



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2024년08월05일
(11) 등록번호 10-2691077
(24) 등록일자 2024년07월29일

- (51) 국제특허분류(Int. Cl.)
G06N 3/0985 (2023.01) G06N 3/045 (2023.01)
G06N 3/0455 (2023.01) G06N 3/0464 (2023.01)
- (52) CPC특허분류
G06N 3/0985 (2023.01)
G06N 3/045 (2023.01)
- (21) 출원번호 10-2023-0046745
(22) 출원일자 2023년04월10일
심사청구일자 2023년04월10일
- (30) 우선권주장
1020230021696 2023년02월17일 대한민국(KR)
- (56) 선행기술조사문헌
CN114565760 A*
KR1020220067931 A*
Ze Liu 등, "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows", arXiv:2103.14030v2, 2021.08.17. 1부.*
*는 심사관에 의하여 인용된 문헌
- (73) 특허권자
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
- (72) 발명자
김승룡
서울특별시 성북구 오패산로 46, 103동 1804호(하월곡동, 월곡두산위브아파트)
홍성환
서울특별시 강북구 한천로109길 57-56, B01 (번동)
조석주
서울특별시 성북구 성북로16가길 7-6 (성북동)
- (74) 대리인
송인호, 윤형근, 최관락

전체 청구항 수 : 총 5 항

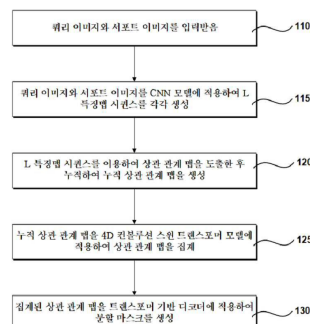
심사관 : 노지명

(54) 발명의 명칭 퓨-샷 세그멘테이션 방법 및 그 장치

(57) 요약

퓨-샷 세그멘테이션 방법 및 그 장치가 개시된다. 퓨-샷 세그멘테이션 방법은, (a) 쿼리 이미지와 서포트 이미지를 컨볼루션 모델에 적용하여 l 레벨 쿼리 특징맵과 서포트 특징맵을 각각 추출하고, 상기 l 레벨 쿼리 특징맵과 서포트 특징맵을 이용하여 마스크된 서포트 특징맵을 생성하는 단계; (b) 상기 l 레벨 쿼리 특징맵과 마스크된 서포트 특징맵을 이용하여 상관 관계 맵을 도출한 후 누적하여 누적 상관 관계 맵을 생성하는 단계; (c) 상기 누적 상관 관계 맵을 4D 윈도우 시퀀스를 기초로 상관 관계 맵을 재구성하고, 상기 재구성된 상관 관계 맵을 스윈 트랜스포머 모델을 통해 쿼리 공간 차원과 서포트 공간 차원이 겹치지 않는 하위 상관 관계 맵으로 균등 분할하고 셀프-어텐션 연산을 수행한 결과로 상기 각 하위 상관 관계 맵을 집계하는 단계; 및 (d) 상기 집계된 상관 관계 맵을 트랜스포머 기반 디코더에 적용하여 예측된 분할 마스크를 생성하는 단계를 포함한다.

대표도



(52) CPC특허분류

G06N 3/0455 (2023.01)

G06N 3/0464 (2023.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711193663
과제번호	2020-0-01819-004
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성(R&D)
연구과제명	ICT명품인재양성(고려대학교)
기 여 율	1/1
과제수행기관명	고려대학교산학협력단
연구기간	2023.01.01 ~ 2023.12.31

공지예외적용 : 있음

명세서

청구범위

청구항 1

- (a) 쿼리 이미지와 서포트 이미지를 컨볼루션 모델에 적용하여 l 레벨 쿼리 특징맵과 l 레벨 서포트 특징맵을 각각 추출하고, 상기 l 레벨 서포트 특징맵과 서포트 마스크를 이용하여 마스크된 서포트 특징맵을 생성하는 단계-상기 서포트 마스크는 세그멘테이션 정보 인코딩 및 백그라운드 필터링을 위한 마스크임;
- (b) 상기 l 레벨 쿼리 특징맵과 상기 마스크된 서포트 특징맵의 내적(inner product)을 이용하여 상기 쿼리 이미지와 상기 서포트 이미지에 대한 상관 관계 맵을 도출한 후 누적하여 누적 상관 관계 맵을 생성하는 단계;
- (c) 상기 누적 상관 관계 맵을 볼륨메트릭 임베딩 모듈 및 볼륨메트릭 트랜스포머 모듈을 통해 확장하여 4D 윈도우 시퀀스로 상관 관계 맵을 재구성하고, 상기 재구성된 상관 관계 맵을 스윈 트랜스포머 모델을 통해 쿼리 공간 차원과 서포트 공간 차원이 겹치지 않는 하위 상관 관계 맵으로 균등 분할하고 셀프-어텐션 연산을 수행한 결과로 상기 각 하위 상관 관계 맵을 집계하는 단계; 및
- (d) 상기 집계된 상관 관계 맵을 트랜스포머 기반 디코더에 적용하여 예측된 분할 마스크를 생성하는 단계를 포함하되,
- 상기 (d) 단계는,
- 상기 집계된 상관 관계 맵을 평균 풀링한 결과에 상기 쿼리 특징맵을 선형 투영한 후 피라미드 구조를 가지는 상기 트랜스포머 기반 디코더에 적용하여 상기 예측된 분할 마스크를 생성하되,
- 상기 4D 윈도우 시퀀스는 쿼리 공간 차원(높이와 너비)과 서포트 공간 차원(높이와 너비)을 포함하는 것을 특징으로 하는 퓨-샷 세그멘테이션 방법.

