

**【서지사항】****【서류명】** 특허출원서**【출원구분】** 특허출원**【출원인】****【명칭】** 연세대학교 산학협력단**【특허고객번호】** 2-2005-009509-9**【대리인】****【성명】** 권성현**【대리인번호】** 9-2012-000114-4**【포괄위임등록번호】** 2020-085395-8**【대리인】****【성명】** 강일신**【대리인번호】** 9-2013-000145-7**【포괄위임등록번호】** 2020-085394-1**【대리인】****【성명】** 김정연**【대리인번호】** 9-2010-001352-0**【포괄위임등록번호】** 2020-085398-0**【대리인】****【성명】** 백두진**【대리인번호】** 9-2010-000842-1**【포괄위임등록번호】** 2020-085396-5

**【대리인】**

**【성명】** 유광철

**【대리인번호】** 9-2013-000581-3

**【포괄위임등록번호】** 2020-085397-2

**【발명의 국문명칭】** 이중 어텐션을 이용한 이미지 세그멘테이션 방법 및 이를 이용한 장치

**【발명의 영문명칭】** IMAGE SEGMENTATION METHOD USING DUAL ATTENTION AND DEVICE USING THE SAME

**【발명자】**

**【성명】** 박상현

**【성명의 영문표기】** SANGHYUN PARK

**【주민등록번호】** 670101-1XXXXXX

**【우편번호】** 08004

**【주소】** 서울특별시 양천구 오목로 300, 204동 3701호

**【발명자】**

**【성명】** 홍승균

**【성명의 영문표기】** SEUNGKYUN HONG

**【주민등록번호】** 671120-1XXXXXX

**【우편번호】** 08083

**【주소】** 서울특별시 양천구 중앙로36길 15, 111동 1902호

**【발명자】**

**【성명】** 안성현

【성명의 영문표기】	SUNGHYUN AHN
【주민등록번호】	000104-3XXXXXX
【우편번호】	21368
【주소】	인천광역시 부평구 원적로269번길 15, 104동 1901호
【발명자】	
【성명】	조영완
【성명의 영문표기】	YOUNGWAN JO
【주민등록번호】	990312-1XXXXXX
【우편번호】	61268
【주소】	광주광역시 북구 서강로 77, 106동 1305호
【출원언어】	국어
【심사청구】	청구
【이 발명을 지원한 국가연구개발사업】	
【과제고유번호】	1711193130
【과제번호】	2017-0-00477-007
【부처명】	과학기술정보통신부
【과제관리(전문)기관명】	정보통신기획평가원
【연구사업명】	SW컴퓨팅산업원천기술개발
【연구과제명】	(SW 스타랩) IoT 환경을 위한 고성능 플래시 메모리 스토리지 기반 인메모리 분산 DBMS 개발
【기여율】	1/1
【과제수행기관명】	연세대학교 산학협력단

**【연구기간】** 2023.01.01 ~ 2023.12.31

**【취지】** 위와 같이 특허청장에게 제출합니다.

대리인 권성현 (서명 또는 인)

대리인 강일신 (서명 또는 인)

대리인 김정연 (서명 또는 인)

대리인 백두진 (서명 또는 인)

대리인 유광철 (서명 또는 인)

### **【수수료】**

**【출원료】** 0 면 46,000 원

**【가산출원료】** 34 면 0 원

**【우선권주장료】** 0 건 0 원

**【심사청구료】** 13 항 829,000 원

**【합계】** 875,000원

**【감면사유】** 전담조직(50%감면)[1]

**【감면후 수수료】** 437,500 원

## 【발명의 설명】

### 【발명의 명칭】

이중 어텐션을 이용한 이미지 세그멘테이션 방법 및 이를 이용한 장치  
{ IMAGE SEGMENTATION METHOD USING DUAL ATTENTION AND DEVICE USING THE SAME }

### 【기술분야】

【0001】 본원 발명은 이미지 세그멘테이션 방법에 관한 것으로, 보다 상세하게는, 이중 어텐션을 이용하여 이미지를 세그멘테이션하는 방법에 관한 것이다.

### 【발명의 배경이 되는 기술】

【0002】 의료 영상 분할은 종양과 같은 병변을 구분하고 다양한 장기를 분할하여 빠르고 쉽게 의료 영상을 분석하고 정확한 진단을 가능하게 하는 분야이다. 특히, 3D CT 및 MRI와 같은 복잡한 의료 영상에서는 여러 장기들을 정확하게 구별하기 위해, 깊은 신경망 모델을 통해 얻은 특징 정보에서 유용한 공간 및 채널 정보를 파악하는 것이 중요하다.

【0003】 기존의 U-Net 기반 깊은 신경망 연구들은 모델의 크기가 증가함에 따라 채널의 수가 증가했음에도 불구하고 공간 정보에만 초점을 맞추어 정확한 의료 영상 분할을 위한 정보를 얻는 데에는 한계가 존재한다. 따라서, 의료 영상에 대한 문맥 정보를 추가적으로 확보하기 위하여 공간 정보뿐만 아니라 채널 정보를 추출하여 이를 융합하는 이중 어텐션을 이용한 기술이 필요하다.

### 【발명의 내용】

## 【해결하고자 하는 과제】

【0004】 본원 발명의 일 과제는 이중 어텐션을 이용한 이미지 세그멘테이션 방법에 관한 것이다.

## 【과제의 해결 수단】

【0005】 일 실시예에 따른 이미지 세그멘테이션 방법은 적어도 하나 이상의 프로세서에 의해 수행되는 이미지 세그멘테이션 방법에 있어서, 입력 데이터로부터 공간 어텐션을 추출하는 단계; 상기 입력 데이터로부터 채널 어텐션을 추출하는 단계; 상기 공간 어텐션 및 상기 채널 어텐션에 기초하여 머징 데이터를 생성하는 단계; 및 상기 머징 데이터에 기초하여 구성 요소별 인덱스를 포함하는 결과 데이터를 획득하는 단계를 포함할 수 있다.

【0006】 여기서, 상기 공간 어텐션은 구성 요소와 관련된 형태 정보를 포함하고, 상기 채널 어텐션은 구성 요소와 관련된 문맥 정보를 포함할 수 있다.

【0007】 여기서, 상기 공간 어텐션을 추출하는 단계는 스윈 트랜스포머(swin transformer)를 이용하는 단계일 수 있다.

【0008】 여기서, 상기 공간 어텐션을 추출하는 단계는, 상기 입력 데이터에 의해 생성된 제1 크기의 패치 및 제1 값의 채널 개수에 기초하여 제1 공간 어텐션을 추출하는 단계; 및 상기 입력 데이터에 의해 생성된 제2 크기의 패치 및 제2 값의 채널 개수에 기초하여 제2 공간 어텐션을 추출하는 단계를 포함하고, 상기 제2 크기는 상기 제1 크기보다 크고, 상기 제1 값은 상기 제2 값보다 작을 수 있다.

【0009】 여기서, 상기 채널 어텐션을 추출하는 단계는, 상기 제1 크기의 패치 및 상기 제1 값의 채널 개수에 기초하여 제1 채널 어텐션을 추출하는 단계; 및 상기 제2 크기의 패치 및 상기 제2 값의 채널 개수에 기초하여 제2 채널 어텐션을 추출하는 단계를 포함할 수 있다.

【0010】 여기서, 상기 머징 데이터를 생성하는 단계는, 제1 공간 어텐션, 제1 채널 어텐션, 제2 공간 어텐션 및 제2 채널 어텐션에 기초하여 제1 바텀업 데이터를 생성하는 단계; 및 상기 제1 바텀업 데이터, 제3 공간 어텐션 및 제3 채널 어텐션에 기초하여 제2 바텀업 데이터를 생성하는 단계를 포함할 수 있다.

