

- 1 -

통해 변형 가능한 부분 박스를 생성하는 단계, 생성된 변형 가능한 부분 박스를 이용하여 영역 분해 및 어셈블리 네트워크(Region Decomposition and Assembly Network; RDA-Net)를 통해 이미지에 대한 객체 탐지 및 분할을 수행하는 단계, 객체 탐지 및 분할된 이미지의 객체 부분 영역의 위치를 지정하기 위해 변형 가능한 부분 영역 탐지기를 통해 관심 영역을 분해하여 복수의 박스로 객체 부분 영역을 정의하고 학습하는 단계 및 상기 변형 가능한 부분 영역 네트워크를 다중-스테이지 탐지기에 통합하여 캐스케이드 방식으로 전체 및 부분 탐지를 반복적으로 수행하는 단계를 포함한다.

(52) CPC특허분류

G06T 7/11 (2017.01)

G06T 7/73 (2017.01)

G06V 20/46 (2022.01)

G06T 2207/20081 (2013.01)

G06T 2210/12 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 1711135532
 과제번호 2021R1F1A1049447
 부처명 과학기술정보통신부
 과제관리(전문)기관명 한국연구재단
 연구사업명 기본연구
 연구과제명 [Ezbaro] 고성능 사물 인지 지능을 위한 사물 특징 학습 원천 기술 개발
 기 여 율 10/100
 과제수행기관명 인하대학교 산학협력단
 연구기간 2021.06.01 ~ 2022.05.31

이 발명을 지원한 국가연구개발사업

과제고유번호 1711154594
 과제번호 2022R1C1C1009208
 부처명 과학기술정보통신부
 과제관리(전문)기관명 한국연구재단
 연구사업명 개인기초연구(과기정통부)
 연구과제명 대규모 확장 가능한 On-Vision AI를 위한 E3-AI 네트워크와 협업 학습 연구
 기 여 율 60/100
 과제수행기관명 인하대학교
 연구기간 2022.03.01 ~ 2023.02.28

이 발명을 지원한 국가연구개발사업

과제고유번호 1711160586
 과제번호 2022-0-00448-001
 부처명 과학기술정보통신부
 과제관리(전문)기관명 정보통신기획평가원
 연구사업명 사람중심인공지능핵심원천기술개발
 연구과제명 인간처럼 회상이 가능한 인공 신경망 지속학습 플랫폼 개발
 기 여 율 10/100
 과제수행기관명 인하대학교
 연구기간 2022.04.01 ~ 2022.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호 1711126259
 과제번호 2020-0-01389-002
 부처명 과학기술정보통신부
 과제관리(전문)기관명 정보통신기획평가원
 연구사업명 인공지능융합연구센터지원사업(국고)
 연구과제명 [Ezbaro][정부] 인공지능융합연구센터지원(2차년도)
 기 여 율 10/100
 과제수행기관명 인하대학교 산학협력단
 연구기간 2021.01.01 ~ 2021.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호 1711174118
 과제번호 00155915
 부처명 과학기술정보통신부
 과제관리(전문)기관명 정보통신기획평가원
 연구사업명 인공지능융합혁신인재양성(R&D)
 연구과제명 인공지능융합혁신인재양성
 기 여 율 10/100
 과제수행기관명 인하대학교 산학협력단
 연구기간 2022.07.01 ~ 2022.12.31

명세서

청구범위

청구항 1

입력 이미지의 객체의 기하학적 변환에 따라 분해된 부분 영역을 변형하기 위해 변형 가능한 부분 영역 네트워크(Deformable Part Region Network; DPR-Net)를 통해 변형 가능한 부분 박스를 생성하는 단계;

생성된 변형 가능한 부분 박스를 이용하여 영역 분해 및 어셈블리 네트워크(Region Decomposition and Assembly Network; RDA-Net)를 통해 이미지에 대한 객체 탐지 및 분할을 수행하는 단계;

객체 탐지 및 분할된 이미지의 객체 부분 영역의 위치를 지정하기 위해 변형 가능한 부분 영역 탐지기를 통해 관심 영역을 분해하여 복수의 박스로 객체 부분 영역을 정의하고 학습하는 단계; 및

상기 변형 가능한 부분 영역 네트워크를 다중-스테이지 탐지기에 통합하여 캐스케이드 방식으로 전체 및 부분 탐지를 반복적으로 수행하는 단계

를 포함하는 객체 검출 및 학습 방법.

청구항 2

제1항에 있어서,

상기 입력 이미지의 객체의 기하학적 변환에 따라 분해된 부분 영역을 변형하기 위해 변형 가능한 부분 영역 네트워크를 통해 변형 가능한 부분 박스를 생성하는 단계는,

앵커 박스 및 부분 박스를 특징 맵에 적용하여 객체 특징을 추출하고, 분류 및 회귀하여 객체 클래스 및 위치를 탐지함으로써 영역 제안(Region Proposal)을 생성하는

객체 검출 및 학습 방법.

청구항 3

제2항에 있어서,

상기 객체 탐지 및 분할된 이미지의 객체 부분 영역의 위치를 지정하기 위해 변형 가능한 부분 영역 탐지기를 통해 관심 영역을 분해하여 복수의 박스로 객체 부분 영역을 정의하고 학습하는 단계는,

생성된 영역 제안을 분류 및 회귀하여 객체 클래스 및 위치를 탐지한 후, 좌표가 있는 부분 박스로 구성된 상기 영역 제안을 부분 박스 변환 계층에서 변환 매개변수를 학습하고 부분 박스 생성 계층에서 예측된 출력을 적용하여 각각의 부분 박스를 변환하는

객체 검출 및 학습 방법.

청구항 4

제1항에 있어서,

상기 생성된 변형 가능한 부분 박스를 이용하여 영역 분해 및 어셈블리 네트워크를 통해 이미지에 대한 객체 탐지 및 분할을 수행하는 단계는,

영역 분해 및 어셈블리부를 통해 영역 분해 및 어셈블리 네트워크에 대한 탐지 및 분할을 수행하고, 부분 모델 특징만으로 정제된 특징과 부분 및 전체 객체 특징으로 정제된 특징을 추출하고,

변형 가능한 부분 영역 네트워크의 학습을 위해 부분 손실을 포함하는 적분 손실을 최소화하기 위해 탐지 및 분할을 위한 영역 분해 및 어셈블리 네트워크의 모든 학습 가능한 매개변수는 모델 복잡성을 줄이기 위해 공유되는

객체 검출 및 학습 방법.

청구항 5

제1항에 있어서,

상기 변형 가능한 부분 영역 네트워크를 다중-스테이지 탐지기에 통합하여 캐스케이드 방식으로 전체 및 부분 탐지를 반복적으로 수행하는 단계는,

상기 변형 가능한 부분 영역 네트워크를 다중-스테이지 탐지기에 통합하여 각각의 스테이지에 대해 이전 스테이지의 예측 탐지에 대한 분류 확률, 경계 박스, 마스크 및 마스크 점수를 세분화하고, 영역 분해 및 어셈블리 네트워크 및 복수의 컨볼루션 계층을 이용하여, 변형 가능한 부분 영역 네트워크의 학습을 위해 부분 손실을 포함하는 적분 손실을 평가하고 모든 손실을 최소화하기 위한 다중-스테이지 탐지기의 파라미터를 업데이트하는

객체 검출 및 학습 방법.

청구항 6

입력 이미지의 객체의 기하학적 변환에 따라 분해된 부분 영역을 변형하기 위해 변형 가능한 부분 영역 네트워크(Deformable Part Region Network; DPR-Net)를 통해 변형 가능한 부분 박스를 생성하는 변형 가능한 부분 영역 탐지부; 및

생성된 변형 가능한 부분 박스를 이용하여 영역 분해 및 어셈블리 네트워크(Region Decomposition and Assembly Network; RDA-Net)를 통해 이미지에 대한 객체 탐지 및 분할을 수행하는 영역 분해 및 어셈블리부

를 포함하고,

상기 변형 가능한 부분 영역 탐지부는,

상기 객체 탐지 및 분할된 이미지의 객체 부분 영역의 위치를 지정하기 위해 변형 가능한 부분 영역 탐지기를 통해 관심 영역을 분해하여 복수의 박스로 객체 부분 영역을 정의하고 학습하고,

상기 변형 가능한 부분 영역 네트워크를 다중-스테이지 탐지기에 통합하여 캐스케이드 방식으로 전체 및 부분 탐지를 반복적으로 수행하는

를 포함하는 객체 검출 및 학습 장치.

청구항 7

제6항에 있어서,

상기 변형 가능한 부분 영역 탐지부는,

앵커 박스 및 부분 박스를 특징 맵에 적용하여 객체 특징을 추출하고, 분류 및 회귀하여 객체 클래스 및 위치를 탐지함으로써 영역 제안(Region Proposal)을 생성하고,

생성된 영역 제안을 분류 및 회귀하여 객체 클래스 및 위치를 탐지한 후, 좌표가 있는 부분 박스로 구성된 상기 영역 제안을 부분 박스 변환 계층에서 변환 매개변수를 학습하고 부분 박스 생성 계층에서 예측된 출력을 적용하여 각각의 부분 박스를 변환하는

객체 검출 및 학습 장치.

