



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2019년04월19일
 (11) 등록번호 10-1970488
 (24) 등록일자 2019년04월15일

- (51) 국제특허분류(Int. Cl.)
 G06T 7/10 (2017.01) G06T 7/50 (2017.01)
 G06T 7/90 (2017.01)
- (52) CPC특허분류
 G06T 7/10 (2017.01)
 G06K 9/00718 (2013.01)
- (21) 출원번호 10-2017-0182798
- (22) 출원일자 2017년12월28일
 심사청구일자 2017년12월28일
- (56) 선행기술조사문헌
 "FuseNet: Incorporating Depth into Semantic Segmentation via Fusion-Based CNN Architecture", ACCV*
 "심층 CNN 모델을 이용한 외부 및 내부 예시영상 기반의 고 해상도 복원 기법", 대한전자공학회*
 RefineNet: Multi-Path Refinement Networks for High-Resolution Semantic Segmentation
 *는 심사관에 의하여 인용된 문헌

- (73) 특허권자
 포항공과대학교 산학협력단
 경상북도 포항시 남구 청암로 77 (지곡동)
- (72) 발명자
 박성진
 경상북도 포항시 남구 청암로 77, 포항공과대학교
 기숙사 11동 314호 (지곡동)
 홍기상
 경상북도 포항시 남구 청암로 77, 포항공과대학교
 공학2동 334호 (지곡동)
 이승용
 경상북도 포항시 남구 청암로 77 포항공과대학교
 공학4동 121호 (지곡동)
- (74) 대리인
 김합곤, 안광석, 박영일

전체 청구항 수 : 총 8 항

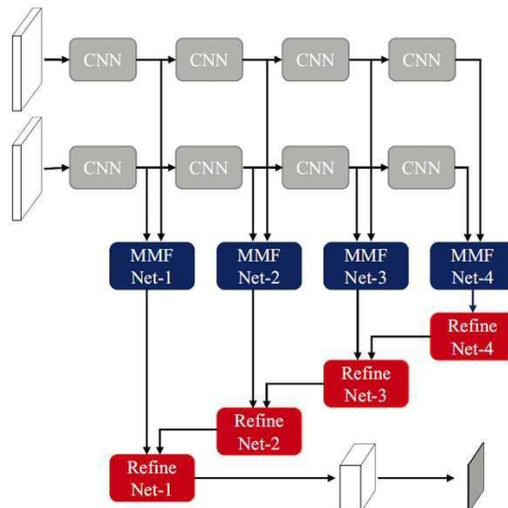
심사관 : 신재철

(54) 발명의 명칭 **실내 의미론적 분할을 위한 컬러-깊이 영상의 단계적 레지듀얼 특성 결합 네트워크 장치**

(57) 요약

본 발명은 한 장의 실내 컬러 영상(RGB)과 그에 대한 깊이 영상(Depth)을 활용하여 딥러닝 기반 픽셀단위 의미론적 분할을 수행하는 네트워크 장치에 관한 것으로, 상위 단계에서 하위 단계로 이어지고 각 단계별로 CNN으로 된 다수의 단계로 구성되어, 컬러 영상으로부터 단계적 특성 정보를 추출하는 제1 레지듀얼 네트워크 모듈; 상위 단

(뒷면에 계속)
대표도 - 도3



계에서 하위 단계로 이어지고 각 단계별로 CNN으로 된 다수의 단계로 구성되어, 깊이 영상으로부터 단계적 특성 정보를 추출하는 제2 레지듀얼 네트워크 모듈; 및 제1 레지듀얼 네트워크 모듈 및 제2 레지듀얼 네트워크 모듈의 각 단계별로 추출된 특성 정보를 결합하는 다중 특성 결합 네트워크(MMFNet) 모듈;을 포함하여 구성되고, 다중 특성 결합 네트워크 모듈은 각 단계별로, 제1 레지듀얼 네트워크 모듈 및 제2 레지듀얼 네트워크 모듈의 대응하는 단계에서 추출된 컬러 특성 정보 및 깊이 특성 정보 각각에 대하여, 특성 정보의 차원을 줄여서 파라미터의 급증을 완화하는 컨볼루션 블록; 형태 결합을 위해 비선형 변형을 수행하는 2개의 레지듀얼 컨볼루션 유닛; 및 다른 형태의 특성 정보를 적응적으로 결합하고, 덧셈을 위해 특성 값의 스케일을 조정하는 컨볼루션 블록;이 차례로 연결되고, 스케일 조정된 컬러 특성 정보와 깊이 특성 정보가 덧셈 결합되는 것을 특징으로 한다.

(52) CPC특허분류

- G06K 9/00744 (2013.01)
- G06T 7/50 (2017.01)
- G06T 7/90 (2017.01)
- G06T 2207/20081 (2013.01)
- G06T 2207/20084 (2013.01)

이 발명을 지원한 국가연구개발사업

- 과제고유번호 1711055313/IITP-2015-0-00174
- 부처명 과학기술정보통신부
- 연구관리전문기관 정보통신기술진흥센터
- 연구사업명 SW 컴퓨팅 산업원천 기술개발 사업
- 연구과제명 (SW 스타랩) 빅 비주얼 데이터 기반의 고품질 사진 메이크업 SW 개발
- 기여율 50/100
- 주관기관 포항공과대학교 산학협력단
- 연구기간 2017.03.01 ~ 2017.12.31

이 발명을 지원한 국가연구개발사업

- 과제고유번호 1711058535/GK17P0300
- 부처명 과학기술정보통신부
- 연구관리전문기관 재단법인 기가코리아산업단
- 연구사업명 법부처 Giga KOREA 사업
- 연구과제명 초실감 서비스를 위한 동적 객체의 실시간 4D 복원 기술 개발
- 기여율 20/100
- 주관기관 포항공과대학교 산학협력단/(세부주관)광주과학기술원
- 연구기간 2017.04.01 ~ 2017.12.31

이 발명을 지원한 국가연구개발사업

- 과제고유번호 1711043394/NRF-2014R1A2A1A11052779
- 부처명 과학기술정보통신부
- 연구관리전문기관 재단법인 한국연구재단
- 연구사업명 중견연구자지원사업(핵심)
- 연구과제명 계층적 구조정보 기반의 실시간 대규모 환경 복원
- 기여율 30/100
- 주관기관 포항공과대학교 산학협력단
- 연구기간 2016.11.01 ~ 2017.10.31

공시예외적용 : 있음

명세서

청구범위

청구항 1

삭제

청구항 2

삭제

청구항 3

한 장의 실내 컬러 영상(RGB)과 그에 대한 깊이 영상(Depth)을 활용하여 딥러닝 기반 픽셀단위 의미론적 분할을 수행하는 네트워크 장치에 있어서,