청구항 2

제1 항에 있어서,

상기 스윈 트랜스포머 모델은 피라미드 구조 기반 인코더를 포함하되,

상기 피라미드 구조 기반 인코더는 이전 레벨의 집계된 상관 관계 맵을 업샘플링한 결과와 현재 레벨 상관 관계 맵을 이용하여 현재 레벨의 상관 관계 맵을 집계하는 것을 특징으로 하는 퓨-샷 세그멘테이션 방법.

청구항 3

삭제

청구항 4

제1 항에 있어서,

상기 (c) 단계에서,

상기 상관 관계 맵은 하기 수학적식을 이용하여 도출되는 것을 특징으로 하는 퓨-샷 세그멘테이션 방법.

$$C^l(i, j) = \text{ReLU} \left(\frac{F_q^l(i) \cdot \hat{F}_s^l(j)}{\|F_q^l(i)\| \|\hat{F}_s^l(j)\|} \right)$$

여기서, i 와 j 는 각각 특징맵의 2차원 공간 위치를 나타내고, F_q^l 는 쿼리 특징맵을 나타내며, $\hat{F}_s^l$ 는 마스크된 서포트 특징맵을 나타내며, $\text{ReLU}()$ 는 활성화 함수를 나타냄.

청구항 5

제1 항에 있어서,

상기 스윈 트랜스포머 모델은,

상기 재구성된 상관 관계 맵을 균등 분할 한 후 각 분할된 하위 상관 관계 맵내에서 셀프-어텐션 계산을 수행하여 셀프-어텐션 맵을 도출한 후 이전 분할된 윈도우 크기의 절반 $(\lfloor \frac{n}{2} \rfloor, \lfloor \frac{n}{2} \rfloor, \lfloor \frac{n}{2} \rfloor, \lfloor \frac{n}{2} \rfloor)$ 변위만큼 이동하여 셀프-어텐션 연산을 수행한 후 롤백하되, n 은 이전 윈도우 크기인 것을 특징으로 하는 퓨-샷 세그멘테이션 방법.

청구항 6

적어도 하나의 명령어를 저장하는 메모리;

상기 메모리에 저장된 명령어를 실행하는 프로세서를 포함하되,

상기 프로세서에 의해 실행된 명령어는,

(a) 쿼리 이미지와 서포트 이미지를 컨볼루션 모델에 적용하여 l 레벨 쿼리 특징맵과 l 레벨 서포트 특징맵을 각각 추출하고, 상기 l 레벨 서포트 특징맵과 서포트 마스크를 이용하여 마스크된 서포트 특징맵을 생성하는 단계-상기 서포트 마스크는 세그멘테이션 정보 인코딩 및 백그라운드 필터링을 위한 마스크임;

(b) 상기 l 레벨 쿼리 특징맵과 상기 마스크된 서포트 특징맵의 내적(inner product)을 이용하여 상기 쿼리 이미지와 상기 서포트 이미지에 대한 상관 관계 맵을 도출한 후 누적하여 누적 상관 관계 맵을 생성하는 단계;

(c) 상기 누적 상관 관계 맵을 볼륨메트릭 임베딩 모듈 및 볼륨메트릭 트랜스포머 모듈을 이용하여 확장하여 4D 윈도우 시퀀스로 상관 관계 맵을 재구성하고, 상기 재구성된 상관 관계 맵을 스윈 트랜스포머 모델을 통해 쿼리 공간 차원과 서포트 공간 차원이 겹치지 않는 하위 상관 관계 맵으로 균등 분할하고 셀프-어텐션 연산을 수행한 결과로 상기 각 하위 상관 관계 맵을 집계하는 단계; 및

(d) 상기 집계된 상관 관계 맵을 트랜스포머 기반 디코더에 적용하여 예측된 분할 마스크를 생성하는 단계를 수행하되,

상기 (d) 단계는,

상기 집계된 상관 관계 맵을 평균 풀링한 결과에 상기 쿼리 특징맵을 선형 투영한 후 피라미드 구조를 가지는 상기 트랜스포머 기반 디코더에 적용하여 상기 예측된 분할 마스크를 생성하되,

상기 4D 윈도우 시퀀스는 쿼리 공간 차원(높이와 너비)과 서포트 공간 차원(높이와 너비)을 포함하는 것을 특징으로 하는 컴퓨팅 장치.

발명의 설명

기술 분야

[0001] 본 발명은 퓨-샷 세그멘테이션 방법 및 그 장치에 관한 것이다.

배경 기술

[0003] 서포트 샘플에서 프로토타입을 추출하는 수 많은 방법들이 제안되었으며, 이는 쿼리 샘플에서 물체들을 세분화하는데 사용된다. 특징 처리에서 단순히 하이 레벨의 특징만 사용하면 성능이 저하되는 문제점을 해결하기 위해 쿼리 샘플에서 대상을 식별하는데 도움이되는 프라이어 맵(prior map)을 계산하기 위한 하이 레벨의 특징을 사용하는 방법이 제안되었다.

[0004] 많은 변형들이 특징맵을 집계(aggregate)하기 위한 추가 정보로 프라이어 맵을 활용하는 이 아이디어를 확장하였나, 프로토타입 또는 프라이어 맵을 기반으로 하는 종래 기술들은 서포트와 쿼리 특징맵의 공간 구조 사이의 쌍별 관계를 무시하는 한계가 있다.

[0005] 시맨틱 세그멘테이션은 해당 클래스로 이미지의 각 픽셀에 레이블을 지정하는 것을 목표로 하는 컴퓨터 비전 어플리케이션이다. 심층 신경망과 실측 세그멘테이션 주석을 포함하는 대규모 데이터 세트에 의해 상당한 진전이 이루어졌으나, 픽셀 단위 세그멘테이션 맵의 수동 라벨링은 상당한 노동을 필요로하므로 새로운 클래스를 추가하기 매우 어렵다. 이로 인해, 레이블이 지정된 데이터에 대한 의존도를 줄이기 위해 퓨-샷 세그멘테이션에 관심이 집중되고 있다.

