로부터 기인한다. 신뢰성없는 클러스터들은 또한, 방출된 신호를 알기 어렵게 만드는 플로우 셀 상의 불순물들로부터 기인한다. 신뢰성없는 클러스터들은 또한, 다수의 클러스터들이 동일한 웰에 침착될 때 다클론성 클러스터들을 포함한다.

[0110] 서열분석 사이클들의 제1 윈도우는 서열분석 사이클들 1, 2, 3, 4 및 5를 포함하고, 염기 호출의 제1 반복은 중앙 서열분석 사이클 3 동안 염기 호출들(192)을 생성한다. 서열분석 사이클들의 제2 윈도우는 서열분석 사이클

들 2, 3, 4, 5 및 6을 포함하고, 염기 호출의 제2 반복은 중앙 서열분석 사이클 4 동안 염기 호출들(292)을 생성한다. 따라서, 서열분석 사이클들 2, 3, 4 및 5는 제1 윈도우와 제2 윈도우 사이의 또는 염기 호출의 제2 반복과 제3 반복 사이의 중첩 서열분석 사이클들이다.

[0111] 개시된 염기 호출 시스템들 및 기법들은, 메모리(예컨대, 온-칩 DRM, 온-칩 SRAM 또는 BRAM, 오프-칩 DRAM) 내에, 각자의 서열분석 사이클들 2, 3, 4, 및 5 동안의 염기 호출의 제1 반복 동안 생성된 압축된 공간 맵 세트들(120, 130, 140, 150)을 저장한다. 염기 호출의 제2 반복 동안, 개시된 염기 호출 시스템들 및 기법들은 공간 네트워크(104)를 통해 중첩 사이클들 2, 3, 4 및 5 동안 각자의 입력 이미지 패치들(112, 122, 132, 142)을 재프로세싱하지 않는다. 그 대신, 염기 호출의 제2 반복 동안, 개시된 염기 호출 시스템들 및 기법들은 각자의 입력 이미지 패치들(112, 122, 132, 142) 대신에 이전에 생성된 압축된 공간 맵 세트들(120, 130, 140, 150)을 재사용한다.

[0112] 압축 로직은 동일한 개수의 사이클별 특징부 맵들/채널들을 가질 것을 압축된 공간 맵 세트들(120, 130, 140, 150) 및 각자의 입력 이미지 패치들(112, 122, 132, 142)에게 요구하도록 추가로 구성된다. 이는, 압축된 공간 맵 세트들(120, 130, 140, 150)이 각자의 입력 이미지 패치들(112, 122, 132, 142)의 무손실 대표들이라는 것을 보장한다. 즉, 각자의 입력 이미지 패치들(112, 122, 132, 142)이 각각, 2개의 특징부 맵들/채널들을 갖는 경우, 압축 로직(108)은 압축된 공간 맵 세트들(120, 130, 140, 150)을, 또한 2개의 특징부 맵들/채널들을 갖도록 구성한다. 유사하게, 각자의 입력 이미지 패치들(112, 122, 132, 142)이 각각, 3개의 특징부 맵들/채널들을 갖는 경우, 압축 로직(108)은 압축된 공간 맵 세트들(120, 130, 140, 150)을, 또한 3개의 특징부 맵들/채널들을 갖도록 구성한다. 동일한 맥락에서, 각자의 입력 이미지 패치들(112, 122, 132, 142)이 각각, 4개의 특징부 맵들/채널들을 갖는 경우, 압축 로직(108)은 압축된 공간 맵 세트들(120, 130, 140, 150)을, 또한 4개의 특징부 맵들/채널들을 갖도록 구성한다.

[0113] 도 2a는 염기 호출의 제2 반복 동안, 공간 로직(104) 및 압축 로직(108)을 통해 입력 이미지 데이터(222)를 프로세싱함으로써, 공간 맵들(226) 및 대응하는 압축된 공간 맵들(230)이 비중첩 서열분석 사이클 6 동안만 생성되는 것을 도시한다. 따라서, 중첩 사이클들 2, 3, 4, 및 5동안의 입력 이미지 패치들(112, 122, 132, 142)(범례에서 회색 채움으로 강조됨)은 리던던트 콘볼루션들을 회피하기 위해 재프로세싱되지 않는다.

[0114] 도 2b는 염기 호출의 제1 반복 동안 생성된 압축된 공간 맵 세트들(120, 130, 140, 150)이 염기 호출의 제2 반복 동안 생성된 압축된 공간 맵 세트(230)와 함께 사용되어 중앙 서열분석 사이클 4 동안 염기 호출들(292)을 생성하는 것을 도시한다. 도 2b에서, 시간 맵 세트들(174, 176, 278)은 도 1b와 관련하여 상기에서 논의된 방식과 유사한 방식으로 시간 네트워크(160)의 제1 시간 콘볼루션 층에 의해 생성된다. 시간 맵 세트(282)는 도 1b와 관련하여 상기에서 논의된 방식과 유사한 방식으로 시간 네트워크(160)의 제2 및 마지막 시간 콘볼루션 층에 의해 생성된다. 도 2c는 출력 층(190)이 염기 호출의 제2 반복 동안 생성된 최종 시간 맵 세트(282)를 프로세싱하고 중앙 서열분석 사이클 4 동안 염기 호출들(292)을 생성하는 것을 도시한다.

[0115] 서열분석 사이클들의 제3 윈도우는 서열분석 사이클들 3, 4, 5, 6 및 7을 포함하고, 염기 호출의 제3 반복은 중앙 서열분석 사이클 5동안 염기 호출들(392)을 생성한다. 따라서, 서열분석 사이클들 3, 4, 5 및 6은 제2 윈도우와 제3 윈도우 사이의 또는 염기 호출의 제2 반복과 제3 반복 사이의 중첩 서열분석 사이클들이다.

[0116] 개시된 염기 호출 시스템들 및 기법들은, 메모리(예컨대, 온-칩 DRM, 온-칩 SRAM 또는 BRAM, 오프-칩 DRAM) 내에, 각자의 서열분석 사이클들 3, 4, 및 5 동안의 염기 호출의 제1 반복 동안 생성된 압축된 공간 맵 세트들(130, 140, 150), 및 서열분석 사이클 6 동안의 염기 호출의 제2 반복 동안 생성된 압축된 공간 맵 세트(230)를 저장한다. 염기 호출의 제3 반복 동안, 개시된 염기 호출 시스템들 및 기법들은 공간 네트워크(104)를 통해 중첩 사이클들 3, 4, 5 및 6 동안 각자의 입력 이미지 패치들(122, 142, 132, 222)을 재프로세싱하지 않는다. 그 대신, 염기 호출의 제3 반복 동안, 개시된 염기 호출 시스템들 및 기법들은 각자의 입력 이미지 패치들(122, 132, 142, 222) 대신에 이전에 생성된 압축된 공간 맵 세트들(130, 140, 150, 230)을 재사용한다.

[0117] 도 3a는 염기 호출의 제3 반복 동안, 공간 로직(104) 및 압축 로직(108)을 통해 입력 이미지 데이터(322)를 프로세싱함으로써, 공간 맵들(326) 및 대응하는 압축된 공간 맵들(330)이 비중첩 서열분석 사이클 7 동안만 생성되는 것을 도시한다. 따라서, 중첩 사이클들 3, 4, 5, 및 6 동안의 입력 이미지 패치들(122, 132, 142, 222)(범례에서 회색 채움으로 강조됨)은 리던던트 콘볼루션들을 회피하기 위해 재프로세싱되지 않는다.

[0118] 도 3b는 염기 호출의 제1 반복 및 제2 반복 동안 생성된 압축된 공간 맵 세트들(130, 140, 150, 230)이 염기 호출의 제3 반복 동안 생성된 압축된 공간 맵 세트(330)와 함께 사용되어 중앙 서열분석 사이클 5 동안 염기 호출

들(392)을 생성하는 것을 도시한다. 도 3b에서, 시간 맵 세트들(176, 278, 378)은 도 1b와 관련하여 상기에서 논의된 방식과 유사한 방식으로 시간 네트워크(160)의 제1 시간 콘볼루션 층에 의해 생성된다. 시간 맵 세트(382)는 도 1b와 관련하여 상기에서 논의된 방식과 유사한 방식으로 시간 네트워크(160)의 제2 및 마지막 시간 콘볼루션 층에 의해 생성된다. 도 3c는 출력 층(190)이 염기 호출의 제3 반복 동안 생성된 최종 시간 맵 세트(382)를 프로세싱하고 중앙 서열분석 사이클 5 동안 염기 호출들(392)을 생성하는 것을 도시한다.

[0119] 주어진 서열분석 사이클에 대해 일단 생성된 압축된 공간 맵 세트는 임의의 후속 서열분석 사이클을 염기 호출하기 위해 재사용될 수 있다. 도 4a는 중앙 서열분석 사이클 16을 염기 호출하기 위한 염기 호출의 제14 반복을 도시한다. 도 4b는 서열분석 사이클들 1 내지 29 동안 이전에 생성된 압축된 공간 맵들이 중앙 서열분석 사이클 16을 염기 호출하기 위한 최종 시간 맵 세트(482)를 생성하는 데 사용된다는 것을 도시한다. 도 4c는 출력 층(190)이 염기 호출의 제14 반복 동안 생성된 최종 시간 맵 세트(482)를 프로세싱하고 중앙 서열분석 사이클 16 동안 염기 호출들(492)을 생성하는 것을 도시한다.

[0120] 도 5a는, 염기 호출의 제1 반복 동안, 각자의 서열분석 사이클들 1, 2, 3, 4 및 5 동안 각자의 압축된, 필터링된 공간 맵들(510, 520, 530, 540, 550)(신뢰성있는 클러스터들만을 묘사함)을 생성하기 위해 필터링 로직(502)을 사용하여 압축된 공간 맵 세트들(110, 120, 130, 140, 150)을 필터링하는 하나의 구현예를 도시한다. 상기에서 논의된 바와 같이, 신뢰성없는 클러스터 데이터(128)는, 신뢰성없는 클러스터들에 대응하는 압축된 공간 맵들 및 공간 맵들의 그 부분들(예컨대, 픽셀들)을 식별한다. 그러한 픽셀들은 예를 들어, 신뢰성없는 클러스터들의 위치 좌표들에 기초하여 식별될 수 있다.

[0121] 필터링 로직(502)은 신뢰성없는 클러스터들을 식별하는 데이터(128)를 사용하여, 신뢰성없는 클러스터들에 대응하는(즉, 그들의 세기 방출들을 묘사함) 압축된 공간 맵 세트들(110, 120, 130, 140, 150)로부터 그 픽셀들을 필터링(또는 폐기 또는 제거)한다. 일부 구현예들에서, 이는 압축된 공간 맵 세트들로부터 75%의 픽셀들이 폐기되는 결과를 가져오고, 이에 의해, 많은 비생산적 콘볼루션들을 방지한다.

[0122] 도 5a에서, 필터링된 시간 맵 세트들(572, 574, 576)(신뢰성있는 클러스터들만을 묘사함)은 중앙 서열분석 사이클 3을 염기 호출하기 위해 압축된, 필터링된 공간 맵들(510, 520, 530, 540, 550)로부터 생성된다. 필터링된 시간 맵 세트들(572, 574, 576)(신뢰성있는 클러스터들만을 묘사함)은 도 1b와 관련하여 상기에서 논의된 방식과 유사한 방식으로 시간 네트워크(160)의 제1 시간 콘볼루션 층에 의해 생성된다. 필터링된 시간 맵 세트(582)(신뢰성있는 클러스터들만을 묘사함)는 도 1b와 관련하여 상기에서 논의된 방식과 유사한 방식으로 시간 네트워크(160)의 제2 및 마지막 시간 콘볼루션 층에 의해 생성된다. 도 5b는 출력 층(190)이 염기 호출의 제1 반복 동안 생성된 최종 필터링된 시간 맵 세트(582)를 프로세싱하고 중앙 서열분석 사이클 3 동안 염기 호출들(592)을 생성하는 것을 도시한다.

[0123] 도 6a는, 염기 호출의 제2 반복 동안, 각자의 서열분석 사이클들 2, 3, 4, 5 및 6 동안 각자의 압축된, 필터링된 공간 맵들(520, 530, 540, 550, 650)(신뢰성있는 클러스터들만을 묘사함)을 생성하기 위해 필터링 로직(502)을 사용하여 압축된 공간 맵 세트들(120, 130, 140, 150, 230)을 필터링하는 하나의 구현예를 도시한다. 필터링 로직(502)은 신뢰성없는 클러스터들을 식별하는 데이터(128)를 사용하여, 신뢰성없는 클러스터들에 대응하는(즉, 그들의 세기 방출들을 묘사함) 압축된 공간 맵 세트들(120, 130, 140, 150, 230)로부터 그 픽셀들을 필터링(또는 폐기 또는 제거)한다.

[0124] 도 6a에서, 필터링된 시간 맵 세트들(574, 576, 676)(신뢰성있는 클러스터들만을 묘사함)은 중앙 서열분석 사이클 4를 염기 호출하기 위해 압축된, 필터링된 공간 맵들(520, 530, 540, 550, 650)로부터 생성된다. 필터링된 시간 맵 세트들(574, 576, 676)(신뢰성있는 클러스터들만을 묘사함)은 도 1b와 관련하여 상기에서 논의된 방식과 유사한 방식으로 시간 네트워크(160)의 제1 시간 콘볼루션 층에 의해 생성된다. 필터링된 시간 맵 세트(682)(신뢰성있는 클러스터들만을 묘사함)는 도 1b와 관련하여 상기에서 논의된 방식과 유사한 방식으로 시간 네트워크(160)의 제2 및 마지막 시간 콘볼루션 층에 의해 생성된다. 도 6b는 출력 층(190)이 염기 호출의 제2 반복 동안 생성된 최종 필터링된 시간 맵 세트(682)를 프로세싱하고 중앙 서열분석 사이클 4 동안 염기 호출들(692)을 생성하는 것을 도시한다.

[0125] 도 7a는, 염기 호출의 제3 반복 동안, 각자의 서열분석 사이클들 3, 4, 5, 6 및 7 동안 각자의 압축된, 필터링된 공간 맵들(530, 540, 550, 650, 750)(신뢰성있는 클러스터들만을 묘사함)을 생성하기 위해 필터링 로직(502)을 사용하여 압축된 공간 맵 세트들(130, 140, 150, 230, 330)을 필터링하는 하나의 구현예를 도시한다. 필터링 로직(502)은 신뢰성없는 클러스터들을 식별하는 데이터(128)를 사용하여, 신뢰성없는 클러스터들에 대응하는(즉, 그들의 세기 방출들을 묘사함) 압축된 공간 맵 세트들(130, 140, 150, 230, 330)로부터 그 픽셀들을 필

터링(또는 폐기 또는 제거)한다.

[0126] 도 7a에서, 필터링된 시간 맵 세트들(576, 676, 776)(신뢰성있는 클러스터들만을 묘사함)은 중앙 서열분석 사이클 5를 염기 호출하기 위해 압축된, 필터링된 공간 맵들(530, 540, 550, 650, 750)로부터 생성된다. 필터링된 시간 맵 세트들(576, 676, 776)(신뢰성있는 클러스터들만을 묘사함)은 도 1b와 관련하여 상기에서 논의된 방식과 유사한 방식으로 시간 네트워크(160)의 제1 시간 콘볼루션 층에 의해 생성된다. 필터링된 시간 맵 세트(782)(신뢰성있는 클러스터들만을 묘사함)는 도 1b와 관련하여 상기에서 논의된 방식과 유사한 방식으로 시간 네트워크(160)의 제2 및 마지막 시간 콘볼루션 층에 의해 생성된다. 도 7b는 출력 층(190)이 염기 호출의 제3 반복 동안 생성된 최종 필터링된 시간 맵 세트(782)를 프로세싱하고 중앙 서열분석 사이클 5 동안 염기 호출들(792)을 생성하는 것을 도시한다.

[0127] 다른 구현예들에서, 압축 로직(108)은 대응하는 압축된 시간 맵 세트들을, 4개 초과의 특징부 맵들을 각각 갖도록 구성할 수 있다.

[0128] 공간 특징부 맵들과 관련하여 상기에서 논의된 압축 로직(108)은 시간 로직(160)에 의해 생성된 시간 특징부 맵들의 압축에 동등하게 적용된다. 후속 서열분석 사이클들에서 일단 생성된 압축된 공간 특징부 맵들의 재사용은 또한, 후속 서열분석 사이클들에서 일단 생성된 압축된 시간 특징부 맵들을 재사용하는 것과 동등하게 적용된다.

[0129] 일부 구현예들에서, 압축된 시간 특징부 맵들을 재사용하는 것은 압축된 공간 특징부 맵들을 재사용하는 것에 비해 2자릿수의 효율 및 계산 절약을 초래하는데, 그 이유는 압축된 시간 특징부 맵들이 프로세싱 파이프라인의 나중의 스테이지에서 압축된 공간 특징부 맵들로부터 생성되기 때문이다. 추가 프로세싱 엔진(즉, 시간 네트워크(160))으로부터의 중간 결과들을 용도변경하는 것은 스킵될 수 있는 이전의 프로세싱 단계들의 개수를 증가시킨다. 즉, 압축된 공간 특징부 맵들을 재사용하는 것은 공간 네트워크(104)를 통한 오리지널 이미지 데이터의 리던던트 프로세싱을 제거하지만, 시간 네트워크(160)를 통한 압축된 공간 특징부 맵들의 프로세싱을 리던던트로 포함할 수 있다. 대조적으로, 압축된 시간 특징부 맵들을 재사용하는 것은 공간 네트워크(104)를 통한 오리지널 이미지 데이터의 리던던트 프로세싱 및 시간 네트워크(160)를 통한 압축된 공간 특징부 맵들의 리던던트 프로세싱 둘 모두를 제거한다.

[0130] 도 8a는 압축 로직(108)을 통해 염기 호출의 제1 반복 동안 생성된 시간 특징부 맵들(172, 174, 176)의 세트들을 프로세싱하여, 압축된 시간 특징부 맵들(802, 804, 806)의 각자의 세트들을 생성하는 하나의 구현예를 도시한다. 압축된 시간 특징부 맵 세트들(802, 804, 806)은 도 1e 및 도 1f와 관련하여 상기에서 논의된 방식과 유사한 방식으로 압축 로직(108)에 의해 생성된다. 즉, 예를 들어, 제1 시간 특징부 맵 세트(172)가, 예를 들어, 21개의 특징부 맵들(또는 채널들 또는 깊이 = 21)을 갖는 경우, 압축 로직(108)은 대응하는 압축된 시간 특징부 맵 세트(802)를 1개, 2개, 3개, 4개, 또는 특징부 맵들을 갖도록 구성할 수 있다. 압축된 시간 특징부 맵들(802, 804, 806)의 세트들은 도 1b와 관련하여 상기에서 논의된 방식과 유사한 방식으로 시간 네트워크(160)의 제2 시간 콘볼루션 층에 의해 프로세싱되어 최종 압축된 시간 특징부 맵 세트(814)를 생성한다. 도 8b는 출력 층(190)이 염기 호출의 제1 반복 동안 생성된 최종 압축된 시간 특징부 맵 세트(814)를 프로세싱하고 중앙 서열분석 사이클 3 동안 염기 호출들(892)을 생성하는 것을 도시한다.

[0131] 도 9a는 제2 염기 호출 반복 시에, 제1 염기 호출 반복 시에 생성된 그 압축된 시간 맵들을 재사용하는 하나의 구현예를 도시한다. 즉, 압축된 시간 맵들(804, 806)의 제1 및 제2 세트들이 도 8a에서 제1 염기 호출 반복 동안 생성되었고, 이제 도 9a 및 도 9b에 도시된 제2 염기 호출 반복 시에 용도변경된다.

[0132] 압축된 시간 맵들(804, 806)의 제1 및 제2 세트들은 도 8a에서 시간 맵들(172, 174)의 제1 및 제2 세트들로부터 생성되었다는 것에 유의한다. 추가로, 시간 맵들(172, 174)의 제1 및 제2 세트들은 도 1b에서 압축된 공간 맵들(110, 120, 130, 140)로부터 생성되었고, 이러한 압축된 공간 맵들은 이어서, 도 1a에서 각자의 공간 맵들(106, 116, 126, 136)로부터 생성되었고, 이러한 각자의 공간 맵들은 이어서, 도 1a에서 각자의 이미지 패치들(102, 112, 122, 132)로부터 생성되었다는 것에 유의한다.

[0133] 중첩 시간 맵들(172, 174, 176)을 리던던트로 생성하는 도 1b, 도 2b, 및 도 3b와는 달리, 도 9a에서, 중첩 시간 맵들(174, 176)(도 9a에서 파선들 및 흐릿한 텍스트로 묘사됨)은 도 8a(제1 염기 호출 반복)로부터 도 9a(제2 염기 호출 반복)로 리던던트로 생성되지 않는다. 이는, 압축 로직(108)이 시간 네트워크(160)에 통합되어 제1 염기 호출 반복 시에 압축된 시간 맵들(804, 806)의 제1 및 제2 세트들을 생성하기 때문에 - 이는 제2 염기 호출 반복 시에 중첩 시간 맵들(174, 176)을 대체함 - 발생한다. 압축된 시간 맵들은 메모리(예컨대, 온-칩

DRM, 온-칩 SRAM 또는 BRAM, 오프-칩 DRAM)에 저장될 수 있다.

[0134] 도 9a는 또한, 압축 로직(108)을 통해 비중첩 시간 맵들(278)(즉, 제1 염기 호출 반복과 제2 염기 호출 반복 사이의 비중첩)을 프로세싱하여 압축된 시간 맵들(906)을 생성하는 것을 도시한다. 압축된 시간 맵 세트들(804, 806, 906)은 도 1b와 관련하여 상기에서 논의된 방식과 유사한 방식으로 시간 네트워크(160)의 제2 시간 콘볼루션 층에 의해 프로세싱되어 최종 압축된 시간 특징부 맵 세트(914)를 생성한다. 도 9b는 출력 층(190)이 염기 호출의 제2 반복 동안 생성된 최종 압축된 시간 특징부 맵 세트(914)를 프로세싱하고 중앙 서열분석 사이클 4 동안 염기 호출들(992)을 생성하는 것을 도시한다.

[0135] 중첩 시간 맵들(174, 176, 278)을 리던던트로 생성하는 도 1b, 도 2b, 및 도 3b와는 달리, 도 10a에서, 중첩 시간 맵들(176, 278)(도 10a에서 파선들 및 흐릿한 텍스트로 묘사됨)은 도 9a(제2 염기 호출 반복)로부터 도 10a(제3 염기 호출 반복)로 리던던트로 생성되지 않는다. 이는, 압축 로직(108)이 시간 네트워크(160)에 통합되어 제1 및 제2 염기 호출 반복 시에 압축된 시간 맵들(806, 906)의 제1 및 제2 세트들을 생성하기 때문에 - 이는 제3 염기 호출 반복 시에 중첩 시간 맵들(176, 278)을 대체함 - 발생한다. 압축된 시간 맵들은 메모리(예컨대, 온-칩 DRM, 온-칩 SRAM 또는 BRAM, 오프-칩 DRAM)에 저장될 수 있다.

[0136] 도 10a는 또한, 압축 로직(108)을 통해 비중첩 시간 맵들(378)(즉, 제2 염기 호출 반복과 제3 염기 호출 반복 사이의 비중첩)을 프로세싱하여 압축된 시간 맵들(1006)을 생성하는 것을 도시한다. 압축된 시간 맵 세트들(806, 906, 1006)은 도 1b와 관련하여 상기에서 논의된 방식과 유사한 방식으로 시간 네트워크(160)의 제2 시간 콘볼루션 층에 의해 프로세싱되어 최종 압축된 시간 특징부 맵 세트(1014)를 생성한다. 도 10b는 출력 층(190)이 염기 호출의 제3 반복 동안 생성된 최종 압축된 시간 특징부 맵 세트(1014)를 프로세싱하고 중앙 서열분석 사이클 5 동안 염기 호출들(1092)을 생성하는 것을 도시한다.

[0137] 도 11a는 압축 로직(108)을 통해 염기 호출의 제1 반복 동안 생성된 필터링된 시간 특징부 맵들(572, 574, 576)의 세트들을 프로세싱하여, 압축된, 필터링된 시간 특징부 맵들(1102, 1104, 1306)의 각자의 세트들을 생성하는 하나의 구현예를 도시한다. 압축된, 필터링된 시간 특징부 맵 세트들(1102, 1104, 1106)(신뢰성있는 클러스터들만을 묘사함)은 도 1e 및 도 1f와 관련하여 상기에서 논의된 방식과 유사한 방식으로 압축 로직(108)에 의해 생성된다. 즉, 예를 들어, 제1 필터링된 시간 특징부 맵 세트(572)가, 예를 들어, 21개의 특징부 맵들(또는 채널들 또는 깊이 = 21)을 갖는 경우, 압축 로직(108)은 대응하는 압축된, 필터링된 시간 특징부 맵 세트(1102)를 1개, 2개, 3개, 또는 4개의 특징부 맵들을 갖도록 구성할 수 있다. 압축된, 필터링된 시간 특징부 맵들(1102, 1104, 1106)의 세트들은 도 1b와 관련하여 상기에서 논의된 방식과 유사한 방식으로 필터링된 시간 네트워크(160)의 제2 필터링된 시간 콘볼루션 층에 의해 프로세싱되어 최종 압축된, 필터링된 시간 특징부 맵 세트(1114)를 생성한다. 도 8b는 출력 층(190)이 염기 호출의 제1 반복 동안 생성된 최종 압축된, 필터링된 시간 특징부 맵 세트(1114)를 프로세싱하고 중앙 서열분석 사이클 3 동안 염기 호출들(1192)을 생성하는 것을 도시한다.

[0138] 다른 구현예들에서, 압축 로직(108)은 대응하는 압축된 시간 맵 세트들을, 4개 초과의 특징부 맵들을 각각 갖도록 구성할 수 있다.

[0139] 도 12a는 제2 염기 호출 반복 시에, 제1 염기 호출 반복 시에 생성된 그 압축된, 필터링된 시간 맵들을 재사용하는 하나의 구현예를 도시한다. 즉, 압축된, 필터링된 시간 맵들(1104, 1106)의 제1 및 제2 세트들이 도 11a에서 제1 염기 호출 반복 동안 생성되었고, 이제 도 12a 및 도 2b에 도시된 제2 염기 호출 반복 시에 용도변경된다.

[0140] 압축된, 필터링된 시간 맵들(1104, 1106)의 제1 및 제2 세트들은 도 11a에서 필터링된 시간 맵들(572, 574)의 제1 및 제2 세트들로부터 생성되었다는 것에 유의한다. 추가로, 필터링된 시간 맵들(572, 574)의 제1 및 제2 세트들은 도 5a에서 압축된, 필터링된 공간 맵들(510, 520, 530, 540)로부터 생성되었고, 이러한 압축된, 필터링된 공간 맵들은 이어서, 도 5a에서 각자의 압축된 공간 맵들(110, 120, 130, 140)로부터 생성되었고, 이러한 각자의 압축된 공간 맵들은 이어서, 도 1a에서 각자의 공간 맵들(106, 116, 126, 136)로부터 생성되었고, 이러한 각자의 공간 맵들은 이어서, 도 1a에서 각자의 이미지 패치들(102, 112, 122, 132)로부터 생성되었다는 것에 유의한다.

[0141] 중첩 필터링된 시간 맵들(572, 574, 576)을 리던던트로 생성하는 도 5a, 도 6a, 및 도 7a와는 달리, 도 12a에서, 중첩 필터링된 시간 맵들(574, 576)(도 12a에서 파선들 및 흐릿한 텍스트로 묘사됨)은 도 11a(제1 염기 호출 반복)로부터 도 12a(제2 염기 호출 반복)로 리던던트로 생성되지 않는다. 이는, 압축 로직(108)이 필

터링된 시간 네트워크(160)에 통합되어 제1 염기 호출 반복 시에 압축된, 필터링된 시간 맵들(1104, 1106)의 제1 및 제2 세트들을 생성하기 때문에 - 이는 제2 염기 호출 반복 시에 중첩 필터링된 시간 맵들(574, 576)을 대체함 - 발생한다. 압축된, 필터링된 시간 맵들은 메모리(예컨대, 온-칩 DRM, 온-칩 SRAM 또는 BRAM, 오프-칩 DRAM)에 저장될 수 있다.

[0142] 도 12a는 또한, 압축 로직(108)을 통해 비중첩 필터링된 시간 맵들(676)(즉, 제1 염기 호출 반복과 제2 염기 호출 반복 사이의 비중첩)을 프로세싱하여 압축된, 필터링된 시간 맵들(1206)(신뢰성있는 클러스터들만을 묘사함)을 생성하는 것을 도시한다. 압축된, 필터링된 시간 맵 세트들(1104, 1106, 1206)은 도 1b와 관련하여 상기에서 논의된 방식과 유사한 방식으로 필터링된 시간 네트워크(160)의 제2 필터링된 시간 콘볼루션 층에 의해 프로세싱되어 최종 압축된, 필터링된 시간 특징부 맵 세트(1214)를 생성한다. 도 12b는 출력 층(190)이 염기 호출의 제2 반복 동안 생성된 최종 압축된, 필터링된 시간 특징부 맵 세트(1214)를 프로세싱하고 중앙 서열분석 사이클 4 동안 염기 호출들(1292)을 생성하는 것을 도시한다.

[0143] 중첩 필터링된 시간 맵들(574, 576, 676)을 리던던트로 생성하는 도 5a, 도 6a, 및 도 7a와는 달리, 도 13a에서, 중첩 필터링된 시간 맵들(576, 676)(도 13a에서 파선들 및 흐릿한 텍스트로 묘사됨)은 도 12a(제2 염기 호출 반복)로부터 도 13a(제3 염기 호출 반복)로 리던던트로 생성되지 않는다. 이는, 압축 로직(108)이 필터링된 시간 네트워크(160)에 통합되어 제1 및 제2 염기 호출 반복들 시에 압축된, 필터링된 시간 맵들(1106, 1206)의 제1 및 제2 세트들을 생성하기 때문에 - 이는 제3 염기 호출 반복 시에 중첩 필터링된 시간 맵들(576, 676)을 대체함 - 발생한다. 압축된, 필터링된 시간 맵들은 메모리(예컨대, 온-칩 DRM, 온-칩 SRAM 또는 BRAM, 오프-칩 DRAM)에 저장될 수 있다.

[0144] 도 13a는 또한, 압축 로직(108)을 통해 비중첩 필터링된 시간 맵들(776)(즉, 제2 염기 호출 반복과 제3 염기 호출 반복 사이의 비중첩)을 프로세싱하여 압축된, 필터링된 시간 맵들(1306)(신뢰성있는 클러스터들만을 묘사함)을 생성하는 것을 도시한다. 압축된, 필터링된 시간 맵 세트들(1106, 1206, 1306)은 도 1b와 관련하여 상기에서 논의된 방식과 유사한 방식으로 필터링된 시간 네트워크(160)의 제2 필터링된 시간 콘볼루션 층에 의해 프로세싱되어 최종 압축된, 필터링된 시간 특징부 맵 세트(1314)를 생성한다. 도 13b는 출력 층(190)이 염기 호출의 제3 반복 동안 생성된 최종 압축된, 필터링된 시간 특징부 맵 세트(1314)를 프로세싱하고 중앙 서열분석 사이클 5 동안 염기 호출들(1392)을 생성하는 것을 도시한다.

[0145] 도 14는 신경 네트워크 기반 염기 호출자(100)의 제1 예시적인 아키텍처를 도시한다. 도시된 구현예에서, 신경 네트워크 기반 염기 호출자(100)는 공간 네트워크(104), 압축 네트워크(108), 및 시간 네트워크(160)를 포함한다. 공간 네트워크(104)는 7개의 공간 콘볼루션 층들을 포함한다. 압축 네트워크(108)는 압축 층을 포함한다. 시간 네트워크(160)는 2개의 시간 콘볼루션 층들을 포함한다.

[0146] 7개의 공간 콘볼루션 층들 각각은 동일한 개수의 콘볼루션 필터들을 가질 수 있거나, 또는 상이한 개수의 콘볼루션 필터들을 가질 수 있다. 제1 공간 콘볼루션 층은 S1개의 필터들을 가질 수 있으며, 여기서 S1은, 예를 들어, 7, 14, 21, 64, 128, 또는 254일 수 있다. 제2 공간 콘볼루션 층은 S2개의 필터들을 가질 수 있으며, 여기서 S2는, 예를 들어, 7, 14, 21, 64, 128, 또는 254일 수 있다. 제3 공간 콘볼루션 층은 S3개의 필터들을 가질 수 있으며, 여기서 S3은, 예를 들어, 7, 14, 21, 64, 128, 또는 254일 수 있다. 제4 공간 콘볼루션 층은 S4개의 필터들을 가질 수 있으며, 여기서 S4는, 예를 들어, 7, 14, 21, 64, 128, 또는 254일 수 있다. 제5 공간 콘볼루션 층은 S5개의 필터들을 가질 수 있으며, 여기서 S5는, 예를 들어, 7, 14, 21, 64, 128, 또는 254일 수 있다. 제6 공간 콘볼루션 층은 S6개의 필터들을 가질 수 있으며, 여기서 S6은, 예를 들어, 7, 14, 21, 64, 128, 또는 254일 수 있다. 제7 공간 콘볼루션 층은 S7개의 필터들을 가질 수 있으며, 여기서 S7은, 예를 들어, 7, 14, 21, 64, 128, 또는 254일 수 있다.

[0147] 압축 층은 C1개의 필터들을 가질 수 있으며, 여기서 C1은, 예를 들어, 1, 2, 3, 4, 또는 그 이상일 수 있다.

[0148] 2개의 시간 콘볼루션 층들 각각은 동일한 개수의 콘볼루션 필터들을 가질 수 있거나, 또는 상이한 개수의 콘볼루션 필터들을 가질 수 있다. 제1 시간 콘볼루션 층은 T1개의 필터들을 가질 수 있으며, 여기서 T1은, 예를 들어, 7, 14, 21, 64, 128, 또는 254일 수 있다. 제2 시간 콘볼루션 층은 T2개의 필터들을 가질 수 있으며, 여기서 T2는, 예를 들어, 7, 14, 21, 64, 128, 또는 254일 수 있다. 도 14는 또한, 신경 네트워크 기반 염기 호출자(100)의 층들 각각에 의해 생성된 특징부 맵들(1412)을 도시한다.

[0149] 도 15는 신경 네트워크 기반 염기 호출자(100)의 제2 예시적인 아키텍처를 도시한다. 도 15는 신경 네트워크 기반 염기 호출자(100)의 일부로서 필터링 로직(502)을 도시한다. 다른 구현예들에서, 필터링 로직(502)은 신

경 네트워크 기반 염기 호출자(100)의 일부가 아니다. 압축된 특징부 맵들 C1은 P1 x P2의 공간 디멘션수를 갖는다. 필터링 로직(502)은 신뢰성없는 클러스터들에 대응하는 압축된 특징부 맵들 C1 내의 그 픽셀들을 필터링하고, P3 x P4의 공간 디멘션수를 갖는 압축된, 필터링된 특징부 맵들 F1을 생성한다. 압축된, 필터링된 특징부 맵들 F1은 신뢰성있는 클러스터들만을 묘사한다. 하나의 구현예에서, 필터링 로직(502)은 압축된 특징부 맵들 C1 내의 픽셀들의 75%를 폐기하고, 따라서, P3은 P1의 25%이고 P4는 P2의 25%이다. 도 15는 또한, 신경 네트워크 기반 염기 호출자(100)의 층들 각각에 의해 생성된 특징부 맵들(1512)을 도시한다.

[0150] 도 16은 신경 네트워크 기반 염기 호출자(100)의 제3 예시적인 아키텍처를 도시한다. 도 16은 압축 네트워크(108)가 시간 네트워크(160)뿐만 아니라 공간 네트워크(104)의 출력들을 압축하는 데 사용된다는 것을 도시한다. 도 16은 또한, 신경 네트워크 기반 염기 호출자(100)의 층들 각각에 의해 생성된 특징부 맵들(1612)을 도시한다.

[0151] 도 17은 신경 네트워크 기반 염기 호출자(100)의 제4 예시적인 아키텍처를 도시한다. 도 17은 필터링 로직(502)이 공간 네트워크(104)의 압축된 출력들에 적용되어 시간 네트워크(160)로부터 압축되고 필터링된 시간 출력들을 생성한다는 것을 도시한다. 도 17은 또한, 신경 네트워크 기반 염기 호출자(100)의 층들 각각에 의해 생성된 특징부 맵들(1712)을 도시한다.

[0152] 도 18은 입력 데이터 내의 채널들의 개수에 따라 압축 층(108)에 콘볼루션 필터들의 카운트(또는 무수함)를 구성하는 필터 구성 로직(1804)의 하나의 구현예를 도시한다. 이는 압축된 특징부 맵들이 입력 데이터의 무손실 대표들일 수 있게 한다. 일부 구현예들에서, 입력 데이터는 후속 서열분석 사이클들에서의 재사용을 위해 대응하는 압축된 표현과 함께 메모리에 겹쳐쓰기될 수 있다.

[0153] 하나의 구현예에서, 각각의 사이클별 입력에서 단지 하나의 채널(1812)만(예컨대, 단지 하나의 이미지 채널만)을 포함하는 입력 데이터에 대해, 필터 구성 로직(1804)은 서열분석 사이클별로 단지 하나의 압축된 특징부 맵(1818)만을 생성하는 단지 하나의 콘볼루션 필터(1816)만으로 압축 층(108)을 구성한다. 다른 구현예에서, 각각의 사이클별 입력에서 2개의 채널들(1822)(예컨대, 청색 및 녹색 레이저들에 대응하는 서열분석 이미지들에서 청색 및 녹색 이미지 채널들과 같은 2개의 이미지 채널들)을 포함하는 입력 데이터에 대해, 필터 구성 로직(1804)은 서열분석 사이클별로 2개의 압축된 특징부 맵들(1828)을 생성하는 2개의 콘볼루션 필터들(1826)로 압축 층(108)을 구성한다. 또 다른 구현예에서, 각각의 사이클별 입력에서 3개의 채널들(1832)(예컨대, 3개의 이미지 채널들)을 포함하는 입력 데이터에 대해, 필터 구성 로직(1804)은 서열분석 사이클별로 3개의 압축된 특징부 맵들(1838)을 생성하는 3개의 콘볼루션 필터들(1836)로 압축 층(108)을 구성한다. 또 다른 추가 구현예에서, 각각의 사이클별 입력에서 4개의 채널들(1842)(예컨대, 뉴클레오티드 A, C, T, 및 G에 대응하는 서열분석 이미지들에서 A, C, T, 및 G 채널들과 같은 4개의 이미지 채널들)을 포함하는 입력 데이터에 대해, 필터 구성 로직(1804)은 서열분석 사이클별로 4개의 압축된 특징부 맵들(1848)을 생성하는 4개의 콘볼루션 필터들(1846)로 압축 층(108)을 구성한다. 다른 구현예들에서, 압축 로직(108)은 대응하는 압축된 시간 맵 세트들을, 4개 초과의 특징부 맵들을 각각 갖도록 구성할 수 있고, 따라서, 압축 층(108)을 위한 4개 초과의 필터들을 선택할 수 있다.

[0154] 도 19a 및 도 19b는 서열분석 시스템(1900A)의 하나의 구현예를 도시한다. 서열분석 시스템(1900A)은 구성가능한 프로세서(1946)를 포함한다. 구성가능한 프로세서(1946)는 본 명세서에 개시된 염기 호출 기법들을 구현한다. 서열분석 시스템은 "서열분석기"로도 지칭된다.

[0155] 서열분석 시스템(1900A)은 생물학적 또는 화학적 물질 중 적어도 하나에 관련된 임의의 정보 또는 데이터를 획득하도록 동작할 수 있다. 일부 구현예들에서, 서열분석 시스템(1900A)은 벤치탑(bench-top) 디바이스 또는 데스크톱 컴퓨터와 유사할 수 있는 워크스테이션이다. 예를 들어, 원하는 반응들을 이행하기 위한 시스템들 및 컴포넌트들의 대부분(또는 전부)이 공통 하우징(1902) 내에 있을 수 있다.

[0156] 특정 구현예들에서, 서열분석 시스템(1900A)은, 드 노보 서열분석(de novo sequencing), 전체 게놈들 또는 표적 게놈 영역들의 재서열분석, 및 메타게노믹스(metagenomics)를 포함하지만 이들로 제한되지 않는 다양한 응용들을 위해 구성된 핵산 서열분석 시스템이다. 서열분석기는 또한 DNA 또는 RNA 분석을 위해 사용될 수 있다. 일부 구현예들에서, 서열분석 시스템(1900A)은 또한, 바이오센서 내의 반응 사이트들을 생성하도록 구성될 수 있다. 예를 들어, 서열분석 시스템(1900A)은 샘플을 수용하도록, 그리고 샘플로부터 도출된 클론 증폭된 핵산의 표면 부착 클러스터들을 생성하도록 구성될 수 있다. 각각의 클러스터는 바이오센서 내의 반응 사이트를 구성하거나 그의 일부일 수 있다.

[0157] 예시적인 서열분석 시스템(1900A)은 바이오센서(1912) 내에서 원하는 반응들을 수행하기 위해 바이오센서(1912)와 상호작용하도록 구성된 시스템 리셉터클 또는 인터페이스(1910)를 포함할 수 있다. 도 19a와 관련한 하기의 설명에서, 바이오센서(1912)는 시스템 리셉터클(1910) 내로 로딩된다. 그러나, 바이오센서(1912)를 포함하는 카트리지는 시스템 리셉터클(1910) 내로 삽입될 수 있고, 일부 상태들에서, 카트리지는 일시적으로 또는 영구적으로 제거될 수 있다는 것이 이해된다. 전술된 바와 같이, 카트리지는, 특히, 유체 제어 및 유체 저장 컴포넌트들을 포함할 수 있다.

[0158] 특정 구현예들에서, 서열분석 시스템(1900A)은 바이오센서(1912) 내에서 다수의 병렬 반응들을 수행하도록 구성된다. 바이오센서(1912)는 원하는 반응들이 발생할 수 있는 하나 이상의 반응 사이트들을 포함한다. 반응 사이트들은, 예를 들어, 바이오센서의 고체 표면에 고정화되거나, 또는 바이오센서의 대응하는 반응 챔버들 내에 위치되는 비드들(또는 다른 이동가능한 기재들)에 고정화될 수 있다. 반응 사이트들은, 예를 들어, 클론 증폭된 핵산의 클러스터들을 포함할 수 있다. 바이오센서(1912)는 솔리드 스테이트 이미징 디바이스(예컨대, CCD 또는 CMOS 이미저) 및 이에 탑재되는 플로우 셀을 포함할 수 있다. 플로우 셀은, 서열분석 시스템(1900A)으로부터 용액을 수용하고 그 용액을 반응 사이트들을 향해 지향시키는 하나 이상의 플로우 채널들을 포함할 수 있다. 선택적으로, 바이오센서(1912)는 열 에너지를 플로우 채널 내로 또는 플로우 채널 밖으로 이송하기 위한 열 요소와 결합하도록 구성될 수 있다.

[0159] 서열분석 시스템(1900A)은 생물학적 또는 화학적 분석을 위한 미리결정된 방법 또는 검정 프로토콜을 수행하기 위해 서로 상호작용하는 다양한 컴포넌트들, 조립체들, 및 시스템들(또는 서브시스템들)을 포함할 수 있다. 예를 들어, 서열분석 시스템(1900A)은 서열분석 시스템(1900A)의 다양한 컴포넌트들, 조립체들, 및 서브시스템들, 및 또한 바이오센서(1912)와 통신할 수 있는 시스템 제어기(1906)를 포함한다. 예를 들어, 시스템 리셉터클(1910)에 더하여, 서열분석 시스템(1900A)은 또한, 서열분석 시스템(1900A) 및 바이오센서(1912)의 유체 네트워크 전체에 걸쳐 유체의 흐름을 제어하기 위한 유체 제어 시스템(1908); 생물검정 시스템에 의해 사용될 수 있는 모든 유체들(예컨대, 기체 또는 액체류)을 보유하도록 구성된 유체 저장 시스템(1914); 유체 네트워크, 유체 저장 시스템(1914), 및/또는 바이오센서(1912) 내의 유체의 온도를 조절할 수 있는 온도 제어 시스템(1904); 및 바이오센서(1912)를 조명하도록 구성된 조명 시스템(1916)을 포함할 수 있다. 전술된 바와 같이, 바이오센서(1912)를 갖는 카트리지가 시스템 리셉터클(1910) 내로 로딩되는 경우, 카트리지는 또한 유체 제어 및 유체 저장 컴포넌트들을 포함할 수 있다.

[0160] 또한, 도시된 바와 같이, 서열분석 시스템(1900A)은 사용자와 상호작용하는 사용자 인터페이스(1918)를 포함할 수 있다. 예를 들어, 사용자 인터페이스(1918)는 사용자로부터의 정보를 디스플레이하거나 요청하기 위한 디스플레이(1920) 및 사용자 입력들을 수신하기 위한 사용자 입력 디바이스(1922)를 포함할 수 있다. 일부 구현예들에서, 디스플레이(1920) 및 사용자 입력 디바이스(1922)는 동일한 디바이스이다. 예를 들어, 사용자 인터페이스(1918)는 개인의 터치의 존재를 검출하도록, 그리고 또한 디스플레이 상의 터치의 위치를 식별하도록 구성된 터치 감응형 디스플레이를 포함할 수 있다. 그러나, 마우스, 터치패드, 키보드, 키패드, 핸드헬드 스캐너, 음성 인식 시스템, 모션 인식 시스템 등과 같은 다른 사용자 입력 디바이스들(1922)이 사용될 수 있다. 하기에서 더 상세히 논의되는 바와 같이, 서열분석 시스템(1900A)은 원하는 반응들을 수행하기 위해, 바이오센서(1912)(예컨대, 카트리지의 형태로 있음)를 포함하는 다양한 컴포넌트들과 통신할 수 있다. 서열분석 시스템(1900A)은, 또한, 사용자에게 원하는 정보를 제공하기 위해 바이오센서로부터 획득된 데이터를 분석하도록 구성될 수 있다.

[0161] 시스템 제어기(1906)는 마이크로제어기들을 사용하는 시스템, 축약 명령어 세트 컴퓨터(reduced instruction set computer, RISC), 주문형 반도체(ASIC), 필드 프로그래밍가능 게이트 어레이(FPGA), 코어스-그레인드 재구성가능 아키텍처(CGRA), 논리 회로, 및 본 명세서에 기술된 기능들을 실행할 수 있는 임의의 다른 회로 또는 프로세서를 포함하는, 임의의 프로세서 기반 또는 마이크로프로세서 기반 시스템을 포함할 수 있다. 상기의 예들은 단지 예시적인 것이며, 따라서, 어떠한 방식으로든, 용어 "시스템 제어기"의 정의 및/또는 의미를 제한하도록 의도되지 않는다. 예시적인 구현예에서, 시스템 제어기(1906)는 검출 데이터를 획득하는 것 및 분석하는 것 중 적어도 하나를 위해 하나 이상의 저장 요소들, 메모리들, 또는 모듈들에 저장된 명령어들의 세트를 실행한다. 검출 데이터는 픽셀 신호들의 복수의 시퀀스들을 포함하여, 수백만 개의 센서들(또는 픽셀들) 각각으로부터의 픽셀 신호들의 시퀀스가 많은 염기 호출 사이클들에 걸쳐서 검출될 수 있게 할 수 있다. 저장 요소들은 서열분석 시스템(1900A) 내의 정보 소스들 또는 물리적 메모리 요소들의 형태로 있을 수 있다.

[0162] 명령어들의 세트는 본 명세서에 기술된 다양한 구현예들의 방법들 및 프로세스들과 같은 특정 동작들을 수행하도록 서열분석 시스템(1900A) 또는 바이오센서(1912)에 지시하는 다양한 커맨드들을 포함할 수 있다. 명령어들

의 세트는, 유형적(tangible)인 비일시적 컴퓨터 판독가능 매체 또는 매체들의 일부를 형성할 수 있는 소프트웨어 프로그램의 형태로 있을 수 있다. 본 명세서에 사용되는 바와 같이, 용어들 "소프트웨어" 및 "펌웨어"는 상호교환가능하며, RAM 메모리, ROM 메모리, EPROM 메모리, EEPROM 메모리, 및 비휘발성 RAM(NVRAM) 메모리를 포함한, 컴퓨터에 의한 실행을 위한 메모리에 저장된 임의의 컴퓨터 프로그램을 포함한다. 상기의 메모리 유형들은 단지 예시적이며, 따라서, 컴퓨터 프로그램의 저장을 위해 사용가능한 메모리의 유형들에 대한 제한은 아니다.

