



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0146429
(43) 공개일자 2024년10월08일

(51) 국제특허분류(Int. Cl.)
G06V 20/52 (2022.01) G06V 10/80 (2022.01)
(52) CPC특허분류
G06V 20/53 (2022.01)
G06V 10/803 (2023.08)
(21) 출원번호 10-2023-0041384
(22) 출원일자 2023년03월29일
심사청구일자 2023년03월29일

(71) 출원인
인하대학교 산학협력단
인천광역시 미추홀구 인하로 100(용현동, 인하대학교)
(72) 발명자
홍성은
서울특별시 성동구 독서당로62길 43, 9동 307호(응봉동, 대림아파트)
장요우지아
인천광역시 미추홀구 인하로134번길 6 (용현동, 해피하우스)
최소윤
경상북도 포항시 북구 법원로 91-12, 601동 201호(장성동, 아름들이)
(74) 대리인
양성보

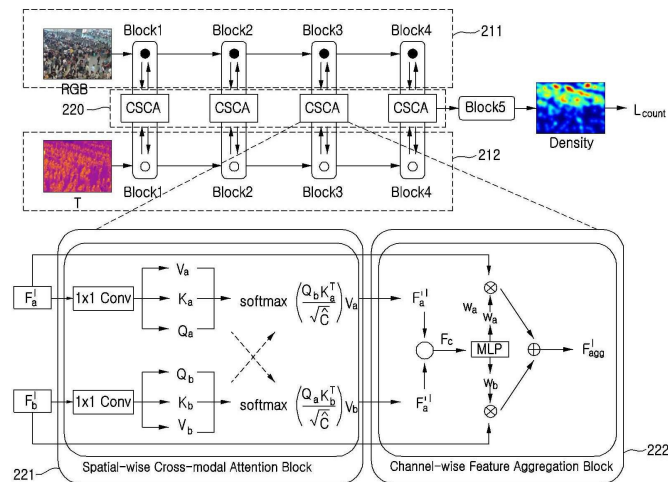
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 교차모달 클라우드 카운팅을 위한 시공간 주의 블록 방법 및 시스템

(57) 요약

교차모달 클라우드 카운팅을 위한 시공간 주의 블록의 동작 방법 및 시스템이 제시된다. 본 발명에서 제안하는 교차모달 클라우드 카운팅을 위한 시공간 주의 블록의 동작 방법은 특징 추출부를 통해 RGB 이미지와 열 이미지 및 깊이 이미지를 포함하는 다중 모달 데이터 쌍으로부터 백본 네트워크를 활용하여 모달리티별 특징을 추출하는 단계, CSCA(Cross-modal Spatio-Channel Attention) 블록의 SCA(Spatial-wise Cross-modal Attention) 블록을 통해 RGB 이미지와 열/깊이 이미지 간의 글로벌 기능 상관 관계를 모델링하는 단계 및 CSCA 블록의 CFA(Channel-wise Feature Aggregation) 블록을 통해 채널 상관관계에 기초하여 교차 모달 정보를 적응적으로 추출하고 융합하는 단계를 포함한다.

대표도



(52) CPC특허분류

G06V 10/806 (2023.08)
G06T 2207/10028 (2013.01)
G06T 2207/10048 (2013.01)
G06T 2207/30242 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711179347
과제번호	00155915
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신인재양성(교육운영수익)
연구과제명	[Ezbaro][정부+민간] 인공지능융합혁신인재양성(1차년도)
기 여 율	1/2
과제수행기관명	인하대학교 산학협력단
연구기간	2022.07.01 ~ 2023.02.28

이 발명을 지원한 국가연구개발사업

과제고유번호	1711186835
과제번호	2022R1A4A1033549
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	기초연구실육성사업(BRL)
연구과제명	[Ezbaro] 인공지능 융합형 메타휴먼 플랫폼 연구실
기 여 율	1/2
과제수행기관명	인하대학교
연구기간	2022.06.01 ~ 2023.02.28

명세서

청구범위

청구항 1

특징 추출부를 통해 RGB 이미지와 열 이미지 및 깊이 이미지를 포함하는 다중 모달 데이터 쌍으로부터 백본 네트워크를 활용하여 모달리티별 특징을 추출하는 단계;

CSCA(Cross-modal Spatio-Channel Attention) 블록의 SCA(Spatial-wise Cross-modal Attention) 블록을 통해 RGB 이미지와 열/깊이 이미지 간의 글로벌 기능 상관 관계를 모델링하는 단계; 및

CSCA 블록의 CFA(Channel-wise Feature Aggregation) 블록을 통해 채널 상관관계에 기초하여 교차 모달 정보를 적응적으로 추출하고 융합하는 단계

를 포함하는 시공간 주의 블록 동작 방법.

청구항 2

제1항에 있어서,

상기 CSCA 블록의 SCA 블록을 통해 RGB 이미지와 열/깊이 이미지 간의 글로벌 기능 상관 관계를 모델링하는 단계는,

비 로컬 기반 모델에서 사용되는 쿼리, 키, 값의 삼중항을 기반으로 하는 주의 메커니즘을 활용하고, RGB 이미지와 열 이미지 및 깊이 이미지 간의 글로벌 기능 상관 관계를 공간 차원으로 모델링하여 정렬 오류를 완화함으로써, 특징 재구성 작업을 통해 계산 오버헤드를 줄이는

시공간 주의 블록 동작 방법.

청구항 3

제2항에 있어서,

상기 CSCA 블록의 SCA 블록을 통해 RGB 이미지와 열/깊이 이미지 간의 글로벌 기능 상관 관계를 모델링하는 단계는,

다중 모달 데이터 쌍에 대한 서로 다른 모달리티에 대하여 유사성을 나타내는 주의 가중치를 곱하여 출력을 계산하고, 서로 다른 모달리티에 대하여 정보를 집계하고 정렬하며, 특징 맵을 공간 차원의 패치로 랜덤으로 분할하고, 채널 차원을 복구하는 재구성 작업을 수행함으로써 계산 오버헤드를 줄이는

시공간 주의 블록 동작 방법.