【0011】 여기서, 상기 머징 데이터를 생성하는 단계는, 상기 제1 바텀업 데이터 상기 제2 바텀업 데이터, 상기 제2 공간 어텐션 및 상기 제2 채널 어텐션에 기초하여 제1 탑다운 데이터를 생성하는 단계; 및 상기 제1 탑다운 데이터, 상기 제1 공간 어텐션 및 상기 제1 채널 어텐션에 기초하여 제2 탑다운 데이터를 생성하는 단계를 포함할 수 있다.

【0012】 여기서, 상기 결과 데이터를 획득하는 단계는, 상기 머징 데이터에 의해 생성된 제3 크기의 패치 및 제3 값의 채널 개수에 기초하여 제1 디코딩 데이터를 생성하는 단계; 및 상기 머징 데이터에 의해 생성된 제4 크기의 패치 및 제4 값의 채널 개수에 기초하여 제2 디코딩 데이터를 생성하는 단계를 포함하고, 상기 제4 크기는 상기 제3 크기보다 작고, 상기 제3 값은 상기 제4 값보다 클 수 있다.

【0013】 여기서, 상기 결과 데이터를 획득하는 단계는, 상기 제1 디코딩 데이터 및 상기 제2 디코딩 데이터에 기초하여 제1 픽셀에 제1 구성 요소에 대응되는 제1 인덱스를 설정하는 단계; 상기 제1 디코딩 데이터 및 상기 제2 디코딩 데이터에 기초하여 제2 픽셀에 제2 구성 요소에 대응되는 제2 인덱스를 설정하는 단계; 및 상기 제1 인덱스가 설정된 픽셀을 포함하는 제1 영역과 상기 제2 인덱스가 설정된 픽셀을 포함하는 제2 영역을 설정하는 단계를 포함할 수 있다.

【0014】 여기서, 상기 제1 디코딩 데이터를 생성하는 단계 및 상기 제2 디코딩 데이터를 생성하는 단계는 스윈 트랜스포머(swin transformer)를 이용하는 단계일 수 있다.

【0015】 여기서, 상기 이미지 세그멘테이션 방법을 실행시키도록 컴퓨터로 판독 가능한 기록 매체에 기록된 컴퓨터 프로그램이 제공될 수 있다.

【0017】 일 실시예에 따른 전자 장치는 전자 장치에 있어서, 메모리; 및 적어도 하나 이상의 프로세서를 포함하고, 상기 메모리는, 상기 적어도 하나 이상의 프로세서에 의해 실행 시, 상기 전자 장치가, 입력 데이터로부터 추출된 공간 어텐션을 저장하고, 상기 입력 데이터로부터 추출된 채널 어텐션을 저장하고, 상기 공간 어텐션 및 상기 채널 어텐션에 기초하여 생성된 머징 데이터를 저장하고, 상기 머징 데이터에 기초하여 획득된 구성 요소별 인덱스를 포함하는 결과 데이터를 저장할 수 있다.

【0019】 일 실시예에 따른 이미지 세그멘테이션 장치는 입력 데이터로부터 공간 어텐션 및 채널 어텐션을 추출하는 인코딩부; 상기 공간 어텐션 및 상기 채널 어텐션에 기초하여 머징 데이터를 생성하는 머징부; 및 상기 머징 데이터에 기초하여 구성 요소별 인덱스를 포함하는 결과 데이터를 생성하는 디코딩부를 포함할 수 있다.

### 【발명의 효과】

【0020】 본원 발명의 일 실시예에 따르면 이중 어텐션을 이용한 이미지 세그멘테이션 방법이 제공될 수 있다.

### 【도면의 간단한 설명】

【0021】 도 1은 일 실시예에 따른 이미지 세그멘테이션 방법의 순서도이다.

도 2는 일 실시예에 따른 이미지 세그멘테이션 방법을 설명하기 위한 도면이다.

도 3은 일 실시예에 따른 공간 어텐션 및 채널 어텐션 병합 방법을 설명하기 위한 도면이다.

도 4 내지 도 9는 본원 발명의 이미지 세그멘테이션 방법에 따른 결과를 종래 기술과 비교하여 설명하기 위한 도면이다.

### 【발명을 실시하기 위한 구체적인 내용】

【0022】 본 명세서에 기재된 실시예는 본원 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 본원 발명의 사상을 명확히 설명하기 위한 것이므로, 본원 발명이 본 명세서에 기재된 실시예에 한정되는 것은 아니며, 본원 발명의 범위는 본원 발명의 사상을 벗어나지 아니하는 수정예 또는 변형예를 포함하는 것으로 해석되어야 한다.

【0023】 본 명세서에서 사용되는 용어는 본원 발명에서의 기능을 고려하여 가능한 현재 널리 사용되고 있는 일반적인 용어를 선택하였으나 이는 본원 발명이 속하는 기술 분야에서 통상의 지식을 가진 자의 의도, 판례 또는 새로운 기술의 출현 등에 따라 달라질 수 있다. 다만, 이와 달리 특정한 용어를 임의의 의미로 정의하여 사용하는 경우에는 그 용어의 의미에 관하여 별도로 기재할 것이다. 따라서 본 명세서에서 사용되는 용어는 단순한 용어의 명칭이 아닌 그 용어가 가진 실질적인 의미와 본 명세서의 전반에 걸친 내용을 토대로 해석되어야 한다.

【0024】 본 명세서에 첨부된 도면은 본원 발명을 용이하게 설명하기 위한 것으로 도면에 도시된 형상은 본원 발명의 이해를 돋기 위하여 필요에 따라 과장되어 표시된 것일 수 있으므로 본원 발명이 도면에 의해 한정되는 것은 아니다.

【0025】 본 명세서에서 본원 발명에 관련된 공지의 구성 또는 기능에 대한 구체적인 설명이 본원 발명의 요지를 흐릴 수 있다고 판단되는 경우에 이에 관한 자세한 설명은 필요에 따라 생략하기로 한다.

【0027】 본원 발명은 추출한 이중 어텐션을 이용하여 이미지를 세그멘테이션 하는 방법 및 장치에 관한 것이다. 구체적으로, 본원 발명은 입력 데이터로부터 공간 어텐션 및 채널 어텐션을 추출할 수 있는 전자 장치, 이미지 세그멘테이션 장치에 관한 것이다. 전자 장치 또는 이미지 세그멘테이션 장치는 컴퓨팅 장치로서 메모리 및 상기 메모리를 실행할 수 있는 프로세서를 포함하는 적어도 하나 이상의 프로세서를 포함할 수 있다. 이하에서 설명하는 이미지 세그멘테이션 방법은 전자 장치 또는 이미지 세그멘테이션 장치에 포함된 적어도 하나 이상의 프로세서에 의해 수행되는 것으로 해석될 수 있다.

【0029】 도 1은 일 실시예에 따른 이미지 세그멘테이션 방법의 순서도이다.

【0030】 도 1을 참조하면, 일 실시예에 따른 이미지 세그멘테이션 방법은 공간 어텐션 및 채널 어텐션을 추출하는 단계(S100), 멀징 데이터를 생성하는 단계(S200) 및 결과 데이터를 생성하는 단계(S300)를 포함할 수 있다. 도 1은 단계 S100 내지 S300이 순차적으로 수행되는 것을 도시하였으나, 이에 한정되지 않고 새로운 단계가 추가되거나 일부 단계가 병합되어 동시에 수행될 수도 있다.