상위 단계에서 하위 단계로 이어지고 각 단계별로 CNN(Convolutional Neural Network)으로 된 다수의 단계로 구성되어, 컬러 영상으로부터 단계적 특성 정보를 추출하는 제1 레지듀얼 네트워크 모듈;

상위 단계에서 하위 단계로 이어지고 각 단계별로 CNN(Convolutional Neural Network)으로 된 다수의 단계로 구성되어, 깊이 영상으로부터 단계적 특성 정보를 추출하는 제2 레지듀얼 네트워크 모듈; 및

상기 제1 레지듀얼 네트워크 모듈 및 상기 제2 레지듀얼 네트워크 모듈의 각 단계별로 추출된 특성 정보를 결합하는 다중 특성 결합 네트워크(MMFNet) 모듈;을 포함하여 구성되고,

상기 다중 특성 결합 네트워크 모듈은 각 단계별로, 상기 제1 레지듀얼 네트워크 모듈 및 상기 제2 레지듀얼 네트워크 모듈의 대응하는 단계에서 추출된 컬러 특성 정보 및 깊이 특성 정보 각각에 대하여,

특성 정보의 차원을 줄여서 파라미터의 급증을 완화하는 컨볼루션 블록;

형태 결합을 위해 비선형 변형을 수행하는 2개의 레지듀얼 컨볼루션 유닛; 및

다른 형태의 특성 정보를 적응적으로 결합하고, 덧셈을 위해 특성 값의 스케일을 조정하는 컨볼루션 블록;이 차례로 연결되고,

스케일 조정된 컬러 특성 정보와 깊이 특성 정보가 덧셈 결합되는 것을 특징으로 하는 실내 의미론적 분할을 위한 컬러-깊이 영상의 단계적 레지듀얼 특성 결합 네트워크 장치.

청구항 4

제3항에 있어서,

컬러 특성 정보와 깊이 특성 정보가 결합된 특성 정보에 문맥상의 정보를 합체시키는 레지듀얼 풀링 블록을 더 포함하는 것을 특징으로 하는 실내 의미론적 분할을 위한 컬러-깊이 영상의 단계적 레지듀얼 특성 결합 네트워크 장치.

청구항 5

제3항에 있어서, 각 단계별로

하위 단계의 정제된 특성 정보와 상기 다중 특성 결합 네트워크 모듈의 대응하는 단계에서 결합된 특성 정보를 입력으로 하여 정제하는 RefineNet을 더 포함함을 특징으로 하는 실내 의미론적 분할을 위한 컬러-깊이 영상의 단계적 레지듀얼 특성 결합 네트워크 장치.

청구항 6

삭제

청구항 7

삭제

청구항 8

한 장의 실내 컬러 영상(RGB)과 그에 대한 깊이 영상(Depth)을 활용하여 딥러닝 기반 픽셀단위 의미론적 분할을 수행하는 방법에 있어서,

- (a) 상위 단계에서 하위 단계로 이어지고 각 단계별로 CNN(Convolutional Neural Network)으로 된 다수의 단계로 구성된 제1 레지듀얼 네트워크 모듈이 컬러 영상을 입력으로 하여, 단계적 특성 정보를 추출하는 단계;
- (b) 상위 단계에서 하위 단계로 이어지고 각 단계별로 CNN(Convolutional Neural Network)으로 된 다수의 단계로 구성된 제2 레지듀얼 네트워크 모듈이 깊이 영상을 입력으로 하여, 단계적 특성 정보를 추출하는 단계; 및
- (c) 다중 특성 결합 네트워크 모듈이 상기 제1 레지듀얼 네트워크 모듈 및 상기 제2 레지듀얼 네트워크 모듈의 각 단계별로 추출된 특성 정보를 결합하는 단계;를 포함하여 구성되고,

상기 (c) 단계는

상기 다중 특성 결합 네트워크 모듈은 각 단계별로, 상기 제1 레지듀얼 네트워크 모듈 및 상기 제2 레지듀얼 네트워크 모듈의 대응하는 단계에서 추출된 컬러 특성 정보 및 깊이 특성 정보 각각에 대하여,

특성 정보의 차원을 줄여서 파라미터의 급증을 완화하는 단계;

형태 결합을 위해 비선형 변형을 수행하는 단계; 및

다른 형태의 특성 정보를 적응적으로 결합하고, 덧셈을 위해 특성 값의 스케일을 조정하는 단계;를 구비하고,

스케일 조정된 컬러 특성 정보와 깊이 특성 정보를 덧셈 결합하는 것을 특징으로 하는 단계적 레지듀얼 특성 결합 네트워크 장치를 이용한 실내 컬러-깊이 영상의 의미론적 분할 방법.

청구항 9

제8항에 있어서, 상기 (c) 단계는

컬러 특성 정보와 깊이 특성 정보가 결합된 특성 정보에 문맥상의 정보를 합체시키는 단계를 더 구비하는 것을 특징으로 하는 단계적 레지듀얼 특성 결합 네트워크 장치를 이용한 실내 컬러-깊이 영상의 의미론적 분할 방법.

청구항 10

제8항에 있어서,

(d) RefineNet가 각 단계별로 하위 단계의 정제된 특성 정보와 상기 다중 특성 결합 네트워크 모듈의 대응하는 단계에서 결합된 특성 정보를 입력으로 하여 정제하는 단계;를 더 포함함을 특징으로 하는 단계적 레지듀얼 특성 결합 네트워크 장치를 이용한 실내 컬러-깊이 영상의 의미론적 분할 방법.

청구항 11

제8항 내지 제10항 중 어느 한 항의 단계적 레지듀얼 특성 결합 네트워크 장치를 이용한 실내 컬러-깊이 영상의 의미론적 분할 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록 매체.

청구항 12

제8항 내지 제10항 중 어느 한 항의 단계적 레지듀얼 특성 결합 네트워크 장치를 이용한 실내 컬러-깊이 영상의 의미론적 분할 방법을 컴퓨터에서 실행시키기 위하여 매체에 저장된 컴퓨터 프로그램.

발명의 설명

기술 분야

본 발명은 한 장의 실내 컬러 영상(RGB)과 그에 대한 깊이 영상(Depth)을 활용하여 딥러닝 기반 픽셀단위 의미론적 분할 (semantic segmentation)을 수행하는 네트워크 장치에 관한 것이다.