청구항 4

제1항에 있어서,

상기 CSCA 블록의 CFA 블록을 통해 채널 상관관계에 기초하여 교차 모달 정보를 적응적으로 추출하고 융합하는 단계는,

공간 상관만을 고려한 비 로컬 기반 모델의 특성을 보완하기 위해 채널 차원에 따라 보완 특징을 동적으로 집계함으로써, 채널 수준에서 공간 상관 특징을 적응적으로 조정하는

시공간 주의 블록 동작 방법.

청구항 5

RGB 이미지와 열 이미지 및 깊이 이미지를 포함하는 다중 모달 데이터 쌍으로부터 백본 네트워크를 활용하여 모달리티별 특징을 추출하는 특징 추출부;

RGB 이미지와 열/깊이 이미지 간의 글로벌 기능 상관 관계를 모델링하는 SCA(Spatial-wise Cross-modal

Attention) 블록 및 채널 상관관계에 기초하여 교차 모달 정보를 적응적으로 추출하고 융합하는 CFA(Channel-wise Feature Aggregation) 블록을 포함하는 CSCA(Cross-modal Spatio-Channel Attention) 블록을 포함하는 시공간 주의 블록 시스템.

청구항 6

제5항에 있어서,

상기 CSCA 블록의 SCA 블록은,

비 로컬 기반 모델에서 사용되는 쿼리, 키, 값의 삼중항을 기반으로 하는 주의 메커니즘을 활용하고, RGB 이미지와 열 이미지 및 깊이 이미지 간의 글로벌 기능 상관 관계를 공간 차원으로 모델링하여 정렬 오류를 완화함으로써, 특징 재구성 작업을 통해 계산 오버헤드를 줄이는

시공간 주의 블록 시스템.

청구항 7

제6항에 있어서,

상기 CSCA 블록의 SCA 블록은,

다중 모달 데이터 쌍에 대한 서로 다른 모달리티에 대하여 유사성을 나타내는 주의 가중치를 곱하여 출력을 계산하고, 서로 다른 모달리티에 대하여 정보를 집계하고 정렬하며, 특징 맵을 공간 차원의 패치로 랜덤으로 분할하고, 채널 차원을 복구하는 재구성 작업을 수행함으로써 계산 오버헤드를 줄이는

시공간 주의 블록 시스템.

청구항 8

제5항에 있어서,

상기 CSCA 블록의 CFA 블록은,

공간 상관만을 고려한 비 로컬 기반 모델의 특성을 보완하기 위해 채널 차원에 따라 보완 특징을 동적으로 집계함으로써, 채널 수준에서 공간 상관 특징을 적응적으로 조정하는

시공간 주의 블록 시스템.

발명의 설명

기술 분야

[0001] 본 발명은 교차모달 클라우드 카운팅을 위한 시공간 주의 블록의 동작 방법 및 시스템에 관한 것이다.

배경 기술

[0002] 군중 카운팅은 사전 지식 없이 제한되지 않은 장면에서 보행자 수를 자동으로 추정하는 것을 목표로 한다. 트래픽 모니터링, 사회적 거리 모니터링 및 기타 보안 관련 시나리오를 포함한 실제 애플리케이션에 의해 주도되는 군중 카운팅은 다양한 분야에서 필수적인 역할을 한다.

[0003] 일반적으로 군중 카운팅을 위한 다양한 접근법은 크게 탐지 기반 방법과 회귀 기반 방법의 두 가지 범주로 나눌 수 있다. 군중 카운팅에 대한 초기 연구는 주로 탐지 기반 접근법을 사용하는데, 드론 물체 탐지 방법에서 볼 수 있는 것처럼 이미지에서 사람의 머리를 탐지한다. 실제로 대부분의 탐지 기반 방법은 이미지에서 군중이 주어진 탐지기에 의해 탐지될 수 있는 개인으로 구성되어 있다고 가정한다. 군중이 겹치고 심각한 폐색이 있는 고밀도 시나리오에서 작은 머리에는 여전히 몇 가지 중요한 한계가 있다. 위의 문제점을 해결하기 위해 일부 기술에서는 이미지에서 카운트 또는 군중 밀도 맵에 대한 매핑 문제로 군중 카운팅을 보는 회귀 기반 방법을 도입한다. 이러한 방식으로 탐지 문제를 피할 수 있다. 최근 기술은 RGB 군중 카운팅에서 밀도 맵 추정의 상당한 발전을 보여주었고, 그것은 복잡한 군중 장면에서 계산 작업의 주류 방법이 되었다.

[0004] 상당한 발전에도 불구하고, 군중 카운팅은 폐색, 높은 혼란(high clutter), 불균일 조명, 불균일 분포 등과 같

은 다양한 상황에 대한 오랜 도전 과제로 남아 있다. 방법의 한계 또는 불충분한 데이터 소스로 인해 대부분의 이전 연구는 RGB 군중 카운팅에 초점을 맞추고 제한되지 않은 시나리오에서 고품질 밀도 맵을 얻는 데 실패했다. RGB 이미지는 어두운 환경에서 거의 보이지 않지만 밝은 조명 조건에서 강력한 증거를 제공할 수 있다. 최근 일부 종래기술에서는 열 특성이 열악한 조명 조건에서 다양한 대상을 더 잘 구별하는 데 도움이 된다고 지적한다.

[0006] 도 1은 종래기술에 따른 RGB-T 및 RGB-D 쌍을 시각화한 도면이다.

[0007] 도 1(a)와 같이 열 이미지(RGB-T)를 사용하여 군중 카운팅 작업의 성능을 향상시킬 수 있다. 그러나, 도 1(b)로부터, 추가적인 가열 물체가 열 이미지를 심각하게 방해한다는 것을 알 수 있다.

[0008] 또한, 광학 단서의 다양한 상황 정보를 활용하는 것 외에도, 깊이 특징은 최근 향상된 군중 밀도 맵을 생성하기 위해 RGB 특징의 보충 정보로 사용되고 있다. 도 1(c)에서, 군중의 움직임이 집합 행동과 머리 크기의 큰 변화로 이어진다는 것을 관찰한다. 따라서, RGB 데이터만을 사용하여 픽셀 단위의 밀도 맵을 생성하는 것은 불충분하다. 깊이 데이터(RGB-D)는 다양한 조명에 강하고 머리의 위치와 크기에 대한 추가 정보를 제공하는 3D 형상을 2D 시각 정보에 제공하여 RGB 정보를 자연스럽게 보완한다. 그러나 도 1(d)와 같이 깊이 정보는 뒤죽박죽인 배경 조건에서 노이즈의 영향을 크게 받는다.

