(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0026195
(43) 공개일자 2021년03월10일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2006.01)
(52) CPC특허분류
G06N 3/082 (2013.01)
G06N 3/08 (2013.01)
(21) 출원번호 10-2019-0106648
(22) 출원일자 2019년08월29일
심사청구일자 2019년08월29일

(71) 출원인
울산과학기술원
울산광역시 울주군 언양읍 유니스트길 50
(72) 발명자
이종은
울산광역시 울주군 언양읍 유니스트길 50
(74) 대리인
리앤목특허법인

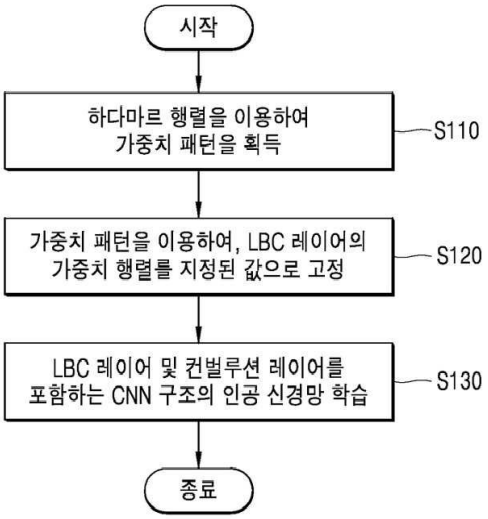
전체 청구항 수 : 총 13 항

(54) 발명의 명칭 **하드 와이어된 가중치 값을 갖는 인공 신경망 학습 방법 및 그 장치**

(57) 요 약

하드 와이어된 가중치 값을 갖는 인공 신경망 학습 방법 및 그 장치가 제공 된다. 본 발명의 일 실시예에 따른 LBC(Local Binary Convolution) 레이어를 포함하는 CNN 구조의 인공 신경망 학습 방법은 하다마르 행렬(Hadamard Matrix)을 이용하여 가중치 행렬을 획득하는 단계, 상기 가중치 행렬을 이용하여, 상기 LBC 레이어의 가중치를 지정된 값으로 하드 와이어(hard-wire)하는 단계 및 상기 LBC 레이어 및 가중치가 고정되지 않은 컨벌루션(Convolution) 레이어를 학습시키는 단계를 포함하고, 상기 컨벌루션 레이어는 상기 LBC 레이어의 출력 데이터를 이용하여 학습시키는 인공 신경망 학습 방법을 포함한다.

대 표 도 - 도2



이 발명을 지원한 국가연구개발사업

과제고유번호	1711093293
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	나노 소재 기술개발(R&D)
연구과제명	재구성로직 아키텍처 기술개발
기 여 율	1/1
과제수행기관명	울산과학기술원
연구기간	2019.02.01 ~ 2019.12.31

명 세 서

청구범위

청구항 1

컴퓨팅 장치에 의해 수행되는 LBC(Local Binary Convolution) 레이어를 포함하는 CNN 구조의 인공 신경망 학습 방법에 있어서,

하다마르 행렬(Hadamard Matrix)을 이용하여 가중치 패턴을 획득하는 단계;

상기 가중치 패턴을 이용하여, 상기 LBC 레이어의 가중치 행렬을 지정된 값으로 고정하는 단계; 및

상기 LBC 레이어의 출력 데이터를 이용하여 가중치가 고정되지 않은 컨벌루션(Convolution) 레이어를 학습시키는 단계;

를 포함하는 인공 신경망 학습 방법.

청구항 2

제1 항에 있어서,

상기 컨벌루션 레이어를 학습시키는 단계는,

1x1 크기의 컨벌루션 레이어를 학습시키는 단계를 포함하는,

인공 신경망 학습 방법.

청구항 3

제1 항에 있어서,

상기 하다마르 행렬을 이용하여 가중치 패턴을 획득하는 단계는,

모든 성분이 +1 또는 -1이며, 행벡터들과 열벡터들이 각각 서로 직교하는 정사각 행렬을 이용하여 상기 가중치 패턴을 획득하는 단계를 포함하는,

인공 신경망 학습 방법.

청구항 4

제1 항에 있어서,

상기 가중치 패턴을 이용하여, 상기 LBC 레이어의 가중치 행렬을 지정된 값으로 고정하는 단계는,

상기 가중치 패턴을 이용하여, LBC 레이어의 가중치 매개 변수에 대응되는 MAC(Multiply and Accumulate)어레이의 입력값 및 출력값을 지정된 값으로 고정하는 단계를 포함하는,

인공 신경망 학습 방법.

청구항 5

제1 항에 있어서,

상기 LBC 레이어의 가중치 행렬은,

상기 가중치 패턴이 반복된 형태인,

인공 신경망 학습 방법.

청구항 6

제1 항에 있어서,

상기 LBC 레이어의 가중치 행렬은,

상기 가중치 패턴이 부정(negation)된 형태인,

인공 신경망 학습 방법.

청구항 7

하다마르 행렬(Hadamard Matrix)을 이용하여 가중치 패턴을 획득하는 가중치 패턴 획득부;

상기 가중치 패턴을 이용하여, 상기 LBC 레이어의 가중치 행렬을 지정된 값으로 고정하는 가중치 고정부; 및

상기 LBC 레이어의 출력 데이터를 이용하여 가중치가 고정되지 않은 컨벌루션(Convolution) 레이어를 학습시키는 신경망 학습부;

를 포함하는 LBC(Local Binary Convolution) 레이어를 포함하는 CNN 구조의 인공 신경망 학습 장치.

청구항 8

제7 항에 있어서,

상기 신경망 학습부는,

1x1 크기의 컨벌루션 레이어를 학습시키는,

인공 신경망 학습 장치.

청구항 9

제7 항에 있어서,

상기 획득부는,

모든 성분이 +1 또는 -1이며, 행벡터들과 열벡터들이 각각 서로 직교하는 정사각 행렬을 이용하여 상기 가중치 패턴을 획득하는,

인공 신경망 학습 장치.

