(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0008045
(43) 공개일자 2024년01월18일

(51) 국제특허분류(Int. Cl.)
G06T 7/11 (2017.01) *G06T 3/40* (2024.01)
G06T 5/20 (2024.01) *G06T 7/00* (2017.01)
(52) CPC특허분류
G06T 7/11 (2017.01)
G06T 3/40 (2024.01)
(21) 출원번호 10-2022-0084911
(22) 출원일자 2022년07월11일
심사청구일자 2022년07월11일

(71) 출원인
세종대학교산학협력단
서울특별시 광진구 능동로 209 (군자동, 세종대학교)
(72) 발명자
유성준
서울특별시 광진구 능동로 209, 대양AI센터 717호(군자동)
구영현
서울특별시 광진구 능동로 209, 대양AI센터 717호(군자동)
박철호
서울특별시 광진구 능동로 209, 세종대학교 대양AI센터 717호(군자동)
(74) 대리인
양성보

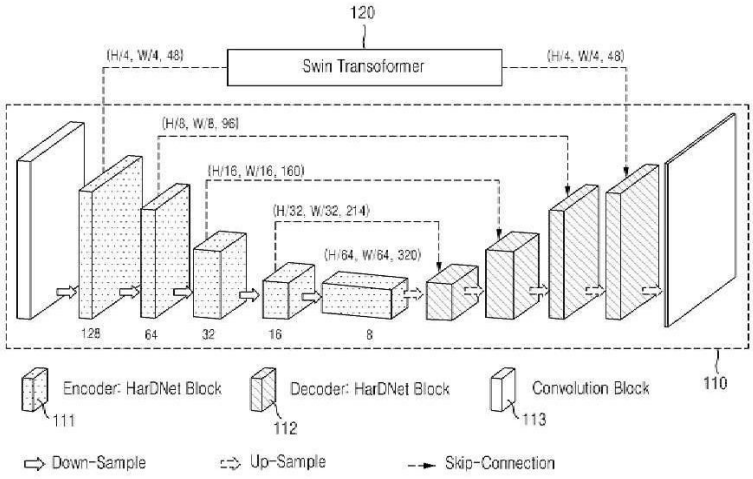
전체 청구항 수 : 총 14 항

(54) 발명의 명칭 **스윈 트랜스포머와 CNN을 결합한 의료영상 이미지 분할 방법 및 시스템**

(57) 요 약

스윈 트랜스포머와 CNN을 결합한 의료영상 이미지 분할 방법 및 시스템이 제시된다. 본 발명에서 제안하는 스윈 트랜스포머와 CNN을 결합한 의료영상 이미지 분할 시스템은 다운 샘플링을 수행하여 입력 이미지의 사이즈를 축소하고, 특징 맵을 추출하는 인코더 및 상기 특징 맵을 입력 이미지와 동일한 사이즈로 업 샘플링하고, 상기 특징 맵에서 질환 영역을 찾아 출력 이미지에 표기하는 디코더를 포함하는 인코더-디코더(encoder-decoder) 구조의 HarDNet 및 상기 인코더-디코더 구조에서 인코더와 디코더 각각에 대한 스킵 연결(skip connection)을 수행하고 입력 이미지로부터 계층적 특징 맵을 추출하고, 쉬프트 윈도우 방식(shifted windows approach)을 이용하여 상기 계층적 특징 맵에서 패치(patch) 간의 관계를 계산함으로써 이미지를 분할하는 스윈 트랜스포머(swin transformer)를 포함한다.

대 표 도



(52) CPC특허분류
G06T 5/20 (2024.01)
G06T 7/0012 (2013.01)
G06T 2207/20084 (2013.01)
G06T 2210/41 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711125819
과제번호	2019-0-00136-003
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	인공지능융합선도프로젝트(R&D)
연구과제명	스마트시티 산업 생산성 혁신을 위한 AI융합 기술 개발
기 여 율	1/1
과제수행기관명	세종대학교 산학협력단
연구기간	2021.01.01 ~ 2021.12.31

명 세 서

청구범위

청구항 1

다운 샘플링을 수행하여 입력 이미지의 사이즈를 축소하고, 특징 맵을 추출하는 인코더 및 상기 특징 맵을 입력 이미지와 동일한 사이즈로 업 샘플링하고, 상기 특징 맵에서 질환 영역을 찾아 출력 이미지에 표기하는 디코더를 포함하는 인코더-디코더(encoder-decoder) 구조의 HarDNet; 및

상기 인코더-디코더 구조에서 인코더와 디코더 각각에 대한 스킵 연결(skip connection)을 수행하고 입력 이미지로부터 계층적 특징 맵을 추출하고, 쉬프트 윈도우 방식(shifted windows approach)을 이용하여 상기 계층적 특징 맵에서 패치(patch) 간의 관계를 계산함으로써 이미지를 분할하는 스윈 트랜스포머(swin transformer)

를 포함하는 의료영상 이미지 분할 시스템.

청구항 2

제1항에 있어서,

상기 HarDNet은,

인코더(encoder)를 포함하고,

상기 인코더는 하나의 컨볼루션 블록(convolution block), 복수의 HardNet 블록 및 복수의 다운 샘플링 블록을 포함하는

의료영상 이미지 분할 시스템.

청구항 3

제2항에 있어서,

상기 하나의 컨볼루션 블록은,

필터=24, 커널 사이즈=3, 스트라이드=2 인 컨볼루션 레이어 및 필터=48, 커널 사이즈=3, 스트라이드=1 인 컨볼루션 레이어를 포함하고,

상기 복수의 다운 샘플링 블록은,

하나의 커널 사이즈=1 인 컨볼루션 레이어와 커널 사이즈=2, 스트라이드=2 인 AvgPoll2d 레이어를 포함하는

의료영상 이미지 분할 시스템.

청구항 4

제2항에 있어서,

상기 인코더의 전환 구간(bottleneck)은 하나의 HardNet 블록을 사용하여 인코더와 디코더 간의 파라미터를 전달하는

의료영상 이미지 분할 시스템.

청구항 5

제1항에 있어서,

상기 HarDNet은,

디코더(decoder)를 포함하고,

상기 디코더는 복수의 HardNet 블록, 복수의 업 샘플링 블록 및 하나의 하나의 컨볼루션 블록을 포함하고,

상기 디코더는 상기 인코더의 전환 구간으로부터 전달 받은 특징 맵을 입력 이미지와 동일한 사이즈로 업 샘플

링하고, 상기 특징 맵에서 질환 영역을 찾아 출력 이미지에 표기하여 최종 이미지를 출력하는

의료영상 이미지 분할 시스템.

청구항 6

제1항에 있어서,

상기 스윈 트랜스포머는,

패치 분할 레이어를 통해 입력 이미지를 패치로 분할하여 패치 토큰을 생성하고, 상기 패치 토큰은 선형 임베딩을 통과한 후 연결되는 두 개의 스윈 트랜스포머 블록에 입력되어 새로운 토큰을 생성하는

의료영상 이미지 분할 시스템.

청구항 7

제6항에 있어서,

상기 스윈 트랜스포머는,

상기 새로운 토큰은 패치 병합을 통해 인접한 패치를 하나의 패치로 병합하고, 상기 새로운 토큰을 다운 샘플링하는 과정을 반복하고,

선형 예측을 통해 특징 맵을 확장한 후 컨볼루션을 거쳐 입력 이미지와 통일한 형태의 특징 맵을 출력하는

의료영상 이미지 분할 시스템.

청구항 8

제6항에 있어서,

상기 스윈 트랜스포머는,

상기 HarDNet의 인코더의 처음 HardNet 블록에서 출력한 특징 맵을 입력으로 받아 상기 HarDNet의 디코더의 마지막 HardNet 블록에 연결되는

의료영상 이미지 분할 시스템.

청구항 9

HarDNet의 인코더가 다운 샘플링을 수행하여 입력 이미지의 사이즈를 축소하고, 특징 맵을 추출하는 단계;

HarDNet의 디코더가 상기 특징 맵을 입력 이미지와 동일한 사이즈로 업 샘플링하고, 상기 특징 맵에서 질환 영역을 찾아 출력 이미지에 표기하는 단계; 및

계층적 구조를 갖는 스윈 트랜스포머(swin transformer)를 통해 입력 이미지로부터 계층적 특징 맵을 추출하고, 쉬프트 윈도우 방식(shifted windows approach)을 이용하여 상기 계층적 특징 맵에서 패치(patch) 간의 관계를 계산함으로써 이미지를 분할하는 단계

를 포함하는 의료영상 이미지 분할 방법.

