



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2022년11월30일

(11) 등록번호 10-2472299

(24) 등록일자 2022년11월25일

(51) 국제특허분류(Int. Cl.)
 $G06N\ 20/00$ (2019.01) $G06N\ 5/04$ (2006.01)
 $G16H\ 50/20$ (2018.01)
 (52) CPC특허분류
 $G06N\ 20/00$ (2021.08)
 $G06N\ 5/04$ (2013.01)
 (21) 출원번호 10-2022-0048760
 (22) 출원일자 2022년04월20일
 심사청구일자 2022년04월20일
 (30) 우선권주장
 1020210149285 2021년11월03일 대한민국(KR)
 (56) 선행기술조사문헌
 KR1020200010971 A
 KR1020200049451 A
 KR102093577 B1
 KR102209086 B1

(73) 특허권자
 주식회사 웨이센
 서울특별시 강남구 삼성로 524, 세화빌딩 5층 (삼성동)
 연세대학교 산학협력단
 서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
 (72) 발명자
 오상일
 서울특별시 송파구 백제고분로15길 14 신우빌 301호
 금지수
 경기도 용인시 수지구 고기로 89, 동천자이아파트 110동 704호
 (뒷면에 계속)
 (74) 대리인
 이준성

전체 청구항 수 : 총 10 항

심사관 : 박승철

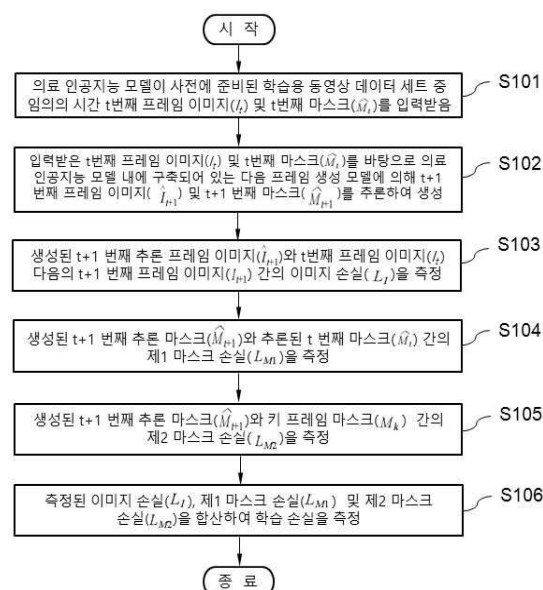
(54) 발명의 명칭 정지영상 데이터로부터 동영상 데이터를 생성하기 위한 의료 인공지능 모델의 학습 방법

(57) 요약

본 발명은 정지영상 데이터로부터 동영상 데이터를 생성하기 위한 의료 인공지능 모델의 학습 방법에 관한 것이다.

본 발명에 따른 정지영상 데이터로부터 동영상 데이터를 생성하기 위한 의료 인공지능 모델의 학습 방법은, 의료 (뒷면에 계속)

대표도



인공지능 모델이 사전에 준비된 학습용 동영상 데이터 세트 중 임의의 시간 t 번째 프레임 이미지(I_t) 및 t 번째 마스크(\hat{M}_t)를 입력받는 단계; t 번째 프레임 이미지(I_t) 및 t 번째 마스크(\hat{M}_t)를 바탕으로 다음 프레임 생성 모델에 의해 시간 $t+1$ 번째 프레임 이미지(\hat{I}_{t+1}) 및 $t+1$ 번째 마스크(\hat{M}_{t+1})를 추론하여 생성하는 단계; 생성된 $t+1$ 번째 추론 프레임 이미지(\hat{I}_{t+1})와 $t+1$ 번째 프레임 이미지(I_{t+1}) 간의 이미지 손실(L_I)을 측정하는 단계; $t+1$ 번째 추론 마스크(\hat{M}_{t+1})와 추론된 t 번째 마스크(\hat{M}_t) 간의 제1 마스크 손실(L_M)을 측정하는 단계; $t+1$ 번째 추론 마스크(\hat{M}_{t+1})와 키 프레임 마스크(M_k) 간의 제2 마스크 손실(L_{M2})을 측정하는 단계; 및 측정된 이미지 손실(L_I)과, 제1 마스크 손실(L_M) 및 제2 마스크 손실(L_{M2})을 합산하여 학습 손실을 측정하는 단계를 포함한다.

(52) CPC특허분류

G16H 50/20 (2018.01)

(72) 발명자

김경남

경기도 수원시 영통구 봉영로 1613 영통하우스토리
아파트 101동 2104호

김지현

서울특별시 강남구 언주로 211 강남세브란스병원

명세서

청구범위

청구항 1

컴퓨팅 장치에 의해 각 단계가 수행되는 의료 인공지능 모델이 정지영상 데이터로부터 동영상 데이터를 생성할 수 있도록 의료 인공지능 모델을 학습시키는 방법으로서,

- a) 상기 의료 인공지능 모델이 사전에 준비된 학습용 동영상 데이터 세트 중 임의의 시간 t 번째 프레임 이미지(I_t) 및 t 번째 마스크(\hat{M}_t)를 입력받는 단계와;
- b) 상기 입력받은 시간 t 번째 프레임 이미지(I_t) 및 t 번째 마스크(\hat{M}_t)를 바탕으로 상기 의료 인공지능 모델 내에 구축되어 있는 다음 프레임 생성 모델에 의해 시간 $t+1$ 번째 프레임 이미지(\hat{I}_{t+1}) 및 $t+1$ 번째 마스크(\hat{M}_{t+1})를 추론하여 생성하는 단계와;
- c) 상기 생성된 시간 $t+1$ 번째 추론 프레임 이미지(\hat{I}_{t+1})와 상기 t 번째 프레임 이미지(I_t) 다음의 $t+1$ 번째 프레임 이미지(I_{t+1}) 간의 이미지 손실(L_I)을 측정하는 단계와;
- d) 상기 생성된 시간 $t+1$ 번째 추론 마스크(\hat{M}_{t+1})와 추론된 t 번째 마스크(\hat{M}_t) 간의 제1 마스크 손실(L_M)을 측정하는 단계와;
- e) 상기 생성된 시간 $t+1$ 번째 추론 마스크(\hat{M}_{t+1})와 키 프레임 마스크(M_k) 간의 제2 마스크 손실(L_{M2})을 측정하는 단계; 및
- f) 상기 단계 c)에서 측정된 이미지 손실(L_I)과, 상기 단계 d)에서 측정된 제1 마스크 손실(L_M) 및 상기 단계 e)에서 측정된 제2 마스크 손실(L_{M2})을 합산하여 학습 손실을 측정하는 단계를 포함하는, 정지영상 데이터로부터 동영상 데이터를 생성하기 위한 의료 인공지능 모델의 학습 방법.

청구항 2

제1항에 있어서,

상기 단계 f) 이후에 상기 측정된 학습 손실 결과를 바탕으로 상기 의료 인공지능 모델의 학습 중단 조건을 만족하는지를 판별하는 단계와;

학습 중단 조건을 만족하지 않으면 프로세스 진행을 상기 단계 a)로 회귀시키고, 학습 중단 조건을 만족하면 프로세스를 종료하는 단계를 더 포함하는, 정지영상 데이터로부터 동영상 데이터를 생성하기 위한 의료 인공지능 모델의 학습 방법.