청구항 8

제6항에 있어서,

상기 영역 분해 및 어셈블리부는,

영역 분해 및 어셈블리부를 통해 영역 분해 및 어셈블리 네트워크에 대한 탐지 및 분할을 수행하고, 부분 모델 특징만으로 정제된 특징과 부분 및 전체 객체 특징으로 정제된 특징을 추출하고,

변형 가능한 부분 영역 네트워크의 학습을 위해 부분 손실을 포함하는 적분 손실을 최소화하기 위해 탐지 및 분할을 위한 영역 분해 및 어셈블리 네트워크의 모든 학습 가능한 매개변수를 모델 복잡성을 줄이기 위해 공유하는

객체 검출 및 학습 장치.

발명의 설명

기술 분야

[0001] 본 발명은 변형 가능한 부분 모델을 기반한 대규모 객체 검출 및 학습 방법 및 장치에 관한 것이다.

배경 기술

[0002] 객체 탐지 또는 인스턴스 분할의 목적은 이미지로부터의 박스 또는 마스크를 사용하여 의미 있는 객체 위치를 확인하는 것이다. 정확하고 빠른 탐지를 위해, 지난 수십 년 동안 강력한 기능[37, 38, 19]과 효과적인 탐지 아키텍처[10, 16, 15]를 개발함으로써 진보를 이루었다. 이 중 컨볼루션 객체 탐지기[35, 34, 32]는 상당한 개선을 보였다. 최근 원-스테이지 탐지기(one-stage detectors) [31, 29, 2, 13] 슬라이딩 앵커(또는 미리 정의된 기본 박스)를 이미지 특징 맵에 설치하여 객체 특징을 추출하고 분류 및 회귀하여 객체 클래스 및 위치를 탐지한다. 슬라이딩 앵커에 의해 추출된 대부분의 특징이 배경이기 때문에 비용 민감 학습(costive-sensitive learning)[28, 23]은 원-스테이지 탐지기의 클래스 불균형을 처리하기 위해 개발되었다.

[0003] 한편, 투-스테이지 탐지기(two-stage detectors)[9, 18, 1]는 먼저 영역 제안 방법[35]을 사용하여 가능한 객체 영역을 찾은 후 다음 스테이지에서 합동 RoI 특징을 분류한다. 보다 정확한 탐지를 위해 다중-스테이지 탐지(multi-stage detection) 방법[4, 5]은 일련의 연속 탐지 헤더를 사용하여 RoI를 반복적으로 개선한다. 더 빠른 탐지를 위해 앵커 프리 탐지기(anchor-free detectors)[23, 33, 40]는 이미지 특징에서 직접 객체 모서리 및 중심에 해당하는 핵심 지점을 추정한다.

[0004] 스케일이 변하지 않는 탐지의 경우 다양한 스케일로 특징 피라미드를 생성하는 다중 스케일 특징 표현이 객체 탐지에 사용된다. 보다 강력한 의미적 특징 맵을 학습하기 위해 하향식 경로[27, 12]에서 전파되는 특징과 측면 연결을 사용하여 상향식 특징을 융합한다. 최근, 특징 피라미드 표현은 다중 스케일 특징 융합[30, 47] 및 교차 스케일 연결[14, 39]과 다중 경로의 특징을 결합함으로써 강화될 수 있다.

[0005] 그러나 이러한 컨볼루션 탐지기는 객체 포즈, 스케일, 관점 및 부분 변형에 의해 야기되는 큰 기하학적 변환과 변형을 처리하도록 여전히 제한된다. 주된 이유는 특징 추출에 사용되는 대부분의 CNN이 종래기술[9]에서 논의된 것처럼 CNN 모듈(즉, 컨볼루션, 풀링 및 RoI 풀링 계층)의 고정 구조를 가지고 있기 때문이다. 결과적으로, 컨볼루션 탐지기를 사용하여 다른 스케일이나 포즈를 가진 비강성(non-rigid) 객체 및 객체를 탐지하는 것은 어렵다. 기하학적 변환의 견고성을 개선하기 위해 분해된 부분 모델을 객체 스케일이나 포즈에 따라 적응적으로 변형할 수 있도록 하는 변형 가능한 부분 영역 네트워크(DPR-Net)를 제안한다. DPR-Net은 변형 가능한 부분 영역 모듈을 도입하며, 객체 스케일 변동에 따라 각 부분 영역의 공간 위치와 스케일을 변환하기 위해 트레이닝 가능한 기하학적 파라미터를 포함하고 있다. 이는 부분의 공간 위치 조정을 위한 파라미터를 학습하는 다른 변형 탐지 방법[17, 42, 9]과 다르다. 또한 부분 모델의 실측 자료(ground truth)를 사용할 수 없기 때문에 DPR-Net을 트레이닝하기 위해 약하게 지도되는 학습을 제시한다. 이를 위해 융합된 부분 특징에서 예측한 점수와 마스크를 상대적인 실측 자료와 비교하여 부분 모델의 분류 및 분할 손실을 평가한다. 이러한 부분 손실을 포함한 필요불가결한 손실을 최소화함으로써, 손실의 경사도(gradient)는 부분 모델의 변환에 더 직접적으로 영향을 미칠 수 있다. 또한 DPR-Net은 별도의 지도 없이 엔드 투 엔드 트레이닝 가능하다.

[0006] 탐지된 박스와 마스크의 품질을 점진적으로 개선하기 위해 캐스케이드 아키텍처를 기반으로 한 변형 가능한 부분 모델 학습을 통해 다중-스테이지 개선을 제시한다. 종래기술[4, 5, 45]와 마찬가지로 캐스케이드 변형 가능한 부분 영역 탐지기(Cascade Deformable Part Region Detector; Cascade D-PRD)는 클래스 또는 위치를 예측하기 위한 일련의 연속 탐지 헤더로 구성된다. 그러나 캐스케이드 D-PRD는 이전 단계에서 개선된 박스를 DPR-Net에 공급하여 전체 및 부분 모델의 이러한 품질을 개선할 수 있다. 그런 다음 개선된 박스를 다시 분해하여 부분 박스를 개선할 수 있다. 그런 다음, 캐스케이드 D-PRD는 개선된 부분과 전체 RoI(Region of Interest)로 더 강력한 의미론적 RoI 특징을 생성하고 이를 연속적인 박스 및 마스크 헤더 트레이닝에 활용할 수 있다.

발명의 내용

해결하려는 과제

[0007] 본 발명이 이루고자 하는 기술적 과제는 분해된 객체 부분을 변환하고 별도의 지도 없이 변환 파라미터를 학습

하기 위한 변형 가능한 부분 모델을 기반한 대규모 객체 검출 및 학습 방법 및 장치를 제공하는데 있다. 변형 가능한 부분 영역 네트워크(Deformable Part Region Network; DPR-Net)를 다중-스테이지 탐지기에 통합하여 변형 가능한 부분 영역 탐지기(Deformable Part Region Detector; D-PRD)를 제안한다. 또한, 전체 및 부분 모델의 박스 및 마스크를 점진적으로 개선하기 위한 캐스케이드 변형 가능한 부분 영역 탐지기(Cascade Deformable Part Region Detector; Cascade D-PRD)를 제안한다.