【0031】 전자 장치 또는 이미지 세그멘테이션 장치는 인코딩부, 멀징부 및 디코딩부를 포함할 수 있다. 인코딩부는 입력 데이터로부터 공간 어텐션 및 채널 어텐션을 추출할 수 있다. 멀징부는 인코딩부가 추출한 공간 어텐션 및 채널 어텐션에 기초하여 멀징 데이터를 생성할 수 있다. 구체적으로, 멀징부는 공간 어텐션

및 채널 어텐션을 병합하여 머징 데이터를 생성할 수 있다. 디코딩부는 머징부가 생성한 머징 데이터에 기초하여 구성 요소별 인덱스를 포함하는 결과 데이터를 생성할 수 있다. 인코딩부는 단계 S100의 주체이고, 머징부는 단계 S200의 주체이고, 디코딩부는 단계 S300의 주체일 수 있다.

**【0032】** 이하에서 도 2 및 도 3을 함께 참조하여 이미지 세그멘테이션 방법에 대해 구체적으로 설명한다.

**【0034】** 도 2는 일 실시예에 따른 이미지 세그멘테이션 방법을 설명하기 위한 도면이다.

**【0035】** 도 2를 참조하면, 일 실시예에 따른 이미지 세그멘테이션 방법은 입력 데이터(10)를 이용하여 인코딩 과정(100), 머징머징 과정(200) 및 디코딩 과정(300)을 포함할 수 있다. 인코딩 과정(100)은 입력 데이터(10)로부터 공간 어텐션을 추출하는 과정(110) 및 입력 데이터(10)로부터 채널 어텐션을 추출하는 과정(120)을 포함할 수 있다. 머징머징 과정(200)은 탑다운 머징 과정(210) 및 바텀업 머징 과정(220)을 포함할 수 있다.

**【0036】** 인코딩 과정(100)은 입력 데이터(10)로부터 공간 어텐션 및 채널 어텐션을 추출하는 단계(S100)를 포함할 수 있다. 공간 어텐션 및 채널 어텐션을 추출하기 전, 프로세서는 입력 데이터(10)를 복수의 패치(patch)로 구분하는 과정을 수행할 수 있다. 이때, 패치의 사이즈 및 패치로부터 추출하는 정보와 관련된 채널

의 개수는 추출 과정마다 상이할 수 있다.

**【0037】** 프로세서는 공간 스트림(예, 공간 어텐션 추출 모듈)을 이용하여 복수의 패치로 나눠진 입력 데이터(10)로부터 공간 어텐션을 추출할 수 있다. 이때, 공간 어텐션은 구성 요소(예, 위, 간 등과 같은 장기)와 관련된 형태 정보 또는 위치 정보를 포함할 수 있다. 또한, 공간 어텐션은 채널 어텐션이 포함하고 있는 맵 정보도 포함할 수 있다. 구체적으로, 프로세서는 입력 데이터(10)에 의해 생성된 제1 크기의 패치 및 제1 값의 채널 개수에 기초하여 제1 공간 어텐션을 추출할 수 있다. 예를 들어, 프로세서는 높이(height), 너비(width), 깊이(depth)가 각각 4, 4, 2 의 크기를 가지는 패치에 대해, 32개의 채널에 기초하여 제1 공간 어텐션을 추출할 수 있다.

**【0038】** 또한, 프로세서는 입력 데이터(10)에 의해 생성된 제2 크기의 패치 및 제2 값의 채널 개수에 기초하여 제2 공간 어텐션을 추출할 수 있다. 이때, 상기 제2 크기는 상기 제1 크기보다 크고, 상기 제1 값은 상기 제2 값보다 작을 수 있다. 예를 들어, 프로세서는 높이, 너비, 깊이가 각각 8, 8, 4 의 크기를 가지는 패치에 대해, 64개의 채널에 기초하여 제2 공간 어텐션을 추출할 수 있다. 이와 같이, 프로세서는 복수의 추출 과정을 가지고, 각 과정마다 상이한 크기의 패치 및 상이한 개수의 채널에 기초하여 공간 어텐션을 추출할 수 있다.

**【0039】** 프로세서는 복수의 추출 과정을 진행할수록 패치의 크기가 커지고 큰 값의 채널을 이용하여 공간 어텐션을 추출하므로, 추출 과정이 진행됨에 따라 더 많은 양의 정보를 획득할 수 있다. 복수의 추출 과정을 진행할수록 공간적 특징

(spatial feature)들은 압축되기 때문에, 이를 통해 프로세서는 로컬 정보(local feature)에서 점점 글로벌 정보(global feature)를 획득할 수 있다. 이때, 공간이 압축된 만큼 소실된 정보들이 발생하는 것을 보완하기 위해 채널 정보를 늘리면서 특징 맵에 대한 다양한 전역적인 문맥 정보를 획득한다.

**【0040】** 또한, 채널을 늘리는 방법은 압축되기 전 공간 정보들의 일부를 채널 방향으로 이동시킨 후, 채널 특징을 선형 변환함으로써 수행될 수 있다. 이에, 각 채널 정보마다 임의의 공간 정보를 갖게될 수 있다. 예를 들어, 제1 채널은 간의 일부에 대한 정보를 갖고, 제2 채널은 심장의 일부에 대한 채널을 갖을 수 있다. 프로세서는 채널 정보 중 구성 요소(예, 간, 심장과 같은 장기)와 관련된 정보들에 집중하기 위해, 채널 어텐션을 추출할 수 있다. 채널 어텐션에 대한 내용은 후술한다.

**【0041】** 프로세서는 공간 어텐션을 추출할 때, 스윈 트랜스포머(swin transformer)를 이용할 수 있다. 즉, 공간 어텐션을 추출하는 단계는 스윈 트랜스포머를 이용하는 단계일 수 있다. 스вин 트랜스포머는 트랜스포머 엔코더를 계층적으로 쌓은 모델로서, 시프티드 윈도우(shifted window) 알고리즘을 통해 윈도우 간 어텐션을 진행할 수 있다.

**【0042】** 본원 발명의 이미지 세그멘테이션 방법은 패치의 사이즈를 줄이고 채널의 개수를 늘리면서 복수의 추출 과정을 진행하므로, 많은 계산양이 요구된다. 이에, 본원 발명은 공간 어텐션을 추출할 때 비전 트랜스포머보다 선형성 (linearity)을 이용하는 스вин 트랜스포머를 이용하여 많은 계산을 할 수 있으며,

종래 대비 로컬과 글로벌 정보를 효율적으로 획득할 수 있다.

**【0043】** 프로세서는 채널 스트림을 이용하여 복수의 패치로 나눠진 입력 데이터(10)로부터 채널 어텐션을 추출할 수 있다. 이때, 채널 어텐션은 구성 요소와 관련된 문맥(context) 정보를 포함할 수 있다. 채널 스트림은 인코딩의 각 단계별로 해상도를 줄임에 따라 정보 손실을 방지하기 위하여 채널을 늘리게 된다. 이에, 채널 어텐션은 그에 따른 구성 정보의 특징 및 다른 구성 요소간의 문맥 정보를 포함할 수 있다. 예를 들어, 채널 어텐션은 색상, 샤프니스, 강도, 다른 구성 요소와의 연관성 등을 포함할 수 있다. 구체적인 예를 들어, 채널 어텐션은 간에 집중한 문맥, 간과 췌장에 대한 문맥 등 컴퓨터 차원에서 인식할 수 있는 이미지의 문맥을 포함할 수 있다.

**【0044】** 구체적으로, 프로세서는 입력 데이터(10)에 의해 생성된 제1 크기의 패치 및 제1 값의 채널 개수에 기초하여 제1 채널 어텐션을 추출할 수 있다. 예를 들어, 프로세서는 높이(height), 너비(width), 깊이(depth)가 각각 4, 4, 2 의 크기를 가지는 패치에 대해, 32개의 채널에 기초하여 제1 채널 어텐션을 추출할 수 있다.

