(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2024년09월26일
(11) 등록번호 10-2711475
(24) 등록일자 2024년09월24일

(51) 국제특허분류(Int. Cl.)
G06N 3/0464 (2023.01) *G06F 17/16* (2006.01)
G06N 3/047 (2023.01) *G06N 3/0495* (2023.01)
G06N 3/08 (2023.01)
(52) CPC특허분류
G06N 3/0464 (2023.01)
G06F 17/16 (2013.01)
(21) 출원번호 10-2023-0053998
(22) 출원일자 2023년04월25일
심사청구일자 2023년04월25일
(56) 선행기술조사문헌
G. Wang 등. "Multi-hop Attention Graph Neural Networks". arXiv:2009.14332v5
F. Shen 등. “GiT: Graph Interactive Transformer for Vehicle Re-identification”. arXiv:2107.05475v3*
H. Cai 등. “Graph Transformer Geometric Learning of Brain Networks Using Multimodal MR Images for Brain Age Estimation”. IEEE Transactions on Medical Imaging*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
이화여자대학교 산학협력단
서울특별시 서대문구 이화여대길 52 (대현동, 이화여자대학교)
(72) 발명자
신태훈
서울특별시 용산구 새창로 70, 111동 903호 (도원동, 도원동삼성래미안)
임희주
서울특별시 관악구 관악로40길 60, 117동 802호 (봉천동, 현대아파트)
(74) 대리인
특허법인(유한)아이시스

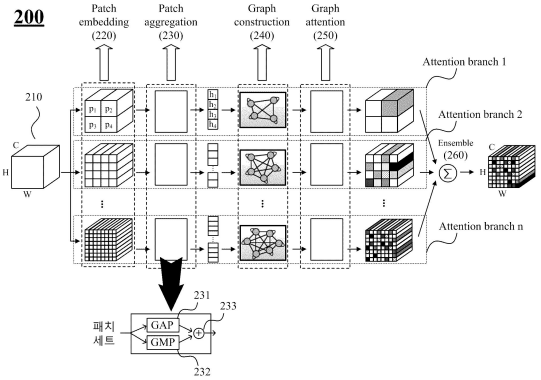
전체 청구항 수 : 총 10 항 심사관 : 성민용

(54) 발명의 명칭 **딥러닝 모델에서의 전역적 특징 증강 방법 및 데이터 처리 장치**

(57) 요 약

딥러닝 모델에서 입력 데이터의 전역적 특징을 증강하는 방법은 데이터 처리 장치가 일정 크기의 패치 단위로 입력되는 특징 텐서를 복수의 패치들로 구분하는 단계, 상기 데이터 처리 장치가 상기 복수의 패치들 각각이 노드에 해당하는 그래프를 구축하는 단계, 상기 데이터 처리 장치가 상기 구축된 그래프에 속한 노드들 사이의 상대적 중요도를 반영하는 그래프 어텐션을 수행하는 단계 및 상기 데이터 처리 장치가 상기 그래프 어텐션 수행으로 업데이트된 패치들을 포함하는 상기 특징 텐서를 다음 계층에 전달하는 단계를 포함한다.

대 표 도



(52) CPC특허분류
G06N 3/047 (2023.01)
G06N 3/0495 (2023.01)
G06N 3/08 (2023.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711174090
과제번호 00155966
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 인공지능융합혁신인재양성
연구과제명 인공지능융합혁신인재양성(이화여자대학교)
기 여 율 1/2
과제수행기관명 이화여자대학교산학협력단
연구기간 2022.07.01 ~ 2022.12.31
이 발명을 지원한 국가연구개발사업
과제고유번호 1345347338
과제번호 2020R1A6A1A03043528
부처명 교육부
과제관리(전문)기관명 한국연구재단
연구사업명 이공학학술연구기반구축
연구과제명 이화여자대학교 뇌융합과학연구원
기 여 율 1/2
과제수행기관명 이화여자대학교
연구기간 2023.03.01 ~ 2024.02.29

명 세 서

청구범위

청구항 1

데이터 처리 장치가 딥러닝 모델의 이전(previous) 계층에서 출력하는 특징 텐서를 입력받는 단계;

상기 데이터 처리 장치가 일정 크기의 패치 단위로 상기 특징 텐서를 복수의 패치들로 구분하는 단계;

상기 데이터 처리 장치가 상기 복수의 패치들 각각이 노드에 해당하는 그래프를 구축하는 단계;

상기 데이터 처리 장치가 상기 구축된 그래프에 속한 노드들 사이의 상대적 중요도를 반영하는 그래프 어텐션을 수행하는 단계; 및

상기 데이터 처리 장치가 상기 그래프 어텐션 수행으로 업데이트된 패치들을 포함하는 상기 특징 텐서를 상기 이전 계층의 다음 계층에 전달하는 단계를 포함하되,

상기 그래프를 구축하는 단계는 상기 데이터 처리 장치가 상기 복수의 패치들을 노드들로 설정하는 단계; 상기 데이터 처리 장치가 상기 노드들을 학습가능한 행렬을 이용하여 임베딩하는 단계; 상기 데이터 처리 장치가 상기 임베딩된 노드들 중 모든 노드 쌍 사이의 거리 유사도 점수를 연산하여 에지 행렬을 구성하는 단계; 및 상기 데이터 처리 장치가 멀티홉 연결성을 표현하도록 확률적 모델을 사용하여 상기 에지 행렬을 업데이트하는 단계를 포함하고, 상기 그래프는 상기 업데이트된 에지 행렬로 정의되는, 딥러닝 모델에서의 전역적 특징 증강 방법.

청구항 2

제1항에 있어서,

상기 데이터 처리 장치가 상기 복수의 패치들 각각에 대한 풀링(pooling)을 수행하여 차원을 축소하는 단계를 더 포함하고,

상기 그래프는 상기 차원이 축소된 복수의 패치들을 노드 집합으로 포함하는,

딥러닝 모델에서의 전역적 특징 증강 방법.

청구항 3

제1항에 있어서,

상기 데이터 처리 장치가 상기 복수의 패치들 각각에 대한 전역 평균 풀링을 수행하여 제1 축소 패치들을 생성하는 단계;

상기 데이터 처리 장치가 상기 복수의 패치들 각각에 대한 전역 최대 풀링을 수행하여 제2 축소 패치들을 생성하는 단계; 및

상기 데이터 처리 장치가 상기 제1 축소 패치들 및 상기 제2 축소 패치들을 결합하여 최종 축소 패치들을 생성하는 단계를 더 포함하고,

상기 그래프는 상기 최종 축소 패치들을 노드 집합으로 포함하는,

딥러닝 모델에서의 전역적 특징 증강 방법.

청구항 4

삭제

청구항 5

제1항에 있어서,

상기 에지 행렬을 업데이트하는 단계는

상기 데이터 처리 장치가 대각 행렬을 이용하여 특정 노드로 진입하는 모든 에지들의 합이 1이 되도록 상기 에지 행렬을 정규화하는 단계;

상기 데이터 처리 장치가 상기 정규화된 에지 행렬의 값을 채프만-콜모고로프 방정식을 이용하여 멀티홉 상태 전이 확률로 변환하는 단계;

상기 데이터 처리 장치가 상기 전이 확률로 변환된 에지 행렬을 1부터 n까지 의 멀티홉 개수를 가중 합산하는 단계; 및

상기 데이터 처리 장치가 상기 가중 합산된 에지 행렬을 대칭 행렬로 변환하는 단계를 포함하는 딥러닝 모델에서의 전역적 특징 증강 방법.

청구항 6

데이터 처리 장치가 딥러닝 모델의 이전(previous) 계층에서 출력하는 특징 텐서를 입력받는 단계;

상기 데이터 처리 장치가 서로 다른 크기의 패치 단위들 각각으로 상기 특징 텐서를 분할하여 복수의 패치 세트들을 생성하는 단계;

상기 데이터 처리 장치가 상기 복수의 패치 세트들 각각에 대하여 그래프 어텐션을 수행하는 단계;

상기 데이터 처리 장치가 상기 복수의 패치 세트들 각각에 대하여 상기 그래프 어텐션 수행으로 업데이트된 패치들을 포함하는 특징 텐서를 생성하고, 상기 복수의 패치 세트들로부터 산출한 복수의 특징 텐서들을 평균화하여 통합하는 단계; 및

상기 데이터 처리 장치가 상기 통합된 특징 텐서를 상기 이전 계층의 다음 계층에 전달하는 단계를 포함하되,

상기 그래프 어텐션을 수행하는 단계는 상기 패치 세트에 속한 복수의 패치들 각각이 노드에 해당하는 그래프를 구축하는 단계 및 상기 구축된 그래프에 속한 노드 임베딩들 사이의 상대적 중요도를 반영하는 그래프 어텐션을 수행하는 단계를 포함하는, 딥러닝 모델에서의 전역적 특징 증강 방법.

청구항 7

제6항에 있어서,

상기 그래프 어텐션을 수행하는 단계는 상기 데이터 처리 장치가 상기 복수의 패치들 각각에 대한 풀링(pooling)을 수행하여 차원을 축소하는 단계를 더 포함하고, 상기 그래프는 상기 차원이 축소된 복수의 패치들을 이용하여 구축되는, 딥러닝 모델에서의 전역적 특징 증강 방법.

청구항 8

제6항에 있어서,

상기 그래프를 구축하는 단계는

상기 데이터 처리 장치가 상기 복수의 패치들을 노드들로 설정하는 단계;

상기 데이터 처리 장치가 상기 노드들을 학습가능한 행렬을 이용하여 임베딩하는 단계;

상기 데이터 처리 장치가 상기 임베딩된 노드들 중 모든 노드 쌍 사이의 거리 유사도 점수를 연산하여 에지 행렬을 구성하는 단계; 및

상기 데이터 처리 장치가 멀티홉 연결성을 표현하도록 확률적 모델을 사용하여 상기 에지 행렬을 업데이트하는 단계를 포함하고,

상기 그래프는 상기 업데이트된 에지 행렬로 정의되는,

딥러닝 모델에서의 전역적 특징 증강 방법.