과제의 해결 수단

- [0008] 일 측면에 있어서, 본 발명에서 제안하는 변형 가능한 부분 모델에 기반한 대규모 객체 검출 및 학습 방법은 입력 이미지의 객체의 기하학적 변환에 따라 분해된 부분 영역을 변형하기 위해 변형 가능한 부분 영역 네트워크(Deformable Part Region Network; DPR-Net)를 통해 변형 가능한 부분 박스를 생성하는 단계, 생성된 변형 가능한 부분 박스를 이용하여 영역 분해 및 어셈블리 네트워크(Region Decomposition and Assembly Network; RDA-Net)를 통해 이미지에 대한 객체 탐지 및 분할을 수행하는 단계, 객체 탐지 및 분할된 이미지의 객체 부분 영역의 위치를 지정하기 위해 변형 가능한 부분 영역 탐지기를 통해 관심 영역을 분해하여 복수의 박스로 객체 부분 영역을 정의하고 학습하는 단계 및 상기 변형 가능한 부분 영역 네트워크를 다중-스테이지 탐지기에 통합하여 캐스케이드 방식으로 전체 및 부분 탐지를 반복적으로 수행하는 단계를 포함한다.
- [0009] 상기 입력 이미지의 객체의 기하학적 변환에 따라 분해된 부분 영역을 변형하기 위해 변형 가능한 부분 영역 네트워크를 통해 변형 가능한 부분 박스를 생성하는 단계는 앵커 박스 및 부분 박스를 특징 맵에 적용하여 객체 특징을 추출하고, 분류 및 회귀하여 객체 클래스 및 위치를 탐지함으로써 영역 제안(Region Proposal)을 생성한다.
- [0010] 상기 객체 탐지 및 분할된 이미지의 객체 부분 영역의 위치를 지정하기 위해 변형 가능한 부분 영역 탐지기를 통해 관심 영역을 분해하여 복수의 박스로 객체 부분 영역을 정의하고 학습하는 단계는 생성된 영역 제안을 분류 및 회귀하여 객체 클래스 및 위치를 탐지한 후, 좌표가 있는 부분 박스로 구성된 상기 영역 제안을 부분 박스 변환 계층에서 변환 매개변수를 학습하고 부분 박스 생성 계층에서 예측된 출력을 적용하여 각각의 부분 박스를 변환한다.
- [0011] 상기 생성된 변형 가능한 부분 박스를 이용하여 영역 분해 및 어셈블리 네트워크를 통해 이미지에 대한 객체 탐지 및 분할을 수행하는 단계는 영역 분해 및 어셈블리부를 통해 영역 분해 및 어셈블리 네트워크에 대한 탐지 및 분할을 수행하고, 부분 모델 특징만으로 정제된 특징과 부분 및 전체 객체 특징으로 정제된 특징을 추출하고, 변형 가능한 부분 영역 네트워크의 학습을 위해 부분 손실을 포함하는 적분 손실을 최소화하기 위해 탐지 및 분할을 위한 영역 분해 및 어셈블리 네트워크의 모든 학습 가능한 매개변수는 모델 복잡성을 줄이기 위해 공유된다.
- [0012] 상기 변형 가능한 부분 영역 네트워크를 다중-스테이지 탐지기에 통합하여 캐스케이드 방식으로 전체 및 부분 탐지를 반복적으로 수행하는 단계는 상기 변형 가능한 부분 영역 네트워크를 다중-스테이지 탐지기에 통합하여 각각의 스테이지에 대해 이전 스테이지의 예측 탐지에 대한 분류 확률, 경계 박스, 마스크 및 마스크 점수를 세분화하고, 영역 분해 및 어셈블리 네트워크 및 복수의 컨볼루션 계층을 이용하며, 변형 가능한 부분 영역 네트워크의 학습을 위해 부분 손실을 포함하는 적분 손실을 평가하고 모든 손실을 최소화하기 위한 다중-스테이지 탐지기의 파라미터를 업데이트한다.
- [0013] 또 다른 일 측면에 있어서, 본 발명에서 제안하는 변형 가능한 부분 모델에 기반한 대규모 객체 검출 및 학습 장치는 입력 이미지의 객체의 기하학적 변환에 따라 분해된 부분 영역을 변형하기 위해 변형 가능한 부분 영역 네트워크(Deformable Part Region Network; DPR-Net)를 통해 변형 가능한 부분 박스를 생성하는 변형 가능한 부분 영역 탐지부 및 생성된 변형 가능한 부분 박스를 이용하여 영역 분해 및 어셈블리 네트워크(Region Decomposition and Assembly Network; RDA-Net)를 통해 이미지에 대한 객체 탐지 및 분할을 수행하는 영역 분해 및 어셈블리부를 포함한다.
- [0014] 본 발명의 실시예에 따른 상기 변형 가능한 부분 영역 탐지부는 상기 객체 탐지 및 분할된 이미지의 객체 부분 영역의 위치를 지정하기 위해 변형 가능한 부분 영역 탐지기를 통해 관심 영역을 분해하여 복수의 박스로 객체 부분 영역을 정의하고 학습하고, 상기 변형 가능한 부분 영역 네트워크를 다중-스테이지 탐지기에 통합하여 캐스케이드 방식으로 전체 및 부분 탐지를 반복적으로 수행한다.

발명의 효과

[0015] 본 발명의 실시예들에 따른 싱글(single) D-PRD는 MSCOCO19에 다른 성능 개선 방법을 사용하지 않고도 우수한 결과를 달성할 수 있다. 별도의 지도 없이 변형 가능한 부분 영역 네트워크를 학습할 수 있으며 각 부분 모델을 객체 스케일 변화에 따라 변형할 수 있다. 또한, 캐스케이드 객체 탐지 및 인스턴스 분할을 개선하기 위해 캐스케이드 변형 부분 영역 아키텍처를 제시하여 캐스케이드 방식으로 전체 및 부분 탐지를 반복적으로 개선할 수 있다. 벨(bells)과 휘슬(whistles)이 없는 캐스케이드 변형 부분 영역 탐지기를 구현하면 최근의 캐스케이드 및 기타 최첨단 탐지기에 비해 COCO 및 VOC 데이터셋에서 더 나은 탐지 및 분할 mAP를 달성할 수 있다.

도면의 간단한 설명

[0016] 도 1은 본 발명의 일 실시예에 따른 변형 가능한 부분 모델에 기반한 대규모 객체 검출 및 학습 장치의 구성을 나타내는 도면이다.

도 2는 본 발명의 일 실시예에 따른 변형 가능한 부분 모델에 기반한 대규모 객체 검출 및 학습 방법을 설명하기 위한 흐름도이다.

도 3은 본 발명의 일 실시예에 따른 변형 가능한 부분 영역 생성 과정을 설명하기 위한 도면이다.

도 4는 본 발명의 일 실시예에 따른 캐스케이드 변형 가능한 부분 영역 검출을 위한 아키텍처를 나타내는 도면이다.

도 5는 본 발명의 일 실시예에 따른 캐스케이드 변형 가능한 부분 영역 검출을 이용한 인스턴스 세분화 결과를 나타내는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0017] 킨볼루션 객체 탐지기에서, 객체의 기하학적 변화나 변환에 의해 야기되는 낮은 특징(feature) 판별성으로 인해 종종 탐지 정확도가 저하될 수 있다. 본 발명에서는 객체의 기하학적 변환에 따라 분해된 부분 영역을 변형할 수 있도록 변형 가능한 부분 영역 학습 방법을 제안한다. 이를 위해 각 부분 모델의 위치에 대해 트레이닝 가능한 기하학적 파라미터를 소개한다. 부분 모델의 실측 자료(ground truth)를 사용할 수 없기 때문에 부분 모델에 대한 분류 및 마스크 손실을 설계하고 이러한 부분 손실을 포함한 필요불가결한 손실을 최소화하여 기하학적 파라미터를 학습한다. 따라서 별도의 지도 없이 변형 가능한 부분 영역 네트워크를 학습할 수 있으며 각 부분 모델을 객체 스케일 변화에 따라 변형할 수 있다. 또한, 캐스케이드 객체 탐지 및 인스턴스 분할을 개선하기 위해 캐스케이드 변형 부분 영역 아키텍처를 제시하여 캐스케이드 방식으로 전체 및 부분 탐지를 반복적으로 개선할 수 있다. 벨(bells)과 휘슬(whistles)이 없는 캐스케이드 변형 부분 영역 탐지기를 구현하면 최근의 캐스케이드 및 기타 최첨단 탐지기에 비해 COCO 및 VOC 데이터셋에서 더 나은 탐지 및 분할 mAP를 달성할 수 있다. 이하, 본 발명의 실시 예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0018] 객체 탐지에는 투-스테이지 탐지기와 원-스테이지 탐지기의 두 가지 주요 스트림이 있다. 투-스테이지 탐지기 [15, 35, 18, 1, 20]에서 관심 영역(Region of Interests; RoIs)이 영역 제안 네트워크(region proposal network)에서 생성되고 객체 클래스와 영역은 후속 R-CNN 또는 그 변형에 RoIs 내에서 추출된 특징을 공급하여 예측된다. 일반적으로 RPN[35]의 사용으로 인해 이러한 투-스테이지 탐지기는 원-스테이지 탐지기보다 정확도는 높지만 속도는 낮다. 한편, 원-스테이지 탐지기[34, 32, 29, 2, 39]는 다중 스케일 특징 맵의 각 픽셀에 대해 미리 정의된 앵커(anchor)를 적용하고 앵커 내에서 추출된 특징으로부터 객체 클래스와 영역을 예측한다.

[0019] 복잡성을 더욱 줄이기 위해, 앵커 프리 탐지기[23, 33, 40, 24]는 클래스별 키 포인트 히트맵(key point heatmap) 내에서 최상위 피크를 찾고 피크를 물체의 중심 위치로 간주한다. 그런 다음 객체 위치와 크기의 예측 오프셋을 사용하여 객체 영역을 회귀하여 결정한다. 또한, 특징 맵을 개선하기 위한 다양한 종래기술들이 있다 [44, 27, 3, 30, 47, 14, 39]. 그 중 특징 피라미드 네트워크(Feature Pyramid Network; FPN)[27]는 상향식 경로와 하향식 경로의 특징을 융합하는 것이 강력한 의미적 특징을 학습하는 데 도움이 된다는 것을 보여준다. 이후, 새로운 경로와 모듈을 추가하여 다중 스케일 기능을 더 개선하는 다양한 방법(예를 들어, PANet [30], M2Det [47], NAS-PFN [14] 및 EfficientDet [39])이 제시되었다.

[0020] 반면에 캐스케이드 아키텍처를 기반으로 하는 다중 스테이지 탐지기는 고품질의 탐지를 위해 설계되었다. 이러한 탐지기는 일련의 탐지 헤더를 사용하여 예측을 점진적으로 조정하여 박스 또는 마스크의 품질을 향상시킨다. 종래기술 중 하나인 캐스케이드 R-CNN[4]은 여러 개의 R-CNN 헤드로 구성된 캐스케이드 방식을 제시하고 후속 헤드에 더 고품질의 탐지를 제공한다. 하이브리드 태스크 캐스케이드 (Hybrid Task Cascade)[5]는 인스턴스 분

할을 더욱 개선하기 위해 이 방식을 확장한다. 캐스케이드 RPN[41]은 반복 RPN을 사용하여 영역 제안 품질을 향상시킨다. 캐스케이드 PSP[7]는 고해상도 이미지 분할을 위한 개선 모델을 제공한다. 사람-객체 상호작용 인식 및 모노클러 3D 휴먼 포즈 추정을 위한 캐스케이드 아키텍처[25, 49]가 제시되어 있다.