청구항 10

제7 항에 있어서,

가중치 고정부는,

상기 가중치 패턴을 이용하여, 상기 LBC 레이어의 가중치 매개 변수에 대응되는 MAC(Multiply and Accumulate) 어레이의 입력값 및 출력값을 지정된 값으로 고정하는,

인공 신경망 학습 장치.

청구항 11

제7 항에 있어서,

상기 LBC 레이어의 가중치 행렬은,

상기 가중치 패턴이 반복된 형태의 행렬을 기초로 획득된 행렬인,

인공 신경망 학습 장치.

청구항 12

제7 항에 있어서,

상기 LBC 레이어의 가중치 행렬은,

상기 가중치 패턴이 부정(negation)된 형태의 행렬을 기초로 획득된 행렬인,

인공 신경망 학습 장치.

청구항 13

제1 항 내지 제6 항 중 어느 한 항에 따른 방법을 수행하기 위한 프로그램이 기록된 컴퓨터 판독 가능한 기록매체.

발명의 설명

기 술 분 야

[0001] 본 발명은 하드 와이어(hard wire)된 가중치 값을 갖는 인공 신경망 학습 방법 및 그 장치 에 관한 것이다. 보다 자세하게는, 가중치 값이 하드 와이어 된 인공 신경망 하드웨어를 이용한 인공 신경망 학습 방법 및 그 장치에 관한 것이다.

배 경 기 술

[0002] 인공 신경망 프로세서의 가중치 매개 변수가 수정 가능하도록 설계된 경우, 신경망의 확장성과 적용성은 높아지나 관련 계산의 로직이 복잡해지므로 많은 양의 컴퓨팅 자원이 필요하다. 이에 반해, 인공 신경망 프로세서의 가중치 매개 변수를 고정하는 경우, 네트워크의 성능은 향상되나 신경망 하드웨어의 적용 범위가 매우 제한되어 문제가 된다.

발명의 내용

해결하려는 과제

[0003] 본 발명이 해결하고자 하는 기술적 과제는, 고정된 가중치를 갖는 인공 신경망 하드웨어를 이용하여 인공 신경망 학습을 수행하는 방법 및 그 장치에 대한 것이다.

[0004] 본 발명의 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 아래의 기재로부터 본 발명의 기술분야에서의 통상의 기술자에게 명확하게 이해 될 수 있을 것이다.

과제의 해결 수단

[0005] 상기 기술적 과제를 해결하기 위한, 본 발명의 일 실시예에 따른 LBC(Local Binary Convolution) 레이어를 포함하는 CNN 구조의 인공 신경망 학습 방법은 하다마르 행렬(Hadamard Matrix)을 이용하여 가중치 패턴을 획득하는 단계, 상기 가중치 패턴을 이용하여, 상기 LBC 레이어의 가중치 행렬을 지정된 값으로 고정하는 단계 및 상기 LBC 레이어 출력 데이터를 이용하여 가중치가 고정되지 않은 컨벌루션(Convolution) 레이어를 학습시키는 단계를 포함할 수 있다.

[0006] 일 실시예에 따른 상기 컨벌루션 레이어를 학습시키는 단계는, 1x1 크기의 컨벌루션 레이어를 학습시키는 단계를 포함할 수 있다.

[0007] 일 실시예에 따른 상기 하다마르 행렬을 이용하여 가중치 패턴을 획득하는 단계는, 모든 성분이 +1 또는 -1이며, 행벡터들과 열벡터들이 각각 서로 직교하는 정사각 행렬을 이용하여 상기 가중치 패턴을 획득하는 단계를 포함할 수 있다.

[0008] 일 실시예에 따른 상기 가중치 패턴을 이용하여, 상기 LBC 레이어의 가중치 행렬을 지정된 값으로 고정하는 단계는, 상기 가중치 패턴을 이용하여, LBC 레이어의 가중치 매개 변수에 대응되는 MAC(Multiply and Accumulate)어레이의 입력값 및 출력값을 지정된 값으로 고정하는 단계를 포함할 수 있다.

[0009] 일 실시예에 따른 상기 LBC 레이어의 가중치 행렬은, 상기 가중치 패턴이 반복된 형태일 수 있다.

[0010] 일 실시예에 따른 상기 LBC 레이어의 가중치 행렬은, 상기 가중치 패턴이 부정(negation)된 형태일 수 있다.

[0011] 상기 기술적 과제를 해결하기 위한, 본 발명의 다른 실시예에 따른 LBC(Local Binary Convolution) 레이어를 포함하는 CNN 구조의 인공 신경망 학습 장치는 하다마르 행렬(Hadamard Matrix)을 이용하여 가중치 패턴을 획득하는 가중치 패턴 획득부, 상기 가중치 패턴을 이용하여, 상기 LBC 레이어의 가중치 행렬을 지정된 값으로 고정하는 가중치 고정부 및 상기 LBC 레이어의 출력 데이터를 이용하여 가중치가 고정되지 않은 컨벌루션

(Convolution) 레이어를 학습시키는 신경망 학습부를 포함할 수 있다.

[0012] 일 실시예에 따른 상기 신경망 학습부는, 1x1 크기의 컨벌루션 레이어를 학습시킬 수 있다.

[0013] 일 실시예에 따른 상기 획득부는, 상기 모든 성분이 +1 또는 -1이며, 행벡터들과 열벡터들이 각각 서로 직교하는 정사각 행렬을 이용하여 상기 가중치 패턴을 획득할 수 있다.

[0014] 일 실시예에 따른 가중치 고정부는, 상기 가중치 패턴을 이용하여, 상기 LBC 레이어의 가중치 매개 변수에 대응되는 MAC(Multiply and Accumulate)어레이의 입력값 및 출력값을 지정된 값으로 고정할 수 있다.

[0015] 일 실시예에 따른 상기 LBC 레이어의 가중치 행렬은, 상기 가중치 패턴이 반복된 형태의 행렬을 기초로 획득된 행렬일 수 있다.

[0016] 일 실시예에 따른 상기 LBC 레이어의 가중치 행렬은, 상기 가중치 패턴이 부정(negation)된 형태의 행렬을 기초로 획득된 행렬일 수 있다.