[0001]

배경 기술

- [0002] 영상을 픽셀 단위로 의미론적 분할하여 인식하는 기술은 많은 영상 처리 기술의 근본이 될 수 있는 기술로, 자율 주행과 같은 분야에서 특히 매우 중요한 영상 정보를 제공할 수 있다.
- [0003] 일반적인 컬러 영상에 대한 의미론적 분할의 최신 기술은 RefineNet이 있다(도 1 참고). RefineNet은 기존의 영상 분류 기술에 사용되었던 Kaiming He의 "Deep Residual Learning for Image Recognition (Residual Network)"를 기반으로, 영상에서 다양한 차원의 특성들을 추출한 뒤, 이를 단계적으로 결합하면서 보다 선명하고 정확한 의미론적 분할을 얻을 수 있는 네트워크 모듈을 설계한다.
- [0004] 도 1에서 (a)는 의미론적 분할을 위한 네트워크 구조이고, (b)는 RefineNet 블록의 상세 구성도이고, (c) 내지 (e)는 각각 RefineNet의 하위 모듈들에 대한 상세 구성도이다. 도 1에서 Conv는 컨볼루션(convolution), ReLU는 Rectified Linear Unit 활성화 함수, Pool은 Pooling 연산을 수행한다.
- [0005] 그런데 실내 영상의 경우 일반적으로 영상의 복잡도가 높아서 좋은 인식 성능을 얻기에 어려움이 있는 반면, Microsoft Kinect와 같은 장비로부터 깊이 영상을 얻기가 용이해 이를 컬러 영상과 함께 사용하여 정확도를 개선할 수 있다.
- [0006] 도 2를 참조하면, 컬러 영상과 깊이 영상을 함께 사용한 기존의 딥러닝 기반 기술들은 다음과 같다. 도 2에서, (a)는 네트워크 모듈의 입력으로 컬러 영상과 깊이 영상을 단순히 연결하여 넣는 방법, (b)는 각각의 영상으로부터 특성을 추출 후 이를 네트워크 모듈의 마지막에서 단순히 결합하는 방법, (c)는 J.Wang의 "Learning common and specific features for rgb-d semantic segmentation with deconvolutional networks", (d)는 C.Hazirbas의"Fusenet: Incorporating depth into semantic segmentation via fusionbased cnn architecture"에 대한 구조를 도식화한 것이다. 여기서, 'C', 'T', '+'는 각각 연결(Concatenation), 변환(Transformation), 요소 단위 덧셈(element-wise Summation)을 의미한다.
- [0007] J.Wang의 네트워크 모듈은 우선 컬러 영상과 깊이 영상 각각으로부터 특성을 어느 정도 추출한다. 이를 특정 단계(layer)에서 각 영상의 공통 특성과 특수한 특성으로 나눈 후 다시 네트워크 모듈을 통과시킨 뒤 특성을 결합하여 최종 분할 결과를 얻는다. C.Hazirbas는 네트워크 모듈의 앞 부분인 인코더(encoder) 파트에서 각각의 영상의 특성을 추출함과 동시에 순차적으로 특성을 더해주고, 얻은 특성 맵을 디컨볼루션(deconvolution) 네트워크 모듈을 통과시켜 최종 분할 결과를 얻는다.
- [0008] 이러한 네트워크들은 각각의 영상으로부터 딥러닝을 통해 추출할 수 있는 여러 단계(layer)의 특성 정보들을 충분히 활용하기 어려우며, 각각의 특성 정보들을 결합하는 부분도 직접 결정해주어야 하는 단점이 있다. 이로 인해 각각의 영상 특성을 효과적으로 추출할 수 있는 네트워크 모듈을 학습시키는 점에 있어서도 어려움이 있을 수 있다.

선행기술문헌

비특허문헌

- [0009] (비특허문헌 0001) Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770-778, 2016.

발명의 내용

해결하려는 과제

- [0010] 본 발명은 실내 컬러 영상과 깊이 영상으로부터 의미론적 분할을 통해 영상을 인식함에 있어서, 효과적으로 다양한 차원의 특성 정보들을 추출함과 동시에 결합함으로써, 이를 효율적으로 학습시킬 수 있고 보다 정확한 의미론적 분할 결과를 얻을 수 있는 실내 의미론적 분할을 위한 컬러-깊이 영상의 단계적 레지듀얼 특성 결합 네트워크 장치 및 단계적 레지듀얼 특성 결합 네트워크 장치를 이용한 실내 컬러-깊이 영상의 의미론적 분할 방법을 제공함을 그 목적으로 한다.