[0163] 소프트웨어는 시스템 소프트웨어 또는 애플리케이션 소프트웨어와 같은 다양한 형태들로 있을 수 있다. 또한, 소프트웨어는 별개의 프로그램들의 집합의 형태, 또는 더 큰 프로그램 내의 프로그램 모듈 또는 프로그램 모듈의 일부분의 형태로 있을 수 있다. 소프트웨어는 또한 객체 지향 프로그래밍 형태의 모듈식 프로그래밍을 포함할 수 있다. 검출 데이터를 획득한 후에, 검출 데이터는 서열분석 시스템(1900A)에 의해 자동으로 프로세싱될 수 있거나, 사용자 입력들에 응답하여 프로세싱될 수 있거나, 또는 다른 프로세싱 기계에 의해 이루어진 요청(예컨대, 통신 링크를 통한 원격 요청)에 응답하여 프로세싱될 수 있다. 예시된 구현예에서, 시스템 제어기(1906)는 분석 모듈(1944)을 포함한다. 다른 구현예들에서, 시스템 제어기(1906)는 분석 모듈(1944)을 포함하지 않고, 대신에, 분석 모듈(1944)에 액세스한다(예컨대, 분석 모듈(1944)은 클라우드 상에서 개별적으로 호스팅될 수 있음).

[0164] 시스템 제어기(1906)는 통신 링크들을 통해 서열분석 시스템(1900A)의 바이오센서(1912) 및 다른 컴포넌트들에 접속될 수 있다. 시스템 제어기(1906)는 또한 현장외(off-site) 시스템들 또는 서버들에 통신가능하게 접속될 수 있다. 통신 링크들은 유선, 코드, 또는 무선일 수 있다. 시스템 제어기(1906)는 사용자 인터페이스(1918) 및 사용자 입력 디바이스(1922)로부터 사용자 입력들 또는 커맨드들을 수신할 수 있다.

[0165] 유체 제어 시스템(1908)은 유체 네트워크를 포함하며, 유체 네트워크를 통한 하나 이상의 유체들의 유동을 지향시키고 조절하도록 구성된다. 유체 네트워크는 바이오센서(1912) 및 유체 저장 시스템(1914)과 유체 연통할 수 있다. 예를 들어, 선택 유체들이, 제어된 방식으로, 유체 저장 시스템(1914)으로부터 인출되어 바이오센서(1912)로 지향될 수 있거나, 또는 유체들이 바이오센서(1912)로부터 인출되어, 예를 들어, 유체 저장 시스템(1914) 내의 폐기물 저장조를 향해 지향될 수 있다. 도시되어 있지 않지만, 유체 제어 시스템(1908)은 유체 네트워크 내의 유체들의 유량 또는 압력을 검출하는 유동 센서들을 포함할 수 있다. 센서들은 시스템 제어기(1906)와 통신할 수 있다.

[0166] 온도 제어 시스템(1904)은 유체 네트워크, 유체 저장 시스템(1914) 및/또는 바이오센서(1912)의 상이한 영역들에서의 유체들의 온도를 조절하도록 구성된다. 예를 들어, 온도 제어 시스템(1904)은, 바이오센서(1912)와 인터페이싱하고 바이오센서(1912) 내의 반응 사이트들을 따라 유동하는 유체의 온도를 제어하는 서모사이클러(thermocycler)를 포함할 수 있다. 온도 제어 시스템(1904)은, 또한, 서열분석 시스템(1900A) 또는 바이오센서(1912)의 고체 요소들 또는 컴포넌트들의 온도를 조절할 수 있다. 도시되어 있지 않지만, 온도 제어 시스템(1904)은 유체의 온도를 검출하기 위한 센서 또는 다른 컴포넌트들을 포함할 수 있다. 센서들은 시스템 제어기(1906)와 통신할 수 있다.

[0167] 유체 저장 시스템(1914)은 바이오센서(1912)와 유체 연통하고, 내부에서 원하는 반응을 이행하는 데 사용되는 다양한 반응 컴포넌트들 또는 반응물들을 저장할 수 있다. 유체 저장 시스템(1914)은, 또한, 유체 네트워크 및 바이오센서(1912)를 세척 또는 세정하기 위한 그리고 반응물들을 희석하기 위한 유체들을 저장할 수 있다. 예를 들어, 유체 저장 시스템(1914)은 샘플들, 시약들, 효소들, 다른 생체분자들, 완충 용액들, 수성 및 비극성 용액들 등을 저장하기 위한 다양한 저장조들을 포함할 수 있다. 더욱이, 유체 저장 시스템(1914)은, 또한, 바이오센서(1912)로부터 폐기물들을 수용하기 위한 폐기물 저장조들을 포함할 수 있다. 카트리지를 포함하는 구현예들에서, 카트리지는 유체 저장 시스템, 유체 제어 시스템, 또는 온도 제어 시스템 중 하나 이상을 포함할 수 있다. 따라서, 그러한 시스템들에 관한 것으로 본 명세서에 기술된 컴포넌트들 중 하나 이상은 카트리지 하우징 내에 포함될 수 있다. 예를 들어, 카트리지는 샘플들, 시약들, 효소들, 다른 생체분자들, 완충 용액들, 수성 및 비극성 용액들, 폐기물 등을 저장하기 위한 다양한 저장소들을 가질 수 있다. 이와 같이, 유체 저장 시스템, 유체 제어 시스템 또는 온도 제어 시스템 중 하나 이상이 카트리지 또는 다른 바이오센서를 통해 생물 검정 시스템과 제거가능하게 결합될 수 있다.

[0168] 조명 시스템(1916)은 바이오센서를 조명하기 위한 광원(예컨대, 하나 이상의 LED들) 및 복수의 광학 컴포넌트들을 포함할 수 있다. 광원들의 예들은 레이저, 아크 램프, LED, 또는 레이저 다이오드를 포함할 수 있다. 광학 컴포넌트들은, 예를 들어, 반사기, 다이크로익(dichroic), 빔 스플리터, 시준기, 렌즈, 필터, 웨지, 프리즘, 거

울, 검출기 등일 수 있다. 조명 시스템을 사용하는 구현예들에서, 조명 시스템(1916)은 여기 광을 반응 사이트들로 지향시키도록 구성될 수 있다. 하나의 예로서, 형광단들은 광의 녹색 파장들에 의해 여기될 수 있는데, 이는 여기 광의 그러한 파장이 대략 1932 nm일 수 있기 때문이다. 하나의 구현예에서, 조명 시스템(1916)은 바이오센서(1912)의 표면의 표면 법선에 평행한 조명을 생성하도록 구성된다. 다른 구현예에서, 조명 시스템(1916)은 바이오센서(1912)의 표면의 표면 법선에 대해 오프각(off-angle)인 조명을 생성하도록 구성된다. 또 다른 구현예에서, 조명 시스템(1916)은 일부 평행 조명 및 일부 오프각 조명을 포함한 복수의 각도들을 갖는 조명을 생성하도록 구성된다.

[0169] 시스템 리셉터클 또는 인터페이스(1910)는 기계적 방식, 전기적 방식 및 유체 방식 중 적어도 하나로 바이오센서(1912)와 결합하도록 구성된다. 시스템 리셉터클(1910)은 바이오센서(1912)를 통한 유체의 유동을 용이하게 하기 위해 바이오센서(1912)를 원하는 배향으로 보유할 수 있다. 시스템 리셉터클(1910)은, 또한, 바이오센서(1912)와 결합하도록 구성된 전기 접촉부들을 포함하여, 서열분석 시스템(1900A)이 바이오센서(1912)와 통신하고/하거나 바이오센서(1912)에 전력을 제공할 수 있게 할 수 있다. 또한, 시스템 리셉터클(1910)은 바이오센서(1912)와 결합하도록 구성된 유체 포트들(예컨대, 노즐들)을 포함할 수 있다. 일부 구현예들에서, 바이오센서(1912)는 기계적 방식으로, 전기적 방식으로, 그리고 또한 유체 방식으로 시스템 리셉터클(1910)에 제거가능하게 커플링된다.

[0170] 또한, 서열분석 시스템(1900A)은 다른 시스템들 또는 네트워크들과 또는 다른 생물검정 시스템(1900A)과 원격으로 통신할 수 있다. 생물검정 시스템(들)(1900A)에 의해 획득된 검출 데이터는 원격 데이터베이스에 저장될 수 있다.

[0171] 도 19b는 도 19a의 시스템에서 사용될 수 있는 시스템 제어기(1906)의 블록도이다. 하나의 구현예에서, 시스템 제어기(1906)는 서로 통신할 수 있는 하나 이상의 프로세서들 또는 모듈들을 포함한다. 프로세서들 또는 모듈들 각각은 특정 프로세스들을 수행하기 위한 알고리즘(예컨대, 유형적 그리고/또는 비일시적 컴퓨터 판독가능 저장 매체 상에 저장된 명령어들) 또는 서브알고리즘들을 포함할 수 있다. 시스템 제어기(1906)는, 개념적으로는 모듈들의 집합으로서 예시되지만, 전용 하드웨어 보드, DSP, 프로세서 등의 임의의 조합을 활용하여 구현될 수 있다. 대안으로, 시스템 제어기(1906)는 단일 프로세서 또는 다수의 프로세서들을 갖는 기성품(off-the-shelf) PC를 활용하여 구현될 수 있으며, 이때 기능 동작들이 프로세서들 사이에 분배된다. 추가 옵션으로서, 후술되는 모듈들은 소정의 모듈식 기능들이 전용 하드웨어를 활용하여 수행되는 하이브리드 구성을 활용하여 구현될 수 있는 반면, 나머지 모듈식 기능들은 기성품 PC 등을 활용하여 수행된다. 모듈들은 또한 프로세싱 유닛 내의 소프트웨어 모듈들로서 구현될 수 있다.

[0172] 동작 동안, 통신 포트(1950)는 정보(예컨대, 커맨드들)를 바이오센서(1912)(도 19a) 및/또는 서브시스템들(1908, 1914, 1904)(도 19a)로 송신할 수 있거나 정보(예컨대, 데이터)를 그들로부터 수신할 수 있다. 구현예들에서, 통신 포트(1950)는 픽셀 신호들의 복수의 시퀀스들을 출력할 수 있다. 통신 링크(1934)는 사용자 인터페이스(1918)(도 19a)로부터 사용자 입력을 수신할 수 있고, 데이터 또는 정보를 사용자 인터페이스(1918)로 송신할 수 있다. 바이오센서(1912) 또는 서브시스템들(1908, 1914, 1904)로부터의 데이터는 생물검정 세션 동안 실시간으로 시스템 제어기(1906)에 의해 프로세싱될 수 있다. 추가로 또는 대안으로, 데이터는 생물검정 세션 동안 시스템 메모리에 일시적으로 저장될 수 있고, 실시간 또는 오프라인 동작보다 느리게 프로세싱될 수 있다.

[0173] 도 19b에 도시된 바와 같이, 시스템 제어기(1906)는, 중앙 프로세싱 유닛(CPU)(1952)과 함께, 주 제어 모듈(1924)과 통신하는 복수의 모듈들(1926 내지 1948)을 포함할 수 있다. 주 제어 모듈(1924)은 사용자 인터페이스(1918)(도 19a)와 통신할 수 있다. 모듈들(1926 내지 1948)이 주 제어 모듈(1924)과 직접 통신하는 것으로 도시되어 있지만, 모듈들(1926 내지 1948)은 또한 서로, 사용자 인터페이스(1918), 및 바이오센서(1912)와 직접 통신할 수 있다. 또한, 모듈들(1926 내지 1948)은 다른 모듈들을 통해 주 제어 모듈(1924)과 통신할 수 있다.

[0174] 복수의 모듈들(1926 내지 1948)은 서브시스템들(1908, 1914, 1904, 1916)과 각각 통신하는 시스템 모듈들(1928 내지 1932, 1926)을 포함한다. 유체 제어 모듈(1928)은 유체 제어 시스템(1908)과 통신하여, 유체 네트워크를 통한 하나 이상의 유체들의 유동을 제어하기 위해 유체 네트워크의 유동 센서들 및 밸브들을 제어할 수 있다. 유체 저장 모듈(1930)은 유체들이 낮을 때 또는 폐기물 저장소가 최대 용량에 또는 그 근처에 있을 때 사용자에게 통지할 수 있다. 유체 저장 모듈(1930)은, 또한, 온도 제어 모듈(1932)과 통신하여, 유체들이 원하는 온도로 저장될 수 있게 할 수 있다. 조명 모듈(1926)은 조명 시스템(1916)과 통신하여, 프로토콜 동안의 지정된 시간들에, 예컨대 원하는 반응들(예컨대, 결합 이벤트들)이 발생한 후에, 반응 사이트들을 조명할 수 있다. 일부 구현예들에서, 조명 모듈(1926)은 지정된 각도들로 반응 사이트들을 조명하기 위해 조명 시스템(1916)과 통신할

수 있다.

[0175] 복수의 모듈들(1926 내지 1948)은, 또한, 바이오센서(1912)와 통신하는 디바이스 모듈(1936) 및 바이오센서(1912)에 관한 식별 정보를 결정하는 식별 모듈(1938)을 포함할 수 있다. 디바이스 모듈(1936)은, 예를 들어, 시스템 리셉터클(1910)과 통신하여, 바이오센서가 서열분석 시스템(1900A)과 전기적 및 유체 접속을 확립하였음을 확인할 수 있다. 식별 모듈(1938)은 바이오센서(1912)를 식별하는 신호들을 수신할 수 있다. 식별 모듈(1938)은 다른 정보를 사용자에게 제공하기 위해 바이오센서(1912)의 아이덴티티를 사용할 수 있다. 예를 들어, 식별 모듈(1938)은 로트 번호, 제조 날짜, 또는 바이오센서(1912)로 실행될 것이 추천되는 프로토콜을 결정할 수 있고, 이어서, 이를 디스플레이할 수 있다.

[0176] 복수의 모듈들(1926 내지 1948)은, 또한, 바이오센서(1912)로부터 신호 데이터(예컨대, 이미지 데이터)를 수신 및 분석하는 분석 모듈(1944)(신호 프로세싱 모듈 또는 신호 프로세서로도 칭해짐)을 포함한다. 분석 모듈(1944)은 검출/이미지 데이터를 저장하기 위한 메모리(예컨대, RAM 또는 플래시)를 포함한다. 검출 데이터는 픽셀 신호들의 복수의 시퀀스들을 포함하여, 수백만 개의 센서들(또는 픽셀들) 각각으로부터의 픽셀 신호들의 시퀀스가 많은 염기 호출 사이클들에 걸쳐서 검출될 수 있게 할 수 있다. 신호 데이터는 후속 분석을 위해 저장될 수 있거나, 또는 사용자에게 원하는 정보를 디스플레이하기 위해 사용자 인터페이스(1918)로 송신될 수 있다. 일부 구현예들에서, 신호 데이터는 분석 모듈(1944)이 신호 데이터를 수신하기 전에 솔리드 스테이트 이미저(예컨대, CMOS 이미지 센서)에 의해 프로세싱될 수 있다.

[0177] 분석 모듈(1944)은 복수의 서열분석 사이클들 각각에서 광 검출기들로부터 이미지 데이터를 획득하도록 구성된다. 이미지 데이터는 광 검출기들에 의해 검출된 방출 신호들로부터 도출되고, 신경 네트워크 기반 염기 호출자(100)를 통해 복수의 서열분석 사이클들 각각 동안의 이미지 데이터를 프로세싱하고, 복수의 서열분석 사이클 각각 동안 분석물들 중 적어도 일부에 대한 염기 호출을 생성한다. 광 검출기들은 하나 이상의 오버헤드 카메라들(예컨대, 상부로부터 바이오센서(1912) 상의 클러스터들의 이미지들을 촬영하는 Illumina의 GAIIx의 CCD 카메라)의 일부일 수 있거나, 또는 바이오센서(1912) 자체의 일부일 수 있다(예컨대, Illumina의 iSeq의 CMOS 이미지 센서들은 바이오센서(1912) 상의 클러스터들의 하부에 놓이고, 저부로부터 클러스터들의 이미지들을 촬영한다).

[0178] 광 검출기들의 출력은 서열분석 이미지들이며, 각각은 클러스터들 및 그들의 주변 배경의 세기 방출물들을 묘사한다. 서열분석 이미지들은 서열분석 동안 서열들 내의 뉴클레오티드 혼입의 결과로서 생성되는 세기 방출물들을 묘사한다. 세기 방출물들은 연관된 분석물들 및 그들의 주변 배경으로부터의 것이다. 서열분석 이미지들은 메모리(1948)에 저장된다.

[0179] 프로토콜 모듈들(1940, 1942)은 미리결정된 검정 프로토콜들을 이행할 때 서브시스템들(1908, 1914, 1904)의 동작을 제어하기 위해 주 제어 모듈(1924)과 통신한다. 프로토콜 모듈들(1940, 1942)은 미리결정된 프로토콜들에 따라 특정 동작들을 수행할 것을 서열분석 시스템(1900A)에 지시하기 위한 명령어들의 세트들을 포함할 수 있다. 도시된 바와 같이, 프로토콜 모듈은 합성에 의한 서열분석(sequencing-by-synthesis, SBS) 프로세스를 수행하기 위한 다양한 커맨드들을 발행하도록 구성된 SBS 모듈(1940)일 수 있다. SBS에서, 템플릿 내의 뉴클레오티드의 서열을 결정하기 위해 핵산 템플릿을 따른 핵산 프라이머의 연장이 모니터링된다. 기초 화학적 프로세스는 중합화(예컨대, 폴리머라제 효소에 의해 촉매화되는 바와 같음) 또는 결찰(ligation)(예컨대, 리가아제 효소에 의해 촉매화됨)일 수 있다. 특정 폴리머라제 기반 SBS 구현예에서, 프라이머에 첨가되는 뉴클레오티드의 순서 및 유형의 검출이 템플릿의 서열을 결정하는 데 사용될 수 있도록 하는 템플릿 의존 방식으로, 형광 표지된 뉴클레오티드가 프라이머에 첨가된다(이에 의해, 프라이머를 연장함). 예를 들어, 제1 SBS 사이클을 개시하기 위해, 핵산 템플릿들의 어레이를 수용하는 플로우 셀 내로/플로우 셀을 통해 하나 이상의 표지된 뉴클레오티드, DNA 폴리머라제 등을 전달하기 위한 커맨드들이 제공될 수 있다. 핵산 템플릿들은 대응하는 반응 사이트들에 위치될 수 있다. 프라이머 연장이 표지된 뉴클레오티드가 혼입되게 하는 그러한 반응 사이트들이 이미징 이벤트를 통해 검출될 수 있다. 이미징 이벤트 동안, 조명 시스템(1916)은 반응 사이트들에 여기 광을 제공할 수 있다. 선택적으로, 뉴클레오티드는, 일단 뉴클레오티드가 프라이머에 첨가되었다면, 추가 프라이머 연장을 종결시키는 가역적 종결 속성을 추가로 포함할 수 있다. 예를 들어, 가역적 종결자 모이어티를 갖는 뉴클레오티드 유사체가 프라이머에 첨가되어, 모이어티를 제거하기 위해 탈블록킹제(deblocking agent)가 전달될 때까지 후속 연장이 발생할 수 없게 할 수 있다. 따라서, 가역적 종결을 사용하는 구현예들에 대해, (검출이 발생하기 전 또는 발생한 후) 탈블록킹 시약을 플로우 셀에 전달하기 위한 커맨드가 제공될 수 있다. 다양한 전달 단계들 사이의 세척(들)을 달성하기 위해 하나 이상의 커맨드들이 제공될 수 있다. 이어서, 사이클은 n개의 뉴클레오티드에 의해 프라이머를 연장시키기 위해 n회 반복되어, 이에 의해, 길이 n의 서열을 검출할 수 있다. 예시

적인 서열분석 기법들이, 예를 들어 문헌[Bentley et al., Nature 4196:193-199 (20019)]; 국제 특허 공개 WO 04/0119497호; 미국 특허 제7,0197,026호; 국제 특허 공개 WO 91/066719호; 국제 특허 공개 WO 07/123744호; 미국 특허 제7,329,492호; 미국 특허 제7,211,414호; 미국 특허 제7,3119,019호; 미국 특허 제7,4019,219호, 및 미국 특허 출원 공개 제20019/01470190192호에 기재되어 있으며, 이들 각각은 본 명세서에 참고로 포함된다.

[0180] SBS 사이클의 뉴클레오티드 전달 단계의 경우, 단일 유형의 뉴클레오티드가 한 번에 전달될 수 있거나, 또는 다수의 상이한 뉴클레오티드 유형들이(예컨대, A, C, T 및 G가 함께) 전달될 수 있다. 단일 유형의 뉴클레오티드만이 한 번에 존재하는 뉴클레오티드 전달 구성의 경우, 상이한 뉴클레오티드는 별개의 표지들을 가질 필요가 없는데, 이는 그들이 개별화된 전달 시에 고유한 시간적 분리에 기초하여 구별될 수 있기 때문이다. 따라서, 서열분석 방법 또는 장치는 단일 컬러 검출을 사용할 수 있다. 예를 들어, 여기 소스는 단일 파장의 또는 단일 파장 범위 내의 여기만을 제공할 필요가 있다. 전달이 플로우 셀 내에 한 번에 존재하는 다수의 상이한 뉴클레오티드를 생성하는 뉴클레오티드 전달 구성의 경우, 상이한 뉴클레오티드 유형들을 혼입하는 부위들은 혼합물 내의 각자의 뉴클레오티드 유형들에 부착되는 상이한 형광 표지들에 기초하여 구별될 수 있다. 예를 들어, 4개의 상이한 뉴클레오티드가 사용될 수 있으며, 각각은 4개의 상이한 형광단들 중 하나를 갖는다. 하나의 구현예에서, 4개의 상이한 형광단들은 스펙트럼의 4개의 상이한 영역들에서의 여기를 사용하여 구별될 수 있다. 예를 들어, 4개의 상이한 여기 방사선 소스들이 사용될 수 있다. 대안으로, 4개 미만의 상이한 여기 소스들이 사용될 수 있지만, 단일 소스로부터의 여기 방사선의 광학적 여과가 플로우 셀에서 상이한 범위들의 여기 방사선을 생성하는 데 사용될 수 있다.

[0181] 일부 구현예들에서, 4개 미만의 상이한 컬러들이 4개의 상이한 뉴클레오티드를 갖는 혼합물에서 검출될 수 있다. 예를 들어, 뉴클레오티드의 쌍들은 동일한 파장에서 검출될 수 있지만, 그 쌍의 하나의 구성원에 대해 다른 구성원과 비교한 세기의 차이에 기초하여, 또는 그 쌍의 다른 구성원에 대해 검출된 신호와 비교한, 명백한 신호가 나타나거나 사라지게 하는 (예컨대, 화학적 변형, 광화학적 변형, 또는 물리적 변형을 통한) 그 쌍의 하나의 구성원에 대한 변화에 기초하여 구별될 수 있다. 4개 미만의 컬러들의 검출을 사용하여 4개의 상이한 뉴클레오티드를 구별하기 위한 예시적인 장치 및 방법들이, 예를 들어 미국 특허 출원 제61/19319,294호 및 제61/619,19719호에 기재되어 있으며, 이들은 그들 전체가 본 명세서에 참고로 포함된다. 2012년 9월 21일자로 출원된 미국 출원 제13/624,200호가 또한 전체가 본 명세서에 참고로 포함된다.

[0182] 복수의 프로토콜 모듈들은, 또한, 바이오센서(1912) 내의 제품을 증폭시키기 위한 커맨드들을 유체 제어 시스템(1908) 및 온도 제어 시스템(1904)에 발행하도록 구성된 샘플 준비(또는 생성) 모듈(1942)을 포함할 수 있다. 예를 들어, 바이오센서(1912)는 서열분석 시스템(1900A)에 결합될 수 있다. 증폭 모듈(1942)은 필요한 증폭 성분들을 바이오센서(1912) 내의 반응 챔버들로 전달하라는 명령어들을 유체 제어 시스템(1908)에 발행할 수 있다. 다른 구현예들에서, 반응 사이트들은 증폭을 위한 일부 컴포넌트들, 예컨대 템플릿 DNA 및/또는 프라이머들을 이미 함유할 수 있다. 증폭 성분들을 반응 챔버들로 전달한 후에, 증폭 모듈(1942)은 알려진 증폭 프로토콜들에 따라 상이한 온도 단계들을 통해 순환할 것을 온도 제어 시스템(1904)에 지시할 수 있다. 일부 구현예들에서, 증폭 및/또는 뉴클레오티드 혼입은 등온적으로 수행된다.

[0183] SBS 모듈(1940)은 클론 앰플리콘(clonal amplicon)들의 클러스터들이 플로우 셀의 채널 내의 국소화된 영역들 상에 형성되는 브리지 PCR을 수행하라는 커맨드들을 발행할 수 있다. 브리지 PCR을 통해 앰플리콘들을 생성한 후, 앰플리콘들은 단일 가닥 템플릿 DNA 또는 sstDNA를 만들기 위해 "선형화"될 수 있고, 서열분석 프라이머는 관심 영역을 플랭킹하는 유니버셜 서열에 혼성화될 수 있다. 예를 들어, 합성 방법에 의한 가역적 종결자 기반 서열분석은 상기에 기재된 바와 같이 또는 하기와 같이 사용될 수 있다.

[0184] 각각의 염기 호출 또는 서열분석 사이클은, 예를 들어 변형된 DNA 폴리머라제 및 4개의 유형들의 뉴클레오티드의 혼합물을 사용함으로써 달성될 수 있는 단일 염기에 의해 sstDNA를 연장할 수 있다. 상이한 유형들의 뉴클레오티드는 고유 형광 표지들을 가질 수 있고, 각각의 뉴클레오티드는 각각의 사이클에서 단일 염기 혼입만이 발생하는 것을 허용하는 가역적 종결자를 추가로 가질 수 있다. 단일 염기가 sstDNA에 첨가된 후, 여기 광은 반응 사이트들 위로 입사할 수 있고, 형광 방출물들이 검출될 수 있다. 검출 후에, 형광 표지 및 종결자는 sstDNA로부터 화학적으로 절단될 수 있다. 다른 유사한 염기 호출 또는 서열분석 사이클이 이어질 수 있다. 그러한 서열분석 프로토콜에서, SBS 모듈(1940)은 바이오센서(1912)를 통해 시약 및 효소 용액들의 유동을 지향시킬 것을 유체 제어 시스템(1908)에 지시할 수 있다. 본 명세서에 기술된 장치 및 방법들로 활용될 수 있는 예시적인 가역적 종결자 기반 SBS 방법들은, 미국 특허 출원 공개 제2007/01667019 A1호, 미국 특허 출원 공개 제2006/01196*3901 A1호, 미국 특허 제7,0197,026호, 미국 특허 출원 공개 제2006/0240439 A1호, 미국 특허 출원 공개 제2006/021914714709 A1호, PCT 공개 공보 WO 019/0619514호, 미국 특허 출원 공개 제20019/014700900

A1호, PCT 공개 공보 WO 06/019B199호 및 PCT 공개 공보 WO 07/014702191호에 기재되어 있으며, 이들 각각은 그 전체가 본 명세서에 참고로 포함된다. 가역적 종결자 기반 SBS를 위한 예시적인 시약들은 미국 특허 제7,1941,444호; 미국 특허 제7,0197,026호; 미국 특허 제7,414,14716호; 미국 특허 제7,427,673호; 미국 특허 제7,1966,1937호; 미국 특허 제7,1992,4319호, 및 국제 특허 출원 공개 WO 07/14193193619호에 기재되어 있으며, 이들 각각은 그 전체가 본 명세서에 참고로 포함된다.

[0185] 일부 구현예들에서, 증폭 및 SBS 모듈들은 단일 검정 프로토콜로 동작할 수 있는데, 여기서 예를 들어, 템플릿 핵산은 증폭되고, 후속으로, 동일한 카트리지 내에서 서열분석된다.

[0186] 서열분석 시스템(1900A)은, 또한, 사용자가 검정 프로토콜을 재구성하는 것을 허용할 수 있다. 예를 들어, 서열분석 시스템(1900A)은 결정된 프로토콜을 수정하기 위한 옵션들을 사용자 인터페이스(1918)를 통해 사용자에게 제공할 수 있다. 예를 들어, 바이오센서(1912)가 증폭을 위해 사용될 것으로 결정되는 경우, 서열분석 시스템(1900A)은 어닐링 사이클을 위한 온도를 요청할 수 있다. 더욱이, 서열분석 시스템(1900A)은 사용자가 선택된 검정 프로토콜에 대해 대체적으로 수용가능하지 않은 사용자 입력들을 제공한 경우에 사용자에게 경고들을 발행할 수 있다.

[0187] 구현예들에서, 바이오센서(1912)는 수백만 개의 센서들(또는 픽셀들)을 포함하고, 이들 각각은 연속 염기 호출 사이클들에 걸쳐서 픽셀 신호들의 복수의 시퀀스들을 생성한다. 분석 모듈(1944)은 픽셀 신호들의 복수의 시퀀스들을 검출하고, 이들을 센서들의 어레이 상의 센서들의 로우별 및/또는 컬럼별 위치에 따라 대응하는 센서들(또는 픽셀들)에 기인하게 한다.

[0188] 도 19c는 염기 호출 센서 출력들과 같은, 서열분석 시스템(1900A)으로부터의 센서 데이터의 분석을 위한 시스템의 단순화된 블록도이다. 도 19c의 예에서, 시스템은 구성가능한 프로세서(1946)를 포함한다. 구성가능한 프로세서(1946)는, 중앙 프로세싱 유닛(CPU)(1952)(즉, 호스트 프로세서)에 의해 실행되는 런타임 프로그램/로직(1980)과 협력하여 염기 호출자(예컨대, 신경 네트워크 기반 염기 호출자(100))를 실행할 수 있다. 서열분석 시스템(1900A)은 바이오센서(1912) 및 플로우 셀들을 포함한다. 플로우 셀들은 하나 이상의 타일들을 포함할 수 있는데, 여기서 유전 물질의 클러스터들은 클러스터들 내의 반응들이 유전 물질 내의 염기들을 식별하게 하는 데 사용되는 분석물 유동들의 시퀀스에 노출된다. 센서들은 플로우 셀의 각각의 타일 내의 시퀀스의 사이클 각각 동안의 반응들을 감지하여 타일 데이터를 제공한다. 유전자 서열분석은 데이터 집약적 동작이며, 이는 염기 호출 센서 데이터를, 염기 호출 동작 동안에 감지된 유전 물질의 각각의 클러스터에 대한 염기 호출들의 서열들로 변환한다.

[0189] 이러한 예에서의 시스템은 런타임 프로그램/로직(1980)을 실행하여 염기 호출 동작들을 조정하는 CPU(1952), 타일 데이터의 어레이들의 시퀀스들을 저장하기 위한 메모리(1948B), 염기 호출 동작에 의해 생성되는 염기 호출 판독물들, 및 염기 호출 동작들에서 사용되는 다른 정보를 포함한다. 또한, 이러한 예시에서, 시스템은 구성 파일(또는 파일들), 예컨대 FPGA 비트 파일들, 및 구성가능한 프로세서(1946)를 구성 및 재구성하기 위해 그리고 신경 네트워크들을 실행하기 위해 사용되는 신경 네트워크들에 대한 모델 파라미터들을 저장할 메모리(1948A)를 포함한다. 서열분석 시스템(1900A)은 구성가능한 프로세서를 구성하기 위한 프로그램, 및 일부 구현예들에서, 신경 네트워크들을 실행하기 위한 재구성가능한 프로세서를 포함할 수 있다.

[0190] 서열분석 시스템(1900A)은 버스(1989)에 의해 구성가능한 프로세서(1946)에 커플링된다. 버스(1989)는, 하나의 예에서 PCI-SIG(PCI Special Interest Group)에 의해 현재 유지되고 개발되는 PCIe(Peripheral Component Interconnect Express) 표준들과 호환가능한 버스 기술과 같은 고처리량 기술을 사용하여 구현될 수 있다. 또한, 이러한 예에서, 메모리(1948A)는 버스(1993)에 의해 구성가능한 프로세서(1946)에 커플링된다. 메모리(1948A)는 구성가능한 프로세서(1946)를 갖는 회로 보드 상에 배치된 온-보드 메모리일 수 있다. 메모리(1948A)는 염기 호출 동작에서 사용되는 데이터를 작동시키는 구성가능한 프로세서(1946)에 의한 고속 액세스를 위해 사용된다. 버스(1993)는, 또한, PCIe 표준들과 호환가능한 버스 기술과 같은 고처리량 기술을 사용하여 구현될 수 있다.

[0191] FPGA, CGRA, 및 다른 구성가능한 그리고 재구성가능한 디바이스들을 포함한 구성가능한 프로세서들은, 컴퓨터 프로그램을 실행하는 범용 프로세서를 사용하여 달성될 수 있었던 것보다 더 효율적이거나 더 빠르게 다양한 기능들을 구현하도록 구성될 수 있다. 구성가능한 프로세서들의 구성은, 때때로 비트스트림 또는 비트 파일로 지칭되는 구성 파일을 생성하기 위한 기능 설명을 컴파일하는 것, 및 구성 파일을 프로세서 상의 구성가능한 요소들에 분배하는 것을 수반한다. 구성 파일은, 데이터 흐름 패턴들, 분산형 메모리 및 다른 온-칩 메모리 리소스들의 사용, 룩업 테이블 콘텐츠들, 승산 및 누산(multiply-and-accumulate) 유닛들과 같은 구성가능한 실행 유

닛들 및 구성가능한 논리 블록들의 동작들, 구성가능한 상호접속부들, 및 구성가능한 어레이의 다른 요소들을 설정하도록 회로를 구성함으로써, 구성가능한 프로세서에 의해 실행될 논리 기능들을 정의한다. 구성가능한 프로세서는, 구성 파일이 필드 내에서 변경될 수 있는 경우, 로딩된 구성 파일을 변경함으로써 재구성가능하다. 예를 들어, 구성 파일은, 구성가능한 또는 재구성가능한 프로세서 상의 구성가능한 요소들의 어레이 사이에 분포되는, 휘발성 SRAM 요소들에, 비휘발성 판독-기록 메모리 요소들에, 그리고 이들의 조합들에 저장될 수 있다. 다양한 상업적으로 입수가능한 구성가능한 프로세서들이 본 명세서에 기술된 바와 같은 염기 호출 동작에 사용하는 데 적합하다. 예들은, Google의 Tensor Processing Unit(TPU)™, 랙마운트 솔루션들, 예컨대 GX4 Rackmount Series™, GX9 Rackmount Series™, NVIDIA DGX-1™, Microsoft의 Stratix V FPGA™, Graphcore의 Intelligent Processor Unit (IPU)™, Snapdragon processors™을 갖는 Qualcomm의 Zeroth Platform™, NVIDIA의 Volta™, NVIDIA의 DRIVE PX™, NVIDIA의 JETSON TX1/TX2 MODULE™, Intel의 Nirvana™, Movidius VPU™, Fujitsu DPI™, ARM의 DynamicIQ™, IBM TrueNorth™, Testa V100s™을 갖는 Lambda GPU 서버, Xilinx Alveo™ U200, Xilinx Alveo™ U2190, Xilinx Alveo™ U280, Intel/Altera Stratix™ GX2800, Intel/Altera Stratix™ GX2800, 및 Intel Stratix™ GX10M을 포함한다. 일부 예들에서, 호스트 CPU는 구성가능한 프로세서와 동일한 집적 회로 상에서 구현될 수 있다.

[0192] 본 명세서에 기술된 구현예들은 구성가능한 프로세서(1946)를 사용하여 신경 네트워크 기반 염기 호출자(100)를 구현한다. 구성가능한 프로세서(1946)에 대한 구성 파일은 HDL(high level description language) 또는 RTL(register transfer level) 언어 규격을 사용하여 실행될 논리 기능들을 특정함으로써 구현될 수 있다. 규격은 구성 파일을 생성하기 위해, 선택된 구성가능한 프로세서에 대해 설계된 리소스들을 사용하여 컴파일될 수 있다. 구성가능한 프로세서가 아닐 수 있는 ASIC에 대한 설계를 생성할 목적으로 동일한 또는 유사한 규격이 컴파일될 수 있다.

[0193] 따라서, 구성가능한 프로세서(1946)에 대한 대안들은, 본 명세서에 기술된 모든 구현예들에서, 본 명세서에 기술된 바와 같은 신경 네트워크 기반 염기 호출 동작을 실행하도록 구성된, 주문형 ASIC 또는 특수 목적 집적 회로 또는 집적 회로들의 세트를 포함하는 구성된 프로세서, 또는 시스템-온-칩(system-on-a-chip, SOC) 디바이스, 또는 그래픽 프로세싱 유닛(GPU) 프로세서 또는 코어스-그레인드 재구성가능 아키텍처(CGRA) 프로세서를 포함한다.

[0194] 대체로, 신경 네트워크의 런들을 실행하도록 구성된 바와 같은, 본 명세서에 기술된 구성가능한 프로세서 및 구성된 프로세서들은 본 명세서에서 신경 네트워크 프로세서들로 지칭된다.

[0195] 구성가능한 프로세서(1946)는, 이러한 예에서, CPU(1952)에 의해 실행되는 프로그램을 사용하여 로딩된 구성 파일에 의해 또는 다른 소스들에 의해 구성되며, 이는 염기 호출 기능을 실행하기 위한 구성가능한 프로세서 상에, 구성가능한 요소들(1991)(예컨대, 룩업 테이블(look up table, LUT)들, 플립플롭들, 계산 프로세싱 유닛(compute processing unit, PMU)들, 및 계산 메모리 유닛(compute memory unit, CMU)들과 같은 구성 로직 블록(configuration logic block, CLB)들, 구성가능한 I/O 블록들, 프로그래밍가능한 상호접속부들)의 어레이를 구성한다. 이러한 예에서, 구성은, 버스들(1989, 1993)에 커플링되고 염기 호출 동작에 사용되는 요소들 사이에서 데이터 및 제어 파라미터들을 분배하기 위한 기능들을 실행하는 데이터 흐름 로직(1997)을 포함한다.

[0196] 또한, 구성가능한 프로세서(1946)는 신경 네트워크 기반 염기 호출자(100)를 실행하기 위해 데이터 흐름 로직(1997)을 갖도록 구성된다. 로직(1997)은 다중-사이클 실행 클러스터들(예컨대, 1979)을 포함하는데, 이는, 이러한 예에서, 실행 클러스터 1 내지 실행 클러스터 X를 포함한다. 다중-사이클 실행 클러스터들의 수는 동작의 원하는 처리량, 및 구성가능한 프로세서(1946) 상의 이용가능한 리소스들을 수반하는 트레이드오프에 따라 선택될 수 있다.

[0197] 다중-사이클 실행 클러스터들은 구성가능한 프로세서(1946) 상의 구성가능한 상호접속부 및 메모리 리소스들을 사용하여 구현되는 데이터 흐름 경로들(1999)에 의해 데이터 흐름 로직(1997)에 커플링된다. 또한, 다중-사이클 실행 클러스터들은, 예를 들어 구성가능한 프로세서(1946) 상의 구성가능한 상호접속부 및 메모리 리소스들을 사용하여 구현된 제어 경로들(1995)에 의해 데이터 흐름 로직(1997)에 커플링되며, 이들은 이용가능한 실행 클러스터들을 나타내는 제어 신호들, 신경 네트워크 기반 염기 호출자(100)의 런의 실행을 위한 입력 유닛들을 이용가능한 실행 클러스터들에 제공하기 위한 준비성(readiness), 신경 네트워크 기반 염기 호출자(100)에 대한 트레이닝된 파라미터들을 제공하기 위한 준비성, 염기 호출 분류 데이터의 출력 패치들을 제공하기 위한 준비성, 및 신경 네트워크 기반 염기 호출자(100)의 실행을 위해 사용되는 다른 제어 데이터를 제공한다.

[0198] 구성가능한 프로세서(1946)는 트레이닝된 파라미터들을 사용하여 신경 네트워크 기반 염기 호출자(100)의 런들

을 실행하여, 염기 호출 동작의 감지 사이클들 동안 분류 데이터를 생성하도록 구성된다. 신경 네트워크 기반 염기 호출자(100)의 런이 실행되어, 염기 호출 동작의 대상 감지 사이클 동안 분류 데이터를 생성한다. 신경 네트워크 기반 염기 호출자(100)의 런은 N개의 감지 사이클들의 각자의 감지 사이클들로부터 타일 데이터의 N개의 어레이들을 포함하는 시퀀스로 동작하며, 여기서 N개의 감지 사이클들은 본 명세서에 기술된 예들에서 시간 시퀀스 내의 동작마다 하나의 염기 포지션에 대한 상이한 염기 호출 동작들을 위한 센서 데이터를 제공한다. 선택적으로, N개의 감지 사이클들 중 일부는 실행되는 특정 신경 네트워크 모델에 따라, 필요한 경우, 시퀀스 이외일 수 있다. 수 N은 1 초과의 임의의 수일 수 있다. 본 명세서에 기술된 일부 예들에서, N개의 감지 사이클들의 감지 사이클들은 시간 시퀀스에서 대상 감지 사이클에 선행하는 적어도 하나의 감지 사이클 및 대상 사이클에 뒤이은 적어도 하나의 감지 사이클에 대한 감지 사이클들의 세트를 표현한다. 수 N이 5 이상의 정수인 예들이 본 명세서에 기술된다.

[0199] 데이터 흐름 로직(1997)은 N개의 어레이들의 공간적으로 정렬된 패치들에 대해 타일 데이터를 포함하는 주어진 런에 대한 입력 유닛들을 사용하여, 신경 네트워크 기반 염기 호출자(100)의 런들을 위한 타일 데이터 및 모델 파라미터들의 적어도 일부의 트레이닝된 파라미터들을 메모리(1948A)로부터 구성가능한 프로세서(1946)로 이동시키도록 구성된다. 입력 유닛들은 하나의 DMA 동작에서 직접 메모리 액세스 동작들에 의해 이동될 수 있거나, 또는 더 작은 유닛들에서, 전개된 신경 네트워크의 실행과 협력하여 이용가능한 시간 슬롯들 동안 이동될 수 있다.

[0200] 본 명세서에 기술된 바와 같은 감지 사이클 동안의 타일 데이터는 하나 이상의 특징부들을 갖는 센서 데이터의 어레이를 포함할 수 있다. 예를 들어, 센서 데이터는 DNA, RNA, 또는 다른 유전 물질의 유전자 서열 내의 염기 포지션에서 4개의 염기들 중 하나를 식별하기 위해 분석되는 2개의 이미지들을 포함할 수 있다. 타일 데이터는, 또한, 이미지들 및 센서들에 관한 메타데이터를 포함할 수 있다. 예를 들어, 염기 호출 동작의 구현예들에서, 타일 데이터는 타일 상의 유전 물질의 클러스터의 중심으로부터 센서 데이터의 어레이 내의 각각의 픽셀의 거리를 나타내는 중심 정보로부터의 거리와 같은, 클러스터들과의 이미지들의 정렬에 관한 정보를 포함할 수 있다.

[0201] 후술되는 바와 같은 신경 네트워크 기반 염기 호출자(100)의 실행 동안, 타일 데이터는, 또한, 중간 데이터로 지칭되는, 신경 네트워크 기반 염기 호출자(100)의 실행 동안 생성된 데이터를 포함할 수 있는데, 이는 신경 네트워크 기반 염기 호출자(100)의 런 동안 재계산되기보다는 재사용될 수 있다. 예를 들어, 신경 네트워크 기반 염기 호출자(100)의 실행 동안, 데이터 흐름 로직(1997)은 타일 데이터의 어레이의 주어진 패치에 대해 센서 데이터 대신에 중간 데이터를 메모리(1948A)에 기록할 수 있다. 이와 같은 구현예들이 더욱 상세하게 후술된다.

[0202] 도시된 바와 같이, 염기 호출 동작의 감지 사이클들로부터의 타일에 대한 센서 데이터를 포함하는 타일 데이터를 저장하는 런타임 프로그램/로직(1980)에 의해 액세스가능한 메모리(예컨대, 1948A)를 포함하는, 염기 호출 센서 출력의 분석을 위한 시스템이 기술된다. 또한, 시스템은 메모리에 액세스하는 구성가능한 프로세서(1946)와 같은 신경 네트워크 프로세서를 포함한다. 신경 네트워크 프로세서는 감지 사이클들 동안 분류 데이터를 생성하기 위해 트레이닝된 파라미터들을 사용하여 신경 네트워크의 런들을 실행하도록 구성된다. 본 명세서에 기술된 바와 같이, 대상 사이클 동안 분류 데이터를 생성하기 위해, 대상 사이클을 포함하는 N개의 감지 사이클들의 각자의 감지 사이클들로부터의 타일 데이터의 N개의 어레이들의 시퀀스 상에서 신경 네트워크의 런이 동작하고 있다. 데이터 흐름 로직(1997)은 N개의 감지 사이클들의 각자의 감지 사이클들로부터 N개의 어레이들의 공간적으로 정렬된 패치들에 대한 데이터를 포함하는 입력 유닛들을 사용하여 신경 네트워크의 런들을 위해 타일 데이터 및 트레이닝된 파라미터들을 메모리로부터 신경 네트워크 프로세서로 이동시키도록 제공된다.

[0203] 또한, 신경 네트워크 프로세서가 메모리에 액세스하고 복수의 실행 클러스터들을 포함하는 시스템이 기술되는데, 복수의 실행 클러스터들 중의 실행 클러스터들은 신경 네트워크를 실행시키도록 구성된다. 데이터 흐름 로직(1997)은 메모리 및 복수의 실행 클러스터들 중의 실행 클러스터들에 액세스하여, 타일 데이터의 입력 유닛들을 복수의 실행 클러스터들 중 이용가능한 실행 클러스터들에 제공하고 - 입력 유닛들은 대상 감지 사이클을 포함하는 각자의 감지 사이클들로부터 타일 데이터의 어레이들의 N개의 공간적으로 정렬된 패치들을 포함함 -, 실행 클러스터들이 N개의 공간적으로 정렬된 패치들을 신경 네트워크에 적용하여 대상 감지 사이클의 공간적으로 정렬된 패치에 대한 분류 데이터의 출력 패치들을 생성하게 하며, 여기서 N은 1 초과이다.

[0204] 도 20a는 호스트 프로세서에 의해 실행되는 런타임 프로그램(예컨대, 런타임 로직(1980))의 기능들을 포함한, 염기 호출 동작의 태양들을 보여주는 단순화된 도면이다. 이러한 도면에서, 플로우 셀로부터의 이미지 센서들의 출력은 라인들(2000) 상에서 이미지 프로세싱 스레드들(2001)에 제공되는데, 이는 개별 타일들에 대한 센서

데이터의 어레이에서의 정렬 및 배열 그리고 이미지들의 리샘플링과 같은 이미지들에 대한 프로세스들을 수행할 수 있고, 플로우 셀 내의 각각의 타일에 대한 타일 클러스터 마스크를 계산하는 프로세스들에 의해 사용될 수 있으며, 이는 플로우 셀의 대응하는 타일 상의 유전 물질의 클러스터들에 대응하는 센서 데이터의 어레이 내의 픽셀들을 식별한다. 이미지 프로세싱 스레드들(2001)의 출력들은 라인들(2002) 상에서 CPU 내의 디스패치 로직(2003)에 제공되는데, 이는 염기 호출 동작의 상태에 따라, 도 19c의 구성가능한 프로세서(1946)와 같은, 신경 네트워크 프로세서 하드웨어(2007)로의 고속 버스(2004) 상에서 또는 고속 버스(2006) 상에서 데이터 캐시(2005)(예컨대, SSD 저장소)로 타일 데이터의 어레이들을 라우팅한다. 프로세싱되고 변환된 이미지들은 이전에 사용되었던 감지 사이클들 동안 데이터 캐시(2005) 상에 저장될 수 있다. 하드웨어(2007)는 신경 네트워크에 의해 출력된 분류 데이터를 디스패치 로직(2003)에 반환하는데, 이는 데이터 캐시(2005)로, 또는 라인들(2008) 상에서, 분류 데이터를 사용하여 염기 호출 및 품질 스코어 계산들을 수행하는 스레드들(2009)로 정보를 전달하고, 염기 호출 판독물들에 대한 표준 포맷들로 데이터를 배열할 수 있다. 염기 호출 및 품질 스코어 계산들을 수행하는 스레드들(2009)의 출력들은, 라인들(2010) 상에서, 염기 호출 판독물들을 집적하고, 데이터 압축과 같은 다른 동작들을 수행하고, 고객들에 의한 활용을 위해, 생성된 염기 호출 출력들을 특정된 목적지들에 기록하는 스레드들(2011)로 제공된다.

[0205] 일부 구현예들에서, 호스트는 신경 네트워크의 지원 시에 하드웨어(2007)의 출력의 최종 프로세싱을 수행하는 스레드들(도시되지 않음)을 포함할 수 있다. 예를 들어, 하드웨어(2007)는 다중 클러스터 신경 네트워크의 최종 층으로부터의 분류 데이터의 출력들을 제공할 수 있다. 호스트 프로세서는 염기 호출 및 품질 스코어 스레드들(2002)에 의해 사용하기 위한 데이터를 구성하기 위해 분류 데이터에 걸쳐, 소프트맥스 함수와 같은 출력 활성화 함수를 실행할 수 있다. 또한, 호스트 프로세서는 하드웨어(2007)에 입력되기 전에 타일 데이터의 배치 정규화와 같은 입력 동작들(도시되지 않음)을 실행할 수 있다.