[0017] 상기 기술적 과제를 해결하기 위한, 본 발명의 또 다른 실시예는 제1 항 내지 제6 항 중 어느 한 항에 따른 방법을 수행하기 위한 프로그램이 기록된 컴퓨터 판독 가능한 기록매체일 수 있다.

도면의 간단한 설명

[0018] 도 1은 본 발명의 일 실시예에 따른 인공 신경망 학습 방법을 설명하기 위한 개념도이다.

도 2는 본 발명의 일 실시예에 따른 인공 신경망 학습 방법의 순서도이다.

도 3은 본 발명의 일 실시예에 따른 인공 신경망 학습 방법에서, 가중치 패턴의 적용예를 설명하기 위한 개념도이다.

도 4는 본 발명의 일 실시예에 따른 인공 신경망 학습 방법에서, 가중치 패턴의 적용예를 설명하기 위한 개념도이다.

도 5는 본 발명의 일 실시예에 따른 인공 신경망 학습 방법에서, 가중치 패턴의 적용예를 설명하기 위한 개념도이다.

도 6은 도 2의 일부 동작을 상세히 설명하기 위한 순서도이다.

도 7은 본 발명의 일 실시예에 따른 인공 신경망 학습 장치를 설명하기 위한 하드웨어 블록도이다.

도 8는 본 발명의 일 실시예에 따른 인공 신경망 학습 방법을 수행하는 프로그램이 기록된 기록매체를 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0019] 후술하는 본 발명에 대한 상세한 설명은, 본 발명이 실시될 수 있는 특정 실시예를 예시로서 도시하는 첨부 도면을 참조한다. 이러한 실시예는 당업자가 본 발명을 실시할 수 있기에 충분하도록 상세히 설명된다. 본 발명의 다양한 실시예는 서로 다르지만 상호 배타적일 필요는 없음이 이해되어야 한다. 예를 들어, 본 명세서에 기재되어 있는 특정 형상, 구조 및 특성은 본 발명의 정신과 범위를 벗어나지 않으면서 일 실시예로부터 다른 실시예로 변경되어 구현될 수 있다. 또한, 각각의 실시예 내의 개별 구성요소의 위치 또는 배치도 본 발명의 정신과 범위를 벗어나지 않으면서 변경될 수 있음이 이해되어야 한다. 따라서, 후술하는 상세한 설명은 한정적인 의미로서 행하여지는 것이 아니며, 본 발명의 범위는 특허청구범위의 청구항들이 청구하는 범위 및 그와 균등한 모든 범위를 포괄하는 것으로 받아들여져야 한다. 도면에서 유사한 참조부호는 여러 측면에 걸쳐서 동일하거나 유사한 구성요소를 나타낸다.

[0020] 이하에서는, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 본 발명을 용이하게 실시할 수 있도록 하기 위하여, 본 발명의 여러 실시예에 관하여 첨부된 도면을 참조하여 상세히 설명하기로 한다.

[0021] 도 1을 참조하여 본 발명의 일 실시예에 따른 인공 신경망 학습 방법에 대해 설명한다. 본 발명의 몇몇 실시예에 따른 학습 대상 인공 신경망(20)은 LBC(Local Binary Convolution) 레이어(21)와 적어도 하나의 컨벌루션(Convolution) 레이어(22)를 포함하는 CNN(Convolutional Neural Network) 구조의 신경망으로, 바람직하게는 LBCNN(Local Binary Convolutional Neural Network) 구조의 신경망일 수 있다.

[0022] 일 실시예에 따른 LBC 레이어(21)의 가중치 매개 변수는 학습이 수행되지 않는 고정값 일 수 있다. 즉, LBC 레

이어(21)의 가중치 매개 변수는 지정된 값으로 하드 와이어(hard wire)될 수 있다.

[0023] 가중치가 하드 와이어 된다는 것은, 신경망에 포함된 레이어의 가중치 매개 변수에 대응되는 MAC(Multiply and Accumulate) 어레이의 입력값 및 출력값이 지정된 값으로 고정되는 것이다. 신경망의 가중치 매개 변수가 하드 와이어 되는 경우, 관련 로직을 단순화할 수 있으므로 DNN(Deep Neural Network) 하드웨어의 효율성이 향상될 수 있다.

[0024] 또한, 일 실시예에 따른 컨벌루션 레이어(22)의 가중치 매개 변수는 학습 가능한 가변값일 수 있다. 컨벌루션 레이어(22)의 가중치 매개 변수가 고정되어 있지 않은 경우, 신경망의 확장성과 프로그래밍성이 높아지고 신경망의 정확도를 보장할 수 있으며, DNN 하드웨어의 적용 가능성이 매우 높아진다.

[0025] 일 실시예에 따른 입력 데이터(10)는 방대한 양의 데이터일 수 있다. 따라서, 방대한 양의 데이터를 이용하여 학습되는 인공 신경망(20)의 LBC 레이어(21)의 가중치 행렬의 크기는 매우 클 수 있다. 가중치 행렬의 크기가 큰 경우 특징 맵(feature map)을 추출하는 연산을 수행하기 위해 많은 양의 컴퓨팅 자원이 필요하다.

[0026] 따라서, 본 발명의 몇몇 실시예에 따른 LBC 레이어(21)의 가중치 행렬은 하다마르(Hadamard) 행렬을 이용하여 획득될 수 있다. 이 경우, 데이터 저장 공간 및 연산 시간을 절약할 수 있다. 상세한 설명은 이하 관련된 도면에서 후술한다.

[0027] 일 실시예에 따른 인공 신경망(20)은 입력 데이터(10)를 입력 받아 LBC 레이어(21)에 입력할 수 있다.

[0028] 또한, 상기 LBC 레이어(21)에서 출력된 데이터를 이용하여 컨벌루션 레이어(22)를 학습시킬 수 있다. 상기 컨벌루션 레이어(22)는 복수개의 컨벌루션 레이어로 구성될 수 있고, 바람직하게는 1x1 크기의 단일 컨벌루션 레이어일 수도 있다.