**【0045】** 또한, 프로세서는 입력 데이터(10)에 의해 생성된 제2 크기의 패치 및 제2 값의 채널 개수에 기초하여 제2 채널 어텐션을 추출할 수 있다. 이때, 상기 제2 크기는 상기 제1 크기보다 크고, 상기 제1 값은 상기 제2 값보다 작을 수 있다. 예를 들어, 프로세서는 높이, 너비, 깊이가 각각 8, 8, 4 의 크기를 가지는 패치에 대해, 64개의 채널에 기초하여 제2 채널 어텐션을 추출할 수 있다.

【0046】 프로세서는 복수의 추출 과정을 진행할수록 패치의 크기가 커지고 큰 값의 채널을 이용하여 공간 어텐션을 추출하므로, 추출 과정이 진행됨에 따라 더 많은 양의 정보를 획득할 수 있다. 예를 들어, 제1 크기의 패치 및 제1 값의 채널 크기에 기초하여, 프로세서는 스원 트랜스포머를 이용하여 제1 공간 어텐션을 추출하고, C-MSA(채널별 어텐션 모듈)를 이용하여 제1 채널 어텐션을 병렬적으로 추출할 수 있다. 그러나, 이에 한정되지 않고 채널 어텐션을 추출할 때에도 스원 트랜스포머를 이용하거나 제1 공간 어텐션을 추출하고 나서 제1 채널 어텐션을 추출하는 등의 변형도 가능하다.

【0047】 머징 과정(200)은 공간 스트림에서 추출한 공간 어텐션과 채널 스트림에서 추출한 채널 어텐션을 병합하는 단계를 포함할 수 있다. 구체적으로, 머징 과정(200)은 복수의 추출 과정의 순서에 따라 순차적으로 병합하는 탑다운 머징 과정(210)과 복수의 추출 과정의 순서와 반대의 순서로 병합하는 바텀업 머징 과정(220)을 포함할 수 있다. 프로세서는 탑다운 머징 과정(210)과 바텀업 머징 과정(220)을 이용하여 두 번의 병합을 함으로써 탑쪽의 로컬 정보와 바텀쪽의 글로벌 정보를 융합하여 많은 정보를 획득할 수 있다.

【0048】 머징 과정(200)은 머징 데이터를 생성하는 단계(S200)일 수 있다. 머징 데이터를 생성하는 단계(S200)는 탑다운 머징 과정(210) 및 바텀업 머징 과정(220)을 포함할 수 있다. 탑다운 머징 과정(210)은 공간 어텐션 및 채널 어텐션을 추출한 단계의 순서대로 공간 어텐션 및 채널 어텐션을 병합하는 과정일 수 있다. 바텀업 머징 과정(220)은 공간 어텐션 및 채널 어텐션을 추출한 단계의 반대 순서

대로 공간 어텐션 및 채널 어텐션을 병합하는 과정일 수 있다. 멀징 데이터는 탑다운 멀징 과정(210) 및 바텀업 멀징 과정(220)에서 생성된 탑다운 데이터들 및 바텀업 데이터들을 포함할 수 있다.

**【0049】** 탑다운 멀징 과정(210) 및 바텀업 멀징 과정(220)을 수행하기 전에, 공간 어텐션 및 채널 어텐션을 병합하여 퓨전(fusion) 데이터를 생성하는 단계가 선행될 수 있다. 예를 들어, 프로세서는 제1 공간 어텐션 및 제1 채널 어텐션을 병합하여 제1 퓨전 데이터를 생성하고, 제2 공간 어텐션 및 제2 채널 어텐션을 병합하여 제2 퓨전 데이터를 생성할 수 있다. 또한, 프로세서는 제3 공간 어텐션 및 제3 채널 어텐션을 병합하여 제3 퓨전 데이터를 생성할 수 있다. 이때, 제3 퓨전 데이터는 제3 공간 어텐션 및 제3 채널 어텐션에 기초하여 생성된 데이터일 수 있다. 또한 이때 제3 공간 어텐션 및 제3 채널 어텐션은 각각 제2 공간 어텐션 및 제2 채널 어텐션 이후 패치의 사이즈를 늘리고 채널 개수를 늘려 추출된 데이터일 수 있다.

**【0050】** 생성된 퓨전 데이터는 탑다운 멀징 과정(210) 및 바텀업 멀징 과정(220)에서 서로 병합될 수 있다. 그러나 이에 한정되지 않고, 퓨전 데이터의 생성 없이 멀징 과정(200)에서 곧바로 공간 어텐션 및 채널 어텐션이 병합될 수도 있다. 아래에서는 퓨전 데이터를 이용한 멀징 과정(200)에 대해서 자세히 설명한다.

**【0051】** 탑다운 멀징 과정(210) 및/또는 바텀업 멀징 과정(220)은 공간 어텐션 및 채널 어텐션을 추출하는 과정에서 다른 사이즈의 패치들을 조정하기 위한 과정(expanding 또는 merging), 학습 가능한 스칼라와 특징맵을 가중합한 뒤 스칼라

의 합으로 나누는 융합 방식인 FnF(Fast normalized fusion) 과정, 합성곱 방식(DSConv)을 포함할 수 있다.

【0052】 프로세서는 바텀업 머징 과정(220)에서 제2 퓨전 데이터 및 제3 퓨전 데이터를 병합하여 제1 바텀업 데이터를 생성할 수 있다. 이후 프로세서는 계속적으로 퓨전 데이터와 이전 단계의 바텀업 데이터를 병합하여 다음 단계의 바텀업 데이터를 생성하게 된다. 예를 들어, 프로세서는 제1 바텀업 데이터 및 제1 퓨전 데이터를 병합하여 제2 바텀업 데이터를 생성할 수 있다.

【0053】 바텀업 머징 과정(220) 이후 탑다운 머징 과정(210)이 수행될 수 있다. 도 2는 바텀업 머징 과정(220) 이후 탑다운 머징 과정(210)이 수행되는 예시를 도시하였으나, 이에 한정되지 않고, 탑다운 머징 과정(210) 이후에 바텀업 머징 과정(220)이 수행될 수도 있다. 프로세서는 탑다운 머징 과정(210)에서 제1 바텀업 데이터, 제2 바텀업 데이터 및 제2 퓨전 데이터를 병합하여 제1 탑다운 데이터를 생성할 수 있다. 이후 프로세서는 계속적으로 바텀업 데이터, 퓨전 데이터 및 이전 단계의 탑다운 데이터를 병합하여 다음 단계의 탑다운 데이터를 생성하게 된다. 예를 들어, 프로세서는 제1 탑다운 데이터, 제3 퓨전 데이터를 병합하여 제2 탑다운 데이터를 생성할 수 있다.

【0054】 도 3은 일 실시예에 따른 공간 어텐션 및 채널 어텐션 추출 방법을 설명하기 위한 도면이다.

【0055】 도 3을 참조하면, 스원 블록(Swin Block)에서 스원 트랜스포머를 이용하여 공간 어텐션을 추출하는 과정 및 C-MSA 블록에서 채널 어텐션을 추출하는 과정을 알 수 있다. 스원 블록 및 C-MSA 모듈에서 각각 N번(예, 3번) 병합할 수 있다. N이 커질수록 병합의 정도가 높아져, 많은 양의 정보가 생성될 수 있다. 예를 들어, 병합은 Depthwise 컨볼루션 후 Pointwise 컨볼루션을 수행하는 효율적인 합성곱 방식인 DSConv 방식을 사용할 수 있으나, 이에 한정되지 않는다.