[0006] 퓨-샷 세그멘테이션의 핵심은 소수의 서포트 샘플을 효과적으로 사용하는 것이다. 종래의 경우 샘플에서 프로토타입 모델을 추출하고 쿼리와 특징 비교에 프로토타입 모델을 사용하였으나, 이러한 접근 방식은 서포트 및 쿼리 특징맵간의 픽셀 수준 쌍별 관계를 무시하여 최적이지 않은 결과를 초래하는 문제점이 있다.

발명의 내용

해결하려는 과제

[0008] 본 발명은 퓨-샷 세그멘테이션 방법 및 그 장치를 제공하기 위한 것이다.

[0009] 또한, 본 발명은 서포트 샘플과 쿼리 샘플간의 상관 관계를 집계하여 작은 수의 영상에서도 쿼리 샘플에 대한 분할 성능이 압도적으로 높은 퓨-샷 세그멘테이션 방법 및 그 장치를 제공하기 위한 것이다.

과제의 해결 수단

[0011] 본 발명의 일 측면에 따르면 퓨-샷 세그멘테이션 방법이 제공된다.

[0012] 본 발명의 일 실시예에 따르면, (a) 쿼리 이미지와 서포트 이미지를 컨볼루션 모델에 적용하여 l 레벨 쿼리 특징맵과 서포트 특징맵을 각각 추출하고, 상기 l 레벨 서포트 특징맵과 서포트 마스크를 이용하여 마스크된 서포트 특징맵을 생성하는 단계; (b) 상기 l 레벨 쿼리 특징맵과 마스크된 서포트 특징맵을 이용하여 상관 관계 맵을 도출한 후 누적하여 누적 상관 관계 맵을 생성하는 단계; (c) 상기 누적 상관 관계 맵을 4D 윈도우 시퀀스로 상관계 맵을 재구성하고, 상기 재구성된 상관 관계 맵을 스윈 트랜스포머 모델을 통해 쿼리 공간 차원과 서포트 공간 차원이 겹치지 않는 하위 상관 관계 맵으로 균등 분할하고 셀프-어텐션 연산을 수행한 결과로 상기 각 하위 상관 관계 맵을 집계하는 단계; 및 (d) 상기 집계된 상관 관계 맵을 트랜스포머 기반 디코더에 적용하여 예측된 분할 마스크를 생성하는 단계를 포함하는 퓨-샷 세그멘테이션 방법이 제공될 수 있다.

[0013] 상기 스윈 트랜스포머 모델은 피라미드 구조 기반 인코더를 포함하되, 상기 피라미드 구조 기반 인코더는 이전 레벨의 집계된 상관 관계 맵을 업샘플링한 결과와 현재 레벨 상관 관계 맵을 이용하여 현재 레벨의 상관 관계

맵을 집계할 수 있다.

[0014] 상기 (d) 단계는, 상기 집계된 상관 관계 맵을 평균 풀링한 결과에 상기 쿼리 특징맵을 선형 투영한 후 상기 피라미드 구조 기반 디코더에 적용하여 상기 예측된 분할 마스크를 생성할 수 있다.

[0015] 상기 (c) 단계에서, 상기 상관 관계 맵은 하기 수식식을 이용하여 도출되되,

$$C^l(i, j) = \text{ReLU} \left(\frac{F_q^l(i) \cdot \hat{F}_s^l(j)}{\|F_q^l(i)\| \|\hat{F}_s^l(j)\|} \right)$$

[0016]

[0017] 여기서, i와 j는 각각 특징맵의 2차원 공간 위치를 나타내고, F_q^l 는 쿼리 특징맵을 나타내며, $\hat{F}_s^l$ 는 마스크된 서포트 특징맵을 나타내며, ReLU()는 활성화 함수를 나타낸다.

[0018] 상기 스윈 트랜스포머 모델은, 상기 재구성된 상관 관계 맵을 균등 분할 한 후 각 분할된 하위 상관 관계 맵내에서 셀프-어텐션 계산을 수행하여 셀프-어텐션 맵을 도출한 후 이전 분할된 윈도우 크기의 절반 $(\lfloor \frac{n}{2} \rfloor, \lfloor \frac{n}{2} \rfloor, \lfloor \frac{n}{2} \rfloor, \lfloor \frac{n}{2} \rfloor)$ 변위만큼 이동하여 셀프-어텐션 연산을 수행한 후 롤백하되, n은 이전 윈도우 크기이다.

[0020] 본 발명의 다른 측면에 따르면, 퓨-샷 세그멘테이션을 위한 장치가 제공된다.

[0021] 본 발명의 일 실시예에 따르면, 적어도 하나의 명령어를 저장하는 메모리; 상기 메모리에 저장된 명령어를 실행하는 프로세서를 포함하되, 상기 프로세서에 의해 실행된 명령어는,

[0022] (a) 쿼리 이미지와 서포트 이미지를 컨볼루션 모델에 적용하여 l 레벨 쿼리 특징맵과 서포트 특징맵을 각각 추출하고, 상기 l 레벨 서포트 특징맵과 서포트 마스크를 이용하여 마스크된 서포트 특징맵을 생성하는 단계; (b) 상기 l 레벨 쿼리 특징맵과 마스크된 서포트 특징맵을 이용하여 상관 관계 맵을 도출한 후 누적하여 누적 상관 관계 맵을 생성하는 단계; (c) 상기 누적 상관 관계 맵을 4D 윈도우 시퀀스를 기초로 상관 관계 맵을 재구성하고, 상기 재구성된 상관 관계 맵을 스윈 트랜스포머 모델을 통해 쿼리 공간 차원과 서포트 공간 차원이 겹치지 않는 하위 상관 관계 맵으로 균등 분할하고 셀프-어텐션 연산을 수행한 결과로 상기 각 하위 상관 관계 맵을 집계하는 단계; 및 (d) 상기 집계된 상관 관계 맵을 트랜스포머 기반 디코더에 적용하여 예측된 분할 마스크를 생성하는 단계를 수행하는 것을 특징으로 하는 컴퓨팅 장치가 제공될 수 있다.