[0029] 최종적으로 출력되는 출력 데이터(30)는 컨벌루션 레이어의 출력 값일 수 있다. 다만, 전술한 인공 신경망의 구조는 본 발명을 설명하기 위해 최대한 간결하게 구성한 일 예시이며, 통상의 기술자가 변형할 수 있는 범위의 구조를 갖는 인공 신경망에 본 발명의 몇몇 실시예에 따른 인공 신경망 학습 방법이 모두 적용될 수 있음은 자명하다.

[0030] 이하, 도 2를 참조하여 본 발명의 일 실시예에 따른 인공 신경망 학습 방법에 대하여 상세히 설명한다.

[0031] 단계 S110에서, 하다마르 행렬을 이용하여 가중치 패턴이 획득될 수 있다. 하다마르 행렬은 모든 성분이 +1 또는 -1이며, 행벡터들과 열벡터들이 각각 서로 직교하는 정사각 행렬이다. 즉, 하다마르 행렬 각각을 원소로 하는 행렬 및 하다마르 행렬의 부정 형태의 행렬도 여전히 하다마르 행렬일 수 있다.

[0032] 따라서, 몇몇 실시예에 따라 상기 하다마르 행렬을 이용하여 가중치 패턴이 획득되는 경우, 상기 가중치 패턴에 대한 저장 공간만 확보하면, 이후 임의 크기의 가중치 행렬을 모두 생성할 수 있다.

[0033] 이에 반해, 종래에는 무작위 값으로 구성된 가중치 행렬을 사용함으로써, 가중치 행렬의 모든 서브 행렬(sub matrix)의 값이 모두 상이하였고, 각각의 서브 행렬 모두를 저장하기 위한 저장공간 확보가 필요하였다.

[0034] 단계 S120에서, 상기 획득된 가중치 패턴을 이용하여, LBC 레이어의 가중치 행렬이 지정된 값으로 고정될 수 있다. 즉, MAC(Multiply and Accumulate) 어레이의 입력값 및 출력값을 조정하여 LBC 레이어의 가중치 매개 변수가 지정된 값으로 하드 와이어로 연결될 수 있고, 바람직하게는 LBC 레이어의 모든 가중치가 하드 와이어로 연결될 수 있다.

[0035] 일 실시예에 따른 가중치 행렬은 상기 획득된 가중치 패턴이 반복된 형태의 행렬일 수 있고, 상기 가중치 패턴이 부정된 형태의 행렬일 수 있으며, 상기 가중치 패턴에 대하여 희소성이 크도록 수정된 형태의 행렬일 수도 있다.

[0036] 따라서, 하드웨어 비용을 줄이기 위하여 행렬 $H_1 = \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix}$ 및 $H_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ 보다는 $H_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ 을 이용하는 것이 더 바람직할 수 있다.

[0037] 단계 S130에서, LBC 레이어 및 컨벌루션 레이어를 포함하는 CNN 구조의 인공 신경망 학습이 수행될 수 있다. 상기 인공 신경망은 바람직하게 LBCNN일 수 있고, 본 발명의 몇몇 실시예에 따른 LBCNN은 적어도 하나의 LBC 레이

어와 컨벌루션 레이어를 포함하며, 상기 컨벌루션 레이어는 1x1 크기일 수 있다. 또한 최초 학습 데이터는 LBC 레이어의 입력데이터일 수 있고, 상기 LBC 레이어의 출력 데이터를 이용하여 컨벌루션 레이어가 학습될 수 있다.

[0038] 본 발명의 몇몇 실시예에 따라 LBC 레이어의 가중치 매개 변수가 하다마르 행렬을 이용하여 획득된 데이터 값으로 고정되는 경우, LBC 레이어의 학습은 수행될 필요가 없고, 컨벌루션 레이어의 학습만 수행될 수 있다.

[0039] 즉, 인공 신경망 중 LBC 레이어의 가중치 매개 변수만 고정하고, 컨벌루션 레이어의 가중치 매개변수는 고정하지 않음으로써, 신경망의 확장성과 로직의 단순화의 이점을 모두 확보할 수 있다.

[0040] 이하 도 3 내지 도 5를 참조하여 LBC 레이어의 가중치 행렬 및 MAC 어레이의 크기에 따른 하다마르 행렬의 적용예에 대하여 상세히 설명한다. MAC 어레이는 벡터 프로세스의 확장 된 형태로 사이클마다 MAC 연산을 수행할 수 있다.

[0041] 종래 MAC 어레이의 모든 가중치 값이 하드 코딩된 경우, 방대한 양의 MAC 연산을 수행하기 위해 타일링(tiling)이 불가하다는 점, 컨벌루션 레이어의 모든 계층의 MAC 연산을 현실적으로 모두 처리할 수 없다는 점, DNN 가속화를 위하여 프로그래밍 가능한 가중치가 필요하다는 점의 문제를 해결할 수 없었다.

[0042] 따라서, 본 발명의 몇몇 실시예에 따른 MAC 어레이의 가중치 값은 하다마르 행렬을 이용하여 수정한 형태의 가중치 패턴을 이용하여 지정될 수 있다. 이 경우, MAC 어레이를 이용해 타일링 수행함으로써 방대한 양의 MAC 연산이 커버될 수 있다. 또한, 본 발명의 몇몇 실시예에 따라 LBCNN 구조의 인공 신경망을 이용함으로써 컨벌루션 레이어의 가중치 매개 변수의 프로그래밍 가능성을 열어둘 수 있고, 상기 LBCNN 구조의 인공 신경망의 가속화 및 확장성 보장이 가능해질 수 있다.

[0043] 도 3을 참조할 때, 일 실시예에 따른 LBC 레이어의 가중치 행렬 W 및 가중치 매개 변수가 지정된 MAC 어레이 $\bar{W}$가 모두 하다마르 행렬의 형태일 수 있다. 이 경우, LBC 레이어의 가중치 행렬 W과 MAC 어레이 $\bar{W}$의 크기(100)가 동일하거나 또는 LBC 레이어의 가중치 행렬 W의 크기가 MAC 어레이 $\bar{W}$의 크기의 배수이므로 상기 MAC 어레이를 수정하지 않고 MAC 연산에 사용될 수 있다.