과제의 해결 수단

- [0011] 상기의 목적을 달성하기 위하여, 본 발명에 의한 실내 의미론적 분할을 위한 컬러-깊이 영상의 단계적 레지듀얼 특성 결합 네트워크 장치는 한 장의 실내 컬러 영상(RGB)과 그에 대한 깊이 영상(Depth)을 활용하여 딥러닝 기반 픽셀단위 의미론적 분할을 수행하는 네트워크 장치에 있어서, 상위 단계에서 하위 단계로 이어지고 각 단계별로 CNN(Convolutional Neural Network)으로 된 다수의 단계로 구성되어, 컬러 영상으로부터 단계적 특성 정보를 추출하는 제1 레지듀얼 네트워크 모듈; 상위 단계에서 하위 단계로 이어지고 각 단계별로 CNN(Convolutional Neural Network)으로 된 다수의 단계로 구성되어, 깊이 영상으로부터 단계적 특성 정보를 추출하는 제2 레지듀얼 네트워크 모듈; 및 상기 제1 레지듀얼 네트워크 모듈 및 상기 제2 레지듀얼 네트워크 모듈의 각 단계별로 추출된 특성 정보를 결합하는 다중 특성 결합 네트워크(MMFNet) 모듈;을 포함하여 구성되고, 상기 다중 특성 결합 네트워크 모듈은 각 단계별로, 상기 제1 레지듀얼 네트워크 모듈 및 상기 제2 레지듀얼 네트워크 모듈의 대응하는 단계에서 추출된 컬러 특성 정보 및 깊이 특성 정보 각각에 대하여, 특성 정보의 차원을 줄여서 파라미터의 급증을 완화하는 콘볼루션 블록; 형태 결합을 위해 비선형 변형을 수행하는 2개의 레지듀얼 콘볼루션 유닛; 및 다른 형태의 특성 정보를 적응적으로 결합하고, 덧셈을 위해 특성 값의 스케일을 조정하는 콘볼루션 블록;이 차례로 연결되고, 스케일 조정된 컬러 특성 정보와 깊이 특성 정보가 덧셈 결합되는 것을 특징으로 한다.
 상기 실내 의미론적 분할을 위한 컬러-깊이 영상의 단계적 레지듀얼 특성 결합 네트워크 장치에 있어서, 컬러 특성 정보와 깊이 특성 정보가 결합된 특성 정보에 문맥상의 정보를 합체시키는 레지듀얼 풀링 블록을 더 포함하여 구성된다.
- [0012] 삭제
- [0013] 삭제
- [0014] 삭제
- [0015] 상기 실내 의미론적 분할을 위한 컬러-깊이 영상의 단계적 레지듀얼 특성 결합 네트워크 장치에 있어서, 각 단계별로 하위 단계의 정제된 특성 정보와 상기 다중 특성 결합 네트워크 모듈의 대응하는 단계에서 결합된 특성 정보를 입력으로 하여 정제하는 RefineNet을 더 포함하여 구성된다.
- [0016] 상기의 목적을 달성하기 위하여, 본 발명에 의한 단계적 레지듀얼 특성 결합 네트워크 장치를 이용한 실내 컬러-깊이 영상의 의미론적 분할 방법은 한 장의 실내 컬러 영상(RGB)과 그에 대한 깊이 영상(Depth)을 활용하여 딥러닝 기반 픽셀단위 의미론적 분할을 수행하는 방법에 있어서, (a) 상위 단계에서 하위 단계로 이어지고 각 단계별로 CNN(Convolutional Neural Network)으로 된 다수의 단계로 구성된 제1 레지듀얼 네트워크 모듈이 컬러 영상을 입력으로 하여, 단계적 특성 정보를 추출하는 단계; (b) 상위 단계에서 하위 단계로 이어지고 각 단계별로 CNN(Convolutional Neural Network)으로 된 다수의 단계로 구성된 제2 레지듀얼 네트워크 모듈이 깊이 영상을 입력으로 하여, 단계적 특성 정보를 추출하는 단계; 및 (c) 다중 특성 결합 네트워크 모듈이 상기 제1 레지듀얼 네트워크 모듈 및 상기 제2 레지듀얼 네트워크 모듈의 각 단계별로 추출된 특성 정보를 결합하는 단계;를 포함하여 구성되고, 상기 (c) 단계는 상기 다중 특성 결합 네트워크 모듈은 각 단계별로, 상기 제1 레지듀얼 네트워크 모듈 및 상기 제2 레지듀얼 네트워크 모듈의 대응하는 단계에서 추출된 컬러 특성 정보 및 깊이 특성 정보 각각에 대하여, 특성 정보의 차원을 줄여서 파라미터의 급증을 완화하는 단계; 형태 결합을 위해 비선형 변형을 수행하는 단계; 및 다른 형태의 특성 정보를 적응적으로 결합하고, 덧셈을 위해 특성 값의 스케일을 조정하는 단계;를 구비하고, 스케일 조정된 컬러 특성 정보와 깊이 특성 정보를 덧셈 결합하는 것을 특징으로 한다.
- [0017] 삭제
- [0018] 삭제
- [0019] 상기 단계적 레지듀얼 특성 결합 네트워크 장치를 이용한 실내 컬러-깊이 영상의 의미론적 분할 방법에 있어서,

상기 (c) 단계는 컬러 특성 정보와 깊이 특성 정보가 결합된 특성 정보에 문맥상의 정보를 합체시키는 단계를 더 구비하는 것을 특징으로 한다.

[0020] 상기 단계적 레지듀얼 특성 결합 네트워크 장치를 이용한 실내 컬러-깊이 영상의 의미론적 분할 방법에 있어서, (d) RefineNet가 각 단계별로 하위 단계의 정제된 특성 정보와 상기 다중 특성 결합 네트워크의 대응하는 단계에서 결합된 특성 정보를 입력으로 하여 정제하는 단계;를 더 포함하여 구성된다.

[0021] 상기의 목적을 달성하기 위하여, 본 발명에 의한 컴퓨터로 읽을 수 있는 기록 매체는 상기 단계적 레지듀얼 특성 결합 네트워크 장치를 이용한 실내 컬러-깊이 영상의 의미론적 분할 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한다.

[0022] 상기의 목적을 달성하기 위하여, 본 발명에 의한 컴퓨터 프로그램은 상기 단계적 레지듀얼 특성 결합 네트워크 장치를 이용한 실내 컬러-깊이 영상의 의미론적 분할 방법을 컴퓨터에서 실행시키기 위하여 매체에 저장된다.

발명의 효과

[0023] 본 발명에 의하면, 실내 의미론적 분할을 위하여 컬러-깊이 영상에 대하여 단순히 순차적으로 특성을 추출하고 결합하고 개선하는 기능을 하는 것뿐만 아니라, 네트워크를 학습할 때 모든 과정이 효과적으로 이루어질 수 있는 방향으로 파라미터들이 자동으로 학습되는 것을 가능하게 한다.

도면의 간단한 설명

- [0024] 도 1은 컬러 영상에 대한 의미론적 분할의 RefineNet의 구조를 보여주는 구조도이다.
- 도 2는 컬러 영상과 깊이 영상을 함께 사용한 기존의 딥러닝 기반 기술들의 구조를 보여준다.
- 도 3은 실내 의미론적 분할을 위한 컬러-깊이 영상의 단계적 레지듀얼 특성 결합 네트워크 장치의 전체 구조를 보여준다.
- 도 4는 2개의 레이어를 갖는 일반적인 CNN 구조(a)와 2개의 레이어에 대해 skip connection 구조가 추가된 레지듀얼 네트워크 구조(b)를 보여주는 도면이다.
- 도 5는 도 3의 MMFNet에 해당하는 부분을 상세히 설명하는 구조도이다.
- 도 6은 의미론적 분할의 정성적인 결과를 보여준다.

발명을 실시하기 위한 구체적인 내용

- [0025] 이하에서, 첨부된 도면을 참조하면서 본 발명의 바람직한 실시예에 대하여 상세히 설명하기로 한다.
- [0026] 본 발명은 컬러(RGB) 영상과 깊이 영상으로부터 단계별 특성 정보를 추출하는 부분(1 단계), 추출된 특성 정보를 단계마다 효과적으로 결합하는 부분 (2 단계), 단계별로 결합하여 얻은 특성 정보를 다시 단계적으로 결합해 가며 특성 정보를 더욱 선명하게 개선하는 부분(3단계)으로 나누어 볼 수 있다.
- [0027] 이렇게 구성된 네트워크의 최종 특성 정보 맵으로부터 실내 영상의 의미론적 분할 결과를 얻을 수 있다.
- [0028] 본 발명에 의한 실내 의미론적 분할을 위한 컬러-깊이 영상의 단계적 레지듀얼 특성 결합 네트워크 장치(RDFNet이라 명명한다)의 전체 구조는 도 3에 도시된다.
- [0029] 컬러-깊이 영상의 의미론적 분할에서 주된 이슈는 칼라 특성을 따라 깊이 특성을 효율적으로 추출하고, 의미론적 분할의 원하는 작업에 그와 같은 특성을 활용하는 것이다. RefineNet은 컬러 영상만에 대해 다른 레벨의 특성을 결합하는 수단을 제시하였지만, 본 발명에 의한 RDFNet은 컬러-깊이 영상에 대해 skip connection의 잇점을 유지하는 다중 형태 CNN 특성 결합을 위한 간단한 구조를 도입한다.
- [0030] RDFNet은 다중 형태 CNN 특성 결합을 처리하도록 하고, 결합된 특성 정제를 위해 RefineNet 블록들을 구비한다. 깊이 정보를 활용하는 기존의 네트워크와는 달리, RDFNet는 ResNet를 기초로 한 추가적인 깊은 깊이 특성 경로를 가지고 MMF 블록을 통해 다중 레벨 깊이 특성 정보를 완전히 활용하도록 설계된다.
- [0032] <1 단계(ResNet)>
- [0033] 도 3에서 CNN(Convolutional Neural Network)으로 표시되어 단계적으로 이어지는 네트워크 모듈은 컬러 영상과 깊이 영상으로부터 단계적 특성 정보를 추출하는 네트워크 모듈이다. 본 발명에서는 최근 영상 분류에서 가장