발명의 효과

[0024] 본 발명은 퓨-샷 세그멘테이션 방법 및 그 장치를 제공함으로써, 서포트 샘플과 쿼리 샘플간의 상관 관계를 집계하여 작은 수의 영상에서도 쿼리 샘플에 대한 분할 성능이 압도적으로 높은 이점이 있다.

도면의 간단한 설명

[0026] 도 1은 본 발명의 일 실시예에 따른 퓨-샷 세그멘테이션 방법을 나타낸 순서도.

도 2는 본 발명의 일 실시예에 따른 전체 아키텍처를 도시한 도면.

도 3은 본 발명의 일 실시예에 따른 4D 윈도우 시프트를 설명하기 위해 도시한 도면.

도 4는 본 발명의 일 실시예에 따른 4D 컨볼루션 스윈 트랜스포머 구조를 도시한 도면.

도 5는 본 발명의 일 실시예에 따른 컴퓨팅 장치의 내부 구성을 개략적으로 도시한 블록도.

발명을 실시하기 위한 구체적인 내용

[0027] 본 명세서에서 사용되는 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "구성된다" 또는 "포함한다" 등의 용어는 명세서상에 기재된 여러 구성 요소들, 또는 여러 단계들을 반드시 모두 포함하는 것으로 해석되지 않아야 하며, 그 중 일부 구성 요소들 또는 일부 단계들은 포함되지 않을 수도 있고, 또는 추가적인 구성 요소 또는 단계들을 더 포함할 수 있는 것으로 해석되어야 한다. 또한, 명세서에 기재된 "...부", "모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합으로 구현될 수 있다.

[0028] 이하, 첨부된 도면들을 참조하여 본 발명의 실시예를 상세히 설명한다.

[0030] 도 1은 본 발명의 일 실시예에 따른 퓨-샷 세그멘테이션 방법을 나타낸 순서도이고 도 2는 본 발명의 일 실시예에 따른 전체 아키텍처를 도시한 도면이고, 도 3은 본 발명의 일 실시예에 따른 4D 윈도우 시프트를 설명하기 위해 도시한 도면이고, 도 4는 본 발명의 일 실시예에 따른 4D 컨볼루션 스윈 트랜스포머 구조를 도시한 도면이다.

[0031] 단계 110에서 컴퓨팅 장치(100)은 쿼리 이미지 x_q 와 서포트 이미지 x_s 를 입력받는다.

[0032] 본 발명의 일 실시예에 따르면, 데이터 셋은 트레이닝 세트 D_{train} 와 테스트 세트 D_{test} 로 구성된다. 트레이닝 세트 D_{train} 와 테스트 세트 D_{test} 의 객체 클래스는 상호 겹치지 않는 것을 가정하기로 한다.

[0033] k-샷 설정에서 여러 에피소드는 두 세트에서 각각 서포트 세트 $S = \{(x_s^k, m_s^k)\}_{k=1}^K$ 가 구성될 수 있다. 여기서, (x_s^k, m_s^k) 는 k번째 서포트 이미지와 이에 대응하는 마스크 페어를 나타낸다. 또한, (x_q, m_q) 는 쿼리 샘플을 나타내고, x_q 는 쿼리 이미지를 나타내고, m_q 는 페어링 마스크를 나타낸다.

[0034] 학습 과정 동안 트레이닝 세트에서 샘플링된 에피소드를 취하고, m_q 를 예측하기 위해 S와 x_q 로의 매핑을 학습할 수 있다. 또한, 학습이 완료되면, 컴퓨팅 장치(100)는 테스트 세트에서 임의로 샘플링된 S와 x_q 가 주어지면 $\hat{m}_q$ 를 예측할 수 있다. 이에 대해서는 하기의 설명에 의해 보다 명확하게 이해될 것이다.

[0035] 단계 115에서 컴퓨팅 장치(100)는 쿼리 이미지 x_q 와 서포트 이미지 x_s 를 CNN 모델(210)에 적용하여 L 특징맵 시퀀스 $\{(\bar{F}_q^l, F_s^l)\}_{l=1}^L$ 를 각각 생성한다(도 2 참조).

[0036] 여기서, F_q^l 와 F_s^l 는 각각 1번째 레벨의 쿼리 특징맵과 서포트 특징맵을 나타낸다. 서포트 마스크 m_s 는 세그멘테이션 정보를 인코딩하고 백그라운드 정보를 필터링하는데 사용될 수 있다.

[0037] 이어, 컴퓨팅 장치(100)는 1번째 레벨의 서포트 특징맵 F_s^l 과 서포트 마스크 m_s 를 이용하여 마스크된 서포트 특징맵을 도출할 수 있다. 마스크된 서포트 특징맵은 $\hat{F}_s^l = F_s^l \odot \psi^l(m_s)$ 와 같이 획득될

수 있다. 여기서, $\odot$ 는 하다마드(Hadamard) 곱을 나타내고, $\psi^l(\cdot)$ 는 텐서의 크기를 조정한 다음 1번째 레이어의 채널 차원을 따라 확장하는 함수를 나타낸다.

[0038] 단계 120에서 컴퓨팅 장치(100)는 L 특징맵 시퀀스를 이용하여 상관 관계 맵을 도출한다(도 2의 215 참조).