[0044] 도 4를 참조할 때, 일 실시예에 따른 가중치 매개 변수가 지정된 MAC 어레이 $\bar{W}$만 하다마르 행렬 형태일 수 있다. LBC 레이어의 가중치 행렬 W이 하다마르 행렬 형태가 아닌 경우, MAC 어레이 $\bar{W}$를 기초로 타일링을 수행하여 상기 LBC 레이어의 가중치 행렬 W보다 모든 차원에서 큰 하다마르 행렬 형태를 가중치 행렬(110)로 이용할 수 있다.

[0045] 상기 가중치 행렬(110)의 적어도 하나의 행 및 적어도 하나의 열이 사용되지 않을 수 있으나, MAC 어레이 $\bar{W}$가 하다마르 행렬 형태인 경우 LBC 레이어의 가중치 행렬 W의 크기는 문제되지 않으며 최소한의 하드웨어 자원을 이용하여 상기 가중치 행렬(110)의 하위 행렬(sub matrix)의 가중치를 생성할 수 있다.

[0046] 도 5를 참조할 때, 일 실시예에 따른 가중치 매개 변수가 지정된 MAC 어레이 $\bar{W}$의 크기가 임의의 크기일 수 있다. 이 경우 MAC 어레이 $\bar{W}$의 크기는 하다마르 행렬 $\tilde{W}$의 배열일 수 있다.

[0047] 도 5를 참조하면 하다마르 행렬 $\tilde{W}$의 부정(negation) 형태인 행렬 $-\tilde{W}$은 개별적으로 수정될 수 있고, 행렬 $-\tilde{W}$ 각각도 역시 하다마르 형태의 가중치 행렬(120)을 구성하기 위해 사용될 수 있다.

[0048] 이하 도 6을 참조하여 인공 신경망의 학습 순서에 대하여 설명한다.

[0049] 단계 S131에서 인공 신경망의 입력 데이터가 지정된 값의 가중치 행렬을 갖는 LBC 레이어에 입력될 수 있다. 상

기 LBC 레이어의 가중치 행렬을 획득하는 방법에 대하여는 전술한바, 중복 기재를 피하기 위해 상세한 설명은 생략한다.

[0050] 단계 S132에서 LBC 레이어에서 출력된 데이터를 이용하여 컨벌루션 레이어가 학습될 수 있다. 즉 상기 LBC 레이어에서 출력된 데이터가 컨벌루션 레이어의 학습 데이터로 이용될 수 있다. 바람직하게는 상기 컨벌루션 레이어의 크기는 1x1 일 수 있다. 따라서 상기 1x1 컨벌루션 레이어의 출력 데이터는 본 발명의 몇몇 실시예에 따른 인공 신경망의 출력 특징 맵(Output Feature Map)일 수 있다.

[0051] 이하 도 7을 참조하여 본 발명의 일 실시예에 따른 인공 신경망 학습 장치의 구성 및 동작에 대하여 상세히 설명한다.

[0052] 일 실시예에 따른 인공 신경망 학습 장치(300)는 가중치 패턴 획득부(301), 가중치 고정부(302), 인공 신경망 학습부(303), 가중치 행렬 DB(304) 및 입력 데이터 DB(305)를 포함할 수 있다.

[0053] 일 실시예에 따른 가중치 패턴 획득부(301)는 하다마르 행렬을 이용하여 상기 가중치 패턴을 획득할 수 있다. 이후 획득한 가중치 패턴을 가중치 고정부(302)에 전달할 수 있다.

[0054] 일 실시예에 따른 가중치 고정부(302)는 가중치 패턴 획득부(301)에서 획득한 가중치 패턴을 이용하여, LBC 레이어의 가중치 행렬을 지정된 값으로 고정할 수 있다. 상기 가중치 행렬은 가중치 패턴을 기초로 생성된 행렬로 가중치 행렬 DB(304)로 부터 획득될 수 있다.

[0055] 일 실시예에 따른 인공 신경망 학습부(303)는 LBC 레이어 및 컨벌루션 레이어를 포함하는 CNN 구조의 인공 신경망 학습을 수행시킬 수 있다. 이 경우 상기 LBC 레이어의 가중치 매개 변수는 가중치 고정부(302)에서 이미 지정된 값으로 고정될 수 있다. 상기 인공 신경망은 바람직하게 LBCNN일 수 있고, 본 발명의 몇몇 실시예에 따른 LBCNN은 적어도 하나의 LBC 레이어와 컨벌루션 레이어를 포함하며, 상기 컨벌루션 레이어는 1x1 크기일 수 있다.

[0056] 일 실시예에 따른 가중치 행렬 DB(304)는 가중치 패턴을 기초로 생성된 가중치 행렬이 저장될 수 있다. 가중치 행렬은 상기 획득된 가중치 패턴이 반복된 형태의 행렬일 수 있고, 상기 가중치 패턴이 부정된 형태의 행렬일 수 있으며, 상기 가중치 패턴에 대하여 희소성이 크도록 수정된 형태의 행렬일 수도 있다.

[0057] 일 실시예에 따른 및 입력 데이터 DB(305)는 인공 신경망의 학습 데이터일 수 있다. 즉, 상기 입력 데이터 DB(305)에 저장된 학습 데이터는 LBC 레이어의 입력데이터일 수 있고, 상기 LBC 레이어의 출력 데이터를 이용하여 컨벌루션 레이어가 학습될 수 있다.

[0058] 이하 도 8을 참조하여 본 발명의 일 실시예에 따른 인공 신경망 학습 방법을 수행하는 프로그램이 기록된 기록매체에 대하여 상세히 설명한다.

[0059] 도 8에 도시된 바와 같이, 본 발명의 일 실시예에 따른 인공 신경망 학습을 수행하는 컴퓨팅 장치(400)는 프로세서(410), 메모리(420), 스토리지(440), 네트워크 인터페이스(430)를 포함하고 몇몇 실시예에서 시스템 버스(450)를 더 포함할 수 있다.

[0060] 메모리(420)에 로드 되어 저장되는 하나 이상의 인스트럭션(421, 422, 423)은 프로세서(410)를 통하여 실행된다. 본 실시예에 따른 인공 신경망 학습 방법을 수행하는 컴퓨팅 장치(400)는 별도의 설명이 없더라도 도 1 내지 도 6을 참조하여 설명한 인공 신경망 학습 방법을 수행할 수 있는 점을 유의한다.