청구항 3

제1항에 있어서,

상기 단계 a)에서 상기 학습용 동영상 데이터 세트는 하나의 동영상에 하나의 병변이 포함된, 정지영상 데이터로부터 동영상 데이터를 생성하기 위한 의료 인공지능 모델의 학습 방법.

청구항 4

제1항에 있어서,

상기 단계 a)에서 상기 동영상 데이터 세트의 입력 방식은 첫 번째 프레임부터 순차적으로 입력하는 방식, 마지막 프레임부터 역순으로 입력하는 방식, 순차적 입력과 역순 입력을 융합하여 입력하는 방식, 키 프레임을 기준으로 순차적 입력+역순 입력 방식 중 어느 하나의 방식인, 정지영상 데이터로부터 동영상 데이터를 생성하기 위한 의료 인공지능 모델의 학습 방법.

청구항 5

제1항에 있어서,

상기 단계 b)에서의 상기 다음 프레임 생성 모델은 다음 프레임 이미지 생성 부분과 마스크 생성 부분이 서로 다른 구조로 구성된, 정지영상 데이터로부터 동영상 데이터를 생성하기 위한 의료 인공지능 모델의 학습 방법.

청구항 6

제1항에 있어서,

상기 단계 b)에서의 상기 다음 프레임 생성 모델은 다음 프레임 이미지 생성 부분과 마스크 생성 부분의 일부 구조는 상호 공유되고, 나머지 구조는 분기되는 형상으로 구성된, 정지영상 데이터로부터 동영상 데이터를 생성하기 위한 의료 인공지능 모델의 학습 방법.

청구항 7

제1항에 있어서,

상기 단계 b)에서의 상기 다음 프레임 생성 모델은 엔코더(encoder)와 디코더(decoder)로 구성되는 생성자(generator) 형태로 구성된, 정지영상 데이터로부터 동영상 데이터를 생성하기 위한 의료 인공지능 모델의 학습 방법.

청구항 8

제1항에 있어서,

상기 단계 e)의 키 프레임 마스크(M_k)에서의 키 프레임은 상기 동영상 데이터 세트 내에서 특정 기준에 의해 선별된 한 장 이상의 프레임인, 정지영상 데이터로부터 동영상 데이터를 생성하기 위한 의료 인공지능 모델의 학습 방법.

청구항 9

제8항에 있어서,

상기 특정 기준은 영상의 선명도, 영상 내 병변의 위치, 밝기 값을 포함하는, 정지영상 데이터로부터 동영상 데이터를 생성하기 위한 의료 인공지능 모델의 학습 방법.

청구항 10

제8항에 있어서,

상기 키 프레임은 의료진 또는 병변 영역을 검출하는 기학습된 인공지능 모델에 의해 병변 영역이 태깅되는, 정지영상 데이터로부터 동영상 데이터를 생성하기 위한 의료 인공지능 모델의 학습 방법.

발명의 설명

기술 분야

[0001] 본 발명은 의료 인공지능 모델의 학습 방법에 관한 것으로서, 더 상세하게는 인공지능 기반 동영상 생성 모델을 도입하여 동영상 데이터 수집 및 동영상 데이터 타겟 영역 태깅 작업이 제한적인 상황에서 동영상 데이터를 활용할 수 있도록 하는정지영상 데이터로부터 동영상 데이터를 생성하기 위한 의료 인공지능 모델의 학습 방법에 관한 것이다.

배경 기술

[0003] CT(Computed Tomography) 및 MRI(Magnetic Resonance Imaging)와 같이 촬영 후 진단에 활용하는 환경과 달리, 내시경이나 초음파 등 실시간 검사가 이루어지는 환경에서는 실시간 검사가 가능한 인공지능 모델이 필요하다. 실시간 검사 환경(진향적 진단)에서는 초당 약 30 프레임의 이미지가 연속적으로 입력되며, 검사자의 움직임, 피검사자의 호흡 등에 의해 크고 작은 노이즈가 지속적으로 촬영된다. 따라서, 실시간 의료 진단 보조를 위한 내시경(또는 초음파 기기 등) 검사 환경에서 동작하는 인공지능 모델의 도입이 필요하다.

[0004] 하지만, 현존하는 대부분의 의료 인공지능 모델의 경우 깨끗(선명)하게 촬영된 정지 영상을 수집하여 인공지능 모델의 학습에 활용하고 있으며, 이에 따라 실시간 검사 환경에서 발생하는 다양한 위협요소를 다루지 못하고 있다. 동영상 데이터를 활용한 대부분의 기존 방법은 수동 프레임 선택(manual frame selection) 과정을 적용하는 수준이다. 동영상 데이터의 효과 및 필요성은 자명하나, 데이터 수집 및 타겟 영역 태깅의 어려움 등으로 의료 인공지능 모델에 동영상 데이터를 도입하는 것이 현실적으로 쉽지 않다.

[0005] 한편, 한국 공개특허공보 제10-2021-0072362호(특허문헌 1)에는 "인공 지능 모델에 대한 학습 데이터를 생성하는 인공 지능 장치 및 그 방법"이 개시되어 있는 바, 이에 따른 학습 데이터를 생성하는 방법은, 마이크론 또는 카메라를 포함하는 입력부를 통해 입력 데이터를 수신하는 단계; 인공 지능 모델을 이용하여 상기 입력 데이터에 대응하는 추론 결과를 생성하는 단계; 상기 추론 결과에 대응하는 피드백을 수신하는 단계; 상기 인공 지능 모델의 갱신에 대한 상기 입력 데이터 및 상기 피드백의 적합성을 판단하는 단계; 및 상기 입력 데이터 및 상기 피드백이 상기 인공 지능 모델의 갱신에 적합한 것으로 판단된 경우에 상기 입력 데이터 및 상기 피드백에 기초하여 학습 데이터를 생성하는 단계를 포함하는 것을 특징으로 한다.

[0006] 이상과 같은 특허문헌 1의 경우, 입력 데이터와 그에 대응하는 피드백이 인공 지능 모델의 갱신에 적합한 데이터인지 판단하고, 인공 지능 모델에 적합한 데이터에 기초하여 학습 데이터를 생성함으로써, 우연히 또는 악의적으로 잘못된 데이터가 생성 및 수집된다 하더라도, 인공 지능 모델의 갱신에 부적절한 데이터를 배제하여 인공 지능 모델이 부적절하게 갱신되는 상황을 방지할 수 있는 장점이 있기는 하나, 이는 기술의 핵심이 인공 지능 모델의 갱신에 적합한 데이터를 생성하는 것에 맞추어져 있어, 동영상 데이터 수집 및 태깅이 제한적인 상황에서 동영상 데이터를 활용할 수 있도록 하는 것에는 적용하기 어려운 한계가 있다.

선행기술문헌

특허문헌

[0007] (특허문헌 0001) 한국 공개특허공보 제10-2021-0072362호(2021.06.17.)