[0021] 최근의 많은 탐지기[15, 35, 32, 8, 18, 39]는 고정된 기하학적 구조를 가진 CNN을 기반으로 하기 때문에 본질적으로 기하학적 변환 모델에만 국한된다. 기하학적 변환의 견고성을 향상시키기 위해 종래기술[17]은 ConvNet과 변형 가능한 부분 모델(deformable part model; DPM)을 결합하여 DeepPyramid DPM을 제시한다[11]. ConvNet과 DPM의 공동 트레이닝을 위해 [42]는 NMS 추론을 통합하여 손실 함수를 설계한다. 또한 주어진 이미지에 대해 CNN 내에서 아핀 변환 파라미터를 학습하기 위한 공간 변환 네트워크(STN)[21]가 제시된다. STN의 모델 용량을 줄이기 위해 역 STN[26]은 왜곡된 영상 대신 왜곡된 파라미터를 전파한다. 변형 가능한 CNN[9]에서 공간 오프셋을 증강하고 학습하여 컨볼루션 필터링 및 RoI 풀링의 공간 위치를 조정한다. [43, 22]는 앵커를 다양한 형태로 생성하는 방법을 학습한다. 이러한 최근 연구에서 영감을 받아 분해된 부분 모델의 기하학적 변환을 수용하기 위해 D-PRD를 제시한다. 또한 엔드 투 엔드 학습에서 부분 모델의 회귀 매개변수를 학습할 수 있다. 종래기술[17, 42, 9]와 비교하여 D-PRD는 공간 크기와 부분 모델의 공간 위치를 조정할 수 있다. 탐지 및 분할 정확도를 더욱 향상시키기 위해 일련의 D-PRD 헤드를 기반으로 전체 및 부분 모델을 사용하여 탐지된 박스와 마스크를 개선하기 위한 캐스케이드 D-PRD 아키텍처를 설계한다.

[0022] 본 발명에서는 분해된 객체 부분을 변환하고 별도의 지도 없이 변환 파라미터를 학습하기 위한 변형 가능한 부분 모델을 기반한 대규모 객체 검출 및 학습 방법 및 장치를 제안한다. 변형 가능한 부분 영역 네트워크(Deformable Part Region Network; DPR-Net)를 투-스테이지 탐지기에 통합하여 변형 가능한 부분 영역 탐지기(Deformable Part Region Detector; D-PRD)를 제안한다. 또한, 전체 및 부분 모델의 박스 및 마스크를 점진적으로 개선하기 위한 캐스케이드 변형 가능한 부분 영역 탐지기(Cascade Deformable Part Region Detector; Cascade D-PRD)를 제안한다.

[0023] 본 발명의 실시예에 따르면, 싱글(single) D-PRD는 MSCOCO19에 다른 성능 개선 방법을 사용하지 않고도 우수한 결과를 달성할 수 있다. 또한, 다양한 특징 추출기를 사용하여 D-PRD를 광범위하게 구현하고 D-PRD의 효과를 입증하기 위한 철저한 어블레이션 연구(ablation study)를 제공한다. 또한 제안하는 캐스케이드 D-PRD는 mAP를 D-PRD로 3%까지 개선할 수 있고, 벨(bells)과 휘슬(whistles)이 없는 캐스케이드 D-PRD는 49.2 박스와 42.4 마스크 AP를 제공할 수 있다. 본 발명의 실시예에 따르면, 동일한 백본의 최근 캐스케이드 R-CNN[4](HTC [5])에 비해 박스 및 마스크 AP를 평균 3.4(2.7) 및 2.9(1.6) 포인트 개선한다. 또 다른 이점은 DPR-Net을 간단히 수정하여 기존 앵커 기반 탐지기에 쉽게 적용할 수 있다는 것이다.

[0025] 도 1은 본 발명의 일 실시예에 따른 변형 가능한 부분 모델에 기반한 대규모 객체 검출 및 학습 장치의 구성을 나타내는 도면이다.

[0026] 본 발명의 실시예에 따른 변형 가능한 부분 모델에 기반한 대규모 객체 검출 및 학습 장치의 변형 가능한 부분 영역 탐지기(Deformable Part Region Detector; Cascade; D-PRD)(120)는 특징 추출기(110)로서 특징 피라미드 네트워크(Feature Pyramid Network; FPN)를 사용한다. 이러한 특징 추출기(110)는 실시예일뿐 이에 한정되지 않으며 다양한 특징 추출기가 사용될 수 있다.

[0027] 본 발명의 실시예에 따른 D-PRD(120)는 추출된 특징 맵에 대해 앵커 박스를 적용하여 영역 제안을 생성한다(121). D-PRD(120)는 앵커 박스 및 부분 박스를 특징 맵에 적용하여 객체 특징을 추출하고 분류(Classification) 및 회귀(Regression)하여 객체 클래스 및 위치를 탐지한다. 탐지한 후, 생성된 영역 제안은 분해된 부분 영역으로 생성한다(122). 각각의 영역 제안은 좌표가 있는 부분 박스로 구성된다. 부분 박스 변환 계층에서는 변환 매개변수를 학습하고 부분 박스 생성 계층에서 예측된 출력을 적용하여 각각의 부분 박스를 변환한다.

[0028] 도 1을 참조하면, DPR-Net에서 각각의 영역 제안(Region Proposal)에 대한 k_p 부분 박스는 부분 박스 변환 계층의 출력에 의해 변환된다. 각각의 RDA-net에 대한 탐지를 수행하는 영역 분해 및 어셈블리부(Region Decomposition and Assembly; RDA)(130)는 탐지를 위한 RDA(131) 및 분할을 위한 RDA(132)를 포함하고, 각각의 RDA-net에 대한 검출을 수행하는 영역 분해 및 어셈블리부(130)는 부분 모델 특징만으로 정제된 특징 $\mathbf{x}_3^{comb}$ 과 부분 및 전체 객체 특징으로 정제된 특징 $\mathbf{x}_4^{comb}$ 를 추출한다. D-PRD(110)의 학습을 위해 부분 손실을 포함하는 적분 손실을 구하는 식(3)을 최소화한다. 탐지 및 분할을 위한 RDA-net의 모든 학습 가능한 매개변수는 모델 복

잡성을 줄이기 위해 공유된다. 이와 같은 과정을 통해 탐지 및 분할된 결과 이미지(140)를 생성할 수 있다.

- [0030] 도 2는 본 발명의 일 실시예에 따른 변형 가능한 부분 모델에 기반한 대규모 객체 검출 및 학습 방법을 설명하기 위한 흐름도이다.
- [0031] 제안하는 변형 가능한 부분 모델에 기반한 대규모 객체 검출 및 학습 방법은 입력 이미지의 객체의 기하학적 변환에 따라 분해된 부분 영역을 변형하기 위해 변형 가능한 부분 영역 네트워크(Deformable Part Region Network; DPR-Net)를 통해 변형 가능한 부분 박스를 생성하는 단계(210), 생성된 변형 가능한 부분 박스를 이용하여 영역 분해 및 어셈블리 네트워크(Region Decomposition and Assembly Network; RDA-Net)를 통해 이미지에 대한 객체 탐지 및 분할을 수행하는 단계(220), 객체 탐지 및 분할된 이미지의 객체 부분 영역의 위치를 지정하기 위해 변형 가능한 부분 영역 탐지기를 통해 관심 영역을 분해하여 복수의 박스로 객체 부분 영역을 정의하고 학습하는 단계(230) 및 상기 변형 가능한 부분 영역 네트워크를 다중-스테이지 탐지기에 통합하여 캐스케이드 방식으로 전체 및 부분 탐지를 반복적으로 수행하는 단계(240)를 포함한다.
- [0032] 단계(210)에서, 입력 이미지의 객체의 기하학적 변환에 따라 분해된 부분 영역을 변형하기 위해 변형 가능한 부분 영역 네트워크(DPR-Net)를 통해 변형 가능한 부분 박스를 생성한다.
- [0033] 앵커 박스 및 부분 박스를 특징 맵에 적용하여 객체 특징을 추출하고, 분류 및 회귀하여 객체 클래스 및 위치를 탐지함으로써 영역 제안(Region Proposal)을 생성한다.
- [0034] 단계(220)에서, 생성된 변형 가능한 부분 박스를 이용하여 영역 분해 및 어셈블리 네트워크(RDA-Net)를 통해 이미지에 대한 객체 탐지 및 분할을 수행한다.
- [0035] 영역 분해 및 어셈블리부를 통해 영역 분해 및 어셈블리 네트워크에 대한 탐지 및 분할을 수행하고, 부분 모델 특징만으로 정제된 특징과 부분 및 전체 객체 특징으로 정제된 특징을 추출한다. 이후, 변형 가능한 부분 영역 네트워크의 학습을 위해 부분 손실을 포함하는 적분 손실을 최소화하기 위해 탐지 및 분할을 위한 영역 분해 및 어셈블리 네트워크의 모든 학습 가능한 매개변수는 모델 복잡성을 줄이기 위해 공유된다.
- [0036] 단계(230)에서, 객체 탐지 및 분할된 이미지의 객체 부분 영역의 위치를 지정하기 위해 변형 가능한 부분 영역 탐지기를 통해 관심 영역을 분해하여 복수의 박스로 객체 부분 영역을 정의하고 학습한다.
- [0037] 생성된 영역 제안을 분류 및 회귀하여 객체 클래스 및 위치를 탐지한 후, 좌표가 있는 부분 박스로 구성된 상기 영역 제안을 부분 박스 변환 계층에서 변환 매개변수를 학습하고 부분 박스 생성 계층에서 예측된 출력을 적용하여 각각의 부분 박스를 변환한다.
- [0038] 단계(240)에서, 변형 가능한 부분 영역 네트워크를 다중-스테이지 탐지기에 통합하여 캐스케이드 방식으로 전체 및 부분 탐지를 반복적으로 수행한다.
- [0039] 상기 변형 가능한 부분 영역 네트워크를 다중-스테이지 탐지기에 통합하여 각각의 스테이지에 대해 이전 스테이지의 예측 탐지에 대한 분류 확률, 경계 박스, 마스크 및 마스크 점수를 세분화한다. 이를 위해, 영역 분해 및 어셈블리 네트워크 및 복수의 컨볼루션 계층을 이용하며, 변형 가능한 부분 영역 네트워크의 학습을 위해 부분 손실을 포함하는 적분 손실을 평가하고 모든 손실을 최소화하기 위한 다중-스테이지 탐지기의 파라미터를 업데이트한다. 도 3 내지 도 4를 참조하여 본 발명의 실시예에 따른 변형 가능한 부분 모델에 기반한 대규모 객체 검출 및 학습 과정을 더욱 상세히 설명한다.
- [0041] 다시 도 1을 참조하면, 본 발명의 실시예에 따른 변형 가능한 부분 모델에 기반한 대규모 객체 검출 및 학습 장치는 영역 분해 및 어셈블리부(Region Decomposition and Assembly; RDA)(130)를 통해 RDA-Net에 대한 탐지 및 분할을 수행한다. 먼저, 본 발명의 실시예에 따른 D-PRD 구현을 위해 RDA-Net[1]을 사용하기 때문에 RDA-Net에 대해 설명한다.
- [0043] 도 3은 본 발명의 일 실시예에 따른 변형 가능한 부분 영역 생성 과정을 설명하기 위한 도면이다.
- [0044] 도 3을 참조하면, $H \times W$ 크기의 특징 맵(310)에 대해 k_A 앵커 박스(311)를 적용할 때 $H \times W \times k_A$ 영역 제안이 생성