[0061] 네트워크 인터페이스(430)는 외부 서버로부터 가중치 패턴 및 인공 신경망의 학습 데이터와 관련된 데이터를 수신할 수 있고, 상기 수신된 정보는 스토리지(440)에 저장되도록 할 수 있다.

[0062] 스토리지(440)는 가중치 행렬 DB(441) 및 입력 데이터 DB(442)를 저장할 수 있다.

[0063] 상기 하나 이상의 인스트럭션은, 가중치 패턴 획득 인스트럭션(421), 가중치 고정 인스트럭션(422) 및 인공 신경망 학습 인스트럭션(423)을 포함할 수 있다.

[0064] 일 실시예에 따른 가중치 패턴 획득 인스트럭션(421)이 실행됨으로써, 하다마르 행렬을 이용하여 가중치 패턴이 획득될 수 있다. 또한 일 실시예에 따라 상기 가중치 패턴을 이용하여 획득된 가중치 행렬(424)이 메모리에 저장될 수 있다.

[0065] 메모리에 저장된 가중치 행렬(424)은 스토리지(440)의 가중치 행렬 DB(441)로 부터 획득된 행렬일 수 있다. 일 실시예에 따른 가중치 행렬(424)은 MAC 어레이의 크기와 동일하거나 배수 크기인 하다마르 행렬 형태일 수 있으

나, 이에 한정되지 않고 MAC 어레이가 하다마르 행렬 형태인 경우 상기 가중치 행렬의 크기는 하다마르 행렬이 아닐 수도 있다.

[0066] 일 실시예에 따른 가중치 고정 인스트럭션(422)이 실행됨으로써, LBC 레이어의 가중치 행렬이 지정된 값으로 고정될 수 있다. 즉, MAC(Multiply and Accumulate) 어레이의 입력값 및 출력값을 조정하여 LBC 레이어의 가중치 매개 변수가 지정된 값으로 하드 와이어로 연결될 수 있고, 바람직하게는 LBC 레이어의 모든 가중치가 하드 와이어로 연결될 수 있다.

[0067] 일 실시예에 따른 인공 신경망 학습 인스트럭션(423)이 실행됨으로써, LBC 레이어 및 컨벌루션 레이어를 포함하는 CNN 구조의 인공 신경망 학습이 수행될 수 있다. 상기 인공 신경망은 바람직하게 LBCNN일 수 있고, 본 발명의 몇몇 실시예에 따른 LBCNN은 적어도 하나의 LBC 레이어와 컨벌루션 레이어를 포함하며, 상기 컨벌루션 레이어는 1x1 크기일 수 있다.

[0068] 본 발명에서 설명하는 특정 실행들은 일 실시 예들로서, 어떠한 방법으로도 본 발명의 범위를 한정하는 것은 아니다. 명세서의 간결함을 위하여, 종래 전자적인 구성들, 제어 시스템들, 소프트웨어, 상기 시스템들의 다른 기능적인 측면들의 기재는 생략될 수 있다. 또한, 도면에 도시된 구성 요소들 간의 선들의 연결 또는 연결 부재들은 기능적인 연결 및/또는 물리적 또는 회로적 연결들을 예시적으로 나타낸 것으로서, 실제 장치에서는 대체 가능하거나 추가의 다양한 기능적인 연결, 물리적인 연결, 또는 회로 연결들로서 나타내어질 수 있다. 또한, “필수적인”, “중요하게” 등과 같이 구체적인 언급이 없다면 본 발명의 적용을 위하여 반드시 필요한 구성 요소가 아닐 수 있다.

[0069] 본 발명의 명세서(특히 특허청구범위에서)에서 “상기”의 용어 및 이와 유사한 지시 용어의 사용은 단수 및 복수 모두에 해당하는 것일 수 있다. 또한, 본 발명에서 범위(range)를 기재한 경우 상기 범위에 속하는 개별적인 값을 적용한 발명을 포함하는 것으로서(이에 반하는 기재가 없다면), 발명의 상세한 설명에 상기 범위를 구성하는 각 개별적인 값을 기재한 것과 같다. 마지막으로, 본 발명에 따른 방법을 구성하는 단계들에 대하여 명백하게 순서를 기재하거나 반하는 기재가 없다면, 상기 단계들은 적당한 순서로 행해질 수 있다. 반드시 상기 단계들의 기재 순서에 따라 본 발명이 한정되는 것은 아니다. 본 발명에서 모든 예들 또는 예시적인 용어(예들 들어, 등등)의 사용은 단순히 본 발명을 상세히 설명하기 위한 것으로서 특허청구범위에 의해 한정되지 않는 이상 상기 예들 또는 예시적인 용어로 인해 본 발명의 범위가 한정되는 것은 아니다. 또한, 당업자는 다양한 수정, 조합 및 변경이 부가된 특허청구범위 또는 그 균등물의 범주 내에서 설계 조건 및 팩터에 따라 구성될 수 있음을 알 수 있다.