청구항 10

제9에 있어서,

상기 HarDNet의 인코더가 다운 샘플링을 수행하여 입력 이미지의 사이즈를 축소하고, 특징 맵을 추출하는 단계는 -상기 인코더는 하나의 컨볼루션 블록(convolution block), 복수의 HardNet 블록 및 복수의 다운 샘플링 블록을 포함함-,

상기 인코더의 전환 구간(bottleneck)이 하나의 HardNet 블록을 사용하여 인코더와 디코더 간의 파라미터를 전달하는

의료영상 이미지 분할 방법.

청구항 11

제9에 있어서,

상기 HarDNet의 디코더가 상기 특징 맵을 입력 이미지와 동일한 사이즈로 업 샘플링하고, 상기 특징 맵에서 질환 영역을 찾아 출력 이미지에 표기하는 단계는 -상기 디코더는 복수의 HardNet 블록, 복수의 업 샘플링 블록 및 하나의 하나의 컨볼루션 블록을 포함함-,

상기 디코더는 상기 인코더의 전환 구간으로부터 전달 받은 특징 맵을 입력 이미지와 동일한 사이즈로 업 샘플링하고, 상기 특징 맵에서 질환 영역을 찾아 출력 이미지에 표기하여 최종 이미지를 출력하는

의료영상 이미지 분할 방법.

청구항 12

제9항에 있어서,

상기 계층적 구조를 갖는 스윈 트랜스포머(swin transformer)를 통해 입력 이미지로부터 계층적 특징 맵을 추출하고, 쉬프트 윈도우 방식(shifted windows approach)을 이용하여 상기 계층적 특징 맵에서 패치(patch) 간의 관계를 계산함으로써 이미지를 분할하는 단계는,

패치 분할 레이어를 통해 입력 이미지를 패치로 분할하여 패치 토큰을 생성하고, 상기 패치 토큰은 선형 임베딩을 통과한 후 연결되는 두 개의 스윈 트랜스포머 블록에 입력되어 새로운 토큰을 생성하는 단계

를 포함하는 의료영상 이미지 분할 방법.

청구항 13

제12항에 있어서,

상기 새로운 토큰은 패치 병합을 통해 인접한 패치를 하나의 패치로 병합하고, 상기 새로운 토큰을 다운 샘플링하는 과정을 반복하는 단계; 및

선형 예측을 통해 특징 맵을 확장한 후 컨볼루션을 거쳐 입력 이미지와 통일한 형태의 특징 맵을 출력하는 단계

를 더 포함하는 의료영상 이미지 분할 방법.

청구항 14

제12항에 있어서,

상기 HarDNet의 인코더의 처음 HardNet 블록에서 출력한 특징 맵을 입력으로 받아 상기 HarDNet의 디코더의 마지막 HardNet 블록에 연결되는

의료영상 이미지 분할 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 스윈 트랜스포머와 CNN을 결합한 의료영상 이미지 분할 방법 및 시스템에 관한 것이다.

배 경 기 술

[0002] 뇌졸중은 발병률, 사망률 및 장애 유발 가능성이 높아 인류의 건강에 위협을 주는 질병이다. 뇌졸중은 여러 첨단 검사 방법을 이용해 진단 할 수 있는데 그 중 주로 뇌전산화 단층 촬영(CT)이나 자기공명영상(MRI) 검사가 많이 사용된다. CT 검사는 급성기 뇌경색과 뇌출혈을 구분하는데 가장 좋은 방법으로, 초기 치료를 결정하기 위해 뇌졸증이 의심된 환자에게서 제일 먼저 시행된다. 뇌경색인 경우 저음영으로 나타나고 뇌출혈인 경우 고음영으로 나타나는데 뇌경색의 초기에는 경색 부위가 잘 나타나지 않는다. MRI 검사는 CT와 비슷하나 영상력이 훨씬 뛰어나 CT에서 찾기 힘들 정도의 작은 병변이나 뇌간 부위의 병소를 정확히 찾을 수 있는 장점이 있다. 또한 MRI 영상은 3D 영상으로 CT 영상 보다 해상도 및 가독성이 더 높다는 장점이 있다. 의사들의 진단을 돕기 위해 컴퓨터 비전(computer vision) 기술을 이용한 뇌졸증 진단 연구가 많이 이루어지고 있다[1,2,3]. 기존의 CT 또

는 MRI 영상 기반 진단은 CNN 구조를 이용한 인코더-디코더(encoder-decoder) 구조를 가진 U-Net[4] 형태의 모델을 많이 사용해 좋은 결과를 얻었다.

[0003] 최근 트랜스포머(transformer)가 컴퓨터 비전 분야에 적용되면서 트랜스포머를 이용한 분할(segmentation) 모델들도 많이 제안되고 있다[5,6]. 트랜스포머는 NLP 분석에서 시퀀스(sequence) 데이터를 처리하는 방법을 컴퓨터 비전 분야에 적용한 성공적인 사례이다. 현재 트랜스포머는 탐지(detection)[7], 분할(segmentation)[8], 분류(classification)[9] 등의 컴퓨터 비전 분야에서 좋은 성능을 보이고 있다. 또한 CNN 모델과 ViT(Vision Transformer) 모델을 결합한 인코더-디코더(encoder-decoder) 구조를 가진 모델들도 의료 영상 분석에서 좋은 성과를 보이고 있다[10,11]. 현재 많은 CNN과 ViT의 결합 모델에서는 트랜스포머를 CNN 기반 인코더-디코더 모델의 보틀넥(bottleneck)에 사용하여 인코더 와 디코더의 파라미터 전달이 더 효과적으로 진행되게 한다[12,13]. 하지만 CNN 기반 인코더-디코더 모델은 모델 사이즈가 크고 계산 속도가 느리며, ViT 모델은 이미지 사이즈가 커짐에 따라 모델의 계산 양이 급증하는 문제가 있다. 기존 CNN과 ViT의 결합 모델에서 트랜스포머를 보틀넥에 사용한 것은 입력 특징 맵(feature map)의 사이즈가 커짐에 따라 트랜스포머가 모델의 전반적인 계산 속도에 주는 영향을 최소화 하기 위한 이유도 있다.

발명의 내용

해결하려는 과제

[0004] 본 발명이 이루고자 하는 기술적 과제는 CNN 모델 중 경량화 모델 구조인 HarDNet 블록과 ViT 모델에서 이미지 사이즈가 커짐에 따라 트랜스포머 모델의 계산 양이 급증하는 문제를 해결한 스윈 트랜스포머(swin transformer) 모델을 결합해 기존의 문제점을 해결함과 동시에 높은 성능을 얻기 위한 스윈 트랜스포머와 CNN을 결합한 의료영상 이미지 분할 방법 및 시스템을 제공하는데 있다.

과제의 해결 수단

[0005] 일 측면에 있어서, 본 발명에서 제안하는 스윈 트랜스포머와 CNN을 결합한 의료영상 이미지 분할 시스템은 다운 샘플링을 수행하여 입력 이미지의 사이즈를 축소하고, 특징 맵을 추출하는 인코더 및 상기 특징 맵을 입력 이미지와 동일한 사이즈로 업 샘플링하고, 상기 특징 맵에서 질환 영역을 찾아 출력 이미지에 표기하는 디코더를 포함하는 인코더-디코더(encoder-decoder) 구조의 HarDNet 및 상기 인코더-디코더 구조에서 인코더와 디코더 각각에 대한 스킵 연결(skip connection)을 수행하고 입력 이미지로부터 계층적 특징 맵을 추출하고, 쉬프트 윈도우 방식(shifted windows approach)을 이용하여 상기 계층적 특징 맵에서 패치(patch) 간의 관계를 계산함으로써 이미지를 분할하는 스윈 트랜스포머(swin transformer)를 포함한다.

[0006] 상기 HarDNet은 인코더(encoder)를 포함하고, 상기 인코더는 하나의 컨볼루션 블록(convolution block), 복수의 HardNet 블록 및 복수의 다운 샘플링 블록을 포함한다.

[0007] 상기 하나의 컨볼루션 블록은 필터=24, 커널 사이즈=3, 스트라이드=2 인 컨볼루션 레이어 및 필터=48, 커널 사이즈=3, 스트라이드=1 인 컨볼루션 레이어를 포함하고, 상기 복수의 다운 샘플링 블록은 하나의 커널 사이즈=1 인 컨볼루션 레이어와 커널 사이즈=2, 스트라이드=2 인 AvgPoll2d 레이어를 포함한다.