좋은 성능을 보였던 Kaiming He의 레지듀얼 네트워크(Residual Network : ResNet) 모듈의 구조를 활용하고 영상 분류 데이터셋에 미리 학습된 파라미터를 의미론적 영상 분할 목적에 맞게 수정하도록 학습 시킨다. 따라서, 본 발명에서 특성 정보를 추출하는 네트워크 모듈은 컬러 영상으로부터 단계적 특성 정보를 추출하는 제1 레지듀얼 네트워크 모듈과 깊이 영상으로부터 단계적 특성 정보를 추출하는 제2 레지듀얼 네트워크 모듈로 구성된다.

[0034] 기존의 CNN 구조들은 역전파(backpropagation) 기법을 이용하여 학습을 진행하는데 CNN 구조가 깊을수록 전 단계의 그라디언트(gradient) 정도에 따라서 다음 단계의 그라디언트 값이 지수적으로 작아지는 현상이 발생할 수 있다. 따라서 출력 단계 근처의 그라디언트들은 값이 존재하지만, 입력 단계에서는 그 값이 0에 가까워지는 문제가 발생해 학습이 더 이상 진행되지 않는 문제가 발생한다(gradient vanishing problem).

[0035] 레지듀얼 네트워크(residual network) 모듈은 이러한 gradient vanishing 문제를 해결하여 깊은 네트워크 구조의 효과적인 학습을 가능하게 한 구조이다. 이 네트워크 모듈에서는 skip connection 이라는 구조를 이용하여 입력 값을 몇 개의 단계에 통과시킨 출력 값(도 4의 F(x))과 입력 값(도 4의 identity)의 합을 계산하여 결과 값을 얻을 수 있도록 구성한다.

$$H(x)$$

[0036] 도 4의 (a)가 기존의 일반적인 네트워크 구조인데, 이러한 구조일 때 두 단계를 거친 후 매핑된 결과는 로 표현할 수 있다. 도 4의 (b)는 레지듀얼 네트워크(residual network) 모듈 구조인데, 2개의 단계를 건너 뛴

$$H(x) = F(x) + x$$

skip connection을 활용한 것이 특징이다. 여기서 는 로 표현할 수 있다. 이러한 구성은 그라디언트를 계산하는데 있어 그 값이 너무 작아지지 않게 해주기 때문에 기존 CNN에서 발생되던 gradient vanishing 문제를 해결할 수 있다. 따라서 보다 많은 단계의 모델을 잘 학습시킬 수 있고 적은 수의 단계를 갖는 모델로 더 높은 성능을 얻을 수 있다.

[0038] <2 단계(MMFNet)>

[0039] 도 3에서 다중 특성 결합 네트워크(Multi Modal Feature Fusion Network: MMFNet) 모듈로 표시되는 부분은 제1 레지듀얼 네트워크 모듈 및 제2 레지듀얼 네트워크 모듈의 각 단계별로 추출된 특성 정보를 결합한다. 컬러 영상과 깊이 영상에서 추출한 특성들을 단순히 연결(concatenation)하거나 더하는 것만으로는 충분히 효과적으로 두 특성을 활용할 수 없다. 따라서 본 네트워크 모듈에서는 컬러 영상에서 여러 단계의 특성을 결합하는 데 우수성이 입증된 구조를 서로 다른 영상에서 추출한 특성 결합에 맞추도록 변형하여 적용한다. 또한 기존의 기술들이 각각의 영상에서 추출한 특성들을 네트워크의 특정 단계에서 결합한 것과 달리, 본 발명에서는 1 단계에서 얻을 수 있는 중간 단계 특성들을 대부분 활용할 수 있도록 네트워크 모듈을 설계한다. 이렇게 설계된 구조는 이를 학습시키는 과정에서 자연스럽게 의미론적 분할에 중요한 특성들이 각각의 영상으로부터 자동적으로 추출, 결합되는 것을 가능하도록 해준다.

[0040] MMFNet의 세부 구성은 도 5에 도시된다.

[0041] 칼라(RGB), 깊이에 대한 제1 레지듀얼 네트워크 모듈 및 제2 레지듀얼 네트워크 모듈의 특성 정보가 주어졌을 때, MMFNet은 먼저 하나의 콜볼루션 블록을 통해 각 특성 정보의 차원을 줄여서, 파라미터의 급증을 완화하면서 효율적인 훈련을 활성화한다.

[0042] 그 다음, 각 특성 정보는 RefineNet에서 처럼 2개의 레지듀얼 콜볼루션 유닛 (Residual Convolution Unit : RCU)와 하나의 콘볼루션 블록을 지나간다. MMFNet과 RefineNet에서는 RCU의 목적에서 차이가 있다. MMFNet에서의 RCU는 형태 결합을 위해 특별히 비선형 변형을 수행하도록 의도된다. 다른 형태(modality)에서의 두 특성은 RCU의 동작을 통해 상호 개선되도록 보상적으로 결합된다. MMFNet에서는, RefineNet에서 처럼 하위 레벨 특성에 고해상도를 도입함으로써 거친 상위 레벨 특성을 정제한다.

[0043] MMFNet에서의 다음의 추가적인 콘볼루션 블록은 다른 형태의 특성들을 적응적으로 결합하고, 덧셈을 위해 특성 값의 스케일을 적절히 조정함에 있어서 중요하다.

[0044] 의미론적 분할에서 칼라 특성은 깊이 특성에 비하여 일반적으로 더 나은 분별력을 가지므로, MMFNet에서의 덧셈 결합은 보충적인 또는 잔여의 깊이 특성을 학습하도록 동작한다. 여기에서의 깊이 특성은 혼란스런 패턴을 식별할 수 있도록 RGB 특성을 개선할 것이다. 각 형태 특성의 중요성은 RCU 이후의 콘볼루션 블록에 있는 학습 가능한 파라미터에 의해 조절될 수 있다.