선행기술문헌

비특허문헌

[0009] (비특허문헌 0001) [1] Ma, Z., Wei, X., Hong, X., Gong, Y.: Bayesian loss for crowd count estimation with point supervision. In: Proc. of Int'l Conf. on Computer Vision (ICCV). pp. 6142-6151 (2019)

(비특허문헌 0002) [2] Li, Y., Zhang, X., Chen, D.: Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes. In: Proc. of Computer Vision and Pattern Recognition (CVPR). pp. 1091-1100 (2018)

발명의 내용

해결하려는 과제

[0010] 본 발명이 이루고자 하는 기술적 과제는 SCA(Spatial-wise Cross-modal Attention) 블록 및 CFA(Channel-wise Feature Aggregation) 블록을 통해 공간적 및 채널적 주의 모두에 기초하여 교차 모달 특징을 적응적으로 집계하는 교차모달 크라우드 카운팅을 위한 시공간 주의 블록 방법 및 시스템을 제공하는데 있다.

과제의 해결 수단

[0011] 일 측면에 있어서, 본 발명에서 제안하는 교차모달 크라우드 카운팅을 위한 시공간 주의 블록의 동작 방법은 특징 추출부를 통해 RGB 이미지와 열 이미지 및 깊이 이미지를 포함하는 다중 모달 데이터 쌍으로부터 백본 네트워크를 활용하여 모달리티별 특징을 추출하는 단계, CSCA(Cross-modal Spatio-Channel Attention) 블록의 SCA(Spatial-wise Cross-modal Attention) 블록을 통해 RGB 이미지와 열/깊이 이미지 간의 글로벌 기능 상관 관계를 모델링하는 단계 및 CSCA 블록의 CFA(Channel-wise Feature Aggregation) 블록을 통해 채널 상관관계에 기초하여 교차 모달 정보를 적응적으로 추출하고 융합하는 단계를 포함한다.

[0012] 상기 CSCA 블록의 SCA 블록을 통해 RGB 이미지와 열/깊이 이미지 간의 글로벌 기능 상관 관계를 모델링하는 단계는 비 로컬 기반 모델에서 사용되는 쿼리, 키, 값의 삼중항을 기반으로 하는 주의 메커니즘을 활용하고, RGB 이미지와 열 이미지 및 깊이 이미지 간의 글로벌 기능 상관 관계를 공간 차원으로 모델링하여 정렬 오류를 완화함으로써, 특징 재구성 작업을 통해 계산 오버헤드를 줄인다.

[0013] 상기 CSCA 블록의 SCA 블록을 통해 RGB 이미지와 열/깊이 이미지 간의 글로벌 기능 상관 관계를 모델링하는 단계는 다중 모달 데이터 쌍에 대한 서로 다른 모달리티에 대하여 유사성을 나타내는 주의 가중치를 곱하여 출력을 계산하고, 서로 다른 모달리티에 대하여 정보를 집계하고 정렬하며, 특징 맵을 공간 차원의 패치로 랜덤으로

분할하고, 채널 차원을 복구하는 재구성 작업을 수행함으로써 계산 오버헤드를 줄인다.

[0014] 상기 CSCA 블록의 CFA 블록을 통해 채널 상관관계에 기초하여 교차 모달 정보를 적응적으로 추출하고 융합하는 단계는 공간 상관관을 고려한 비 로컬 기반 모델의 특성을 보완하기 위해 채널 차원에 따라 보완 특징을 동적으로 집계함으로써, 채널 수준에서 공간 상관 특징을 적응적으로 조정한다.

[0015] 또 다른 일 측면에 있어서, 본 발명에서 제안하는 교차모달 클라우드 카운팅을 위한 시공간 주의 블록 시스템은 RGB 이미지와 열 이미지 및 깊이 이미지를 포함하는 다중 모달 데이터 쌍으로부터 백본 네트워크를 활용하여 모달리티별 특징을 추출하는 특징 추출부, RGB 이미지와 열/깊이 이미지 간의 글로벌 기능 상관 관계를 모델링하는 SCA(Spatial-wise Cross-modal Attention) 블록 및 채널 상관관계에 기초하여 교차 모달 정보를 적응적으로 추출하고 융합하는 CFA(Channel-wise Feature Aggregation) 블록를 포함하는 CSCA(Cross-modal Spatio-Channel Attention) 블록를 포함한다.

발명의 효과

[0016] 본 발명의 실시예들에 따르면 교차모달 클라우드 카운팅을 위한 시공간 주의 블록 방법 및 시스템은 SCA(Spatial-wise Cross-modal Attention) 블록 및 CFA(Channel-wise Feature Aggregation) 블록를 통해 공간적 및 채널적 주의 모두에 기초하여 교차 모달 특징을 적응적으로 집계할 수 있다. 또한, 제안하는 SCA 블록은 기존의 비 로컬 블록에 비해 낮은 복잡성으로 교차 모달 특징 변동에 강한 이점을 가질 수 있다.

도면의 간단한 설명

[0017] 도 1은 종래기술에 따른 RGB-T 및 RGB-D 쌍을 시각화한 도면이다.

도 2는 본 발명의 일 실시예에 따른 교차모달 클라우드 카운팅을 위한 시공간 주의 블록 시스템의 구성을 나타내는 도면이다.

도 3은 본 발명의 일 실시예에 따른 교차모달 클라우드 카운팅을 위한 시공간 주의 블록의 동작 방법을 설명하기 위한 흐름도이다.

도 4는 본 발명의 일 실시예에 따른 SCA 블록을 종래기술과 비교하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0018] 군중 카운팅 연구는 실제 응용 분야에서 상당한 발전을 이루었지만 교차 모드 설정에서 여전히 어려운 과제로 남아 있다. 대부분의 기존 방법은 열 및 깊이 이미지와 같은 다른 모달리티의 실현 가능성을 무시하고 RGB 이미지의 광학 특징에만 의존한다. 모델 아키텍처에 대한 서로 다른 모달리티와 설계 선택의 다양성 사이의 본질적으로 중요한 차이는 교차 모달리티 군중 카운팅을 더 어렵게 만든다.