청구항 9

제8항에 있어서,

상기 에지 행렬을 업데이트하는 단계는

상기 데이터 처리 장치가 대각 행렬을 이용하여 특정 노드로 진입하는 모든 에지들의 합이 1이 되도록 상기 에지 행렬을 정규화하는 단계;

상기 데이터 처리 장치가 상기 정규화된 에지 행렬의 값을 채프만-콜모고로프 방정식을 이용하여 멀티홉 상태 전이 확률로 변환하는 단계;

상기 데이터 처리 장치가 상기 전이 확률로 변환된 에지 행렬을 1부터 n까지 의 멀티홉 개수를 가중 합산하는 단계; 및

상기 데이터 처리 장치가 상기 가중 합산된 에지 행렬을 대칭 행렬로 변환하는 단계를 포함하는 딥러닝 모델에서의 전역적 특징 증강 방법.

청구항 10

입력데이터를 입력받는 인터페이스 장치;

적어도 하나의 전역적 특징 증강 모듈을 포함하는 딥러닝 모델을 저장하는 저장장치; 및

상기 딥러닝 모델을 이용하여 상기 입력데이터로부터 일정한 데이터 내지 정보를 산출하는 연산장치를 포함하되,

상기 적어도 하나의 전역적 특징 증강 모듈은 상기 딥러닝 모델의 이전(previous) 계층에서 전달되는 특징 텐서로부터 추출한 복수의 패치들의 관계를 그래프로 구축하고, 상기 그래프에 대한 어텐션을 수행하여 업데이트된 패치들을 포함하는 상기 특징 텐서를 상기 이전 계층의 다음 계층에 전달하고,

상기 그래프는 복수의 패치들 각각이 노드에 대응되고, 학습가능한 행렬을 이용하여 임베딩된 노드들 사이의 거리 유사도 점수로 산출되는 값을 에지 값으로 갖고,

상기 에지 값은 상기 임베딩된 노드들 중 모든 노드 쌍 사이의 거리 유사도 점수를 멀티홉 연결성을 표현하도록 확률적 모델을 사용하여 업데이트한 값인, 전역적 특징 표현 가능한 딥러닝 모델을 이용하는 데이터 처리 장치.

청구항 11

제10항에 있어서,

상기 적어도 하나의 전역적 특징 증강 모듈은 상기 복수의 패치들을 입력받는 전역 평균 풀링 계층; 및 상기 복수의 패치들을 입력받는 전역 최대 풀링 계층을 포함하고,

상기 그래프는 상기 전역 평균 풀링 계층 및 상기 전역 최대 풀링 계층 각각에서 출력하는 패치들을 결합하여 생성되는 최종 축소 패치들을 기준으로 구축되는 전역적 특징 표현 가능한 딥러닝 모델을 이용하는 데이터 처리 장치.

청구항 12

삭제

청구항 13

삭제

발명의 설명

기 술 분 야

[0001] 이하 설명하는 기술은 딥러닝 모델에서 추출하는 특징을 추출하는 기법에 관한 것이다. 특히 이하 설명하는 기술은 데이터의 전역적 특징을 추출하는 기법에 관한 것이다.

배 경 기 술

[0002] 인공신경망(artificial neural network)은 기계학습 모델 중 하나로 최근 다양한 분야에서 활용되고 있다. 영상처리 분야에서 많이 사용되는 CNN(convolution neural network)은 대표적인 인공신경망이다. 나아가, 합성곱 계

층(convolution layer)은 다양한 딥러닝 모델에서도 이용된다.

선행기술문헌

비특허문헌

[0003] (비특허문헌 0001) Dong, Y., et al., Weighted feature fusion of convolutional neural network and graph attention network for hyperspectral image classification. IEEE Transactions on Image Processing, 2022. 31: p. 1559-1572.

발명의 내용

해결하려는 과제

[0004] 합성곱 계층은 기본적으로 국소 특징(local feature) 추출에 장점이 있다. 합성곱은 일정한 크기를 갖는 윈도우(window)를 사용하여 특징을 추출하기 때문이다. 반대로 합성곱 계층은 입력 데이터에서 서로 멀리 떨어진 영역들간의 관계성(전역적 특징)을 추출하기 어렵다.

[0005] 이하 설명하는 기술은 합성곱 계층에서 출력되는 특징 텐서에서 전역적 특징을 증강하는 기법을 제공하고자 한다. 이하 설명하는 기술은 종래 합성곱 계층이 추출하지 못하는 먼 영역 사이의 관계성을 포함하는 특징 텐서를 제공하고자 한다.

과제의 해결 수단

[0006] 딥러닝 모델에서의 전역적 특징 증강 방법은 데이터 처리 장치가 딥러닝 모델의 이전(previous) 계층에서 출력하는 특징 텐서를 입력받는 단계, 상기 데이터 처리 장치가 일정 크기의 패치 단위로 상기 특징 텐서를 복수의 패치들로 구분하는 단계, 상기 데이터 처리 장치가 상기 복수의 패치들 각각이 노드에 해당하는 그래프를 구축하는 단계, 상기 데이터 처리 장치가 상기 구축된 그래프에 속한 노드들 사이의 상대적 중요도를 반영하는 그래프 어텐션을 수행하는 단계 및 상기 데이터 처리 장치가 상기 그래프 어텐션 수행으로 업데이트된 패치들을 포함하는 상기 특징 텐서를 상기 이전 계층의 다음 계층에 전달하는 단계를 포함한다.

[0007] 다른 측면에서 딥러닝 모델에서의 전역적 특징 증강 방법은 데이터 처리 장치가 딥러닝 모델의 이전(previous) 계층에서 출력하는 특징 텐서를 입력받는 단계, 상기 데이터 처리 장치가 서로 다른 크기의 패치 단위들 각각으로 상기 특징 텐서를 분할하여 복수의 패치 세트들을 생성하는 단계, 상기 데이터 처리 장치가 상기 복수의 패치 세트들 각각에 대하여 그래프 어텐션을 수행하는 단계, 상기 데이터 처리 장치가 상기 복수의 패치 세트들 각각에 대하여 상기 그래프 어텐션 수행으로 업데이트된 패치들을 포함하는 특징 텐서를 생성하고, 상기 복수의 패치 세트들로부터 산출한 복수의 특징 텐서들을 평균화하여 통합하는 단계 및 상기 데이터 처리 장치가 상기 통합된 특징 텐서를 상기 이전 계층의 다음 계층에 전달하는 단계를 포함한다. 상기 그래프 어텐션을 수행하는 단계는 상기 패치 세트에 속한 복수의 패치들 각각이 노드에 해당하는 그래프를 구축하는 단계 및 상기 구축된 그래프에 속한 노드 임베딩들 사이의 상대적 중요도를 반영하는 그래프 어텐션을 수행하는 단계를 포함한다.

[0008] 전역적 특징 표현 가능한 딥러닝 모델을 이용하는 데이터 처리 장치는 입력데이터를 입력받는 인터페이스 장치, 적어도 하나의 전역적 특징 증강 모듈을 포함하는 딥러닝 모델을 저장하는 저장장치 및 상기 딥러닝 모델을 이용하여 상기 입력데이터로부터 일정한 데이터 내지 정보를 산출하는 연산장치를 포함한다. 상기 적어도 하나의 전역적 특징 증강 모듈은 상기 딥러닝 모델의 이전(previous) 계층에서 전달되는 특징 텐서로부터 추출한 복수의 패치들의 관계를 그래프로 구축하고, 상기 그래프에 대한 어텐션을 수행하여 업데이트된 패치들을 포함하는 상기 특징 텐서를 상기 이전 계층의 다음 계층에 전달한다.

발명의 효과

[0009] 이하 설명하는 기술은 종래 합성곱 계층의 한계로 고려하지 못한 전역적 특징을 추출하여 딥러닝 모델의 성능을 증대한다. 나아가, 이하 설명하는 기술은 종래 다양한 딥러닝 모델의 계층 사이에 삽입하여 구현하기 용이하다.

도면의 간단한 설명

[0010] 도 1은 전역적 특징 추출이 가능한 합성곱 기반의 딥러닝 모델의 예이다.

도 2는 멀티홉 그래프 어텐션(MGA) 모듈의 동작을 설명하는 예이다.

도 3은 임베딩된 패치를 노드로 갖는 그래프에 대한 예이다.

도 4는 MGA 모듈이 사용되는 다양한 예이다.

도 5는 MGA 모듈이 처리하는 다양한 데이터 유형의 예이다.

도 6은 다차원 입력데이터를 처리하는 MGA 모듈의 동작을 설명하는 예이다.

도 7은 뇌 MRI로부터 뇌 연령을 예측하는 딥러닝 모델의 예이다.

도 8은 MGA 모듈을 사용하여 전역적 특징을 추출하는 데이터 처리 장치의 예이다.

발명을 실시하기 위한 구체적인 내용

[0011] 이하 설명하는 기술은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세하게 설명하고자 한다. 그러나, 이는 이하 설명하는 기술을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 이하 설명하는 기술의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.

[0012] 제1, 제2, A, B 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 해당 구성요소들은 상기 용어들에 의해 한정되지는 않으며, 단지 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 이하 설명하는 기술의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. 및/또는 이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.

[0013] 본 명세서에서 사용되는 용어에서 단수의 표현은 문맥상 명백하게 다르게 해석되지 않는 한 복수의 표현을 포함하는 것으로 이해되어야 하고, "포함한다" 등의 용어는 설명된 특징, 개수, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함을 의미하는 것이지, 하나 또는 그 이상의 다른 특징들이나 개수, 단계 동작 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 배제하지 않는 것으로 이해되어야 한다.

[0014] 도면에 대한 상세한 설명을 하기에 앞서, 본 명세서에서의 구성부들에 대한 구분은 각 구성부가 담당하는 주기능 별로 구분한 것에 불과함을 명확히 하고자 한다. 즉, 이하에서 설명할 2개 이상의 구성부가 하나의 구성부로 합쳐지거나 또는 하나의 구성부가 보다 세분화된 기능별로 2개 이상으로 분화되어 구비될 수도 있다. 그리고 이하에서 설명할 구성부 각각은 자신이 담당하는 주기능 이외에도 다른 구성부가 담당하는 기능 중 일부 또는 전부의 기능을 추가적으로 수행할 수도 있으며, 구성부 각각이 담당하는 주기능 중 일부 기능이 다른 구성부에 의해 전담되어 수행될 수도 있음은 물론이다.