【0056】 다시 도 2를 참조하면, 프로세서는 인코딩 과정(100) 및 머징 과정(200) 이후에 디코딩 과정(300)을 수행할 수 있다. 디코딩 과정(300)은 결과 데이터를 생성하는 단계(S300)일 수 있다. 구체적으로, 결과 데이터를 획득하는 단계(S300)는 머징 데이터(구체적으로, 제2 탑다운 데이터)에 의해 생성된 제3 크기의 패치 및 제3 값의 채널 개수에 기초하여 제1 디코딩 데이터를 생성하는 단계를 포함할 수 있다. 예를 들어, 프로세서는 높이, 너비, 깊이가 각각 16, 16, 8 의 크기를 가지는 패치에 대해, 128개의 채널에 기초하여 머징 데이터를 디코딩(patch expanding)할 수 있다.

【0057】 또한, 결과 데이터를 획득하는 단계(S300)는 머징 데이터(구체적으로, 제1 탑다운 데이터)에 의해 생성된 제4 크기의 패치 및 제4 값의 채널 개수에 기초하여 제2 디코딩 데이터를 생성하는 단계를 포함할 수 있다. 예를 들어, 프로세서는 높이, 너비, 깊이가 각각 8, 8, 4 의 크기를 가지는 패치에 대해, 64개의 채널에 기초하여 머징 데이터를 디코딩할 수 있다. 프로세서는 디코딩 과정을 진행 할수록 작아지는 패치의 크기 및 작아지는 채널의 개수에 기초하여 디코딩을 수행하

여 결과 데이터(20)를 출력할 수 있다. 이때, 제1 디코딩 데이터를 생성하는 단계 및 제2 디코딩 데이터를 생성하는 단계는 스원 트랜스포머를 이용하는 단계일 수 있으나, 이에 한정되지 않는다.

**【0058】** 또한, 결과 데이터를 획득하는 단계(S300)는 제1 디코딩 데이터 및 상기 제2 디코딩 데이터에 기초하여 제1 핵셀에 제1 구성 요소에 대응되는 제1 인덱스를 설정하는 단계를 포함할 수 있다. 또한, 결과 데이터를 획득하는 단계(S300)는 제1 디코딩 데이터 및 상기 제2 디코딩 데이터에 기초하여 제2 핵셀에 제2 구성 요소에 대응되는 제2 인덱스를 설정하는 단계를 포함할 수 있다.

**【0059】** 예를 들어, 프로세서는 결과 데이터(20)에 포함된 제1 핵셀이 간과 관련된 경우, 간과 관련된 제1 인덱스(예, 1)를 설정할 수 있다. 또한 예를 들어, 프로세서는 결과 데이터(20)에 포함된 제2 핵셀이 심장과 관련된 경우, 심장과 관련된 제2 인덱스(예, 2)를 설정할 수 있다.

**【0060】** 프로세서는 제1 인덱스가 설정된 핵셀들에 기초하여 제1 영역을 설정하고, 제2 인덱스가 설정된 핵셀을 포함하는 제2 영역을 설정할 수 있다. 예를 들어, 프로세서는 1로 설정된 핵셀들의 값을 변경하여 제1 색상을 이용해 간과 관련된 영역인 제1 영역을 설정할 수 있다. 또한 예를 들어, 프로세서는 2로 설정된 핵셀들의 값을 변경하여 제2 색상을 이용해 간과 관련된 영역인 제2 영역을 설정할 수 있다. 이와 같이, 프로세서는 디코딩의 결과인 결과 데이터(20)를 획득하고, 이를 이용하여 구성 요소 별 세그멘테이션을 수행할 수 있다.

【0062】 도 4 내지 도 9는 본원 발명의 이미지 세그멘테이션 방법에 따른 결과를 종래 기술과 비교하여 설명하기 위한 도면이다.

【0063】 도 4는 종래 기술과 본원 발명을 이용하여 각 구성 요소(장기)별 Dice, HD95, Params 및 FLOPs를 나타낸 표이다.

【0064】 도 4를 참조하면, 본원 발명에 따른 이미지 세그멘테이션 방법(DS-UNETR)은 구성 요소별 정확도가 종래 발명과 비교했을 때 가장 좋거나 상위권인 것을 확인할 수 있다. 구체적으로, 본원 발명에 따른 이미지 세그멘테이션 방법은 특히 비장(spleen)에 대해 높은 점수를 확보한 것을 알 수 있다. 또한, Params와 FLOPs를 보면 알 수 있듯이, 하드웨어(메모리 등) 점유량 및 계산량 등이 종래 발명보다 훨씬 낮은 것을 확인할 수 있다.

【0065】 도 5는 종래 기술과 본원 발명을 이용한 이미지 세그멘테이션 결과를 나타낸 도면이다.

【0066】 도 5를 참조하면, 본원 발명의 이미지 세그멘테이션 방법(DS-UNETR)은 첫번째 행과 같이 오른쪽 신장(right kidney)을 캐치한 것을 확인할 수 있다. 또한, 본원 발명은 두번째 행과 같이 간(liver) 사이에 위치한 쓸개(gallbladder)를 캐치한 것을 확인할 수 있다. 또한, 본원 발명은 세번째 행과 같이 간(liver) 사이에 겹쳐있는 위(stomach)를 잘 분할하였고, 네 번째 행과 같이 다른 장기와 겹친 영역이 없는 것을 확인할 수 있다. 이에, 본원 발명은 위(stomach), 콩팥(pancreas)에 대한 영상 분할 성능이 높은 것을 확인할 수 있다.

【0067】 도 6은 분할 결과를 3D 형태로 나타낸 도면이다.

【0068】 도 6을 참조하면, 본원 발명(DS-UNETR)은 다른 기술보다 Ground Truth와 가장 유사한 결과를 보인 것을 알 수 있다.

【0069】 도 7은 종래와 본원 발명의 다양한 변형의 성능을 비교한 표이다.

【0070】 도 7을 참조하면, 본원 발명과 그 변형(예, 머징부인 Bi-ASF가 없는 것 등)은 종래 기술인 Swin-Unet, Swin UNETR보다 좋은 성능을 보이는 것을 확인할 수 있다.

【0071】 도 8은 우심실, 심근, 좌심실에 대한 성능을 나타낸 표이다.

【0072】 도 8을 참조하면, 본원 발명은 다른 기술보다 적은 파라미터 수와 연산량으로 좋은 성능을 보이는 것을 확인할 수 있다.

【0073】 도 9는 파라미터 수 및 다이스 스코어에 대한 종래와 본원 발명의 비교 그래프이다.

【0074】 도 9를 참조하면, 본원 발명은 다른 기술에 비해 적은 파라미터를 가지면서 높은 dice score를 가지는 것을 확인할 수 있다.

【0076】 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 관독 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 룸(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. 상기된 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

**【0077】** 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 대체되거나 치환되더라도 적절한 결과가 달성될 수 있다.

**【0078】** 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

## 【청구범위】

### 【청구항 1】

적어도 하나 이상의 프로세서에 의해 수행되는 이미지 세그멘테이션 방법에 있어서,

입력 데이터로부터 공간 어텐션을 추출하는 단계;

상기 입력 데이터로부터 채널 어텐션을 추출하는 단계;

상기 공간 어텐션 및 상기 채널 어텐션에 기초하여 머징 데이터를 생성하는 단계; 및

상기 머징 데이터에 기초하여 구성 요소별 인덱스를 포함하는 결과 데이터를 획득하는 단계를 포함하는

이미지 세그멘테이션 방법.

### 【청구항 2】

제1항에 있어서,

상기 공간 어텐션은 구성 요소와 관련된 형태 정보를 포함하고,

상기 채널 어텐션은 구성 요소와 관련된 문맥 정보를 포함하는

이미지 세그멘테이션 방법.

### 【청구항 3】

제1항에 있어서,

상기 공간 어텐션을 추출하는 단계는 스윈 트랜스포머(swin transformer)를 이용하는 단계인

이미지 세그멘테이션 방법.