된다.

- [0045] 앵커 박스(311) 및 부분 박스(312)를 특징 맵(310)에 적용하여 객체 특징을 추출하고 분류(321) 및 회귀(322)하여 객체 클래스 및 위치를 탐지한다.
- [0046] 분류(321) 및 회귀(322)하여 객체 클래스 및 위치를 탐지한 후, 생성된 영역 제안(331)은 분해된 부분 영역(332)으로 생성한다. 각각의 영역 제안(331)은 k_R 좌표가 있는 k_P 부분 박스로 구성된다. 부분 박스 변환 계층에서 $k_P \times k_R$ 변환 매개변수를 학습(323)하고 부분 박스 생성 계층에서 예측된 출력을 적용하여 각각의 부분 박스를 변환한다(333).
- [0047] 예를 들어, $d = (x, y, w, h)$ 를 경계 박스라고 하자. 여기서 x, y, w, h 는 중심 위치, 너비 및 높이이다. 그런 다음 d 를 도 3과 같이 여러 부분 영역으로 나누어 더 작게 분해된 영역 $\{d^p | p \in \{\text{left, right, bottom, upper}\}\}$, 다시 말해 왼쪽, 오른쪽, 아래쪽, 위쪽 박스로 생성할 수 있다. 강력한 의미적 특징을 생성하기 위해 RDA-Net은 서로 다른 지역의 RoI 특징에 대해 다중 스테이지 개선을 수행한다. 스테이지 $l (> 1)$ 에서는 서로 다른 영역의 특징을 비교하고 보다 강력한 의미론적 특징 x_l 만 남는다. 구체적으로, RoIAlign[18] 이후 박스 크기를 고려하여 다중 스케일 피라미드 특징(예를 들어, FPN [27], PANet [30], BiFPN [39])의 해당 스케일(또는 피라미드) 수준에서 d 에 대해 뒤틀린(warped) 특징 x^{whole} 을 추출할 수 있다. 마찬가지로 각 부분 박스 $\{d^p | p \in \{\text{left, right, bottom, upper}\}\}$ 에 대해 동일한 크기의 뒤틀린 부분 특징 맵 $\{x_l^p | p \in \{\text{left, right, bottom, upper}\}\}$ 을 추출할 수 있다.
- [0048] 각각의 l 스테이지에서 서로 다른 영역의 양방향 특징 x_l^p 과 x_l^q 을 병합하여 도 1과 같이 각 채널에 대해 4개의 연속 3x3 컨볼루션 필터, ReLU 함수 및 하나의 요소 별 최대 단위로 구성된 영역 어셈블리 블록(Region Assembly Block; RAB)을 사용하여 보다 강력한 특징 x_l^r 을 생성한다. 결과적으로 결합된 특징 x_l^r 은 다음과 같이 입력 특징 x_l^p 과 x_l^q 에 대한 RAB의 출력이다.
- $$x_l^r = \text{RAB}(x_l^p, x_l^q) \quad (1)$$
- [0049]
- [0050] p, q, r 은 또한 각각 부분(왼쪽, 오른쪽, 아래쪽, 위쪽) 또는 결합된 부분(왼쪽-오른쪽(l/r), 아래쪽-위쪽(b/u), 혼합)을 나타낸다. 반면에 x^{whole} 은 여러 컨볼루션 계층을 통과하며 x_3^{comb} 과 비교하여 x_4^{comb} 를 생성하며, 마지막으로 개선된 특징 x_4^{comb} 는 박스와 마스크 헤드의 입력으로 사용된다. x^{whole} , x_l^p 및 x_l^r 의 해상도는 모든 스테이지에서 동일한 크기이다.
- [0052] 다시 도 1을 참조하면, 본발명의 실시예에 따른 변형 가능한 부분 모델에 기반한 대규모 객체 검출 및 학습 장치의 변형 가능한 부분 영역 탐지기(Deformable Part Region Detector; Cascade; D-PRD)(120)는 기하학적 변환에 대한 컨볼루션 탐지기의 견고성을 개선할 수 있다.
- [0053] 도 1의 D-PRD(120)는 분해된 부분 박스의 공간 위치와 크기를 변환하기 위해 변형 가능한 부분 회귀 계층(121)을 도입한다. 또한 부분 모델에 대한 분류 및 분할 손실은 부분 영역의 결합된 특징의 예측 출력(즉, 클래스 레이블과 마스크)과 그 실측 자료 사이의 차이를 최소화하도록 설계된다. 부분 모델의 출력 특성은 변형 가능한 부분 계층의 변환 파라미터의 영향을 받기 때문에, 분해된 부분 모델의 파라미터는 이러한 손실을 줄임으로써 학습된다.
- [0054] 본 발명의 실시예에 따른 변형 가능한 부분 영역 네트워크(DPR-Net)에서, D-PRD(120)는 크기 $H \times W$ 의 특징맵이 주어지면 위치별로 k_A 앵커(또는 기준 박스)를 적용한다. 앞설 설명된 바와 같이, 각각의 앵커 $d = (x, y, w, h)$ 를 k_P 부분(예를 들어, $k_P = 4$)으로 분해할 수 있다. 따라서 HWk_A 앵커와 HWk_Ak_P 부분 박스가 모두 존재한다. NMS 또는 점수 임계값에 의해 앵커가 제거될 때 앵커의 k_P 부분 박스도 제거된다고 가정한다. 이 가정은 부분 박스당 분류 점수(또는 객체성 점수)를 예측하기 위한 추가 트레이닝을 방지하고 탐지 속도를 향상시킨다. 또한 각 부

분 박스를 독립적으로 변환할 수 있다고 가정한다. 이 경우 부분 박스의 좌표를 인코딩하려면 $k_R = 4$ 개의 파라미터가 필요하다. 따라서 부분 박스 변환 계층은 각 위치에 대한 $k_A \times k_P$ 부분 박스의 k_R 좌표를 변환하기 위해 $k_A \times k_R \times k_P$ 출력을 갖는다.

[0055] $k_A \times k_R \times k_P$ 출력을 $k_A \times k_P$ 부분 박스의 좌표 오프셋으로 사용한다. 가장 낮은 부분부터 가장 높은 부분까지, 이 좌표는 도 3과 같이 왼쪽, 오른쪽, 위쪽, 아래쪽 부분 박스의 (x, y, w, h) 좌표에 해당한다. 부분 박스 변환을 위해, 본 발명에서는 박스 회귀의 역 파라미터화를 사용하여 각 부분 박스의 좌표에 대해 오프셋 $\{\Delta x^p, \Delta y^p, \Delta w^p, \Delta h^p\}$ 을 다음과 같이 적용한다.

$$\begin{aligned} \hat{x}^p &= x + \Delta x^p \cdot w^p, & \hat{y}^p &= y + \Delta y^p \cdot h^p, \\ \hat{w}^p &= \exp(\Delta w^p) \cdot w^p, & \hat{h}^p &= \exp(\Delta h^p) \cdot h^p, \end{aligned} \quad (2)$$

[0057] 여기서 $\hat{x}$ 는 변환된 박스이고 p 는 왼쪽, 오른쪽, 위쪽, 아래쪽 부분을 나타낸다. 이미지 경계에 부분 박스를 클립하여 경계 내에 배치한다. 또한, $\hat{d}^p$ 가 $\hat{d}^p$ 에 대한 낮은 오버랩 비율을 가질 때, $\hat{d}^p$ 를 사용하는 대신 d^p 를 부분 박스로 사용한다. 이렇게 하면 변환된 부분 박스가 객체 영역에서 너무 많이 벗어나는 것을 방지할 수 있다. 앞서 설명된 바와 같이, 도 3은 변형 가능한 부분 박스의 생성 프로세스를 보여준다.