[0019] 본 발명에서는 교차 모드 시공간 채널 주의(Cross-modal Spatio-Channel Attention; CSCA) 블록를 제안하며, 이는 모든 모달리티 특정 아키텍처에 쉽게 통합될 수 있다. CSCA 블록은 먼저 공간별 교차 모달 주의(Spatial-wise Cross-modal Attention; SCA) 블록를 통해 더 적은 오버헤드로 다중 모달 간의 글로벌 기능 상관관계를 공간적으로 포착한다. 공간적 주의를 기울인 교차 모달 특징은 이후 적응형 채널별 특징 집계(Channel-wise Feature Aggregation; CFA) 블록를 통해 개선된다. 제안하는 블록은 다양한 백본 네트워크에서 지속적으로 상당한 성능 향상을 보여 RGB-T(다시 말해, 열 이미지) 및 RGB-D(다시 말해, 깊이 이미지) 군중 카운팅에서 우수한 결과를 초래한다. 이하, 본 발명의 실시 예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0021] 군중 카운팅의 초기 작업은 탐지 기반 접근법에 초점을 맞춘다. 초기 방법은 주로 수작업 기능을 기반으로 하는 탐지기를 사용한다. 일반적인 전통적인 방법은 웨이블릿, HOG, 에지, 전신에서 추출된 기타 특징을 사용하여 보행자를 탐지하도록 분류기를 훈련시키며, 이는 매우 혼잡한 장면에서 심각하게 저하된다. 군중 밀도가 증가함에 따라, 사람들 사이의 폐색은 점점 더 심각해진다. 따라서 군중 카운팅 문제를 다루기 위해 부분적인 신체 탐지(예를 들어, 머리, 어깨 등)를 기반으로 하는 방법이 제안되었다. 전체적인 탐지와 비교하여, 이 방법은 효과가 약간 향상되었다. 최근의 탐지 기반 방법도 더 진보된 물체 탐지기의 도입과 함께 더욱 개선되었다. 예를 들어, 일부 종래기술은 작은 머리 탐지를 위한 탐지기의 견고성을 향상시켜 군중 카운팅의 성능을 향상시키기 위해 Faster RCNN 또는 RetinaNet과 같은 특정 아키텍처를 설계한다. 그러나, 이러한 종래기술은 여전히 매우 높은 밀도, 상당한 폐색 및 매우 밀집된 군중의 배경 혼란이 있는 이미지에 대해 불만족스러운 결과를 나타내고, 이

는 군중 카운팅 작업에서 상당히 일반적이다.

- [0022] 최근의 접근법 중 일부 종래기술은 탐지 문제를 완전히 우회하여 이 문제를 해결하려고 시도한다. 이러한 종래 기술은 이미지에서 카운트 또는 군중 밀도 맵에 이르는 매핑 문제로 직접 해결한다. 최근에는 수많은 모델이 기본 CNN 계층, 다중 열 기반 모델 또는 단일 열 기반 모델을 사용하여 밀도 추정 및 군중 카운팅을 위한 풍부한 정보를 추출한다. 다양한 해상도에서 다중 스케일 정보를 캡처하기 위해 MCNN은 세 가지 브랜치 아키텍처를 제안한다. 그리고, CrowdNet은 낮은 수준의 특징과 높은 수준의 의미 정보를 결합한다. 다중 열 방법으로 인한 정보 중복을 줄이기 위해 SANet는 인셉션(Inception) 아키텍처를 기반으로 구축된다. 그리고, CSRNet은 수용 영역을 확장하기 위해 확장된 컨볼루션 층을 채택한다. 회귀 기반 방법은 다중 스케일 연산 또는 확장 컨볼루션을 통해 낮은 수준의 특징과 높은 수준의 의미 정보를 캡처하여 군중 카운팅 작업에서 놀라운 성공을 보여준다. 그러나 대부분의 방법은 RGB 이미지에만 초점을 맞추고 제한되지 않은 시나리오에서 의미론적 객체를 정확하게 인식하지 못할 수 있다.
- [0023] 최근, 다중 모달 특징 융합이 다양한 분야에 걸쳐 광범위하게 연구되고 있다. 교차 모달 특징 융합은 여러 모달 간의 상보성을 활용하여 더 나은 특징 표현을 학습하는 것을 목표로 한다. 일부 교차 모달 융합 방법은 RGB 소스와 깊이 소스를 "초기 융합" 또는 "후기 융합" 방식으로 융합한다. 또한, 다양한 컴퓨터 비전 작업에서 멀티태스킹 학습에 기초하여, 일부 종래기술에서는 RGB 군중 카운팅 작업의 성능을 향상시키기 위한 보조 정보로 깊이 특징을 활용한다. 마찬가지로, 밀도 추정 절차 중의 투시 정보도 RGB 군중 카운팅의 성능을 향상시키기 위한 보조 데이터로 취급된다. 그러나 이러한 방법은 다중 모달 데이터를 다른 모달리티의 가중치를 예측하는 것과 융합하는 대신 수작업 특징에 의존한다. 또한 일부 종래기술에서는 탐지기의 성능에 의해 제한될 수 있는 군중 카운팅 작업의 탐지 절차에 깊이 정보를 결합한다. IADM은 RGB와 열 이미지를 공동으로 표현할 것을 제안한다. 그러나 RGB 데이터와 열 이미지 데이터의 다른 특성과 특징 공간 정렬 오류를 완전히 고려하지는 않는다. 본 발명에서는 종래기술과 달리 공간적 및 채널적 주의를 모두 기반으로 교차 모달 특징을 적응적으로 집계하는 새로운 교차 모달 시공간 채널 주의 블록을 제안한다. 또한, 제안하는 SCA 블록은 기존의 비 로컬 블록에 비해 낮은 복잡성으로 교차 모달 특징 변동에 강하다.
- [0025] 본 발명에서는 보다 강력한 교차 모달 군중 카운팅을 위해 다음과 같은 두 가지 주요 문제에 중점을 둔다.
- [0026] 먼저, 군중 카운팅 데이터는 서로 다른 양식에서 얻는데, 이는 분포와 속성이 크게 다르다는 것을 의미한다. 다중 모달 데이터 간에는 상당한 차이가 있기 때문에, 그 차이를 효과적으로 식별하고 여러 유형의 정보를 통합하는 방법은 교차 모달 군중 카운팅에서 여전히 문제이다. 또한, RGB 이외의 추가 모달리티는 고정되지 않기 때문에(열 또는 깊이) 모달 내 및 모달 간 상관관계를 적응적으로 포착하는 것이 매우 중요하다.
- [0027] 또한, 교차 모달 군중 카운팅의 경우 RGB 외에도 다양한 유형이 추가로 제공될 수 있다. 이 때, 처음부터 교차 모달 아키텍처를 구성하는 것은 보통 많은 시행착오를 수반한다. 본 발명에서는 잘 탐구된 모델 아키텍처가 전이 학습 또는 모델 튜닝을 통해 다양한 분야에서 우수한 결과를 보여주는 것을 확인하였다. 이러한 측면에서, 널리 사용되는 RGB 기반 단일 모달 아키텍처가 벨(bells)과 휘슬(whistles) 없이 다중 모달 아키텍처로 확장될 수 있다면, 모델 아키텍처 설계 선택의 수고를 덜 수 있다.
- [0028] 본 발명에서는 두 개의 주요 모듈로 구성된 교차 모달 시공간 채널 주의(Cross-modal Spatio-Channel Attention; CSCA) 블록이라는 새로운 플러그 앤 플레이 모듈을 제안한다. 첫째, 공간별 교차 모달 주의(Spatial-wise Cross-modal Attention; SCA) 모듈은 비 로컬 기반 모델에서 널리 사용되는 '쿼리', '키', '값'의 삼중항을 기반으로 하는 주의 메커니즘을 활용한다. 특히 기존의 비 로컬 블록은 계산 복잡성이 높고 단일 모달 내에서 자기 주의를 활용하도록 설계되었기 때문에 교차 모달 작업에 적합하지 않다. 이러한 문제를 해결하기 위해 SCA 모듈은 교차 모달 기능 간의 상관 관계를 이해하도록 의도적으로 설계되었으며 제안된 특징 재구성 체계를 통해 계산 오버헤드를 줄인다. 크기가 $C \times H \times W$ 인 입력 특징 맵이 주어지면, 일반적인 비 로컬 블록의 계산 복잡도는 $O(CH^2W^2)$ 인 반면, SCA 블록은 G_1 에서 $O(CH^2W^2/G_1)$ 로 복잡도를 줄일 수 있다. 여기서 G_1 은 백본 네트워크의 1번째 블록에 있는 공간 재구성 계수를 의미한다.
- [0029] 중요한 것은 다양한 비전 작업에 널리 사용되는 비 로컬 블록이 특징의 공간 상관 관계에 초점을 맞추지만 채널 수준에서 특징 재보정을 직접 수행하지 않는다는 점이다. 공간 상관만을 고려한 비 로컬 기반 블록의 특성을 보완하기 위해 채널별 특징 집계(Channel-wise Feature Aggregation; CFA) 모듈은 채널 수준에서 공간 상관 특징을 적응적으로 조정한다. 피라미드 풀링 특징 차이를 사용하여 다중 모달 특징 변화를 측정하거나 밀도 맵을 미