[0015] 또, 방법 또는 동작 방법을 수행함에 있어서, 상기 방법을 이루는 각 과정들은 문맥상 명백하게 특정 순서를 기재하지 않은 이상 명기된 순서와 다르게 일어날 수 있다. 즉, 각 과정들은 명기된 순서와 동일하게 일어날 수도 있고 실질적으로 동시에 수행될 수도 있으며 반대의 순서대로 수행될 수도 있다.

[0017] 이하 설명하는 기술은 합성곱 계층을 사용하는 딥러닝 모델에서 전역적 특징을 추출하는 기법이다. 전역적 특징은 전체 입력 데이터에서 서로 멀리 떨어진 영역들 간의 관계성을 나타내는 특징을 포함한다. 여기서, 멀리 떨어진 영역들은 합성곱 과정에서 하나의 윈도우에 속하지 못하는 영역들을 의미한다.

[0018] 이하 데이터 처리 장치가 딥러닝 모델을 이용하여 입력 데이터를 처리한다고 설명한다. 데이터 처리 장치는 PC, 네트워크상의 서버, 스마트 기기, 전용 프로그램이 임베딩된 칩셋 등으로 구현될 수 있다.

[0019] 이하 딥러닝 모델은 합성곱 계층을 포함하는 다양한 유형의 모델 중 어느 하나를 의미한다. 다만, 이하 설명의 편의를 위하여 CNN을 중심으로 설명한다.

[0020] 딥러닝 모델이 처리하는 데이터 유형은 다양할 수 있다. 또한, 딥러닝 모델이 처리하는 데이터의 차원은 다양할 수 있다. 다만, 이하 설명의 편의를 위하여 데이터 유형은 2차원 영상 데이터를 중심으로 설명한다.

[0022] 도 1은 전역적 특징 추출이 가능한 합성곱 기반의 딥러닝 모델(100)의 예이다. 도 1은 CNN을 예로 설명한다.

[0023] CNN은 H×W 크기의 2차원 영상을 입력받는다.

[0024] CNN은 일반적으로 다수의 합성곱 계층을 반복적으로 사용하여 입력 영상의 복잡하고 다양한 특징들을 추출한다. 합성곱 계층은 설정된 파라미터(필터 크기, 스트라이드 간격, 채널수)에 따라 특정 크기의 특징 텐서(tensor)를 출력한다. 도 1은 i번째 합성곱 계층이 출력하는 텐서의 사이즈를 W_i, H_i 및 C_i로 표시하였다. C는 채널을 의미한다. 한편, 합성곱 계층은 필터의 진행 방향에 따라 1D 합성곱, 2D 합성곱, 3D 합성곱 등 다양한 방식의 합성곱을 수행할 수 있다.

[0025] 첫 번째 합성곱 계층(110)은 입력 영상을 입력받아 $H_1 \times W_1 \times C_1$ 크기의 특징 텐서를 출력한다.

[0026] 도 1에서 멀티홉 그래프 어텐션(Multi-hop Graph Attention, 이하 MGA라 함, 120) 모듈은 첫 번째 합성곱 계층(110) 다음에 위치한다. MGA 모듈(120)은 직전에 위치한 합성곱 계층에서 출력되는 특징 텐서를 입력받는다. MGA 모듈(120)은 입력되는 특징 텐서를 그래프 구조의 데이터로 변환하고, 그래프에 대한 어텐션을 통해 특징 텐서를 업데이트한다. MGA 모듈(120)은 다음 계층에 업데이트된 특징 텐서를 전달한다. 업데이트된 특징 텐서는 전역적 특징이 포함된다. 따라서, MGA 모듈(120)은 전역적 특징을 증강하는 모델 내지 전역적 특징 증강 모듈이라고도 명명할 수 있다. MGA 모듈의 상세한 동작은 후술한다.

[0027] 두 번째 합성곱 계층(130)은 MGA 모듈(120)로부터 업데이트된 특징 텐서를 입력받아 $H_2 \times W_2 \times C_2$ 크기의 특징 텐서를 출력한다.

[0028] n 번째 합성곱 계층(140)은 직전 합성곱 계층으로부터 특징 텐서를 입력받아 $H_3 \times W_3 \times C_3$ 크기의 특징 텐서를 출력한다.

[0029] 전연결 계층(fully connected layers, 150)은 n 번째 합성곱 계층(140)의 출력 텐서를 입력받아 일정한 정보를 출력한다. 예컨대, 전연결 계층(150)는 객체 분류 결과, 영상 분석 결과 등을 출력할 수 있다.

[0031] 도 2는 멀티홉 그래프 어텐션 모듈(MGA)의 동작(200)을 설명하는 예이다. MGA 모듈은 소프트웨어적인 모듈이다. 따라서, 실제 물리적인 데이터 처리 장치가 해당 모듈의 동작을 수행한다. 다만, 이하 도 2의 과정은 MGA 모듈이 수행한다고 설명한다.

[0032] MGA 모듈은 크게 두 가지 동작을 수행한다. MGA 모듈은 특징 텐서를 그래프 자료 구조로 변환하는 과정 및 그래프 어텐션을 통해 그래프의 노드값을 업데이트하여 특징 텐서를 재조절(업데이트)하는 과정을 포함한다.

[0033] MGA 모듈은 특징 텐서를 입력받는다(210). 특징 텐서는 $H \times W \times C$ 크기를 갖는다.

[0034] MGA 모듈은 다수의 어텐션 가지(attention branch)들을 포함할 수 있다. 도 2는 n개의 어텐션 가지들을 도시한 예이다. 이하 설명하는 과정 220~250은 각 가지에서 수행될 수 있다. 하나의 가지를 기준으로 설명한다.

[0035] MGA 모듈은 특징 텐서를 동일한 크기의 패치들로 나눈다(patch embedding, 220). 하나의 패치 p의 크기는 아래 수학식 1과 같이 정의될 수 있다.

수학식 1

[0037] $$p = \frac{H}{\Gamma} \times \frac{W}{\Gamma} \times C$$

[0038] Γ는 패치의 크기를 설정하는 하이퍼파라미터이다. Γ가 크다면, 하나의 패치의 크기는 작아지고, 생성되는 전체 패치들의 개수는 증가한다.

[0040] 하나의 특징 텐서는 일정한 크기의 패치들로 분할될 수 있다. 하나의 패치 세트는 하나의 특징 텐서를 일정한 크기로 나눈 패치들로 구성될 수 있다. 패치 세트 $P = \{p_1, p_2, \ldots, p_{Np}\}$이고, N_p는 패치 개수이다. 이하 N_p개의 패치들로 분할된 특징 텐서를 기준으로 설명한다.

[0041] MGA 모듈은 N_p개의 패치들로 분할된 특징 텐서의 차원을 축소하고 통합한다(patch aggregation, 230).

[0042] 경우에 따라서, MGA 모듈에서 패치의 차원 축소 및 통합은 선택적 구성일 수 있다. 패치의 차원 축소가 없는 경우, MGA 모듈은 220 과정에서 분할된 패치를 기준으로 그래프를 구축할 수 있다(240).

[0043] 도 2의 하단에 분할된 특징 텐서를 통합하는 과정을 도시한다. MGA 모듈은 전역 평균 풀링 계층 및 전역 최대

풀링 계층을 포함한다. 각 풀링 계층은 입력되는 패치들 각각에 대한 풀링을 수행한다. MGA 모듈은 패치들 각각에 전역 평균 풀링(global average pooling, GAP)을 수행하여 풀링된 제1 축소 패치들을 생성한다(231). 또한, MGA 모듈은 패치들 각각에 전역 최대 풀링(global max pooling, GMP)을 사용하여 풀링된 제2 축소 패치들을 생성한다(232). MGA 모듈은 제1 축소 패치들 및 제2 축소 텐서들을 합산(summation) 또는 결합(concatenation)하여 최종 축소 패치들을 생성한다(233). 최종 축소 패치 h_i는 아래 수학식 2와 같다. i는 패치 세트 중 인덱스를 의미한다.

수학식 2

[0045] $$h_i = GAP(p_i) \oplus GMP(p_i)$$

[0047] MGA 모듈은 전체 패치 세트에 대하여 통합 과정을 수행하여 축소 패치 세트 $H = \{H_1, H_2, \ldots, H_{Np}\}$를 생성한다.

[0049] MGA 모듈은 최종 축소 패치들 각각을 하나의 노드로 삼는 그래프를 구축한다(240).

[0050] 그래프 구축 과정에 대하여 설명한다. 그래프 G = (N, E)로 정의된다. N는 노드의 집합이고, E는 에지의 집합이다. 노드 집합은 수학식 2에 따라 산출되는 축소 패치 세트 H에 대응된다. 한편, 각 노드에 대응하는 패치의 크기는 하이퍼파라미터 Γ로 결정된다. Γ가 크다면, 보다 많은 지역적 패치들이 많은 개수의 노드로 임베딩된다.

[0051] 에지는 노드 간의 연결성을 나타낸다. 에지의 값은 두 개의 노드 임베딩 사이의 유클리디언 거리 유사도(Euclidean distance similarity) 점수로 정의될 수 있다. 동일한 노드의 유사도 점수는 1이다. MGA 모듈은 유클리디언 거리 유사도를 이용하여 노드 임베딩 값들 간의 관계성을 연산하여 초기 에지 행렬을 생성한다. 두 개의 노드 h_i와 h_j사이의 에지 e_{ij}는 아래 수학식 3으로 정의될 수 있다.

[0052] 한편, 노드 임베딩 값들 간의 관계성은 유클리디언 거리가 아닌 다른 형태의 거리가 활용될 수도 있다.

수학식 3

[0054] $$e_{ij} := 1/(exp(\|\mathbf{V}\mathbf{h}_i - \mathbf{V}\mathbf{h}_j\|_2))$$

[0055] h_i는 i 번째 노드이고, h_j는 j 번째 노드이다. 하나의 노드는 특징 벡터에 해당한다. V는 노드임베딩을 위한 행렬(학습가능한 행렬)을 의미하며 학습에 의해 결정되는 파라미터이다.

[0056] MGA 모듈은 가능한 모든 노드 쌍의 유클리디언 거리 유사도 점수를 연산하여 초기 에지 행렬 E를 생성한다. 엣지 계산에 관여하는 두 노드들의 순서가 상관이 없으므로 (즉 $e_{ij}=e_{ji}$ 이므로) E는 대칭 행렬이다.