[0008] 상기 인코더의 전환 구간(bottleneck)은 하나의 HardNet 블록을 사용하여 인코더와 디코더 간의 파라미터를 전달한다.

[0009] 상기 HarDNet은 디코더(decoder)를 포함하고, 상기 디코더는 복수의 HardNet 블록, 복수의 업 샘플링 블록 및 하나의 하나의 컨볼루션 블록을 포함하고, 상기 디코더는 상기 인코더의 전환 구간으로부터 전달 받은 특징 맵을 입력 이미지와 동일한 사이즈로 업 샘플링하고, 상기 특징 맵에서 질환 영역을 찾아 출력 이미지에 표기하여 최종 이미지를 출력한다.

[0010] 상기 스윈 트랜스포머는 패치 분할 레이어를 통해 입력 이미지를 패치로 분할하여 패치 토큰을 생성하고, 상기 패치 토큰은 선형 임베딩을 통과한 후 연결되는 두 개의 스윈 트랜스포머 블록에 입력되어 새로운 토큰을 생성한다.

[0011] 상기 스윈 트랜스포머는 상기 새로운 토큰은 패치 병합을 통해 인접한 패치를 하나의 패치로 병합하고, 상기 새로운 토큰을 다운 샘플링하는 과정을 반복하고, 선형 예측을 통해 특징 맵을 확장한 후 컨볼루션을 거쳐 입력 이미지와 통일한 형태의 특징 맵을 출력한다.

[0012] 상기 스윈 트랜스포머는 상기 HarDNet의 인코더의 처음 HardNet 블록에서 출력한 특징 맵을 입력으로 받아 상기 HarDNet의 디코더의 마지막 HardNet 블록에 연결된다.

[0013] 또 다른 일 측면에 있어서, 본 발명에서 제안하는 스윈 트랜스포머와 CNN을 결합한 의료영상 이미지 분할 방법은 HarDNet의 인코더가 다운 샘플링을 수행하여 입력 이미지의 사이즈를 축소하고, 특징 맵을 추출하는 단계, HarDNet의 디코더가 상기 특징 맵을 입력 이미지와 동일한 사이즈로 업 샘플링하고, 상기 특징 맵에서 질환 영역을 찾아 출력 이미지에 표기하는 단계 및 계층적 구조를 갖는 스윈 트랜스포머(swin transformer)를 통해 입력 이미지로부터 계층적 특징 맵을 추출하고, 쉬프트 윈도우 방식(shifted windows approach)을 이용하여 상기 계층적 특징 맵에서 패치(patch) 간의 관계를 계산함으로써 이미지를 분할하는 단계를 포함한다.

발명의 효과

[0014] 본 발명의 실시예들에 따르면 CNN 모델 중 경량화 모델 구조인 HarDNet 블록과 ViT 모델에서 이미지 사이즈가 커짐에 따라 트랜스포머 모델의 계산 양이 급증하는 문제를 해결한 스윈 트랜스포머(swin transformer) 모델을 결합해 기존의 문제점을 해결함과 동시에 높은 성능을 얻을 수 있다. 본 발명의 실시예들에 따르면 HarDNet 블록[14]으로 U-Net 형태의 인코더-디코더(encoder-decoder) 모델을 구축하고, 인코더 모델의 처음 스킵 연결 레이어(skip connection layer)에 스윈 트랜스포머 모델을 사용함으로써 스윈 트랜스포머 모델은 보틀넥에 사용할 때 보다 큰 특징 맵(feature map)에 적용되어 셀프-어텐션 기반 쉬프트 윈도우 접근 방식(shifted windows approach base on self-attention) 및 계층별 특징 추출의 장점을 유지할 수 있다.

도면의 간단한 설명

[0015] 도 1은 본 발명의 일 실시예에 따른 스윈 트랜스포머와 CNN을 결합한 의료영상 이미지 분할 시스템의 구성을 나타내는 도면이다.

도 2는 본 발명의 일 실시예에 따른 인코더-디코더 구조를 포함하는 HarDNet의 구성을 나타내는 도면이다.

도 3은 본 발명의 일 실시예에 따른 스윈 트랜스포머를 설명하기 위한 도면이다.

도 4는 본 발명의 일 실시예에 따른 스윈 트랜스포머와 CNN을 결합한 의료영상 이미지 분할 방법을 설명하기 위한 흐름도이다.

도 5는 본 발명의 일 실시예에 따른 데이터 셋을 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0016] 본 발명에서는 CNN 모델 중 경량화 모델 구조인 HarDNet 블록과 ViT 모델에서 이미지 사이즈가 커짐에 따라 트랜스포머 모델의 계산 양이 급증하는 문제를 해결한 스윈 트랜스포머(swin transformer) 모델을 결합해 기존의 문제점을 해결함과 동시에 높은 성능을 얻고자 한다. 이를 위해 우선 HarDNet 블록[14]으로 U-Net 형태의 인코더-디코더(encoder-decoder) 모델을 구축한다. 그리고 기존 연구에서 인코더-디코더의 보틀넥에 트랜스포머 레이어(transformer layer)를 사용하는 방법과 달리 인코더 모델의 처음 스킵 연결 레이어(skip connection layer)에 스윈 트랜스포머[15] 모델을 사용한다. 이로서 스윈 트랜스포머 모델은 보틀넥에 사용할 때 보다 큰 특징 맵(feature map)에 적용되어 셀프-어텐션 기반 쉬프트 윈도우 접근 방식(shifted windows approach base on self-attention) 및 계층별 특징 추출의 장점을 유지한다. 본 발명에서 제안하는 STHarDNet 모델은 다음과 같은 특성이 있다.

[0017] 본 발명의 실시예에 따르면, HarDNet 블록을 사용해 기존 CNN 기반 인코더-디코더 구조 모델의 계산 속도가 느린 단점을 개선한 스윈 트랜스포머를 사용하여 ViT 모델에서 이미지의 사이즈가 커짐에 따라 메모리 사용량 및 계산량이 급증하는 문제를 해결한다.

[0018] 본 발명의 실시예에 따르면, U-Net구조를 가진 인코더-디코더 모델에서 스윈 트랜스포머 모델을 인코더의 처음 스킵 연결 레이어에 사용해 보틀넥보다 큰 특징 맵을 입력 받아 스윈 트랜스포머 모델의 셀프-어텐션 기반 쉬프트 윈도우 접근 방식 및 계층적 트랜스포머(hierarchical transformer)의 특징을 유지한다.

[0019] 본 발명에서는 제안하는 STHarDNet의 우수성을 증명하기 위해 CT및 MRI 영상에서 기존 뇌질환 분할 SOTA 모델들과 성능비교 실험을 했다. CT 영상은 서울성모병원 외 4개 기관에서 제공한 5개자 뇌출혈 질환을 포함한 82,636장 영상 데이터를 사용했다. 기존 SOTA 모델과 성능 비교 실험 결과 STHarDNet의 dice, IoU, precision,

Recall은 각각 0.728, 0.612, 0.791, 0.718로 제일 우수했다.

[0020] MRI 스캔 영상은 ATLAS(Anatomical Tracings of Lesions After Stroke) 데이터셋을 사용했다. ATLAS 데이터셋은 229명 뇌졸중 환자의 MRI 영상에서 수동으로 병변 위치를 분할한 다양한 알고리즘 성능 비교를 위해 구축된 표준화된 오픈 데이터셋(open dataset)이다. ATLAS 데이터에서 177명 환자의 MRI 영상을 학습데이터로 사용하고, 나머지 52명 환자의 데이터를 검증 데이터로 사용해 기존 SOTA 모델과 성능 비교 실험을 진행한 결과, 제안하는 모델의 Dice, IoU, Precision, Recall은 각각 0.5470, 0.4174, 0.7279, 0.4902로 기존 모델에 비해 우수했다. 마지막 모델의 계산 속도 비교 실험을 한결과 STHarDNet은 10만장 이미지를 계산할 때 FPS가 299.943으로 HarDNet 보다 2.48% 느려 두 번째로 빠른 성능을 보였다. 이하, 본 발명의 실시 예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0022] 도 1은 본 발명의 일 실시예에 따른 스윈 트랜스포머와 CNN을 결합한 의료영상 이미지 분할 시스템의 구성을 나타내는 도면이다.