### 【청구항 4】

제1항에 있어서,

상기 공간 어텐션을 추출하는 단계는,

상기 입력 데이터에 의해 생성된 제1 크기의 패치 및 제1 값의 채널 개수에 기초하여 제1 공간 어텐션을 추출하는 단계; 및

상기 입력 데이터에 의해 생성된 제2 크기의 패치 및 제2 값의 채널 개수에 기초하여 제2 공간 어텐션을 추출하는 단계를 포함하고,

상기 제2 크기는 상기 제1 크기보다 크고, 상기 제1 값은 상기 제2 값보다 작은

이미지 세그멘테이션 방법.

### 【청구항 5】

제1항에 있어서,

상기 채널 어텐션을 추출하는 단계는,  
제1 크기의 패치 및 제1 값의 채널 개수에 기초하여 제1 채널 어텐션을 추출  
하는 단계; 및  
제2 크기의 패치 및 제2 값의 채널 개수에 기초하여 제2 채널 어텐션을 추출  
하는 단계를 포함하는  
이미지 세그멘테이션 방법.

### 【청구항 6】

제1항에 있어서,  
상기 머징 데이터를 생성하는 단계는,  
제1 공간 어텐션, 제1 채널 어텐션, 제2 공간 어텐션 및 제2 채널 어텐션에  
기초하여 제1 바텀업 데이터를 생성하는 단계; 및  
상기 제1 바텀업 데이터, 제3 공간 어텐션 및 제3 채널 어텐션에 기초하여  
제2 바텀업 데이터를 생성하는 단계를 포함하는  
이미지 세그멘테이션 방법.

### 【청구항 7】

제6항에 있어서,  
상기 머징 데이터를 생성하는 단계는,

상기 제1 바텀업 데이터 상기 제2 바텀업 데이터, 상기 제2 공간 어텐션 및  
 상기 제2 채널 어텐션에 기초하여 제1 탑다운 데이터를 생성하는 단계; 및  
 상기 제1 탑다운 데이터, 상기 제1 공간 어텐션 및 상기 제1 채널 어텐션에  
 기초하여 제2 탑다운 데이터를 생성하는 단계를 포함하는  
 이미지 세그멘테이션 방법.

### 【청구항 8】

제1항에 있어서,  
 상기 결과 데이터를 획득하는 단계는,  
 상기 머징 데이터에 의해 생성된 제3 크기의 패치 및 제3 값의 채널 개수에  
 기초하여 제1 디코딩 데이터를 생성하는 단계; 및  
 상기 머징 데이터에 의해 생성된 제4 크기의 패치 및 제4 값의 채널 개수에  
 기초하여 제2 디코딩 데이터를 생성하는 단계를 포함하고,  
 상기 제4 크기는 상기 제3 크기보다 작고, 상기 제3 값은 상기 제4 값보다  
 큼  
 이미지 세그멘테이션 방법.

### 【청구항 9】

제8항에 있어서,

상기 결과 데이터를 획득하는 단계는,

상기 제1 디코딩 데이터 및 상기 제2 디코딩 데이터에 기초하여 제1 픽셀에  
제1 구성 요소에 대응되는 제1 인덱스를 설정하는 단계;

상기 제1 디코딩 데이터 및 상기 제2 디코딩 데이터에 기초하여 제2 픽셀에  
제2 구성 요소에 대응되는 제2 인덱스를 설정하는 단계; 및

상기 제1 인덱스가 설정된 픽셀을 포함하는 제1 영역과 상기 제2 인덱스가  
설정된 픽셀을 포함하는 제2 영역을 설정하는 단계를 포함하는  
이미지 세그멘테이션 방법.

### 【청구항 10】

제8항에 있어서,

상기 제1 디코딩 데이터를 생성하는 단계 및 상기 제2 디코딩 데이터를 생성  
하는 단계는 스윈 트랜스포머(swin transformer)를 이용하는 단계인  
이미지 세그멘테이션 방법.

### 【청구항 11】

제1항 내지 제5항 중 어느 한 항에 기재된 이미지 세그멘테이션 방법을 실행  
시키도록 컴퓨터로 판독 가능한 기록 매체에 기록된 컴퓨터 프로그램.

## 【청구항 12】

전자 장치에 있어서,

메모리; 및

적어도 하나 이상의 프로세서를 포함하고,

상기 메모리는, 상기 적어도 하나 이상의 프로세서에 의해 실행 시, 상기 전자 장치가,

입력 데이터로부터 추출된 공간 어텐션을 저장하고,

상기 입력 데이터로부터 추출된 채널 어텐션을 저장하고,

상기 공간 어텐션 및 상기 채널 어텐션에 기초하여 생성된 머징 데이터를 저장하고,

상기 머징 데이터에 기초하여 획득된 구성 요소별 인덱스를 포함하는 결과 데이터를 저장하는

전자 장치.

## 【청구항 13】

입력 데이터로부터 공간 어텐션 및 채널 어텐션을 추출하는 인코딩부;

상기 공간 어텐션 및 상기 채널 어텐션에 기초하여 머징 데이터를 생성하는 머징부; 및

상기 머징 데이터에 기초하여 구성 요소별 인덱스를 포함하는 결과 데이터를

생성하는 디코딩부를 포함하는

이미지 세그멘테이션 장치.

## 【요약서】

### 【요약】

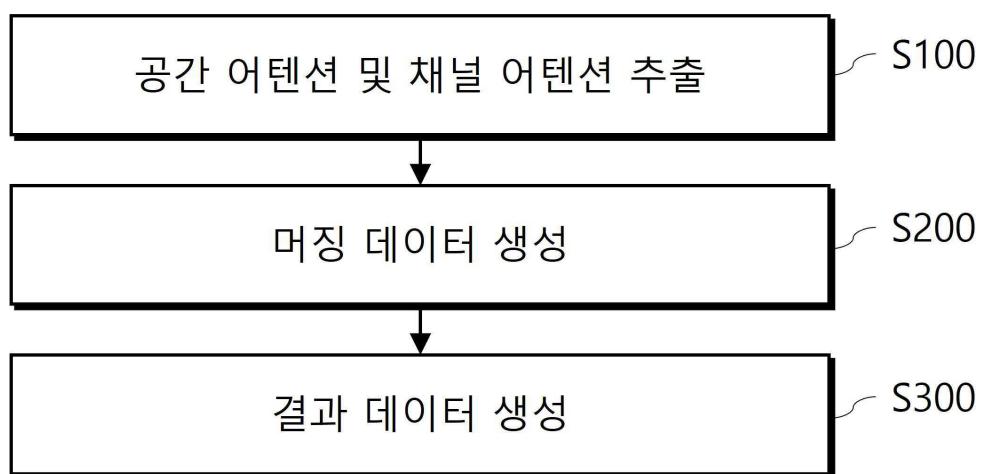
본원 발명의 이미지 세그멘테이션 방법은 입력 데이터로부터 공간 어텐션을 추출하는 단계; 상기 입력 데이터로부터 채널 어텐션을 추출하는 단계; 상기 공간 어텐션 및 상기 채널 어텐션에 기초하여 머징 데이터를 생성하는 단계; 및 상기 머징 데이터에 기초하여 구성 요소별 인덱스를 포함하는 결과 데이터를 획득하는 단계를 포함할 수 있다.