- [0045] 마지막으로, MMFNet에서는 결합된 특성 정보에 문맥상의 정보(Contextual Information)가 합체되도록 추가적인 레지듀얼 풀링 동작을 수행한다. 각 단계의 MMFNet에서 하나의 레지듀얼 풀링이면 충분하다. 더 강한 문맥 정보는 RefineNet 블록을 통하여 이어지는 다중 레벨 결합에서 더 결합될 수 있다.
- [0046] MMFNet은 모든 단계를 통하여 skip connection을 가진 레지듀얼 학습을 도입함으로써 효율적인 다중 레벨 칼라 (RGB), 깊이 특성 추출과 효율적인 종간 훈련(end-to-end training)이 가능하게 한다.
- [0048] <3 단계(RefineNet)>
- [0049] 도 3에서 RefineNet으로 표시되는 부분은 하위 단계의 정제된 특성 정보와 MMFNet의 대응하는 단계에서 결합된 특성 정보를 입력으로 하여 다시 높은 해상도의 의미론적 분할 결과를 얻기 위해 개선해가는 부분이다. 단계별로 얻은 특성들은 서로 다른 해상도와 차원의 정보를 가지고 있으므로 이들을 단계적으로 잘 결합하는 것이 중요하다.
- [0050] 도 3의 RefineNet은 종래 기술에 의한 RefineNet의 구조를 그대로 활용할 수 있다. 하지만 네트워크 모듈의 입력으로 컬러 영상에서 뽑은 특성을 넣는 대신, 컬러 영상과 깊이 영상에서 효과적으로 추출되고 결합된 특성을 넣어준다는 점에서 차이가 있다.
- [0051] 도 3의 RefineNet은 다중 레벨 MMFNet의 컬러 영상과 깊이 영상에서 추출되고 결합된 특성 정보와 이전에 정제된 특성 정보를 입력으로 받아들인다. 그와 같은 특성들은 도 1의 (b)에 도시된 일련의 하위 구성들(레지듀얼 콜볼루션 유닛, 다중 해상도 결합 블록, 체인드 레지듀얼 풀링)에 의해 정제되고 결합된다.
- [0052] 도 1의 (c)에 도시된 레지듀얼 콜볼루션 유닛(Residual Convolution Unit : RCU)은 의미론적 분할을 위해 미리 훈련된 ResNet의 웨이트를 미세 조정하는 적응적인 콘볼루션 세트이다.
- [0053] 도 1의 (d)에 도시된 다중 해상도 결합 블록(Multiresolution Fusion Block)은 고해상도 특성 맵으로의 다중 경로 입력을 결합한다. 블록에서의 하나의 콘볼루션은 입력 적응을 위한 것으로, 특성 채널의 수를 맞추고 덧셈에 맞추어 특성 값의 스케일을 조정한다.
- [0054] 도 1의 (e)에 도시된 체인드 레지듀얼 풀링(Chained Residual Pooling : CRP)의 목적은 큰 영역으로부터의 문맥상의 정보를 인코드 하기 위한 것이다. 그 블록은 다중 풀링 블록의 체인으로 구성된다. 풀링 블록은 각각 하나의 맥스 풀링(max-pooling) 레이어와 하나의 콘볼루션 레이어로 구성된다. 풀링 동작은 근처로부터 액세스되는 큰 활성화 값을 문맥상의 특성으로 전파하는 효과를 가진다. 추가적인 콘볼루션 레이어는 풀 된 특성의 중요성을 학습하는데, 풀 된 특성은 레지듀얼 커넥션(residual connection)을 통해 원래의 특성과 결합된다.
- [0055] RefineNet의 끝에는 결합된 특성 맵 상에서 비선형 동작을 도입하기 위한 추가적인 RCU가 있다.
- [0057] 본 발명에서 네트워크 모듈의 학습은 확률적 그라디언트 하강(stochastic Gradient descent)를 이용한 역 전파(back-propagation) 알고리즘으로 이루어진다.
- [0059] <평가 결과>
- [0060] 본 발명에서는 NYUDv2와 SUN-RGBD 두 종류의 실내 영상 데이터셋에 대해 네트워크 모듈을 학습 시키고 이를 통해 얻은 의미론적 영상 분할 결과를 세 가지의 평가 기준으로 평가하였다.
- [0061] 첫 번째로는 픽셀별 분류 정확도를 측정하는 Pixel accuracy, 물체 종류별 정확도의 평균 값인 mean accuracy, 그리고 각 물체 종류별 실제 라벨과 겹치는 정도의 평균 값인 mean IoU를 사용하였다.
- [0062] 또한 이 수치를 표 1과 표 2에서 다른 주요 연구 그룹들의 결과와 비교하였다.

표 1

	data	pixel acc.	mean acc.	IoU
Gupta <i>et al.</i> [11]	RGB-D	-	35.1	-
Eigen <i>et al.</i> [7]	RGB-D	65.6	45.1	34.1
FCN [30]	RGB-D	65.4	46.1	34.0
Wang <i>et al.</i> [39]	RGB-D	-	47.3	-
Context [27]	RGB	70.0	53.6	40.6
Refine-101 [26]	RGB	72.8	57.8	44.9
Refine-152 [26]	RGB	73.6	58.9	46.5
RDF-152 (ours)	RGB-D	76.0	62.8	50.1

[0063]

[0064]

<NYUDv2 데이터셋에 대한 의미론적 영상 분할 결과 비교>

[0065]

표 1은 NYUDv2 데이터셋에 대한 결과를 보여준다. 표 1에서 볼 수 있듯이, 본 발명(RDF-152 (ours))의 성능이 컬러 영상과 깊이 영상을 같이 활용한 기존 연구 결과(data: RGB-D)들보다 월등한 성능을 나타냈다. 또한 컬러 영상만을 사용하여 좋은 성능을 냈던 RefineNet[26]보다 컬러 영상과 깊이 영상을 동시에 활용한 본 발명이 훨씬 좋은 성능을 얻었음을 확인할 수 있다.