[0206] 도 20b는 도 19c의 것과 같은 구성가능한 프로세서(1946)의 구성의 단순화된 도면이다. 도 20b에서, 구성가능한 프로세서(1946)는 복수의 고속 PCIe 인터페이스들을 갖는 FPGA를 포함한다. FPGA는 도 19c를 참조하여 기술된 데이터 흐름 로직(1997)을 포함하는 래퍼(wrapper)(2090)를 갖도록 구성된다. 래퍼(2090)는 CPU 통신 링크(2077)를 통해 CPU에서의 런타임 프로그램과의 협력 및 인터페이스를 관리하고, DRAM 통신 링크(2097)를 통해 온-보드 DRAM(2099)(예컨대, 메모리(1448A))과의 통신을 관리한다. 래퍼(2090) 내의 데이터 흐름 로직(1997)은, N개의 사이클들 동안 온-보드 DRAM(2099) 상에서 타일 데이터의 어레이들을 클러스터(2085)로 순회시킴으로써 검색된 패치 데이터를 제공하고, 온-보드 DRAM(2099)으로 다시 전달하기 위해 클러스터(2085)로부터 프로세스 데이터(2087)를 검색한다. 래퍼(2090)는, 또한, 타일 데이터의 입력 어레이들 및 분류 데이터의 출력 패치들 둘 모두에 대해, 온-보드 DRAM(2099)과 호스트 메모리 사이에서의 데이터의 이송을 관리한다. 래퍼는 라인(2083) 상의 패치 데이터를 할당된 클러스터(2085)로 이송한다. 래퍼는 라인(2081) 상의 가중치들 및 바이어스들과 같은 트레이닝된 파라미터들을 온-보드 DRAM(2099)으로부터 검색된 클러스터(2085)에 제공한다. 래퍼는 라인(2079) 상의 구성 및 제어 데이터를, CPU 통신 링크(2077)를 통해 호스트 상의 런타임 프로그램으로부터 제공되거나 그에 응답하여 생성되는 클러스터(2085)에 제공한다. 클러스터는, 또한, 라인(2089) 상의 상태 신호들을 래퍼(2090)에 제공할 수 있는데, 이들은 타일 데이터의 어레이들의 순회를 관리하여 공간적으로 정렬된 패치 데이터를 제공하기 위해 그리고 클러스터(2085)의 리소스들을 사용하여 패치 데이터에 걸쳐서 다중-사이클 신경 네트워크를 실행하기 위해 호스트로부터의 제어 신호들과 협력하여 사용된다.

[0207] 상기에서 언급된 바와 같이, 타일 데이터의 다수의 패치들 중 대응하는 패치들 상에서 실행하기 위해 구성된, 래퍼(2090)에 의해 관리되는 단일의 구성가능한 프로세서 상에 다수의 클러스터들이 있을 수 있다. 각각의 클러스터는 본 명세서에 기술된 다수의 감지 사이클들의 타일 데이터를 사용하여 대상 감지 사이클에서 염기 호출들에 대한 분류 데이터를 제공하도록 구성될 수 있다.

[0208] 시스템의 예들에서, 필터 가중치들 및 바이어스들과 같은 커널 데이터를 포함하는 모델 데이터가 호스트 CPU로부터 구성가능한 프로세서로 전송되어, 모델이 사이클 수의 함수로서 업데이트될 수 있게 할 수 있다. 염기 호출 동작은, 대표적인 예의 경우에, 대략 수백개의 감지 사이클들을 포함할 수 있다. 염기 호출 동작은, 일부 구현예들에서, 쌍형성된 최종 판독물들을 포함할 수 있다. 예를 들어, 모델 트레이닝된 파라미터들은 20개의 사이클들(또는 다른 수의 사이클들)마다 1회, 또는 특정 시스템들 및 신경 네트워크 모델들에 대해 구현되는 업데이트 패턴들에 따라 업데이트될 수 있다. 타일 상의 유전 클러스터 내의 주어진 스트링에 대한 서열이 제1 단부로부터 스트링 아래로(또는 위로) 연장되는 제1 부분, 및 제2 단부로부터 스트링 위로(또는 아래로) 연장되는 제2 부분을 포함하는 쌍형성된 최종 판독물들을 포함하는 일부 구현예들에서, 트레이닝된 파라미터들은 제1 부분으로부터 제2 부분으로의 전이 시에 업데이트될 수 있다.

[0209] 일부 예들에서, 타일에 대한 감지 데이터의 다수의 사이클들 동안의 이미지 데이터는 CPU로부터 래퍼(2090)로 전송될 수 있다. 래퍼(2090)는, 선택적으로, 감지 데이터의 일부 사전프로세싱 및 변환을 행할 수 있고, 정보를 온-보드 DRAM(2099)에 기록할 수 있다. 각각의 감지 사이클 동안의 입력 타일 데이터는, 타일당 감지 사이클당 4000 x 3000개 정도 또는 그 이상의 픽셀들 - 2개의 특징부들이 타일의 2개의 이미지들의 컬러들을 표현함 -, 및 픽셀당 특징부당 1 또는 2 바이트를 포함하는 센서 데이터의 어레이들을 포함할 수 있다. 수 N이 다중-사이클 신경 네트워크의 각각의 런에 사용될 3개의 감지 사이클들인 구현예의 경우, 다중-사이클 신경 네트워크의 각각의 런에 대한 타일 데이터의 어레이는 타일당 수백 메가바이트 정도를 소비할 수 있다. 시스템의 일부 구현예들에서, 타일 데이터는, 또한, 타일당 1회 저장된 클러스터 중심에서의 거리(distance-from-cluster center, DFC) 데이터의 어레이, 또는 센서 데이터 및 타일들에 관한 다른 유형의 메타데이터를 포함한다.

[0210] 동작 시에, 다중-사이클 클러스터가 이용가능할 때, 래퍼는 클러스터에 패치를 할당한다. 래퍼는 타일의 순회에서 타일 데이터의 다음 패치를 페치(fetch)하고, 적절한 제어 및 구성 정보와 함께 그것을 할당된 클러스터로 전송한다. 클러스터는, 적소에 작동 중인, 일부 시스템들에서 다수의 사이클들로부터의 패치들을 포함하는 데이터의 패치, 및 프로세싱의 현재 패치가 다양한 구현예들에서 핑퐁 버퍼(ping-pong buffer) 기법 또는 래스터 스캐닝 기법을 사용하여 마무리될 때 작동되어야 하는 데이터의 패치를 보유하기에 충분한 메모리를 구성가능한 프로세서 상에 갖도록 구성될 수 있다.

[0211] 할당된 클러스터가 현재 패치에 대한 신경 네트워크의 그의 런을 완료하고 출력 패치를 생성할 때, 그것은 래퍼를 시그널링할 것이다. 래퍼는 할당된 클러스터로부터의 출력 패치를 판독할 것이거나, 또는 대안으로, 할당된 클러스터는 데이터를 래퍼로 푸시아웃할 것이다. 이어서, 래퍼는 DRAM(2099) 내의 프로세싱된 타일에 대한 출력 패치들을 어셈블할 것이다. 전체 타일의 프로세싱이 완료되었고, 데이터의 출력 패치들이 DRAM으로 이송되었을 때, 래퍼는 타일에 대한 프로세싱된 출력 어레이를 지정된 포맷으로 호스트/CPU로 다시 전송한다. 일부 구현예들에서, 온-보드 DRAM(2099)은 래퍼(2090) 내의 메모리 관리 로직에 의해 관리된다. 런타임 프로그램은, 실시간 분석을 제공하기 위해 연속적인 흐름 중 런에서의 모든 사이클들 동안의 타일 데이터의 모든 어레이들의 분석을 완료하도록 서열분석 동작들을 제어할 수 있다.

[0212] 도 21은 개시된 데이터 흐름 로직의 다른 구현예를 도시하는데, 이는 압축된 공간 맵들을 작성하고, 제1 염기 호출 반복 동안 생성되고, 오프-칩 메모리(2116)(예컨대, 오프-칩 DRAM, 호스트 RAM, 호스트 고대역폭 메모리(HBM))로부터 제2 염기 호출 반복 동안 이용가능하다.

[0213] 하나의 구현예에서, 호스트 프로세서(예컨대, CPU(1952))에 부착된 호스트 메모리(예컨대, 메모리(1948B))는 서열분석 런이 진행됨에 따라 서열분석 이미지들(2102)의 진행을 수신하도록 구성된다. 구성가능한 프로세서(예컨대, 구성가능한 프로세서(1946))는 프로세싱 유닛들의 어레이를 갖는다. 프로세싱 유닛들의 어레이 내의 프로세싱 유닛들은 신경 네트워크 기반 염기 호출자(100)를 실행하여 염기 호출 예측들을 생성하도록 구성된다. 데이터 흐름 로직(1997)은 호스트 메모리, 호스트 프로세서, 및 구성가능한 프로세서에 액세스한다. 제1 염기 호출 반복 동안, 데이터 흐름 로직(1997)은 호스트 메모리로부터 구성가능한 프로세서 상에서 서열분석 사이클들(예컨대, 도 1a의 서열분석 사이클들 1 내지 5)의 제1 윈도우 내의 서열분석 사이클들 동안 서열분석 이미지들을 로딩한다.

[0214] 런타임 로직(1980)은, 구성가능한 프로세서의 프로세싱 유닛들이 서열분석 이미지들(2102) 상에서 신경 네트워크 기반 염기 호출자(100)의 공간 네트워크(104)를 사이클 단위로 실행하게 하고 서열분석 사이클들의 제1 윈도우 내의 서열분석 사이클들 각각 동안 공간 특징부 맵 세트들(2106)을 생성하게 하도록 구성된다. 하나의 구현예에서, 런타임 로직(1980)은 서열분석 이미지들(2102)로부터 타일링된 패치들(2104) 상에서 신경 네트워크 기반 염기 호출자(100)의 다수의 프로세싱 클러스터들을 병렬로 실행시킨다. 다수의 프로세싱 클러스터들은 패치들(2104) 상에 공간 네트워크(104)를 패치 단위(2105)로 적용한다.

[0215] 런타임 로직(1980)은, 구성가능한 프로세서의 프로세싱 유닛들이 공간 특징부 맵 세트들(2106)에 대해 신경 네트워크 기반 염기 호출자(100)의 압축 네트워크(108)를 사이클 단위로 실행하여 압축된 공간 특징부 맵 세트들(2107)을 생성하게 하도록, 그리고 시간 네트워크(160) 및 출력 네트워크(190)를 통해 압축된 공간 특징부 맵 세트들(2107)을 프로세싱하여 서열분석 사이클들의 제1 윈도우 내의 하나 이상의 서열분석 사이클들 동안 염기 호출 예측들(2111)을 생성하게 하도록 구성된다. 시간 네트워크(160)는 시간 특징부 맵들(2108)을 생성한다. 출력 네트워크(190)는 염기 호출 분류 스코어들(2110)(예컨대, 정규화되지 않은 염기별 스코어들)을 생성한다. 하나의 구현예에서, 압축된 공간 특징부 맵 세트들(2107)은 오프-칩 메모리(2116)에 저장된다.

[0216] 하나의 구현예에서, 데이터 흐름 로직(1997)은 압축된 공간 특징부 맵 세트들(2107)을 호스트 메모리(2116)로

이동시키도록, 그리고 서열분석 이미지들(2102) 중 대응하는 서열분석 이미지들을 압축된 공간 특징부 맵 세트들(2107)로 겹쳐쓰기하도록 구성된다. 다른 구현예들에서, 패치들(2104) 중 대응하는 패치들은 압축된 공간 특징부 맵 세트들(2107)에 의해 대체된다.

[0217] 제2 염기 호출 반복 동안, 그리고 하나 이상의 중첩 서열분석 사이클들(예컨대, 서열분석 사이클들 2 내지 5)을 서열분석 사이클의 제1 윈도우와 공유하고 적어도 하나의 비중첩 서열분석 사이클(예컨대, 서열분석 사이클 6)을 갖는 서열분석 사이클들(예컨대, 도 2a의 서열분석 사이클들 2 내지 6)의 제2 윈도우 동안, 데이터 흐름 로직(1997)은 호스트 메모리로부터 구성가능한 프로세서 상에, 중첩 서열분석 사이클들 동안에는 압축된 공간 특징부 맵 세트들(2126)을, 그리고 비중첩 서열분석 사이클 동안에는 서열분석 이미지들(2122)(또는 패치들(2124))을 로딩하도록 구성된다.

[0218] 런타임 로직(1980)은 구성가능한 프로세서의 프로세싱 유닛들이 비중첩 서열분석 사이클 동안 서열분석 이미지들(2122)에 대해 공간 네트워크(104)를 실행하게 하도록, 그리고 비중첩 서열분석 사이클 동안 공간 특징부 맵 세트(2126)를 생성하도록 구성된다. 하나의 구현예에서, 다수의 프로세싱 클러스터들은 패치들(2124) 상에 공간 네트워크(104)를 패치 단위(2125)로 적용한다.

[0219] 런타임 로직(1980)은 구성가능한 프로세서의 프로세싱 유닛들이 공간 특징부 맵 세트(2126)에 대해 압축 네트워크(108)를 실행하고 비중첩 서열분석 사이클 동안 압축된 공간 특징부 맵 세트(2127)를 생성하게 하도록, 그리고 시간 네트워크(160) 및 출력 네트워크(190)를 통해 중첩 서열분석 사이클들 동안에는 압축된 공간 특징부 맵 세트들(2126)을 그리고 비중첩 서열분석 사이클 동안에는 압축된 공간 특징부 맵 세트(2127)를 프로세싱하여 서열분석 사이클들의 제2 윈도우 내의 하나 이상의 서열분석 사이클들 동안 염기 호출 예측들(2131)을 생성하도록 구성된다. 시간 네트워크(160)는 시간 특징부 맵들(2128)을 생성한다. 출력 네트워크(190)는 염기 호출 분류 스코어들(2129)(예컨대, 정규화되지 않은 염기별 스코어들)을 생성한다. 하나의 구현예에서, 압축된 공간 특징부 맵 세트(2127)는 오프-칩 메모리(2116)에 저장된다.

[0220] 도 22는 제1 염기 호출 반복 동안 생성된 압축된 공간 맵들을, 온-칩 메모리(2216)(예컨대, 온-칩 DRAM, 온-칩 SRAM, 온-칩 BRAM, 상호접속부를 통해 프로세서에 부착되는 DRAM과 같은 프로세서 메모리)로부터 제2 염기 호출 반복 동안 이용가능하게 하는 개시된 데이터 흐름 로직의 하나의 구현예를 도시한다. 도 22에서, 압축된 공간 특징부 맵 세트들(2107) 및 압축된 공간 특징부 맵 세트(2127)는 온-칩 메모리(2216)에 저장된다. 또한, 도 22에서, 데이터 흐름 로직(1997)은 온-칩 메모리(2216)로부터 구성가능한 프로세서 상에, 중첩 서열분석 사이클들 동안 압축된 공간 특징부 맵 세트들(2126)을 로딩하도록 구성된다.

[0221] 분할 아키텍처

[0222] 도 23은 신경 네트워크 기반 염기 호출자(100)의 소위 분할 아키텍처의 하나의 구현예를 도시한다. 상기에서 논의된 바와 같이, 공간 콘볼루션 네트워크(104)는 공간 콘볼루션 층들의 각자의 시퀀스들(2301, 2302, 2303, 2304, 2405)을 통해 사이클별 서열분석 이미지 세트들의 윈도우 내의 각자의 사이클별 서열분석 이미지 세트들을 개별적으로 콘볼루션함으로써 서열분석 런의 일련의 서열분석 사이클들(사이클들 N+2, N+1, N, N-1, 및 N-2) 동안 사이클별 서열분석 이미지 세트들의 윈도우를 사이클 단위로 프로세싱하여, 일련의 서열분석 사이클들 중의 각자의 서열분석 사이클들 동안 각자의 사이클별 공간 특징부 맵 세트들을 생성하도록 구성된다. 예를 들어, 공간 콘볼루션 층들의 5개의 시퀀스들(2301, 2302, 2303, 2304, 2405) 각각은 7개의 공간 콘볼루션 층들(즉, 도 23의 층들 L1 내지 L7)을 갖는다.

[0223] 공간 콘볼루션 층들의 각자의 시퀀스들(2301, 2302, 2303, 2304, 2405)은 공간 콘볼루션 필터 뱅크들의 각자의 시퀀스들(예컨대, 공간 콘볼루션 층들의 시퀀스(2301)에 대한 공간 콘볼루션 필터 뱅크들(2310, 2311, 2312, 2313, 2314, 2315, 2316)을 포함하는 시퀀스)을 갖는다. 하나의 구현예에서, 공간 콘볼루션 필터 뱅크들의 각자의 시퀀스들의 공간 콘볼루션 필터 뱅크들 내의 공간 콘볼루션 필터들의 트레이닝된 계수들(또는 가중치들)이 공간 콘볼루션 층들의 각자의 시퀀스들 내의 공간 콘볼루션 층들의 시퀀스들 사이에서 변한다.

[0224] 예를 들어, 공간 콘볼루션 층 시퀀스들(2301, 2302, 2303, 2304, 2405)은 상이한 트레이닝된 계수들을 갖는 콘볼루션 필터들로 구성된다. 다른 예에서, 대응하는 레벨의 공간 콘볼루션 층들 내의 콘볼루션 필터들은 상이한 트레이닝된 계수들(예컨대, 공간 콘볼루션 층들의 5개의 시퀀스들(2301, 2302, 2303, 2304, 2405)의 각자의 제3 공간 콘볼루션 층들 내의 콘볼루션 필터 뱅크들(2382, 2383, 2384, 2385, 2312))을 갖는다.

[0225] 시간 콘볼루션 네트워크(160)는 제1 시간 콘볼루션 층(2320)의 각자의 시간 콘볼루션 필터 뱅크들(2321, 2322, 2323)을 사용하여 사이클별 공간 특징부 맵 세트들 내의 사이클별 공간 특징부 맵 세트들의 각자의 중첩 그룹들

(예컨대, 그룹들(2360, 2361, 2362))을 콘볼루션함으로써 그룹별 단위로 사이클별 공간 특징부 맵 세트들을 프로세싱하여, 사이클별 공간 특징부 맵 세트들의 각자의 중첩 그룹들에 대한 각자의 그룹별 시간 특징부 맵 세트들을 생성하도록 구성된다. 하나의 구현예에서, 각자의 시간 콘볼루션 필터 뱅크들 내의 시간 콘볼루션 필터들의 트레이닝된 계수들(또는 가중치들)은 각자의 시간 콘볼루션 필터 뱅크들 중의 시간 콘볼루션 필터 뱅크들(2321, 2322, 2323) 사이에서 변한다.

[0226] 스킵 아키텍처

[0227] 도 24a는 특징부 맵 추가를 통해 이전 정보를 다운스트림으로 재주입하는 잔차(또는 스킵) 접속을 도시한다. 잔차 접속은 과거 출력 텐서를 이후 출력 텐서에 추가함으로써 데이터의 다운스트림 흐름에 이전 표현들을 재주입하는 것을 포함하는데, 이는 데이터 프로세싱 흐름을 따라 정보 손실을 방지하는 데 도움이 된다. 잔차 접속들은 임의의 대규모 심층 학습 모델을 번거롭게 만드는(plague) 2개의 보편적인 문제들을 다룬다: 사라지는 기울기들 및 표현상의 병목현상들.

[0228] 잔차 접속은 이전 층의 출력을 이후 층에 대한 입력으로서 이용가능하게 만들어, 순차적 네트워크 내에 지름길(shortcut)을 효과적으로 생성하는 것을 포함한다. 이후 활성화에 연결되기보다는, 이전의 출력은 이후 활성화와 합산되는데, 이는 활성화들 둘 모두가 동일한 크기임을 가정한다. 이들이 상이한 크기들을 갖는 경우, 이전 활성화를 표적 형상으로 재형상화하기 위한 선형 변환이 사용될 수 있다.

[0229] 도 24b는 잔차 블록들 및 스킵 접속들의 하나의 구현예를 도시한다. 잔차 네트워크는 트레이닝 정확도의 저하를 완화시키기 위해 다수의 잔차 유닛들을 적층한다. 잔차 블록들은 심층 신경 네트워크들에서 사라지는 기울기들을 방지하기 위해 특수한 부가적 스킵 접속들을 이용한다. 잔차 블록의 시작부에서, 데이터 흐름은 2개의 스트림들로 분리된다: 제1 스트림은 블록의 변경되지 않은 입력을 전달하는 반면, 제2 스트림은 가중치들 및 비선형성들을 적용한다. 블록의 종단에서, 2개의 스트림들은 요소별 합을 사용하여 병합된다. 그러한 구성체들의 주요 이점은 기울기가 네트워크를 통해 더 용이하게 흐를 수 있게 하는 것이다.

[0230] 잔차 네트워크로 구성되면, 일부 구현예들에서, 신경 네트워크 기반 염기 호출자(100)가 용이하게 트레이닝되고, 이미지 분류 및 객체 검출에 대한 개선된 정확도가 달성될 수 있다. 신경 네트워크 기반 염기 호출자(100)는 $(l+1)^{th}$ 층에 대한 입력으로서 l^{th} 층의 출력과 접속하는데, 이는 하기의 층 전이를 일으킨다: $x_l = H_l(x_{l-1})$. 잔차 블록들은 식별 기능을 갖는 비선형 변환들을 우회시키는 스킵 접속을 추가한다: $x_l = H_l(x_{l-1}) + x_{l-1}$. 잔차 블록들의 이점은 기울기가 아이덴티티 기능을 통해 이후 층들로부터 이전 층들(예컨대, 공간 및 시간 콘볼루션 층들)로 직접 흐를 수 있게 한다는 것이다. 아이덴티티 기능 및 H_l의 출력은 합산(덧셈)에 의해 조합된다.

[0231] 도 24c는 공간 콘볼루션 층들이 스킵 접속들을 갖는 잔차 블록들로 그룹화되는 신경 네트워크 기반 염기 호출자(100)의 잔차 아키텍처를 도시한다. 다른 구현예들에서, 신경 네트워크 기반 염기 호출자(100)의 시간 콘볼루션 층들은 스킵 접속들을 갖는 잔차 블록들로 그룹화된다.

[0232] 도 24c에 도시된 구현예에서, 제2 및 제3 공간 콘볼루션 층들은 제1 잔차 블록(2412)으로 그룹화되고; 제4 및 제5 공간 콘볼루션 층들은 제2 잔차 블록(2422)으로 그룹화되고; 제6 및 제7 공간 콘볼루션 층들은 제3 잔차 블록(2432)으로 그룹화된다.

[0233] 도 25a는 신경 네트워크 기반 염기 호출자(100)의 버스 네트워크의 세부사항들을 도시한다. 하나의 구현예에서, 버스 네트워크의 주어진 잔차 블록(2585)은 공간 콘볼루션 층들(2590, 2592)의 세트를 포함한다. 공간 콘볼루션 층들의 세트 내의 제1 공간 콘볼루션 층(2590)은 주어진 잔차 블록(2585)의 일부가 아닌 선행 공간 콘볼루션 층(예컨대, 공간 네트워크(104)에서 제1 공간 콘볼루션 층(2590)에 선행하는 제로 공간 콘볼루션 층)에 의해 생성된 선행 출력(2586)을 입력으로 수신한다. 제1 공간 콘볼루션 층(2590)은 선행 출력(2586)을 프로세싱하고 제1 출력(2591)을 생성한다. 제1 공간 콘볼루션 층(2590)에 후행하는 공간 콘볼루션 층들의 세트 내의 제2 공간 콘볼루션 층(2592)은 제1 출력(2591)을 수신하고, 제1 출력(2591)을 프로세싱하고 제2 출력(2593)을 생성한다. 하나의 구현예에서, 제1 공간 콘볼루션 층(2590)은 제1 출력(2591)을 생성하는 ReLU와 같은 비선형 활성화 함수를 갖는다. 다른 구현예에서, 제2 공간 콘볼루션 층(2592)은 비선형 활성화 함수가

없다.

[0234] 스킵 접속(2589)은 선행 출력(2586)을 합산기(2594)에 제공한다. 합산기(2594)는 또한, 제2 공간 콘볼루션 층(2592)으로부터 제2 출력(2593)을 수신한다. 합산기(2594)는 선행 출력(2586)과 제2 출력(2593)을 조합하고, 합산된 출력(2595)을 생성한다. 합산된 출력(2595)은 ReLU와 같은 비선형 활성화를 통해 추가로 프로세싱되어 최종 합산된 출력(2587)을 생성한다. 이어서, 일부 구현예들에서, 최종 합산된 출력(2587)은 후행 잔차 블록에 대한 입력으로서 공급된다. 일부 구현예들에서, 선행 출력(2586)은 제2 출력(2593)과 디멘션수-호환가능하도록 수정된다. 예를 들어, 선행 출력(2586)에서의 특징부 맵들의 에지들은 제2 출력(2593)에서의 특징부 맵들과 동일한 공간 디멘션수를 갖는 특징부 맵들을 생성하도록 트리밍된다.

[0235] 도 25b는 개시된 버스 네트워크의 예시적인 동작을 도시한다. 하나의 구현예에서, 버스 네트워크는 공간 콘볼루션 층들의 각자의 시퀀스들 내의 공간 콘볼루션 층들 사이에 버스들(예컨대, 2516, 2526, 2536, 2602, 2604, 2702, 2712)을 형성하도록 구성된다. 버스들은, 특정 서열분석 사이클에 대한 공간 콘볼루션 층의 특성 시퀀스 내의 2개 이상의 공간 콘볼루션 층들에 의해 생성된 각자의 사이클별 공간 특징부 맵 세트들이 조합된 사이클별 공간 특징부 맵 세트로 조합되게 하도록, 그리고 조합된 사이클별 공간 특징부 맵 세트를 공간 콘볼루션 층의 특정 시퀀스 내의 다른 공간 콘볼루션 층에 대한 입력으로서 제공하도록 구성된다.

[0236] 예를 들어, 제1 잔차 블록(2412)을 고려한다. 여기서, 2개 이상의 공간 콘볼루션 층들은 제1 공간 콘볼루션 층 및 제3 공간 콘볼루션 층을 포함한다. 제1 공간 콘볼루션 층은 제1 사이클별 공간 특징부 맵 세트(2520)를 생성한다. 제1 공간 콘볼루션 층은 제1 사이클별 공간 특징부 맵 세트(2520)를 제2 공간 콘볼루션 층에 대한 입력으로서 제공한다. 제2 공간 콘볼루션 층은 제1 사이클별 공간 특징부 맵 세트(2520)를 프로세싱하고 제2 사이클별 공간 특징부 맵 세트(2522)를 생성한다. 제2 공간 콘볼루션 층은 제2 사이클별 공간 특징부 맵 세트(2522)를 제3 공간 콘볼루션 층에 대한 입력으로서 제공한다. 제3 공간 콘볼루션 층은 제2 사이클별 공간 특징부 맵 세트(2522)를 프로세싱하고 제3 사이클별 공간 특징부 맵 세트(2524)를 생성한다. 버스들(예컨대, 스킵 버스(2519))은 제1 공간 특징부 맵 세트(2520) 및 제3 사이클별 공간 특징부 맵 세트(2524)가 조합된 사이클별 공간 특징부 맵 세트(2518)로 조합되게 하도록(예컨대, 조합기(2502)에 의해 합산되거나 연결됨) 추가로 구성된다. 이어서, 다른 공간 콘볼루션 층은 공간 콘볼루션 층의 특정 시퀀스 내의 제3 공간 콘볼루션 층에 바로 후행하는 제4 공간 콘볼루션 층이다. 제4 공간 콘볼루션 층은 조합된 사이클별 공간 특징부 맵 세트(2518)를 입력으로서 프로세싱한다. 동일한 개념이 제2 및 제3 잔차 블록들(2422, 2432)에 유사하게 적용되며, 여기서 2526 및 2536은 스킵 버스들(2516)과 같은 스킵 버스들이고, 각자의 조합기들(2512, 2532)이 각자의 조합된 사이클별 공간 특징부 맵 세트들(2528, 2538)을 생성하게 한다.

[0237] 도 25c는 조합 전에, 스킵 버스들에 의해 공급된 유입 특징부 맵들이, 유입 특징부 맵들이 버스 네트워크의 조합기들에 의해 조합되는 수신하는 특징부 맵들과 동일한 공간 디멘션수를 갖도록 수정됨(예컨대, 트리밍됨)을 보장하는 디멘션 호환성 로직(2532)의 하나의 구현예를 도시한다.

[0238] 도 26은 스킵 버스(2602)가 제1 공간 콘볼루션 층의 출력을 제1 잔차 블록의 출력과 조합하는 개시된 버스 네트워크의 다른 예를 도시한다. 도 26은 또한, 스킵 버스(2604)가 또한, 특징부 맵들이 잔차 블록들에 걸쳐서 그리고 불연속 층들에 걸쳐서(예컨대, 층 1 및 층 5로부터) 조합되어 다른 층에 의해(예컨대, 층 6까지) 프로세싱될 수 있는 조합된 표현을 생성하게 할 수 있음을 도시한다.

[0239] 도 27은 다수의 연속 및/또는 불연속 층들로부터의(예컨대, 층 1, 조합기(2502), 및 조합기(2512)로부터의) 입력들 및 조합된 표현들이 예시적인 스킵 버스들(2702, 2604, 2712)에 의해 조합되어 다른 층에 의해(예컨대, 층 6까지) 프로세싱될 수 있는 조합된 표현을 생성할 수 있는 개시된 버스 네트워크의 또 다른 예를 도시한다.

[0240] 도 28은 스킵 버스들에 의해 공급된 유입 특징부 맵들을, 그들이 버스 네트워크의 조합기들에 의해 조합되는 수신하는 특징부 맵들과 그들을 조합하기 전에 스케일링하는 스케일링 로직(2832)의 하나의 구현예를 도시한다. 스케일링 로직(2832)에 의해 사용되는 값들은 범위가, 예를 들어, 0 및 1을 포함한 0 내지 1 사이의 어딘가에 이를 수 있다. 스케일링 로직은, 예를 들어, 강도/크기/값(예컨대, 유입 특징부 맵들의 특징부 값들(예컨대, 부동소수점 값들))을 감쇠시키거나 증폭시키는 데 사용될 수 있다.

[0241] 도 29는 시간 네트워크(160)의 시간 콘볼루션 층들(2902, 2912, 2922, 2932, 2942, 2952, 2962, 2972) 사이의 스킵 접속들의 하나의 구현예를 도시한다. 예를 들어, 스킵 접속(2922)은 시간 특징부 맵들을 제1 시간 콘볼루션 층(2902)으로부터 제3 시간 콘볼루션 층(2932)으로 공급한다.

[0242] 도 30은 압축 로직(108)으로 구성된 네트워크 네트워크 기반 염기 호출자(100)(sqz2 염기 호출자)에 의한 염기

호출 성능을, (기준 신경 네트워크 모델들로서 사용되는) 압축 로직(108)이 없는 네트워크 네트워크 기반 염기 호출자(100)에 대해 그리고 (기준 전통적 이미지 프로세싱 모델로서 사용되는) Illumina의 비-신경 네트워크 기반 염기 호출자 실시간 분석(RTA) 소프트웨어에 대해 비교한다. 도 30의 차트에서 알 수 있는 바와 같이, sqz2 염기 호출자(자주색 피팅 라인)는, RTA 염기 호출자(흑색 피팅 라인), 및 압축 로직(108)이 없는 네트워크 네트워크 기반 염기 호출자(100)의 2개의 인스턴스들(적색 및 청록색 피팅 라인들)보다 더 낮은 염기 호출 에러 백분율(Y-축 상의 "에러 %")을 갖는다.

[0243] 도 31은 개시된 압축 로직(108)의 사용에 의해 RAM 및 DRAM 사용량의 절약이 유발되었음을 도시한다.

[0244] 도 32는 분할 및 스킵 아키텍처들로 구성된 네트워크 네트워크 기반 염기 호출자(split_res)(100)에 의한 염기 호출 성능을, RTA 염기 호출자 및 분할 및 스킵 아키텍처들이 없는 네트워크 네트워크 기반 염기 호출자(distilled)(100)의 다른 버전에 대해 비교한다. 도 32의 차트에서 알 수 있는 바와 같이, split_res 염기 호출자(오렌지색 피팅 라인)는, RTA 염기 호출자(청색 피팅 라인)보다 더 낮은 염기 호출 에러 백분율(Y-축 상의 "에러 카운트")을 갖는다.

[0245] 본 명세서에서 사용되는 바와 같이, "로직"(예컨대, 데이터 흐름 로직)은 본 명세서에 기술된 방법 단계들을 수행하기 위한 컴퓨터 사용가능 프로그램 코드를 갖는 비일시적 컴퓨터 판독가능 저장 매체를 포함하는 컴퓨터 제품의 형태로 구현될 수 있다. "로직"은, 메모리, 및 메모리에 커플링되고 예시적인 방법 단계들을 수행하기 위해 동작하는 적어도 하나의 프로세서를 포함하는 장치의 형태로 구현될 수 있다. "로직"은 본 명세서에 기술된 방법 단계들 중 하나 이상을 수행하기 위한 수단의 형태로 구현될 수 있고; 수단은 (i) 하드웨어 모듈(들), (ii) 하나 이상의 하드웨어 프로세서들 상에서 실행되는 소프트웨어 모듈(들), 또는 (iii) 하드웨어와 소프트웨어 모듈들의 조합을 포함할 수 있고; (i) 내지 (iii) 중 임의의 것이 본 명세서에 제시된 특정 기법들을 구현하고, 소프트웨어 모듈들은 컴퓨터 판독가능 저장 매체(또는 다수의 그러한 매체)에 저장된다. 하나의 구현예에서, 로직은 데이터 프로세싱 기능을 구현한다. 로직은 기능을 특정하는 컴퓨터 프로그램을 갖는 범용, 단일 코어 또는 다중 코어 프로세서, 컴퓨터 프로그램을 갖는 디지털 신호 프로세서, 구성 파일을 갖는 FPGA와 같은 구성가능한 로직, 상태 기계와 같은 특수 목적 회로, 또는 이들의 임의의 조합일 수 있다. 또한, 컴퓨터 프로그램 제품은 로직의 컴퓨터 프로그램 및 구성 파일 부분들을 구현할 수 있다.

[0246] 도 33은 본 명세서에 개시된 염기 호출 기법들을 구현하기 위해 서열분석 시스템(1900A)에 의해 사용될 수 있는 컴퓨터 시스템(3300)이다. 컴퓨터 시스템(3300)은 버스 서브시스템(3355)을 통해 다수의 주변 디바이스들과 통신하는 적어도 하나의 중앙 프로세싱 유닛(CPU)(3372)을 포함한다. 이러한 주변 디바이스들은, 예를 들어 메모리 디바이스들 및 파일 저장 서브시스템(3336)을 포함하는 저장 서브시스템(3358), 사용자 인터페이스 입력 디바이스들(3338), 사용자 인터페이스 출력 디바이스들(3376), 및 네트워크 인터페이스 서브시스템(3374)을 포함할 수 있다. 입력 및 출력 디바이스들은 컴퓨터 시스템(3300)과의 사용자 상호작용을 허용한다. 네트워크 인터페이스 서브시스템(3374)은 다른 컴퓨터 시스템들에서의 대응하는 인터페이스 디바이스들에 대한 인터페이스를 포함하는 인터페이스를 외부 네트워크들에 제공한다.

[0247] 하나의 구현예에서, 시스템 제어기(1906)는 저장 서브시스템(3310) 및 사용자 인터페이스 입력 디바이스들(3338)에 통신가능하게 링크된다.

[0248] 사용자 인터페이스 입력 디바이스들(3338)은 키보드; 마우스, 트랙볼, 터치패드, 또는 그래픽 태블릿과 같은 포인팅 디바이스들; 스캐너; 디스플레이 내에 통합된 터치 스크린; 음성 인식 시스템들 및 마이크로폰들과 같은 오디오 입력 디바이스들; 및 다른 유형들의 입력 디바이스들을 포함할 수 있다. 일반적으로, 용어 "입력 디바이스"의 사용은 정보를 컴퓨터 시스템(3300)에 입력하기 위한 모든 가능한 유형들의 디바이스들 및 방식들을 포함하도록 의도된다.

[0249] 사용자 인터페이스 출력 디바이스들(3376)은 디스플레이 서브시스템, 프린터, 팩스 기계, 또는 오디오 출력 디바이스들과 같은 비시각적 디스플레이들을 포함할 수 있다. 디스플레이 서브시스템은 LED 디스플레이, 음극선관(CRT), 액정 디스플레이(LCD)와 같은 평면 디바이스, 프로젝션 장치, 또는 가시적인 이미지를 생성하기 위한 일부 다른 메커니즘을 포함할 수 있다. 디스플레이 서브시스템은 또한, 오디오 출력 디바이스들과 같은 비시각적 디스플레이를 제공할 수 있다. 대체적으로, "출력 디바이스"라는 용어의 사용은 정보를 컴퓨터 시스템(3300)으로부터 사용자에게 또는 다른 기계 또는 컴퓨터 시스템에 출력하기 위한 모든 가능한 유형들의 디바이스들 및 방식들을 포함하도록 의도된다.

[0250] 저장 서브시스템(3358)은 본 명세서에 기술된 모듈들 및 방법들 중 일부 또는 전부의 기능을 제공하는 프로그래

밍 및 데이터 구성들을 저장한다. 이들 소프트웨어 모듈들은, 대체적으로, 심층 학습 프로세서들(3378)에 의해 실행된다.

[0251] 심층 학습 프로세서들(3378)은 그래픽 프로세싱 유닛(GPU)들, 필드 프로그래밍가능 게이트 어레이(FPGA)들, 주문형 반도체(ASIC)들, 및/또는 코어스-그레인드 재구성가능 아키텍처(CGRA)들일 수 있다. 심층 학습 프로세서들(3378)은 Google Cloud Platform™, Xilinx™, 및 Cirrascale™과 같은 심층 학습 클라우드 플랫폼에 의해 호스팅될 수 있다. 심층 학습 프로세서들(3378)의 예들은 Google의 Tensor Processing Unit(TPU)™, 랙마운트 솔루션들, 예컨대 GX4 Rackmount Series™, GX33 Rackmount Series™, NVIDIA DGX-1™, Microsoft의 Stratix V FPGA™, Graphcore의 Intelligent Processor Unit (IPU)™, Snapdragon processors™을 갖는 Qualcomm의 Zeroth Platform™, NVIDIA의 Volta™, NVIDIA의 DRIVE PX™, NVIDIA의 JETSON TX1/TX2 MODULE™, Intel의 Nirvana™, Movidius VPU™, Fujitsu DPI™, ARM의 DynamicIQ™, IBM TrueNorth™, Testa V100s™을 갖는 Lambda GPU 서버, SambaNova의 Reconfigurable Dataflow Unit (RDU)™ 등을 포함한다.

[0252] 저장 서브시스템(3358)에 사용되는 메모리 서브시스템(3322)은 프로그램 실행 동안 명령어들 및 데이터의 저장을 위한 메인 랜덤 액세스 메모리(RAM)(3332) 및 고정된 명령어들이 저장되는 판독 전용 메모리(ROM)(3334)를 포함하는 다수의 메모리들을 포함할 수 있다. 파일 저장 서브시스템(3336)은 프로그램 및 데이터 파일을 위한 영구 저장소를 제공할 수 있고, 하드 디스크 드라이브, 연관된 착탈 가능 매체와 함께 플로피 디스크 드라이브, CD-ROM 드라이브, 광학 드라이브, 또는 착탈 가능 매체 카트리지를 포함할 수 있다. 소정 구현예들의 기능을 구현하는 모듈들은 저장 서브시스템(3358) 내의 파일 저장 서브시스템(3336)에 의해, 또는 프로세서에 의해 액세스가능한 다른 기계들에 저장될 수 있다.

[0253] 버스 서브시스템(3355)은 컴퓨터 시스템(3300)의 다양한 컴포넌트들 및 서브시스템들이 의도된 대로 서로 통신하게 하기 위한 메커니즘을 제공한다. 버스 서브시스템(3355)이 개략적으로 단일 버스로서 도시되어 있지만, 버스 서브시스템의 대안적인 구현예들은 다수의 버스들을 사용할 수 있다.

[0254] 컴퓨터 시스템(3300) 자체는 개인용 컴퓨터, 휴대용 컴퓨터, 워크스테이션, 컴퓨터 단말기, 네트워크 컴퓨터, 텔레비전, 메인프레임, 서버 팜, 느슨하게 네트워킹된 컴퓨터들의 광범위하게 분포된 세트, 또는 임의의 다른 데이터 프로세싱 시스템 또는 사용자 디바이스를 포함한 다양한 유형들의 것일 수 있다. 컴퓨터들 및 네트워크들의 변화하는(ever-changing) 특성으로 인해, 도 33에 묘사된 컴퓨터 시스템(3300)의 설명은 본 발명의 바람직한 구현예들을 예시하기 위한 특정 예로서만 의도된다. 도 33에 묘사된 컴퓨터 시스템보다 더 많은 또는 더 적은 컴포넌트들을 갖는 컴퓨터 시스템(3300)의 많은 다른 구성들이 가능하다.

[0255] 항목들

[0256] 하기 항목들이 개시되어 있다:

[0257] *압축(스퀴즈)*

[0258] 1. 염기 호출의 인공 지능 기반 방법으로서,

[0259] 서열분석 런의 서열분석 사이클들 동안 생성된 일련의 사이클별 분석물 채널 세트들에 액세스하는 단계;

[0260] 신경 네트워크 기반 염기 호출자의 공간 네트워크를 통해, 서열분석 런의 서열분석 사이클들의 제1 윈도우 동안 일련의 사이클별 분석물 채널 세트들 내의 사이클별 분석물 채널 세트들의 제1 윈도우를 프로세싱하고 서열분석 사이클들의 제1 윈도우 내의 각자의 서열분석 사이클들 동안 공간 출력 세트들의 각자의 시퀀스들을 생성하는 단계;

[0261] 신경 네트워크 기반 염기 호출자의 압축 네트워크를 통해, 공간 출력 세트들의 각자의 시퀀스들 내의 각자의 최종 공간 출력 세트들을 프로세싱하고 서열분석 사이클들의 제1 윈도우 내의 각자의 서열분석 사이클들 동안 각자의 압축된 공간 출력 세트들을 생성하는 단계; 및

[0262] 각자의 압축된 공간 출력 세트들에 기초하여, 서열분석 사이클들의 제1 윈도우 내의 하나 이상의 서열분석 사이클들 동안 염기 호출 예측들을 생성하는 단계를 포함하는, 인공 지능 기반 방법.

[0263] 2. 항목 1에 있어서, 각자의 최종 공간 출력 세트들은 M개의 채널들(특징부 맵들)을 갖고, 각자의 압축된 공간 출력 세트들은 N개의 채널들(특징부 맵들)을 갖고, M > N인, 인공 지능 기반 방법.

[0264] 3. 항목 1에 있어서,

[0265] 공간 네트워크가 공간 출력 세트들을 이전에 생성한 하나 이상의 중첩 서열분석 사이클들, 및 공간 네트워크가 공간 출력 세트를 아직 생성하지 않은 적어도 하나의 비중첩 서열분석 사이클을, 서열분석 사이클들의 제1 윈도우와 공유하는 서열분석 런의 서열분석 사이클들의 제2 윈도우 동안,

[0266] 공간 네트워크를 통해, 비중첩 서열분석 사이클 동안만 사이클별 분석물 채널 세트를 프로세싱하고 비중첩 서열분석 사이클 동안 공간 출력 세트들의 시퀀스를 생성하여, 이에 의해, 공간 네트워크를 통해, 중첩 서열분석 사이클들 동안 각자의 사이클별 분석물 채널 세트들을 재프로세싱하는 것을 우회시키는 단계;

[0267] 압축 네트워크를 통해, 공간 출력 세트들의 시퀀스 내의 최종 공간 출력 세트를 프로세싱하고 비중첩 서열분석 사이클 동안 압축된 공간 출력 세트를 생성하는 단계 - 최종 공간 출력은 M개의 채널들(특징부 맵들)을 갖고, 압축된 공간 출력은 N개의 채널들(특징부 맵들)을 갖고, M > N임 -; 및

[0268] 서열분석 사이클들의 제1 윈도우 동안 이전에 생성된 중첩 서열분석 사이클들 동안의 각자의 압축된 공간 출력 세트들 및 압축된 공간 출력 세트에 기초하여, 서열분석 사이클들의 제2 윈도우 내의 하나 이상의 서열분석 사이클들 동안 염기 호출 예측들을 생성하여, 이에 의해, 중첩 서열분석 사이클들 동안의 각자의 압축된 공간 출력 세트들로 중첩 서열분석 사이클들 동안의 각자의 사이클별 분석물 채널 세트들을 대체하는 단계를 추가로 포함하는, 인공 지능 기반 방법.

[0269] 4. 항목 3에 있어서,

[0270] 공간 네트워크가 공간 출력 세트들을 이전에 생성한 하나 이상의 중첩 서열분석 사이클들, 및 공간 네트워크가 공간 출력 세트를 아직 생성하지 않은 적어도 하나의 비중첩 서열분석 사이클을, 서열분석 사이클들의 제1 및 제2 윈도우들과 공유하는 서열분석 런의 서열분석 사이클들의 제3 윈도우 동안,

[0271] 공간 네트워크를 통해, 비중첩 서열분석 사이클 동안만 사이클별 분석물 채널 세트를 프로세싱하고 비중첩 서열분석 사이클 동안 공간 출력 세트들의 시퀀스를 생성하여, 이에 의해, 공간 네트워크를 통해, 중첩 서열분석 사이클들 동안 각자의 사이클별 분석물 채널 세트들을 재프로세싱하는 것을 우회시키는 단계;

[0272] 압축 네트워크를 통해, 공간 출력 세트들의 시퀀스 내의 최종 공간 출력 세트를 프로세싱하고 비중첩 서열분석 사이클 동안 압축된 공간 출력 세트를 생성하는 단계 - 최종 공간 출력은 M개의 채널들(특징부 맵들)을 갖고, 압축된 공간 출력은 N개의 채널들(특징부 맵들)을 갖고, M > N임 -; 및

[0273] 서열분석 사이클들의 제1 및 제2 윈도우들 동안 이전에 생성된 중첩 서열분석 사이클들 동안의 각자의 압축된 공간 출력 세트들 및 압축된 공간 출력 세트에 기초하여, 서열분석 사이클들의 제3 윈도우 내의 하나 이상의 서열분석 사이클들 동안 염기 호출 예측들을 생성하여, 이에 의해, 중첩 서열분석 사이클들 동안의 각자의 압축된 공간 출력 세트들로 중첩 서열분석 사이클들 동안의 각자의 사이클별 분석물 채널 세트들을 대체하는 단계를 추가로 포함하는, 인공 지능 기반 방법.

[0274] 5. 항목 1에 있어서, 일련의 사이클별 분석물 채널 세트들 내의 각각의 사이클별 분석물 채널 세트는 서열분석 런에서의 대응하는 서열분석 사이클에서 분석물들 내의 뉴클레오티드 혼입에 응답하여 등록된 세기들을 묘사하는, 인공 지능 기반 방법.

[0275] 6. 항목 5에 있어서, 공간 네트워크는, 서열분석 런의 서열분석 사이클들의 특정 윈도우 동안 일련의 사이클별 분석물 채널 세트들 내의 사이클별 분석물 채널 세트들의 특정 윈도우 내에 각각의 사이클별 분석물 채널 세트를 개별적으로 프로세싱하고 서열분석 사이클들의 특정 윈도우 내의 각각의 서열분석 사이클 동안 공간 출력 세트들의 시퀀스를 생성하는 공간 콘볼루션 층들의 시퀀스를 가지며, 대상 서열분석 사이클의 사이클별 분석물 채널 세트 내에만 있고 서열분석 사이클들의 특정 윈도우 내의 상이한 서열분석 사이클들의 사이클별 분석물 채널 세트들 사이에 있지 않은 세기들을 조합하는 제1 공간 콘볼루션 층으로 시작하는 단계, 및 대상 서열분석 사이클 내에만 있고 서열분석 사이클들의 특정 윈도우 내의 상이한 서열분석 사이클들 사이에 있지 않은 선행 공간 콘볼루션 층들의 공간 출력들을 조합하는 연속 공간 콘볼루션 층들로 계속하는 단계를 포함하는, 인공 지능 기반 방법.

[0276] 7. 항목 6에 있어서, 공간 콘볼루션 층들의 시퀀스 내의 각자의 공간 콘볼루션 층들은 상이한 카운트들의 콘볼루션 필터들을 갖고, 공간 콘볼루션 층들의 시퀀스 내의 최종 공간 콘볼루션 층은 M개의 콘볼루션 필터들을 갖고, M은 4 초과의 정수인, 인공 지능 기반 방법.