[0039] 예를 들어, 컴퓨팅 장치(100)는 쿼리 특징맵 F_q^l 과 서포트 특징맵 F_s^l 이 주어지면, 1-2 정규화된 특징맵들 사이의 내적(inner product)를 이용하여 상관 관계 맵을 도출할 수 있다. 이를 수학식으로 나타내면, 수학식 1과 같다.

수학식 1

$$C^l(i, j) = \text{ReLU} \left(\frac{F_q^l(i) \cdot \hat{F}_s^l(j)}{\|F_q^l(i)\| \|\hat{F}_s^l(j)\|} \right)$$

[0040]

여기서, i와 j는 각각 특징맵의 2차원 공간 위치를 나타낸다. 동일한 공간 크기의 모든 중간 특징맵으로부터 상관 관계 맵을 수집한 후 누적한 누적 상관 관계 맵은

$C^p \in \mathbb{R}^{h_q \times w_q \times h_s \times w_s \times |\mathcal{L}^p|}$ 과 같이 획득될 수 있다. 여기서, h_q , w_q 는 각각 쿼리 특징맵의 높이와 너비를 나타낸다. 또한, h_s 와 w_s 는 각각 서

포트 특징맵의 높이와 너비를 나타낸다. $\mathcal{L}^p$ 는 동일한 공간 크기의 상관 관계 맵을 포함하는 피라미드 계층

p에서 CNN 계층 인덱스 $\{1, \dots, L\}$ 의 서브 셋을 나타낸다.

[0042] 단계 125에서 컴퓨팅 장치(100)는 누적 상관 관계 맵을 4D 컨볼루션 스윈 트랜스포머 모델에 적용하여 상관 관계 맵을 집계한다(도2의 220 참조).

[0043] 이에 대해 설명하기로 한다.

[0044] 트랜스포머 모델을 통해 상관 관계 맵을 처리하기 위해서는 토큰 감소 수단이 필수적이거나, 이미지, 텍스트 또는 특징이 다른 상관 관계 맵에 대한 토큰을 정의하는 것은 어려운 일이다.

[0045] 상관 관계 맵을 처리하기 위해 4차원 상관 관계 맵을 2차원 맵으로 재구성하고, 2차원에서 셀프-어텐션(self-attention)을 수행하는 CAT(Cost Aggregation Transformers) 기법이 있다. 그러나, CAT 기법은 성능을 제한할 수 있는 서포트 및 쿼리 모드에 대한 상관 관계 맵의 공간 구조를 무시하는 단점이 있다.

이를 해결하기 위해 모든 공간 엔트리 $h_q \times w_q \times h_s \times w_s$ 를 토큰으로 취급하고, $\mathcal{L}^p$ 를 토큰의 특징 타원으로 취급하는 방법이 있다. 그러나 이는 더 큰 상관 관계 맵으로 증가하는 상당한 계산 부담을 초래한다.

[0047] 따라서, 본 발명의 일 실시예에서는 상관 관계 맵의 공간 구조를 유지하기 위해 효율적인 셀프-어텐션 계산을 제공하고, 장거리 셀프-어텐션을 제공하면서도 상관 관계 맵의 부드러운 속성을 유지할 수 있는 스윈 트랜스포머 구조(도 4 참조)를 이용한다.

- [0048] 종래의 패치 임베딩 절차를 따라 상관 관계 맵을 임베딩하는 경우, 많은 수의 토큰으로 인해 스윈 트랜스포머
- $$\mathcal{M}^P = \text{VEM}(C^P)$$
- 모델에서 처리할 수 없다. 따라서, 패치 임베딩 모듈을 $\mathcal{M}^P$ 와 같이 더 높은 차원의 입력을 처리하는 볼륨메트릭 임베딩 모듈로 확장한다. 패치 임베딩 절차와 유사한 절차에 따라 16 x 16 x 16과 같은 대형 컨볼루션 커널을 사용하여 4D 윈도우 시퀀스로 상관 관계 맵을 재구성할 수 있다.
- [0049] 이와 같이 상관 관계 맵을 재구성한 후 $\mathcal{M}^P$ 의 쿼리와 서포트 공간 차원을 겹치지 않는 하위 상관 맵 $\bar{\mathcal{M}}'^{P} \in \mathbb{R}^{n \times n \times n \times n \times D}$ 으로 균등 분할하여 셀프-어텐션 연산을 확장한다. 분할된 각 하위 상관 관계 맵 내에서 셀프-어텐션 맵을 도출한다.
- [0050] 그런 다음 이전에 분할된 윈도우에서 $(\lfloor \frac{n}{2} \rfloor, \lfloor \frac{n}{2} \rfloor, \lfloor \frac{n}{2} \rfloor, \lfloor \frac{n}{2} \rfloor)$ 의 변위만큼 윈도우를 이동한 다음 새로 생성된 윈도우에서 셀프-어텐션을 계산한다(도 2 참조).
- [0051] 그런 다음 스윈 트랜스포머 모델은 상관 관계 맵을 원래 형식으로 롤백한다. 셀프-어텐션을 계산할 때 상대 위치 편향을 사용하고, 확장된 매개변수 편향 매트릭스로부터 값을 가져온다. 스윈 트랜스포머 블록의 다른 구성 요소(LN 및 MLP 레이어)는 변경하지 않는다.
- [0052] 본 발명의 일 실시예에서는 이러한 확장을 볼륨메트릭 트랜스포머 모듈(VTM: Volumetric Transformer Module)이라 칭하기로 한다. 해당 프로세스는 수학식 2와 같이 나타낼 수 있다.
- 수학식 2**
- $$\mathcal{A}^P = \text{VTM}(\mathcal{M}^P)$$
- [0053]
- [0054] 스윈 트랜스포머를 사용한 상관 관계 맵의 집계는 CNN 모델 사용 문제와 표준 트랜스포머 사용의 높은 계산 부담을 해결할 수 있으나, 변환 증가 부족을 피할 수 없다. 이는 주로 셀프-어텐션 계산 이전에 중첩되지 않은 작업을 활용하기 때문에 발생한다.
- [0055] 스윈 트랜스포머 모델은 상대 포지셔닝 편향을 사용하여 문제를 어느 정도 완화할 수 있으나, 근사치가 충분하지 않다.
- [0056] 볼륨메트릭 임베딩 모듈은 겹치지 않은 큰 컨볼루션 커널을 사용하면 제한된 유도 편향만 제공된다. 중첩되는 작업에 비해 중첩되지 않은 작업에서 상대적으로 낮은 변환 증가가 달성된다. 이로 인해, 일반화 능력과 성능이 상대적으로 낮아지며, 밀도가 높은 예측 작업의 경우 중첩되지 않는 커널로 인해 윈도우 경계를 무시하면 불연속성으로 인해 전반적인 성능이 저하되는 문제점이 있다.
- [0057] 따라서, 본 발명의 일 실시예에서는 볼륨메트릭 컨볼루션 모듈을 통해 상관 관계 맵을 재구성한다. 즉, 볼륨메트릭 컨볼루션 모듈은 3D 공간 맥스 풀링, 중첩 4D 컨볼루션, ReLU 및 GN(Group Normalization)을 적용하여 지원 및 쿼리 공간 차원을 순차적으로 줄일 수 있다. 예를 들어, 각 위치에서 멀티-레벨 유사성 벡터를 D로 표시되는 임의의 고정 차원으로 투영할 수 있다. $m \times m \times m \times m$ 과 같은 4D 윈도우로서 수용 필드를 고려하면, $C^P \in \mathbb{R}^{h'_q{}^P \times w'_q{}^P \times h'_s{}^P \times w'_s{}^P \times D}$ 를 획득한다. 여기서, $h'_s{}^P, w'_s{}^P, h'_q{}^P$ 및 $w'_q{}^P$ 는 프로세스된 크기이다. 서포트 및 쿼리 공간 차원에 대해 다른 크기의 m 을 선택할 수 있다. 볼륨메트릭 컨볼루션 모듈에 대한 전체 개요는 도 2에 도시된 바와 같다.