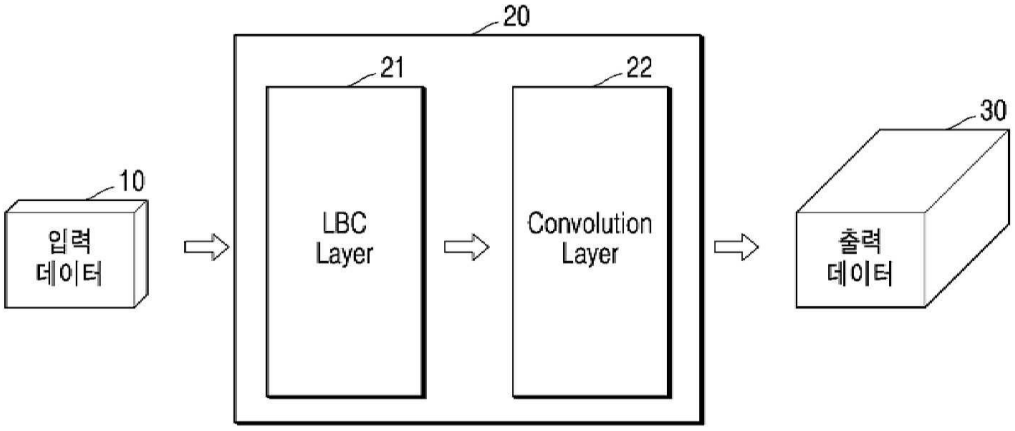
[0070] 이상 설명된 본 발명에 따른 실시예는 다양한 컴퓨터 구성요소를 통하여 실행될 수 있는 프로그램 명령어의 형태로 구현되어 컴퓨터 판독 가능한 기록 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능한 기록 매체는 프로그램 명령어, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 컴퓨터 판독 가능한 기록 매체에 기록되는 프로그램 명령어는 본 발명을 위하여 특별히 설계되고 구성된 것이거나 컴퓨터 소프트웨어 분야의 당업자에게 공지되어 사용 가능한 것일 수 있다. 컴퓨터 판독 가능한 기록 매체의 예에는, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD-ROM 및 DVD와 같은 광기록 매체, 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical medium), 및 ROM, RAM, 플래시 메모리 등과 같은, 프로그램 명령어를 저장하고 실행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령어의 예에는, 컴파일러에 의하여 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용하여 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드도 포함된다. 하드웨어 장치는 본 발명에 따른 처리를 수행하기 위하여 하나 이상의 소프트웨어 모듈로 변경될 수 있으며, 그 역도 마찬가지이다.

[0071] 이상에서 본 발명이 구체적인 구성요소 등과 같은 특정 사항과 한정된 실시예 및 도면에 의하여 설명되었으나, 이는 본 발명의 보다 전반적인 이해를 돕기 위하여 제공된 것일 뿐, 본 발명이 상기 실시예에 한정되는 것은 아니며, 본 발명이 속하는 기술분야에서 통상적인 지식을 가진 자라면 이러한 기재로부터 다양한 수정과 변경을 꾀할 수 있다.

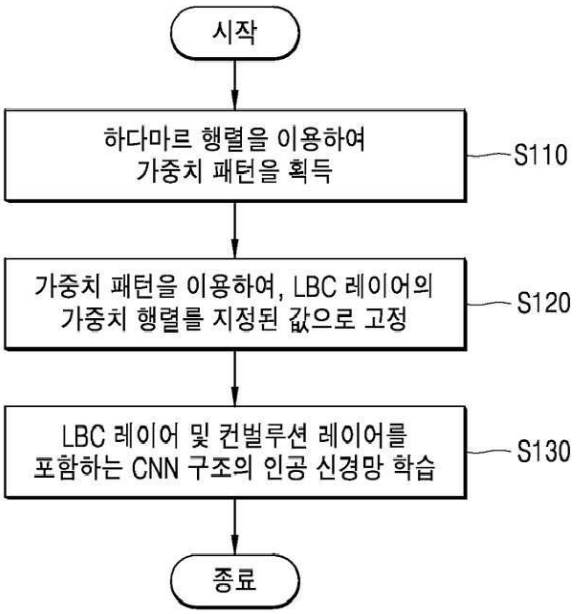
[0072] 따라서, 본 발명의 사상은 상기 설명된 실시예에 국한되어 정해져서는 아니 되며, 후술하는 특허청구범위뿐만 아니라 이 특허청구범위와 균등한 또는 이로부터 등가적으로 변경된 모든 범위는 본 발명의 사상의 범주에 속한다고 할 것이다.

도면

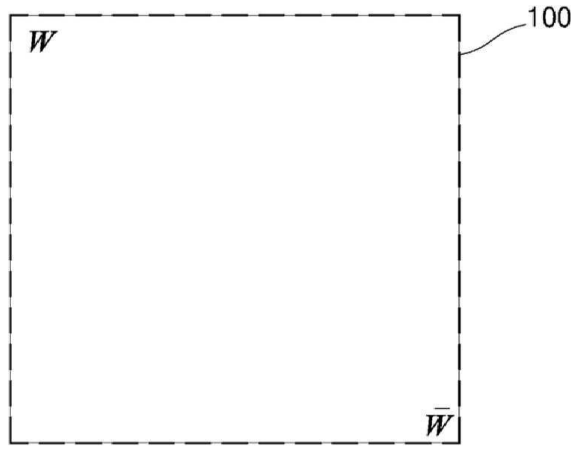
도면1


20
21
22
10
30
입력
데이터
LBC
Layer
Convolution
Layer
출력
데이터

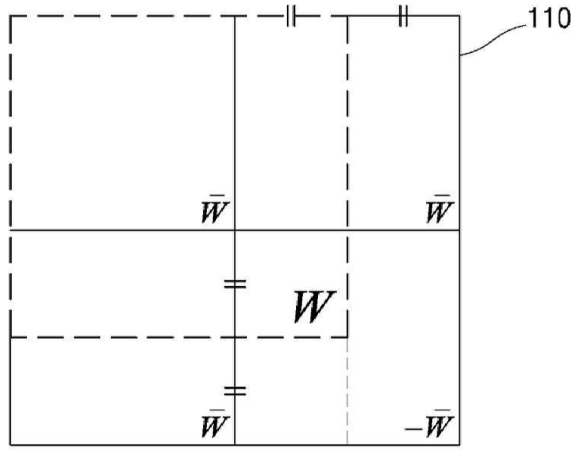
도면2


시작
하다마르 행렬을 이용하여
가중치 패턴을 획득
S110
가중치 패턴을 이용하여, LBC 레이어의
가중치 행렬를 지정된 값으로 고정
S120
LBC 레이어 및 컨벌루션 레이어를
포함하는 CNN 구조의 인공 신경망 학습
S130
종료