발명의 내용

해결하려는 과제

[0008] 본 발명은 상기와 같은 사항을 종합적으로 감안하여 창출된 것으로서, 인공지능 기반 동영상 생성 모델을 도입하여 동영상 데이터 수집 및 동영상 데이터 타겟 영역 태깅 작업이 제한적인 상황에서 동영상 데이터를 활용할

수 있도록 하는 정지영상 데이터로부터 동영상 데이터를 생성하기 위한 의료 인공지능 모델의 학습 방법을 제공함에 그 목적이 있다.

[0009] 본 발명의 다른 목적은 동영상 데이터는 확보했으나, 타겟 병변 영역 태깅 작업이 필요한 경우에도 활용할 수 있는 정지영상 데이터로부터 동영상 데이터를 생성하기 위한 의료 인공지능 모델의 학습 방법을 제공함에 있다.

과제의 해결 수단

[0010] 상기의 목적을 달성하기 위하여 본 발명에 따른 정지영상 데이터로부터 동영상 데이터를 생성하기 위한 의료 인공지능 모델의 학습 방법은,

[0011] 컴퓨팅 장치에 의해 각 단계가 수행되는 의료 인공지능 모델이 정지영상 데이터로부터 동영상 데이터를 생성할 수 있도록 의료 인공지능 모델을 학습시키는 방법으로서,

[0012] a) 상기 의료 인공지능 모델이 사전에 준비된 학습용 동영상 데이터 세트 중 임의의 시간 t 번째 프레임 이미지 (I_t) 및 t 번째 마스크(\hat{M}_t)를 입력받는 단계와;

[0013] b) 상기 입력받은 시간 t 번째 프레임 이미지(I_t) 및 t 번째 마스크(\hat{M}_t)를 바탕으로 상기 의료 인공지능 모델 내에 구축되어 있는 다음 프레임 생성 모델에 의해 시간 $t+1$ 번째 프레임 이미지(\hat{I}_{t+1}) 및 $t+1$ 번째 마스크(\hat{M}_{t+1})를 추론하여 생성하는 단계와;

[0014] c) 상기 생성된 시간 $t+1$ 번째 추론 프레임 이미지(\hat{I}_{t+1})와 상기 t 번째 프레임 이미지(I_t) 다음의 $t+1$ 번째 프레임 이미지(I_{t+1}) 간의 이미지 손실(L_I)을 측정하는 단계와;

[0015] d) 상기 생성된 시간 $t+1$ 번째 추론 마스크(\hat{M}_{t+1})와 추론된 t 번째 마스크(\hat{M}_t) 간의 제1 마스크 손실(L_M)을 측정하는 단계와;

[0016] e) 상기 생성된 시간 $t+1$ 번째 추론 마스크(\hat{M}_{t+1})와 키 프레임 마스크(M_k) 간의 제2 마스크 손실(L_M)을 측정하는 단계; 및

[0017] f) 상기 단계 c)에서 측정된 이미지 손실(L_I)과, 상기 단계 d)에서 측정된 제1 마스크 손실(L_M) 및 상기 단계 e)에서 측정된 제2 마스크 손실(L_M)을 합산하여 학습 손실을 측정하는 단계를 포함하는 점에 그 특징이 있다.

[0018] 여기서, 상기 단계 f) 이후에 상기 측정된 학습 손실 결과를 바탕으로 상기 의료 인공지능 모델의 학습 중단 조건을 만족하는지를 판별하는 단계와; 학습 중단 조건을 만족하지 않으면 프로세스 진행을 상기 단계 a)로 회귀시키고, 학습 중단 조건을 만족하면 프로세스를 종료하는 단계를 더 포함할 수 있다.

[0019] 또한, 상기 단계 a)에서 상기 학습용 동영상 데이터 세트는 하나의 동영상에 하나의 병변이 포함될 수 있다.

[0020] 또한, 상기 단계 a)에서 상기 동영상 데이터 세트의 입력 방식은 첫 번째 프레임부터 순차적으로 입력하는 방식, 마지막 프레임부터 역순으로 입력하는 방식, 순차적 입력과 역순 입력을 융합하여 입력하는 방식, 키 프레임을 기준으로 순차적 입력+역순 입력 방식 중 어느 하나의 방식일 수 있다.

[0021] 또한, 상기 단계 b)에서의 상기 다음 프레임 생성 모델은 다음 프레임 이미지 생성 부분과 마스크 생성 부분으로 서로 다른 구조로 구성될 수 있다.

[0022] 또한, 상기 단계 b)에서의 상기 다음 프레임 생성 모델은 다음 프레임 이미지 생성 부분과 마스크 생성 부분의 일부 구조는 상호 공유되고, 나머지 구조는 분기되는 형상으로 구성될 수 있다.

[0023] 또한, 상기 단계 b)에서의 상기 다음 프레임 생성 모델은 엔코더(encoder)와 디코더(decoder)로 구성되는 생성자(generator) 형태로 구성될 수 있다.

- [0024] 또한, 상기 단계 e)의 키 프레임 마스크(M_k)에서의 키 프레임은 상기 동영상 데이터 세트 내에서 특정 기준에 의해 선별된 한 장 이상의 프레임일 수 있다.
- [0025] 이때, 상기 특정 기준은 영상의 선명도, 영상 내 병변의 위치, 밝기 값 등을 포함할 수 있다.
- [0026] 이때, 또한 상기 키 프레임은 의료진 또는 병변 영역을 검출하는 기학습된 인공지능 모델에 의해 병변 영역이 태깅될 수 있다.

발명의 효과

- [0028] 이와 같은 본 발명에 의하면, 인공지능 기반 동영상 생성 모델을 도입하여 동영상 데이터 수집 및 동영상 데이터 타겟 영역 태깅 작업이 제한적인 상황에서 동영상 데이터를 활용할 수 있도록 함으로써, 동영상 데이터 활용에 요구되는 시간과 인력을 최소화 할 수 있고, 수집된 정지영상 DB를 통해 동영상 학습 모델의 효과를 얻을 수 있으며, 실시간 진단 환경에 효과적으로 의료 인공지능 모델을 적용할 수 있는 장점이 있다.