[0058] 본 발명의 실시예에 따른 RDA-Nets에서, RDA(130)는 강력한 특징 응답을 조합하여 추출할 수 있다. 본 발명에서는 $\mathbf{x}_3^{comb}$ 와 $\mathbf{x}_4^{comb}$ 를 각각 박스와 마스크 헤드에 공급한다. 여기서 $\mathbf{x}_3^{comb}$ 은 부분 특징만 병합하여 개선한 특징이며, $\mathbf{x}_4^{comb}$ 는 전체 객체 모델의 $\mathbf{x}_3^{comb}$ 과 $\mathbf{x}^{whole}$ 을 혼합하여 개선한다. 본 발명의 실시예에 따른 박스 헤드에는 1024개의 뉴런이 있는 두 개의 완전 연결(Fully Connect; FC) 계층이 있다. $\mathbf{x}_4^{comb}$ 에 대한 마지막 FC 계층의 출력은 $\text{cls} + 1$ 뉴런과 $4 \times \text{cls}$ 뉴런으로 객체 분류 및 박스 회귀 계층에 연결된다. 여기서 cls 는 객체 클래스 수이고 배경 클래스로 인해 하나가 추가된다. 반면에 $\mathbf{x}_3^{comb}$ 의 FC 출력은 트레이닝 중에만 동일한 분류 계층에 연결된다.

[0059] 인스턴스 분할에 대해 RDA-Net의 출력 $\mathbf{x}_3^{comb}$ 과 $\mathbf{x}_4^{comb}$ 를 마스크 헤드로 결합한다. 픽셀 단위 라벨링이 필요하므로, 유연성과 견고성 및 트레이닝 및 추론의 빠른 속도 때문에 일반적으로 마스크 헤드로 완전 컨볼루션 네트워크(Fully Convolutional network; FCN)[36] 스택을 사용한다. 본 발명에서는 마스크 헤드에 대해 [18]의 구현을 따른다.

[0060] 본 발명의 실시예에 따르면, 연속 3×3 컨볼루션 4개의 스택과 입력의 공간 해상도를 업샘플링하기 위한 스트라이드 2가 있는 2×2 디컨볼루션 계층을 가진다. 그런 다음, cls 마스크를 생성하기 위해 1×1 컨볼루션이 이어진다. 은닉 계층에는 ReLU를 사용한다. 또한 입력 $\mathbf{x}_3^{comb}$ 에서 cls 마스크는 동일한 마스크 헤드를 사용하여 추출된다. 부분 특징에 대한 이러한 마스크는 트레이닝 중에만 사용된다. 또한 예측 마스크와 상대 실측 자료 사이의 픽셀 수준 IoU를 계산하여 각 마스크의 신뢰도를 더 정확하게 평가할 수 있기 때문에 마스크 헤드 뒤에 MaskIoU 헤드를 부착한다. [20]을 구현한 후 마스크 헤드의 입력 RoI 특징과 출력 마스크를 채널 지수를 따라 연결한 다음 연결된 특징에서 클래스당 MaskIoU 점수를 예측하기 위해 4개의 컨볼루션 및 2개의 완전 연결 계층을 따른다.

[0061] 본 발명의 실시예에 따른 D-PRD(120)는 객체 부분의 실측 자료없이 객체 부분 영역을 위치 지정해야 하기 때문에 준지도 문제로 간주할 수 있다. 이 문제를 해결하기 위해 RoI 영역을 분해하여 4개의 직사각형 박스로 객체 부분을 정의한다. 그런 다음 기준 부분 박스로 사용한다. 즉, 부분 박스의 공간 위치와 스케일은 기준 박스에 비례하여 변환된다. 대신, 이러한 분해 박스는 복잡한 준감독 문제를 해결할 때 이점을 제공한다.

[0062] 부분 박스의 변환 파라미터는 손실을 줄이기 위해, 기존 박스 회귀 손실[16]을 직접 적용할 수 없다. 이는 부분 영역의 실측 자료를 사용할 수 없기 때문에 실측 자료와 예측 박스 사이의 불일치를 최소화한다. 그러나 부분 박스에서 추출된 특징은 여전히 객체 분류 및 분할에 기여한다. 따라서 $\hat{d}^p$ 에 관하여 박스 분류 및 마스크 분할

손실을 최소화하여 준지도 문제를 해결할 수 있다. 먼저 IoU를 평가하여 각 박스 d와 실측 자료 박스 $\mathbf{d}^*$ 를 일치시키고 d가 임의의 $\mathbf{d}^*$ 에 비해 0.5 이상 IoU를 가진 경우 d를 양의 레이블 $o^* \in \{1...cls\}$ 에 할당한다. 0.1과 0.5 사이의 IoU를 가진 d에 음의 레이블($o^* = 0$)을 할당한다. $\mathbf{p} = (p^0, ..., p^{cls})$ 및 $\mathbf{p}^{prt} = (p^{prt,0}, ..., p^{prt,cls})$ 는 박스 헤드에 $\mathbf{x}_3^{comb}$ 및 $\mathbf{x}_4^{comb}$ 를 공급하고 헤드의 출력에 소프트맥스를 적용하여 계산되는 cls + 1에 대한 확률 분포를 나타낸다. 그런 다음 변형 가능한 부분 및 전체 모델 $L_{cls}^{prt}(\mathbf{p}^{prt}, \mathbf{o}^*)$ 와 $L_{cls}(\mathbf{p}, \mathbf{o}^*)$ 의 분류 손실을 정의하고 이러한 손실은 교차 엔트로피 손실을 사용하여 등급 확률 예측과 실측 자료 레이블 간의 차이를 평가한다. 메모리 부담을 줄이기 위해 p와 $\mathbf{p}^{prt}$ 계산에 사용되는 박스 헤드의 파라미터 또한 공유된다.

[0063]

또한 다중 작업 손실을 위해 마스크 손실 $L_{mask}(\mathbf{m}, \mathbf{m}^*)$ 및 $L_{mask}^{prt}(\mathbf{m}^{prt}, \mathbf{m}^*)$ 를 추가한다. 이들은 m 및 $\mathbf{m}^{prt}$ 를 실측 자료 마스크 $\mathbf{m}^*$ 과 비교한다. 여기서 $\mathbf{m}^{prt}$ 과 m은 입력 $\mathbf{x}_3^{comb}$ 및 $\mathbf{x}_4^{comb}$ 에 대한 공유 마스크 헤드의 출력을 각각 표시한다. 부분 및 전체 모델에 대한 MaskIoU 손실 L_{miou} 를 평가하기 위해, 이진 마스크와 그 대응 실측 자료 사이의 MaskIoU를 계산한 다음 MaskIoU 대상 $\mathbf{s}^*$ 로 간주한다. L_2 손실을 계산하여 전체 s 및 부분 $\mathbf{s}^{prt}$ 로부터 MaskIoU 대상을 통해 예측 MaskIoU 모델을 회귀시킨다. 따라서 다음과 같이 모든 손실을 전체 및 부분 모델과 결합하여 새롭게 필요불가결한 총 손실을 제시한다.

$$\begin{aligned} L(\mathbf{p}, \mathbf{p}^{prt}, \mathbf{o}^*, \mathbf{t}, \mathbf{t}^*, \mathbf{m}, \mathbf{m}^{prt}, \mathbf{m}^*, \mathbf{s}, \mathbf{s}^{prt}, \mathbf{s}^*) \\ = L_{cls}(\mathbf{p}, \mathbf{o}^*) + \lambda[o \geq 1] L_{reg}(\mathbf{t}, \mathbf{t}^*) + L_{mask}(\mathbf{m}, \mathbf{m}^*) \\ + L_{miou}(\mathbf{s}, \mathbf{s}^*) + L_{cls}^{prt}(\mathbf{p}^{prt}, \mathbf{o}^*) \\ + L_{mask}^{prt}(\mathbf{m}^{prt}, \mathbf{m}^*) + L_{miou}^{prt}(\mathbf{s}^{prt}, \mathbf{s}^*) \end{aligned} \quad (3)$$

[0064]

[0065]

L_{mask} 및 L_{mask}^{prt} 는 평균 이진 교차 엔트로피로 정의된 마스크 손실이다. 마스크 헤드는 cls 클래스에 대해 해상도 $h_{roi}^{mask} \times w_{roi}^{mask}$ 의 $\{\mathbf{m}^1, ..., \mathbf{m}^{cls}\}$ 및 $\mathbf{m}^{prt} = \{\mathbf{m}^{prt,1}, ..., \mathbf{m}^{prt,cls}\}$ 를 생성한다. L_{mask} 및 L_{mask}^{prt} 를 평가할 때 $\mathbf{o}^*$ 와 연관된 RoI에서 예측한 마스크를 실측 자료 마스크 $\mathbf{m}^*$ 과만 비교한다. 또한 MaskIoU 헤드는 cls 클래스에 대해 $\mathbf{s} = (s^0, ..., s^{cls})$ 및 $\mathbf{s}^{prt} = (s^{prt,0}, ..., s^{prt,cls})$ 를 생성한다. $p^{prt,u}$ 는 클래스 u에 대해 예측된 클래스 확률이다. L_{reg} 의 경우 예측 박스 d를 클래스 $\mathbf{o}^*$ 에 대한 앵커 및 실측 자료 박스와 비교하여 박스 회귀 목표 t 및 $\mathbf{t}^*$ 를 평가한다. 추론 중에 부분 모델의 분류, 마스크 및 MaskIoU 예측은 활용되지 않는다. 마스크 점수를 세분화하기 위해 예측 MaskIoU 점수에 동일한 클래스 마스크에 대한 분류 점수를 곱한다. 우리의 구현에서 $\lambda = 1$ 이다.