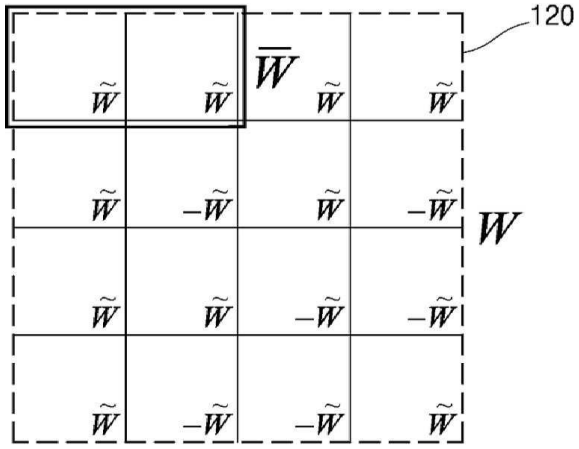
도면3



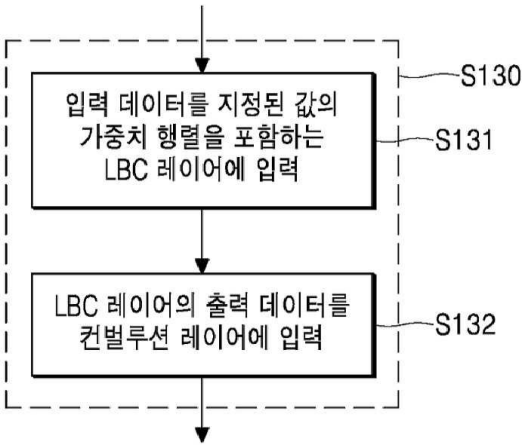
도면4



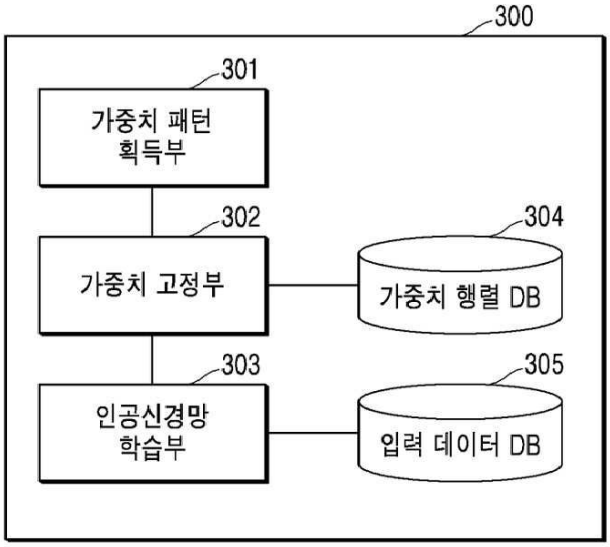
도면5



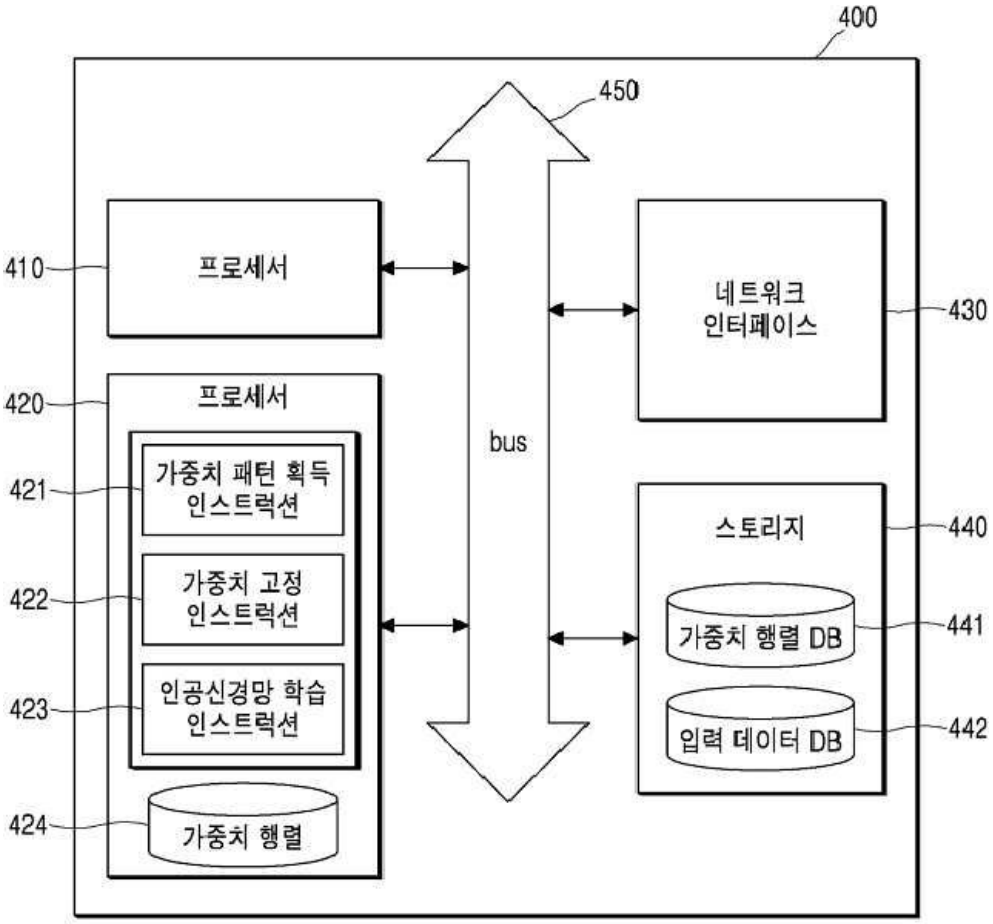
도면6


S130
입력 데이터를 지정된 값의 가중치 행렬을 포함하는 LBC 레이어에 입력
S131
LBC 레이어의 출력 데이터를 컨벌루션 레이어에 입력
S132

도면7


300
301
가중치 패턴 획득부
302
가중치 고정부
303
인공신경망 학습부
304
가중치 행렬 DB
305
입력 데이터 DB

도면8


400
450
410
프로세서
420
프로세서
421
가중치 패턴 획득
인스트럭션
422
가중치 고정
인스트럭션
423
인공신경망 학습
인스트럭션
424
가중치 행렬
bus
네트워크
인터페이스
430
스토리지
440
가중치 행렬 DB
441
입력 데이터 DB
442