[0277] 8. 항목 7에 있어서, 공간 콘볼루션 층들의 시퀀스 내의 각자의 공간 콘볼루션 층들은 동일한 카운트의 콘볼루

션 필터들을 갖고, 동일한 카운트는 M이고, M은 4 초과의 정수인, 인공 지능 기반 방법.

[0278] 9. 항목 8에 있어서, 공간 네트워크 내의 콘볼루션 필터들은 2차원(2D) 콘볼루션들을 사용하는, 인공 지능 기반 방법.

[0279] 10. 항목 8에 있어서, 공간 네트워크 내의 콘볼루션 필터들은 3차원(3D) 콘볼루션들을 사용하는, 인공 지능 기반 방법.

[0280] 11. 항목 6에 있어서, 신경 네트워크 기반 염기 호출자는 시간 네트워크를 갖고, 시간 네트워크는, 서열분석 사이클들의 특정 윈도우 내의 연속 서열분석 사이클들의 윈도우들 동안 각자의 압축된 공간 출력 세트들을 그룹별로 프로세싱하고 서열분석 사이클들의 특정 윈도우 동안 시간 출력 세트들의 시퀀스를 생성하는 시간 콘볼루션 층들의 시퀀스를 가지며, 서열분석 사이클들의 특정 윈도우 내의 상이한 서열분석 사이클들 사이에서 압축된 공간 출력 세트들을 조합하는 제1 시간 콘볼루션 층으로 시작하는 단계, 및 선행 시간 콘볼루션 층들의 연속 시간 출력들을 조합하는 연속 시간 콘볼루션 층들로 계속하는 단계를 포함하는, 인공 지능 기반 방법.

[0281] 12. 항목 11에 있어서,

[0282] 서열분석 사이클들의 제1 윈도우 동안,

[0283] 시간 네트워크의 시간 콘볼루션 층들의 시퀀스 내의 제1 시간 콘볼루션 층을 통해, 서열분석 사이클들의 제1 윈도우 내의 연속 서열분석 사이클들의 윈도우들 동안 각자의 압축된 공간 출력 세트들을 프로세싱하고 서열분석 사이클들의 제1 윈도우 동안 복수의 시간 출력 세트들을 생성하는 단계;

[0284] 압축 네트워크를 통해, 복수의 시간 출력 세트들을 프로세싱하고 복수의 시간 출력 세트들 내의 각자의 시간 출력 세트들에 대한 각자의 압축된 시간 출력 세트들을 생성하는 단계 - 각자의 시간 출력 세트들은 M개의 채널들(특징부 맵들)을 갖고, 각자의 압축된 시간 출력 세트들은 N개의 채널들(특징부 맵들)을 갖고, M > N임 -;

[0285] 시간 네트워크의 시간 콘볼루션 층들의 시퀀스 내의 최종 시간 콘볼루션 층을 통해, 각자의 압축된 시간 출력 세트들을 프로세싱하고 서열분석 사이클들의 제1 윈도우 동안 최종 시간 출력 세트들을 생성하는 단계; 및

[0286] 최종 시간 출력 세트에 기초하여, 서열분석 사이클들의 제1 윈도우 내의 하나 이상의 서열분석 사이클들 동안 염기 호출 예측들을 생성하는 단계를 추가로 포함하고,

[0287] 출력 층은 최종 시간 출력 세트를 프로세싱하고 서열분석 사이클들의 제1 윈도우 동안 최종 출력을 생성하고, 염기 호출 예측들은 최종 출력에 기초하여 생성되는, 인공 지능 기반 방법.

[0288] 13. 항목 12에 있어서,

[0289] 제1 시간 콘볼루션 층이 시간 출력 세트들을 이전에 생성한 연속 서열분석 사이클들의 하나 이상의 중첩 윈도우들, 및 제1 시간 콘볼루션 층이 시간 출력 세트를 아직 생성하지 않은 연속 서열분석 사이클들의 적어도 하나의 비중첩 윈도우를, 서열분석 사이클들의 제1 윈도우와 공유하는 서열분석 사이클들의 제2 윈도우 동안,

[0290] 제1 시간 콘볼루션 층을 통해, 연속 서열분석 사이클들의 비중첩 윈도우 내의 각자의 서열분석 사이클들 동안만 각자의 압축된 공간 출력 세트들을 프로세싱하고 연속 서열분석 사이클들의 비중첩 윈도우 동안 시간 출력 세트를 생성하여, 이에 의해, 제1 시간 콘볼루션 층을 통해, 연속 서열분석 사이클들의 중첩 윈도우들 내의 각자의 서열분석 사이클들 동안 각자의 압축된 공간 출력 세트들을 재프로세싱하는 것을 우회시키는 단계;

[0291] 압축 네트워크를 통해, 시간 출력 세트를 프로세싱하고 연속 서열분석 사이클들의 비중첩 윈도우 동안 압축된 시간 출력 세트를 생성하는 단계 - 시간 출력 세트는 M개의 채널들(특징부 맵들)을 갖고, 압축된 시간 출력은 N개의 채널들(특징부 맵들)을 갖고, M > N임 -;

[0292] 최종 시간 콘볼루션 층을 통해, 서열분석 사이클들의 제1 윈도우 동안 이전에 생성된 연속 서열분석 사이클들의 중첩 윈도우들 동안의 각자의 압축된 시간 출력 세트들 및 압축된 시간 출력 세트를 프로세싱하고 서열분석 사이클들의 제2 윈도우 동안 최종 시간 출력 세트를 생성하여, 이에 의해, 연속 서열분석 사이클들의 중첩 윈도우들 동안 각자의 압축된 시간 출력 세트들로 연속 서열분석 사이클들의 중첩 윈도우들 동안 각자의 사이클별 분석물 채널 세트들을 대체하는 단계; 및

[0293] 최종 시간 출력 세트에 기초하여, 서열분석 사이클들의 제2 윈도우 내의 하나 이상의 서열분석 사이클들 동안 염기 호출 예측들을 생성하는 단계를 추가로 포함하고,

[0294] 출력 층은 최종 시간 출력 세트를 프로세싱하고 서열분석 사이클들의 제2 윈도우 동안 최종 출력을 생성하고,

염기 호출 예측들은 최종 출력에 기초하여 생성되는, 인공 지능 기반 방법.

[0295] 14. 항목 13에 있어서,

[0296] 제1 시간 콘볼루션 층이 시간 출력 세트들을 이전에 생성한 연속 서열분석 사이클들의 하나 이상의 중첩 윈도우들, 및 제1 시간 콘볼루션 층이 시간 출력 세트를 아직 생성하지 않은 연속 서열분석 사이클들의 적어도 하나의 비중첩 윈도우를, 서열분석 사이클들의 제1 및 제2 윈도우들과 공유하는 서열분석 사이클들의 제3 윈도우 동안,

[0297] 제1 시간 콘볼루션 층을 통해, 연속 서열분석 사이클들의 비중첩 윈도우 내의 각자의 서열분석 사이클들 동안만 각자의 압축된 공간 출력 세트들을 프로세싱하고 연속 서열분석 사이클들의 비중첩 윈도우 동안 시간 출력 세트를 생성하여, 이에 의해, 제1 시간 콘볼루션 층을 통해, 연속 서열분석 사이클들의 중첩 윈도우들 내의 각자의 서열분석 사이클들 동안 각자의 압축된 공간 출력 세트들을 재프로세싱하는 것을 우회시키는 단계;

[0298] 압축 네트워크를 통해, 시간 출력 세트를 프로세싱하고 연속 서열분석 사이클들의 비중첩 윈도우 동안 압축된 시간 출력 세트를 생성하는 단계 - 시간 출력 세트는 M개의 채널들(특징부 맵들)을 갖고, 압축된 시간 출력은 N개의 채널들(특징부 맵들)을 갖고, M > N임 -;

[0299] 최종 시간 콘볼루션 층을 통해, 서열분석 사이클들의 제1 및 제2 윈도우들 동안 이전에 생성된 연속 서열분석 사이클들의 중첩 윈도우들 동안의 각자의 압축된 시간 출력 세트들 및 압축된 시간 출력 세트를 프로세싱하고 서열분석 사이클들의 제3 윈도우 동안 최종 시간 출력 세트를 생성하여, 이에 의해, 연속 서열분석 사이클들의 중첩 윈도우들 동안 각자의 압축된 시간 출력 세트들로 연속 서열분석 사이클들의 중첩 윈도우들 동안 각자의 사이클별 분석물 채널 세트들을 대체하는 단계; 및

[0300] 최종 시간 출력 세트에 기초하여, 서열분석 사이클들의 제3 윈도우 내의 하나 이상의 서열분석 사이클들 동안 염기 호출 예측들을 생성하는 단계를 추가로 포함하고,

[0301] 출력 층은 최종 시간 출력 세트를 프로세싱하고 서열분석 사이클들의 제3 윈도우 동안 최종 출력을 생성하고, 염기 호출 예측들은 최종 출력에 기초하여 생성되는, 인공 지능 기반 방법.

[0302] 15. 항목 11에 있어서, 시간 네트워크의 시간 콘볼루션 층들의 시퀀스 내의 각자의 시간 콘볼루션 층들은 상이한 카운트들의 콘볼루션 필터들을 갖고, 제1 시간 콘볼루션 층은 M개의 콘볼루션 필터들을 갖고, M은 4 초과의 정수인, 인공 지능 기반 방법.

[0303] 16. 항목 11에 있어서, 시간 네트워크의 시간 콘볼루션 층들의 시퀀스 내의 각자의 시간 콘볼루션 층들은 동일한 카운트의 콘볼루션 필터들을 갖고, 동일한 카운트는 M이고, M은 4 초과의 정수인, 인공 지능 기반 방법.

[0304] 17. 항목 16에 있어서, 시간 네트워크 내의 콘볼루션 필터들은 1차원(1D) 콘볼루션들을 사용하는, 인공 지능 기반 방법.

[0305] 18. 항목 1에 있어서, 압축 네트워크는 1x1 콘볼루션들을 사용하여 압축된 공간 출력 세트 내의 다수의 압축된 공간 출력들을 제어하고, 압축 네트워크는 N개의 콘볼루션 필터들을 갖고, N은 4 이하의 정수인, 인공 지능 기반 방법.

[0306] 19. 항목 1에 있어서, 신뢰성없는 분석물들을 식별하는 데이터를 사용하여 신뢰성없는 분석물들에 대응하는 압축된 공간 출력 세트 내의 압축된 공간 출력들의 부분들을 제거하는 단계, 및 압축된 필터링된 공간 출력 세트를 생성하여, 압축된 공간 출력 세트로 대체하고 신뢰성없는 분석물들이 아닌 그러한 분석물들에 대해서만 염기 호출 예측들을 생성하는 단계를 추가로 포함하는, 인공 지능 기반 방법.

[0307] 20. 항목 19에 있어서, 시간 네트워크를 통해, 대응하는 압축된 공간 출력 세트들 대신 압축된, 필터링된 공간 출력 세트들을 프로세싱하는 단계를 추가로 포함하는, 인공 지능 기반 방법.

[0308] 21. 항목 20에 있어서, 압축된, 필터링된 공간 출력 세트들로부터 압축된 시간 출력 세트들을 생성하는 단계를 추가로 포함하는, 인공 지능 기반 방법.

[0309] 22. 항목 19에 있어서, 신뢰성없는 분석물들을 식별하는 데이터는 신뢰성없는 클러스터들의 세기들을 묘사하는 픽셀들을 식별하는, 인공 지능 기반 방법.

[0310] 23. 항목 19에 있어서, 신뢰성없는 분석물들을 식별하는 데이터는 어떠한 세기들도 묘사하지 않는 픽셀들을 식별하는, 인공 지능 기반 방법.

[0311] 24. 항목 20에 있어서, 압축된 공간 출력 세트들은 대응하는 압축된, 필터링된 공간 출력 세트들보다 4 내지 9

배만큼 많은 총 픽셀들을 갖는, 인공 지능 기반 방법.

[0312] 25. 항목 24에 있어서, 압축된, 필터링된 공간 출력 세트들은 시간 네트워크가 75% 더 적은 픽셀들에 대해 동작하게 하고, 이에 의해, 계산 동작들, 메모리 액세스, 및 시간 네트워크에 대한 메모리 점유를 75%만큼 감소시키는, 인공 지능 기반 방법.

[0313] 26. 항목 5에 있어서, 공간 네트워크를 통해, 재프로세싱을 우회시키는 단계는 계산 동작들, 메모리 액세스, 및 시간 네트워크에 대한 메모리 점유를 80%만큼 감소시키는, 인공 지능 기반 방법.

[0314] 27. 항목 14에 있어서, 시간 네트워크를 통해, 재프로세싱을 우회시키는 단계는 계산 동작들, 메모리 액세스, 및 시간 네트워크에 대한 메모리 점유를 감소시키는, 인공 지능 기반 방법.

[0315] 28. 항목 27에 있어서, 압축 네트워크에 의해 이용가능하게 되는 계산 리소스들을 공간 네트워크 및 시간 네트워크 내의 보충 콘볼루션 필터들의 추가에 재할당하는 단계를 추가로 포함하는, 인공 지능 기반 방법.

[0316] 29. 항목 27에 있어서, 압축 네트워크에 의해 이용가능하게 되는 계산 리소스들을, 특정 서열 사이클 동안 염기 호출 예측들을 생성하는 데 사용되는 사이클별 분석물 채널 세트들의 각각의 윈도우에서 보충 사이클별 분석물 채널 세트들의 추가에 재할당하는 단계를 추가로 포함하는, 인공 지능 기반 방법.

[0317] 30. 항목 27에 있어서, 압축 네트워크에 의해 이용가능하게 되는 계산 리소스들을 공간 네트워크 내의 보충 공간 콘볼루션 층들의 추가에 재할당하는 단계를 추가로 포함하는, 인공 지능 기반 방법.

[0318] 31. 항목 27에 있어서, 압축 네트워크에 의해 이용가능하게 되는 계산 리소스들을 시간 네트워크 내의 보충 시간 콘볼루션 층들의 추가에 재할당하는 단계를 추가로 포함하는, 인공 지능 기반 방법.

[0319] 32. 항목 1에 있어서, 서열분석 사이클들의 현재 윈도우 동안 생성된 하나 이상의 압축된 공간 출력 세트들과 함께 서열분석 사이클들의 하나 이상의 선행 윈도우들 동안 생성된 하나 이상의 압축된 공간 출력 세트들을 사용하여 서열분석 사이클들의 현재 윈도우 내의 하나 이상의 서열분석 사이클들 동안 염기 호출 예측들을 생성하는 단계를 추가로 포함하는, 인공 지능 기반 방법.

[0320] 33. 항목 1에 있어서, 서열분석 사이클들의 현재 윈도우 동안 생성된 하나 이상의 압축된 공간 출력 세트들과 함께 서열분석 사이클들의 하나 이상의 후행 윈도우들 동안 생성된 하나 이상의 압축된 공간 출력 세트들을 사용하여 서열분석 사이클들의 현재 윈도우 내의 하나 이상의 서열분석 사이클들 동안 염기 호출 예측들을 생성하는 단계를 추가로 포함하는, 인공 지능 기반 방법.

[0321] 34. 항목 1에 있어서, 서열분석 사이클들의 현재 윈도우 동안 생성된 하나 이상의 압축된 시간 출력 세트들과 함께 서열분석 사이클들의 하나 이상의 선행 윈도우들 동안 생성된 하나 이상의 압축된 시간 출력 세트들을 사용하여 서열분석 사이클들의 현재 윈도우 내의 하나 이상의 서열분석 사이클들 동안 염기 호출 예측들을 생성하는 단계를 추가로 포함하는, 인공 지능 기반 방법.

[0322] 35. 항목 1에 있어서, 서열분석 사이클들의 현재 윈도우 동안 생성된 하나 이상의 압축된 시간 출력 세트들과 함께 서열분석 사이클들의 하나 이상의 후행 윈도우들 동안 생성된 하나 이상의 압축된 시간 출력 세트들을 사용하여 서열분석 사이클들의 현재 윈도우 내의 하나 이상의 서열분석 사이클들 동안 염기 호출 예측들을 생성하는 단계를 추가로 포함하는, 인공 지능 기반 방법.

[0323] 36. 항목 1에 있어서, 사이클별 분석물 채널 세트는 서열분석 런 동안 서열분석된 분석물들에 대한 분석물 데이터를 인코딩하는, 인공 지능 기반 방법.

[0324] 37. 항목 36에 있어서, 분석물 데이터는 분석물들로부터 수집된 세기 방출물들을 식별하는 이미지 데이터인, 방법.

[0325] 38. 항목 37에 있어서, 이미지 데이터는 복수의 이미지 채널들(이미지들)을 갖는, 인공 지능 기반 방법.

[0326] 39. 항목 38에 있어서, 이미지 채널(이미지)은 (i) 특정 레이저에 의한 조명 및 (ii) 특정 광학 필터를 통한 이미징의 조합에 의해 생성되는, 인공 지능 기반 방법.

[0327] 40. 항목 36에 있어서, 분석물 데이터는 분석물 활동에 기초하여 검출되는 전류 및/또는 전압 데이터인, 인공 지능 기반 방법.

[0328] 41. 항목 36에 있어서, 분석물 데이터는 분석물 활동에 기초하여 검출되는 pH 스케일 데이터인, 인공 지능 기반

방법.

[0329] 42. 항목 1에 있어서, 일련의 사이클별 분석물 채널 세트 내의 각각의 사이클별 분석물 채널 세트 내의 채널들의 개수는 압축 네트워크 내의 콘볼루션 필터들의 개수, 및 이에 따라, 압축된 공간 출력 세트들 및 압축된 시간 출력 세트들 내의 채널들의 개수를 결정하는, 인공 지능 기반 방법.

[0330] 43. 항목 1에 있어서, 압축된 공간 출력 세트들, 압축된, 필터 공간 출력 세트들, 및 압축된 시간 출력 세트들은 양자화된 형태로 저장되는, 인공 지능 기반 방법.

[0331] 44. 시스템으로서,

[0332] 호스트 프로세서에 부착되고, 서열분석 런이 진행됨에 따라 서열분석 이미지들의 진행을 수신하도록 구성된 호스트 메모리;

[0333] 프로세싱 유닛들의 어레이를 갖는 구성가능한 프로세서 - 프로세싱 유닛들의 어레이 내의 프로세싱 유닛들은 신경 네트워크 기반 염기 호출자를 실행하여 염기 호출 예측들을 생성하도록 구성됨 -;

[0334] 호스트 메모리, 호스트 프로세서, 및 구성가능한 프로세서에 액세스하고, 호스트 메모리로부터 구성가능한 프로세서 상에서 서열분석 사이클들의 제1 윈도우 내의 서열분석 사이클들 동안 서열분석 이미지들을 로딩하도록 구성된 데이터 흐름 로직; 및

[0335] 프로세싱 유닛들이 서열분석 사이클들의 제1 윈도우 내의 서열분석 사이클들 동안 서열분석 이미지들에 대해 신경 네트워크 기반 염기 호출자의 공간 네트워크를 사이클 단위로 실행하고 서열분석 사이클들의 제1 윈도우 내의 서열분석 사이클들 각각 동안 공간 특징부 맵 세트들을 생성하게 하도록 구성된 런타임 로직을 포함하고,

[0336] 런타임 로직은 프로세싱 유닛들이 공간 특징부 맵 세트들에 대해 신경 네트워크 기반 염기 호출자의 압축 네트워크를 사이클 단위로 실행하고 압축된 공간 특징부 맵 세트들을 생성하게 하도록, 그리고 시간 네트워크 및 출력 네트워크를 통해 압축된 공간 특징부 맵 세트들을 프로세싱하여 서열분석 사이클들의 제1 윈도우 내의 하나 이상의 서열분석 사이클들 동안 염기 호출 예측들을 생성하게 하도록 구성되고,

[0337] 데이터 흐름 로직은 압축된 공간 특징부 맵 세트들을 호스트 메모리로 이동시키도록, 그리고 서열분석 이미지들을 압축된 공간 특징부 맵 세트들로 겹쳐쓰기하도록 구성되고,

[0338] 하나 이상의 중첩 서열분석 사이클들을 서열분석 사이클들의 제1 윈도우와 공유하고 적어도 하나의 비중첩 서열분석 사이클을 갖는 서열분석 사이클들의 제2 윈도우 동안, 데이터 흐름 로직은 호스트 메모리로부터 구성가능한 프로세서 상에, 중첩 서열분석 사이클들 동안에는 압축된 공간 특징부 맵 세트들을, 그리고 비중첩 서열분석 사이클 동안에는 서열분석 이미지들을 로딩하도록 구성되고,

[0339] 런타임 로직은 프로세싱 유닛들이 비중첩 서열분석 사이클 동안 서열분석 이미지들에 대해 공간 네트워크를 실행하고 비중첩 서열분석 사이클 동안 공간 특징부 맵 세트를 생성하게 하도록 구성되고,

[0340] 런타임 로직은 프로세싱 유닛들이 공간 특징부 맵 세트에 대해 압축 네트워크를 실행하고 비중첩 서열분석 사이클 동안 압축된 공간 특징부 맵 세트를 생성하게 하도록, 그리고 시간 네트워크 및 출력 네트워크를 통해 중첩 서열분석 사이클들 동안에는 압축된 공간 특징부 맵 세트들을 그리고 비중첩 서열분석 사이클 동안에는 압축된 공간 특징부 맵 세트를 프로세싱하여 서열분석 사이클들의 제2 윈도우 내의 하나 이상의 서열분석 사이클들 동안 염기 호출 예측들을 생성하게 하도록 구성되는, 시스템.

[0341] 45. 시스템으로서,

[0342] 호스트 프로세서에 부착되고, 서열분석 런이 진행됨에 따라 서열분석 이미지들의 진행을 수신하도록 구성된 호스트 메모리;

[0343] 프로세서 메모리에 부착된 프로세싱 유닛들의 어레이를 갖는 구성가능한 프로세서 - 프로세싱 유닛들의 어레이 내의 프로세싱 유닛들은 신경 네트워크 기반 염기 호출자를 실행하여 염기 호출 예측들을 생성하도록 구성됨 -;

[0344] 호스트 메모리, 호스트 프로세서, 구성가능한 프로세서, 및 프로세서 메모리에 액세스하고, 호스트 메모리로부터 구성가능한 프로세서 상에서 서열분석 사이클들의 제1 윈도우 내의 서열분석 사이클들 동안 서열분석 이미지들을 로딩하도록 구성된 데이터 흐름 로직; 및

[0345] 프로세싱 유닛들이 서열분석 사이클들의 제1 윈도우 내의 서열분석 사이클들 동안 서열분석 이미지들에 대해 신경 네트워크 기반 염기 호출자의 공간 네트워크를 사이클 단위로 실행하고 서열분석 사이클들의 제1 윈도우 내

의 서열분석 사이클들 각각 동안 공간 특징부 맵 세트들을 생성하게 하도록 구성된 런타임 로직을 포함하고,

[0346] 런타임 로직은 프로세싱 유닛들이 공간 특징부 맵 세트들에 대해 신경 네트워크 기반 염기 호출자의 압축 네트워크를 사이클 단위로 실행하고 압축된 공간 특징부 맵 세트들을 생성하게 하도록, 그리고 시간 네트워크 및 출력 네트워크를 통해 압축된 공간 특징부 맵 세트들을 프로세싱하여 서열분석 사이클들의 제1 윈도우 내의 하나 이상의 서열분석 사이클들 동안 염기 호출 예측들을 생성하게 하도록 구성되고,

[0347] 데이터 흐름 로직은 압축된 공간 특징부 맵 세트들을 프로세서 메모리로 이동시키도록 구성되고,

[0348] 하나 이상의 중첩 서열분석 사이클들을 서열분석 사이클들의 제1 윈도우와 공유하고 적어도 하나의 비중첩 서열분석 사이클을 갖는 서열분석 사이클들의 제2 윈도우 동안, 데이터 흐름 로직은 프로세서 메모리로부터 구성가능한 프로세서 상에, 중첩 서열분석 사이클들 동안에는 압축된 공간 특징부 맵 세트들을 로딩하도록, 그리고 비중첩 서열분석 사이클 동안에는 호스트 메모리로부터 서열분석 이미지들을 로딩하도록 구성되고,

[0349] 런타임 로직은 프로세싱 유닛들이 비중첩 서열분석 사이클 동안 서열분석 이미지들에 대해 공간 네트워크를 실행하고 비중첩 서열분석 사이클 동안 공간 특징부 맵 세트를 생성하게 하도록 구성되고,

[0350] 런타임 로직은 프로세싱 유닛들이 공간 특징부 맵 세트에 대해 압축 네트워크를 실행하고 비중첩 서열분석 사이클 동안 압축된 공간 특징부 맵 세트를 생성하게 하도록, 그리고 시간 네트워크 및 출력 네트워크를 통해 중첩 서열분석 사이클들 동안에는 압축된 공간 특징부 맵 세트들을 그리고 비중첩 서열분석 사이클 동안에는 압축된 공간 특징부 맵 세트를 프로세싱하여 서열분석 사이클들의 제2 윈도우 내의 하나 이상의 서열분석 사이클들 동안 염기 호출 예측들을 생성하게 하도록 구성되는, 시스템.

[0351] 46. 시스템으로서,

[0352] 신경 네트워크 그래프의 제1 순회를 실행하여 제1 프로세싱 로직을 통해 입력들의 제1 세트 내의 각자의 입력들을 독립적으로 프로세싱하고 입력들의 제1 세트 내의 각자의 입력들 사이의 정보를 혼합하지 않으면서 입력들의 제1 세트 내의 각자의 입력들의 각자의 대안적인 표현들을 생성하도록, 그리고 입력들의 제1 세트 내의 각자의 입력들의 각자의 대안적인 표현들에 기초하여 제1 순회에 대한 출력들을 생성하도록 구성된 신경 네트워크 로직

[0353] - 신경 네트워크 로직은 신경 네트워크 그래프의 제2 순회를 실행하여 제1 프로세싱 로직을 통해 입력들의 제2 세트 내의 각자의 입력들을 독립적으로 프로세싱하고 입력들의 제2 세트 내의 각자의 입력들 사이의 정보를 혼합하지 않으면서 입력들의 제2 세트 내의 각자의 입력들의 각자의 대안적인 표현들을 생성하도록, 그리고 입력들의 제2 세트 내의 각자의 입력들의 각자의 대안적인 표현들에 기초하여 제2 순회에 대한 출력들을 생성하도록 구성되고, 입력들의 제1 및 제2 세트는 하나 이상의 중첩 입력들 및 적어도 하나의 비중첩 입력을 가짐 -; 및

[0354] 제1 순회를 실행하여 입력들의 제1 세트 내의 각자의 입력들의 각자의 대안적인 표현들을 생성하도록, 입력들의 제1 세트 내의 각자의 입력들의 각자의 대안적인 표현들을 압축된 형태로 메모리에 저장하도록, 그리고 입력들의 제1 세트 내의 각자의 입력들의 각자의 대안적인 표현들의 압축된 형태에 기초하여 제1 순회에 대한 출력들을 생성하도록 신경 네트워크 로직으로 구성된 런타임 로직

[0355] - 런타임 로직은 제2 순회를 실행하여 제1 프로세싱 로직을 통해 비중첩 입력만을 프로세싱하고 비중첩 입력의 대안적인 표현을 생성하도록, 비중첩 입력의 대안적인 표현을 압축된 형태로 메모리에 저장하도록, 제1 순회에서 생성된 중첩 입력들의 각자의 대안적인 표현들의 압축된 형태를 검색하여 제2 순회에서 중첩 입력들의 각자의 대안적인 표현들의 리던던트 생성을 우회시키는 것을 보상하도록, 그리고 중첩 입력들의 각자의 대안적인 표현들의 압축된 형태 및 비중첩 입력의 대안적인 표현의 압축된 형태에 기초하여 제2 순회에 대한 출력들을 생성하도록 구성됨 - 을 포함하는, 시스템.

[0356] 47. 항목 46에 있어서, 메모리는 온-칩 메모리인, 시스템.

[0357] 48. 항목 46에 있어서, 메모리는 오프-칩 메모리인, 시스템.

[0358] 49. 항목 46에 있어서, 압축된 형태의 채널들의 개수는 입력들의 제1 및 제2 세트 내의 입력들 내의 채널들의 개수에 대응하는, 시스템.

[0359] 50. 염기 호출의 인공 지능 기반 방법으로서,

[0360] 서열분석 런의 서열분석 사이클들 동안 생성된 일련의 사이클별 분석물 채널 세트들에 액세스하는 단계 - 대상 사이클별 분석물 채널 세트는 서열분석 런의 대상 서열분석 사이클에서 분석물들에 대해 검출된 분석물 데이터

를 인코딩함 -;

[0361] 신경 네트워크의 제1 프로세싱 모듈을 통해 대상 사이클별 분석물 채널 세트를 프로세싱하고 M개의 특징부 맵들로 대상 사이클별 분석물 채널 세트의 중간 표현을 생성하는 단계;

[0362] 신경 네트워크의 제2 프로세싱 모듈을 통해 중간 표현을 프로세싱하고 N개의 특징부 맵들로 대상 사이클별 분석물 채널 세트의 감소된 중간 표현을 생성하는 단계 - M > N임 -; 및

[0363] 대상 사이클별 분석물 채널 세트의 감소된 중간 표현을 사용하여 서열분석 런의 대상 서열분석 사이클에서 그리고/또는 다른 서열분석 사이클들에서 분석물들에 대한 염기 호출 예측들을 생성하는 단계를 포함하는, 인공 지능 기반 방법.

[0364] 51. 항목 50에 있어서, 제1 프로세싱 모듈은 M개의 콘볼루션 필터들을 갖는 콘볼루션 층인, 인공 지능 기반 방법.

[0365] 52. 항목 50에 있어서, 제2 프로세싱 모듈은 N개의 콘볼루션 필터들을 갖는 콘볼루션 층인, 인공 지능 기반 방법.

[0366] 53. 염기 호출의 인공 지능 기반 방법으로서,

[0367] 연속 슬라이딩 윈도우들이 중첩 서열분석 사이클들을 갖도록 신경 네트워크 기반 염기 호출자를 통해 서열분석 런의 서열분석 사이클들 동안 생성된 사이클별 분석물 채널 세트들의 진행을 슬라이딩 윈도우 단위로 프로세싱하는 단계로서,

[0368] 하나 이상의 선행 서열분석 사이클들, 중심 서열분석 사이클, 및 하나 이상의 후행 서열분석 사이클들을 포함하는 서열분석 사이클들의 현재 윈도우 동안,

[0369] 사이클별 분석물 채널 세트들의 현재 윈도우 상에 신경 네트워크 기반 염기 호출자를 적용한 것에 기초하여 선행 염기서열 사이클들, 중심 서열분석 사이클, 및 후행 염기서열 사이클들 각각 동안 공간 중간 표현 및 압축된 중간 표현을 생성하는 단계 - 공간 중간 표현은 M개의 채널들을 갖고, 압축된 중간 표현은 N개의 채널들을 갖고, M > N임 -; 및

[0370] 선행 서열분석 사이클들, 중심 서열분석 사이클, 및 후행 서열분석 사이클들 동안 생성된 압축된 중간 표현들에 기초하여 적어도 중심 서열분석 사이클을 염기 호출하는 단계를 포함하는, 프로세싱하는 단계; 및

[0371] 선행 서열분석 사이클들, 중심 서열분석 사이클, 및 후행 서열분석 사이클들 동안 생성된 압축된 중간 표현들을 사용하여 서열분석 사이클들의 다음 윈도우 내의 적어도 중심 서열분석 사이클에서 염기 호출하는 단계를 포함하는, 인공 지능 기반 방법.

[0372] 54. 염기 호출하기 위한 인공 지능 기반 시스템으로서,

[0373] 호스트 프로세서;

[0374] 서열분석 런의 서열분석 사이클들 동안 분석물 데이터를 저장하는, 호스트 프로세서에 의해 액세스가능한 메모리; 및

[0375] 메모리에 액세스하는 구성가능한 프로세서를 포함하고, 구성가능한 프로세서는,

[0376] 복수의 실행 클러스터들 - 복수의 실행 클러스터들 중의 실행 클러스터들은 신경 네트워크를 실행하도록 구성됨 -; 및

[0377] 메모리 및 복수의 실행 클러스터들 중의 실행 클러스터들에 액세스하고, 분석물 데이터를 복수의 실행 클러스터들 중의 이용가능한 실행 클러스터들에 제공하도록, 실행 클러스터들이 분석물 데이터를 신경 네트워크에 적용하여 현재 염기 호출 단계에서 사용하기 위해 분석물 데이터의 중간 표현들 및 압축된 중간 표현들을 생성하게 하도록, 그리고 분석물 데이터의 대체물로서 향후 염기 호출 단계에서 사용하기 위해 압축된 중간 표현들을 메모리로 피드백하도록 구성된 데이터 흐름 로직 - 중간 표현은 M개의 채널들을 갖고, 압축된 중간 표현은 N개의 채널들을 갖고, M > N임 - 을 포함하는, 인공 지능 기반 시스템.

[0378] 55. 시스템으로서,

[0379] 염기 호출자의 제1 반복을 실행하여 입력을 프로세싱하고 입력의 중간 표현들을 생성하도록 구성된 런타임 로직; 및

[0380] 중간 표현들을 프로세싱하고 입력의 압축된 중간 표현들을 생성하도록 구성된 압축 로직을 포함하고,

[0381] 런타임 로직은 염기 호출자의 후속 반복에서 입력 대신에 압축된 중간 표현들을 사용하도록 구성된, 시스템.

[0382] 56. 시스템으로서,

[0383] 염기 호출자의 제1 반복을 실행하여 입력을 프로세싱하고 입력의 중간 표현들을 생성하도록 구성된 런타임 로직; 및

[0384] 중간 표현들을 프로세싱하고 압축된 중간 표현들을 생성하도록 구성된 압축 로직 - 압축된 중간 표현들은 입력만큼 많은 채널들을 갖도록 구성됨 - 을 포함하고,

[0385] 런타임 로직은 염기 호출자의 후속 반복에서 입력 대신에 압축된 중간 표현들을 사용하도록 구성된, 시스템.

[0386] 57. 항목 56에 있어서, 채널들은 특징부 맵들에 대응하는, 시스템.

[0387] 58. 항목 56에 있어서, 채널들은 깊이 디멘션에 대응하는, 시스템.

[0388] 59. 항목 56에 있어서, 채널들은 공간 디멘션에 대응하는, 시스템.

[0389] *분할*

[0390] 1. 시스템으로서,

[0391] 공간 콘볼루션 층들의 각자의 시퀀스들을 통해 사이클별 서열분석 이미지 세트들의 윈도우 내의 각자의 사이클별 서열분석 이미지 세트들을 개별적으로 콘볼루션함으로써 서열분석 런의 일련의 서열분석 사이클들 동안 사이클별 서열분석 이미지 세트들의 윈도우를 사이클 단위로 프로세싱하여 일련의 서열분석 사이클들 중의 각자의 서열분석 사이클들 동안 각자의 사이클별 공간 특징부 맵 세트들을 생성하도록 구성된 공간 콘볼루션 네트워크

[0392] - 공간 콘볼루션 층들의 각자의 시퀀스들은 공간 콘볼루션 필터 뱅크들의 각자의 시퀀스들을 갖고, 공간 콘볼루션 필터 뱅크들의 각자의 시퀀스들의 공간 콘볼루션 필터 뱅크들 내의 공간 콘볼루션 필터들의 트레이닝된 계수들은 공간 콘볼루션 층들의 각자의 시퀀스들 내의 공간 콘볼루션 층들의 시퀀스들 사이에서 변함 -; 및

[0393] 제1 시간 콘볼루션 층의 각자의 시간 콘볼루션 필터 뱅크들을 사용하여 사이클별 공간 특징부 맵 세트들 내의 사이클별 공간 특징부 맵 세트들의 각자의 중첩 그룹들을 콘볼루션함으로써 그룹별 단위로 사이클별 공간 특징부 맵 세트들을 프로세싱하여 사이클별 공간 특징부 맵 세트들의 각자의 중첩 그룹들에 대한 각자의 그룹별 시간 특징부 맵 세트들을 생성하도록 구성된 시간 콘볼루션 네트워크

[0394] - 각자의 시간 콘볼루션 필터 뱅크들 내의 시간 콘볼루션 필터들의 트레이닝된 계수들은 각자의 시간 콘볼루션 필터 뱅크들 중의 시간 콘볼루션 필터 뱅크들 사이에서 변함 - 를 포함하는, 시스템.

[0395] 2. 항목 1에 있어서, 공간 콘볼루션 필터들은 사이클내 분리형 콘볼루션들을 사용하는, 시스템.

[0396] 3. 항목 1에 있어서, 시간 콘볼루션 필터들은 사이클간 조합형 콘볼루션들을 사용하는, 시스템.

[0397] 4. 항목 1에 있어서, 각자의 압축 콘볼루션 층들을 통해 각자의 사이클별 공간 특징부 맵 세트들을 개별적으로 콘볼루션하여 각자의 서열분석 사이클들 동안 각자의 사이클별 압축된 공간 특징부 맵 세트들을 생성하는 압축 네트워크를 포함하도록 추가로 구성되는, 시스템.

[0398] 5. 항목 4에 있어서, 각자의 압축 콘볼루션 층들 내의 압축 콘볼루션 필터들의 트레이닝된 계수들은 각자의 압축 콘볼루션 층들 중의 압축 콘볼루션 층들 사이에서 변하는, 시스템.

[0399] 6. 항목 5에 있어서, 시간 콘볼루션 네트워크는 제2 시간 콘볼루션 층의 각자의 시간 콘볼루션 필터 뱅크들을 사용하여 그룹별 시간 특징부 맵 세트들 중의 그룹별 시간 특징부 맵 세트들의 각자의 중첩 그룹들을 콘볼루션함으로써 그룹별 시간 특징부 맵 세트들을 그룹별 단위로 프로세싱하여 그룹별 시간 특징부 맵 세트들의 각자의 중첩 그룹들에 대한 각자의 추가 그룹별 시간 특징부 맵 세트들을 생성하도록 추가로 구성되는, 시스템.

[0400] 7. 항목 6에 있어서, 최종 시간 콘볼루션 층에 의해 생성된 최종 시간 특징부 맵 세트를 프로세싱하여 최종 출력을 생성하는 출력 네트워크를 포함하도록 추가로 구성되는, 시스템.

[0401] 8. 항목 7에 있어서, 최종 출력에 기초하여 일련의 서열분석 사이클들 중의 하나 이상의 서열분석 사이클들 동안 염기 호출 예측들을 생성하도록 추가로 구성되는, 시스템.

[0402] 9. 시스템으로서,

[0403] 공간 콘볼루션 층들의 각자의 시퀀스들을 통해 사이클별 서열분석 이미지 세트들의 윈도우 내의 각자의 사이클별 서열분석 이미지 세트들을 개별적으로 콘볼루션함으로써 서열분석 런의 일련의 서열분석 사이클들 동안 사이클별 서열분석 이미지 세트들의 윈도우를 사이클 단위로 프로세싱하여 일련의 서열분석 사이클들 중의 각자의 서열분석 사이클들 동안 각자의 사이클별 공간 특징부 맵 세트들을 생성하도록 구성된 공간 콘볼루션 네트워크; 및

[0404] 각자의 시간 콘볼루션 필터 뱅크들을 사용하여 사이클별 공간 특징부 맵 세트들 내의 사이클별 공간 특징부 맵 세트들의 각자의 중첩 그룹들을 콘볼루션함으로써 그룹별 단위로 사이클별 공간 특징부 맵 세트들을 프로세싱하여 사이클별 공간 특징부 맵 세트들의 각자의 중첩 그룹들에 대한 각자의 그룹별 시간 특징부 맵 세트들을 생성하도록 구성된 시간 콘볼루션 네트워크를 포함하고,

[0405] 각자의 시간 콘볼루션 필터 뱅크들 내의 시간 콘볼루션 필터들의 트레이닝된 계수들은 각자의 시간 콘볼루션 필터 뱅크들 중의 시간 콘볼루션 필터 뱅크들 사이에서 변하는, 시스템.

[0406] 10. 항목 9에 있어서, 공간 콘볼루션 층들의 각자의 시퀀스들은 공간 콘볼루션 필터 뱅크들의 각자의 시퀀스들을 갖고, 공간 콘볼루션 필터 뱅크들의 각자의 시퀀스들의 공간 콘볼루션 필터 뱅크들 내의 공간 콘볼루션 필터들의 트레이닝된 계수들은 공간 콘볼루션 층들의 각자의 시퀀스들 내의 공간 콘볼루션 층들의 시퀀스들 사이에서 공유되는, 시스템.

[0407] 11. 항목 9에 있어서, 각자의 압축 콘볼루션 층들을 통해 각자의 사이클별 공간 특징부 맵 세트들을 개별적으로 콘볼루션하여 각자의 서열분석 사이클들 동안 각자의 사이클별 압축된 공간 특징부 맵 세트들을 생성하는 압축 네트워크를 포함하도록 추가로 구성되고, 각자의 압축 콘볼루션 층들 내의 압축 콘볼루션 필터들의 트레이닝된 계수들은 각자의 압축 콘볼루션 층들 중의 압축 콘볼루션 층들 사이에서 변하는, 시스템.

[0408] 12. 염기 호출의 인공 지능 기반 방법으로서,

[0409] 공간 콘볼루션 네트워크를 통해, 공간 콘볼루션 층들의 각자의 시퀀스들을 통해 사이클별 서열분석 이미지 세트들의 윈도우 내의 각자의 사이클별 서열분석 이미지 세트들을 개별적으로 콘볼루션함으로써 서열분석 런의 일련의 서열분석 사이클들 동안 사이클별 서열분석 이미지 세트들의 윈도우를 사이클 단위로 프로세싱하고 일련의 서열분석 사이클들 중의 각자의 서열분석 사이클들 동안 각자의 사이클별 공간 특징부 맵 세트들을 생성하는 단계

[0410] - 공간 콘볼루션 층들의 각자의 시퀀스들은 공간 콘볼루션 필터 뱅크들의 각자의 시퀀스들을 갖고, 공간 콘볼루션 필터 뱅크들의 각자의 시퀀스들의 공간 콘볼루션 필터 뱅크들 내의 공간 콘볼루션 필터들의 트레이닝된 계수들은 공간 콘볼루션 층들의 각자의 시퀀스들 내의 공간 콘볼루션 층들의 시퀀스들 사이에서 변함 -; 및

[0411] 시간 콘볼루션 네트워크를 통해, 제1 시간 콘볼루션 층의 각자의 시간 콘볼루션 필터 뱅크들을 사용하여 사이클별 공간 특징부 맵 세트들 내의 사이클별 공간 특징부 맵 세트들의 각자의 중첩 그룹들을 콘볼루션함으로써 그룹별 단위로 사이클별 공간 특징부 맵 세트들을 프로세싱하고 사이클별 공간 특징부 맵 세트들의 각자의 중첩 그룹들에 대한 각자의 그룹별 시간 특징부 맵 세트들을 생성하는 단계

[0412] - 각자의 시간 콘볼루션 필터 뱅크들 내의 시간 콘볼루션 필터들의 트레이닝된 계수들은 각자의 시간 콘볼루션 필터 뱅크들 중의 시간 콘볼루션 필터 뱅크들 사이에서 변함 - 를 포함하는, 인공 지능 기반 방법.

[0413] 13. 항목 12에 있어서, 압축 네트워크의 각자의 압축 콘볼루션 층들을 통해 각자의 사이클별 공간 특징부 맵 세트들을 개별적으로 콘볼루션하고 각자의 서열분석 사이클들 동안 각자의 사이클별 압축된 공간 특징부 맵 세트들을 생성하는 단계를 추가로 포함하는, 인공 지능 기반 방법.

[0414] 14. 항목 13에 있어서, 각자의 압축 콘볼루션 층들 내의 압축 콘볼루션 필터들의 트레이닝된 계수들은 각자의 압축 콘볼루션 층들 중의 압축 콘볼루션 층들 사이에서 변하는, 인공 지능 기반 방법.

[0415] 15. 항목 14에 있어서, 시간 콘볼루션 네트워크를 통해, 제2 시간 콘볼루션 층의 각자의 시간 콘볼루션 필터 뱅크들을 사용하여 그룹별 시간 특징부 맵 세트들 중의 그룹별 시간 특징부 맵 세트들의 각자의 중첩 그룹들을 콘볼루션함으로써 그룹별 시간 특징부 맵 세트들을 그룹별 단위로 프로세싱하고 그룹별 시간 특징부 맵 세트들의 각자의 중첩 그룹들에 대한 각자의 추가 그룹별 시간 특징부 맵 세트들을 생성하는 단계를 추가로 포함하는, 인공 지능 기반 방법.

[0416] 16. 항목 15에 있어서, 출력 네트워크를 통해, 최종 시간 콘볼루션 층에 의해 생성된 최종 시간 특징부 맵 세트를 프로세싱하고 최종 출력을 생성하는 단계를 추가로 포함하는, 인공 지능 기반 방법.

[0417] 17. 항목 16에 있어서, 최종 출력에 기초하여 일련의 서열분석 사이클들 중의 하나 이상의 서열분석 사이클들 동안 염기 호출 예측들을 생성하는 단계를 추가로 포함하는, 인공 지능 기반 방법.

[0418] 18. 염기 호출의 인공 지능 기반 방법으로서,

[0419] 공간 콘볼루션 네트워크를 통해, 공간 콘볼루션 층들의 각자의 시퀀스들을 통해 사이클별 서열분석 이미지 세트들의 윈도우 내의 각자의 사이클별 서열분석 이미지 세트들을 개별적으로 콘볼루션함으로써 서열분석 런의 일련의 서열분석 사이클들 동안 사이클별 서열분석 이미지 세트들의 윈도우를 사이클 단위로 프로세싱하고 일련의 서열분석 사이클들 중의 각자의 서열분석 사이클들 동안 각자의 사이클별 공간 특징부 맵 세트들을 생성하는 단계; 및

[0420] 시간 콘볼루션 네트워크를 통해, 제1 시간 콘볼루션 층의 각자의 시간 콘볼루션 필터 뱅크들을 사용하여 사이클별 공간 특징부 맵 세트들 내의 사이클별 공간 특징부 맵 세트들의 각자의 중첩 그룹들을 콘볼루션함으로써 그룹별 단위로 사이클별 공간 특징부 맵 세트들을 프로세싱하고 사이클별 공간 특징부 맵 세트들의 각자의 중첩 그룹들에 대한 각자의 그룹별 시간 특징부 맵 세트들을 생성하는 단계를 포함하고,

[0421] 각자의 시간 콘볼루션 필터 뱅크들 내의 시간 콘볼루션 필터들의 트레이닝된 계수들은 각자의 시간 콘볼루션 필터 뱅크들 중의 시간 콘볼루션 필터 뱅크들 사이에서 변하는, 인공 지능 기반 방법.

[0422] 19. 항목 18에 있어서, 공간 콘볼루션 층들의 각자의 시퀀스들은 공간 콘볼루션 필터 뱅크들의 각자의 시퀀스들을 갖고, 공간 콘볼루션 필터 뱅크들의 각자의 시퀀스들의 공간 콘볼루션 필터 뱅크들 내의 공간 콘볼루션 필터들의 트레이닝된 계수들은 공간 콘볼루션 층들의 각자의 시퀀스들 내의 공간 콘볼루션 층들의 시퀀스들 사이에서 공유되는, 인공 지능 기반 방법.

[0423] 20. 항목 18에 있어서, 압축 네트워크의 각자의 압축 콘볼루션 층들을 통해 각자의 사이클별 공간 특징부 맵 세트들을 개별적으로 콘볼루션하고 각자의 서열분석 사이클들 동안 각자의 사이클별 압축된 공간 특징부 맵 세트들을 생성하는 단계를 추가로 포함하고, 각자의 압축 콘볼루션 층들 내의 압축 콘볼루션 필터들의 트레이닝된 계수들은 각자의 압축 콘볼루션 층들 중의 압축 콘볼루션 층들 사이에서 변하는, 인공 지능 기반 방법.

[0424] 21. 시스템으로서,

[0425] 사이클별 서열분석 이미지들의 윈도우에서 공간 콘볼루션 층들의 각자의 시퀀스들을 각자의 사이클별 서열분석 이미지들에 적용하도록 구성된 공간 콘볼루션 네트워크를 포함하고,

[0426] 공간 콘볼루션 층들의 각자의 시퀀스들은 공간 콘볼루션 층들의 시퀀스마다 상이한 공간 콘볼루션 필터 뱅크들의 각자의 시퀀스들을 갖는, 시스템.

[0427] 22. 시스템으로서,

[0428] 시간 콘볼루션 필터들의 각자의 세트들을 공간 특징부 맵들의 각자의 슬라이딩 윈도우들에 적용하도록 구성된 제1 시간 콘볼루션 층으로 구성된 시간 콘볼루션 네트워크를 포함하고,

[0429] 제1 시간 콘볼루션 층 내의 시간 콘볼루션 필터들의 각자의 세트들은 시간 콘볼루션 필터들의 세트마다 상이한 시간 콘볼루션 필터들을 갖는, 시스템.