[0067]

도 4는 본 발명의 일 실시예에 따른 캐스케이드 변형 가능한 부분 영역 검출을 위한 아키텍처를 나타내는 도면이다.

[0068]

도 4를 참조하면, 여기서 "C"는 분류, "B"는 상자, "S"는 마스크, "M"은 마스크 채점 헤드를 의미한다. "CP", "SP" 및 "MP"는 각각 "C", "S" 및 "M"과 공유된다. 하지만, 각각의 캐스케이드 단계 n, 다시 말해 스테이지-1(430), 스테이지-2(440) 및 스테이지-3(450)에서 전체 $\mathbf{x}_{n,4}^{comb}$ 및 일부 특징 $\mathbf{x}_{n,3}^{comb}$ 을 원본("C", "B", "S" 및 "M")과 부분("CP", "SP" 및 "MP")에 공급한다.

[0069]

본 발명의 실시예에 따르면 특징 맵(410)을 이용하여 객체 클래스 및 위치를 탐지한 후 변형 가능한 부분 영역

네트워크(Deformable Part Region Network; DPR-Net)(420)를 다중-스테이지 탐지기에 통합한다.

- [0070] 본 발명의 실시예에 따른 캐스케이드 D-PRD는 각 캐스케이드 단계 n 에 대해 이전 D-PRD의 예측 탐지 $\mathbf{d}_n (\mathbf{d}_n^p)$ 에 대한 분류 확률 $\mathbf{P}_n (\mathbf{P}_n^{prt})$, 경계 박스 $\mathbf{d}_n$, 마스크 $\mathbf{m}_n (\mathbf{m}_n^{prt})$, MaskIoU 점수 $\mathbf{s}_n (\mathbf{s}_n^{prt})$ 를 세분화한다. 스테이지-1(430) 및 스테이지-3(450)에서는 서로 다른 영역 특징을 융합하기 위해 RDA 네트워크를 사용하고, 도 3에 나타난 것과 같이 결합된 $\mathbf{x}_{n,3}^{comb}$ 및 $\mathbf{x}_{n,4}^{comb}$ 를 사용한다. 스테이지-2(430)에서는 여러 컨볼루션 계층을 사용한다. 각 스테이지에서 사용되는 RDA와 다른 헤드 네트워크의 트레이닝 가능한 파라미터는 공유되지 않는다.
- [0071] 본 발명에서는 스테이지-1(430)부터 스테이지-3(450)까지 IOU 임계값을 (0.5, 0.6, 0.7)로 설정하고 다음 박스와 마스크 헤드를 권장하여 더 높은 품질의 결과를 도출한다. 처음 두 스테이지의 출력은 DPR 네트워크에 공급된다. 필요불가결한 손실 식(3)을 사용하여 스테이지 당 손실을 평가하고 모든 손실을 최소화하여 캐스케이드 탐지기의 파라미터를 업데이트한다.
- [0073] 도 5는 본 발명의 일 실시예에 따른 캐스케이드 변형 가능한 부분 영역 검출을 이용한 인스턴스 세분화 결과를 나타내는 도면이다.
- [0074] 도 5를 참조하면, X101-FPN과 함께 캐스케이드 D-PRD를 사용한 인스턴스 세분화 결과를 도시하였다. 경계 상자와 마스크를 같은 색으로 표시하였고 단, 변형 가능한 부분 박스는 주황색으로 표시하였다.
- [0075] 객체 특징 판별성을 개선하기 위해 분명한 방법은 객체 주변의 글로벌 컨텍스트와 객체 내의 로컬 세부 정보를 학습하는 것이다. 그러나 주요 과제는 이러한 컨텍스트를 학습할 의미 있는 영역을 찾는 것이다. 의미 영역은 상호 작용하는 다른 객체나 배경을 포함하여 더 큰 영역일 수도 있고, 로컬 컨텍스트에 중요한 객체를 포함하는 더 작은 영역일 수도 있다. 글로벌 컨텍스트를 학습하기 위해 종래기술[3, 6, 43]은 전체적인 이미지와 객체 RoI 특징을 통합한다. 그러나 이러한 방법은 글로벌 특징 융합 후 객체 부분 세부 정보를 놓치는 경우가 많다. 반면에 종래기술[1, 42, 46, 48]은 로컬 컨텍스트를 학습할 수 있지만, 부분 모델이 객체 내에서 고정되거나 제한된 변형성이기 때문에 글로벌 컨텍스트를 학습하기가 어렵다. 이러한 작업과 비교하여 DPM은 객체 내부 또는 주변 부분 모델의 높은 변형성으로 인해 두 가지 컨텍스트를 모두 학습할 수 있다. 변형성은 부분 박스의 레이블[46, 48] 없이 상기 식(3)의 다중 작업 학습 동안 각 객체 클래스와 크기에 대해 조정된다.
- [0076] 도 5는 각 객체의 탐지된 전체 및 부분 경계 박스를 도시하였다. 표시된 것처럼 객체 부분 영역은 객체 범주, 스케일 및 포즈에 따라 변형 가능한 것으로 학습된다. 부분 박스의 변형성은 기하학적 변화에 대한 견고성을 향상시킬 수 있다. 또한 객체 부분 세부 정보와 객체 주변의 글로벌 컨텍스트를 학습하기 위해 추출해야 하는 중요한 영역을 나타낸다.
- [0078] 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, FPA(field programmable 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.
- [0079] 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.

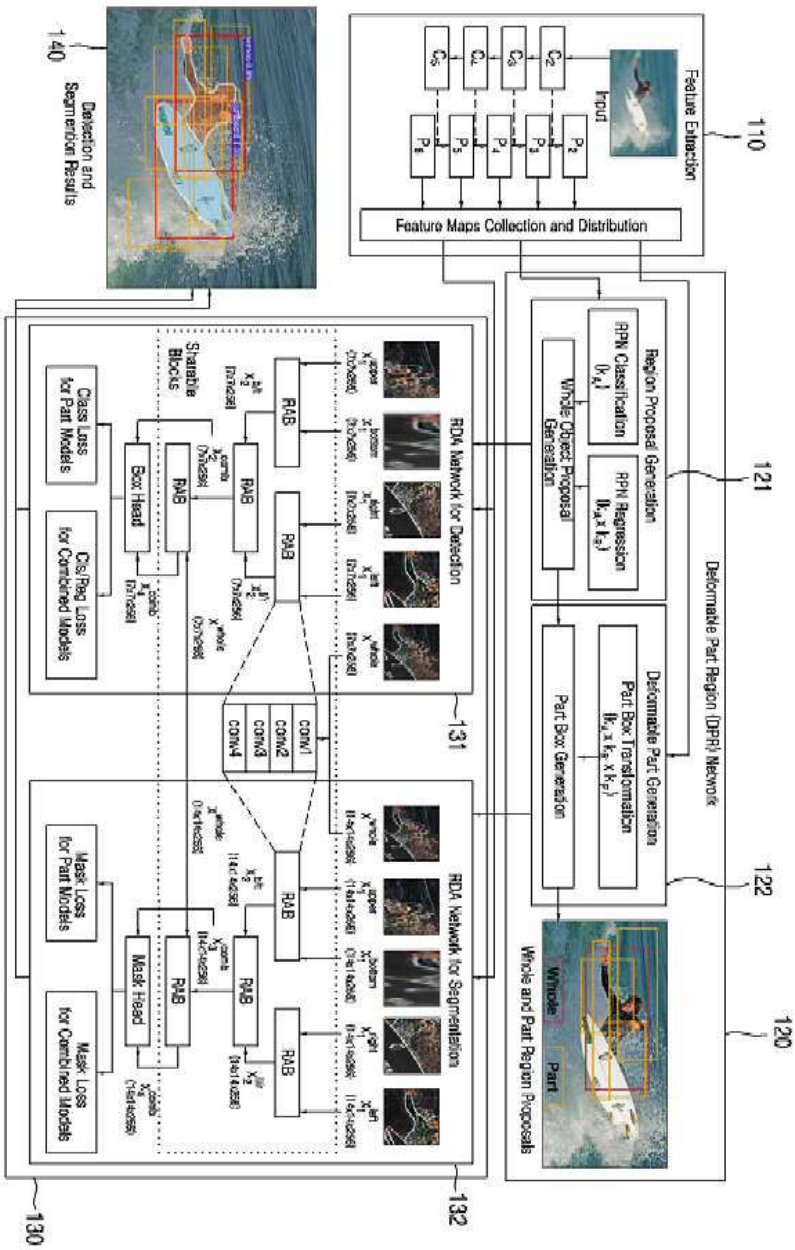
- [0080] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic 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.
- [0081] 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.
- [0082] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.
- [0084] <참고자료>
- [0085] [1] Seung-Hwan Bae. Object detection based on region decomposition and assembly. In AAAI, volume 33, pages 8094- 8101, 2019. 1, 2, 3, 6
- [0086] [2] Daniel Bolya, Chong Zhou, Fanyi Xiao, and Yong Jae Lee. Yolact: Real-time instance segmentation. In ICCV, pages 9157-9166, 2019. 1, 2
- [0087] [3] Zhaowei Cai, Quanfu Fan, Rogerio Schmidt Feris, and Nuno Vasconcelos. A unified multi-scale deep convolutional neural network for fast object detection. In ECCV, pages 354- 370, 2016. 2, 6
- [0088] [4] Zhaowei Cai and Nuno Vasconcelos. Cascade r-cnn: Delving into high quality object detection. In CVPR, June 2018. 1, 2
- [0089] [5] Kai Chen, Jiangmiao Pang, Jiaqi Wang, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jianping Shi, Wanli Ouyang, Chen Change Loy, and Dahua Lin. Hybrid task cascade for instance segmentation. In CVPR, June 2019. 1, 2
- [0090] [6] Zhao-Min Chen, Xin Jin, Borui Zhao, Xiu-Shen Wei, and Yanwen Guo. Hierarchical context embedding for regionbased object detection. ECCV, 2020. 6
- [0091] [7] Ho Kei Cheng, Jihoon Chung, Yu-Wing Tai, and Chi-Keung Tang. Cascadepsp: Toward class-agnostic and very highresolution segmentation via global and local refinement. In CVPR, 2020. 2
- [0092] [8] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun. R-FCN: object detection via region-based fully convolutional networks. In NIPS, pages 1-9, 2016. 2
- [0093] [9] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In ICCV, pages 764-773, 2017. 1, 2
- [0094] [10] Piotr Dollar, Ron Appel, Serge J. Belongie, and Pietro Perona. Fast feature pyramids for object detection. TPAMI, 36(8):1532-1545, 2014. 1