[0058] 해당 과정은 수학식 3와 같이 나타낼 수 있다.

수학식 3

$$\mathcal{M}^p = \text{VCM}(\mathcal{C}^p)$$

[0059]

[0060] 본 발명의 일 실시예에 따른 4D 컨볼루션 스원 트랜스포머 모델을 통해 윈도우 경계에서 더 나은 처리 뿐만 아니라 추가 유도 편향의 이점을 얻을 수 있다.

[0061] 도 2에 도시된 바와 같이, 4D 컨볼루션 스원 트랜스포머 모델은 시맨틱 매칭 및 퓨-샷 세그멘테이션에서 피라미드 구조가 사용된다.

[0062] 미세한 레벨(finer-level) 상관 관계 맵 $\mathcal{A}^p$ 의 집계는 이전 레벨 상관 관계 맵 $\mathcal{A}^{p+1}$ 의 집계된 상관 관계 맵에 의해 유도된다. 보다 상세하게, 집계된 상관 관계 맵 $\mathcal{A}^{p+1}$ 은 업-샘플링되고, 가이드로서 역할을 위해 다음 레벨 상관 관계 맵 $\mathcal{A}^p$ 에 추가된다. 가장 높은 수준 집계된 맵이 도출될때까지 반복되고, 이후 디코더 모듈(225)로 전달된다.

[0063] 피라미드 구조를 사용하는 트랜스포머 인코더는 수학식 4와 같이 정의될 수 있다.

수학식 4

$$\mathcal{A}^p = \text{VTM}(\text{VCM}(\mathcal{C}^p) + \text{up}(\mathcal{A}^{p+1})),$$

[0064]

[0065] 여기서, $\text{up}(\cdot)$ 는 바이리니어(bilinear) 업샘플링 함수를 나타낸다.

[0066] 단계 130에서 컴퓨팅 장치(100)는 집계된 상관 관계 맵을 트랜스포머 기반 디코더에 적용하여 분할 마스크를 생성한다(도 2의 225 참조).

[0067] 이에 대해 보다 상세히 설명하기로 한다.

[0068] 본 발명의 일 실시예에 따른 트랜스포머 기반 디코더는 성능 향상을 위해 쿼리 특징 맵에서 획득된 외관 정보(appearance) 임베딩을 통해 추가 집계를 수행할 수 있다. 쿼리 특징맵의 외관 정보 임베딩은 두가지 방식으로 도움이 된다.

[0069] 우선, CVF(Cost Volume Filtering)와 같은 스테레오 매칭에서 입증된 바와 같이 외관 친화도 정보는 매칭 점수에서 노이즈를 필터링하는데 효과적인 가이드 역할을 수행한다.

[0070] 또한, 외관 정보 임베딩에 의해 제공되는 고해상도 공간 구조를 활용하여 업샘플링 품질을 개선할 수 있어 미세한 세부 사항이 보존되어 매우 정확한 예측 마스크 $\hat{m}^q$ 를 생성할 수도 있다.

[0071] 트랜스포머 기반 디코더는 집계된 상관 관계 맵 $\mathcal{A}^p$ 의 서포트 이미지 차원에 대한 평균을 취하여 쿼리 특징 맵에서 임베딩된 외관 정보와 연결한 다음 쌍선형 보간을 가지는 트랜스포머로 집계한다.

[0072] 이를 수학식 5와 같이 정의될 수 있다.