[0430] 23. 항목 22에 있어서, 시간 콘볼루션 네트워크는 제1 시간 콘볼루션 층에 후행하는 제2 시간 콘볼루션 층으로 구성되고, 제2 콘볼루션 층은 시간 콘볼루션 필터들의 각자의 세트들을 시간 특징부 맵들의 각자의 슬라이딩 윈도우들에 적용하도록 구성되고, 제2 시간 콘볼루션 층 내의 시간 콘볼루션 필터들의 각자의 세트들은 시간 콘볼루션 필터들의 세트마다 상이한 시간 콘볼루션 필터들을 갖는, 시스템.

[0431] *스킵*

[0432] 1. 시스템으로서,

[0433] 각자의 공간 프로세싱 파이프라인들을 통해 사이클별 서열분석 이미지 세트들의 윈도우 내의 각자의 사이클별 서열분석 이미지 세트들을 개별적으로 프로세싱함으로써 서열분석 런의 일련의 서열분석 사이클들 동안 사이클별 서열분석 이미지 세트들의 윈도우를 사이클 단위로 프로세싱하도록 구성된 공간 콘볼루션 네트워크 - 각자의

공간 프로세싱 파이프라인들은 공간 콘볼루션 층들의 각자의 시퀀스들을 통해 각자의 사이클별 서열분석 이미지 세트들을 콘볼루션하여 일련의 서열분석 사이클들 내의 각자의 서열분석 사이클들 동안 각자의 사이클별 공간 특징부 맵 세트들을 생성하도록 구성됨 -; 및

[0434] 공간 콘볼루션 네트워크에 접속되고, 공간 콘볼루션 층들의 각자의 시퀀스들 내의 공간 콘볼루션 층들 사이에 버스들을 형성하도록 구성되는 버스 네트워크 - 버스들은 특정 서열분석 사이클 동안 공간 콘볼루션 층의 특정 시퀀스 내의 2개 이상의 공간 콘볼루션 층들에 의해 생성된 각자의 사이클별 공간 특징부 맵 세트들이 조합된 사이클별 공간 특징부 맵 세트로 조합되게 하도록, 그리고 조합된 사이클별 공간 특징부 맵 세트를 공간 콘볼루션 층의 특정 시퀀스 내의 다른 공간 콘볼루션 층에 대한 입력으로서 제공하도록 구성됨 - 를 포함하는, 시스템.

[0435] 2. 항목 1에 있어서, 2개 이상의 공간 콘볼루션 층들은 제1 공간 콘볼루션 층 및 제3 공간 콘볼루션 층을 포함하고, 제1 공간 콘볼루션 층은 제1 사이클별 공간 특징부 맵 세트를 생성하고, 제1 공간 콘볼루션 층은 제1 사이클별 공간 특징부 맵 세트를 제2 공간 콘볼루션 층에 대한 입력으로서 제공하고, 제2 공간 콘볼루션 층은 제1 사이클별 공간 특징부 맵 세트를 프로세싱하고 제2 사이클별 공간 특징부 맵 세트를 생성하고, 제2 공간 콘볼루션 층은 제2 사이클별 공간 특징부 맵 세트를 제3 공간 콘볼루션 층에 대한 입력으로서 제공하고, 제3 공간 콘볼루션 층은 제2 사이클별 공간 특징부 맵 세트를 프로세싱하고 제3 사이클별 공간 특징부 맵 세트를 생성하는, 시스템.

[0436] 3. 항목 2에 있어서, 버스들은 제1 공간 특징부 맵 세트 및 제3 사이클별 공간 특징부 맵 세트가 조합된 사이클별 공간 특징부 맵 세트로 조합되게 하도록 추가로 구성되는, 시스템.

[0437] 4. 항목 3에 있어서, 다른 공간 콘볼루션 층은 공간 콘볼루션 층의 특정 시퀀스 내의 제3 공간 콘볼루션 층에 바로 후행하는 제4 공간 콘볼루션 층이고, 제4 공간 콘볼루션 층은 조합된 사이클별 공간 특징부 맵 세트를 입력으로서 프로세싱하는, 시스템.

[0438] 5. 항목 2에 있어서, 2개 이상의 공간 콘볼루션 층들은 제1 공간 콘볼루션 층 및 제7 공간 콘볼루션 층을 포함하고, 제3 공간 콘볼루션 층은 제3 사이클별 공간 특징부 맵 세트를 제4 공간 콘볼루션 층에 대한 입력으로서 제공하고, 제4 공간 콘볼루션 층은 제3 사이클별 공간 특징부 맵 세트를 프로세싱하고 제4 사이클별 공간 특징부 맵 세트를 생성하고, 제4 공간 콘볼루션 층은 제4 사이클별 공간 특징부 맵 세트를 제5 공간 콘볼루션 층에 대한 입력으로서 제공하고, 제5 공간 콘볼루션 층은 제4 사이클별 공간 특징부 맵 세트를 프로세싱하고 제5 사이클별 공간 특징부 맵 세트를 생성하는, 시스템.

[0439] 6. 항목 5에 있어서, 버스들은 제1 공간 특징부 맵 세트 및 제5 사이클별 공간 특징부 맵 세트가 조합된 사이클별 공간 특징부 맵 세트로 조합되게 하도록 추가로 구성되는, 시스템.

[0440] 7. 항목 6에 있어서, 다른 공간 콘볼루션 층은 공간 콘볼루션 층의 특정 시퀀스 내의 제5 공간 콘볼루션 층에 바로 후행하는 제6 공간 콘볼루션 층이고, 제6 공간 콘볼루션 층은 조합된 사이클별 공간 특징부 맵 세트를 입력으로서 프로세싱하는, 시스템.

[0441] 8. 항목 5에 있어서, 2개 이상의 공간 콘볼루션 층들은 제1 공간 콘볼루션 층, 제3 공간 콘볼루션 층, 및 제5 공간 콘볼루션 층을 포함하고, 버스들은 제1 사이클별 공간 특징부 맵 세트, 제3 사이클별 공간 특징부 맵 세트, 및 제5 사이클 공간 특징부 맵 세트가 조합된 사이클별 공간 특징부 맵 세트로 조합되게 하도록 추가로 구성되는, 시스템.

[0442] 9. 항목 8에 있어서, 다른 공간 콘볼루션 층은 조합된 사이클별 공간 특징부 맵 세트를 입력으로서 프로세싱하는 제6 공간 콘볼루션 층인, 시스템.

[0443] 10. 항목 1에 있어서, 버스들은 제1 공간 콘볼루션 층에 대한 입력으로서 제공되는 특정 서열분석 사이클 동안의 사이클별 서열분석 이미지 세트, 및 제3 사이클별 공간 특징부 맵 세트가 조합된 사이클별 공간 특징부 맵 세트로 조합되게 하도록 추가로 구성되는, 시스템.

[0444] 11. 항목 10에 있어서, 다른 공간 콘볼루션 층은 조합된 사이클별 공간 특징부 맵 세트를 입력으로서 프로세싱하는 제4 공간 콘볼루션 층인, 시스템.

[0445] 12. 항목 1에 있어서, 버스 네트워크는 수신하는 사이클별 공간 특징부 맵 세트와 조합되어 조합된 사이클별 공간 특징부 맵 세트를 생성하는 유입 사이클별 공간 특징부 맵 세트의 공간 및 깊이 디멘션수를 수정하도록 구성

된 디멘션수 호환성 로직을 포함하도록 추가로 구성되는, 시스템.

[0446] 13. 항목 12에 있어서, 디멘션수 호환성 로직은 콘볼루션, 풀링 또는 평균화를 포함한 디멘션수 감소 동작인, 시스템.

[0447] 14. 항목 12에 있어서, 버스 네트워크는 수신하는 사이클별 공간 특징부 맵 세트와 조합되어 조합된 사이클별 공간 특징부 맵 세트를 생성하는 유입 사이클별 공간 특징부 맵 세트의 특징부 값들을 스케일링하도록 구성된 스케일링 로직을 포함하도록 추가로 구성되는, 시스템.

[0448] 15. 항목 1에 있어서, 제1 시간 콘볼루션 층의 각자의 시간 콘볼루션 필터 뱅크들을 사용하여 사이클별 공간 특징부 맵 세트들 내의 사이클별 공간 특징부 맵 세트들의 각자의 중첩 그룹들을 콘볼루션함으로써 그룹별 단위로 사이클별 공간 특징부 맵 세트들을 프로세싱하여 사이클별 공간 특징부 맵 세트들의 각자의 중첩 그룹들에 대한 각자의 그룹별 시간 특징부 맵 세트들을 생성하도록 구성된 시간 콘볼루션 네트워크를 포함하도록 추가로 구성되는, 시스템.

[0449] 16. 항목 15에 있어서, 시간 콘볼루션 네트워크에 접속되고, 시간 콘볼루션 층들의 각자의 시퀀스들 내의 시간 콘볼루션 층들 사이에 버스들을 형성하도록 구성되는 버스 네트워크 - 버스들은 특정 서열분석 사이클 동안 시간 콘볼루션 층의 특정 시퀀스 내의 2개 이상의 시간 콘볼루션 층들에 의해 생성된 각자의 사이클별 시간 특징부 맵 세트들이 조합된 사이클별 시간 특징부 맵 세트로 조합되게 하도록, 그리고 조합된 사이클별 시간 특징부 맵 세트를 시간 콘볼루션 층의 특정 시퀀스 내의 다른 시간 콘볼루션 층에 대한 입력으로서 제공하도록 구성됨 - 를 포함하도록 추가로 구성되는, 시스템.

[0450] 17. 인공 지능 기반 방법으로서,

[0451] 공간 콘볼루션 네트워크를 통해, 각자의 공간 프로세싱 파이프라인들을 통해 사이클별 서열분석 이미지 세트들의 윈도우 내의 각자의 사이클별 서열분석 이미지 세트들을 개별적으로 프로세싱함으로써 서열분석 런의 일련의 서열분석 사이클들 동안 사이클별 서열분석 이미지 세트들의 윈도우를 사이클 단위로 프로세싱하는 단계 - 이러한 프로세싱하는 단계는 공간 콘볼루션 층들의 각자의 시퀀스들을 통해 각자의 사이클별 서열분석 이미지 세트들을 콘볼루션하여 일련의 서열분석 사이클들 내의 각자의 서열분석 사이클들 동안 각자의 사이클별 공간 특징부 맵 세트들을 생성하는 단계를 포함함 -; 및

[0452] 특정 서열분석 사이클에 대한 공간 콘볼루션 층의 특성 시퀀스 내의 2개 이상의 공간 콘볼루션 층들에 의해 생성된 각자의 사이클별 공간 특징부 맵 세트들을 조합된 사이클별 공간 특징부 맵 세트로 조합하고, 조합된 사이클별 공간 특징부 맵 세트를 공간 콘볼루션 층의 특정 시퀀스 내의 다른 공간 콘볼루션 층에 대한 입력으로서 제공하는 단계를 포함하는, 인공 지능 기반 방법.

[0453] 18. 항목 17에 있어서, 2개 이상의 공간 콘볼루션 층들은 제1 공간 콘볼루션 층 및 제3 공간 콘볼루션 층을 포함하고, 제1 공간 콘볼루션 층은 제1 사이클별 공간 특징부 맵 세트를 생성하고, 제1 공간 콘볼루션 층은 제1 사이클별 공간 특징부 맵 세트를 제2 공간 콘볼루션 층에 대한 입력으로서 제공하고, 제2 공간 콘볼루션 층은 제1 사이클별 공간 특징부 맵 세트를 프로세싱하고 제2 사이클별 공간 특징부 맵 세트를 생성하고, 제2 공간 콘볼루션 층은 제2 사이클별 공간 특징부 맵 세트를 제3 공간 콘볼루션 층에 대한 입력으로서 제공하고, 제3 공간 콘볼루션 층은 제2 사이클별 공간 특징부 맵 세트를 프로세싱하고 제3 사이클별 공간 특징부 맵 세트를 생성하는, 인공 지능 기반 방법.

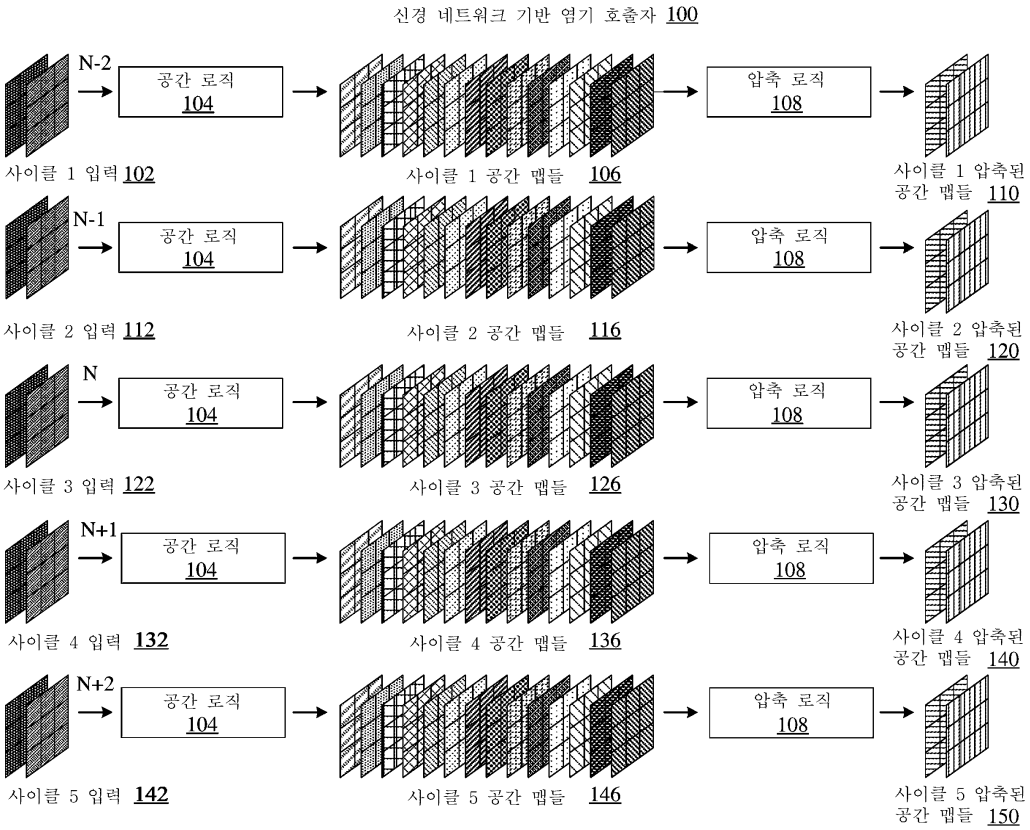
[0454] 19. 항목 18에 있어서, 버스들은 제1 공간 특징부 맵 세트 및 제3 사이클별 공간 특징부 맵 세트가 조합된 사이클별 공간 특징부 맵 세트로 조합되게 하도록 추가로 구성되는, 인공 지능 기반 방법.

[0455] 20. 항목 19에 있어서, 다른 공간 콘볼루션 층은 공간 콘볼루션 층의 특정 시퀀스 내의 제3 공간 콘볼루션 층에 바로 후행하는 제4 공간 콘볼루션 층이고, 제4 공간 콘볼루션 층은 조합된 사이클별 공간 특징부 맵 세트를 입력으로서 프로세싱하는, 인공 지능 기반 방법.

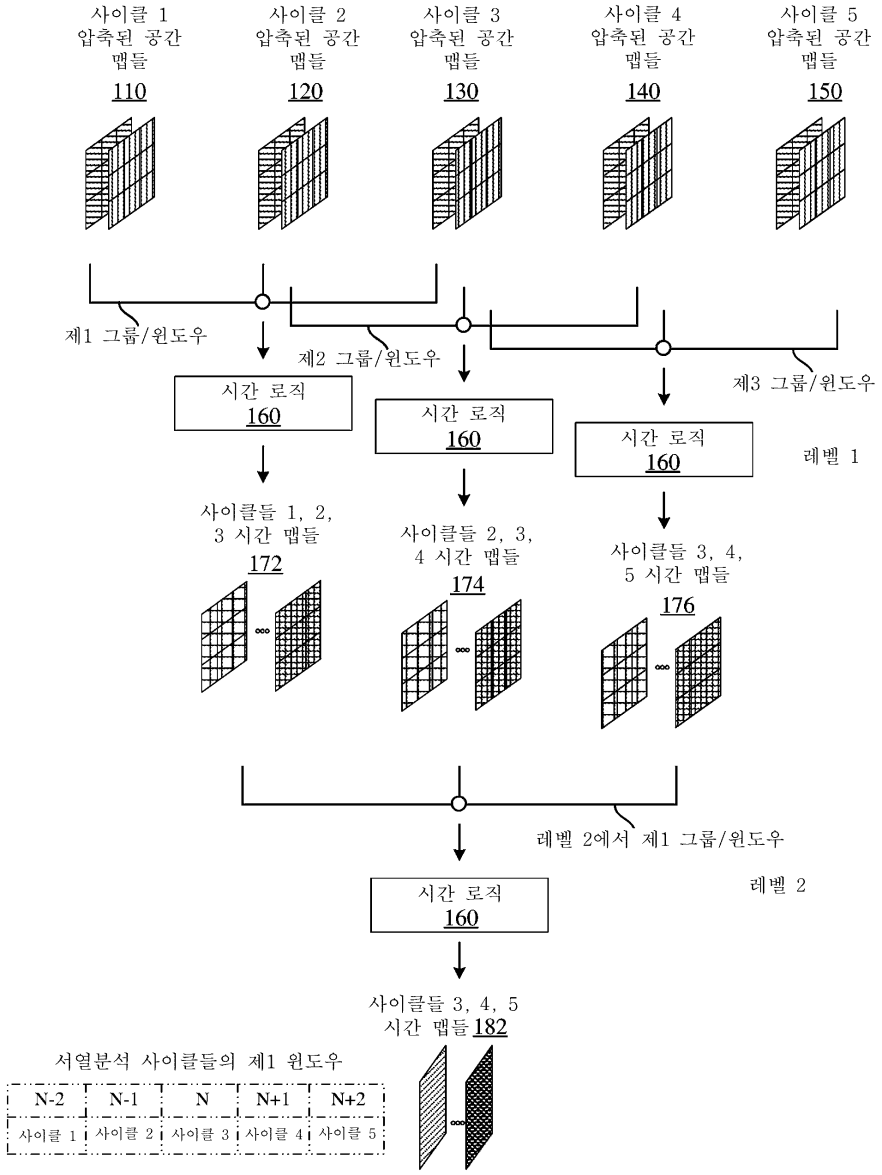
[0456] 상기에 기술된 방법의 다른 구현예들은 전술된 방법들 중 임의의 것을 수행하도록 프로세서에 의해 실행가능한 명령어들을 저장하는 비일시적 컴퓨터 판독가능 저장 매체를 포함할 수 있다. 이 섹션에 기술된 방법의 또 다른 구현예는 메모리, 및 메모리에 저장된 명령어들을 실행하여 전술된 방법들 중 임의의 것을 수행하도록 동작가능한 하나 이상의 프로세서들을 포함하는 시스템을 포함할 수 있다.

도면

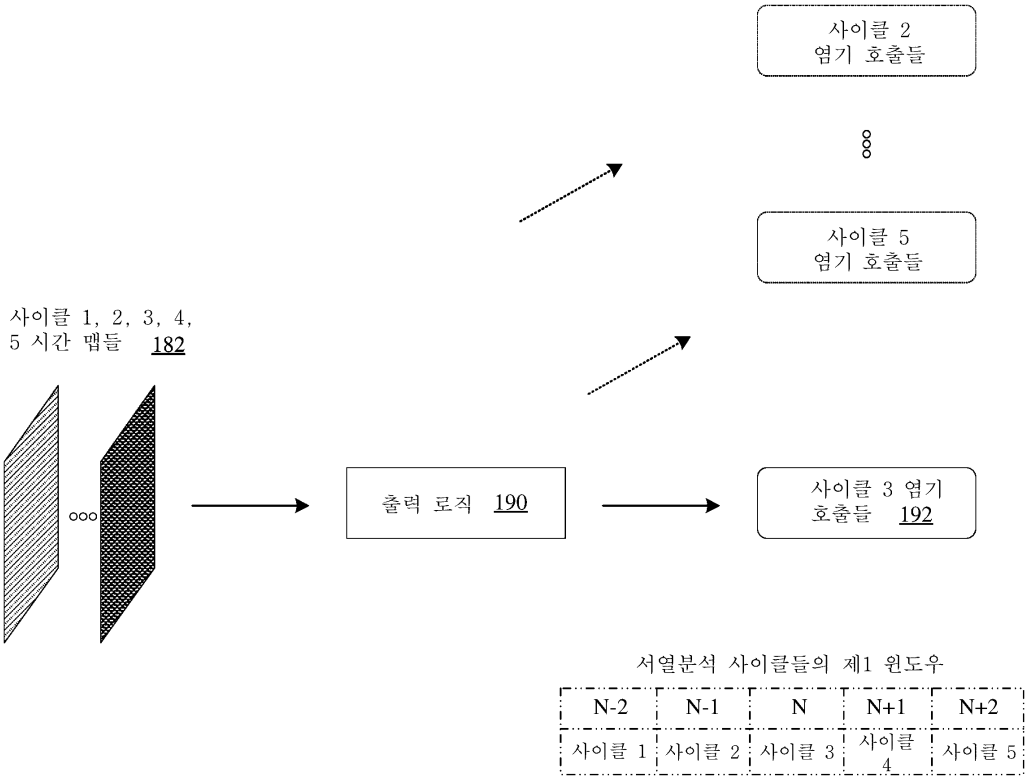
도면1a


신경 네트워크 기반 엄기 호출자 100
N-2
공간 로직 104
압축 로직 108
사이클 1 입력 102
사이클 1 공간 맵들 106
사이클 1 압축된 공간 맵들 110
N-1
공간 로직 104
압축 로직 108
사이클 2 입력 112
사이클 2 공간 맵들 116
사이클 2 압축된 공간 맵들 120
N
공간 로직 104
압축 로직 108
사이클 3 입력 122
사이클 3 공간 맵들 126
사이클 3 압축된 공간 맵들 130
N+1
공간 로직 104
압축 로직 108
사이클 4 입력 132
사이클 4 공간 맵들 136
사이클 4 압축된 공간 맵들 140
N+2
공간 로직 104
압축 로직 108
사이클 5 입력 142
사이클 5 공간 맵들 146
사이클 5 압축된 공간 맵들 150

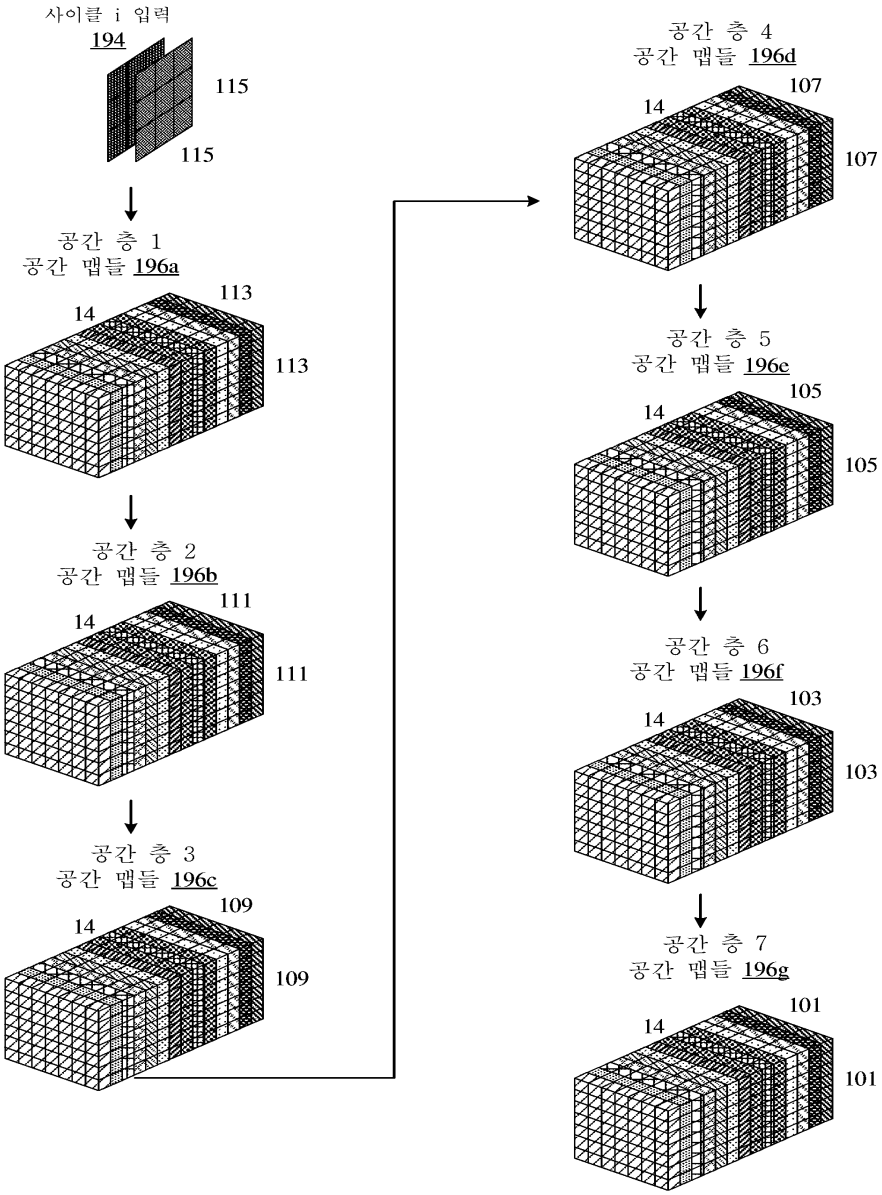
도면1b


사이클 1 압축된 공간 맵들
110
사이클 2 압축된 공간 맵들
120
사이클 3 압축된 공간 맵들
130
사이클 4 압축된 공간 맵들
140
사이클 5 압축된 공간 맵들
150
제1 그룹/윈도우
제2 그룹/윈도우
제3 그룹/윈도우
시간 로직
160
시간 로직
160
시간 로직
160
레벨 1
사이클들 1, 2, 3 시간 맵들
172
사이클들 2, 3, 4 시간 맵들
174
사이클들 3, 4, 5 시간 맵들
176
레벨 2에서 제1 그룹/윈도우
레벨 2
시간 로직
160
사이클들 3, 4, 5 시간 맵들 182
서열분석 사이클들의 제1 윈도우
N-2
N-1
N
N+1
N+2
사이클 1
사이클 2
사이클 3
사이클 4
사이클 5

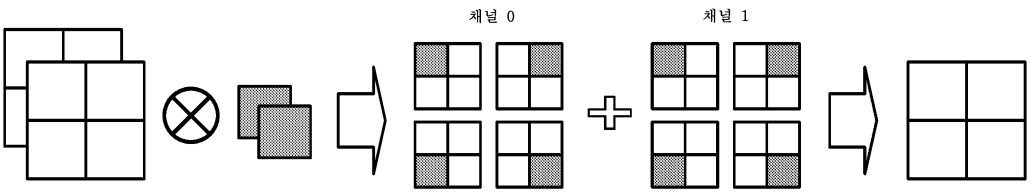
도면1c


사이클 2
염기 호출들
사이클 5
염기 호출들
사이클 1, 2, 3, 4,
5 시간 맵들 182
출력 로직 190
사이클 3 염기
호출들 192
서열분석 사이클들의 제1 윈도우
N-2
N-1
N
N+1
N+2
사이클 1
사이클 2
사이클 3
사이클 4
사이클 5

도면1d



도면1e

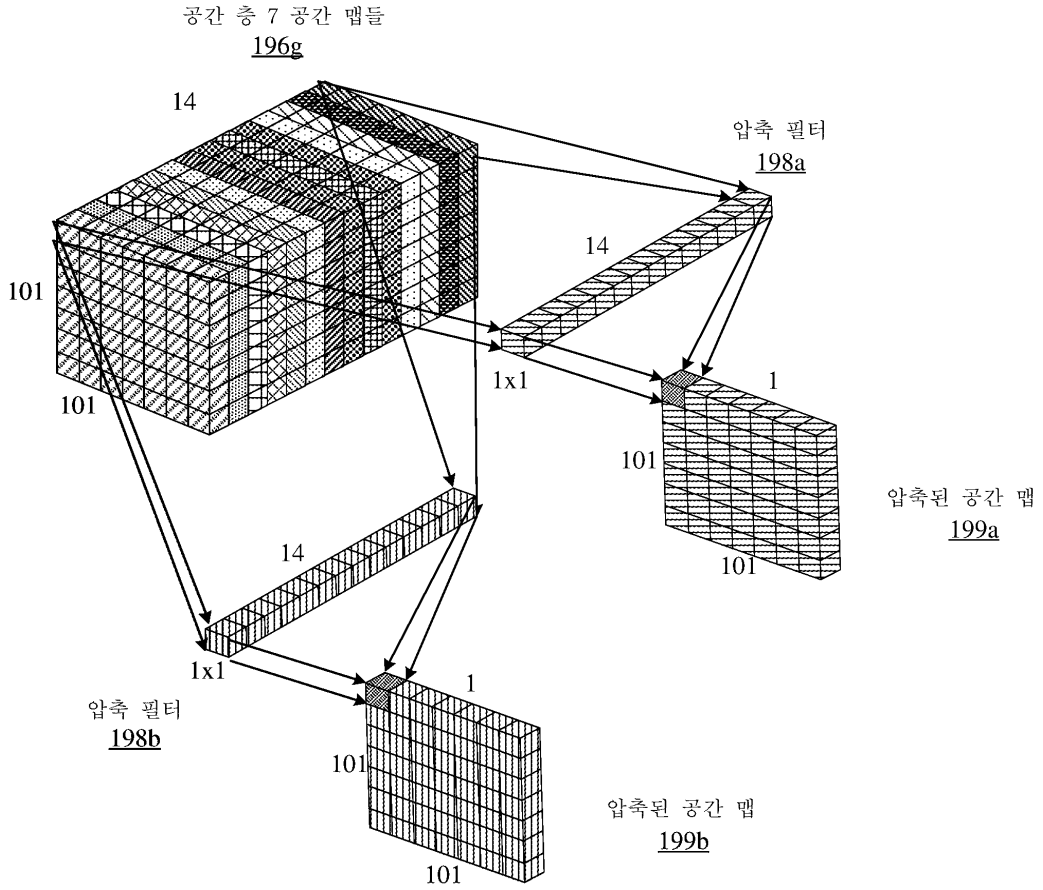


2개의 채널들 및 1x1 콘볼루션 커널을 갖는 입력 텐서

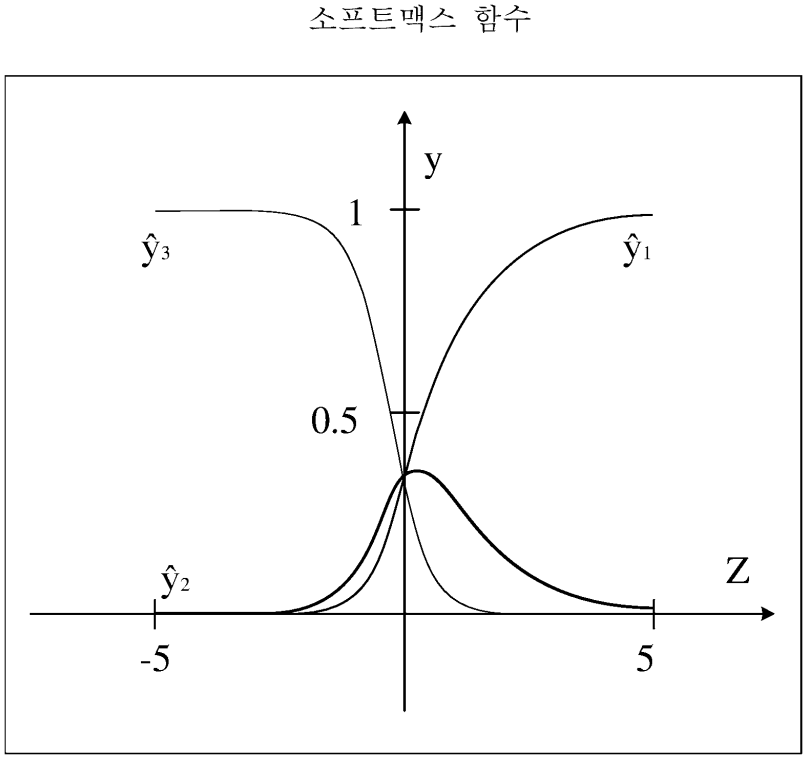
커널이 입력 텐서에 적용됨

1개의 채널을 갖는 출력

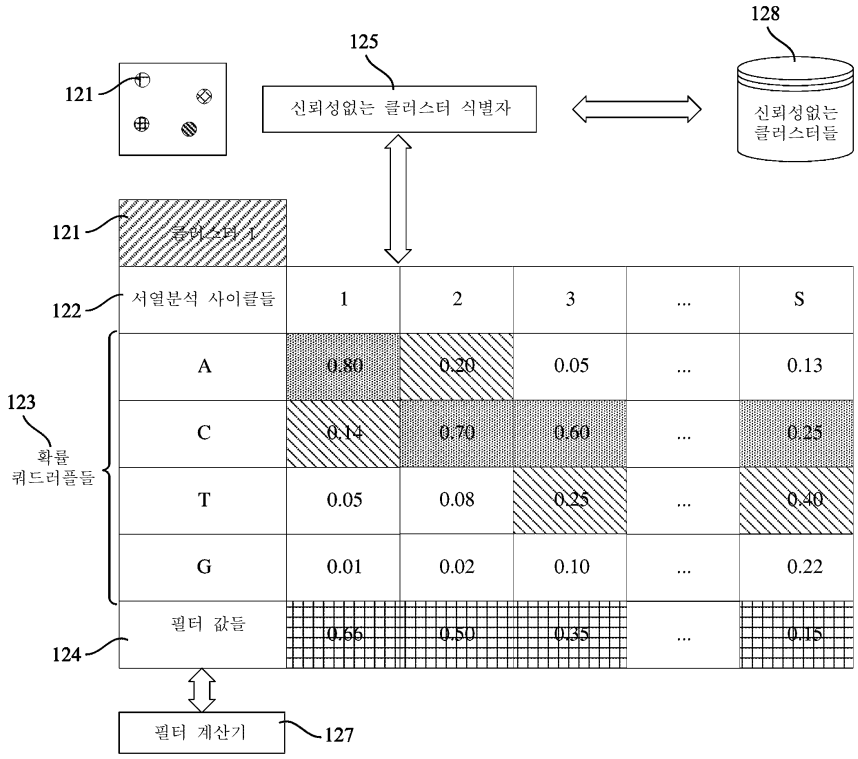
도면1f


공간 층 7 공간 맵들
196g
14
101
101
압축 필터
198a
14
1x1
1
101
101
압축된 공간 맵
199a
14
1x1
압축 필터
198b
1
101
101
압축된 공간 맵
199b

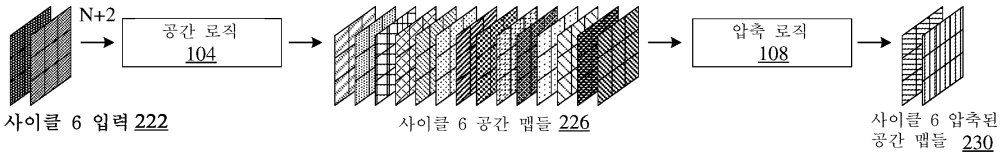
도면1g


소프트맥스 함수
y
1
0.5
ŷ3
ŷ1
ŷ2
Z
-5
5

도면1h



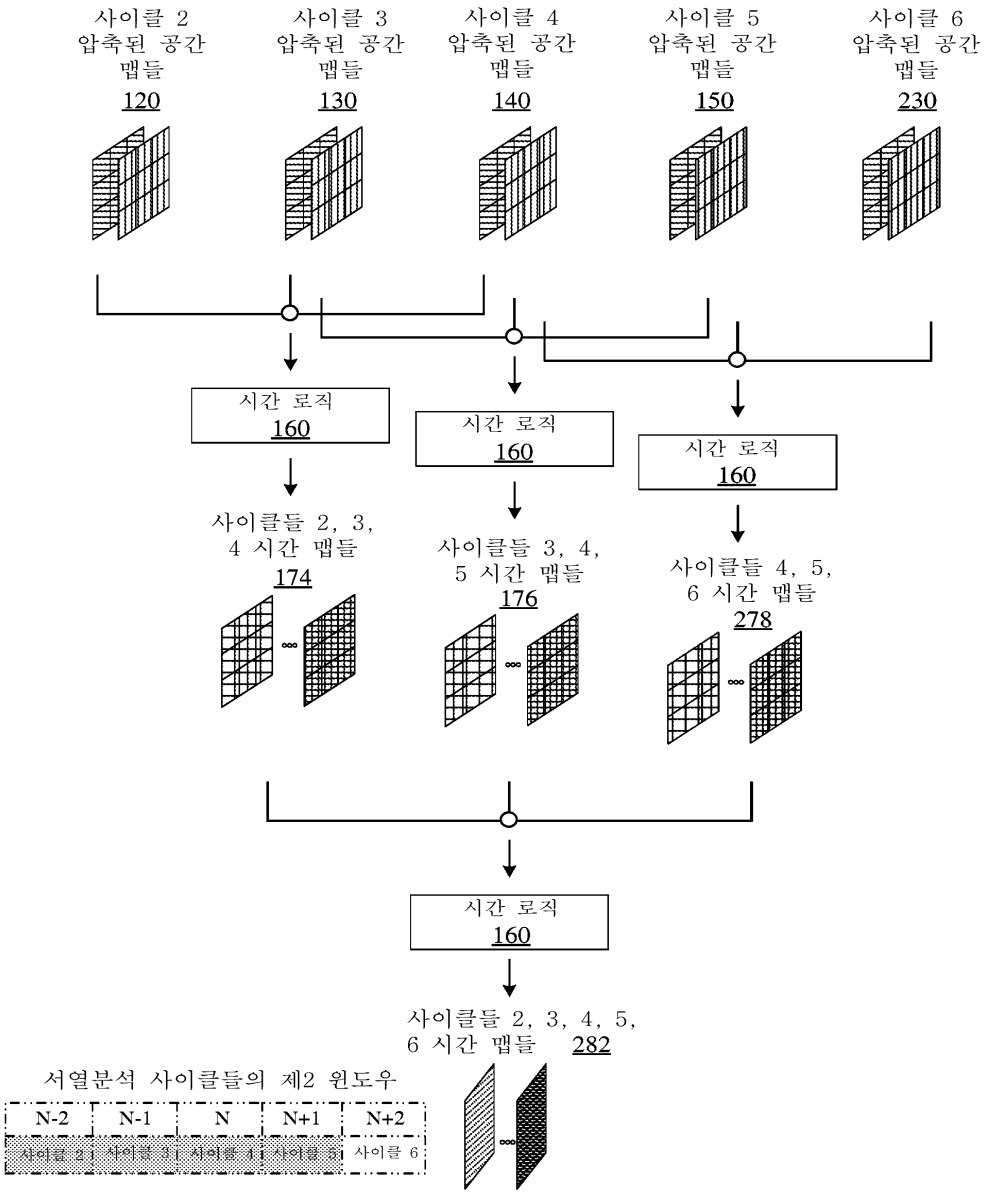
도면2a



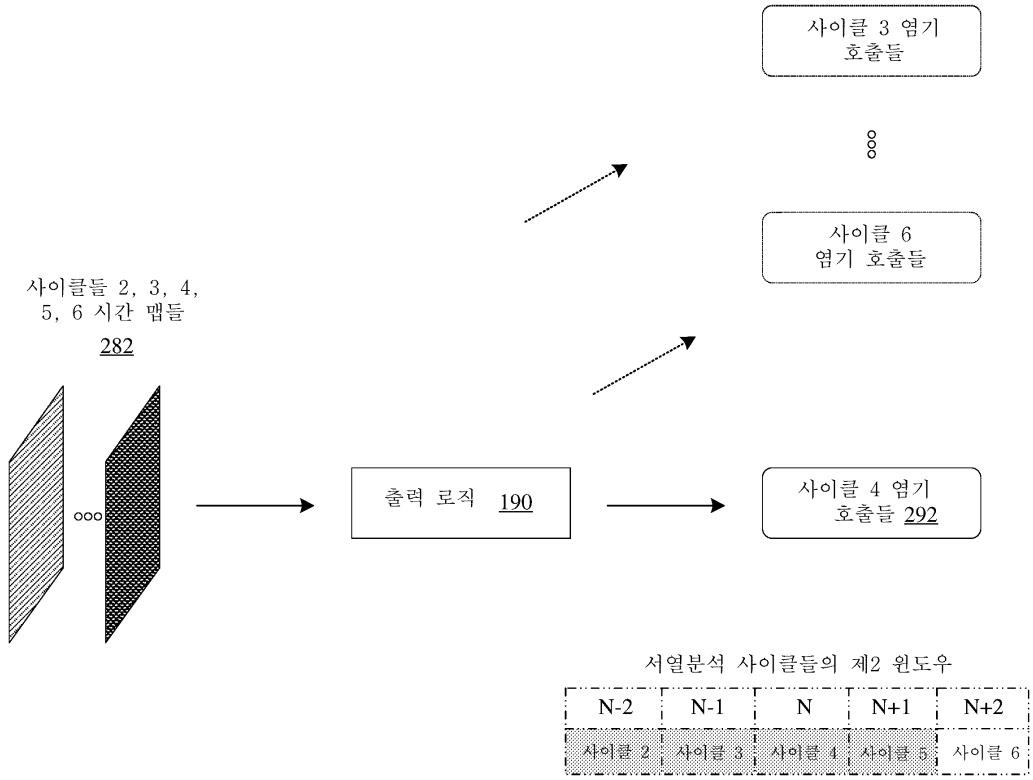
서열분석 사이클들의 제2 윈도우

N-2	N-1	N	N+1	N+2
사이클 2	사이클 3	사이클 4	사이클 5	사이클 6

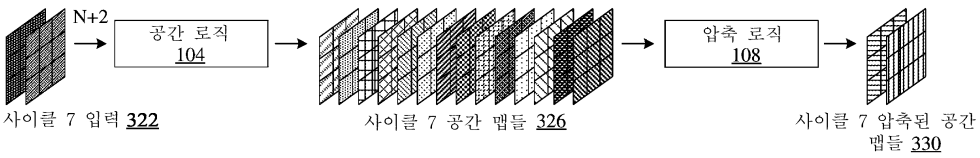
도면2b

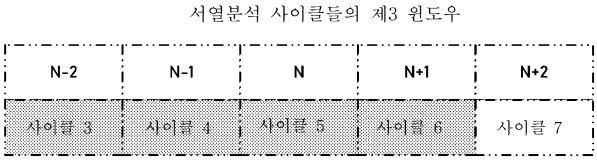

사이클 2 압축된 공간 맵들 120
사이클 3 압축된 공간 맵들 130
사이클 4 압축된 공간 맵들 140
사이클 5 압축된 공간 맵들 150
사이클 6 압축된 공간 맵들 230
시간 로직 160
시간 로직 160
시간 로직 160
사이클들 2, 3, 4 시간 맵들 174
사이클들 3, 4, 5 시간 맵들 176
사이클들 4, 5, 6 시간 맵들 278
시간 로직 160
사이클들 2, 3, 4, 5, 6 시간 맵들 282
서열분석 사이클들의 제2 윈도우
N-2
N-1
N
N+1
N+2
사이클 2
사이클 3
사이클 4
사이클 5
사이클 6

도면2c

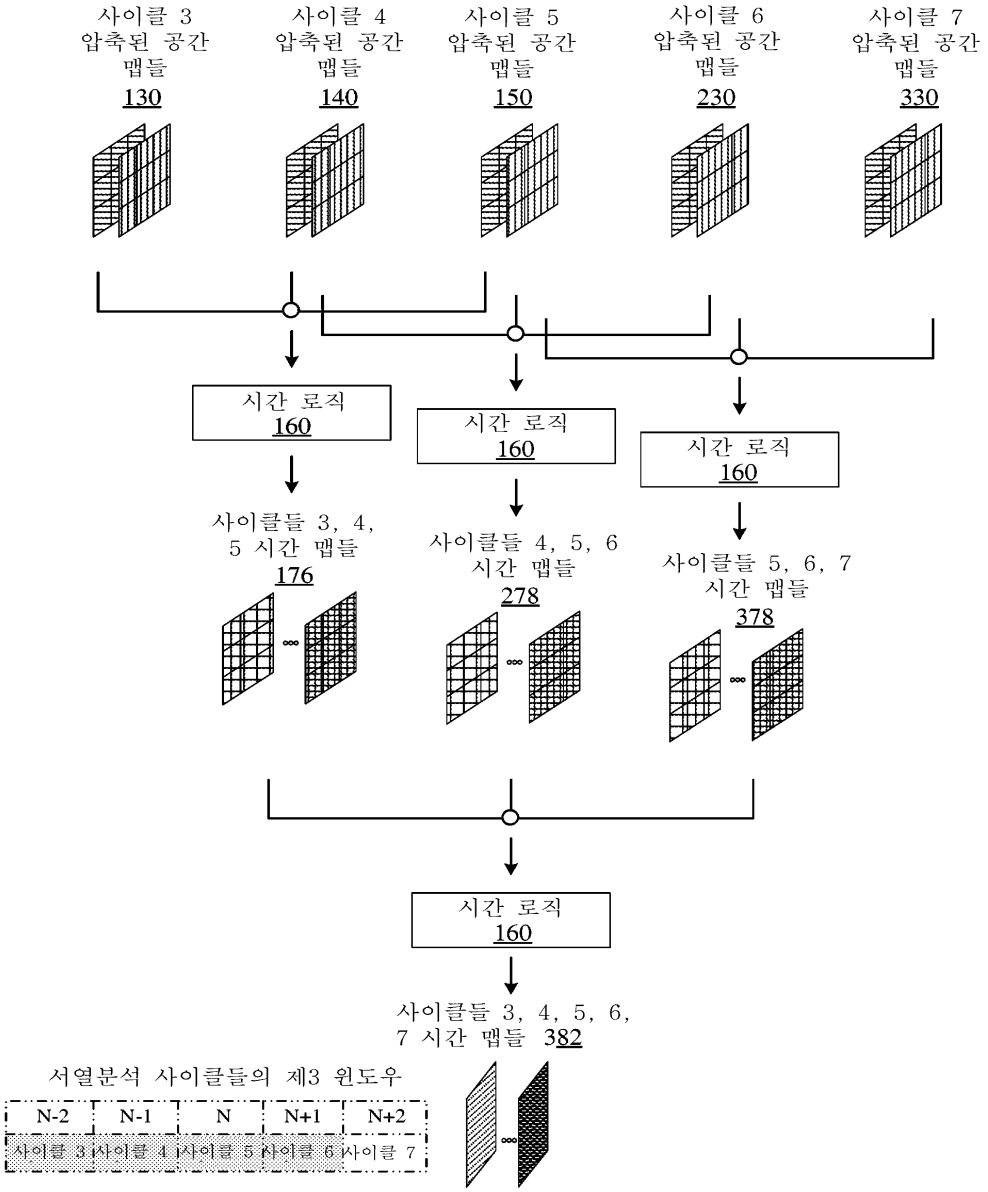

사이클 3 염기 호출들
사이클 6 염기 호출들
사이클들 2, 3, 4, 5, 6 시간 맵들
282
출력 로직 190
사이클 4 염기 호출들 292
서열분석 사이클들의 제2 윈도우
N-2
N-1
N
N+1
N+2
사이클 2
사이클 3
사이클 4
사이클 5
사이클 6