[0023] 제안하는 스윈 트랜스포머와 CNN을 결합한 의료영상 이미지 분할 시스템은 u-net 형태의 인코더-디코더 구조를 가진 HarDNet(110) 및 인코더와 디코더를 연결하는 스킵 연결에 사용되는 스윈 트랜스포머(swin transformer)(120)를 포함한다.

[0024] 본 발명의 실시예에 따른 HarDNet(110)는 다운 샘플링을 수행하여 입력 이미지의 사이즈를 축소하고, 특징 맵을 추출하는 인코더 및 상기 특징 맵을 입력 이미지와 동일한 사이즈로 업 샘플링하고, 상기 특징 맵에서 질환 영역을 찾아 출력 이미지에 표기하는 디코더를 포함하는 인코더-디코더(encoder-decoder) 구조를 갖는다.

[0025] 본 발명의 실시예에 따른 HarDNet(110)는 인코더를 포함한다.

[0026] 본 발명의 실시예에 따른 인코더는 하나의 컨볼루션 블록(convolution block)(113), 복수의 HardNet 블록(111) 및 복수의 다운 샘플링 블록을 포함한다.

[0027] 본 발명의 실시예에 따른 하나의 컨볼루션 블록(113)은 필터=24, 커널 사이즈=3, 스트라이드=2 인 컨볼루션 레이어 및 필터=48, 커널 사이즈=3, 스트라이드=1 인 컨볼루션 레이어를 포함할 수 있다.

[0028] 본 발명의 실시예에 따른 복수의 다운 샘플링 블록은 하나의 커널 사이즈=1 인 컨볼루션 레이어와 커널 사이즈=2, 스트라이드=2 인 AvgPoll2d 레이어를 포함할 수 있다.

[0029] 본 발명의 실시예에 따른 인코더(111)의 전환 구간(bottleneck)은 하나의 HardNet 블록을 사용하여 인코더와 디코더 간의 파라미터를 전달할 수 있다.

[0030] 본 발명의 실시예에 따른 HarDNet(110)는 디코더를 포함한다.

[0031] 본 발명의 실시예에 따른 디코더는 복수의 HardNet 블록(112), 복수의 업 샘플링 블록 및 하나의 하나의 컨볼루션 블록(113)을 포함한다.

[0032] 본 발명의 실시예에 따른 디코더는 상기 인코더의 전환 구간으로부터 전달 받은 특징 맵을 입력 이미지와 동일한 사이즈로 업 샘플링하고, 상기 특징 맵에서 질환 영역을 찾아 출력 이미지에 표기하여 최종 이미지를 출력한다.

[0033] 본 발명의 실시예에 따른 스윈 트랜스포머(120)는 상기 인코더-디코더 구조에서 인코더와 디코더 각각에 대한 스킵 연결(skip connection)을 수행하고 입력 이미지로부터 계층적 특징 맵을 추출하고, 쉬프트 윈도우 방식(shifted windows approach)을 이용하여 상기 계층적 특징 맵에서 패치(patch) 간의 관계를 계산함으로써 이미지를 분할한다.

[0034] 하지만 계층적 트랜스포머(hierarchical transformer)는 인접한 2×2의 패치를 병합하면서 특징 맵의 사이즈를 다운 샘플링하여 계층적 특징 맵을 생성한다. 그러므로 입력한 특징 맵의 사이즈가 작으면 스윈 트랜스포머는 깊은 계층적 특징 맵을 추출할 수 없다. 그런데 기존 CNN과 특징 맵의 결합 모델에서는 트랜스포머를 인코더와 디코더의 전환 구간(bottleneck)에 사용했다. 하지만 전환 구간은 인코더의 제일 마지막에서 출력한 특징 맵을 입력으로 받으므로 사이즈가 작다(예를 들어, 7×7). 이는 본 발명에서 스윈 트랜스포머의 계층적 트랜스포머를 사용하는 목적에 적합하지 않는다. 그래서 본 발명에서는 스윈 트랜스포머를 인코더와 디코더의 전환 구간에서 사용하지 않고 스킵 연결에서 특징 맵의 사이즈가 제일 큰 처음 스킵 연결에 사용한다.

[0035] 본 발명의 실시예에 따른 스윈 트랜스포머(120)는 패치 분할 레이어를 통해 입력 이미지를 패치로 분할하여 패치 토큰을 생성하고, 상기 패치 토큰은 선형 임베딩을 통과한 후 연결되는 두 개의 스윈 트랜스포머 블록에 입력되어 새로운 토큰을 생성한다.

[0036] 이후, 상기 새로운 토큰은 패치 병합을 통해 인접한 패치를 하나의 패치로 병합하고, 상기 새로운 토큰을 다운 샘플링하는 과정을 반복하고, 선형 예측을 통해 특징 맵을 확장한 후 컨볼루션을 거쳐 입력 이미지와 통일한 형태의 특징 맵을 출력한다.

[0037] 본 발명의 실시예에 따른 스윈 트랜스포머(120)는 상기 HarDNet의 인코더의 처음 HardNet 블록에서 출력한 특징 맵을 입력으로 받아 상기 HarDNet의 디코더의 마지막 HardNet 블록에 연결된다. 도 2 및 도 3을 참조하여 본 발명의 실시예에 따른 스윈 트랜스포머와 CNN을 결합한 의료영상 이미지 분할 시스템의 구성을 더욱 상세히 설명한다.

[0039] 도 2는 본 발명의 일 실시예에 따른 인코더-디코더 구조를 포함하는 HarDNet의 구성을 나타내는 도면이다.

[0040] 본 발명의 실시예에 따른 HarDNet은 HardNet 블록을 백본(backbone)으로 구축된 인코더-디코더구조를 가진 u-net형태의 모델이다.

[0041] 본 발명의 실시예에 따른 HarDNet의 인코더(210)는 다운 샘플링(212) 과정을 통해 이미지의 사이즈를 축소하면서 특징 맵을 추출한다.

[0042] 본 발명의 실시예에 따른 인코더(210)는 하나의 컨볼루션 블록(230), 4개의 HardNet 블록(211) 및 4개의 다운 샘플링 블록(212)으로 구성되었다. 본 발명의 실시예에 따른 컨볼루션 블록(230)은 필터(filter)=24, 커널 사이즈(kernel size)=3, 스트라이드(stride)=2인 컨볼루션 레이어와 필터(filter)=48, 커널 사이즈(kernel size)=3, 스트라이드(stride)=1인 컨볼루션 레이어로 구성되었다. 다운 샘플링 블록(212)은 하나의 커널 사이즈(kernel size)=1인 컨볼루션 레이어와 커널 사이즈(kernel size)=2, 스트라이드(stride)=2인 AvgPoll2d 레이어로 구성되었다. 그리고 전환 구간(bottleneck)은 하나의 HardNet 블록을 사용하여 인코더와 디코더의 파라미터 전달을 완성한다.

[0043] 본 발명의 실시예에 따른 디코더(220)는 전환 구간에서 전달받은 특징 맵을 입력과 동일한 사이즈로 업 샘플링함과 동시에 특징 맵에서 질환 영역을 찾아 출력 이미지에 표기한다.

[0044] 본 발명의 실시예에 따른 디코더(220)는 4개의 HarDNet 블록(221) 및 5개의 업 샘플링 블록(222) 및 마지막 커널 사이즈(kernel size)=1인 컨볼루션 레이어(230)를 거쳐 최종 (W, H) 사이즈의 이미지를 출력한다. 업 샘플링 블록(222)은 '바이리니어(bilinear)' 모드를 사용한 보간 함수(interpolate function)와 커널 사이즈=1 인 컨볼루션 레이어로 구성되었다.

[0045] 본 발명에서는 스윈 트랜스포머를 인코더와 디코더의 전환 구간에서 사용하지 않고 스킵 연결(240)에서 특징 맵의 사이즈가 제일 큰 처음 스킵 연결에 사용한다.

[0046] 본 발명의 실시예에 따르면, 입력 이미지를 (W(너비), H(높이), C(채널))로 표현하면 인코더의 컨볼루션 블록을 통해 계산된 특징 맵의 형태는 (W/2, H/2, 48)이다. 그리고 4개의 HardNet 블록 및 4개의 다운 샘플링 블록을 통해(W/32, H/32, 286)의 특징 맵을 출력한다. 디코더에서는 전환 구간에서 출력한 (W/32, H/32, 320)의 특징 맵을 입력으로 받아 (W, H, class)의 특징 맵을 출력한다(class는 데이터의 질환 카테고리 수를 의미함). HarDNet의 상세한 구조 및 출력 예시는 표 1에 표기한 바와 같다. 표 1에서 출력 사이즈 컬럼은 이미지 사이즈가 224224인 그레이 스케일(gray scale) 이미지를 입력으로 한 예시이다.