[0057] MGA 모듈은 마르코프 성질을 지닌다는 가정하에 E를 확률적 모델링을 하여 업데이트한다. MGA 모듈은 이러한 확률적 모델을 사용하여 직접적 연결성(직접 연결된 노드와의 연결성) 및 간접적 연결성(다른 노드를 거쳐 연결된 노드와의 연결성, 멀티홉 연결성)을 고려한다.

[0059] 도 3은 임베딩된 패치를 노드로 갖는 그래프에 대한 예이다. 도 3은 상태 공간 $\Omega = \{H_1, H_2, \ldots, H_{Np}\}$에서 하나의 노드에서 다른 노드로의 연결성을 나타낸다. 유사도 점수는 상태 h_j로부터 h_i로 상태 전이(transition)하는 조건 확률 $P(X(t)= h_i | X(t-1) = h_j)$로 모델링될 수 있다. X(t)는 t번째 상태 전이의 결과를 나타낸다. $t \in T=\{0, 1, 2, \cdots, m\}$이다. 상태 전이는 단순하게 2개의 노드 사이에 발생할 수 있고, 연속된 3개 이상의 노드에서 발생할 수도 있다.

[0060] 특정 노드로 진입하는 모든 에지들의 합을 1로 만들기 위하여, 에지 행렬은 $\tilde{\mathbf{E}}=\mathbf{D}^{-1}\mathbf{E}$와 같이 정규화될 수 있다. D는 대각 요소 $d_{ii} = \sum_j e_{ij}$를 갖는 대각 행렬(diagonal matrix)이다. 정규화된 에지 행렬 $\tilde{\mathbf{E}}$는 모든 i 및 $j = \{1, 2, \ldots, N_p\}$에서 아래 수학식 4의 조건을 만족하는 확률 행렬에 해당한다.

수학식 4

[0062] $$\begin{cases} 1 \geq \tilde{e}_{ij} \geq 0 \\ \sum_{j=0}^{N_p} \tilde{e}_{ij} = 1 \end{cases}$$

[0064] 후술하겠지만 기존의 그래프 어텐션(graph attention)은 그래프 구조에서 하나의 노드가 직접 연결된 인접 노드들과의 관계성만을 학습하여 업데이트 된다. 그래프 네트워크의 표현능력을 증대하기 위하여, 연구자는 m차(m^{th} order) 이웃 정보까지 고려하도록 에지 행렬을 변환하였다.

[0065] m-홉 상태 전이 확률 $\tilde{e}_{ij}^{(m)}$은 현재 상태 h_j가 m(≥1)번의 상태 전이 후에 상태 h_j가 되는 과정에 해당한다. m-홉 상태 전이 확률은 그래프에서 노드 사이의 멀티홉 연결성에 대응된다. 상태 전이는 마르코프 특성을 만족한다고 가정한다. $\tilde{e}_{ij}^{(m)}$는 아래 수학식 5와 같이 채프만-콜모고로프 방정식(Chapman-Kolmogorov equation)으로 표현될 수 있다.

수학식 5

[0067] $$\begin{aligned} \tilde{e}_{ij}^{(m)} &= P\left(X(m) = \mathbf{h}_i \middle| X(0) = \mathbf{h}_j\right) \\ &= \sum_{k=0}^{N_p} P\left(X(m) = \mathbf{h}_i,\ X(m-1) = \mathbf{h}_k \middle| X(0) = \mathbf{h}_j\right) \\ &= \sum_{k=0}^{N_p} P(X(m) = \mathbf{h}_i | X(m-1) = \mathbf{h}_k) \cdot P\left(X(m-1) = \mathbf{h}_k \middle| X(0) = \mathbf{h}_j\right) \\ &= \sum_{k=0}^{N_p} \tilde{e}_{ik} \cdot \tilde{e}_{kj}^{(m-1)} \end{aligned}$$

[0068] $\tilde{e}_{ij}^{(m)}$는 초기 $\tilde{e}_{ij}^{(0)}$로 설정된다. 수학식 5는 m-홉 상태 전이의 확률이 재귀적으로 전체 중간 상태 전이들의 곱의 합산과 같다는 의미이다. 아래 수학식 6은 수학식 5를 행렬 형태로 다시 표현한 것이다.

수학식 6

[0070] $$\tilde{E}^{(m)} = \tilde{E}\,\tilde{E}^{(m-1)} = \tilde{E}^{(2)}\,\tilde{E}^{(m-2)} = \cdots = \tilde{E}^{m}$$

[0072] m-홉까지의 상태 전이를 고려하여, m차 에지 행렬 $\tilde{E}_{\forall}^{(m)}$는 추가적인 하이퍼파라미터 β를 갖는 형태로 아래 수학식 7과 같이 표현할 수 있다. β는 노드 연결의 홉 개수에 따라 에지 행렬에 가중치를 다르게 부여하는 하이퍼파라미터이다. β는 0에서 1사이의 값을 갖는다. 아래 수학식 7은 멀티홉 상태 전이 확률로 변환된 에지 행렬을 1부터 n까지 멀티홉 개수에 따라 가중치를 주어 합산하는 것이다.

수학식 7

[0074] $$\begin{aligned} \tilde{E}_{\forall}^{(m)} &= \tilde{E} + \beta\tilde{E}^{(2)} + \beta^2\tilde{E}^{(3)} + \cdots + \beta^{m-1}\tilde{E}^{(m)} \\ &= \sum_{k=1}^{m} \beta^{k-1}\tilde{E}^{k} \end{aligned}$$

[0075] $\tilde{E}_{\forall}^{(m)}$는 대칭 행렬이 아니다. 즉, $\tilde{E}_{\forall}^{(m)}$는 무방향 그래프의 요구 사항을 만족하지 못한다. 따라서, 최종 멀티홉 에지 행렬은 아래 수학7식 8과 같이 $\tilde{E}_{\forall}^{(m)}$와 $\tilde{E}_{\forall}^{(m)T}$를 합산하여 정의함으로써 대칭성을 갖도록 한다.

수학식 8

[0077] $$E_{\forall}^{(m)} := (\tilde{E}_{\forall}^{(m)} + \tilde{E}_{\forall}^{(m)T})/2$$

[0079] m이 큰 값 일수록 노드 세트 H에서 보다 전역적(먼 거리) 관계를 평가하게 한다. 다만, MGA 모듈이 먼거리의 전역적 거리를 고려하면 합성곱 계층이 추출하는 지역 특징이 저감될 수 있다. 따라서, 입력 데이터의 유형 또는 애플리케이션의 유형에 따라 데이터 처리 장치는 적절한 m값을 사용해야 한다.

[0081] MGA 모듈은 구축된 그래프에 대한 어텐션을 수행한다(250). 그래프 어텐션은 에지 정보를 활용하여 노드들 간의 상대적 중요도를 계산하여 노드를 업데이트하는 알고리즘이다.

[0082] 그래프에서 h_i와 h_j는 각각 i번째 노드와 j번째 노드이다. 이때 그래프 어텐션 계수 α_{ij}는 i번째 노드에 대한 j번째 노드의 중요성을 의미하며 아래 수학식 9와 같이 표현된다.

수학식 9

[0084] $$\alpha_{ij} = \frac{exp\left(LeakyReLU\left(\boldsymbol{W}^T[\boldsymbol{Vh_i} \| \boldsymbol{Vh_j}]\right)\right)}{\sum_{k \in N_i} exp(LeakyReLU(\boldsymbol{W}^T[\boldsymbol{Vh_i} \| \boldsymbol{Vh_k}])}$$

[0085] 수학식 9에서 ‖는 결합(concatenation) 연산이다. W 및 V는 학습에 의해 결정되는 가중치 행렬이다. 수학식 9에서 N_i는 노드 h_i를 업데이트하기 위해 참조될 이웃 노드들의 집합으로 일반적인 그래프 어텐션 네트워크에서는 h_i와 직접 연결된 노드들로 구성된다. 이에 반하여 전술한 바와 같이 MGA 모듈은 멀티홉 연결까지 고려할 수 있도록 변환된 에지 행렬을 사용하였기에, 간접적으로 연결된 노드들까지 고려하여 어텐션을 수행하게 된다.

[0086] 일반적인 그래프에서 에지는 바이너리 값으로 정의된다. 즉, 직접 연결된 에지는 1의 값으로 정의되고, 직접 연결되지 않는 에지는 0의 값으로 정의된다. 그러나, 전술한 패치 임베딩을 노드로 갖는 그래프에서 에지는 유클리디언 거리 유사도 기반의 임의의 실수값을 갖는다. 따라서, 그래프 어텐션을 위하여 그래프의 에지는 아래 수학식 10과 같이 일정한 임계값 Θ를 기준으로 이진화되어야 한다.

수학식 10

[0088] $$\boldsymbol{E}_{\forall}^{(m)} = \begin{cases} 1, & if\ \boldsymbol{E}_{\forall}^{(m)} > \theta \\ 0, & otherwise \end{cases}$$

[0089] Θ는 그래프 전체의 에지의 평균값을 사용할 수 있다. 또는 Θ는 실험적으로 결정될 수 있다.

[0091] MGA 모듈은 그래프 어텐션 계수를 연산하고, 아래 수학식 11을 이용하여 노드를 업데이트한다. MGA 모듈은 N_i에 속한 패치 p_j에 가중된 선형 변형을 수행하여 출력 패치 p'_i를 생성한다.

수학식 11

[0093] $$\boldsymbol{p'}_i = \sum_{j \in N_i}(\alpha_{ij} \cdot \boldsymbol{p}_j)$$

[0095] MGA 모듈은 전체 입력 패치 세트 P에 속한 패치들 각각에 대한 그래프 어텐션을 수행하여 출력 패치 세트 P'로 업데이트한다. 출력 패치 세트 P'는 전역적 특징이 증강된 특징 텐서를 구성한다.

[0097] 도 2와 달리 MGA 모듈은 하나의 어텐션 가지(예컨대, 도 2의 어텐션 가지들 중 하나)로 구성될 수 있다. 이 경우 MGA 모듈은 특징 텐서를 일정 크기의 패치로 분할하여 그래프 어텐션을 수행하여 산출되는 H×W×C 크기의 특징 텐서를 다음 계층에 출력할 수 있다.