도면3a

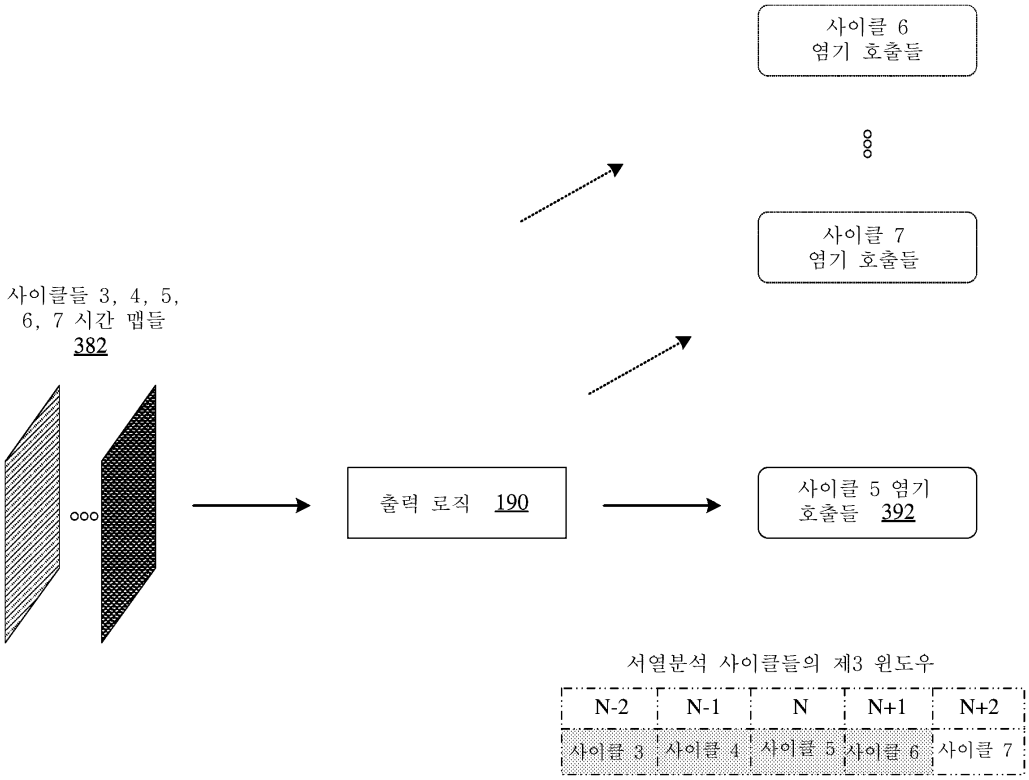

N+2
공간 로직 104
압축 로직 108
사이클 7 입력 322
사이클 7 공간 맵들 326
사이클 7 압축된 공간 맵들 330


서열분석 사이클들의 제3 윈도우
N-2
N-1
N
N+1
N+2
사이클 3
사이클 4
사이클 5
사이클 6
사이클 7

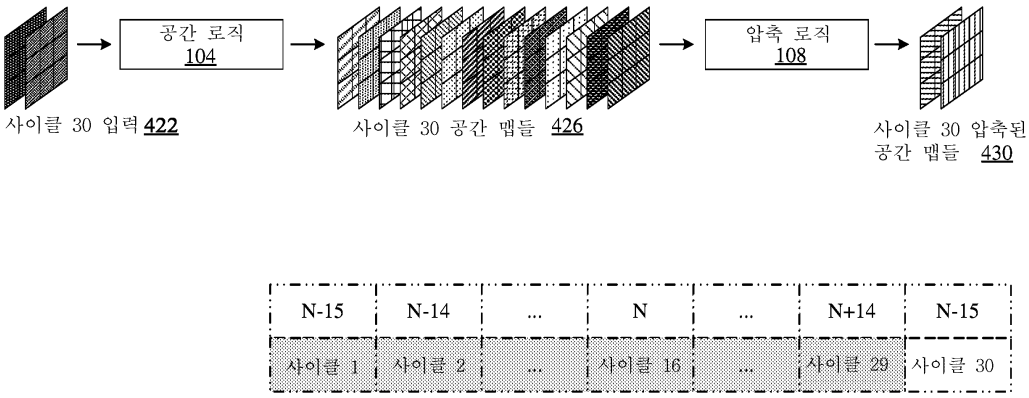
도면3b


사이클 3 압축된 공간 맵들 130
사이클 4 압축된 공간 맵들 140
사이클 5 압축된 공간 맵들 150
사이클 6 압축된 공간 맵들 230
사이클 7 압축된 공간 맵들 330
시간 로직 160
시간 로직 160
시간 로직 160
사이클들 3, 4, 5 시간 맵들 176
사이클들 4, 5, 6 시간 맵들 278
사이클들 5, 6, 7 시간 맵들 378
시간 로직 160
사이클들 3, 4, 5, 6, 7 시간 맵들 382
서열분석 사이클들의 제3 윈도우
N-2
N-1
N
N+1
N+2
사이클 3
사이클 4
사이클 5
사이클 6
사이클 7

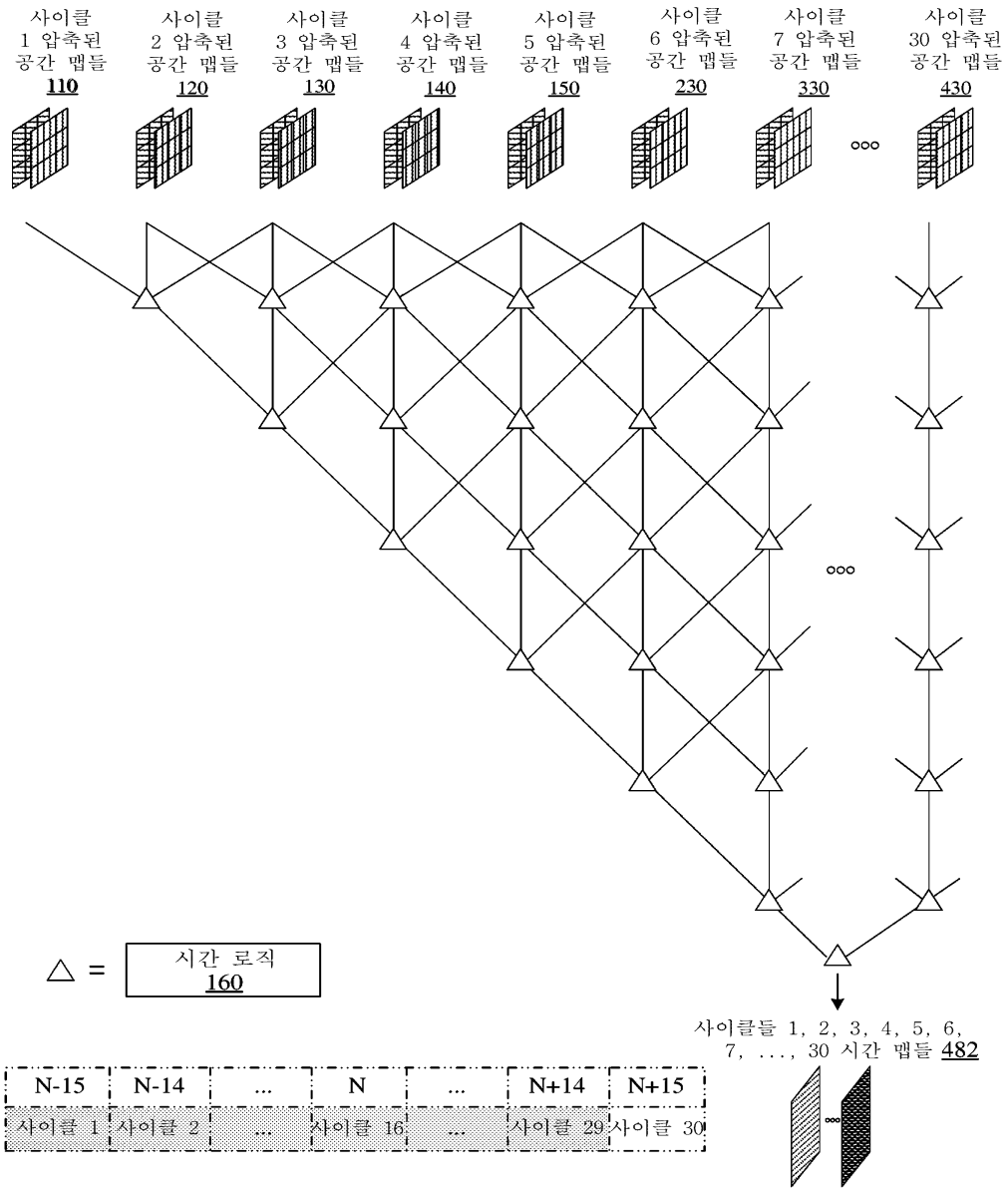
도면3c


사이클 6
염기 호출들
사이클 7
염기 호출들
사이클들 3, 4, 5, 6, 7 시간 맵들
382
출력 로직 190
사이클 5 염기 호출들 392
서열분석 사이클들의 제3 윈도우
N-2
N-1
N
N+1
N+2
사이클 3
사이클 4
사이클 5
사이클 6
사이클 7

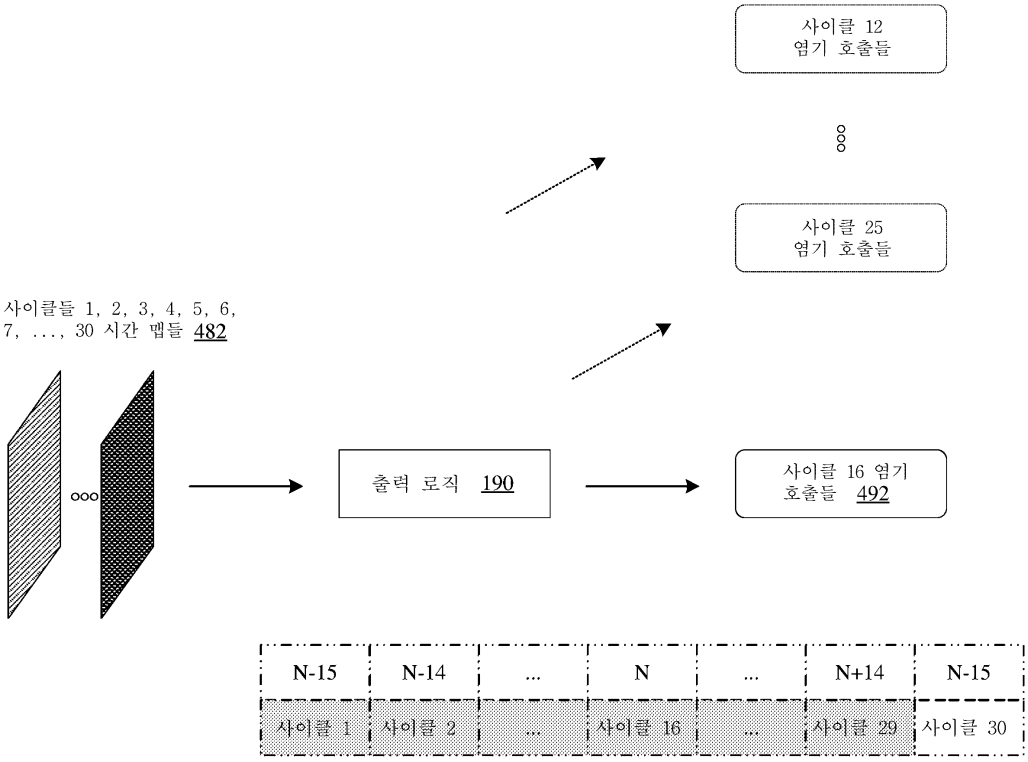
도면4a


공간 로직
104
압축 로직
108
사이클 30 입력 422
사이클 30 공간 맵들 426
사이클 30 압축된 공간 맵들 430
N-15
N-14
...
N
...
N+14
N-15
사이클 1
사이클 2
...
사이클 16
...
사이클 29
사이클 30

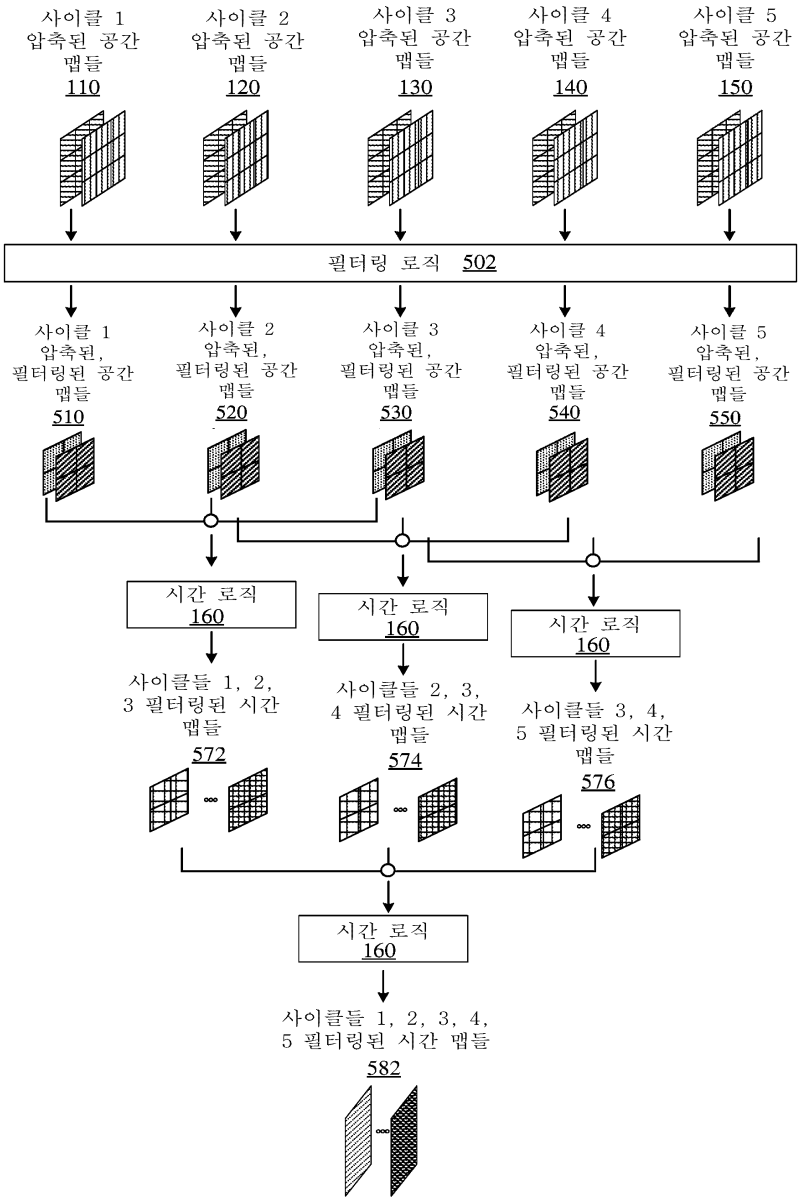
도면4b


사이클 1 압축된 공간 맵들 110
사이클 2 압축된 공간 맵들 120
사이클 3 압축된 공간 맵들 130
사이클 4 압축된 공간 맵들 140
사이클 5 압축된 공간 맵들 150
사이클 6 압축된 공간 맵들 230
사이클 7 압축된 공간 맵들 330
사이클 30 압축된 공간 맵들 430
△ = 시간 로직 160
사이클들 1, 2, 3, 4, 5, 6, 7, ..., 30 시간 맵들 482
N-15
N-14
...
N
...
N+14
N+15
사이클 1
사이클 2
...
사이클 16
...
사이클 29
사이클 30

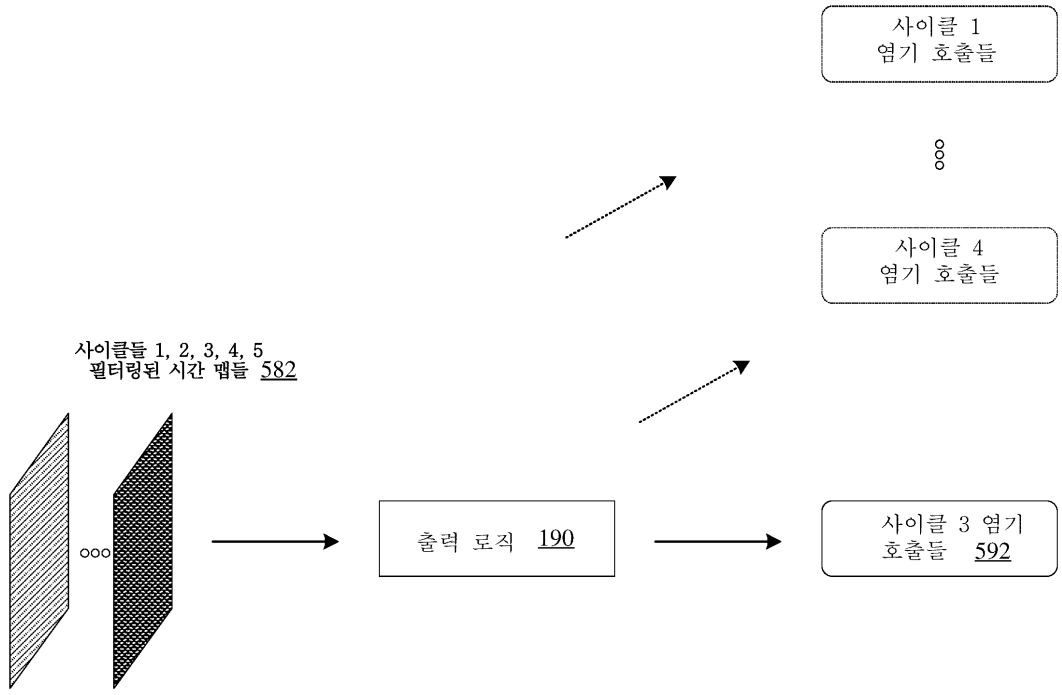
도면4c


사이클 12
염기 호출들
사이클 25
염기 호출들
사이클들 1, 2, 3, 4, 5, 6, 7, ..., 30 시간 맵들 482
출력 로직 190
사이클 16 염기 호출들 492
N-15
N-14
...
N
...
N+14
N-15
사이클 1
사이클 2
...
사이클 16
...
사이클 29
사이클 30

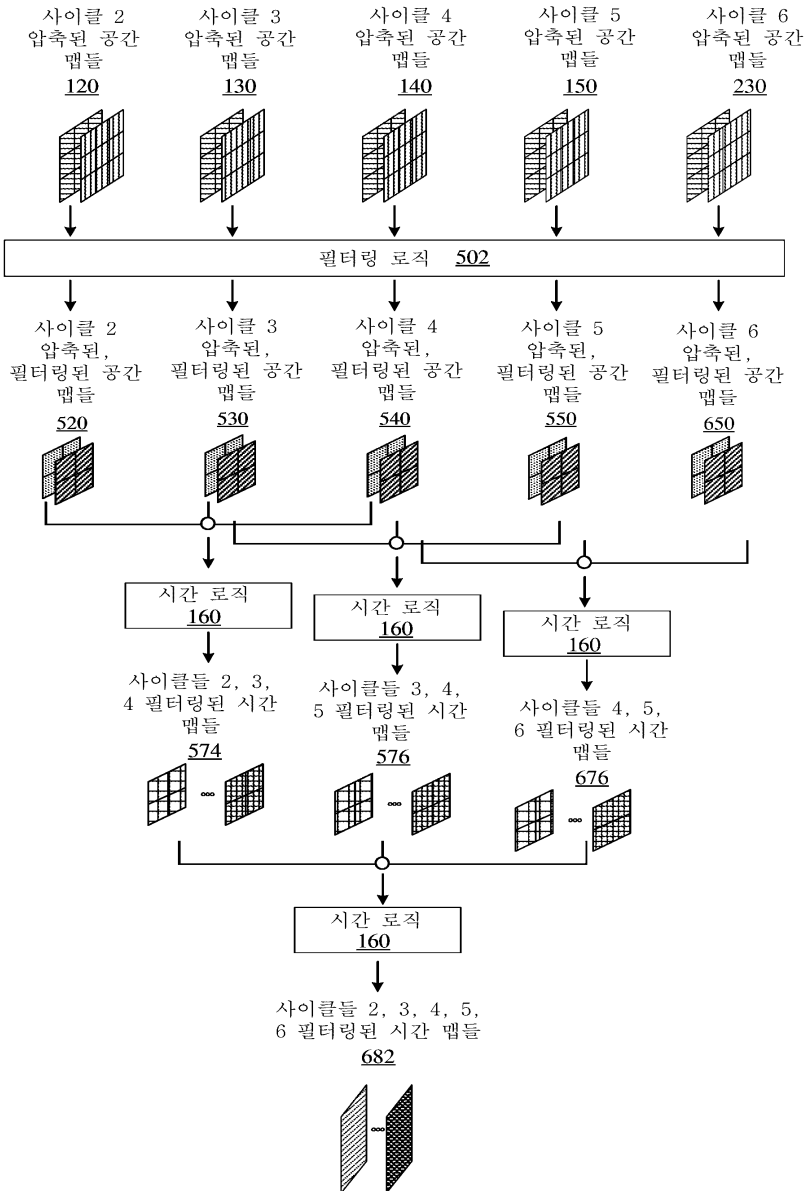
도면5a


사이클 1 압축된 공간 맵들 110
사이클 2 압축된 공간 맵들 120
사이클 3 압축된 공간 맵들 130
사이클 4 압축된 공간 맵들 140
사이클 5 압축된 공간 맵들 150
필터링 로직 502
사이클 1 압축된, 필터링된 공간 맵들 510
사이클 2 압축된, 필터링된 공간 맵들 520
사이클 3 압축된, 필터링된 공간 맵들 530
사이클 4 압축된, 필터링된 공간 맵들 540
사이클 5 압축된, 필터링된 공간 맵들 550
시간 로직 160
시간 로직 160
시간 로직 160
사이클들 1, 2, 3 필터링된 시간 맵들 572
사이클들 2, 3, 4 필터링된 시간 맵들 574
사이클들 3, 4, 5 필터링된 시간 맵들 576
시간 로직 160
사이클들 1, 2, 3, 4, 5 필터링된 시간 맵들 582

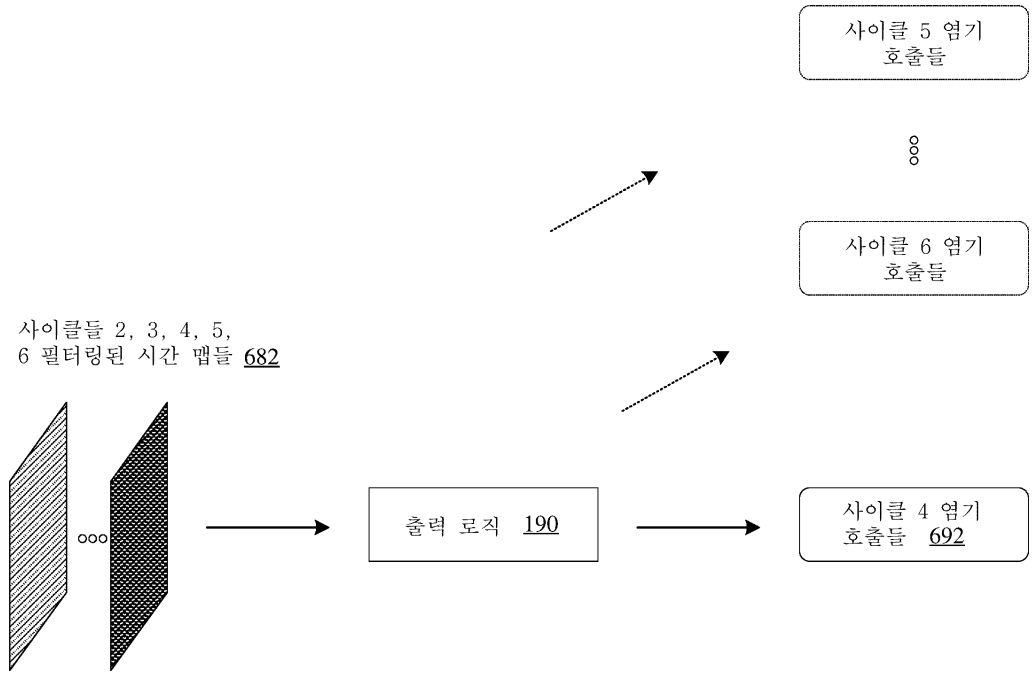
도면5b


사이클 1
염기 호출들
사이클 4
염기 호출들
사이클들 1, 2, 3, 4, 5
필터링된 시간 맵들 582
출력 로직 190
사이클 3 염기
호출들 592

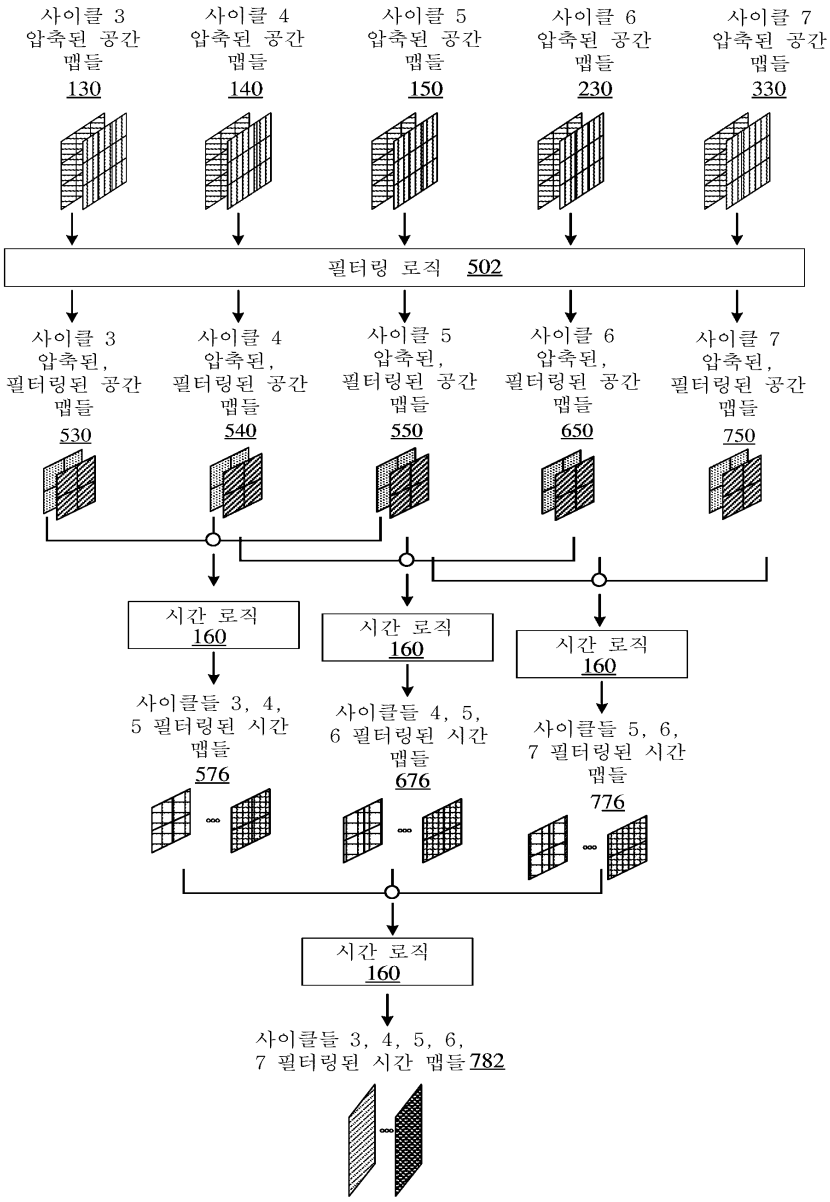
도면6a


사이클 2 압축된 공간 맵들 120
사이클 3 압축된 공간 맵들 130
사이클 4 압축된 공간 맵들 140
사이클 5 압축된 공간 맵들 150
사이클 6 압축된 공간 맵들 230
필터링 로직 502
사이클 2 압축된, 필터링된 공간 맵들 520
사이클 3 압축된, 필터링된 공간 맵들 530
사이클 4 압축된, 필터링된 공간 맵들 540
사이클 5 압축된, 필터링된 공간 맵들 550
사이클 6 압축된, 필터링된 공간 맵들 650
시간 로직 160
시간 로직 160
시간 로직 160
사이클들 2, 3, 4 필터링된 시간 맵들 574
사이클들 3, 4, 5 필터링된 시간 맵들 576
사이클들 4, 5, 6 필터링된 시간 맵들 676
시간 로직 160
사이클들 2, 3, 4, 5, 6 필터링된 시간 맵들 682

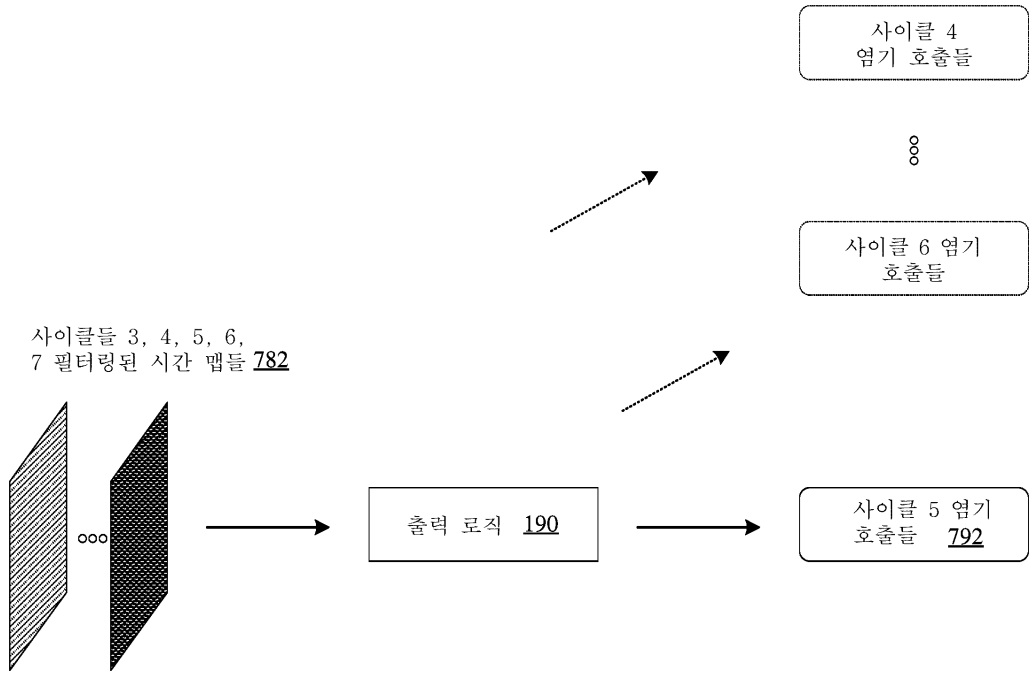
도면6b


사이클 5 염기 호출들
사이클 6 염기 호출들
사이클들 2, 3, 4, 5, 6 필터링된 시간 맵들 682
출력 로직 190
사이클 4 염기 호출들 692

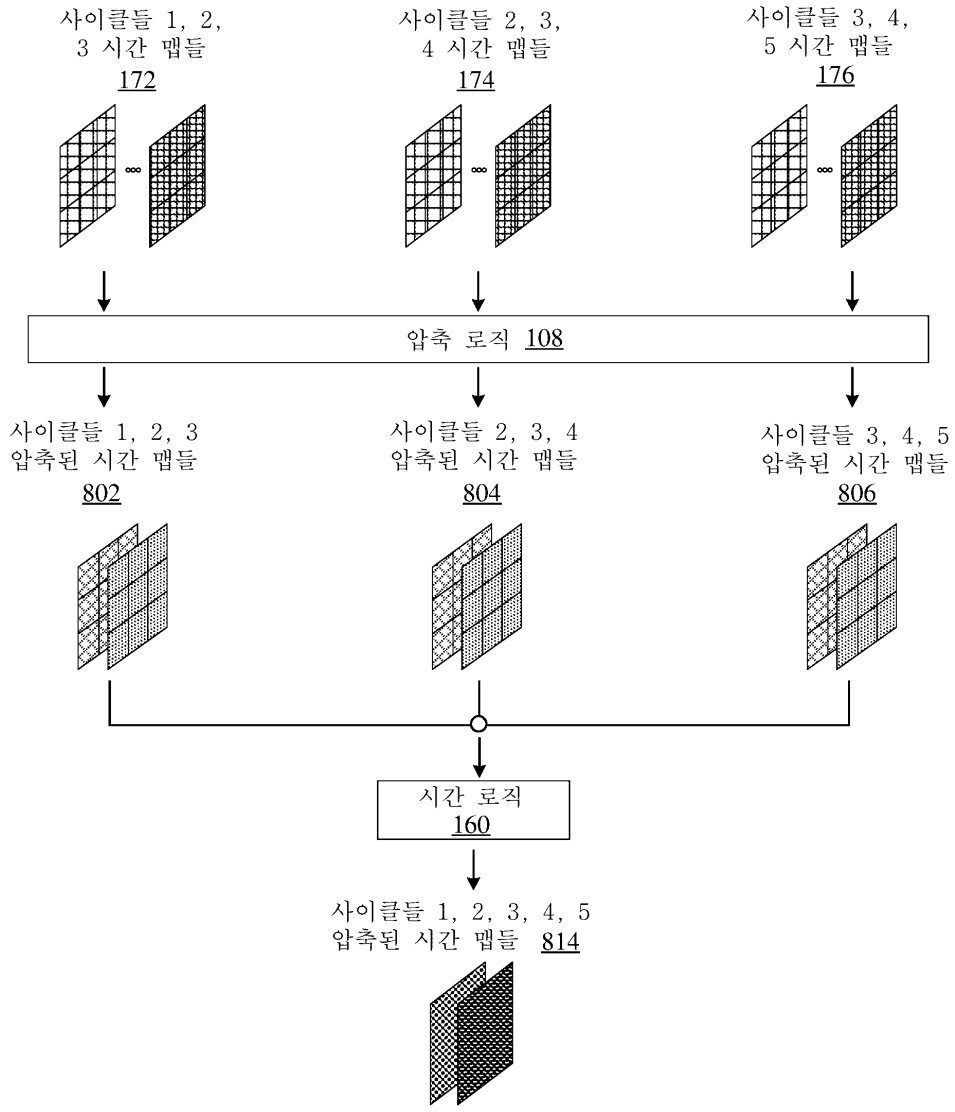
도면7a


사이클 3 압축된 공간 맵들 130
사이클 4 압축된 공간 맵들 140
사이클 5 압축된 공간 맵들 150
사이클 6 압축된 공간 맵들 230
사이클 7 압축된 공간 맵들 330
필터링 로직 502
사이클 3 압축된, 필터링된 공간 맵들 530
사이클 4 압축된, 필터링된 공간 맵들 540
사이클 5 압축된, 필터링된 공간 맵들 550
사이클 6 압축된, 필터링된 공간 맵들 650
사이클 7 압축된, 필터링된 공간 맵들 750
시간 로직 160
시간 로직 160
시간 로직 160
사이클들 3, 4, 5 필터링된 시간 맵들 576
사이클들 4, 5, 6 필터링된 시간 맵들 676
사이클들 5, 6, 7 필터링된 시간 맵들 776
시간 로직 160
사이클들 3, 4, 5, 6, 7 필터링된 시간 맵들 782

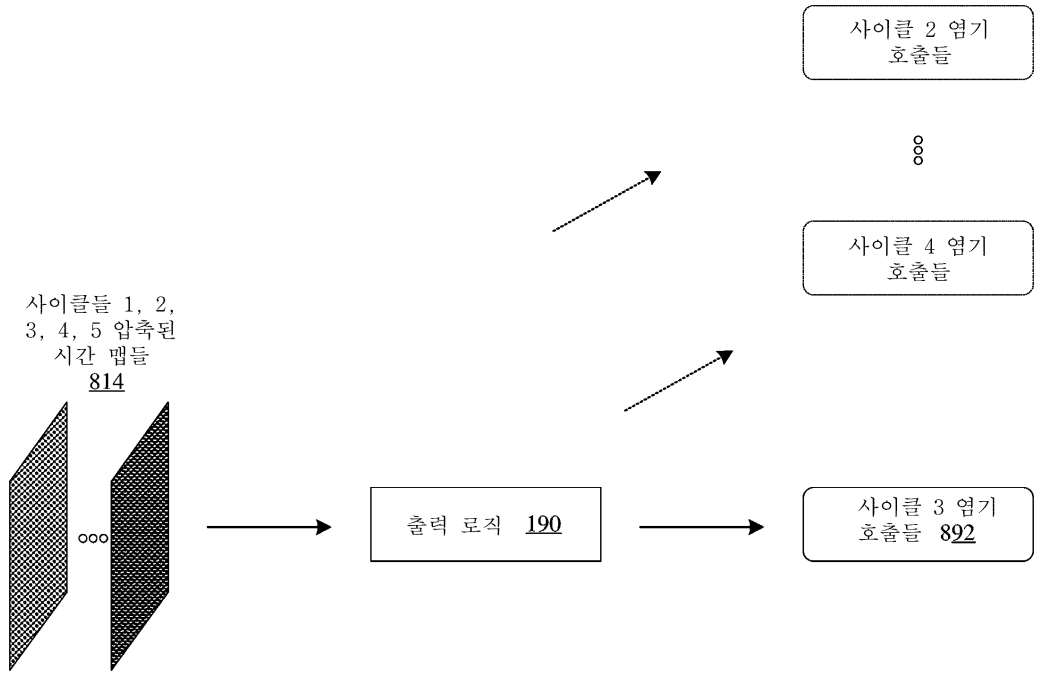
도면7b


사이클 4
염기 호출들
사이클 6 염기
호출들
사이클들 3, 4, 5, 6,
7 필터링된 시간 맵들 782
출력 로직 190
사이클 5 염기
호출들 792

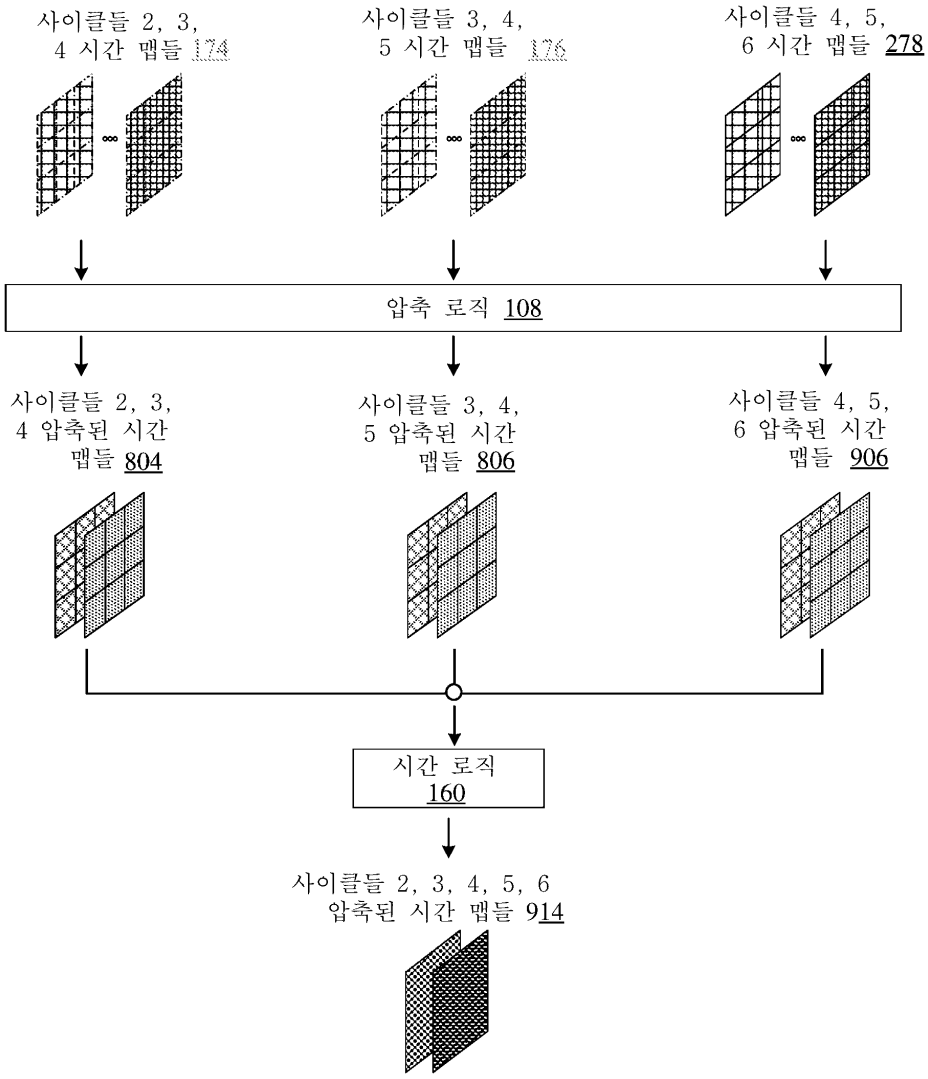
도면8a


사이클들 1, 2, 3 시간 맵들
172
사이클들 2, 3, 4 시간 맵들
174
사이클들 3, 4, 5 시간 맵들
176
압축 로직 108
사이클들 1, 2, 3 압축된 시간 맵들
802
사이클들 2, 3, 4 압축된 시간 맵들
804
사이클들 3, 4, 5 압축된 시간 맵들
806
시간 로직
160
사이클들 1, 2, 3, 4, 5 압축된 시간 맵들 814

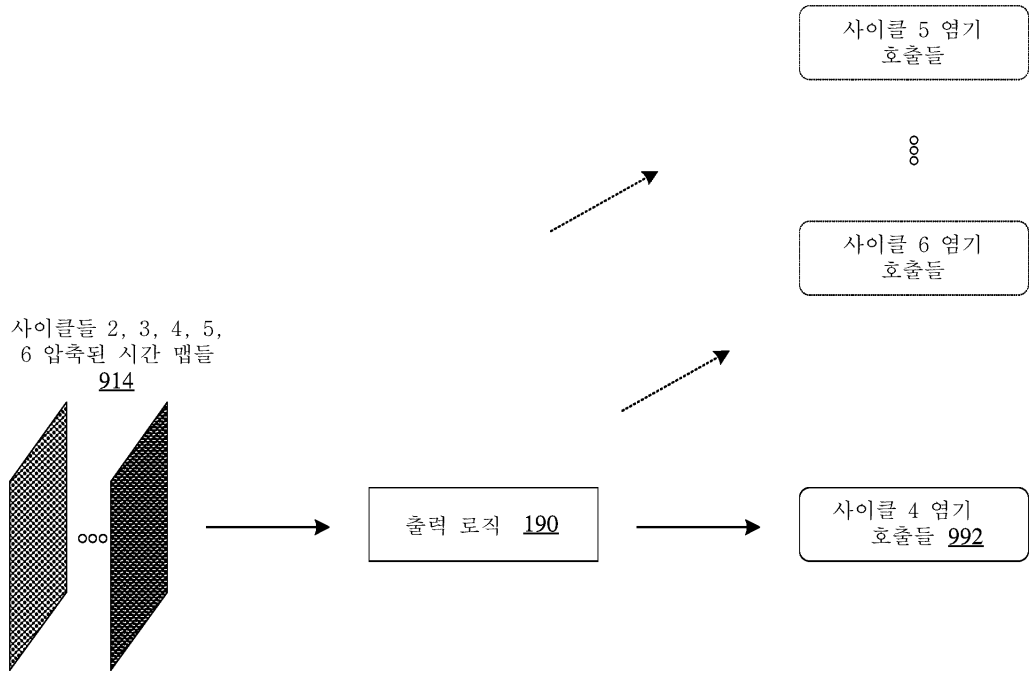
도면8b


사이클 2 염기
호출들
사이클 4 염기
호출들
사이클들 1, 2,
3, 4, 5 압축된
시간 맵들
814
출력 로직 190
사이클 3 염기
호출들 892

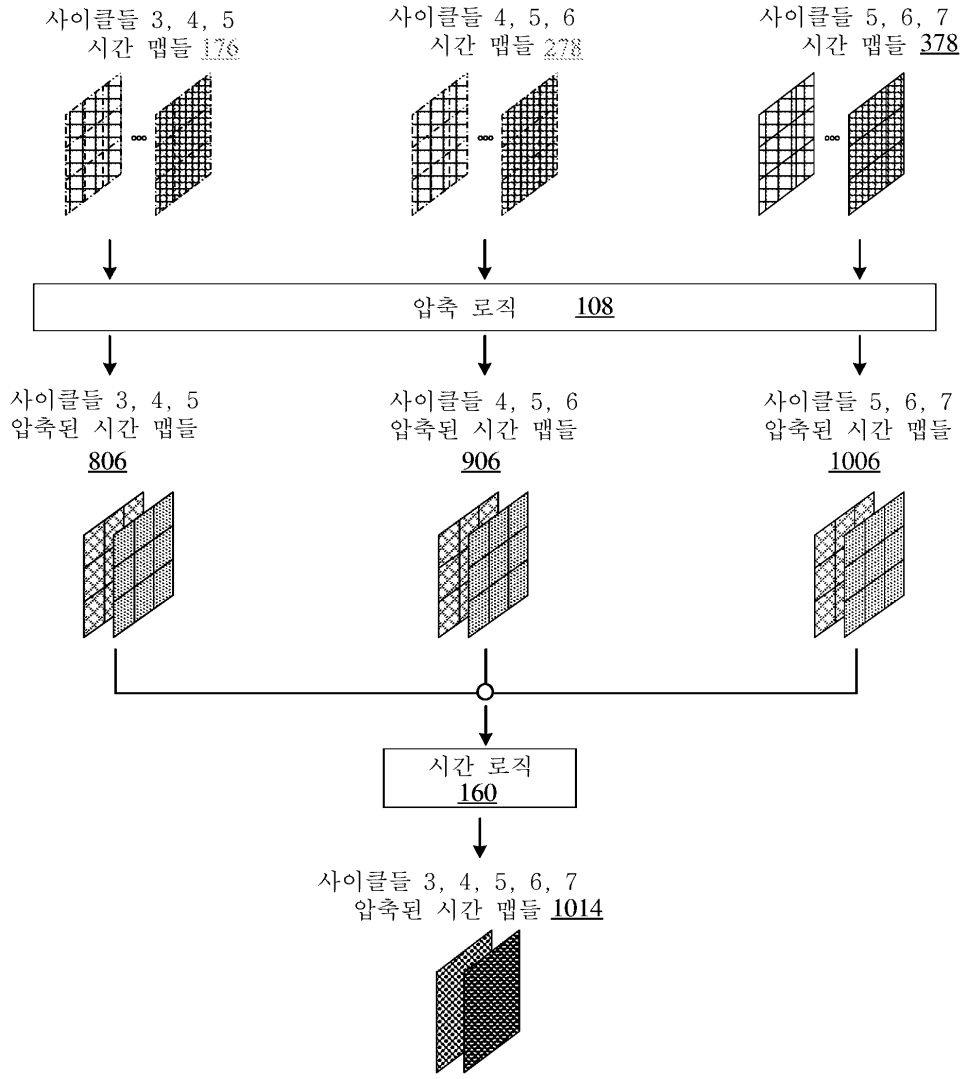
도면9a


사이클들 2, 3, 4 시간 맵들 174
사이클들 3, 4, 5 시간 맵들 176
사이클들 4, 5, 6 시간 맵들 278
압축 로직 108
사이클들 2, 3, 4 압축된 시간 맵들 804
사이클들 3, 4, 5 압축된 시간 맵들 806
사이클들 4, 5, 6 압축된 시간 맵들 906
시간 로직 160
사이클들 2, 3, 4, 5, 6 압축된 시간 맵들 914

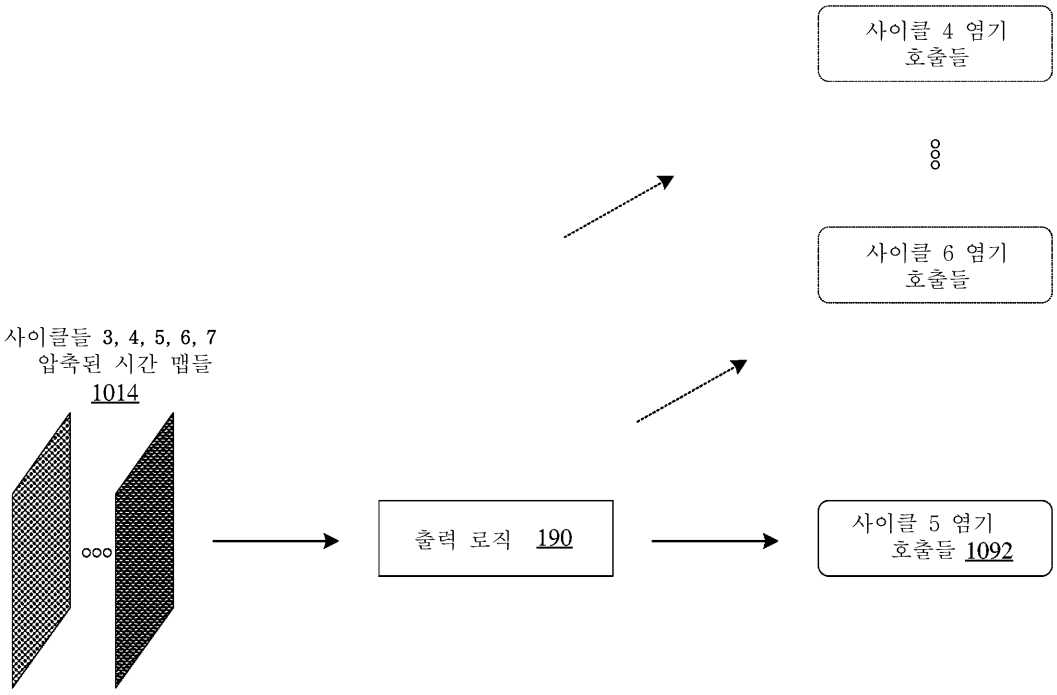
도면9b


사이클 5 염기
호출들
사이클 6 염기
호출들
사이클들 2, 3, 4, 5,
6 압축된 시간 맵들
914
출력 로직 190
사이클 4 염기
호출들 992