[0047] <표 1>

Stage	Block Name	Details	Output Size
Input			224*224*1
Encoder	Conv Block	2*Convolution	112*112*48
	HarDNet Block	4*Convolution	112*112*48
	Down-sampling Block	1*Convolution 1*AvgPool2d	56*56*64
	HarDNet Block	4*Convolution	56*56*78
	Down-sampling Block	1*Convolution 1*AvgPool2d	28*28*96
	HarDNet Block	8*Convolution	28*28*160
	Down-sampling Block	1*Convolution 1*AvgPool2d	14*14*160
	HarDNet Block	8*Convolution	14*14*214
	Down-sampling Block	1*Convolution 1*AvgPool2d	7*7*286
Bottle Neck	HarDNet Block	8*Convolution	7*7*320
Decoder	Up-sampling Block	1*Upsample 1*Convolution	-
	HarDNet Block	8*Convolution	14*14*214
	Up-sampling Block	1*Upsample 1*Convolution	-
	HarDNet Block	8*Convolution	28*28*160
	Up-sampling Block	1*Upsample 1*Convolution	-
	HarDNet Block	4*Convolution	56*56*78
	Up-sampling Block	1*Upsample 1*Convolution	-
	HarDNet Block	4*Convolution	112*112*48
	Up-sampling Block	1*Upsample 1*Convolution	-
Output	Conv Block	1*Convolution	224*224*2

[0048]

[0050] 도 3은 본 발명의 일 실시예에 따른 스윈 트랜스포머를 설명하기 위한 도면이다.

[0051] 본 발명의 일 실시예에 따른 스윈 트랜스포머는 입력 이미지가 패치 분할 레이어(patch partition layer)(311)를 통과해 4×4 사이즈를 가진 패치로 분리되어 (W/4, H/4, 4×4×channel)의 형태를 가진 패치 토큰(patch tokens)을 생성한다. 이렇게 생성된 패치 토큰은 stage 1(310)에서 선형 임베딩(linear embedding)(312)을 통과한 후 연결된 2개의 스윈 트랜스포머 블록(313)에 입력되어 (W/4, H/4, C)의 토큰을 생성하게 된다. 여기서 C는 임의의 단위(arbitrary dimension)를 의미한다. stage 2(320), stage 3(330)은 패치 병합(patch merging)(321, 331) 및 스윈 트랜스포머 블록(swin transformer block)(331, 332)으로 구성된다. 패치 병합(321, 331) 에서는 인접한 2×2 패치를 하나의 패치로 합치면서 토큰을 1/2로 다운 샘플링하고 C를 2배로 늘린다. stage 2(320), stage 3(330)에서의 토큰의 형태는 각각 (W/8, H/8, 2C), (W/16, H/16, 4C) 이다. 스윈 트랜스포머의 마지막 선형 예측(Linear Projection)에서 특징 맵을 입력 이미지 단위의 8배로 확장한 다음 1×1의 컨볼루션을 거쳐 입력과 동일한 형태의 특징 맵을 출력한다. 스윈 트랜스포머 모델은 인코더의 처음 HardNet 블록에서 출력한 112×112×48의 특징 맵을 입력으로 받아 디코더의 마지막 HardNet 블록에 연결된다. 본 발명의 실시예에 따른 스윈 트랜스포머 모델의 윈도우 사이즈는 7이다.

[0053] 도 4는 본 발명의 일 실시예에 따른 스윈 트랜스포머와 CNN을 결합한 의료영상 이미지 분할 방법을 설명하기 위한 흐름도이다.

[0054] 제안하는 스윈 트랜스포머와 CNN을 결합한 의료영상 이미지 분할 방법은 HarDNet의 인코더가 다운 샘플링을 수행하여 입력 이미지의 사이즈를 축소하고, 특징 맵을 추출하는 단계(410), HarDNet의 디코더가 상기 특징 맵을 입력 이미지와 동일한 사이즈로 업 샘플링하고, 상기 특징 맵에서 질환 영역을 찾아 출력 이미지에 표기하는 단계(420) 및 계층적 구조를 갖는 스윈 트랜스포머(swin transformer)를 통해 입력 이미지로부터 계층적 특징 맵을 추출하고, 쉬프트 윈도우 방식(shifted windows approach)을 이용하여 상기 계층적 특징 맵에서 패치(patch) 간의 관계를 계산함으로써 이미지를 분할하는 단계(430)를 포함한다.

[0055] 단계(410)에서, HarDNet의 인코더가 다운 샘플링을 수행하여 입력 이미지의 사이즈를 축소하고, 특징 맵을 추출

한다.

[0056] 본 발명의 실시예에 따른 HarDNet는 인코더를 포함한다.

[0057] 본 발명의 실시예에 따른 인코더는 하나의 컨볼루션 블록(convolution block), 복수의 HardNet 블록 및 복수의 다운 샘플링 블록을 포함한다.

[0058] 본 발명의 실시예에 따른 인코더의 전환 구간(bottleneck)은 하나의 HardNet 블록을 사용하여 인코더와 디코더 간의 파라미터를 전달할 수 있다.

[0059] 단계(420)에서, HarDNet의 디코더가 상기 특징 맵을 입력 이미지와 동일한 사이즈로 업 샘플링하고, 상기 특징 맵에서 질환 영역을 찾아 출력 이미지에 표기한다.

[0060] 본 발명의 실시예에 따른 HarDNet는 디코더를 포함한다.

[0061] 본 발명의 실시예에 따른 디코더는 복수의 HardNet 블록, 복수의 업 샘플링 블록 및 하나의 하나의 컨볼루션 블록을 포함한다.

[0062] 본 발명의 실시예에 따른 디코더는 상기 인코더의 전환 구간으로부터 전달 받은 특징 맵을 입력 이미지와 동일한 사이즈로 업 샘플링하고, 상기 특징 맵에서 질환 영역을 찾아 출력 이미지에 표기하여 최종 이미지를 출력한다.

[0063] 단계(430)에서, 계층적 구조를 갖는 스윈 트랜스포머(swin transformer)를 통해 입력 이미지로부터 계층적 특징 맵을 추출하고, 쉬프트 윈도우 방식(shifted windows approach)을 이용하여 상기 계층적 특징 맵에서 패치(patch) 간의 관계를 계산함으로써 이미지를 분할한다.

[0064] 본 발명의 실시예에 따른 스윈 트랜스포머는 상기 인코더-디코더 구조에서 인코더와 디코더 각각에 대한 스킵 연결(skip connection)을 수행하고 입력 이미지로부터 계층적 특징 맵을 추출하고, 쉬프트 윈도우 방식(shifted windows approach)을 이용하여 상기 계층적 특징 맵에서 패치(patch) 간의 관계를 계산함으로써 이미지를 분할한다.

[0065] 하지만 계층적 트랜스포머(hierarchical transformer)는 인접한 2×2의 패치를 병합하면서 특징 맵의 사이즈를 다운 샘플링하여 계층적 특징 맵을 생성한다. 그러므로 입력한 특징 맵의 사이즈가 작으면 스윈 트랜스포머는 깊은 계층적 특징 맵을 추출할 수 없다. 그런데 기존 CNN과 특징 맵의 결합 모델에서는 트랜스포머를 인코더와 디코더의 전환 구간(bottleneck)에 사용했다. 하지만 전환 구간은 인코더의 제일 마지막에서 출력한 특징 맵을 입력으로 받으므로 사이즈가 작다(예를 들어, 7×7). 이는 본 발명에서 스윈 트랜스포머의 계층적 트랜스포머를 사용하는 목적에 적합하지 않는다. 그래서 본 발명에서는 스윈 트랜스포머를 인코더와 디코더의 전환 구간에서 사용하지 않고 스킵 연결에서 특징 맵의 사이즈가 제일 큰 처음 스킵 연결에 사용한다.

[0066] 본 발명의 실시예에 따른 스윈 트랜스포머는 패치 분할 레이어를 통해 입력 이미지를 패치로 분할하여 패치 토큰을 생성하고, 상기 패치 토큰은 선형 임베딩을 통과한 후 연결되는 두 개의 스윈 트랜스포머 블록에 입력되어 새로운 토큰을 생성한다.