세 조정하는 종래기술과 달리, 본 발명의 실시예에 따른 SCA는 다중 모달 특징 사이의 글로벌 특징 상관 관계를 공간적으로 포착하여 다중 모달 데이터의 정렬 오류를 완화하는 것을 목표로 한다. 또한, 다른 모달 정보에 대한 보조 작업을 추가하는 대신, 본 발명의 실시예에 따른 방법은 CFA 블록에 의해 채널 차원을 따라 보완 특징을 동적으로 집계한다. 본 발명의 실시예에 따르면, 제안된 두 개의 주요 모듈이 두 가지 까다로운 RGB-T와 RGB-D에서 서로를 보완한다는 것을 보여준다.

- [0030] 본 발명의 실시예에 따르면, RGB 기반 군중 카운팅 모델과 통합될 때 성능을 크게 향상시킬 수 있는 새로운 플러그 앤 플레이 모듈을 제안한다.
- [0031] 본 발명의 실시예에 따른 특징 재구성을 통한 교차 모달 주의는 계산적으로 효율적이며 기존의 비 로컬 주의보다 높은 성능을 가지고 있다. 또한 채널 수준 재보정은 성능을 더욱 향상시킨다.
- [0032] 본 발명의 실시예에 따른 CSCA 블록은 다양한 백본 네트워크에서 지속적으로 상당한 성능 향상을 보여주며 RGB-T 및 RGB-D 군중 카운팅에서 우수한 성능을 보여준다.
- [0034] 도 2는 본 발명의 일 실시예에 따른 교차모달 크라우드 카운팅을 위한 시공간 주의 블록 시스템의 구성을 나타내는 도면이다.
- [0035] 본 발명의 실시예에 따르면, 기존의 기준 모델을 단일 모달 군중 카운팅에서 다중 모달 장면으로 확장하기 위한 통합 프레임워크를 제안한다.
- [0036] 도 2에 도시된 바와 같이, 교차 모달 군중 카운팅을 위한 프레임워크는 두 부분으로 구성된다. 모달별 브랜치에서 특징을 추출하는 특징 추출부(211, 212)와 교차 모달 시공간 채널 주의(Cross-modal Spatio-Channel Attention; CSCA) 블록(220)이다.
- [0037] 특징 추출부(211, 212)는 RGB-T(다시 말해, 열 이미지) 또는 RGB-D(다시 말해, 깊이 이미지) 쌍과 같은 다중 모달 데이터 쌍이 주어지면 먼저 백본 네트워크(예를 들어, BL[1], CSRNNet[2] 등)를 활용하여 RGB 이미지와 열 이미지 및 깊이 이미지를 포함하는 다중 모달 데이터 쌍에 대하여 모달리티별 특징을 추출한다. 다중 모달 상보성을 완전히 활용하기 위해 두 모달리티별 브랜치의 서로 다른 컨볼루션 블록 뒤에 CSCA(220) 블록을 내장하여 특징 맵을 추가로 업데이트하고 보완 융합 정보를 얻는다.
- [0038] 모달리티 간의 유익한 특징 상호 작용을 보장하기 위해 CSCA는 도 2와 같이 SCA(Spatial-wise Cross-modal Attention) 블록(221) 및 CFA(Channel-wise Feature Aggregation) 블록(222)의 두 가지 작업으로 구성된다. SCA 블록(221)은 개선된 교차 모달 주의 블록을 활용하여 RGB 이미지와 열/깊이 이미지 사이의 글로벌 기능 상관 관계를 모델링하며, 이는 일반 CNN 모듈보다 수용 필드가 크고 상호 보완적인 집계 작업을 용이하게 한다. 그리고, CFA 블록(222)은 채널 상관관계에 기초하여 교차 모달 정보를 적응적으로 추출하고 융합한다. 결과적으로, 프레임워크는 비 로컬 종속성을 포착하고 교차 모달 데이터의 보완 정보를 집계하여 단일 모달 군중 카운팅에서 다중 모달 군중 카운팅으로 확장할 수 있다.
- [0040] 도 3은 본 발명의 일 실시예에 따른 교차모달 크라우드 카운팅을 위한 시공간 주의 블록의 동작 방법을 설명하기 위한 흐름도이다.
- [0041] 제안하는 교차모달 크라우드 카운팅을 위한 시공간 주의 블록의 동작 방법은 특징 추출부를 통해 RGB 이미지와 열 이미지 및 깊이 이미지를 포함하는 다중 모달 데이터 쌍으로부터 백본 네트워크를 활용하여 모달리티별 특징을 추출하는 단계(310), CSCA(Cross-modal Spatio-Channel Attention) 블록의 SCA(Spatial-wise Cross-modal Attention) 블록을 통해 RGB 이미지와 열/깊이 이미지 간의 글로벌 기능 상관 관계를 모델링하는 단계(320) 및 CSCA 블록의 CFA(Channel-wise Feature Aggregation) 블록을 통해 채널 상관관계에 기초하여 교차 모달 정보를 적응적으로 추출하고 융합하는 단계(330)를 포함한다.
- [0042] 단계(310)에서, 특징 추출부를 통해 RGB 이미지와 열 이미지 및 깊이 이미지를 포함하는 다중 모달 데이터 쌍으로부터 백본 네트워크를 활용하여 모달리티별 특징을 추출한다.
- [0043] 단계(320)에서, CSCA 블록의 SCA 블록을 통해 RGB 이미지와 열/깊이 이미지 간의 글로벌 기능 상관 관계를 모델링한다.