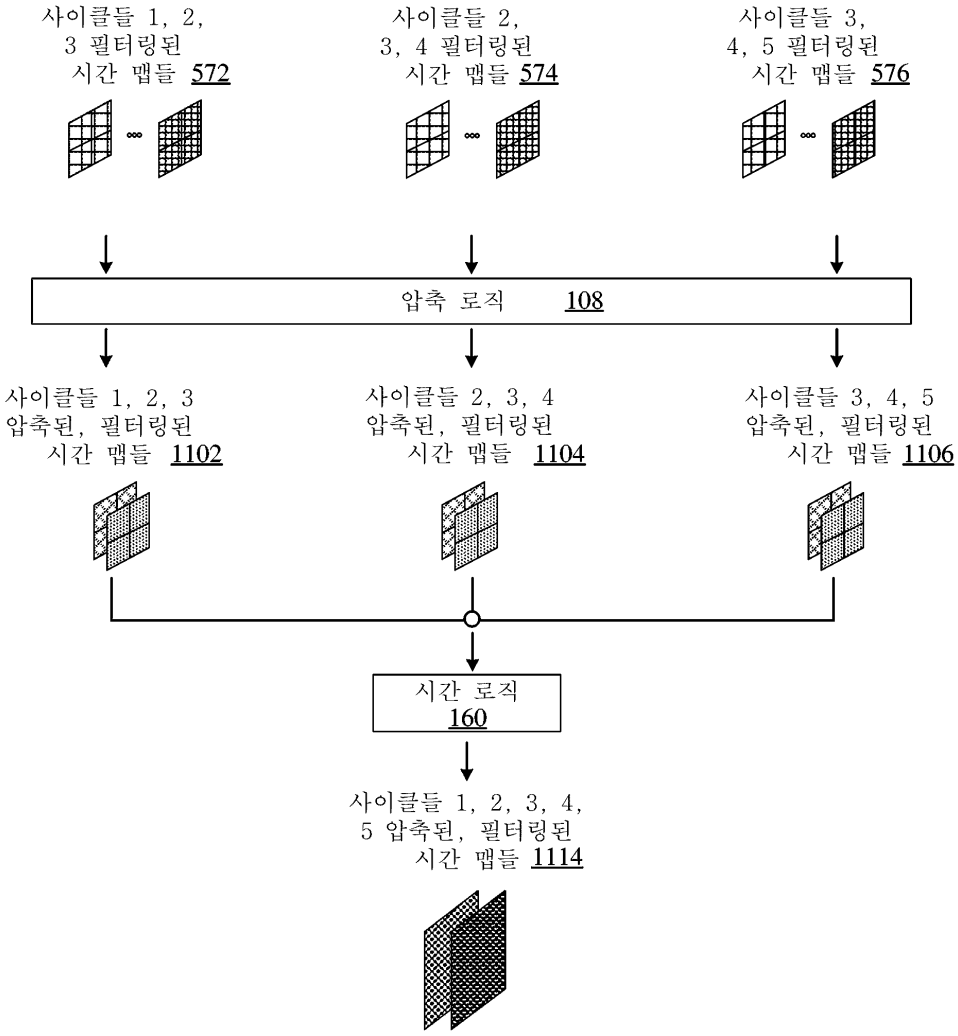
도면10a


사이클들 3, 4, 5
시간 맵들 176
사이클들 4, 5, 6
시간 맵들 278
사이클들 5, 6, 7
시간 맵들 378
압축 로직 108
사이클들 3, 4, 5
압축된 시간 맵들
806
사이클들 4, 5, 6
압축된 시간 맵들
906
사이클들 5, 6, 7
압축된 시간 맵들
1006
시간 로직
160
사이클들 3, 4, 5, 6, 7
압축된 시간 맵들 1014

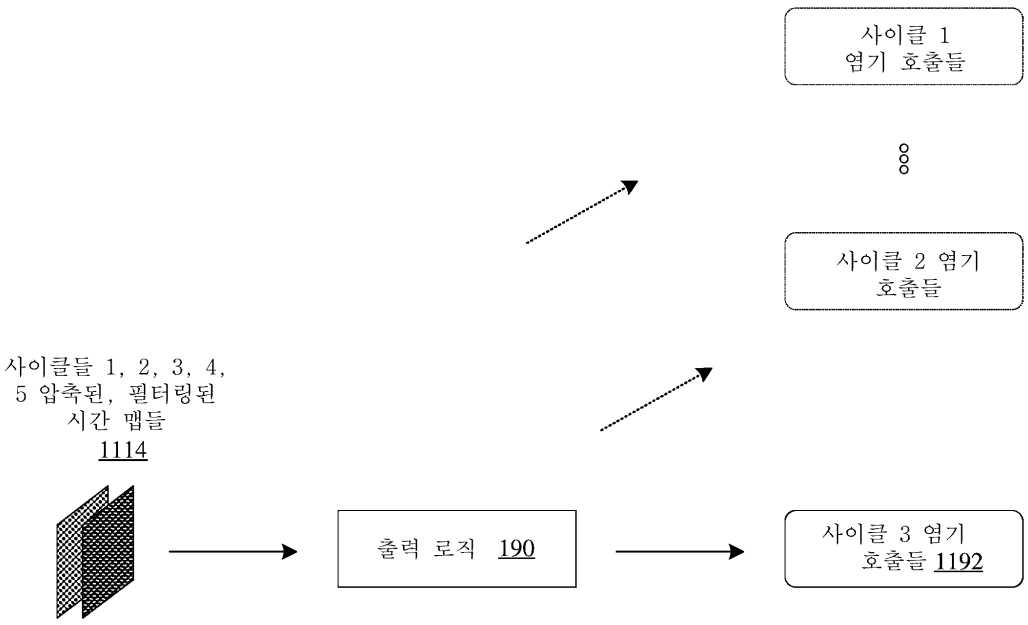
도면10b


사이클 4 염기
호출들
사이클 6 염기
호출들
사이클들 3, 4, 5, 6, 7
압축된 시간 맵들
1014
출력 로직 190
사이클 5 염기
호출들 1092

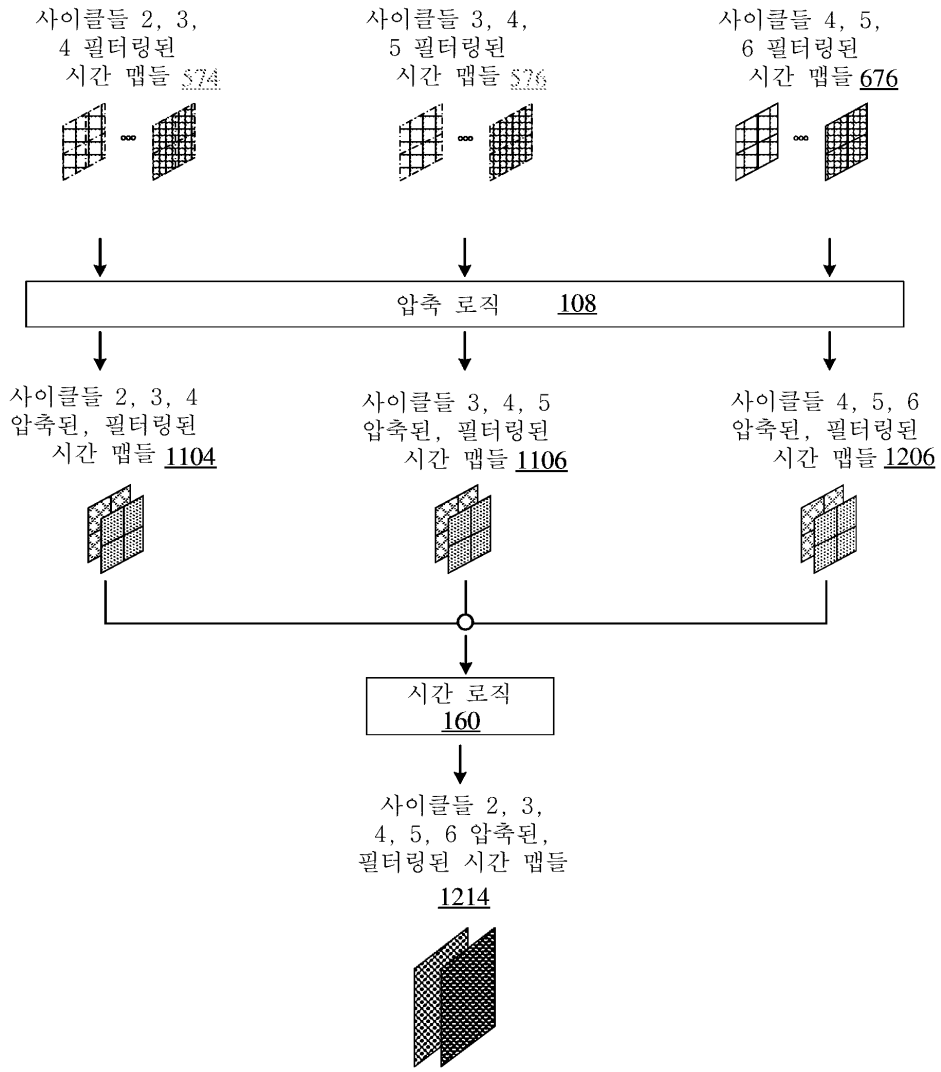
도면11a


사이클들 1, 2, 3 필터링된 시간 맵들 572
사이클들 2, 3, 4 필터링된 시간 맵들 574
사이클들 3, 4, 5 필터링된 시간 맵들 576
압축 로직 108
사이클들 1, 2, 3 압축된, 필터링된 시간 맵들 1102
사이클들 2, 3, 4 압축된, 필터링된 시간 맵들 1104
사이클들 3, 4, 5 압축된, 필터링된 시간 맵들 1106
시간 로직 160
사이클들 1, 2, 3, 4, 5 압축된, 필터링된 시간 맵들 1114

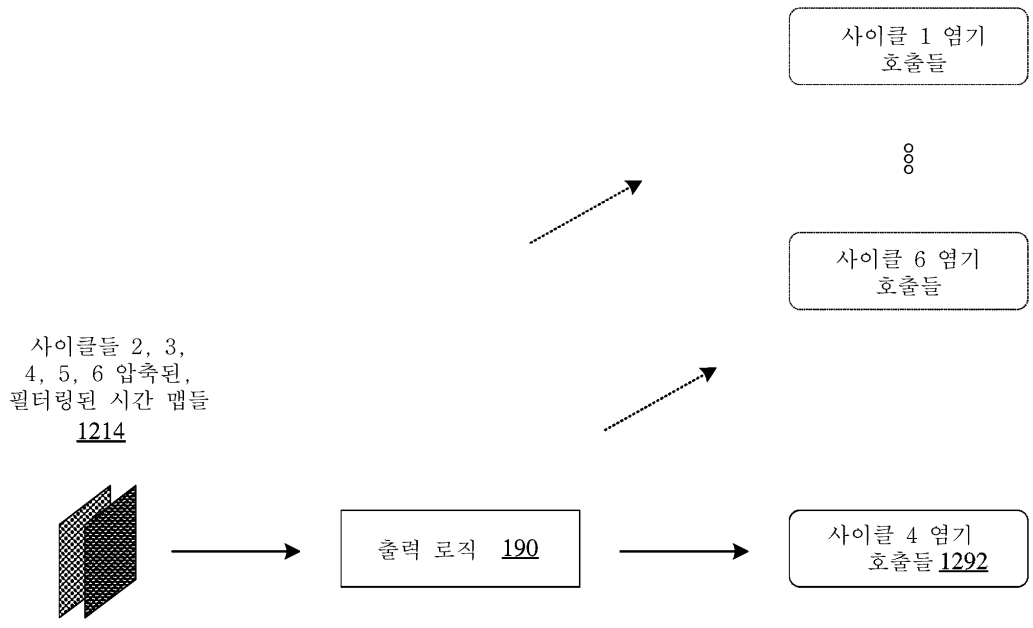
도면11b


사이클 1 염기 호출들
사이클 2 염기 호출들
사이클들 1, 2, 3, 4, 5 압축된, 필터링된 시간 맵들 1114
출력 로직 190
사이클 3 염기 호출들 1192

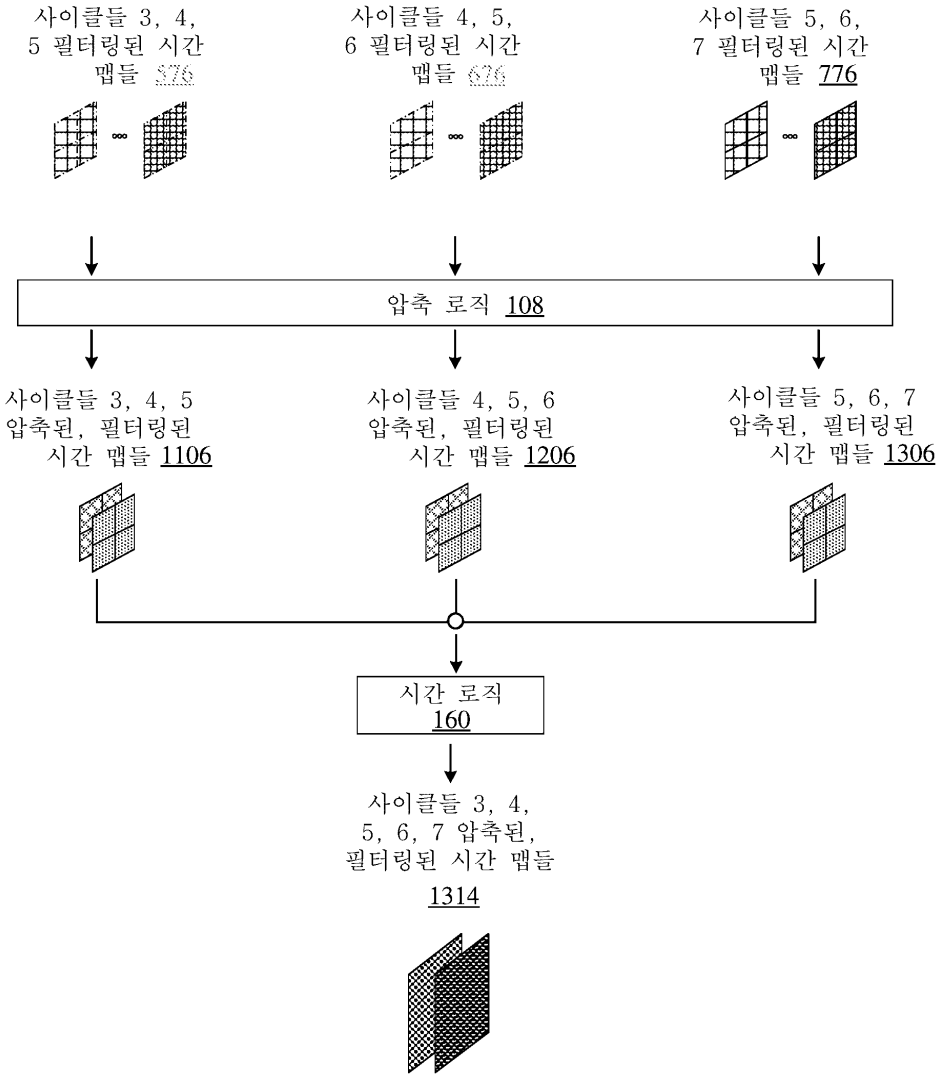
도면12a


사이클들 2, 3, 4 필터링된 시간 맵들 674
사이클들 3, 4, 5 필터링된 시간 맵들 676
사이클들 4, 5, 6 필터링된 시간 맵들 676
압축 로직 108
사이클들 2, 3, 4 압축된, 필터링된 시간 맵들 1104
사이클들 3, 4, 5 압축된, 필터링된 시간 맵들 1106
사이클들 4, 5, 6 압축된, 필터링된 시간 맵들 1206
시간 로직 160
사이클들 2, 3, 4, 5, 6 압축된, 필터링된 시간 맵들 1214

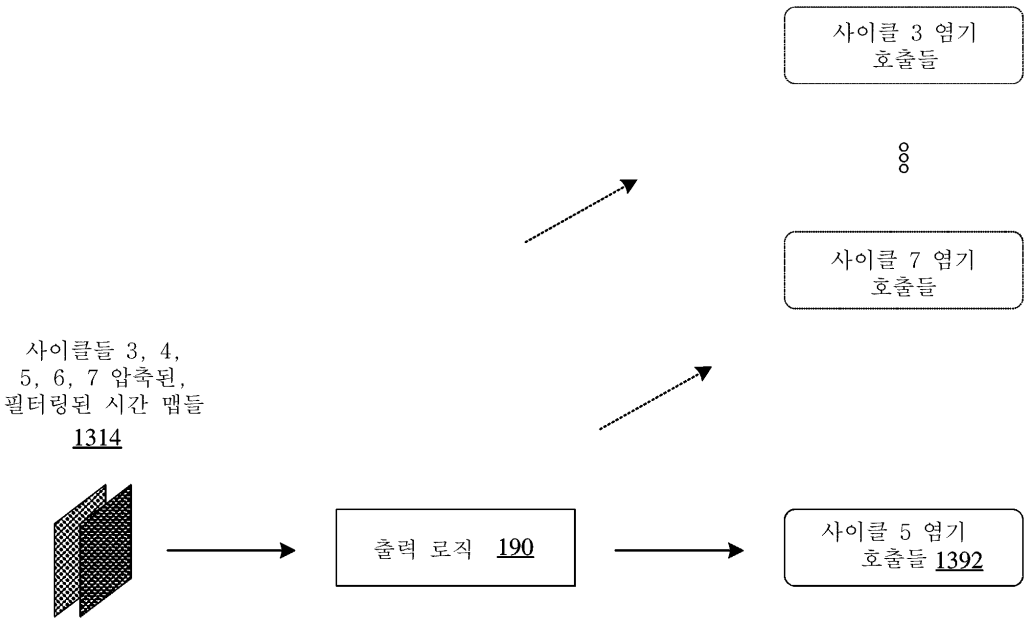
도면12b


사이클 1 염기 호출들
사이클 6 염기 호출들
사이클들 2, 3, 4, 5, 6 압축된, 필터링된 시간 맵들 1214
출력 로직 190
사이클 4 염기 호출들 1292

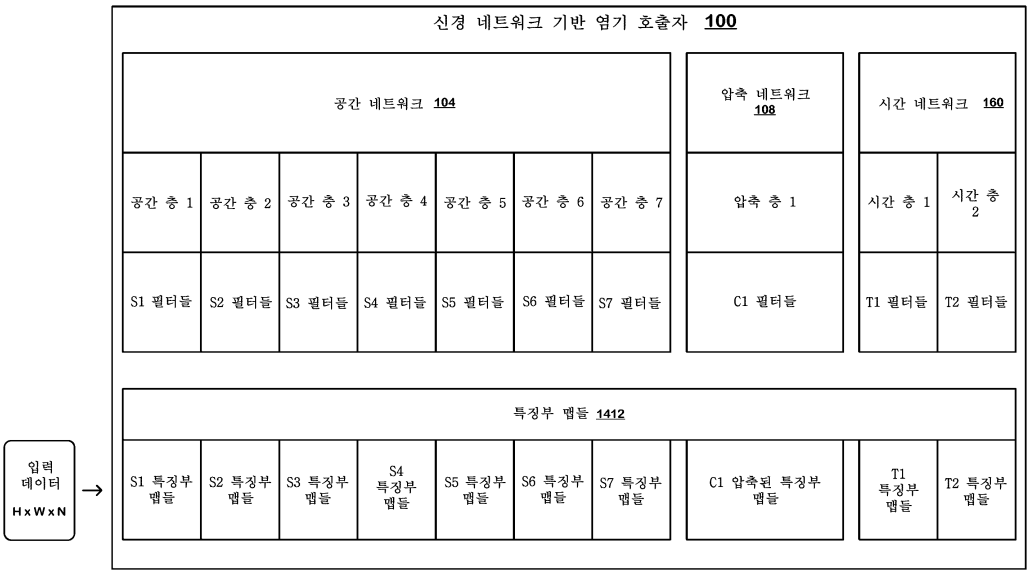
도면13a


사이클들 3, 4, 5 필터링된 시간 맵들 576
사이클들 4, 5, 6 필터링된 시간 맵들 676
사이클들 5, 6, 7 필터링된 시간 맵들 776
압축 로직 108
사이클들 3, 4, 5 압축된, 필터링된 시간 맵들 1106
사이클들 4, 5, 6 압축된, 필터링된 시간 맵들 1206
사이클들 5, 6, 7 압축된, 필터링된 시간 맵들 1306
시간 로직 160
사이클들 3, 4, 5, 6, 7 압축된, 필터링된 시간 맵들 1314

도면13b


사이클들 3, 4, 5, 6, 7 압축된, 필터링된 시간 맵들 1314
출력 로직 190
사이클 3 염기 호출들
사이클 7 염기 호출들
사이클 5 염기 호출들 1392

도면14



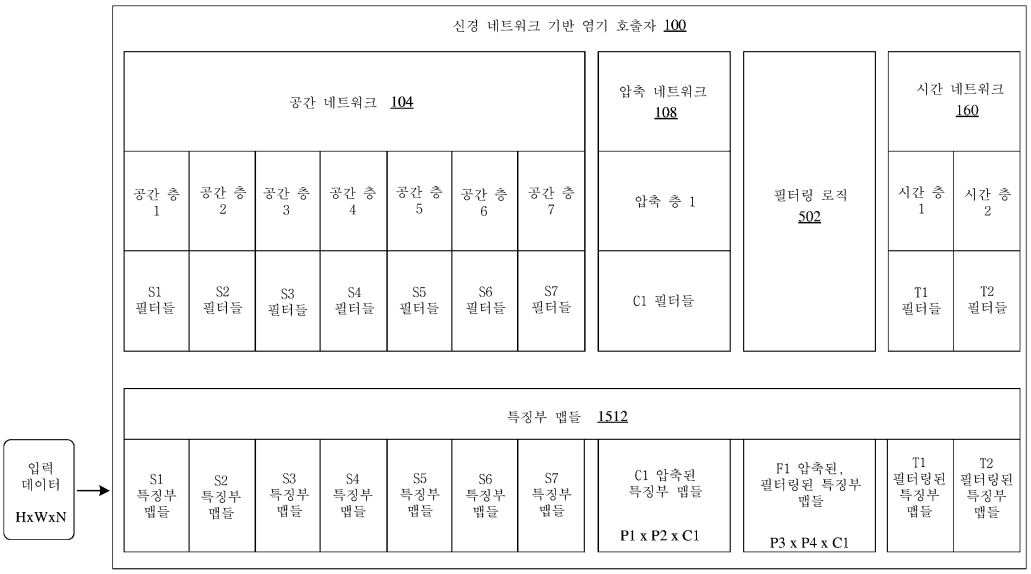
H = 64, 78, 115, 230
W = 64, 78, 115, 230
N = 1, 2, 3, 4

S1 = 7, 14, 21, 64, 128, 254
S2 = 7, 14, 21, 64, 128, 254
S3 = 7, 14, 21, 64, 128, 254
S4 = 7, 14, 21, 64, 128, 254
S5 = 7, 14, 21, 64, 128, 254
S6 = 7, 14, 21, 64, 128, 254
S7 = 7, 14, 21, 64, 128, 254

C1 = 1, 2, 3, 4, 이상

T1 = 7, 14, 21, 64, 128, 254
T2 = 7, 14, 21, 64, 128, 254

도면15



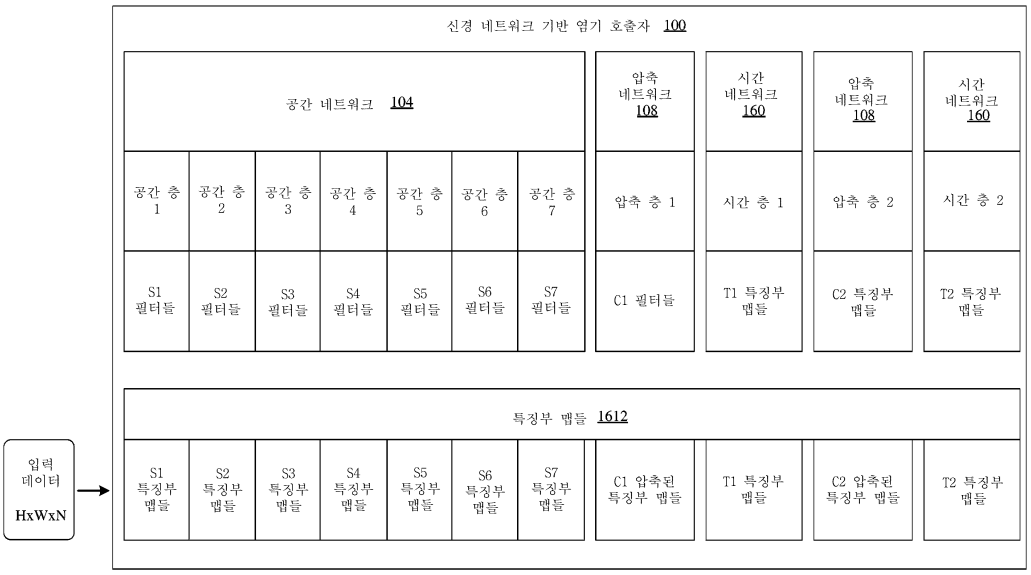
H = 64, 78, 115, 230
W = 64, 78, 115, 230
N = 1, 2, 3, 4

S1 = 7, 14, 21, 64, 128, 254
S2 = 7, 14, 21, 64, 128, 254
S3 = 7, 14, 21, 64, 128, 254
S4 = 7, 14, 21, 64, 128, 254
S5 = 7, 14, 21, 64, 128, 254
S6 = 7, 14, 21, 64, 128, 254
S7 = 7, 14, 21, 64, 128, 254

C1 = 1, 2, 3, 4, 이상
P3 = 25% 의 P1
P2 = 25% 의 P4

T1 = 7, 14, 21, 64, 128, 254
T2 = 7, 14, 21, 64, 128, 254

도면16



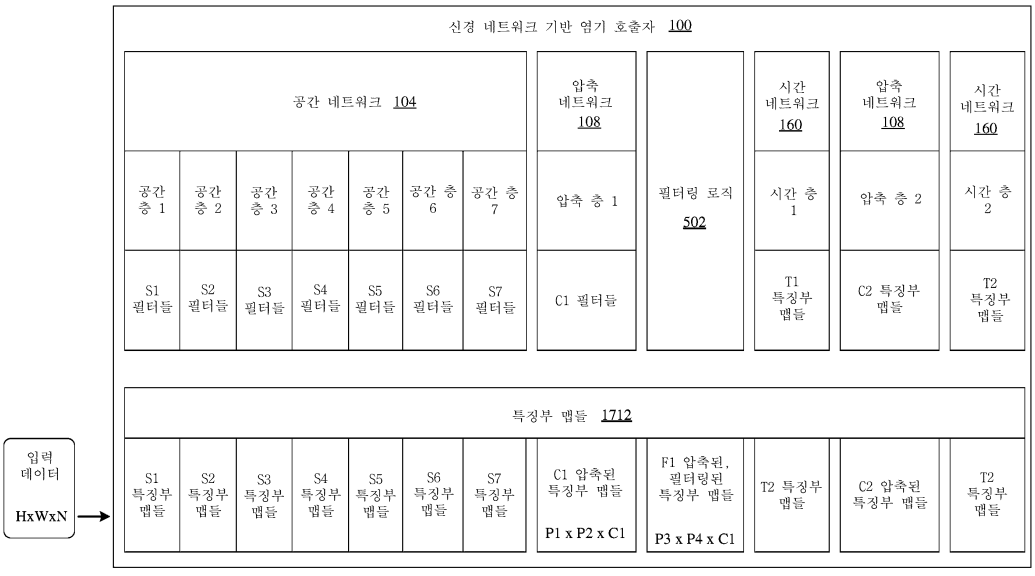
H = 64, 78, 115, 230
W = 64, 78, 115, 230
N = 1, 2, 3, 4

S1 = 7, 14, 21, 64, 128, 254
S2 = 7, 14, 21, 64, 128, 254
S3 = 7, 14, 21, 64, 128, 254
S4 = 7, 14, 21, 64, 128, 254
S5 = 7, 14, 21, 64, 128, 254
S6 = 7, 14, 21, 64, 128, 254
S7 = 7, 14, 21, 64, 128, 254

C1 = 1, 2, 3, 4, 이상
C2 = 1, 2, 3, 4, 이상

T1 = 7, 14, 21, 64, 128, 254
T2 = 1, 2, 3, 4

도면17



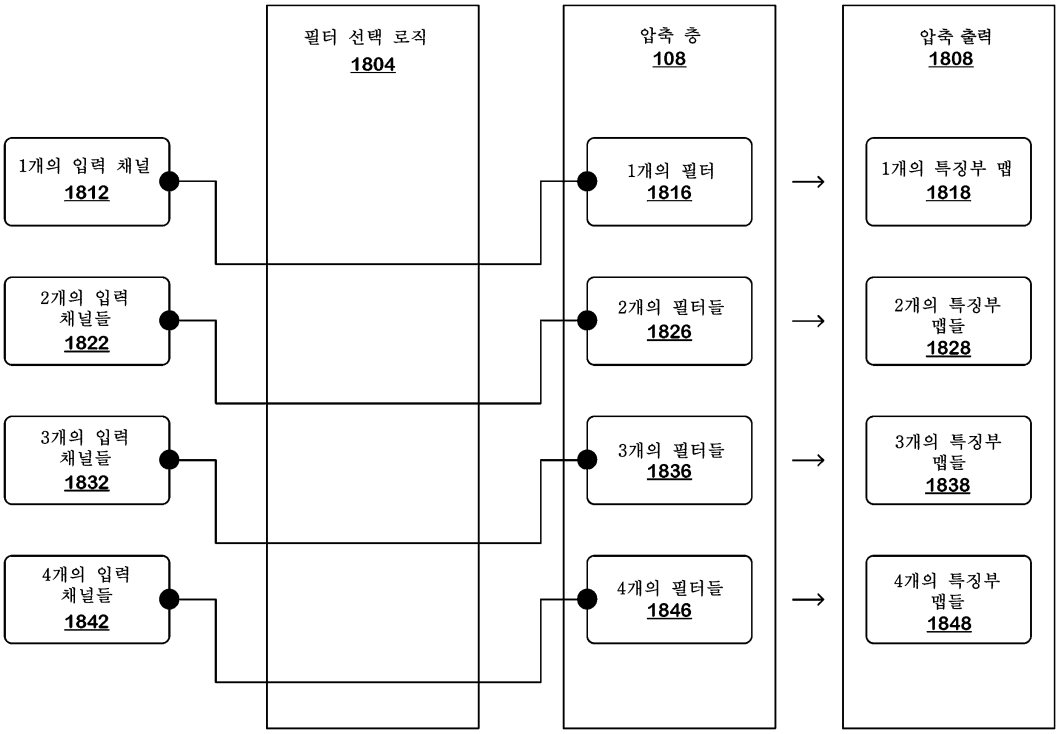
H = 64, 78, 115, 230
W = 64, 78, 115, 230
N = 1, 2, 3, 4

S1 = 7, 14, 21, 64, 128, 254
S2 = 7, 14, 21, 64, 128, 254
S3 = 7, 14, 21, 64, 128, 254
S4 = 7, 14, 21, 64, 128, 254
S5 = 7, 14, 21, 64, 128, 254
S6 = 7, 14, 21, 64, 128, 254
S7 = 7, 14, 21, 64, 128, 254

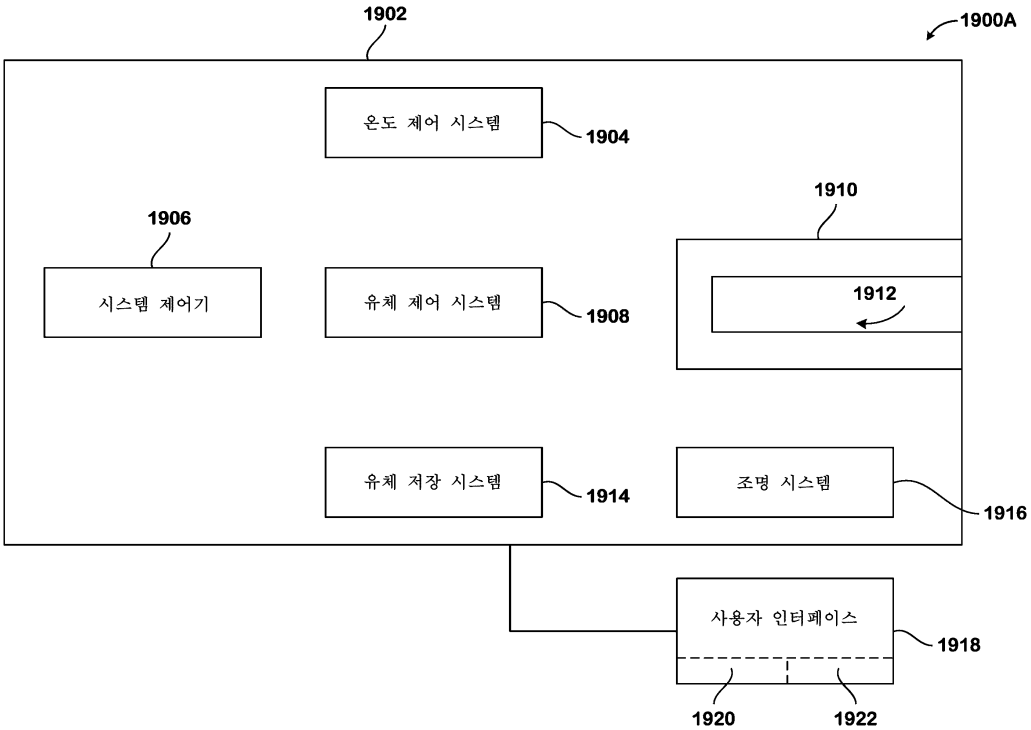
C1 = 1, 2, 3, 4, 이상
C2 = 1, 2, 3, 4, 이상
P3 = 25% 의 P1
P2 = 25% 의 P4

T1 = 7, 14, 21, 64, 128, 254
T2 = 1, 2, 3, 4

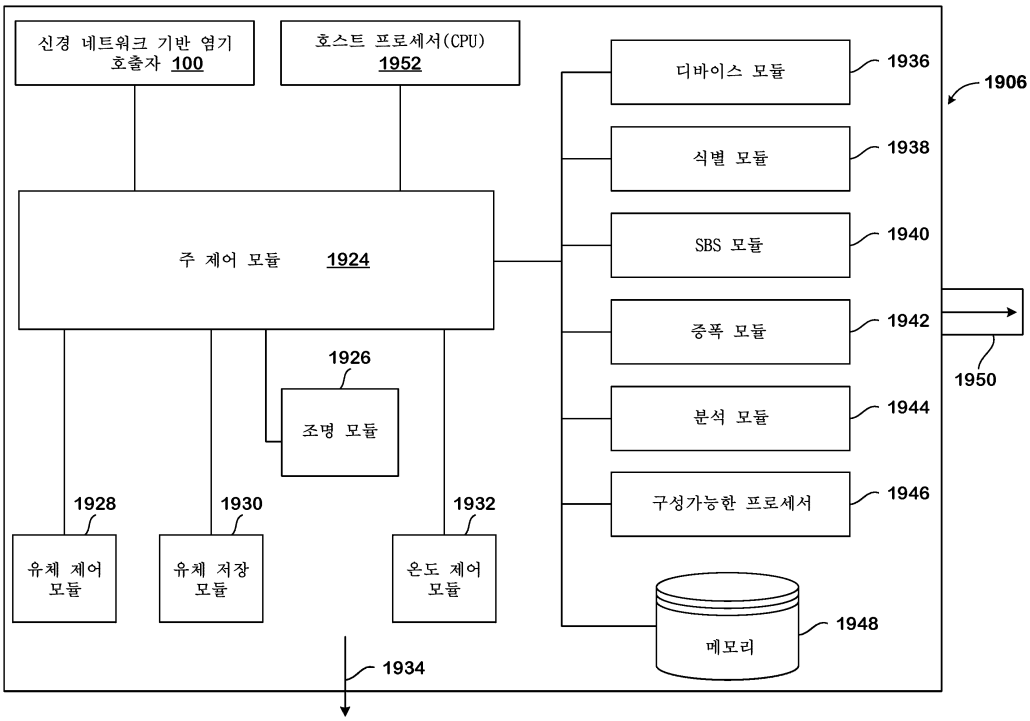
도면18


필터 선택 로직
1804
압축 층
108
압축 출력
1808
1개의 입력 채널
1812
2개의 입력 채널들
1822
3개의 입력 채널들
1832
4개의 입력 채널들
1842
1개의 필터
1816
2개의 필터들
1826
3개의 필터들
1836
4개의 필터들
1846
1개의 특징부 맵
1818
2개의 특징부 맵들
1828
3개의 특징부 맵들
1838
4개의 특징부 맵들
1848

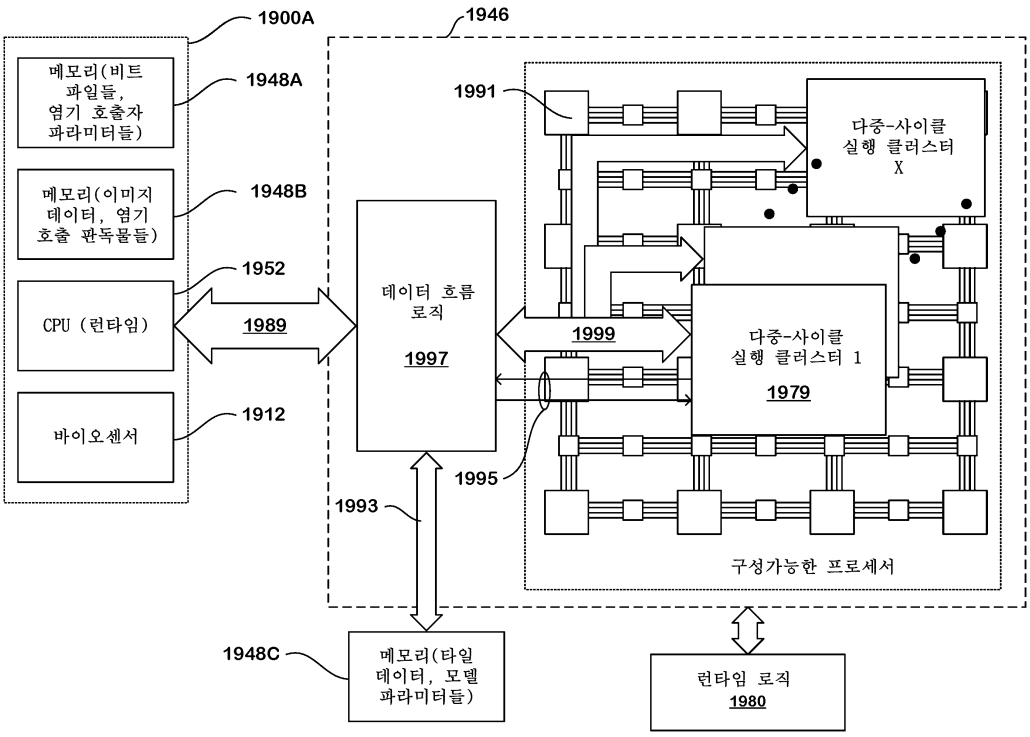
도면19a


1902
1900A
온도 제어 시스템
1904
1906
시스템 제어기
유체 제어 시스템
1908
1910
1912
유체 저장 시스템
1914
조명 시스템
1916
사용자 인터페이스
1918
1920
1922

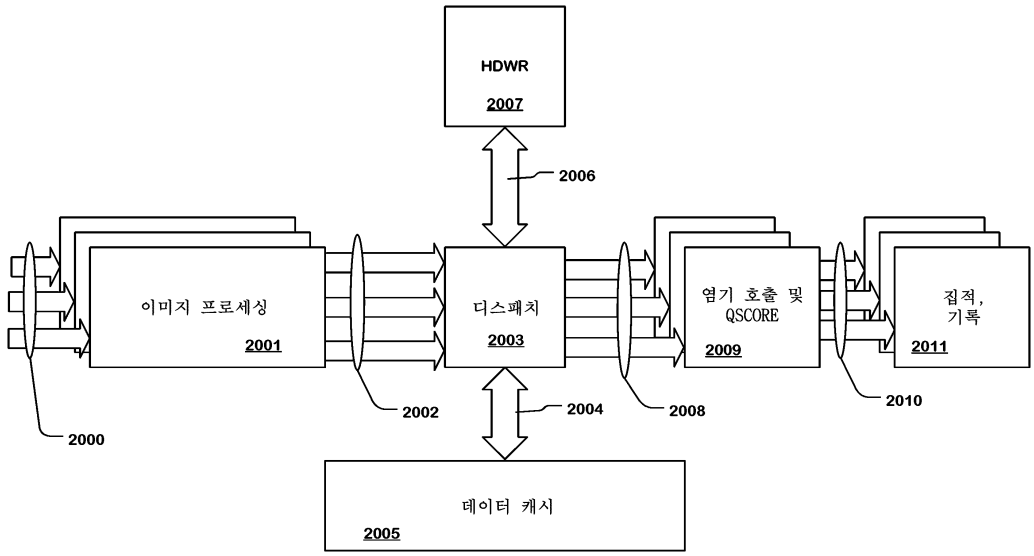
도면19b


신경 네트워크 기반 염기 호출자 100
호스트 프로세서(CPU) 1952
주 제어 모듈 1924
디바이스 모듈
1936
식별 모듈
1938
SBS 모듈
1940
증폭 모듈
1942
분석 모듈
1944
구성가능한 프로세서
1946
1906
1950
1926
조명 모듈
1928
유체 제어 모듈
1930
유체 저장 모듈
1932
온도 제어 모듈
메모리
1948
1934

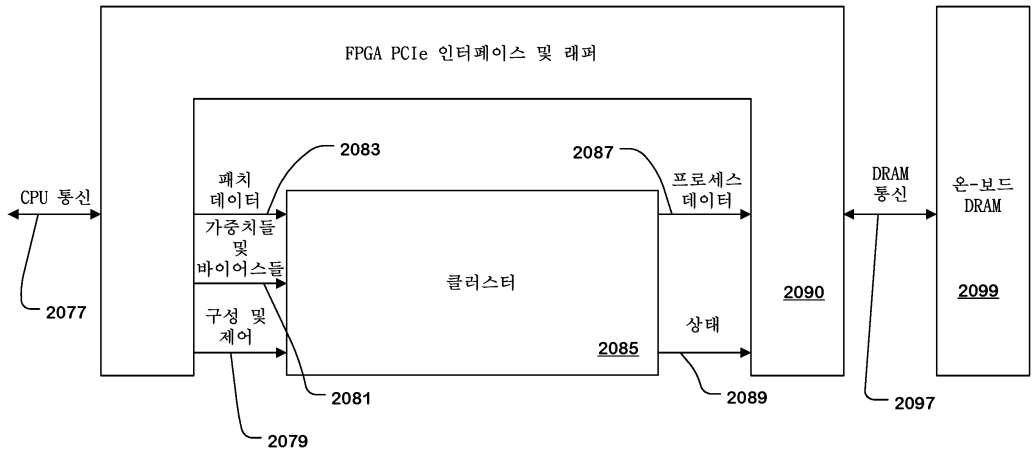
도면19c


1900A
메모리(비트 파일들, 염기 호출자 파라미터들)
1948A
메모리(이미지 데이터, 염기 호출 판독물들)
1948B
CPU (런타임)
1952
1989
바이오센서
1912
1946
1991
다중-사이클 실행 클러스터 X
데이터 흐름 로직
1997
1999
다중-사이클 실행 클러스터 1
1979
1995
1993
구성가능한 프로세서
1948C
메모리(타일 데이터, 모델 파라미터들)
런타임 로직
1980

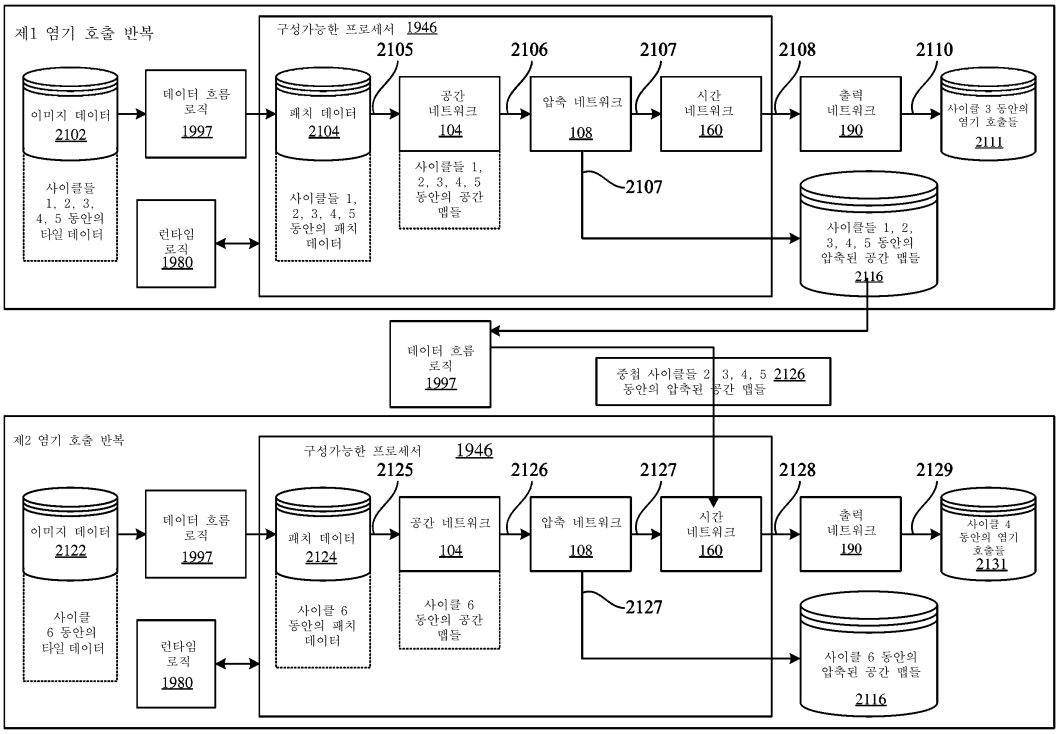
도면20a


HDWR
2007
2006
이미지 프로세싱
2001
디스패치
2003
염기 호출 및
QSCORE
2009
집적,
기록
2011
2000
2002
2004
2008
2010
데이터 캐시
2005

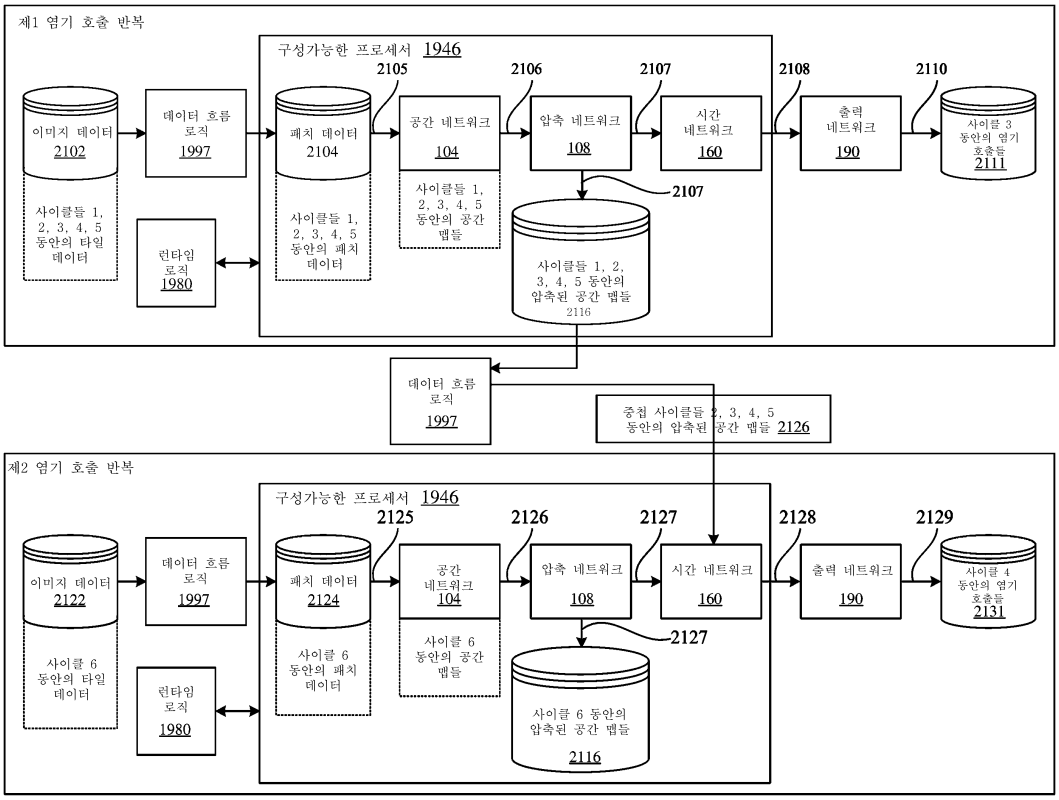
도면20b


FPGA PCIe 인터페이스 및 래퍼
CPU 통신
2077
패치
데이터
가중치들
및
바이어스들
구성 및
제어
2083
2087
클러스터
2085
프로세스
데이터
상태
2090
DRAM
통신
온-보드
DRAM
2099
2079
2081
2089
2097