[0067] 이후, 상기 새로운 토큰은 패치 병합을 통해 인접한 패치를 하나의 패치로 병합하고, 상기 새로운 토큰을 다운 샘플링하는 과정을 반복하고, 선형 예측을 통해 특징 맵을 확장한 후 컨볼루션을 거쳐 입력 이미지와 통일한 형태의 특징 맵을 출력한다.

[0068] 본 발명의 실시예에 따른 스윈 트랜스포머는 상기 HarDNet의 인코더의 처음 HardNet 블록에서 출력한 특징 맵을 입력으로 받아 상기 HarDNet의 디코더의 마지막 HardNet 블록에 연결된다.

[0070] 도 5는 본 발명의 일 실시예에 따른 데이터 셋을 설명하기 위한 도면이다.

[0071] 본 발명의 실시예에 에서는 뇌출혈 CT 영상 및 뇌졸중 MRI 영상을 사용했다.

[0072] 본 발명의 실시예에 따른 CT 데이터 셋은 ICH(Intracerbral Hemorrhage) 영상을 사용했다. 본 발명의 실시예에서 서울성모병원, 중앙대학교, 인제대학교, 부산백병원, 건국대학교병원 등 5개 기관에서 제공한 82,636장 뇌출혈 CT 영상을 기반으로 연구를 진행했다(The dataset is published on AIHub[38]). 모든 영상은 512×512의 고해상도 이미지이고, 뇌출혈 영상은 출혈 위치에 따라 EDH(Epidural Hemorrhage), IPH(Intraparenchy-mal Hemorrhage), IVH(Intraventricular Hemorrhage), SAH(Subarachnoid Hemorrhage), SDH(Subdural Hemor-rhage)

등 5개의 카테고리로 분류되었다. 도 5는 각 카테고리별 뇌출혈 CT 영상의 예시이다. 도 5(a)는 IPH CT 영상을 의미하고, 도 5(b)는 IVH CT 영상을 의미하고, 도 5(c)는 SAH CT 영상을 의미하고, 도 5(d)는 SDH CT 영상을 의미하고, 도 5(e)는 EDH CT 영상을 의미하고, 도 5(f) multiple은 CT 영상에 한가지 이상의 뇌출혈 병변이 있는 영상을 의미한다.

[0073] 본 발명에서는 뇌출혈 CT 영상 EDH, IPH, IVH, SAH, SDH 및 multiple 카테고리에서 각자 200장을 선택해, 총 1,200장 뇌출혈 영상을 테스트 데이터로 사용한다. 그리고 나머지 81,436장 뇌출혈 영상을 질환 카테고리 단위에서 8:2의 비율로 분할해 65,151장 훈련 데이터셋, 16,285장 검증 데이터셋을 구축한다. <표 2>는 카테고리별 모델 학습 및 테스트에 사용한 데이터의 통계이다.

[0074] <표 2>

[0075]

	EDH	IVH	SDH	SAH	IPH	Multiple	Sum
Train	2,286	5,352	27,413	13,421	17,605	15,359	81,436
Test	200	200	200	200	200	200	1,200
Sum	2,486	5,552	27,613	13,621	17,805	15,559	82,636

[0076] ATLAS(Anatomical Tracings of Lesions After Stroke) 데이터셋은 뇌졸중의 병변 분할 알고리즘을 훈련 및 테스트하고 성능 비교를 위해 구축된 표준화된 오픈 데이터 셋(open dataset)이다[21,22]. ATLAS 데이터셋은 229명의 환자로부터 각각 197×233 해상도 MRI scan 영상 189장을 수집해 만들었다. ATLAS 데이터셋은 총 43,281장의 MRI 스캔 영상에 질환인 픽셀과 정상인 픽셀 2개의 클래스로 어노테이션(annotation)되었다. 본 발명에서는 전체 데이터에서 랜덤으로 약 80%인 177명 환자의 MRI 영상을 학습 데이터로 사용하고, 나머지 52명 환자의 데이터를 검증 데이터로 사용했다.

[0077] 본 발명의 실시예에서는 스윈 트랜스포머(swin transformer)와 HarDNet을 조합한 STHarDNet 구조를 제안하여 뇌졸중 MRI 스캔 영상의 분할에 적용했다. STHarDNet은 HarDNet 블록으로 구성된 경량화 U-Net형태의 인코더-디코더 모델과, 처음 스킵 연결에 인코더와 디코더를 연결하는 스윈 트랜스포머 블록으로 구성된 스윈 트랜스포머 모델로 구성되었다. STHarDNet은 CNN 모델의 구속적인 부분을 동시에 바라보며 태스크를 완성하는 특징과 트랜스포머의 이미지를 시퀀스 데이터(sequence data)로 바라보며 태스크를 완성하는 특징을 모두 가지고 있다. 그리고 스윈 트랜스포머를 모델의 첫 번째 스킵 연결에 적용함으로써 전환 구간보다 큰 사이즈의 특징 맵을 입력으로 받아 스윈 트랜스포머의 쉬프트 윈도우 접근 방식 및 계층별 특징 추출의 장점을 유지했다.

[0078] 제안하는 모델의 우수성능 증명하기 위해 뇌 CT, MRI 영상에 기존 SOTA 모델과의 분할 성능 비교 실험을 하였다. 실험 결과 5가지 뇌출혈 질환을 포함한 82,636장 CT 영상에서의 분할 성능은 dice, IoU, precision, recall 각각 0.728, 0.618, 0.791, 721로 U-Net, U-Net++, SegNet, PSPNet, HarDNet, TransHarDNet, Swin-UNet, 스윈 트랜스포머 기반 분할(Swin Transformer based Segmentation) 등 분할 SOTA 모델들 보다 성능이 우수했다.

[0079] 그리고 ATLAS 데이터셋에서 177명 환자의 MRI 스캔 영상을 학습에 사용하고, 52명 환자의 영상을 검증에 사용해 기존 분할 분야에서의 SOTA 모델과 성능 비교 실험을 진행했다. 실험 결과 STHarDNet은 224×224×1의 그레이스케일이미지를 입력하였을 때의 dice는 0.517, IoU는 0.387, precision은 0.622, recall은 0.498로 기존 SOTA 모델보다 우수했다. 그리고 동일한 조건에서 4장의 연속된 그레이 스케일 이미지를 모은 224×224×4인 3D 이미지를 입력했을 때 제안하는 모델의 dice는 0 0.5457, IoU는 0.4191, precision은 0.6470, re-call은 0.5180로 2D 이미지를 입력하였을 때 보다 성능이 더 우수했다. 또한 기존 SOTA 모델과 영상처리 속도 비교 실험을 진행한 결과 100,000장의 이미지를 처리할 때 STHarDNet의 FPS는 299.943장/s로 HarDNet보다 2.48% 느려 10개의 모델 중 2번째로 빠르다. 하지만 STHarDNet의 분할 성능은 HarDNet보다 dice에서 7.17%, IoU에서 9.95% 높아 성능과 속도를 모두 고려했을 때 가장 우수하다고 할 수 있다.

[0080] 본 발명에서는 실험을 통해 스윈 트랜스포머를 HarDNet의 처음 스킵 연결에 조합한 STHarDNet의 분할 성능의 우수성을 증명했다. 그리고 트랜스포머를 전환 구간뿐만 아니라 스킵 연결에 연결해도 모델의 성능을 높임과 동시에 빠른 계산 속도를 유지할 수 있다는 것을 증명했다.

[0082] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어

구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0083] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치에 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[0084] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

[0085] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

[0086] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

[0088] <참고문헌>

[0089] 1. Badza, M.; Barjaktarovi'c, M. Segmentation of Brain Tumors from MRI Images Using Convolutional Autoencoder. Appl. Sci. 2021, 11, 4317. [CrossRef]

[0090] 2. Ghosh, S.; Huo, M.; Shawkat, M.S.A.; McCalla, S. Using Convolutional Encoder Networks to Determine the Optimal Magnetic Resonance Image for the Automatic Segmentation of Multiple Sclerosis. Appl. Sci. 2021, 11, 8335. [CrossRef]

[0091] 3. Wu, S.; Wu, Y.; Chang, H.; Su, F.T.; Liao, H.; Tseng, W.; Liao, C.; Lai, F.; Hsu, F.; Xiao, F. Deep Learning-Based Segmentation of Various Brain Lesions for Radiosurgery. Appl. Sci. 2021, 11, 9180. [CrossRef]

[0092] 4. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-

Assisted Intervention, Munich, Germany, 5-9 October 2015; Springer: Cham, Switzerland, 18 November 2015; pp. 234-241.