- [0044] 본 발명의 실시예에 따른 SCA 블록은 비 로컬 기반 모델에서 사용되는 쿼리, 키, 값의 삼중항을 기반으로 하는 주의 메커니즘을 활용하고, RGB 이미지와 열 이미지 및 깊이 이미지 간의 글로벌 기능 상관 관계를 공간 차원으로 모델링하여 정렬 오류를 완화함으로써, 특징 재구성 작업을 통해 계산 오버헤드를 줄일 수 있다.
- [0045] 본 발명의 실시예에 따른 SCA 블록은 다중 모달 데이터 쌍에 대한 서로 다른 모달리티에 대하여 유사성을 나타내는 주의 가중치를 곱하여 출력을 계산하고, 서로 다른 모달리티에 대하여 정보를 집계하고 정렬하며, 특징 맵을 공간 차원의 패치로 랜덤으로 분할하고, 채널 차원을 복구하는 재구성 작업을 수행함으로써 계산 오버헤드를 줄일 수 있다.
- [0047] 도 4는 본 발명의 일 실시예에 따른 SCA 블록을 종래기술과 비교하기 위한 도면이다.
- [0048] 도 4(a)는 종래기술에 따른 비 로컬 기반 모델을 나타내는 도면이고, 도 4(b)는 본 발명의 실시예에 따른 SCA 블록을 나타내는 도면이다.
- [0049] 본 발명의 실시예에 따른 SCA 블록은 교차 모달 주의를 기반으로 더 적은 오버헤드로 다중 모달 데이터 간의 글로벌 기능 상관 관계를 포착하는 것을 목표로 한다.
- [0050] 최근에, 변환기 기반 아키텍처는 컴퓨터 비전 작업에서 동등하거나 심지어 더 우수한 성능을 달성했다. 변환기의 가장 중요한 부분 중 하나인 다중 헤드 주의를 병렬로 실행되는 여러 비 로컬 주의 계층으로 구성된다. 전형적인 비 로컬 블록은 도 4(a)에 나타나 있으며, 입력 이미지의 모든 위치 간의 글로벌 관계를 캡처하는 것을 목표로 한다. 입력 특징 맵 $X \in \mathcal{R}^{C \times H \times W}$ 를 생각해 보자. 여기서 C, H, W는 각각 채널 번호, 공간 높이, 폭을 나타낸다. 특징 맵이 먼저 투영되고 세 개의 벡터로 평평해진다: 쿼리 $Q \in \mathcal{R}^{N \times C'}$, 키 $K \in \mathcal{R}^{N \times C'}$, 값 $V \in \mathcal{R}^{N \times C'}$ (세 가지 다른 1×1 컨볼루션). 여기서 $N = H \times W$, C' 은 새로운 임베딩 채널인 $C' = C/2$ 를 참조한다. 그런 다음 쿼리, 키, 값의 삼중항을 사용하여 특징 맵의 임의의 위치에서 관계를 인코딩함으로써 글로벌 컨텍스트 정보를 캡처한다. 출력은 해당 키와의 쿼리 호환성에 의해 가중치가 각 값에 할당되는 가중치 합으로 계산될 수 있다. 이러한 절차는 행렬 곱셈을 사용하여 다음과 같이 효율적으로 계산할 수 있다:
- [0051]
$$Z = f(QK^T)V, \quad (1)$$
- [0052] 여기서 $f(\cdot)$ 는 소프트맥스, 리스케일링 등과 같은 정규화 함수이다. 마지막으로, 비 로컬 블록은 추가적인 1×1 컨볼루션 층에 의해 C' 에서 C로 채널 차원을 복구한다.
- [0053] 그러나 일반적인 비 로컬 블록은 단일 데이터에서만 작동할 수 있으며 행렬 곱셈에 의해 도입된 높은 복잡성은 사용을 방해한다. 이러한 문제를 완화하기 위해 본 발명에서는 SCA 블록을 제안한다. 일반적인 비 로컬 블록과 달리 SCA 블록의 입력은 모달리티가 다른 이미지 쌍이다. 키와 값은 동일한 형식에서 가져온 것이고 쿼리는 다른 형식에서 가져온 것이다. 예를 들어 키와 값은 RGB 이미지 양식에서 가져온 것이고 쿼리는 열 이미지 또는 깊이 이미지 양식에서 생성된 것이며 그 반대도 마찬가지이다.
- [0054]
$$Z_a = \text{softmax}\left(\frac{Q_b K_a^T}{\sqrt{\hat{C}}}\right) V_a, \quad (2)$$
- [0055] 여기서 a는 RGB 모달리티를 의미하고 b는 서로 다른 모달리티(즉, 열 또는 깊이 모달리티)를 의미한다. SCA 블록(도 4(b))의 출력은 값에 주의 가중치를 곱하여 계산되며, 이 가중치는 양식 b의 해당 쿼리와 다른 양식 a의 모든 키 사이의 유사성을 나타낸다. 즉, SCA 블록은 두 가지 모달리티에서 정보를 집계하고 정렬한다.
- [0056] 교차 주의를 오정렬을 해결하기 위해 사용해 왔다. 그러나 행렬 곱셈 연산은 컨볼루션 및 활성화 블록에 비해 시간과 메모리가 매우 많이 소모된다. 일부 종래기술은 풀링 작업을 통해 키와 값의 차원을 줄임으로써 문제를 완화한다. 그러나, 많은 글로벌 정보를 잃음으로써 그 성능이 낮아질 수 있다. 풀링 기능 대신 SCA 블록에서 재구성 작업을 수행하여 특징 맵을 공간 차원의 G1 패치로 랜덤으로 분할한다. 여기서 1은 백본 네트워크의 1번째