도면의 간단한 설명

- [0030] 도 1은 본 발명에 따른 정지영상 데이터로부터 동영상 데이터를 생성하기 위한 의료 인공지능 모델의 학습 방법의 실행 과정을 나타낸 흐름도이다.
- 도 2는 본 발명에 따른 의료 인공지능 모델의 학습 방법에 있어서, 동영상과 정답영역을 생성하기 위한 인공지능 모델의 학습 과정을 나타낸 도면이다.
- 도 3은 본 발명의 방법에 채용되는 다음 프레임 생성 모델에 의해 임의의 시간 t 의 이미지와 마스크를 제공받아 시간 $t+1$ 의 이미지와 마스크를 각각 추론하는 개요를 나타낸 도면이다.
- 도 4는 본 발명에 따른 정지영상 데이터로부터 동영상 데이터를 생성하기 위한 의료 인공지능 모델의 학습 방법의 프로세스 진행 과정을 나타낸 흐름도이다.
- 도 5는 동영상 데이터가 없는 경우 정지 영상으로부터 동영상 데이터 및 정답 영역 마스크를 생성하는 프로세스 진행 과정을 나타낸 흐름도이다.
- 도 6은 동영상 데이터는 있으나, 병변 태깅이 없는 경우 동영상 데이터를 태깅을 위해 사용하는 프로세스 진행 과정을 나타낸 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0031] 본 명세서 및 청구범위에 사용된 용어나 단어는 통상적이거나 사전적인 의미로 한정되어 해석되지 말아야 하며, 발명자는 그 자신의 발명을 가장 최선의 방법으로 설명하기 위해 용어의 개념을 적절하게 정의할 수 있다는 원칙에 입각하여 본 발명의 기술적 사상에 부합하는 의미와 개념으로 해석되어야 한다.
- [0032] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있다는 것을 의미한다. 또한, 명세서에 기재된 "...부", "...기", "모듈", "장치" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.
- [0033] 이하 첨부된 도면을 참조하여 본 발명의 실시예를 상세히 설명한다.
- [0034] 도 1은 본 발명의 실시예에 따른 정지영상 데이터로부터 동영상 데이터를 생성하기 위한 의료 인공지능 모델의 학습 방법의 실행 과정을 나타낸 흐름도이고, 도 2은 본 발명에 따른 의료 인공지능 모델의 학습 방법에 있어서, 동영상과 정답영역을 생성하기 위한 인공지능 모델의 학습 과정을 나타낸 도면이며, 도 3은 본 발명의 방법에 채용되는 다음 프레임 생성 모델에 의해 임의의 시간 t 의 이미지와 마스크를 제공받아 시간 $t+1$ 의 이미지와 마스크를 각각 추론하는 개요를 나타낸 도면이다.
- [0035] 도 1 내지 도 3을 참조하면, 본 발명에 따른 정지영상 데이터로부터 동영상 데이터를 생성하기 위한 의료 인공지능 모델의 학습 방법은, 컴퓨팅 장치에 의해 각 단계가 수행되는 의료 인공지능 모델이 정지영상 데이터로부터 동영상 데이터를 생성할 수 있도록 의료 인공지능 모델을 학습시키는 방법으로서, 먼저 상기 의료 인공지능 모델(210)이 사전에 준비된 학습용 동영상 데이터 세트 중 임의의 시간 t 번째 프레임 이미지(I_t)(205) 및 t

번째 마스크(\hat{M}_t)(206)를 입력받는다(단계 S101). 여기서, 상기 학습용 동영상 데이터 세트의 구성 조건은 하나의 동영상에 하나의 병변이 포함될 수 있다. 물론, 하나의 장면(scene)에 동일한 형태의 병변이 하나 이상인 경우도 가능하다. 또한, 상기 동영상 데이터 세트의 입력 방식은 첫 번째 프레임부터 순차적으로 입력하는 방식, 마지막 프레임부터 역순으로 입력하는 방식, 순차적 입력과 역순 입력을 융합하여 입력하는 방식, 키 프레임(key frame)을 기준으로 순차적 입력+역순 입력 방식 중 어느 하나의 방식일 수 있다.

[0036] 이렇게 하여 시간 t 번째 프레임 이미지(I_t)(205) 및 t 번째 마스크(\hat{M}_t)(206)가 입력되면, 의료 인공지능 모델(210)은 입력받은 시간 t 번째 프레임 이미지(I_t) 및 t 번째 마스크(\hat{M}_t)를 바탕으로 내부에 구축되어 있는 다음 프레임 생성 모델(211)에 의해 시간 t+1 번째 프레임 이미지(\hat{I}_{t+1})(212) 및 t+1 번째 마스크(\hat{M}_{t+1})(213)를 추론하여 생성한다(단계 S102). 여기서, 상기 다음 프레임 생성 모델(211)은 다음 프레임 이미지 생성 부분과 마스크 생성 부분이 서로 다른 구조로 구성될 수 있다. 또한, 상기 다음 프레임 생성 모델(211)은 다음 프레임 이미지 생성 부분과 마스크 생성 부분의 일부 구조는 상호 공유되고, 나머지 구조는 분기되는 형상으로 구성될 수 있다. 또한, 상기 다음 프레임 생성 모델(211)은 인코더(encoder)(211a)와 디코더(decoder)(211b)(211c)로 구성되는 생성자(generator) 형태로 구성될 수 있다. 이때, 상기 인코더(211a)는 convolutional layer 그룹으로 구성될 수 있다. 그리고 상기 디코더(211b)(211c)는 deconvolutional layer 혹은 upscale layer 그룹으로 구성될 수 있다. 또한, 이상과 같은 다음 프레임 생성 모델(211)은 학습 시 생성자(generator)와 감별자(discriminator)로 구성되는 GAN(Generative Adversarial Network) 방식이 사용될 수 있다. 이때, 생성자(generator)와 감별자(discriminator)의 구조에는 제한이 없다.

[0037] 이상에 의해 시간 t+1 번째 프레임 이미지(\hat{I}_{t+1})(212) 및 t+1 번째 마스크(\hat{M}_{t+1})(213)의 추론 생성이 완료되면, 그 생성된 시간 t+1 번째 추론 프레임 이미지(\hat{I}_{t+1})(212)와 상기 t 번째 프레임 이미지(I_t) 다음의 t+1 번째 프레임 이미지(I_{t+1}) 간의 이미지 손실(L_I)(214)을 측정한다(단계 S103). 여기서, 이와 같은 이미지 손실(L_I)(214)은 실제의 t 번째 프레임 이미지(I_t) 다음의 t+1 번째 프레임 이미지(I_{t+1})와 생성된 t+1 번째 추론 프레임 이미지(\hat{I}_{t+1})(212) 간의 유사도를 측정함으로써 이미지 손실의 정도를 알 수 있다. 예를 들면, 실제의 이미지(I_{t+1})와 추론 이미지(\hat{I}_{t+1})(212) 간의 유사도 값이 크면 이미지 손실이 작은 것으로 볼 수 있고, 유사도 값이 작으면 이미지 손실이 큰 것으로 볼 수 있다.

[0038] 이상과 같은 이미지 손실(L_I)은 다음과 같은 수식 관계로 나타낼 수 있다.

수학적식 1

$$L_I(I_{t+1}, \hat{I}_{t+1}) = \frac{1}{N_V N_F} \sum_{i=1}^{N_V} \sum_{f=1}^{N_{F_i}-1} \phi(I_{t+1}, \hat{I}_{t+1})$$

[0039]

[0040] 위의 수식에서 N_V 는 학습 동영상 수, N_{F_i} 는 동영상 V_i 의 전체 프레임 수를 나타내며, $\phi(I_{t+1}, \hat{I}_{t+1})$ 의 해를 구하기 위해 Dice coefficient, L1 loss, L2 loss 등을 포함하는 2차원 데이터 유사도 측정 방법이 사용될 수 있다.