[0093] 5. Wang, D.; Wu, Z.; Yu, H. TED-Net: Convolution-Free T2T Vision Transformer-Based Encoder-Decoder Dilation Network for Low-Dose CT Denoising. arXiv 2021, arXiv:2106.04650.

[0094] 6. Zhou, H.-Y.; Guo, J.; Zhang, Y.; Yu, L.; Wang, L.; Yu, Y. nnFormer: Interleaved Transformer for Volumetric Segmentation. arXiv 2021, arXiv:2109.03201.

[0095] 7. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-End Object Detection with Trans-formers. In European Conference on Computer Vision; Springer: Cham, Switzerland, 2020; Volume 12346, pp. 213-229. [CrossRef]

[0096] 8. Zheng, S.; Lu, J.; Zhao, H.; Zhu, X.; Luo, Z.; Wang, Y.; Fu, Y.; Feng, J.; Xiang, T.; Torr, P.H.; et al. Rethinking Semantic Segmentation from a Sequence-to-Sequence Perspective with Transformers. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 21-24 June 2021; pp. 6877-6886.

[0097] 9. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv 2020, arXiv:2010.11929.

[0098] 10. Xu, G.; Wu, X.; Zhang, X.; He, X. LeViT-UNet: Make Faster Encoders with Transformer for Medical Image Segmentation. arXiv 2021, arXiv:2107.08623.

[0099] 11. Chang, Y.; Menghan, H.; Guangtao, Z.; Xiao-Ping, Z. TransClaw U-Net: Claw U-Net with Transformers for Medical Image Segmentation. arXiv 2021, arXiv:2107.05188.

[0100] 12. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. arXiv 2021, arXiv:2102.04306.

[0101] 13. Wang, W.; Chen, C.; Ding, M.; Yu, H.; Zha, S.; Li, J. TransBTS: Multimodal Brain Tumor Segmentation Using Transformer. In Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2021; pp. 109-119. [CrossRef]

[0102] 14. Chao, P.; Kao, C.-Y.; Ruan, Y.; Huang, C.-H.; Lin, Y.-L. HarDNet: A Low Memory Traffic Network. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October-2 November 2019; pp. 3551-3560. [CrossRef]

[0103] 15. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. arXiv 2021, arXiv:2103.14030.

[0104] 16. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation. arXiv 2021, arXiv:2105.05537.

[0105] 17. Qi, K.; Yang, H.; Li, C.; Liu, Z.; Wang, M.; Liu, Q.; Wang, S. X-Net: Brain Stroke Lesion Segmentation Based on Depthwise Separable Convolution and Long-Range Dependencies. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13-17 October 2019; Springer: Cham, Switzerland, 2019; Volume 11766, pp. 247-255. [CrossRef]

[0106] 18. Zhou, Y.; Huang, W.; Dong, P.; Xia, Y.; Wang, S. D-UNet: A Dimension-Fusion U Shape Network for Chronic Stroke Lesion Segmentation. IEEE/ACM Trans. Comput. Biol. Bioinform. 2019, 18, 940-950. [CrossRef] [PubMed]

[0107] 19. Basak, H.; Hussain, R.; Rana, A. DFENet: A Novel Dimension Fusion Edge Guided Network for Brain MRI Segmentation. SN Comput. Sci. 2021, 2, 1-11. [CrossRef]

[0108] 20. Zhang, Y.; Wu, J.; Liu, Y.; Chen, Y.; Wu, E.X.; Tang, X. MI-UNet: Multi-Inputs UNet Incorporating Brain Parcellation for Stroke Lesion Segmentation From T1-Weighted Magnetic Resonance Images. IEEE J. Biomed. Health Inform. 2021, 25, 526-535. [CrossRef] [PubMed]

[0109] 21. Anatomical Tracings of Lesions after Stroke (ATLAS) R1.1. Available online: http://fcon_1000.projects.nitrc.org/indi/retro/atlas. html (accessed on 15 February 2021).

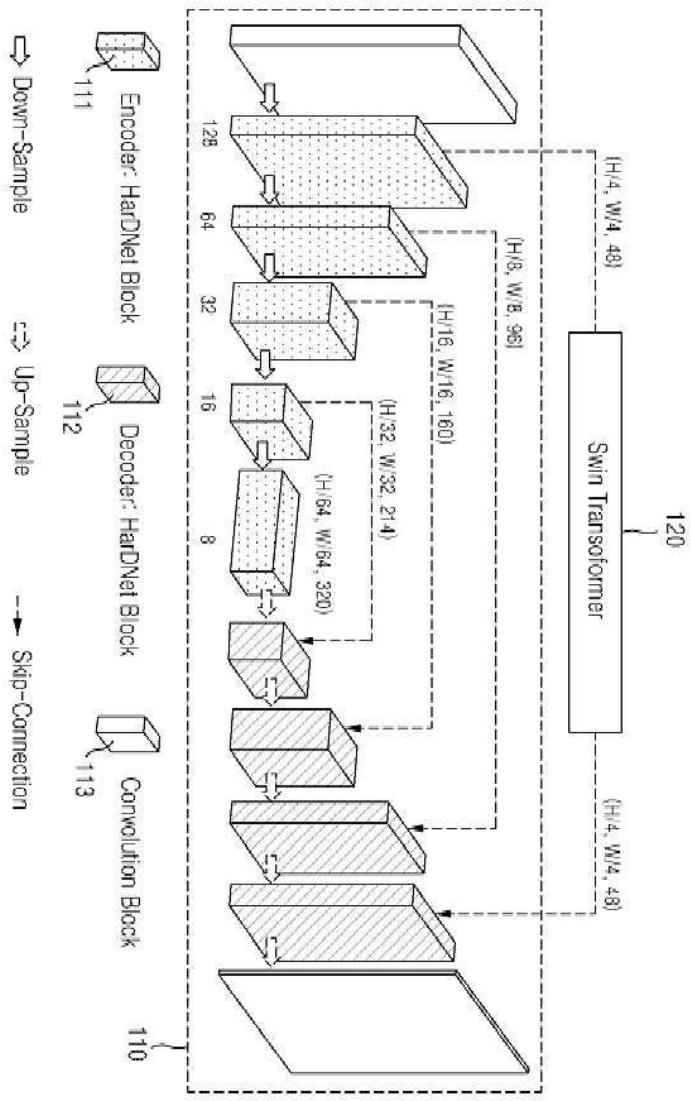
[0110] 22. Liew, S.-L.; Anglin, J.M.; Banks, N.W.; Sondag, M.; Ito, K.; Kim, H.; Chan, J.; Ito, J.; Jung, C.; Khoshab, N.; et al. A large, open source dataset of stroke anatomical brain images and manual lesion segmentations. Sci. Data 2018, 5, 180011. [CrossRef] [PubMed]

[0111] 23. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 2017, 39, 2481-2495. [CrossRef] [PubMed]

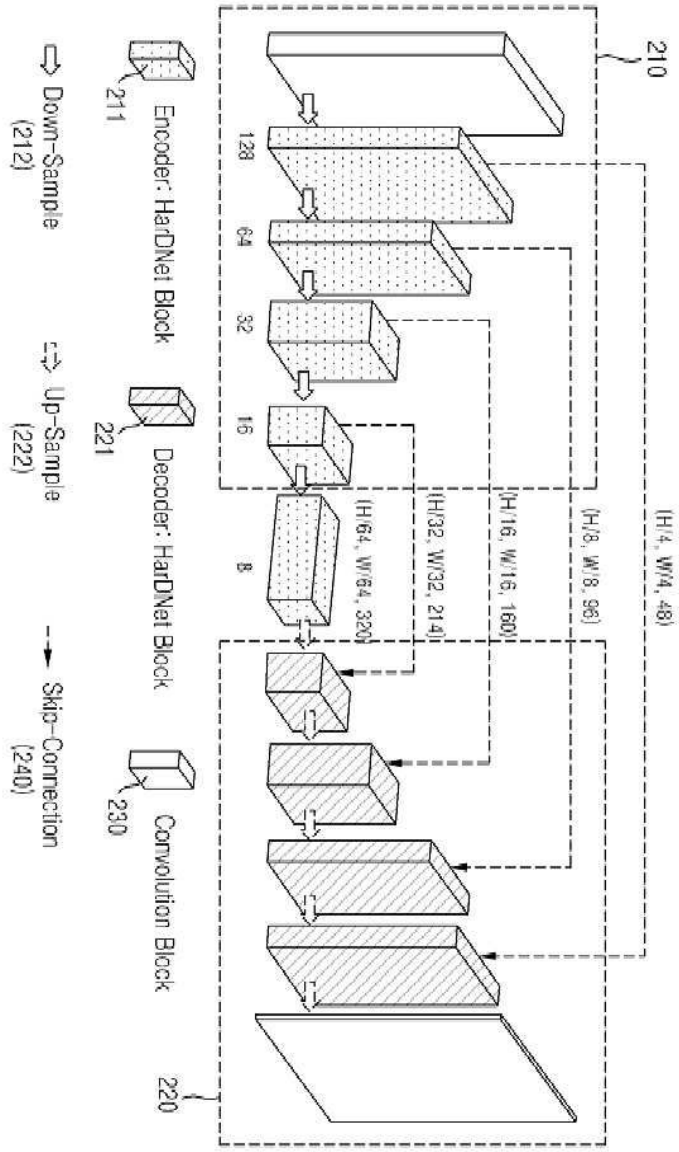
[0112] 24. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21-26 July 2017; pp. 2881-2890. [CrossRef]