표 2

	data	pixel acc.	mean acc.	IoU
Ren <i>et al.</i> [32]	RGB-D	-	36.3	-
B-SegNet [19]	RGB	71.2	45.9	30.7
LSTM [25]	RGB-D	-	48.1	-
FuseNet [13]	RGB-D	76.3	48.3	37.3
Context [28]	RGB	78.4	53.4	42.3
Refine-152 [26]	RGB	80.6	58.5	45.9
RDF-152 (ours)	RGB-D	81.5	60.1	47.7

[0066]

[0067]

<SUN-RGBD 데이터셋에 대한 의미론적 영상 분할 결과 비교>

[0068]

표 2는 SUN-RGBD 데이터셋에 대해 실험한 다른 연구 그룹들의 결과와 본 발명의 결과 비교를 보여준다. 표 2에서도 마찬가지로 컬러 영상과 깊이 영상을 활용한 기존 연구(FuseNet)보다 본 발명이 월등한 성능을 보여주며, 컬러 영상만을 사용한 RefineNet보다 나은 성능으로 최고 성능을 보여주는 것을 알 수 있다.

[0069]

표 1과 표 2에서 사용된 문헌은 다음과 같다.

[0070]

[문헌 11] S. Gupta, R. Girshick, P. Arbel' aez, and J. Malik. Learning rich features from rgb-d images for object detection and segmentation. In Proc. ECCV, pages 345-360. Springer, 2014

[0071]

[문헌 7] D. Eigen and R. Fergus. Predicting depth, surface normal and semantic labels with a common multi-scale convolutional architecture. In Proc. ICCV, pages 2650-2658, 2015

[0072]

[문헌 30] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In Proc. CVPR, pages 3431-3440, 2015

[0073]

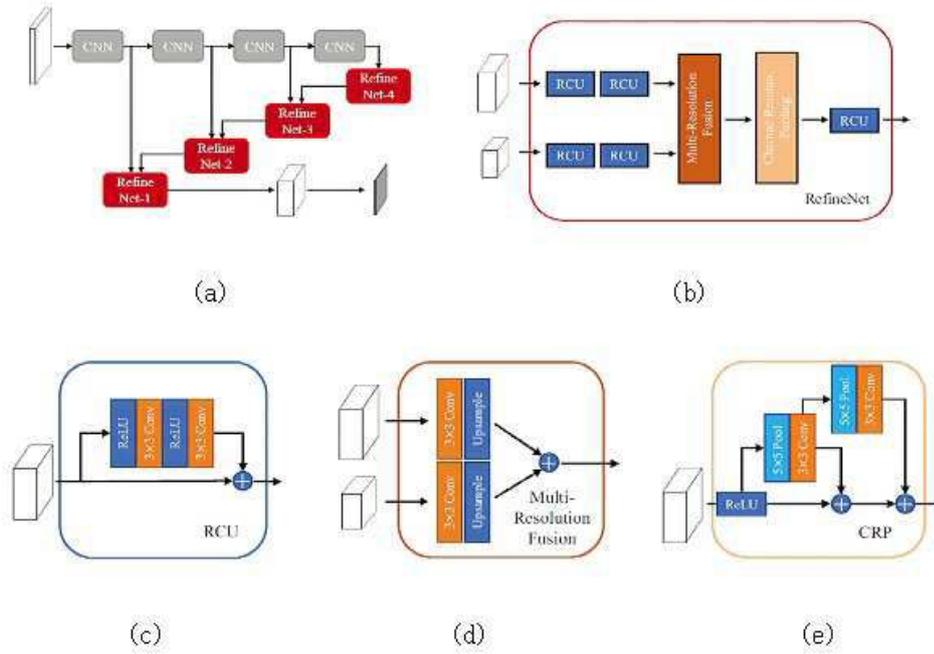
[문헌 39] J. Wang, Z. Wang, D. Tao, S. See, and G. Wang. Learning common and specific features for rgb-d semantic segmentation with deconvolutional networks. In Proc. ECCV, pages 664-679. Springer,

2016.

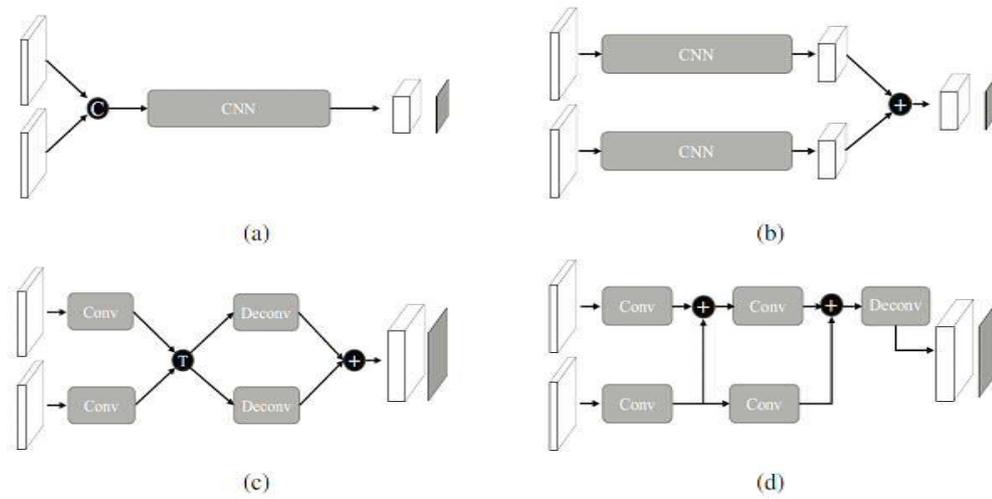
- [0074] [문헌 27] G. Lin, C. Shen, A. v. d. Hengel, and I. Reid. Exploring context with deep structured models for semantic segmentation. arXiv preprint arXiv:1603.03183, 2016
- [0075] [문헌 26] G. Lin, A. Milan, C. Shen, and I. Reid. RefineNet: Multipath refinement networks for high-resolution semantic segmentation. In CVPR, July 2017
- [0076] [문헌 32] X. Ren, L. Bo, and D. Fox. Rgb-(d) scene labeling: Features and algorithms. In Proc. CVPR, pages 2759-2766. IEEE, 2012.
- [0077] [문헌 19] A. Kendall, V. Badrinarayanan, and R. Cipolla. Bayesian segnet: Model uncertainty in deep convolutional encoderdecoder architectures for scene understanding. arXiv preprint arXiv:1511.02680, 2015.
- [0078] [문헌 25] Z. Li, Y. Gan, X. Liang, Y. Yu, H. Cheng, and L. Lin. Lstmcf: Unifying context modeling and fusion with lstms for rgb-d scene labeling. In Proc. ECCV, pages 541-557. Springer, 2016.
- [0079] [문헌 13] C. Hazirbas, L. Ma, C. Domokos, and D. Cremers. Fusetnet: Incorporating depth into semantic segmentation via fusionbased cnn architecture. In Proc. ACCV, volume 2, 2016.
- [0080] [문헌 28] G. Lin, C. Shen, A. van den Hengel, and I. Reid. Efficient piecewise training of deep structured models for semantic segmentation. In Proc. CVPR, pages 3194-3203, 2016
- [0082] 도 6은 의미론적 분할의 정성적인 결과를 보여준다.
- [0083] 각 영상별로 왼쪽부터 차례대로 입력 영상, 실제 라벨, 컬러 영상만을 이용한 RefineNet의 결과, 본 발명의 결과이다. 물체의 분류에 따라 색상을 다르게 표현하였다. 컬러 영상만을 사용했을 때보다 훨씬 정확한 의미론적 분할 결과를 얻었음을 확인할 수 있다.
- [0085] 한편, 상술한 본 발명의 실시예는 개인용 컴퓨터를 포함한 범용 컴퓨터에서 사용되는 매체에 기록될 수 있다. 상기 매체는 마그네틱 기록매체(예를 들면, 롬, 플로피 디스크, 하드 디스크 등), 광학적 판독매체(예를 들면, 씨디롬, 디브이디 등) 및 전기적 기록매체(예를 들면, 플래쉬 메모리, 메모리 스틱 등)와 같은 기록매체를 포함한다.
- [0086] 이제까지 본 발명에 대하여 그 바람직한 실시예들을 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 실시예는 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