[0098] 또한, 도 2와 같이 MGA 모듈은 다수의 어텐션 가지들로 구성될 수 있다. 이 경우 MGA 모듈은 어텐션 가지마다 서로 다른 크기의 패치 단위로 동일한 특징 텐서를 분할한다. 즉, MGA 모듈은 서로 다른 값을 갖는 Γ를 각 어텐션 가지에 적용하여 특징 텐서를 분할한다.

[0099] MGA 모듈은 제1 패치 단위, 제2 패치 단위, ... , 제n 패치 단위로 동일한 특징 텐서를 분할할 수 있다. 이때 패치의 크기는 제1 패치> 제2 패치> ... > 제n 패치라고 가정한다. 하나의 패치는 그래프에서 하나의 노드에 대응된다. 따라서, MGA 모듈은 입력 데이터에서 서로 다른 이격 거리에 있는 영역들 사이의 관계를 특징으로 추출할 수 있다.

[0100] 예컨대, 도 2에서 제1 어텐션 가지(attention branch 1)는 4개의 패치들 {p_1, p_2, p_3, p_4}로 분할된 특징 텐서를 처리하는 경로이다. 제2 어텐션 가지(attention branch 2)는 16개의 패치들로 분할된 특징 텐서를 처리하는 경로이다.

[0101] MGA 모듈은 각 어텐션 가지에서 동일한 과정을 수행한다. 다만, 각 어텐션 가지에서 처리하는 패치들의 개수가 다르기 때문에 MGA 모듈이 각 어텐션 가지에서 연산하는 유클리디언 거리 유사도 점수, 어텐션 계수 등은 서로 다른 값을 갖는다.

[0102] 다수의 어텐션 가지들을 갖는 경우, MGA 모듈은 각 어텐션 가지(어텐션 가지 1 ~ n)에서 산출되는 각 출력 패치 세트를 모두 통합하여 H×W×C 크기의 특징 텐서를 산출한다(260). 예컨대, MGA 모듈은 전체 어텐션 가지에서 산출되는 출력 패치 세트를 평균하여 최종 특징 텐서를 산출할 수 있다.

[0104] 도 4는 MGA 모듈이 사용되는 다양한 예이다. 도 4는 입력 데이터가 2차원 이미지인 예이다.

[0105] 도 4(A)는 하나의 MGA 모듈이 사용되는 예이다. 합성곱 계층(Conv)은 입력 이미지로부터 특징 텐서를 추출한다. MGA 모듈은 특징 텐서를 입력받아 도 2의 과정에서 설명한 그래프 어텐션으로 특징 텐서를 업데이트한다. MGA 모듈은 업데이트된 특징 텐서를 전연결 계층(FC)으로 전달한다. 이때 딥러닝 모델은 건너뛴 연결 (skip connection)을 통해 합성곱 계층이 출력하는 초기 특징 텐서와 MGA 모듈이 출력하는 업데이트된 특징 텐서를 합하여 다음 계층으로 전달할 수도 있다.

[0106] 도 4(B)는 다수의 MGA 모듈이 사용되는 예이다. 제1 합성곱 계층(Conv 1)은 입력 이미지로부터 제1 특징 텐서를 추출한다. 제1 MGA 모듈은 제1 특징 텐서를 입력받아 그래프 어텐션을 통해 제1 업데이트된 특징 텐서를 출력한다. 제2 합성곱 계층(Conv 2)은 제1 업데이트된 특징 텐서로부터 제2 특징 텐서를 추출한다. 제2 MGA 모듈은 제2 특징 텐서를 입력받아 그래프 어텐션을 통해 제2 업데이트된 특징 텐서를 출력한다. 제2 MGA 모듈은 제2 업데이트된 특징 텐서를 전연결 계층(FC)으로 전달한다. 도 4(B)는 2개의 MGA 모듈을 예로 도시하였지만, 딥러닝 모델의 구조 및 구현에 따라 더 많은 MGA 모듈들이 사용될 수도 있다.

[0107] 도 4(C)는 MGA 모듈과 다른 모듈이 혼합 사용되는 예이다. 도 4(C)에서 다른 모듈은 SE (Squeeze and Excitation) 블록을 예로 도시한다. 제1 합성곱 계층(Conv 1)은 입력 이미지로부터 제1 특징 텐서를 추출한다. 제1 MGA 혼합 모듈은 제1 특징 텐서를 입력받아 처리하는 SE 블록과 SE 블록의 출력을 입력받아 그래프 어텐션을 통해 제1 업데이트된 특징 텐서를 출력하는 MGA 모듈로 구성된다. 제2 합성곱 계층(Conv 2)은 제1 업데이트된 특징 텐서로부터 제2 특징 텐서를 추출한다. 제2 MGA 혼합 모듈은 제2 특징 텐서를 입력받아 처리하는 SE 블록과 병렬적으로 제2 특징 텐서를 입력받아 그래프 어텐션을 수행하는 MGA 모듈로 구성된다. 제2 MGA 혼합 모듈은 SE 블록이 출력하는 특징 텐서와 MGA 모듈이 출력하는 특징 텐서를 결합하여 제2 업데이트된 특징 텐서를 출력한다. 제2 MGA 혼합 모듈은 제2 업데이트된 특징 텐서를 전연결 계층(FC)으로 전달한다. 도 4(C)는 MGA 모듈과 복합적으로 이용할 수 있는 모듈로 SE 블록을 도시하지만, MGA 모듈은 동일 크기의 특징 텐서를 출력하는 다양한 모듈과 함께 활용될 수 있다.

[0109] 전술한 설명에서 입력 데이터는 2차원 이미지를 가정하였다. 그러나, 딥러닝 모델은 애플리케이션에 따라 다른 다양한 유형의 데이터를 처리할 수 있다. MGA 모듈도 다양한 유형의 데이터를 처리하는 딥러닝 모델에 삽입되어 활용될 수 있다.

[0110] 도 5는 MGA 모듈이 처리하는 다양한 데이터 유형의 예이다.

[0111] 도 5(A)는 원-핫(one-hot) 벡터로 구성되는 행렬을 입력받는 예이다. 원-핫 인코딩은 텍스트, 숫자 등으로 구성된 소스데이터를 벡터화한다. 합성곱 계층은 원-핫 인코딩된(one-hot encoded) 행렬을 입력받는다. MGA 모듈은 합성곱 계층이 출력하는 특징 텐서를 그래프 어텐션을 통해 업데이트한다.

[0112] 도 5(B)는 3차원 이미지를 입력받는 예이다. 합성곱 계층은 3차원 이미지를 입력받는다. 합성곱 계층은 3D 영상을 처리할 수 있는 다양한 구조 중 어느 하나일 수 있다. MGA 모듈은 합성곱 계층이 출력하는 특징 텐서를 그래프 어텐션을 통해 업데이트한다.

[0113] 도 5(C)는 시계열 데이터(sequential data)를 입력받는 예이다. 시계열 데이터는 X축과 Y축 외에 시간축 t를 갖는 3차원 형태일 수 있다. 시계열 데이터는 시간축에 따라 연속 또는 일정한 간격으로 수집된다. 시계열 데이터는 다양한 유형의 데이터 중 어느 하나일 수 있다. 예컨대, 시계열 데이터는 연속되는 2D 이미지(동영상), 연속되는 스펙트로그램(오디오 신호), 일정 간격으로 측정되는 센서 데이터 등일 수 있다. 합성곱 계층은 시계열 데이터를 입력받는다. MGA 모듈은 합성곱 계층이 출력하는 특징 텐서를 그래프 어텐션을 통해 업데이트한다.

[0115] 도 6은 다차원 입력데이터를 처리하는 MGA 모듈의 동작(300)을 설명하는 예이다. 도 6은 도 2에서 하나의 어텐션 가지에서의 동작에 해당한다. 따라서, MGA 모듈은 도 6과 같은 동작을 수행하는 다수의 어텐션 가지로 구성될 수 있다. 도 6의 각 과정은 도 2에서 대응되는 과정과 동일 내지 유사하다. 따라서, 도 6의 동작 중 도 2의 동작과 동일한 부분은 간략하게 설명한다. 도 6은 다차원 특징 텐서를 각 차원의 축을 기준으로 패치를 분할한 예이다.

[0116] MGA 모듈은 다차원 데이터에 대한 특징 텐서를 입력받는다(310). 특징 텐서는 a, b 및 c라는 차원을 갖는다. 특징 텐서는 a×b×c 크기를 갖는다. a, b 및 c는 각각 다수의 값을 갖는다.

[0117] MGA 모듈은 특징 텐서를 동일한 크기의 패치들로 나눈다(320). 도 6은 $\Gamma = 2$로 설정한 예이다. 도 2와 달리 MGA 모듈은 입력된 특징 텐서를 각 차원의 축에 따라 일정한 크기의 패치로 구분한다. 도 6에서 패치 세트 $P = \{p_1, p_2, \ldots, p_7, p_8\}$이다. MGA 모듈은 패치들 각각의 차원을 축소하고 통합하여 축소 패치 세트 $H = \{H_1, H_2, \ldots, H_8\}$를 생성한다(330).

[0118] MGA 모듈은 최종 축소 패치들 각각을 임베딩 한 후 하나의 노드로 삼는 그래프를 구축한다(340). MGA 모듈은 구축된 그래프에 대한 어텐션을 수행한다(350). 이 과정을 통해 MGA 모듈 업데이트된 특징 텐서를 산출한다.

[0120] 연구자는 종전 연구된 딥러닝 애플리케이션에 MGA 모듈을 적용하여 성능을 검증하였다. 종전 딥러닝 애플리케이션은 뇌 MRI를 해석하여 뇌 연령(brain age)을 예측하는 기술이다. 연구자는 공개 데이터 세트(OpenNeuro dataset, COBRE dataset 등)를 이용하여 2,788명의 건강한 모집단에 대한 3차원 T1-가중 MR 영상을 수집하였다. 연구자는 수집한 영상에 N4 편향 필드 보정 및 정규화를 진행하였다. 연구자는 수집한 데이터 세트에서 70%는 학습 데이터로 이용하고, 15%는 검증 데이터로 이용하고, 나머지 15%는 테스트 데이터로 이용하였다.