### 【대표도】

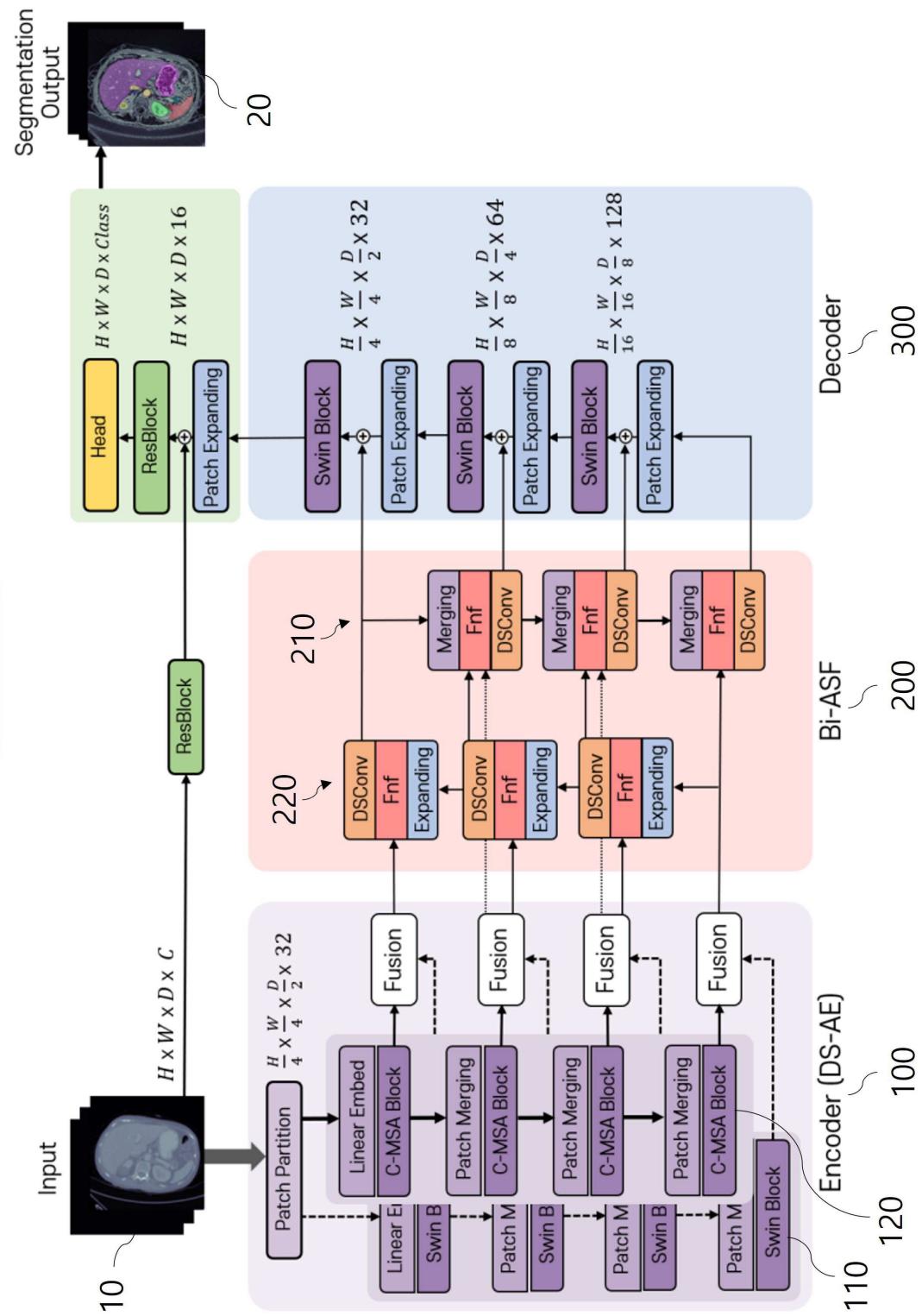
도 2

【도면】

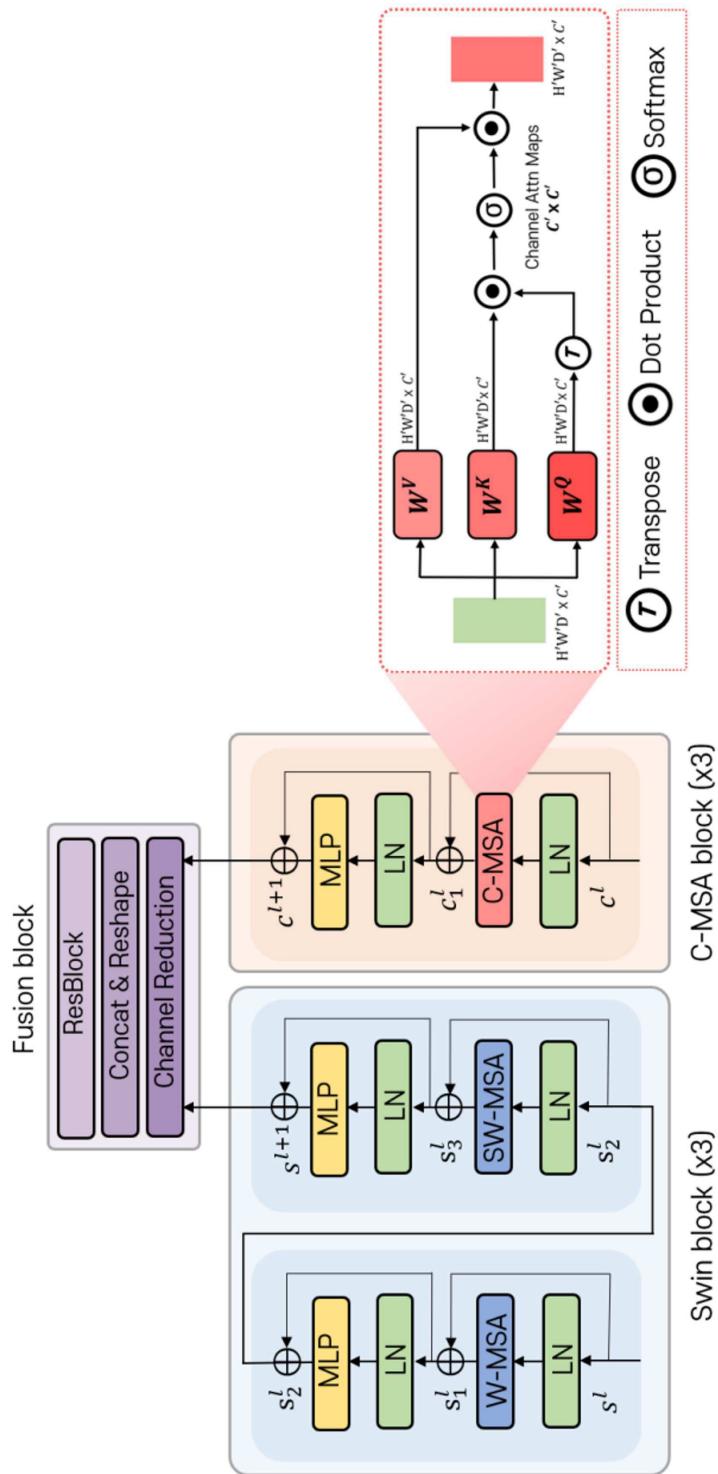
【도 1】



【Fig 2】



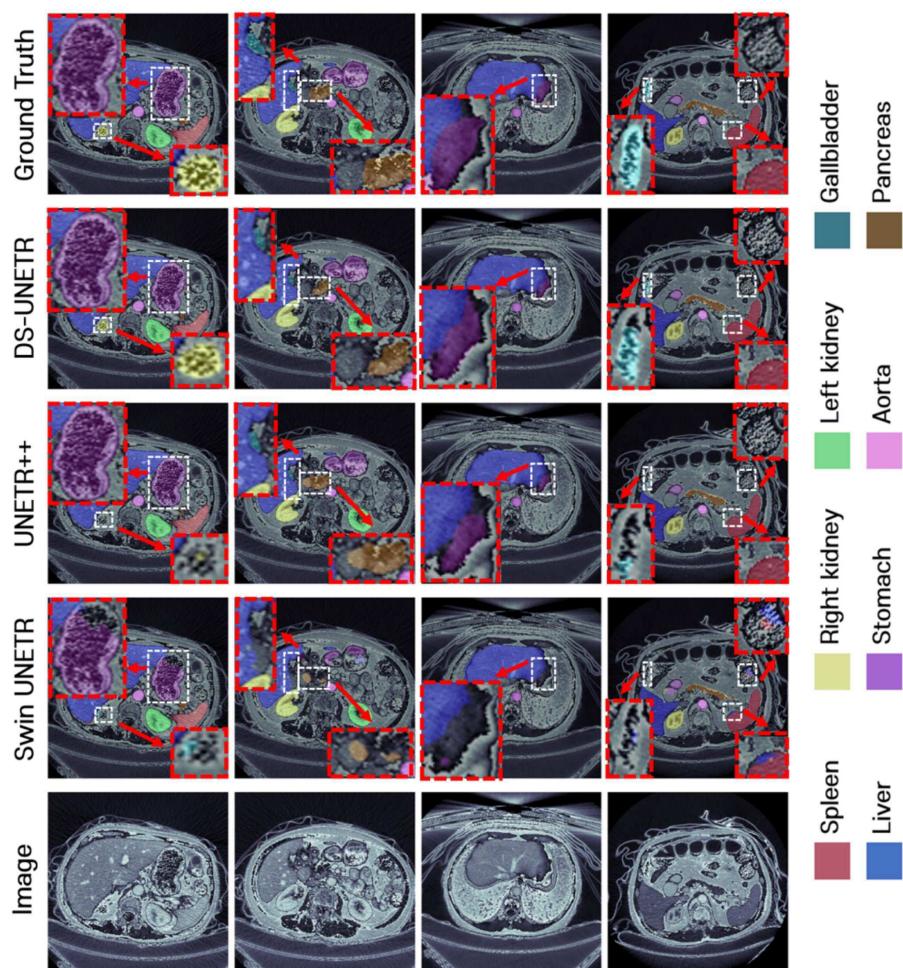
【Fig. 3】



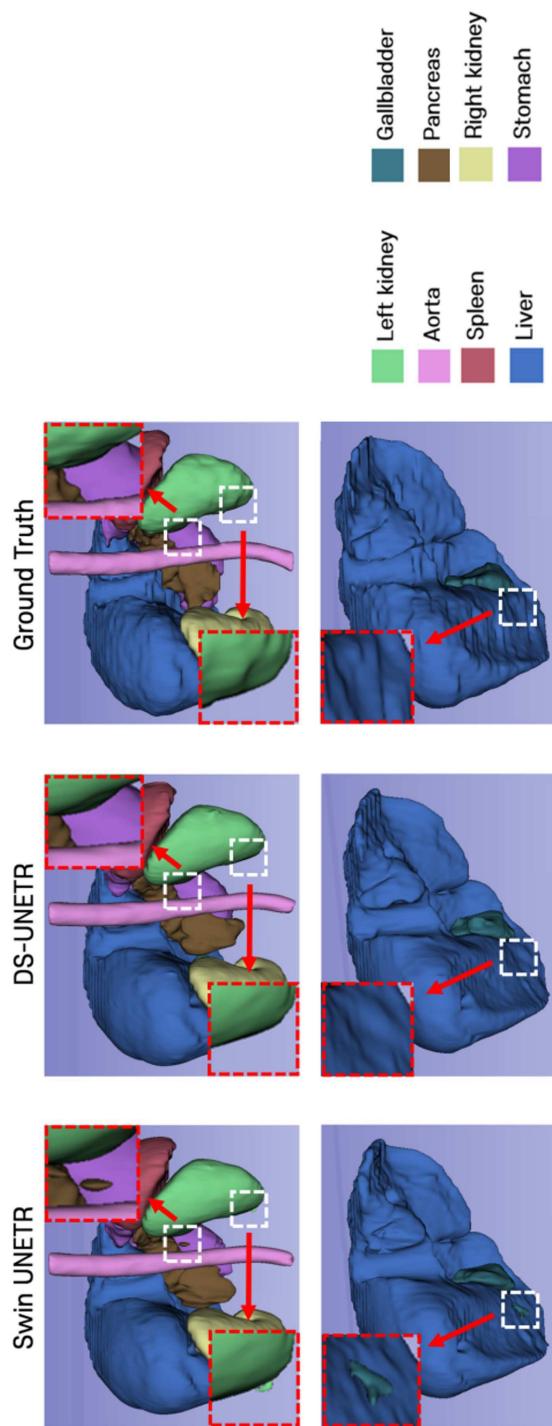
## 【Fig. 4】

Methods	Spl	RKid	LKid	Gal	Liv	Sto	Aor	Pan	Average		Params (M)	FLOPs (G)
									Dice↑	HD95↓		
TranUNet [4]	85.08	77.02	81.87	63.16	94.08	75.62	87.23	55.86	77.48	31.69	96.07	88.91
Swin-Unet[2]	90.66	79.61	83.28	66.53	94.29	76.60	85.47	56.58	79.13	21.55	-	-
UNETR [13]	87.81	84.80	85.66	60.56	94.46	73.99	89.99	59.25	79.56	22.97	92.49	75.76
MISSFormer [18]	91.92	82.00	85.21	68.65	94.41	80.81	86.99	65.67	81.96	18.20	-	-
Swin UNETR [12]	95.37	86.26	86.99	66.54	95.72	77.01	91.12	68.80	83.48	10.55	62.19	350.60
nnFormer [44]	90.51	86.25	86.57	70.17	<b>96.84</b>	<b>86.83</b>	92.04	<b>83.35</b>	86.57	10.63	150.50	213.40
UNETR++ [32]	95.77	87.18	<b>87.54</b>	71.25	<u>96.42</u>	86.01	<b>92.52</b>	81.10	<u>87.22</u>	<b>7.53</b>	42.96	47.98
DS-UNETR	<b>95.81</b>	<b>87.52</b>	87.31	<b>74.65</b>	96.38	<u>86.04</u>	91.55	80.02	<b>87.41</b>	<u>9.84</u>	<b>29.00</b>	<b>45.11</b>

【F 5】



【E 6】