수학식 5

$$\hat{m}_q = \text{ATD}([\mathcal{A}'^p, \mathcal{P}(F_q)]).$$

[0073]

[0074] 여기서, $\mathcal{A}'^p \in \mathbb{R}^{h_q' \times w_q' \times D}$ 는 서포트 이미지의 공간 차원에 걸쳐 $\mathcal{A}^p$ 에 대한 평균 풀

$$\mathcal{P}(F_q) \in \mathbb{R}^{h_q' \times w_q' \times c}$$

링에 의해 추출되며, $\mathcal{A}^p$ 는 쿼리 특징맵의 선형 투영을 나타내며, $[\cdot, \cdot]$ 는 이들의 연결 함수를 나타낸다. 바이리니어 업샘플링 직후 출력을 순차적으로 미세 조정하여 세부 정보를 다시 캡처하고 외관 정보를 통합할 수 있다.

[0075] K 쌍의 서포트 이미지와 마스크 $\{(x_s^i, m_s^i)\}_{i=1}^K$ 와 쿼리 이미지 x_q 가 주어지면, 4D 컨볼루션 스원

트랜스포머 모델은 K개의 다른 쿼리 마스크 $\hat{m}_q^k$ 를 얻기 위해 K번 정방향 통과한다. 각 공간 위치에서 모든 K

개의 예측을 합산하고 K로 나눈 합계가 임계값 τ 를 초과하면 위치가 전경으로 예측되고 그렇지 않으면 배경으로 예측될 수 있다.

[0077] 도 5는 본 발명의 일 실시예에 따른 컴퓨팅 장치의 내부 구성을 개략적으로 도시한 블록도이다.

[0078] 도 5를 참조하면, 본 발명의 일 실시예에 따른 컴퓨팅 장치(100)는 메모리(510) 및 프로세서(520)를 포함하여 구성된다.

[0079] 메모리(510)는 본 발명의 일 실시예에 따른 퓨-샷 세그멘테이션 방법을 수행하기 위한 명령어를 저장한다.

[0080] 프로세서(520)는 본 발명의 일 실시예에 따른 컴퓨팅 장치(100)의 내부 구성 요소들(예를 들어, 메모리(510))를 제어하기 위한 수단이다.

[0081] 또한, 프로세서(520)는 메모리(510)에 저장된 명령어를 실행할 수 있다. 프로세서(520)에 의해 실행된 명령어는 도 1에서 설명된 각각의 단계를 수행할 수 있다. 이는 이미 기술한 바와 동일하므로 중복되는 설명은 생략하기로 한다.

[0083] 본 발명의 실시 예에 따른 장치 및 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 컴퓨터 판독 가능 매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 분야 통상의 기술자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기매체(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.

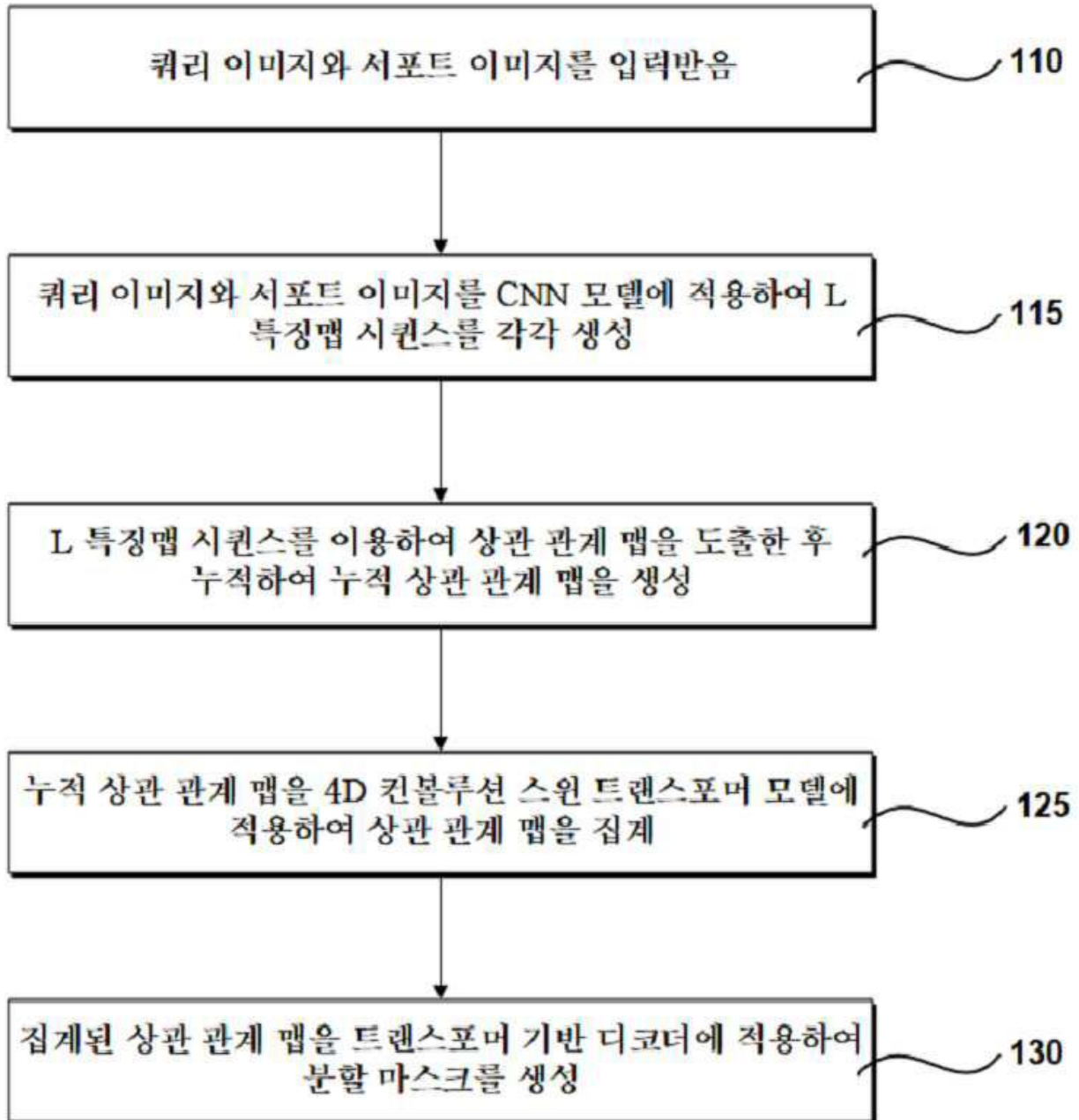
[0084] 상술한 하드웨어 장치는 본 발명의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0085] 이제까지 본 발명에 대하여 그 실시 예들을 중심으로 살펴보았다. 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로 개시된 실시 예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야