[0041] 이후, 상기 단계 S102에서 생성된 시간 t+1 번째 추론 마스크(\hat{M}_{t+1})(213)와, 추론된 t 번째 마스크(\hat{M}_t)(206b)

간의 제1 마스크 손실(L_{M1})(215)을 측정한다(단계 S104). 여기서, 이와 같은 제1 마스크 손실(L_{M1})은 입력된 (추론된) t 번째 마스크(\hat{M}_t)(206b)와 t+1 번째 추론 마스크(\hat{M}_{t+1})(213) 간의 유사도를 측정함으로써 제1 마스크 손실의 정도를 알 수 있다. 예를 들면, 입력된(추론된) t 번째 마스크 (\hat{M}_t)(206b)와 t+1 번째 추론 마스크(\hat{M}_{t+1})(213) 간의 유사도 값이 크면 제1 마스크 손실이 작은 것으로 볼 수 있고, 유사도 값이 작으면 제1 마스크 손실이 큰 것으로 볼 수 있다.

[0042] 여기서, 이상과 같은 제1 마스크 손실(L_{M1})은 다음과 같은 수식 관계로 나타낼 수 있다.

수학식 2

$$L_{M1}(\hat{M}_t, \hat{M}_{t+1}) = \frac{1}{N_V N_F} \sum_{i=1}^{N_V} \sum_{f=1}^{N_{F_i}-1} \theta(\phi(M_A, M_B), d_{AB})$$

[0043]

위의 수식에서, $\theta(\phi(M_A, M_B), d_{AB})$ 는 프레임 A와 B 사이의 유사도 d_{AB} 에 따라 $\phi(M_A, M_B)$ (이미지 손실과 동일)에 페널티(penalty)를 주는 방식이다. 이는 두 마스크의 원본 이미지 간의 유사도에 따라 페널티를 적용하는 방식이다. 즉, 원본 이미지가 유사할수록 두 마스크도 유사해야 한다.

[0044]

이렇게 하여 제1 마스크 손실(L_{M1})의 측정이 완료되면, 상기 단계 S102에서 생성된 시간 t+1 번째 추론 마스크(\hat{M}_{t+1})와 키 프레임 마스크(M_k) 간의 제2 마스크 손실(L_{M2})을 측정한다(단계 S105). 여기서, 상기 키 프레임 마스크(M_k)에서의 키 프레임은 상기 동영상 데이터 세트 내에서 특정 기준에 의해 선별된 한 장 이상의 프레임일 수 있다. 이때, 상기 특정 기준은 영상의 선명도, 영상 내 병변의 위치, 밝기 값 등을 포함할 수 있다. 이때, 또한 상기 키 프레임은 제3자, 예를 들면, 의료진 또는 병변 영역을 검출하는 기학습된 인공지능 모델에 의해 병변 영역이 태깅될 수 있다.

[0045]

여기서, 또한 상기 제2 마스크 손실(L_{M2})은 키 프레임 마스크(M_k)와 t+1 번째 추론 마스크(\hat{M}_{t+1})(213) 간의 유사도를 측정함으로써 제2 마스크 손실의 정도를 알 수 있다. 예를 들면, 키 프레임 마스크(M_k)와 t+1 번째 추론 마스크(\hat{M}_{t+1})(213) 간의 유사도 값이 크면 제2 마스크 손실이 작은 것으로 볼 수 있고, 유사도 값이 작으면 제2 마스크 손실이 큰 것으로 볼 수 있다.

[0046]

여기서, 이상과 같은 제2 마스크 손실(L_{M2})은 다음과 같은 수식 관계로 나타낼 수 있다.

수학식 3

$$L_{M2}(\hat{M}_k, \hat{M}_{t+1}) = \frac{1}{N_V N_F} \sum_{i=1}^{N_V} \sum_{f=1}^{N_{F_i}-1} \theta(\phi(M_A, M_B), d_{AB})$$

[0048]

위의 수식에서, 마찬가지로 $\theta(\phi(M_A, M_B), d_{AB})$ 는 프레임 A와 B 사이의 유사도 d_{AB} 에 따라 $\phi(M_A, M_B)$ (이미지 손

[0049]

실과 동일)에 페널티(penalty)를 주는 방식이다. 전술한 바와 같이, 이는 두 마스크의 원본 이미지 간의 유사도에 따라 페널티를 적용하는 방식이다. 다시 말해서, 원본 이미지가 유사할수록 두 마스크도 유사해야 한다.

[0050] 이상과 같은 제1 마스크 손실(L_{M1})과 제2 마스크 손실(L_{M2})에 공통으로 적용되는 $\theta(\phi(M_A, M_B), d_{AB})$ 은 다음과 같은 값으로 주어질 수 있다.

수식 4

$$\theta(\phi(M_A, M_B), d_{AB}) = \begin{cases} \phi(M_A, M_B), & \text{if } d_{AB} \geq \gamma_1 \\ \max(0, \gamma_2 - \phi(M_A, M_B)) & \text{else} \end{cases}$$

[0051]

[0052] 위의 수식에서 γ_1 는 포지티브 쌍(positive pair)과 네거티브 쌍(negative pair)을 나누는 마진(margin) 값이며, d_{AB} 가 클수록 비슷한 프레임이다. 또한, γ_2 는 거리 $\phi(\cdot, \cdot)$ 의 마진 값을 나타낸다.

[0053] 이상에 의해 제2 마스크 손실(L_{M2})의 측정이 완료되면, 상기 단계 S103에서 측정된 이미지 손실(L_I)과, 상기 단계 S104에서 측정된 제1 마스크 손실(L_{M1}) 및 상기 단계 S105에서 측정된 제2 마스크 손실(L_{M2})을 합산하여 학습 손실을 측정한다(단계 S106). 여기서, 이와 같은 학습 손실은 그 손실값이 작을수록 학습 결과(효과)가 양호하다고 할 수 있으며, 이는 곧 의료 인공지능 모델의 성능이 우수함을 의미하는 것으로 볼 수 있다.

[0054] 여기서, 또한 상기 단계 S106 이후에 상기 측정된 학습 손실 결과를 바탕으로 상기 의료 인공지능 모델의 학습 중단 조건을 만족하는지를 판별하는 단계와; 학습 중단 조건을 만족하지 않으면 프로세스 진행을 상기 단계 S101로 회귀시키고, 학습 중단 조건을 만족하면 프로세스를 종료하는 단계를 더 포함할 수 있다. 이 부분에 대해서는 아래에서 다시 설명하기로 한다.

[0055] 이하에서는 이상과 같은 본 발명에 따른 정지영상 데이터로부터 동영상 데이터를 생성하기 위한 의료 인공지능 모델의 학습 방법과 관련하여 조금 더 설명을 부가해 보기로 한다.

[0056] 도 4는 본 발명에 따른 정지영상 데이터로부터 동영상 데이터를 생성하기 위한 의료 인공지능 모델의 학습 방법의 프로세스 진행 과정을 나타낸 흐름도이다.

[0057] 도 4를 참조하면, 먼저 상기 의료 인공지능 모델(210)이 사전에 준비된 학습용 동영상 데이터 세트 중 임의의 시간 t 번째 프레임 이미지(I_t)(205)와 t 번째 마스크(\hat{M}_t)(206)를 입력받는다(단계 S401, S403). 이때, 시간 t=0인지를 판별하여(단계 S402), t=0이 아니면 t 번째 마스크, 즉 추론된 t 번째 마스크(\hat{M}_t)(206b)(도 2 참조)를 입력받고(단계 S403), t=0이면 랜덤으로 t 번째 마스크(\hat{M}_t)(206a)(도 2 참조)를 입력받는다(단계 S404).