(a) Comparison of 3D multi-organ image segmentation results between Swin UNETR and our proposed model (DS-UNETR).

【E 7】

Table 3: Ablation study of the DS-UNETR with Swin-Unet [2] and Swin UNETR [12].

Methods	Swin Transformer in encoder	Swin Transformer in decoder	DS-AE	Bi-ASF module	Average Dice Score
Swin-Unet [2]	✓	✓			79.13
Swin UNETR [12]	✓	✓			83.48
Ours (w/o DS-AE & Bi-ASF)	✓	✓			
Ours (w/o DS-AE)	✓	✓			84.10
Ours (w/o Bi-ASF)	✓	✓	✓	✓	85.10
Ours (w/o DS-UNETR)	✓	✓	✓	✓	86.68
Ours (DS-UNETR)					<b>87.41</b>

【H 8】

Table 2: Comparison on automatic cardiac diagnosis (ACDC). Best results are bolded. Best seconds are underlined.

Methods	Right Ventricle	Myocardium	Left Ventricle	Average Dice score	Params (M)	FLOPS (G)
TransUNet [4]	88.86	84.54	95.73	89.71	-	-
Swin-Unet [2]	88.55	85.62	<u>95.83</u>	90.00	-	-
UNETR [13]	85.29	86.52	94.02	90.32	92.69	33.83
MISSFormer [18]	86.36	85.75	91.59	87.90	-	-
mFormer [44]	90.94	<u>89.58</u>	95.65	<u>92.06</u>	37.16	47.73
UNETR++ [32]	<u>91.89</u>	<b>90.61</b>	<b>96.00</b>	<b>92.83</b>	66.80	43.71
DS-UNETR	<b>91.96</b>	87.06	94.32	91.16	<b>28.06</b>	<b>25.67</b>

【E 9】