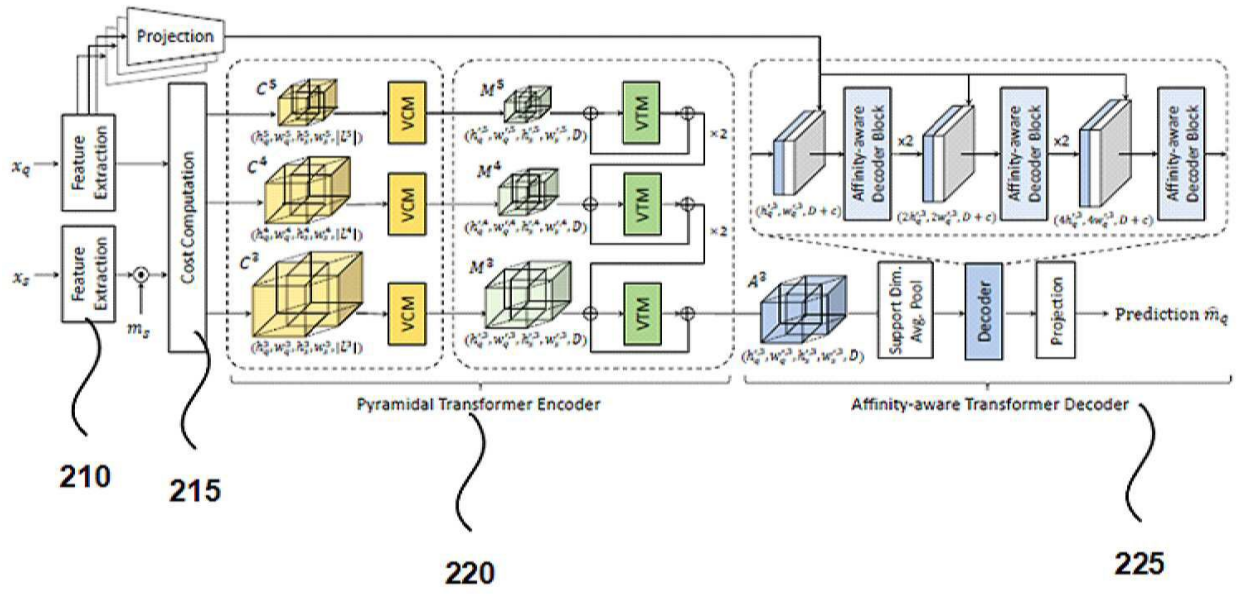
한다. 본 발명의 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

도면

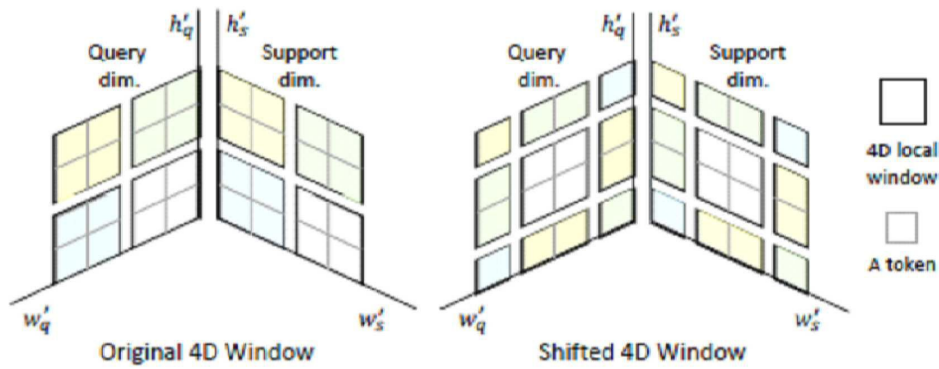
도면1



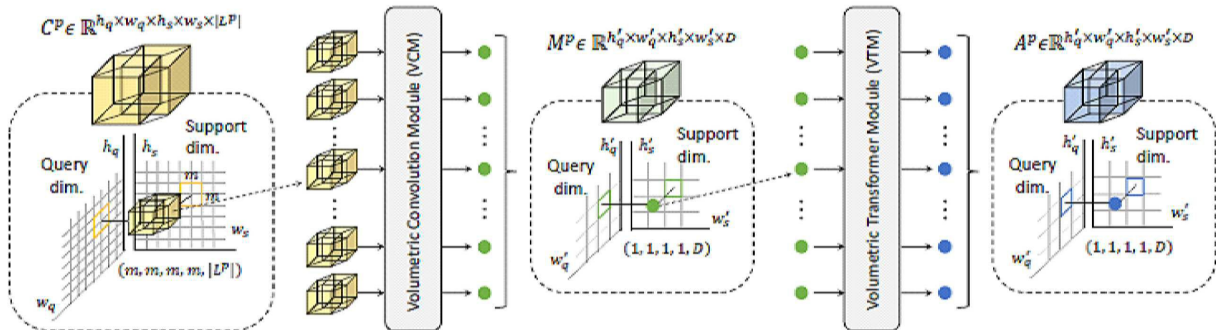
도면2



도면3



도면4



도면5