[0058] 이렇게 하여 시간 t 번째 프레임 이미지(I_t)(205) 및 t 번째 마스크(\hat{M}_t)(206)가 입력되면, 의료 인공지능 모델(210)은 입력받은 시간 t 번째 프레임 이미지(I_t) 및 t 번째 마스크(\hat{M}_t)를 바탕으로 다음 프레임 생성 모델(211)에 의해 시간 t+1 번째 프레임 이미지(\hat{I}_{t+1})(212) 및 t+1 번째 마스크(\hat{M}_{t+1})(213)를 추론하여 생성한다(단계 S405~S407).

[0059] 이상에 의해 시간 t+1 번째 프레임 이미지(\hat{I}_{t+1})(212) 및 t+1 번째 마스크(\hat{M}_{t+1})(213)의 추론 생성이 완료되면, 그 생성된 시간 t+1 번째 추론 프레임 이미지(\hat{I}_{t+1})(212)(S406)와 상기 t 번째 프레임 이미지(I_t) 다음의 t+1

번째 프레임 이미지(I_{t+1})(S408) 간의 이미지 손실(L_I)을 측정한다(단계 S409).

[0060] 그리고 상기 단계 S407에서 생성된 시간 t+1 번째 추론 마스크(\hat{M}_{t+1})(213)와, 상기 단계 S403에서의 추론된 t 번째 마스크(\hat{M}_t)(206b) 간의 제1 마스크 손실 (L_{M1})을 측정한다(단계 S410).

[0061] 또한, 상기 단계 S407에서 생성된 시간 t+1 번째 추론 마스크(\hat{M}_{t+1})와 키 프레임 마스크(M_k)(S411) 간의 제2 마스크 손실(L_{M2})을 측정한다(단계 S412).

[0062] 이상에 의해 제2 마스크 손실(L_{M2})까지 측정이 완료되면, 상기 단계 S409에서 측정된 이미지 손실(L_I)과, 상기 단계 S410에서 측정된 제1 마스크 손실(L_{M1}) 및 상기 단계 S412에서 측정된 제2 마스크 손실(L_{M2})을 합산하여 학습 손실을 측정한다(단계 S413).

[0063] 이후, 상기 측정된 학습 손실 결과를 바탕으로 상기 의료 인공지능 모델의 학습 중단 조건을 만족하는지를 판별한다(단계 S414). 이 판별에서 학습 중단 조건을 만족하지 않으면 프로세스 진행을 S401로 회귀시키고, 학습 중단 조건을 만족하면 프로세스를 종료한다.

[0064] 도 5는 동영상 데이터가 없는 경우 정지 영상으로부터 동영상 데이터 및 정답 영역 마스크를 생성하는 프로세스 진행 과정을 나타낸 흐름도이다.

[0065] 도 5를 참조하면, 이는 본 발명의 학습 방법에 따라 학습된 의료 인공지능 모델을 적용하여, 동영상 데이터가 없는 경우 정지 영상으로부터 동영상 데이터 및 정답 영역 마스크를 생성하는 프로세스 진행 과정으로서, 먼저 시간 t=0인지를 판별하여(단계 S501), t=0이면 t=0일 때의 이미지(I_t)와 랜덤으로 선정된 마스크(\hat{M}_t)를 입력하고(단계 S502, S503), t=0이 아니면 시간 t에서의 추론 이미지 (\hat{I}_t) 및 추론 마스크(\hat{M}_t)를 입력한다(단계 S504, S505).

[0066] 그러면, 의료 인공지능 모델(210)은 입력받은 시간 t=0일 때의 이미지(I_t)와 랜덤으로 선정된 마스크(\hat{M}_t) 또는 시간 t에서의 추론 이미지(\hat{I}_t) 및 추론 마스크(\hat{M}_t)를 바탕으로 다음 프레임 생성 모델에 의해 시간 t+1 번째 프레임 이미지 (\hat{I}_{t+1}) 및 t+1 번째 마스크(\hat{M}_{t+1})를 추론하여 생성한다(단계 S506~S508).

[0067] 이상과 같은 일련의 과정은 수 회 또는 수십 회 반복적으로 수행된다. 따라서 싸이클이 반복 수행될 때마다 매번 t=t+n(미리 설정된 시점 혹은 횟수)에 도달했는지를 판별하여(단계 S509), t=t+n(미리 설정된 시점 혹은 횟수)에 도달하지 않았으면 프로세스 진행을 단계 S501로 회귀시키고, t=t+n(미리 설정된 시점 혹은 횟수)에 도달했으면 프로세스 진행을 종료한다.

[0068] 도 6은 동영상 데이터는 있으나, 병변 태깅이 없는 경우 동영상 데이터를 태깅을 위해 사용하는 프로세스 진행 과정을 나타낸 흐름도이다.

[0069] 도 6을 참조하면, 이는 본 발명의 학습 방법에 따라 학습된 의료 인공지능 모델을 적용하여, 동영상 데이터는 있으나, 병변 태깅이 없는 경우 동영상 데이터를 태깅을 위해 사용하는 프로세스 진행 과정으로서, 먼저 의료 인공지능 모델(210)이 사전에 준비된 학습용 동영상 데이터 세트 중 임의의 시간 t 번째 프레임 이미지(I_t)(205)와 t 번째 마스크(\hat{M}_t)(206)를 입력받는다(단계 S601, S603). 이때, 시간 t=0인지를 판별하여(단계 S602), t=0이 아니면 t 번째 마스크, 즉 추론된 t 번째 마스크(\hat{M}_t) (206b)(도 2 참조)를 입력받고(단계 S603), t=0이면 랜덤으로 t 번째 마스크(\hat{M}_t)(206a)(도 2 참조)를 입력받는다(단계 S604).