블록을 의미한다. 따라서 SCA 블록의 키, 값, 쿼리 벡터는 $K, V, Q \in \mathcal{R}^{S \times \hat{C}}$ 가 되며, 여기서 $S = HW/G_l$, $\hat{C} = C'G_l$ 이다. 마지막으로, 채널 차원을 복구하고 결과를 $C \times H \times W$ 크기로 재구성한다. 따라서 SCA 블록의 계산 비용은 $O(CH^2W^2/G_l)$ 이며, 이는 일반적인 비 로컬 블록의 계산 복잡도 $O(CH^2W^2)$ 보다 G_l 배 낮다.

[0058] 단계(330)에서, CSCA 블록의 CFA 블록을 통해 채널 상관관계에 기초하여 교차 모달 정보를 적응적으로 추출하고 융합한다.

[0059] 본 발명의 실시예에 따른 CSCA 블록은 공간 상관만을 고려한 비 로컬 기반 모델의 특성을 보완하기 위해 채널 차원에 따라 보완 특징을 동적으로 집계함으로써, 채널 수준에서 공간 상관 특징을 적응적으로 조정한다.

[0060] 본 발명의 실시예에 따른 SCA 블록은 공간적 주의를 기울여 다중 모달 이미지의 모든 위치에서 글로벌 특징을 캡처할 수 있다. 다중 모달 데이터의 상보성을 최대한 활용하려면 특성화 기능에 따라 교차 모달 특징을 보완적으로 집계해야 한다. 이를 달성하기 위해 교차 모드 특징을 융합하고 채널별 상호 작용을 캡처하는 CFA 블록을 제안한다.

[0061] 본 발명의 실시예에 따른 CFA 블록은 SCA 블록의 출력을 연결하여 F_c 를 획득한다. 그런 다음 MLP 층과 소프트맥스 함수를 사용하여 채널 전체에서 RGB 특징과 열 이미지 및 깊이 이미지 특징에 가중치를 다시 부여하는 데 사용되는 가중치 벡터 $w_a, w_b \in \mathcal{R}^{C \times H \times W}$ 를 학습한다.

[0062]
$$[w_a, w_b] = \text{softmax}(\text{MLP}(F_c)). \quad (3)$$

[0063] 집계된 정보를 완전히 활용하고 특징 노이즈와 중복성을 동시에 억제하기 위해 두 개의 채널별 적응 가중치를 사용하여 교차 모드 특징을 융합한다. l^{th} 블록에서 최종 집계 특징 맵 F_{agg}^l 는 다음 공식으로 얻을 수 있다:

[0064]
$$F_{agg}^l = w_a * F_a^l + w_b * F_b^l. \quad (4)$$

[0065] SCA 블록을 통해 네트워크는 중요 채널의 정보 특징에 대한 민감도를 높이고 최종 특징 맵을 향상시킬 수 있다. 실제로 각 계층 l 에 대해 업데이트된 특징 맵 $f_a^l = (F_{agg}^l + F_a^l)/2$ 및 $f_b^l = (F_{agg}^l + F_b^l)/2$ 는 마지막 계층을 제외한 다음 블록으로 전파된다. 따라서 모듈은 보다 차별적이고 풍부한 특성을 적응적으로 제공하여 교차 모달 군중 카운팅 성능을 높일 수 있다.

[0067] 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.

[0068] 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.

[0069] 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.

[0070] 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.