[0121] 연구자는 ResNet18에 공간 SE(spatial squeeze-and-excitation) 모듈을 부가한 모델(이하 sSE-ResNet18라 함)을 백본 (backbone) 으로 삼아 MGA를 추가한 모델 (이하 MGA-sSE-ResNet18 라 함)을 구축하였다. 도 7은 뇌 MRI로부터 뇌 연령을 예측하는 딥러닝 모델의 예이다. 도 7이 연구자가 구축한 모델이다. 해당 딥러닝 모델은 최초 합성곱 계층과 전연결 계층 사이에 점선으로 표시한 블록을 8개 반복한 구조를 갖는다. 점선으로 표시한 블록은 연속된 2개의 합성곱 계층 및 MGA 혼합 모듈을 포함한다. MGA 혼합 모듈은 sSE 모듈과 MGA 모듈을 병렬로 배치한 형태이다.

[0122] 연구자는 구축한 MGA-sSE-ResNet18과 다른 모델들의 성능을 비교하였다. 다른 모델들은 MGA 모듈을 사용하지 않는 CNN 기반의 모델들이다. 다른 모델들은 대표적인 5개의 CNN 모델(ResNet18, sSE-ResNet18, DenseNet121 및 MobileNetV2) 및 1개의 뇌 연령 예측 모델(SFCN)로 구성된다. SFCN(simple-fully-convolutional-network)은 뇌 연령 예측 용도로 연구된 모델이다.

[0124] 다른 모델들도 MGA-sSE-ResNet18과 동일한 환경에서 동일한 데이터 세트를 이용하여 뇌 연령을 예측하도록 학습하였다. 아래 표 1은 MGA-sSE-ResNet18과 다른 모델들의 성능을 비교한 결과이다. 성능 지표는 MAE(mean absolute error), RMSE(root mean squared error) 및 PCC(Pearson correlation coefficient)를 사용하였다. 성능 평가 결과 MGA-sSE-ResNet18이 검증에 사용된 모델들 중 가장 낮은 에러를 보였다.

표 1

[0126]

Model	MAE	RMSE	PCC
ResNet18	3.249	4.499	0.948
sSE-ResNet18	3.239	4.296	0.956
DenseNet121	3.340	4.437	0.961
MobileNetV2	3.295	4.533	0.950
SFCN	3.233	4.406	0.949
MGA-sSE-ResNet18	2.822	4.195	0.968

[0128] 도 8은 MGA 모듈을 사용하여 전역적 특징을 추출하는 데이터 처리 장치(500)의 예이다. 데이터 처리 장치(500)는 MGA 모듈을 포함하는 딥러닝 모델을 이용하여 입력 데이터를 처리하는 장치이다. 데이터 처리 장치(500)는 딥러닝 모델을 이용하여 일정한 서비스를 제공하는 장치일 수 있다. 데이터 처리 장치(500)는 물리적으로 다양한 형태로 구현될 수 있다. 예컨대, 데이터 처리 장치(500)는 PC와 같은 컴퓨터 장치, 네트워크의 서버, 데이터 처리 전용 칩셋 등의 형태를 가질 수 있다.

[0129] 데이터 처리 장치(500)는 저장장치(510), 메모리(520), 연산장치(530), 인터페이스 장치(540), 통신장치(550) 및 출력장치(560)를 포함할 수 있다.

[0130] 저장장치(510)는 MGA 모듈을 저장할 수 있다.

[0131] 저장장치(510)는 MGA 모듈을 포함하는 학습된 딥러닝 모델을 저장할 수 있다.

[0132] 저장장치(510)는 MGA 모듈을 통하여 전역 특징을 추출하는 과정을 제어하는 코드 내지 프로그램을 저장할 수 있다.

[0133] 메모리(520)는 데이터 처리 장치(500)가 MGA 모듈을 통하여 전역 특징을 추출하는 과정 및 딥러닝 모델이 동작하는 과정에서 생성되는 텐서, 데이터 및 정보 등을 저장할 수 있다.

[0134] 인터페이스 장치(540)는 외부로부터 일정한 명령 및 데이터를 입력받는 장치이다.

[0135] 인터페이스 장치(540)는 물리적으로 연결된 입력 장치 또는 외부 저장장치로부터 입력 데이터를 입력받을 수 있다.

[0136] 인터페이스 장치(540)는 물리적으로 연결된 입력 장치 또는 외부 저장장치로터 MGA 모듈에 대한 하이퍼파라미터를 입력받을 수 있다.

[0137] 인터페이스 장치(540)는 MGA 모듈이 업데이트한 특징 텐서를 딥러닝 모델의 다음 계층에 전달할 수 있다.

[0138] 인터페이스 장치(540)는 입력데이터를 해석하는 딥러닝 모델이 출력하는 결과를 외부 객체에 전달할 수도 있다.

[0139] 통신장치(550)는 유선 또는 무선 네트워크를 통해 일정한 정보를 수신하고 전송하는 구성을 의미한다.

[0140] 통신장치(550)는 외부 객체로부터 입력 데이터를 수신할 수 있다.

[0141] 통신장치(550)는 외부 객체로부터 MGA 모듈에 대한 하이퍼파라미터를 수신할 수 있다.

[0142] 통신장치(550)는 입력데이터를 해석하는 딥러닝 모델이 출력하는 결과를 외부 객체에 송신할 수 있다.

[0143] 인터페이스 장치(540)는 통신장치(550)를 통해 수신한 데이터 및 정보를 데이터 처리 장치(500) 내부로 전달받는 장치일 수도 있다.

[0144] 출력장치(560)는 일정한 정보를 출력하는 장치이다. 출력장치(560)는 데이터 처리 과정에 필요한 인터페이스, 딥러닝 모델의 분석 결과 등을 출력할 수 있다.

[0145] 연산 장치(530)는 입력 데이터를 일정하게 전처리할 수 있다. 예컨대, 입력 데이터를 벡터화하거나 정규화할 수 있다.

[0146] 연산 장치(530)는 딥러닝 모델의 합성곱 계층에 입력 데이터를 전달하여 특징 텐서를 생성할 수 있다. 또는, 연산 장치(530)는 딥러닝 모델의 이전 합성곱 계층서 출력하는 특징 텐서를 다음 합성곱 계층에 전달하여 다른 특징 텐서를 생성할 수 있다. 또는, 연산 장치(530)는 딥러닝 모델의 다른 모듈(예컨대, SE 모듈)이 출력하는 특

징 텐서를 합성곱 계층에 전달하여 특징 텐서를 업데이트할 수도 있다.

[0147] 연산 장치(530)는 일정한 특징 텐서를 입력받아 처리하는 MGA 모듈의 동작을 제어할 수 있다.

[0148] 연산 장치(530)는 입력되는 특징 텐서를 일정한 크기의 패치로 분할할 수 있다(패치 임베딩). 연산 장치(530)는 하나의 특징 텐서를 다수의 패치들(패치 세트 P)로 분할할 수 있다. 나아가, 연산 장치(530)는 패치 크기를 결정하는 하이퍼파라미터에 따라 동일한 특징 텐서를 서로 다른 개수(서로 다른 크기)의 패치들로 분할할 수 있다.

[0149] 연산 장치(530)는 분할된 하나의 패치에 대한 차원을 축소할 수 있다. 예컨대, 연산 장치(530)는 패치에 대한 전역 평균 풀링(GAP)을 수행할 수 있다. 연산 장치(530)는 패치에 대한 전역 최대 풀링(GMP)을 수행할 수 있다. 나아가, 연산 장치(530)는 서로 다른 풀링 기법으로 차원을 축소한 패치들을 결합(concatenation)할 수 있다. 연산 장치(530)는 패치 세트에 속한 각 패치에 대한 차원 축소 및 결합을 수행할 수 있다. 이 과정을 통해 연산 장치(530)는 특징 텐서에 대하여 차원이 축소된 패치 세트 H를 생성할 수 있다.

[0150] 연산 장치(530)는 패치 세트 H에 속한 패치들을 이용하여 그래프를 구축할 수 있다. 연산 장치(530)는 패치 세트 H에 속한 패치들 각각을 하나의 노드로 설정하고, 노드 사이의 에지의 연결성을 결정한다.

[0151] 연산 장치(530)는 두 개의 노드들 사이의 유클리디언 거리 유사도 점수를 해당 노드들 사이의 에지 값으로 결정한다. 연산 장치(530)는 가능한 모든 노드 쌍 사이의 유클리디언 거리 유사도 점수를 연산하여 초기 에지 행렬을 구성한다.

[0152] 연산 장치(530)는 에지 행렬이 마르코프 성질을 지닌다는 가정하에 멀티홉 거리의 노드들 사이의 관계성을 파악하기 위하여 통계적 모델을 사용하여 초기 에지 행렬을 업데이트한다. 이 과정은 전술한 수학식 5부터 수학식 8까지 설명한 바와 같다. 연산 장치(530)는 수학식 8과 같은 멀티홉 에지 행렬을 생성한다.

[0153] 연산 장치(530)는 멀티홉 에지 행렬로 정의되는 그래프에 대한 어텐션을 수행한다. 연산 장치(530)는 수학식 9와 같이 어텐션 계수를 연산하고, 연산된 어텐션 계수를 이용하여 그래프 어텐션을 수행한다. 이를 통해 연산 장치(530)는 멀리 있는 노드들 간의 관계성(전역 특징)이 반영된 업데이트된 특징 텐서(패치 세트)를 산출할 수 있다.

[0154] MGA 모듈이 다수의 어텐션 가지들을 갖는다면, 연산 장치(530)는 각 어텐션 가지마다 동일한 동작을 수행하여 각각 업데이트된 특징 텐서를 산출한다. 이 경우, 연산 장치(530)는 전체 어텐션 가지에서 산출되는 특징 텐서를 평균하여 최종 특징 텐서를 산출할 수 있다.

[0155] 연산 장치(530)는 최종 특징 텐서와 직전 합성곱 계층이 출력하는 특징 텐서를 결합할 수 있다.

[0156] 또한, 연산 장치(530)는 MGA 모듈의 동작으로 산출되는 최종 특징 텐서와 다른 기능을 수행하는 모듈(예컨대, SE 모듈)이 출력하는 특징 텐서를 결합할 수도 있다.

[0157] 연산 장치(530)는 데이터를 처리하고, 일정한 연산을 처리하는 프로세서, AP, 프로그램이 임베디드된 칩과 같은 장치일 수 있다.