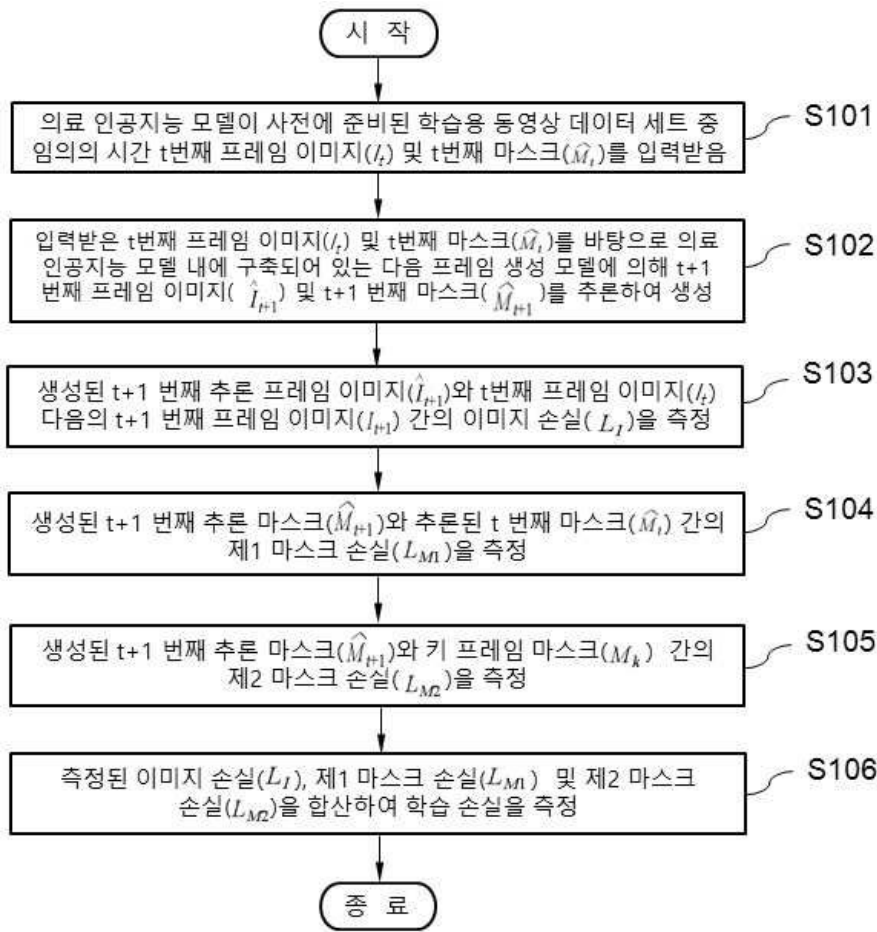
- [0070] 이렇게 하여 시간 t 번째 프레임 이미지(I_t)(205) 및 t 번째 마스크(\hat{M}_t)(206)가 입력되면, 의료 인공지능 모델(210)은 입력받은 시간 t 번째 프레임 이미지(I_t) 및 t 번째 마스크(\hat{M}_t)를 바탕으로 다음 프레임 생성 모델(211)에 의해 시간 $t+1$ 번째 프레임 이미지(\hat{I}_{t+1})(212) 및 $t+1$ 번째 마스크(\hat{M}_{t+1})(213)를 추론하여 생성한다(단계 S605~S607). 여기서, 이때 생성된 $t+1$ 번째 프레임 이미지(\hat{I}_{t+1})(212)는 동영상 데이터가 이미 있으므로 사용하지 않는다.
- [0071] 이상과 같은 일련의 과정은 마찬가지로 수 회 또는 수십 회 반복적으로 수행된다. 따라서 사이클이 반복 수행될 때마다 매번 $t=t+n$ (미리 설정된 시점 혹은 횟수)에 도달했는지를 판별하여(단계 S608), $t=t+n$ (미리 설정된 시점 혹은 횟수)에 도달하지 않았으면 프로세스 진행을 단계 S501로 회귀시키고, $t=t+n$ (미리 설정된 시점 혹은 횟수)에 도달했으면 프로세스 진행을 종료한다.
- [0072] 이상에 의해 생성된 모든 프레임 이미지와 병변 태깅 마스크는 의료 현장에서 사용 가능하며, 특정 스텝 크기(step size)를 기준으로 샘플링하여 활용할 수 있다. 그리고 특정 조건을 기준으로 선택하여 활용할 수 있다. 예를 들면, 프레임 간의 유사도를 측정하여 유사하지 않은 프레임 세트로만 구성하여 사용할 수도 있고, 유사한 프레임 세트로만 구성하여 사용할 수도 있다.
- [0073] 이상의 설명과 같이, 본 발명에 따른 정지영상 데이터로부터 동영상 데이터를 생성하기 위한 의료 인공지능 모델의 학습 방법은, 인공지능 기반 동영상 생성 모델을 도입하여 동영상 데이터 수집 및 동영상 데이터 타겟 영역 태깅 작업이 제한적인 상황에서 동영상 데이터를 활용할 수 있도록 함으로써, 동영상 데이터 활용에 요구되는 시간과 인력을 최소화 할 수 있고, 수집된 정지영상 DB를 통해 동영상 학습 모델의 효과를 얻을 수 있으며, 실시간 진단 환경에 효과적으로 의료 인공지능 모델을 적용할 수 있는 장점이 있다.
- [0074] 이상, 바람직한 실시예를 통하여 본 발명에 관하여 상세히 설명하였으나, 본 발명은 이에 한정되는 것은 아니며, 본 발명의 기술적 사상을 벗어나지 않는 범위 내에서 다양하게 변경, 응용될 수 있음은 당해 기술분야의 통상의 기술자에게 자명하다. 따라서, 본 발명의 진정한 보호 범위는 다음의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술적 사상은 본 발명의 권리 범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

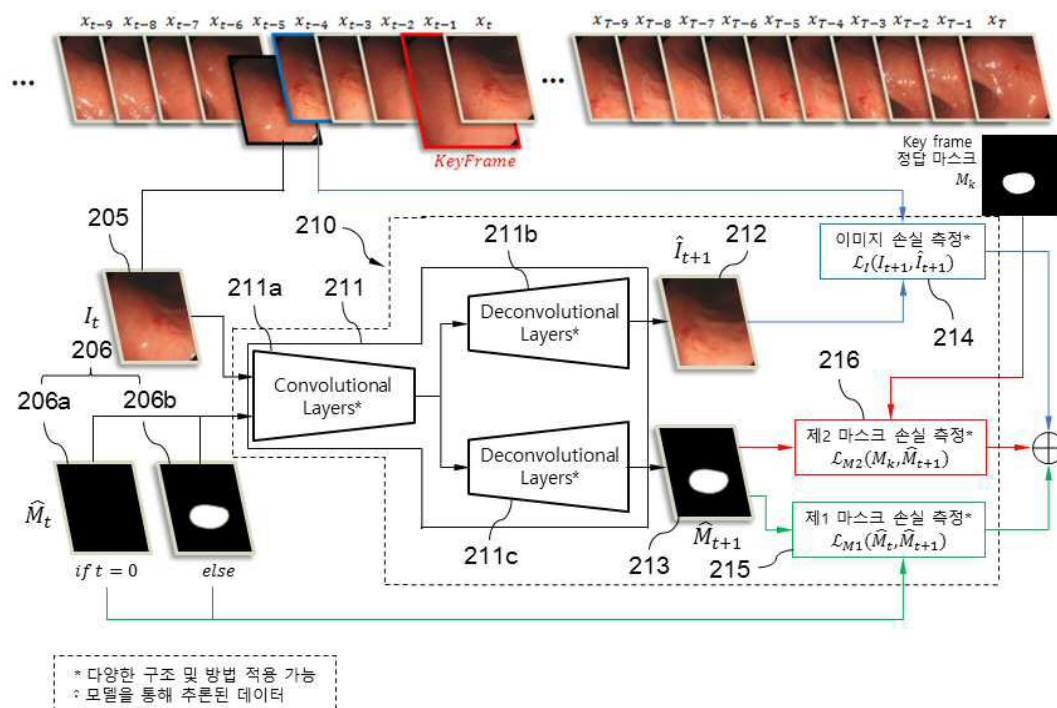
- [0075]
- | | |
|--|--------------------------------------|
| 205: t 번째 프레임 이미지(I_t) | 206: t 번째 마스크(\hat{M}_t) |
| 210: 의료 인공지능 모델 | 211: 다음 프레임 생성 모델 |
| 212: $t+1$ 번째 프레임 이미지(\hat{I}_{t+1}) | 213: $t+1$ 번째 마스크(\hat{M}_{t+1}) |