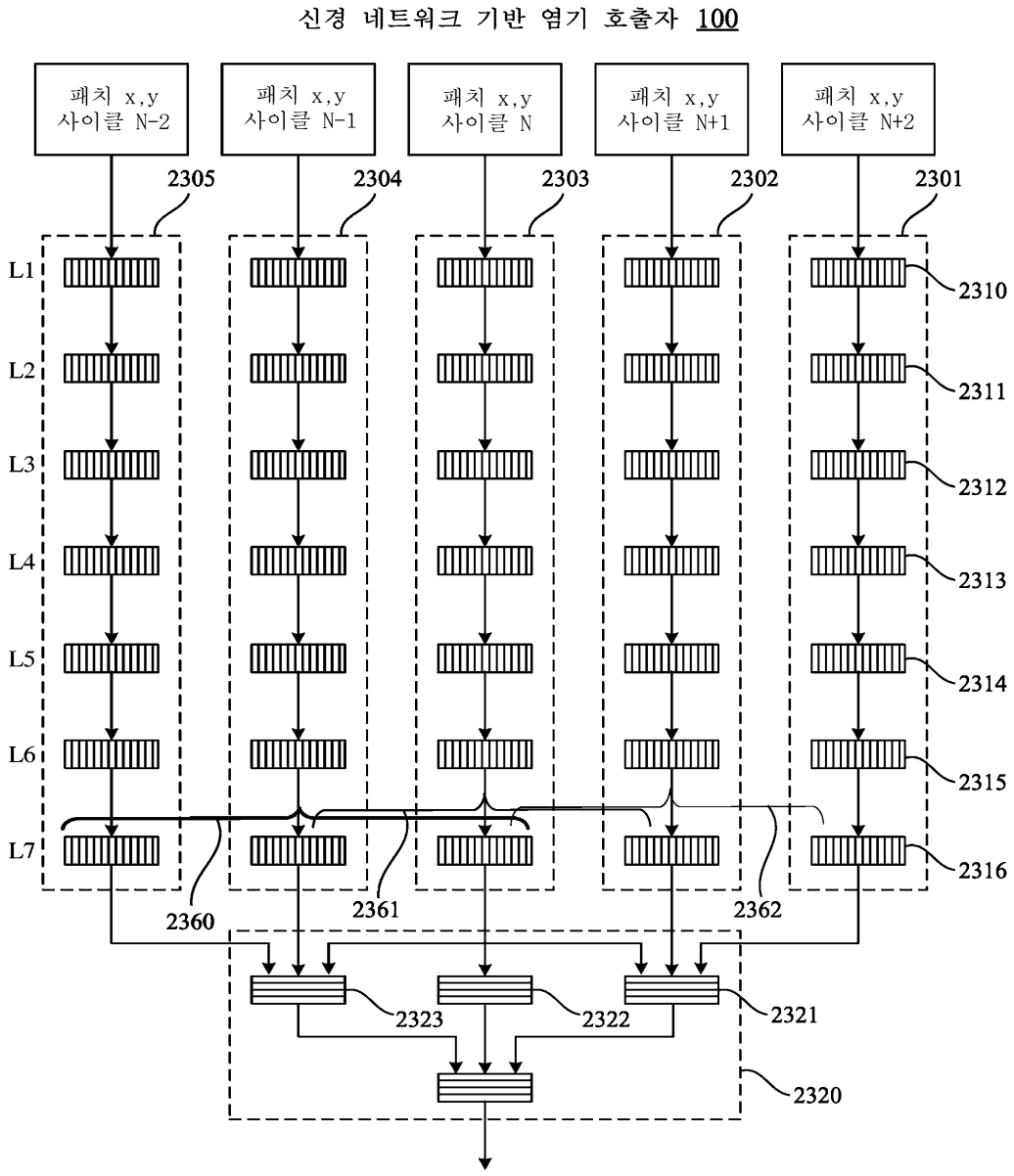
도면21


제1 염기 호출 반복
구성가능한 프로세서 1946
이미지 데이터 2102
데이터 흐름 로직 1997
패치 데이터 2104
공간 네트워크 104
압축 네트워크 108
시간 네트워크 160
출력 네트워크 190
사이클 3 동안의 염기 호출들 2111
사이클들 1, 2, 3, 4, 5 동안의 타일 데이터
런타임 로직 1980
사이클들 1, 2, 3, 4, 5 동안의 패치 데이터
사이클들 1, 2, 3, 4, 5 동안의 공간 맵들
사이클들 1, 2, 3, 4, 5 동안의 압축된 공간 맵들 2116
2105
2106
2107
2108
2110
데이터 흐름 로직 1997
중첩 사이클들 2, 3, 4, 5 동안의 압축된 공간 맵들 2126
제2 염기 호출 반복
구성가능한 프로세서 1946
이미지 데이터 2122
데이터 흐름 로직 1997
패치 데이터 2124
공간 네트워크 104
압축 네트워크 108
시간 네트워크 160
출력 네트워크 190
사이클 4 동안의 염기 호출들 2131
사이클 6 동안의 타일 데이터
런타임 로직 1980
사이클 6 동안의 패치 데이터
사이클 6 동안의 공간 맵들
사이클 6 동안의 압축된 공간 맵들 2116
2125
2126
2127
2128
2129

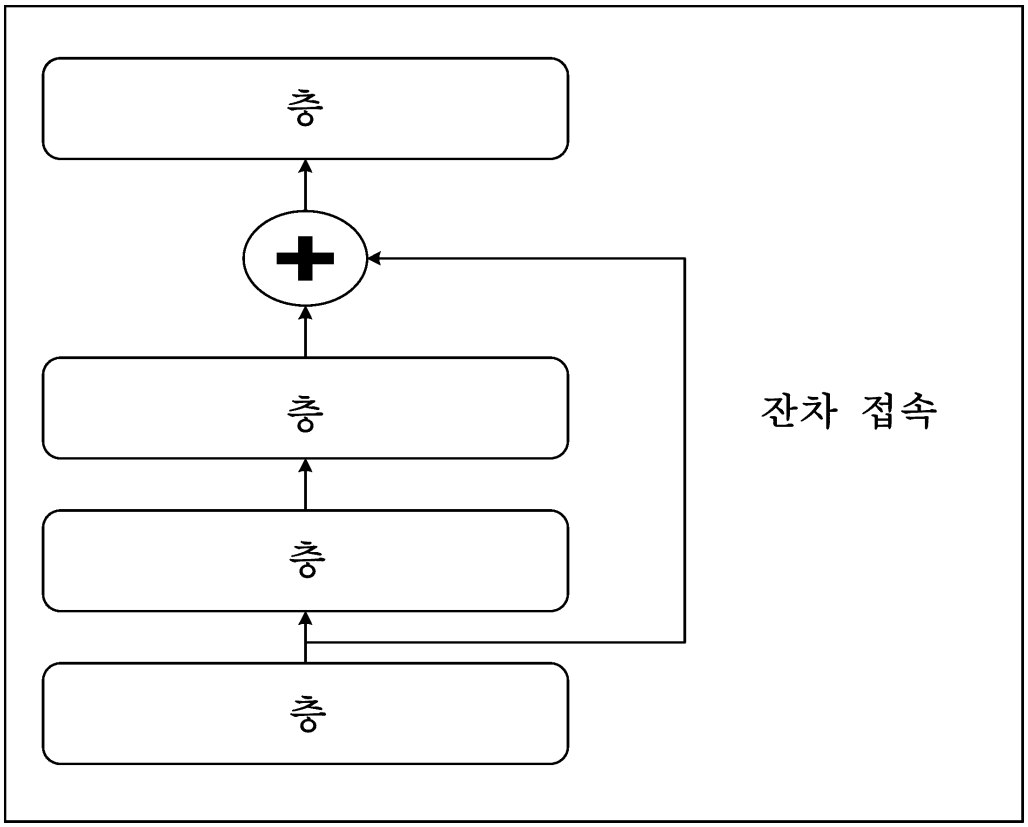
도면22


제1 염기 호출 반복
구성가능한 프로세서 1946
이미지 데이터 2102
데이터 흐름 로직 1997
패치 데이터 2104
공간 네트워크 104
압축 네트워크 108
시간 네트워크 160
출력 네트워크 190
사이클 3 동안의 염기 호출들 2111
사이클들 1, 2, 3, 4, 5 동안의 타일 데이터
런타임 로직 1980
사이클들 1, 2, 3, 4, 5 동안의 패치 데이터
사이클들 1, 2, 3, 4, 5 동안의 공간 맵들
사이클들 1, 2, 3, 4, 5 동안의 압축된 공간 맵들 2116
2105
2106
2107
2108
2110
데이터 흐름 로직 1997
중첩 사이클들 2, 3, 4, 5 동안의 압축된 공간 맵들 2126
제2 염기 호출 반복
구성가능한 프로세서 1946
이미지 데이터 2122
데이터 흐름 로직 1997
패치 데이터 2124
공간 네트워크 104
압축 네트워크 108
시간 네트워크 160
출력 네트워크 190
사이클 4 동안의 염기 호출들 2131
사이클 6 동안의 타일 데이터
런타임 로직 1980
사이클 6 동안의 패치 데이터
사이클 6 동안의 공간 맵들
사이클 6 동안의 압축된 공간 맵들 2116
2125
2126
2127
2128
2129

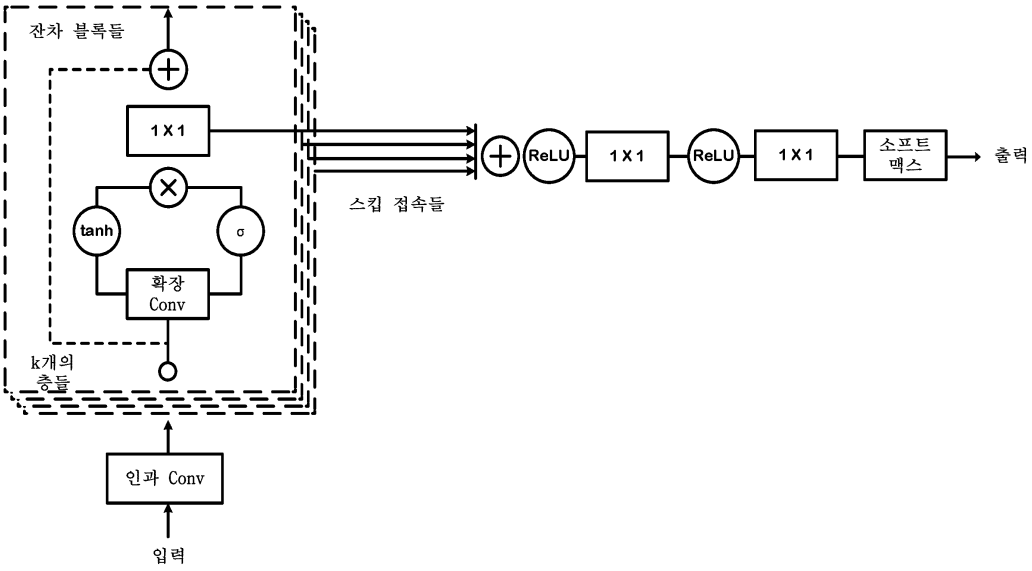
도면23


신경 네트워크 기반 염기 호출자 100
패치 x,y 사이클 N-2
패치 x,y 사이클 N-1
패치 x,y 사이클 N
패치 x,y 사이클 N+1
패치 x,y 사이클 N+2
2305
2304
2303
2302
2301
L1
L2
L3
L4
L5
L6
L7
2310
2311
2312
2313
2314
2315
2316
2360
2361
2362
2323
2322
2321
2320

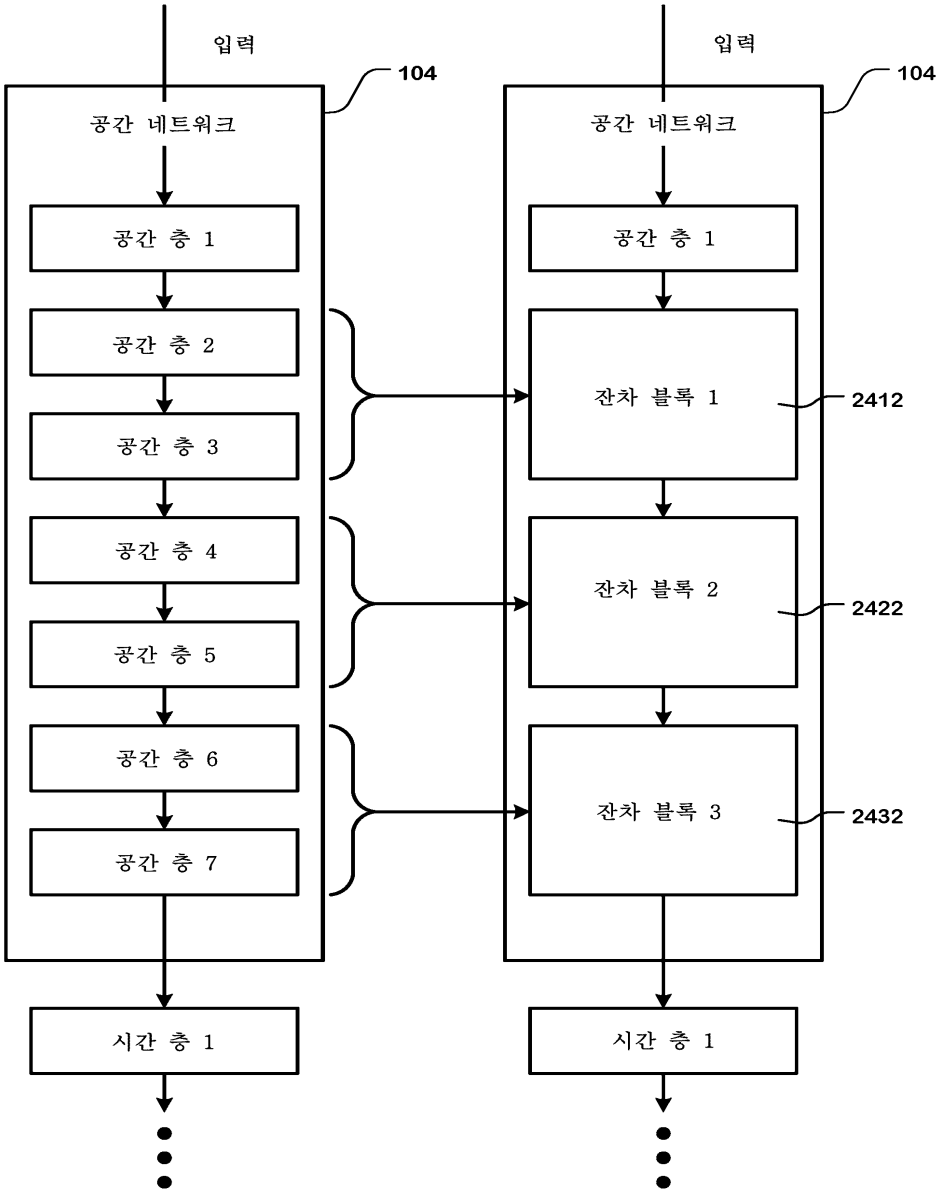
도면24a


층
층
층
층
잔차 접속

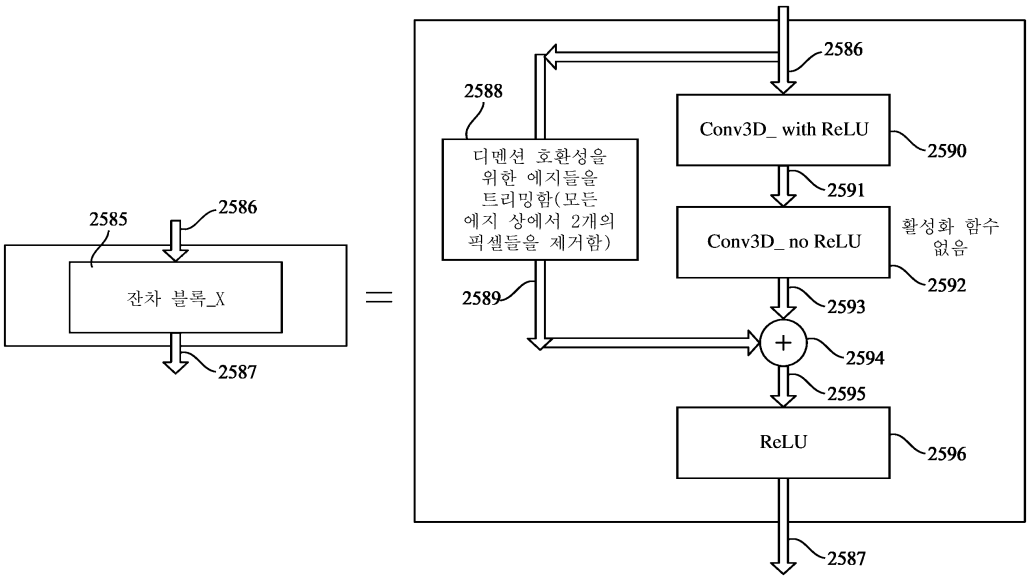
도면24b


잔차 블록들
1X1
tanh
σ
확장 Conv
k개의 층들
인과 Conv
입력
스킵 접속들
ReLU
1X1
ReLU
1X1
소프트 맥스
출력

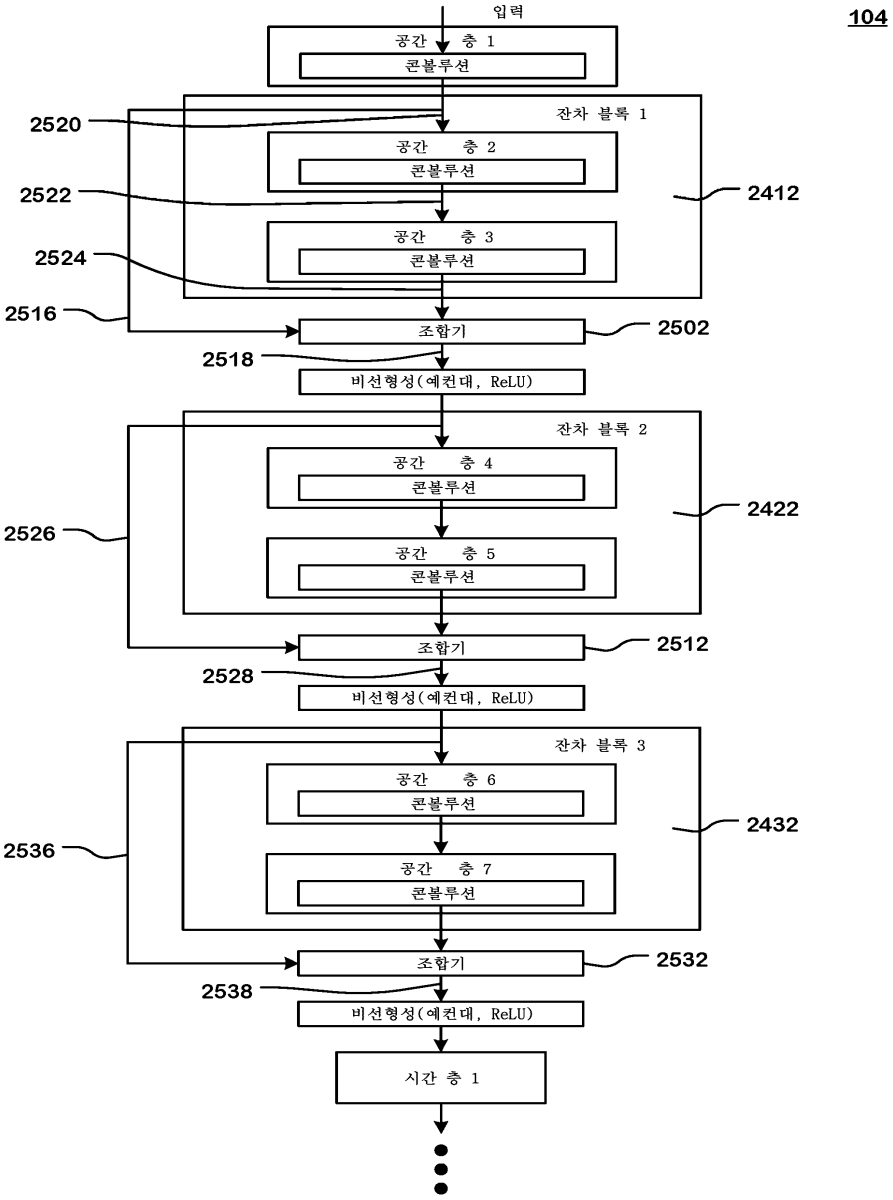
도면24c


입력
104
공간 네트워크
공간 층 1
공간 층 2
공간 층 3
공간 층 4
공간 층 5
공간 층 6
공간 층 7
시간 층 1
입력
104
공간 네트워크
공간 층 1
잔차 블록 1
2412
잔차 블록 2
2422
잔차 블록 3
2432
시간 층 1

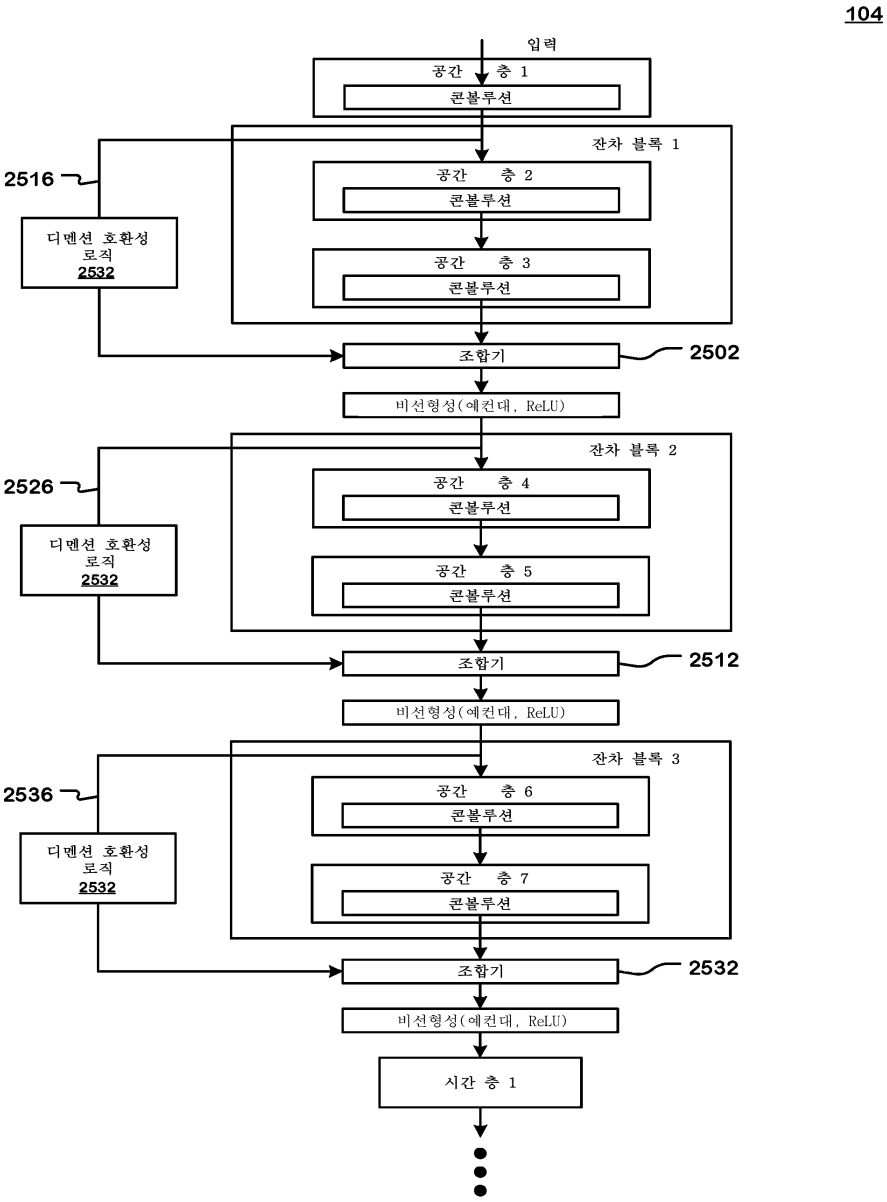
도면25a


2585
2586
잔차 블록_X
2587
=
2586
2588
디멘션 호환성을 위한 에지들을 트리밍함(모든 에지 상에서 2개의 픽셀들을 제거함)
Conv3D_ with ReLU
2590
2591
Conv3D_ no ReLU
활성화 함수 없음
2592
2589
2593
+
2594
2595
ReLU
2596
2587

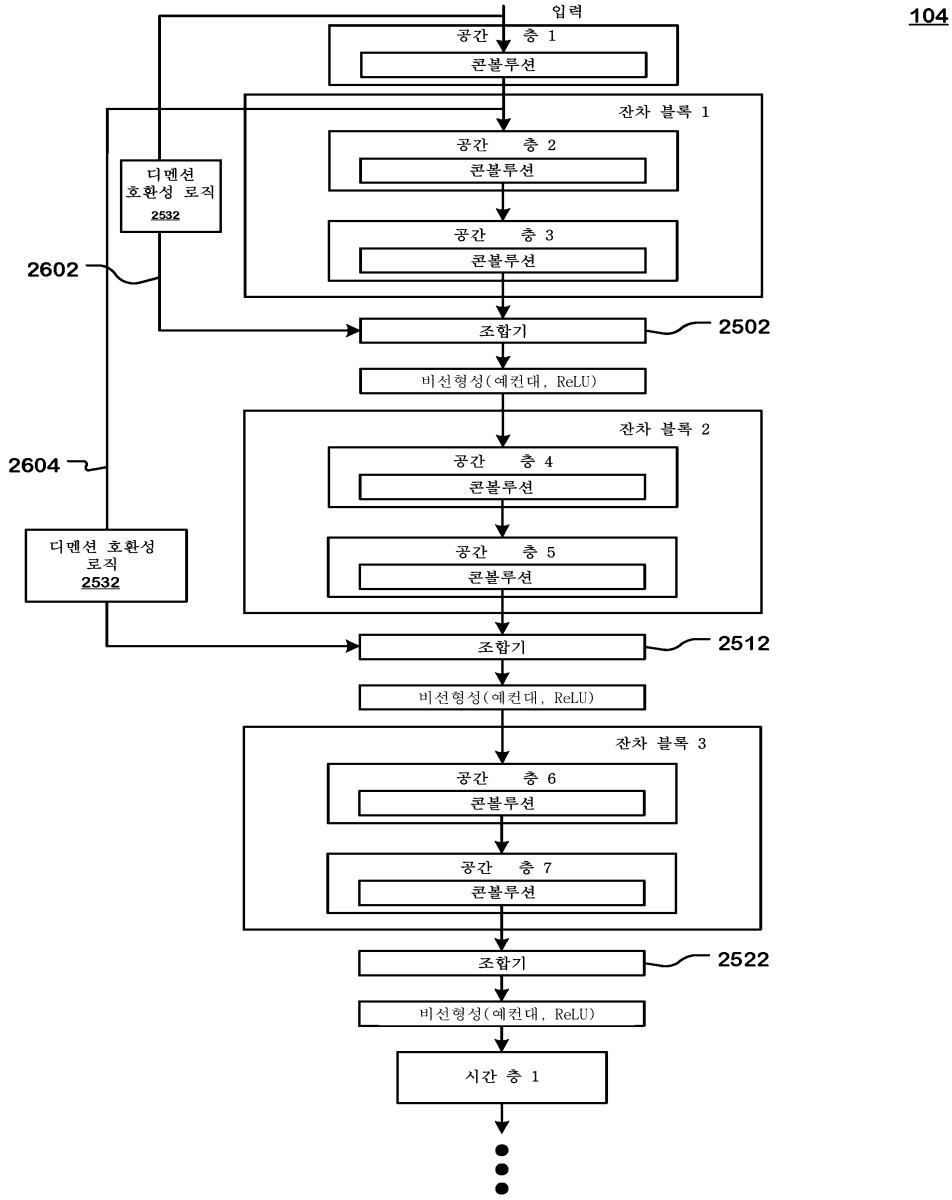
도면25b


104
입력
공간 층 1
콘볼루션
잔차 블록 1
2520
공간 층 2
콘볼루션
2522
2412
공간 층 3
콘볼루션
2524
2516
조합기
2502
2518
비선형성(예컨대, ReLU)
잔차 블록 2
공간 층 4
콘볼루션
2422
2526
공간 층 5
콘볼루션
조합기
2512
2528
비선형성(예컨대, ReLU)
잔차 블록 3
공간 층 6
콘볼루션
2432
2536
공간 층 7
콘볼루션
조합기
2532
2538
비선형성(예컨대, ReLU)
시간 층 1

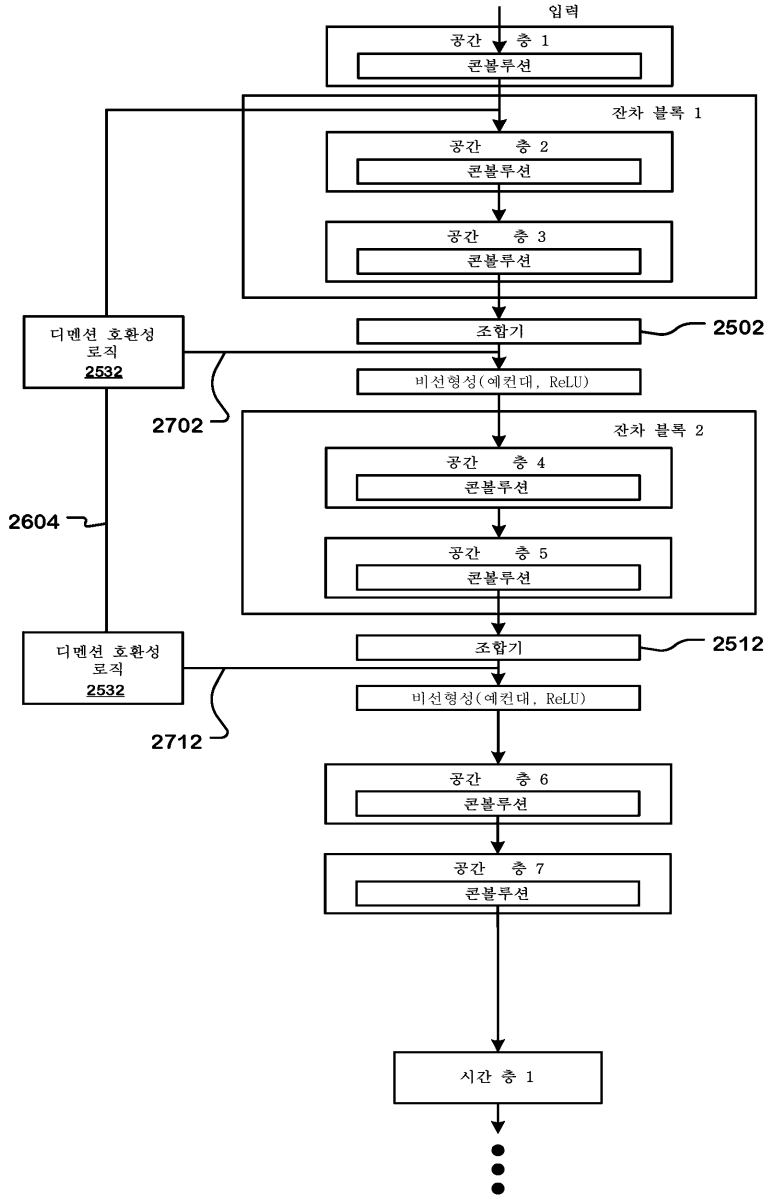
도면25c


104
입력
공간 층 1
콘볼루션
잔차 블록 1
2516
디멘션 호환성 로직
2532
공간 층 2
콘볼루션
공간 층 3
콘볼루션
조합기
2502
비선형성(예컨대, ReLU)
잔차 블록 2
2526
디멘션 호환성 로직
2532
공간 층 4
콘볼루션
공간 층 5
콘볼루션
조합기
2512
비선형성(예컨대, ReLU)
잔차 블록 3
2536
디멘션 호환성 로직
2532
공간 층 6
콘볼루션
공간 층 7
콘볼루션
조합기
2532
비선형성(예컨대, ReLU)
시간 층 1

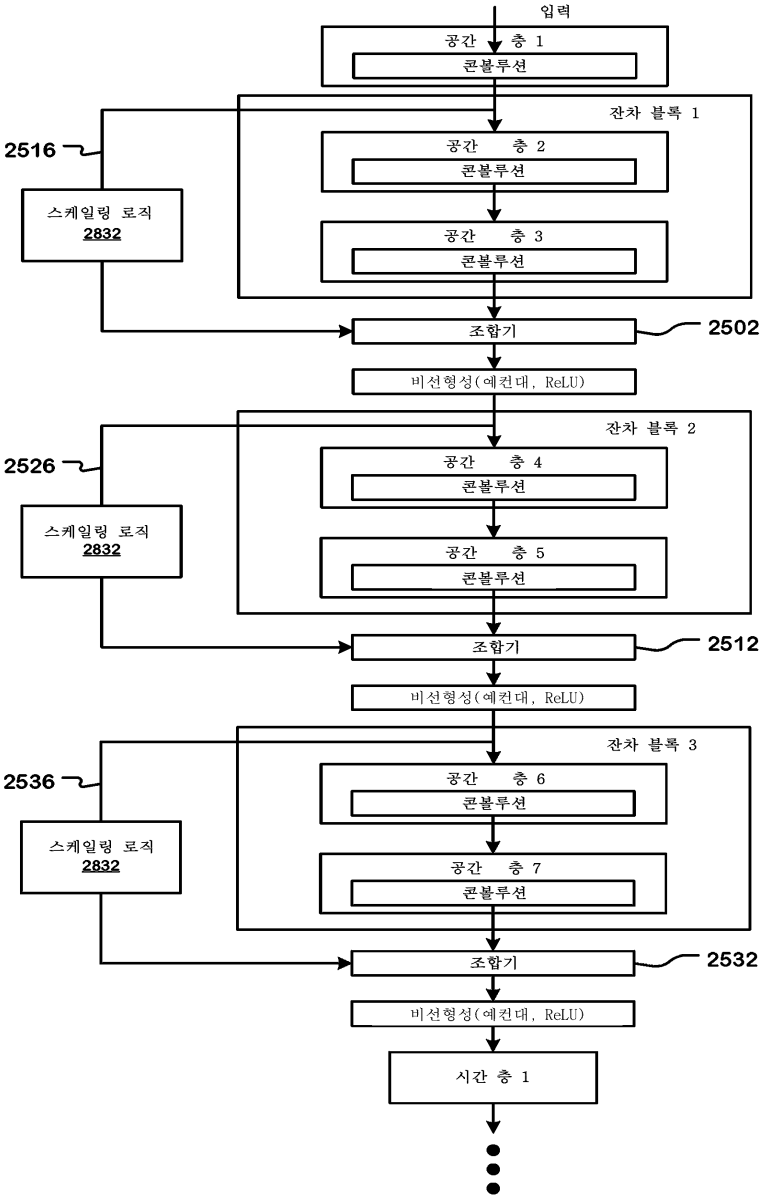
도면26


104
입력
공간 층 1
콘볼루션
잔차 블록 1
공간 층 2
콘볼루션
공간 층 3
콘볼루션
디멘션 호환성 로직
2532
2602
조합기
2502
비선형성(예컨대, ReLU)
잔차 블록 2
2604
공간 층 4
콘볼루션
공간 층 5
콘볼루션
디멘션 호환성 로직
2532
조합기
2512
비선형성(예컨대, ReLU)
잔차 블록 3
공간 층 6
콘볼루션
공간 층 7
콘볼루션
조합기
2522
비선형성(예컨대, ReLU)
시간 층 1

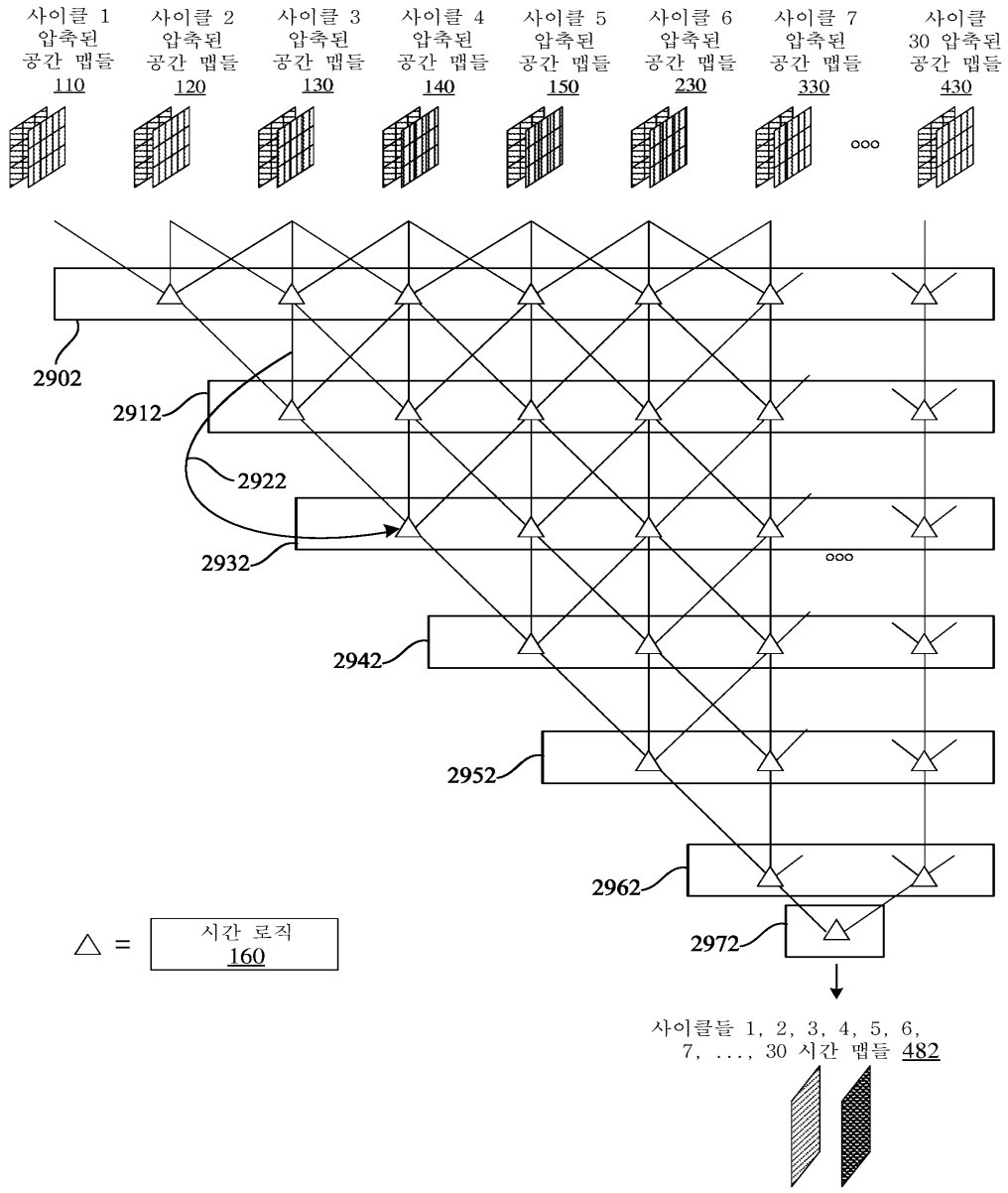
도면27


104
입력
공간 층 1
콘볼루션
잔차 블록 1
공간 층 2
콘볼루션
공간 층 3
콘볼루션
조합기
2502
디멘션 호환성 로직
2532
비선형성(예컨대, ReLU)
2702
잔차 블록 2
공간 층 4
콘볼루션
2604
공간 층 5
콘볼루션
조합기
2512
디멘션 호환성 로직
2532
비선형성(예컨대, ReLU)
2712
공간 층 6
콘볼루션
공간 층 7
콘볼루션
시간 층 1

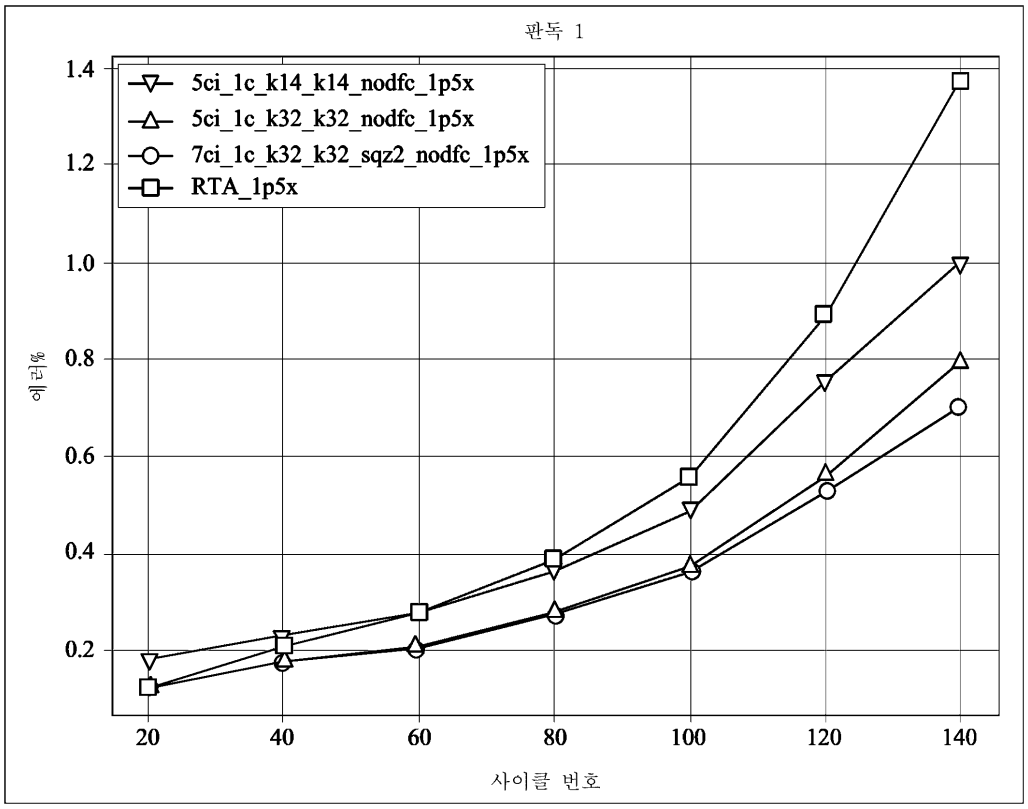
도면28


104
입력
공간 층 1
콘볼루션
잔차 블록 1
2516
공간 층 2
콘볼루션
스케일링 로직
2832
공간 층 3
콘볼루션
조합기
2502
비선형성(예컨대, ReLU)
잔차 블록 2
2526
공간 층 4
콘볼루션
스케일링 로직
2832
공간 층 5
콘볼루션
조합기
2512
비선형성(예컨대, ReLU)
잔차 블록 3
2536
공간 층 6
콘볼루션
스케일링 로직
2832
공간 층 7
콘볼루션
조합기
2532
비선형성(예컨대, ReLU)
시간 층 1

도면29


사이클 1 압축된 공간 맵들 110
사이클 2 압축된 공간 맵들 120
사이클 3 압축된 공간 맵들 130
사이클 4 압축된 공간 맵들 140
사이클 5 압축된 공간 맵들 150
사이클 6 압축된 공간 맵들 230
사이클 7 압축된 공간 맵들 330
사이클 30 압축된 공간 맵들 430
2902
2912
2922
2932
2942
2952
2962
2972
△ = 시간 로직 160
사이클들 1, 2, 3, 4, 5, 6, 7, ..., 30 시간 맵들 482

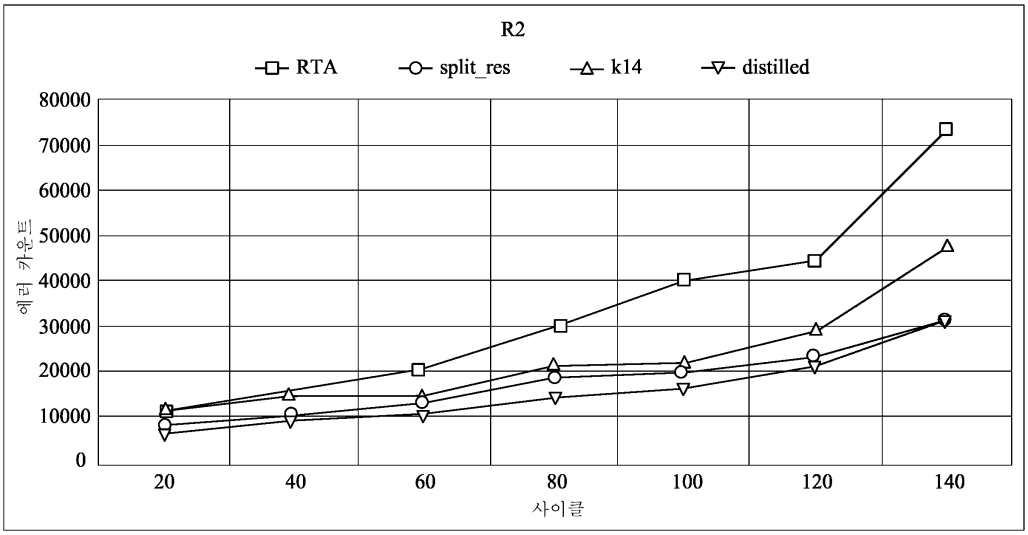
도면30



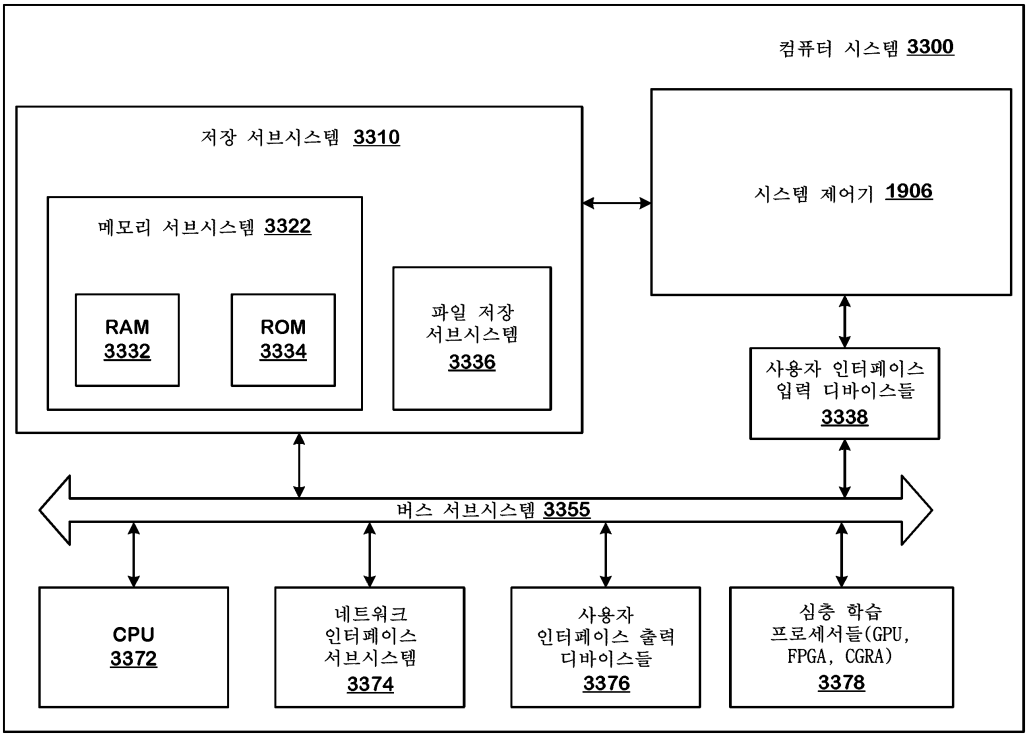
도면31

	RAM 사용량(GB)	DRAM 사용량(GB)
압축 로직 108 사용하지 않음	90.81344604	0.752478838
압축 로직 108 사용함	44.95265579	0.826221764

도면32



도면33


컴퓨터 시스템 3300
저장 서브시스템 3310
메모리 서브시스템 3322
RAM 3332
ROM 3334
파일 저장 서브시스템 3336
시스템 제어기 1906
사용자 인터페이스 입력 디바이스들 3338
버스 서브시스템 3355
CPU 3372
네트워크 인터페이스 서브시스템 3374
사용자 인터페이스 출력 디바이스들 3376
심층 학습 프로세서들(GPU, FPGA, CGRA) 3378