[0071] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

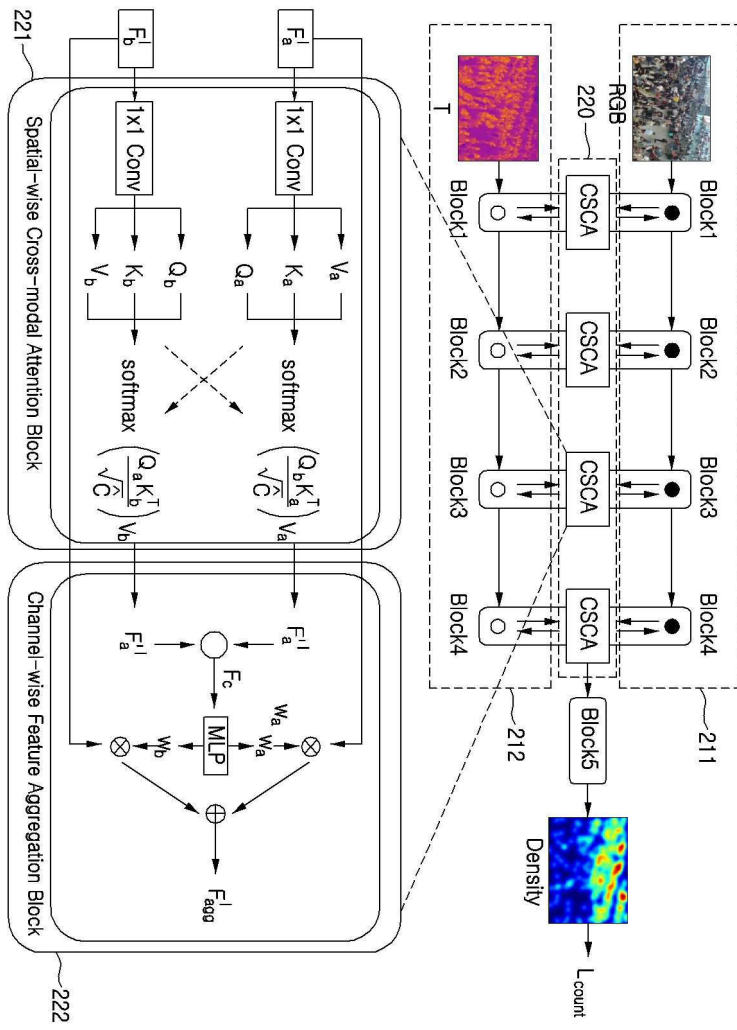
도면

도면1

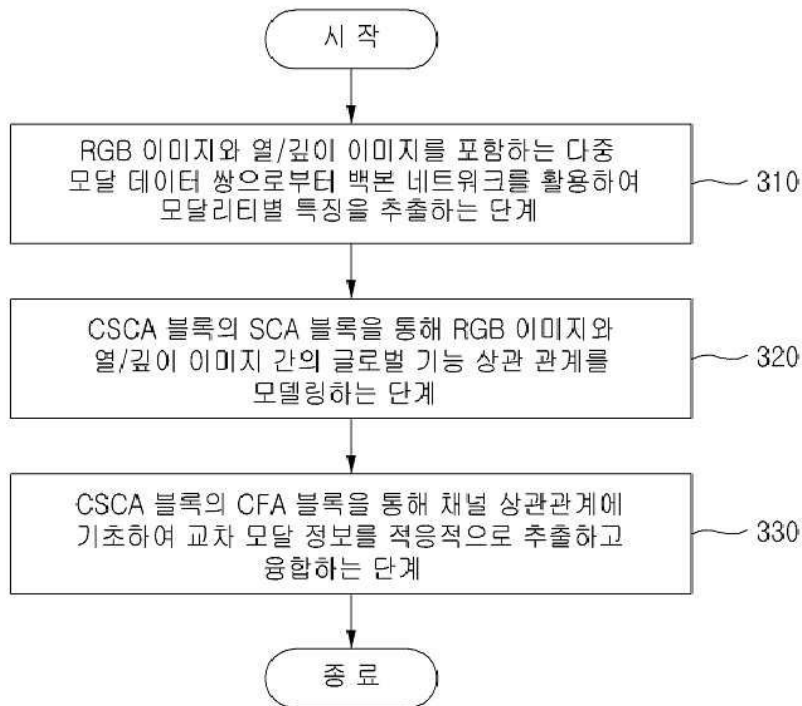


(a) RGB-T (RGB<T) (b) RGB-T (RGB>T) (c) RGB-D (RGB<D) (d) RGB-D (RGB>D)

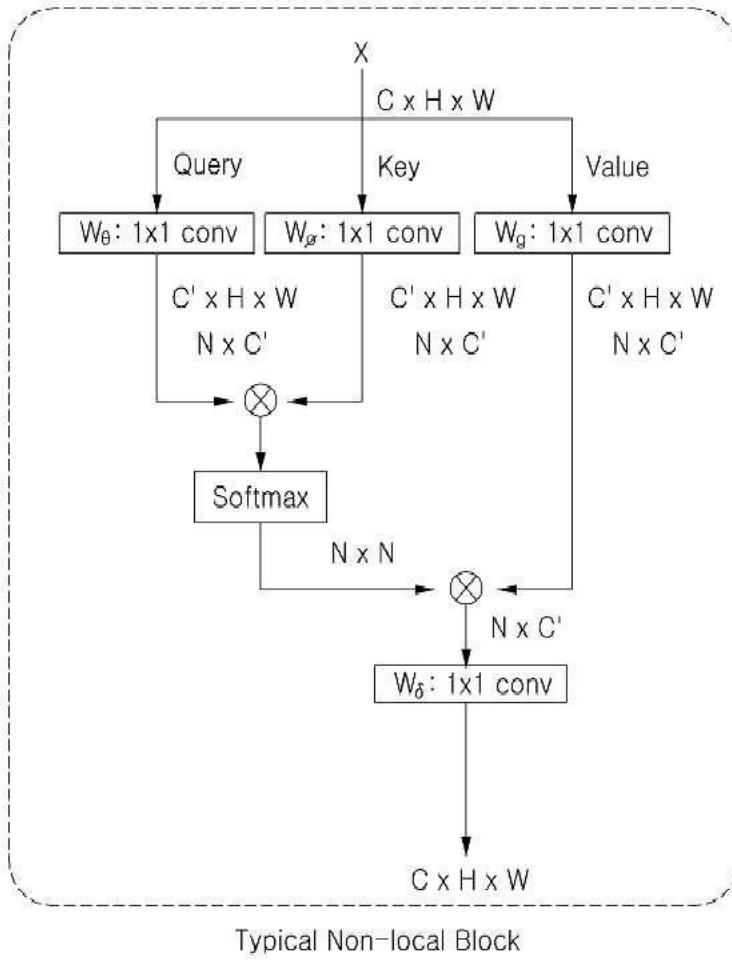
도면2



도면3



도면4a



도면4b