[0159] 또한, 상술한 바와 같은 MGA 모듈의 동작 방법, 입력 데이터의 전역 특징 추출 방법, 딥러닝 모델의 동작 방법 등은 컴퓨터에서 실행될 수 있는 실행가능한 알고리즘을 포함하는 프로그램(또는 어플리케이션)으로 구현될 수 있다. 상기 프로그램은 일시적 또는 비일시적 판독 가능 매체(non-transitory computer readable medium)에 저장되어 제공될 수 있다.

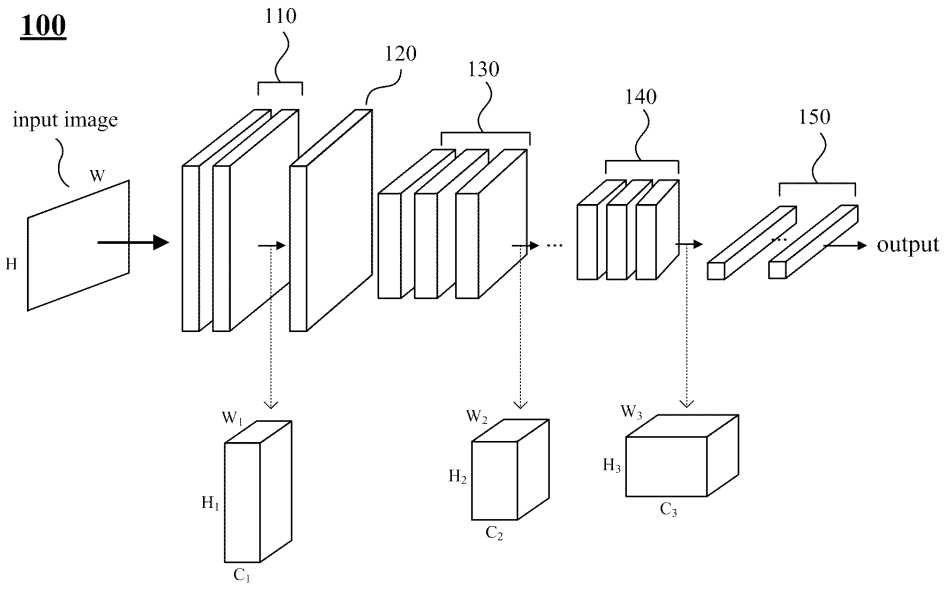
[0160] 비일시적 판독 가능 매체란 레지스터, 캐쉬, 메모리 등과 같이 짧은 순간 동안 데이터를 저장하는 매체가 아니라 반영구적으로 데이터를 저장하며, 기기에 의해 판독(reading)이 가능한 매체를 의미한다. 구체적으로는, 상술한 다양한 어플리케이션 또는 프로그램들은 CD, DVD, 하드 디스크, 블루레이 디스크, USB, 메모리카드, ROM (read-only memory), PROM (programmable read only memory), EPROM(Erasable PROM, EPROM) 또는 EEPROM(Electrically EPROM) 또는 플래시 메모리 등과 같은 비일시적 판독 가능 매체에 저장되어 제공될 수 있다.

[0161] 일시적 판독 가능 매체는 스태틱 램(Static RAM, SRAM), 다이내믹 램(Dynamic RAM, DRAM), 싱크로너스 디램(Synchronous DRAM, SDRAM), 2배속 SDRAM(Double Data Rate SDRAM, DDR SDRAM), 증강형 SDRAM(Enhanced SDRAM, ESDRAM), 동기화 DRAM(Synclink DRAM, SLDRAM) 및 직접 램버스 램(Direct Rambus RAM, DRRAM) 과 같은 다양한 RAM을 의미한다.

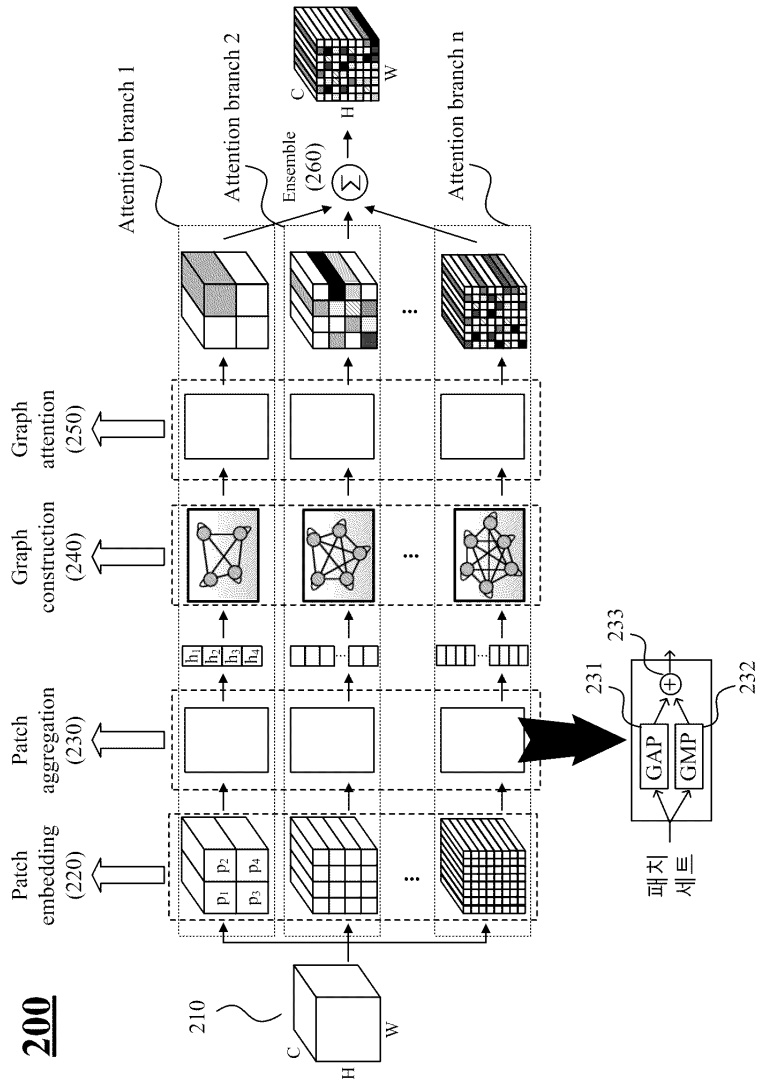
[0162] 본 실시예 및 본 명세서에 첨부된 도면은 전술한 기술에 포함되는 기술적 사상의 일부를 명확하게 나타내고 있는 것에 불과하며, 전술한 기술의 명세서 및 도면에 포함된 기술적 사상의 범위 내에서 당업자가 용이하게 유추할 수 있는 변형 예와 구체적인 실시례는 모두 전술한 기술의 권리범위에 포함되는 것이 자명하다고 할 것이다.