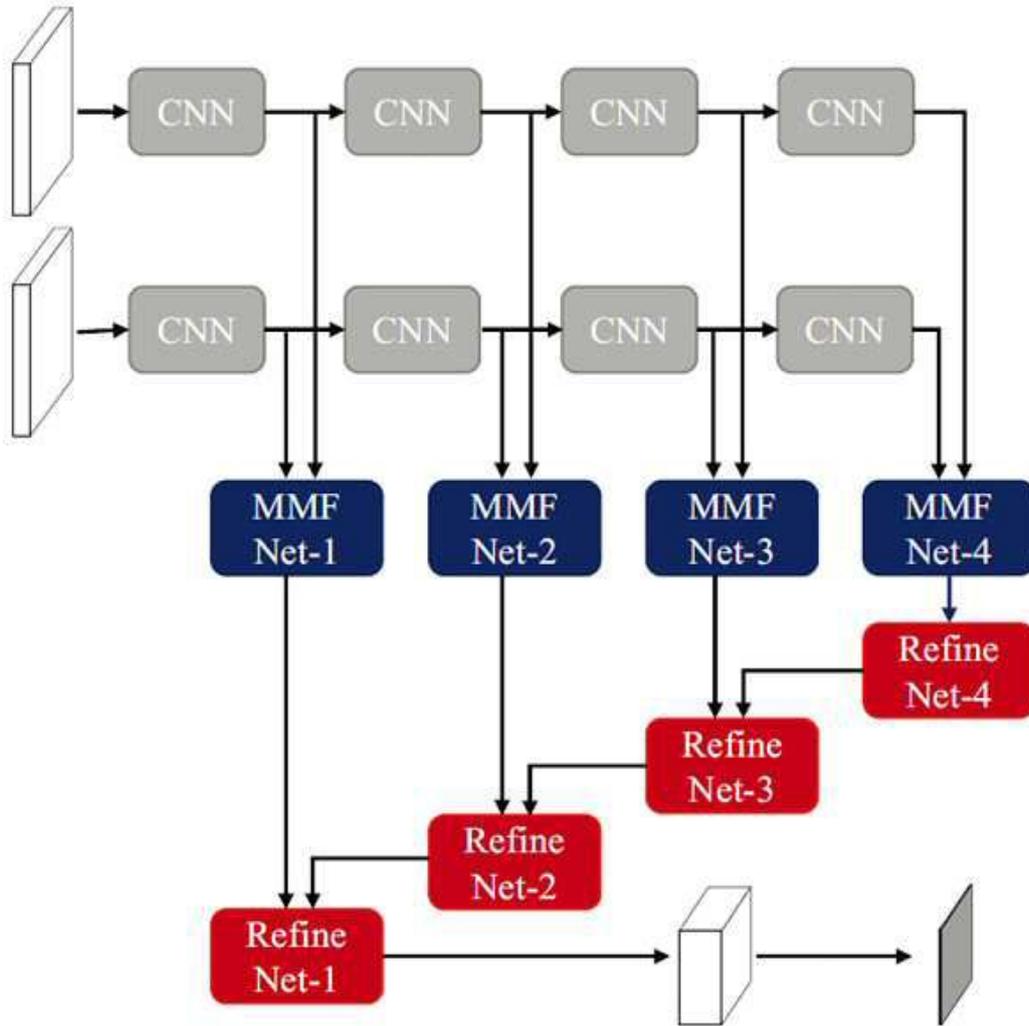
도면1



도면2

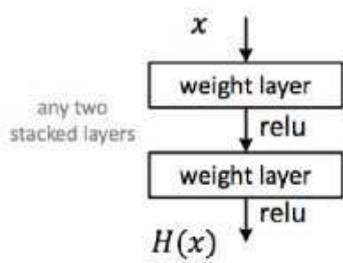


도면3



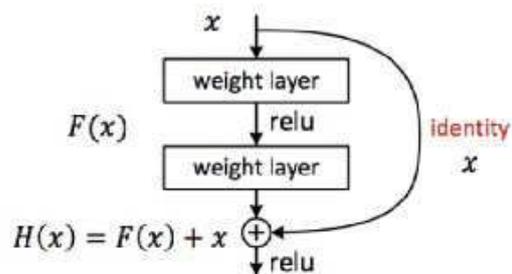
도면4

• Plain net



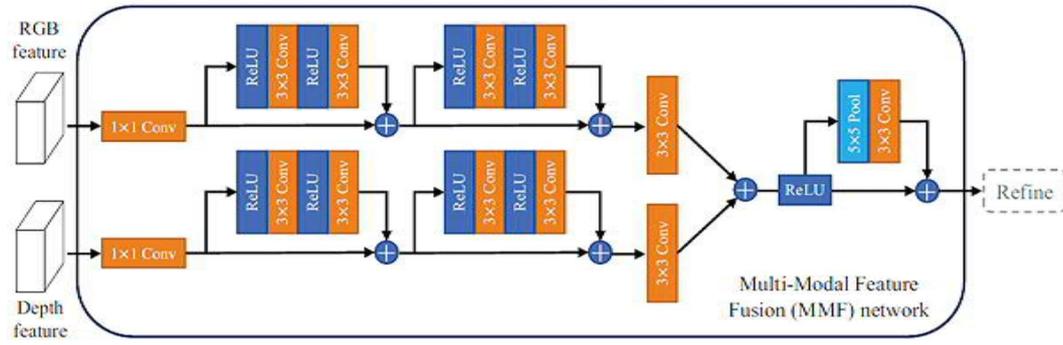
(a)

• Residual net



(b)

도면5



도면6