도면

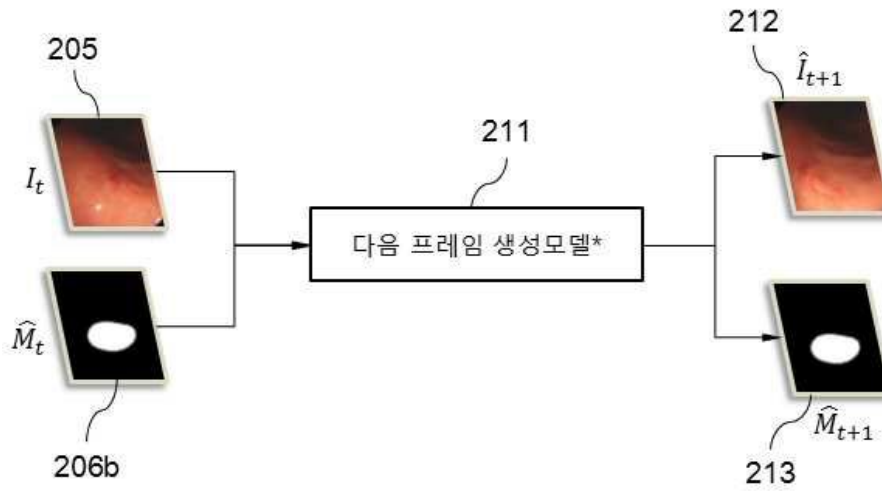
도면1



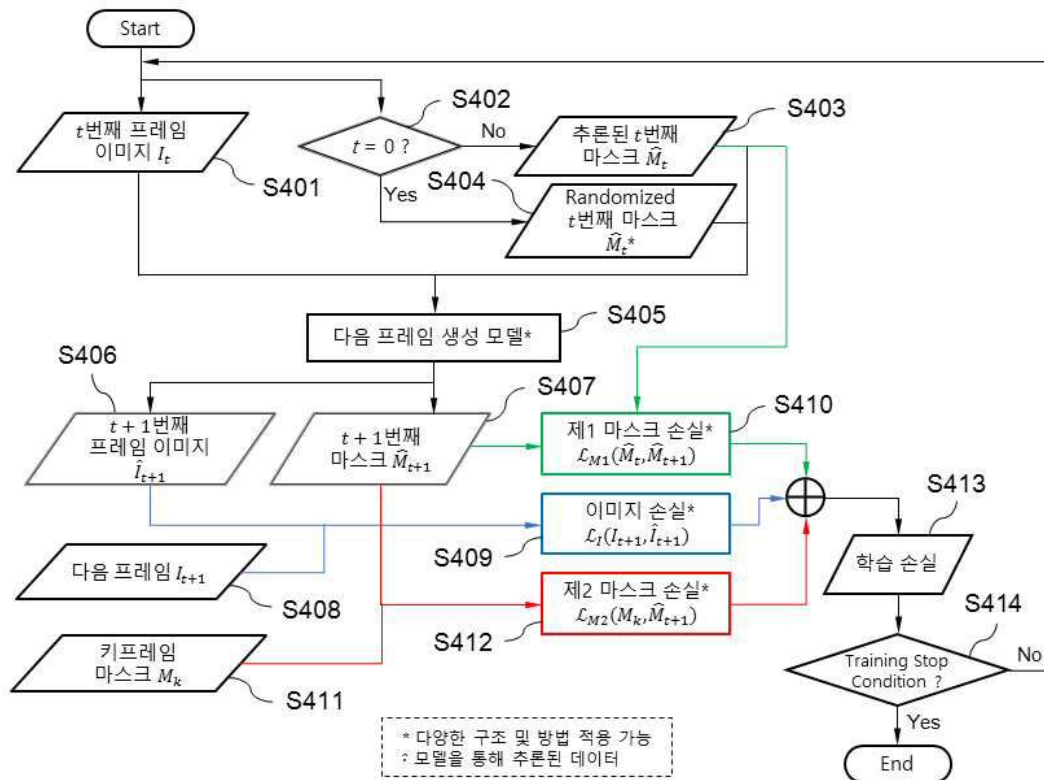
도면2



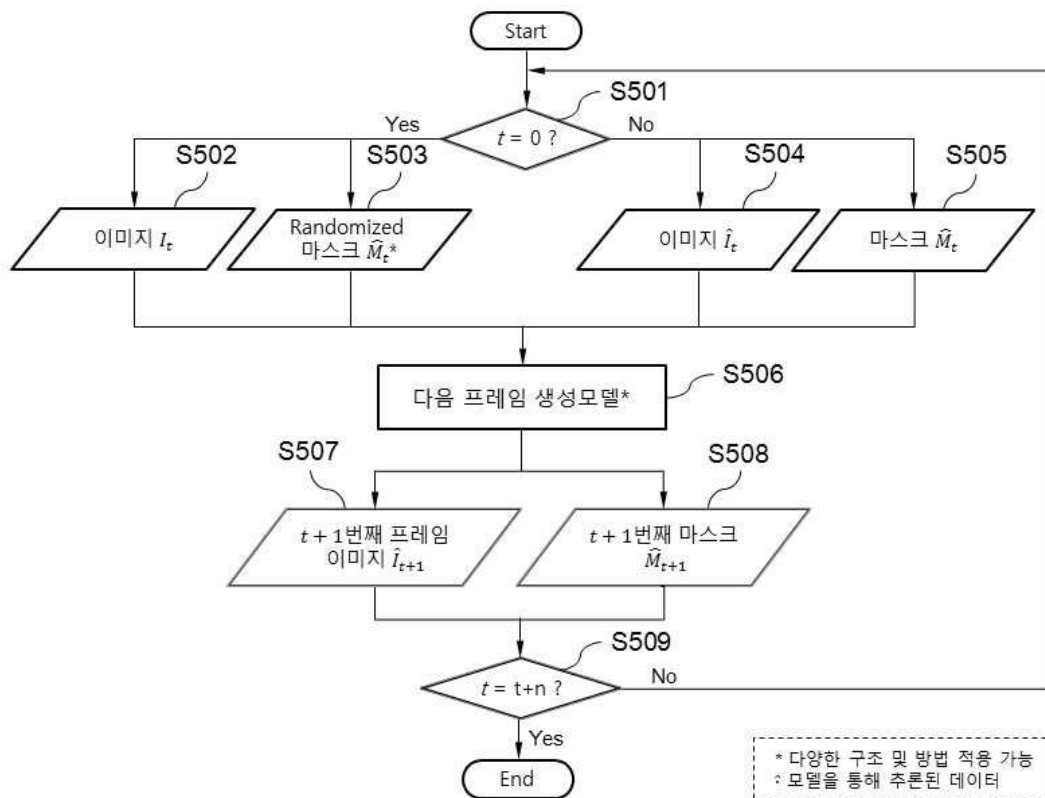
도면3



도면4



도면5



도면6