도면

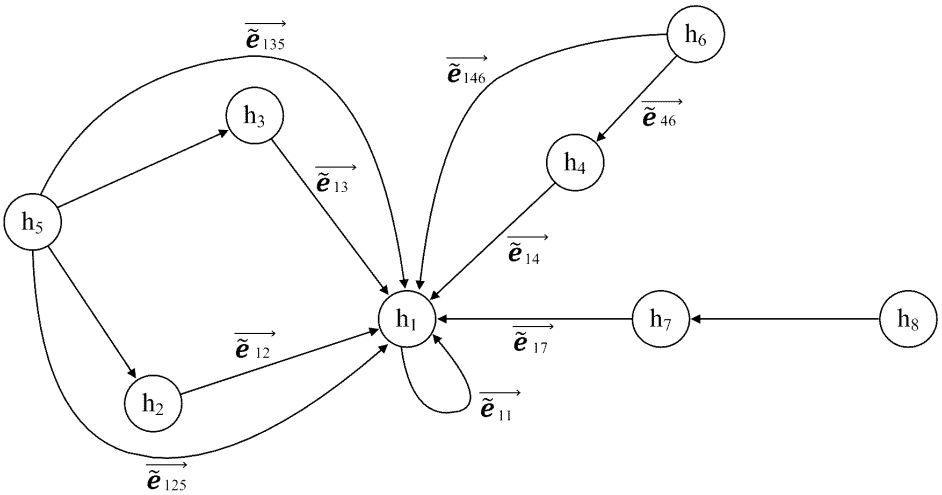
도면1



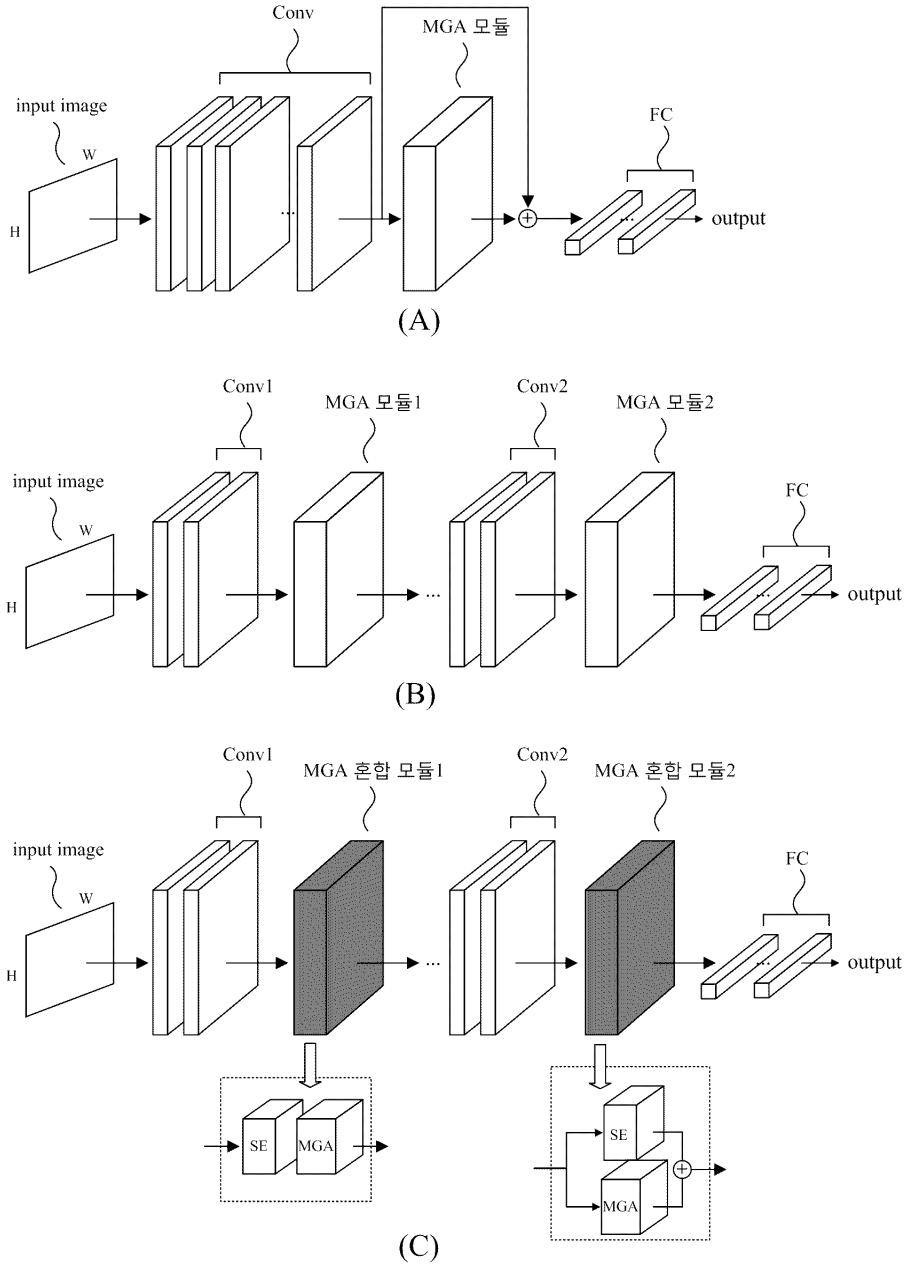
도면2


200
210
C
H
W
Patch embedding (220)
Patch aggregation (230)
Graph construction (240)
Graph attention (250)
Attention branch 1
Attention branch 2
Attention branch n
Ensemble (260)
231
232
233
GAP
GMP
패치 세트

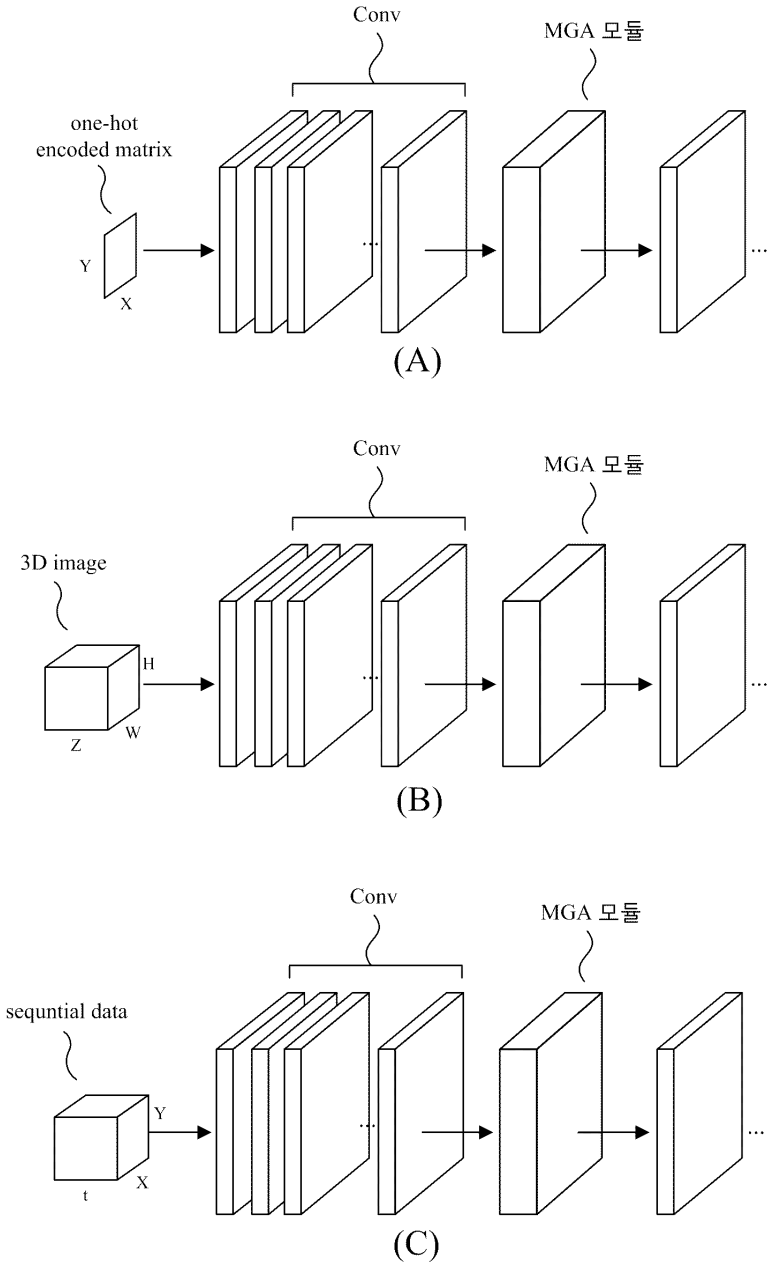
도면3


h1
h2
h3
h4
h5
h6
h7
h8

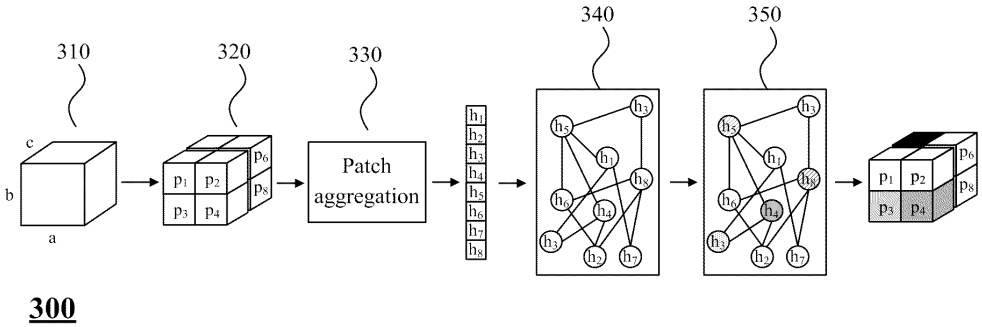
도면4


Conv
MGA 모듈
input image
W
H
FC
output
(A)
Conv1
MGA 모듈1
Conv2
MGA 모듈2
input image
W
H
FC
output
(B)
Conv1
MGA 혼합 모듈1
Conv2
MGA 혼합 모듈2
input image
W
H
FC
output
SE
MGA
SE
MGA
(C)

도면5

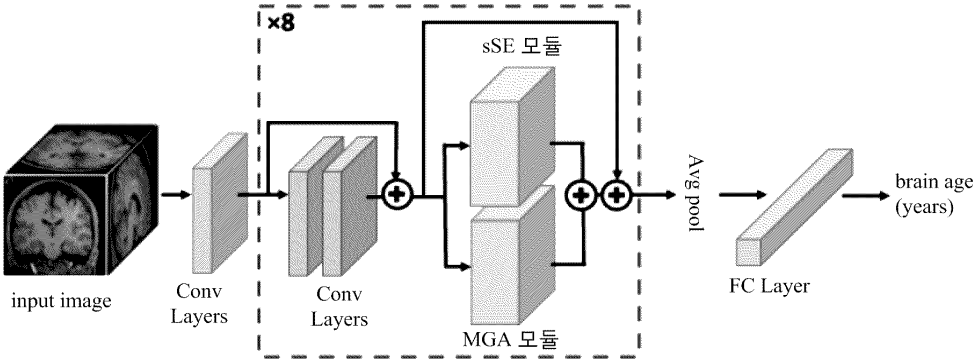

Conv
MGA 모듈
one-hot
encoded matrix
Y
X
(A)
Conv
MGA 모듈
3D image
H
Z
W
(B)
Conv
MGA 모듈
sequntial data
Y
t
X
(C)

도면6

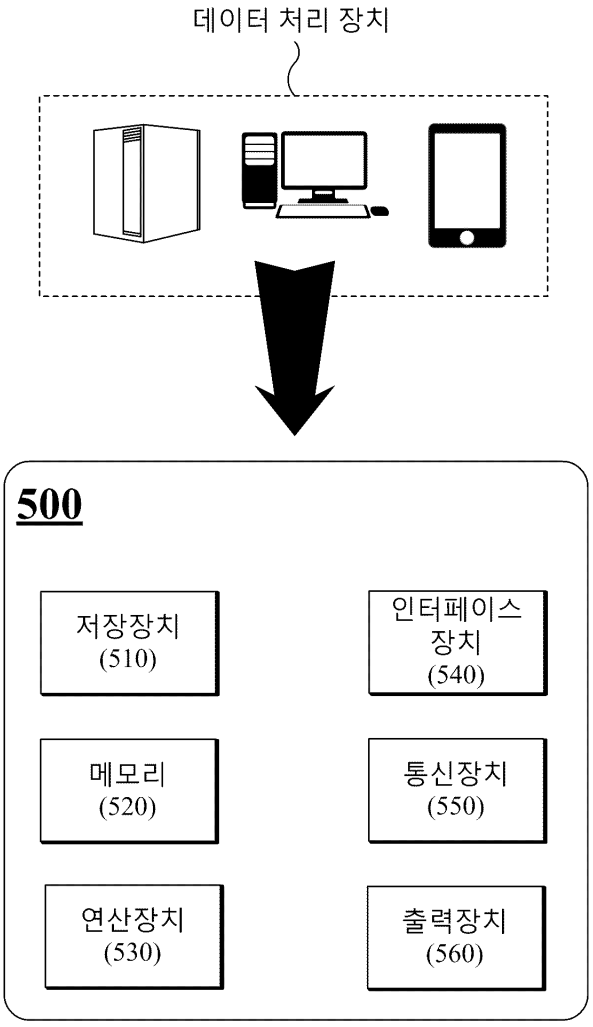

310
320
330
340
350
c
b
a
Patch
aggregation

300

도면7


×8
sSE 모듈
input image
Conv Layers
Conv Layers
MGA 모듈
Avg pool
FC Layer
brain age (years)

도면8


데이터 처리 장치
500
저장장치 (510)
인터페이스 장치 (540)
메모리 (520)
통신장치 (550)
연산장치 (530)
출력장치 (560)