도면

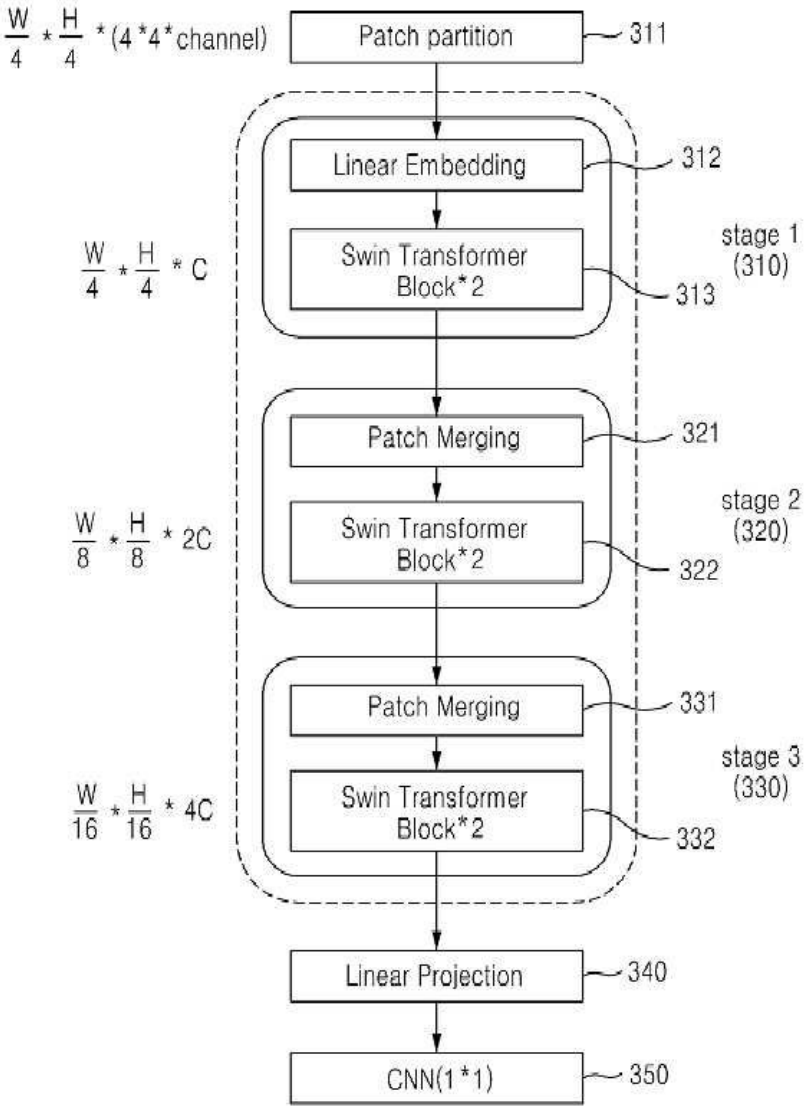
도면1



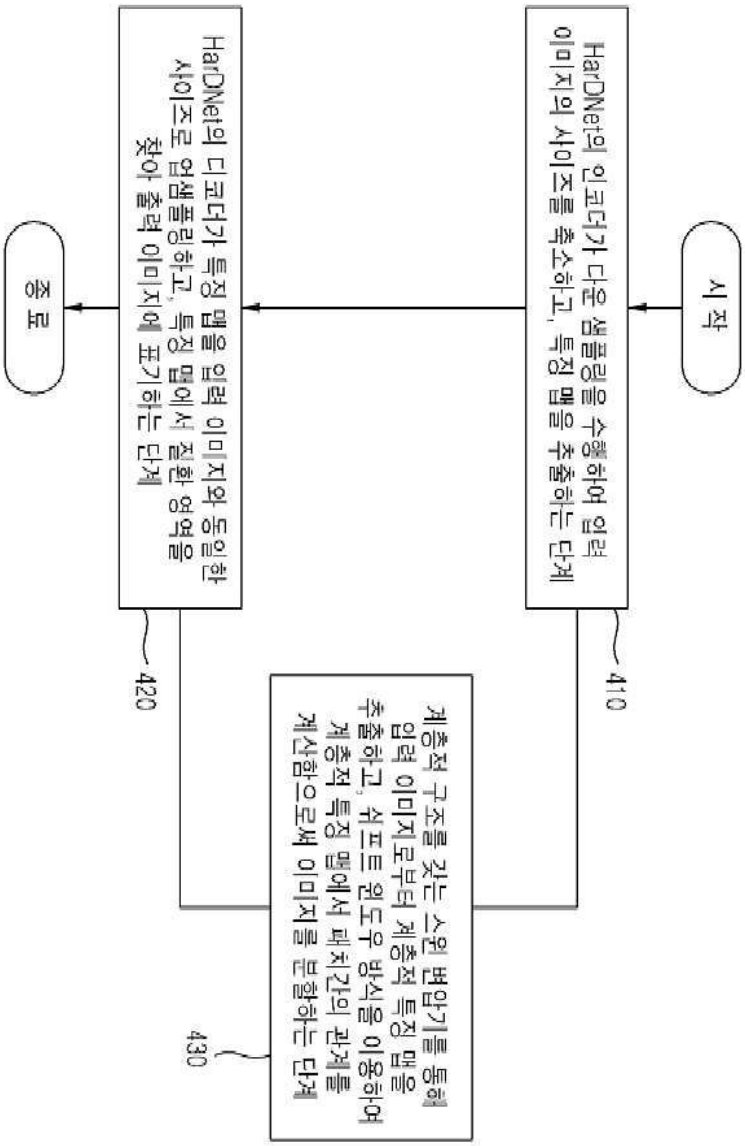
도면2


210
128
64
32
16
8
(H/4, W/4, 48)
(H/8, W/8, 96)
(H/16, W/16, 160)
(H/32, W/32, 214)
(H/64, W/64, 320)
211
Encoder: HarDNet Block
221
Decoder: HarDNet Block
230
Convolution Block
220
Down-Sample (212)
Up-Sample (222)
Skip-Connection (240)

도면3


W/4 * H/4 * (4*4*channel)
Patch partition
311
Linear Embedding
312
W/4 * H/4 * C
Swin Transformer Block*2
313
stage 1 (310)
Patch Merging
321
W/8 * H/8 * 2C
Swin Transformer Block*2
322
stage 2 (320)
Patch Merging
331
W/16 * H/16 * 4C
Swin Transformer Block*2
332
stage 3 (330)
Linear Projection
340
CNN(1*1)
350

도면4


시 작
HarDNet의 인코더가 다운 샘플링을 수행하여 입력 이미지의 사이즈를 축소하고, 특징 맵을 추출하는 단계
410
계층적 구조를 갖는 스윈 변압기를 통해 입력 이미지로부터 계층적 특징 맵을 추출하고, 쉬프트 윈도우 방식을 이용하여 계층적 특징 맵에서 패치간의 관계를 계산함으로써 이미지를 분할하는 단계
430
HarDNet의 디코더가 특징 맵을 입력 이미지와 동일한 사이즈로 업샘플링하고, 특징 맵에서 질환 영역을 찾아 출력 이미지에 표기하는 단계
420
종 료

도면5