- [0095] [11] Pedro F. Felzenszwalb, Ross B. Girshick, David A. McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. TPAMI, 32(9):1627-1645, 2010. 2
- [0096] [12] Cheng-Yang Fu, Wei Liu, Ananth Ranga, Ambrish Tyagi, and Alexander C. Berg. DSSD : Deconvolutional single shot detector. CoRR, abs/1701.06659, 2017. 1 7
- [0097] [13] Cheng-Yang Fu, Mykhailo Shvets, and Alexander C Berg. Retinamask: Learning to predict masks improves state-of-the-art single-shot detection for free. arXiv preprint arXiv:1901.03353, 2019. 1
- [0098] [14] Golnaz Ghiasi, Tsung-Yi Lin, and Quoc V. Le. Nas-fpn: Learning scalable feature pyramid architecture for object detection. In CVPR, June 2019. 1, 2
- [0099] [15] Ross B. Girshick. Fast R-CNN. In ICCV, pages 1440-1448, 2015. 1, 2
- [0100] [16] Ross B. Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In CVPR, pages 580-587, 2014. 1, 4
- [0101] [17] Ross B. Girshick, Forrest N. Iandola, Trevor Darrell, and Jitendra Malik. Deformable part models are convolutional neural networks. In CVPR, pages 437-446, 2015. 1, 2
- [0102] [18] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross B. Girshick. Mask R-CNN. In ICCV, pages 2980-2988, 2017. 1, 2, 3, 4
- [0103] [19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In CVPR, pages 770-778, 2016. 1
- [0104] [20] Zhaojin Huang, Lichao Huang, Yongchao Gong, Chang Huang, and Xinggang Wang. Mask Scoring R-CNN. In CVPR, 2019. 2, 4
- [0105] [21] Max Jaderberg, Karen Simonyan, Andrew Zisserman, and Koray Kavukcuoglu. Spatial transformer networks. In NIPS, pages 2017-2025, 2015. 2
- [0106] [22] Wei Ke, Tianliang Zhang, Zeyi Huang, Qixiang Ye, Jianzhuang Liu, and Dong Huang. Multiple anchor learning for visual object detection. In CVPR, June 2020. 2
- [0107] [23] Hei Law and Jia Deng. Cornernet: Detecting objects as paired keypoints. In ECCV, pages 734-750, 2018. 1, 2
- [0108] [24] Youngwan Lee and Jongyoul Park. Centermask: Real-time anchor-free instance segmentation. In CVPR, pages 13906- 13915, 2020. 2
- [0109] [25] Shichao Li, Lei Ke, Kevin Pratama, Yu-Wing Tai, Chi-Keung Tang, and Kwang-Ting Cheng. Cascaded deep monocular 3d human pose estimation with evolutionary training data. In CVPR, pages 6172-6182, 2020. 2
- [0110] [26] Chen-Hsuan Lin and Simon Lucey. Inverse compositional spatial transformer networks. In CVPR, 2017. 2
- [0111] [27] Tsung-Yi Lin, Piotr Dollar, Ross B. Girshick, Kaiming He, Bharath Hariharan, and Serge J. Belongie. Feature pyramid networks for object detection. In CVPR, pages 936-944, 2017. 1, 2, 3
- [0112] [28] Tsung-Yi Lin, Priya Goyal, Ross B. Girshick, Kaiming He, and Piotr Dollar. Focal loss for dense object detection. In ICCV, pages 2999-3007, 2017. 1
- [0113] [29] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. Focal loss for dense object detection. In ICCV, pages 2980-2988, 2017. 1, 2
- [0114] [30] Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, and Jiaya Jia. Path aggregation network for instance segmentation. In CVPR, June 2018. 1, 2, 3
- [0115] [31] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In ECCV, pages 21-37. Springer, 2016. 1

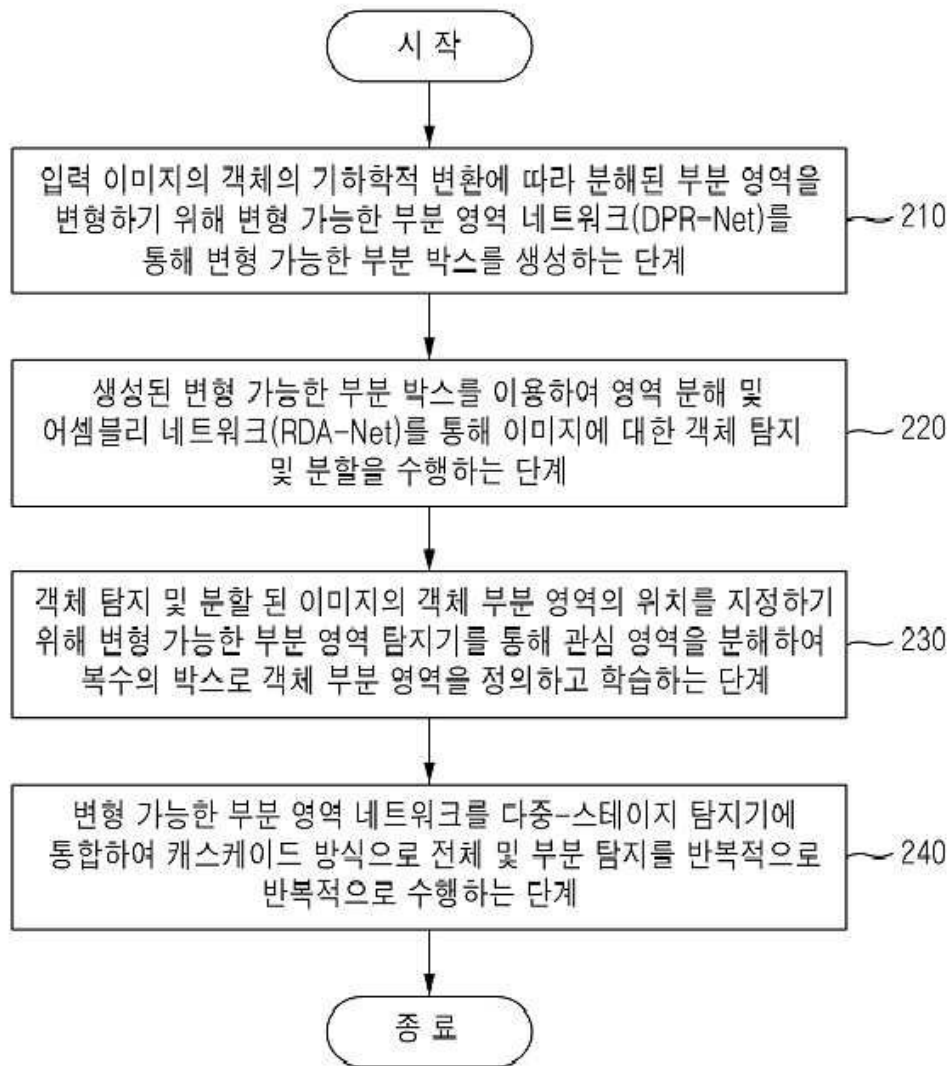
- [0116] [32] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott E. Reed, Cheng-Yang Fu, and Alexander C. Berg. SSD: single shot multibox detector. In ECCV, pages 21-37, 2016. 1, 2
- [0117] [33] Xin Lu, Buyu Li, Yuxin Yue, Quanquan Li, and Junjie Yan. Grid r-cnn. In CVPR, pages 7363-7372, 2019. 1, 2
- [0118] [34] Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In CVPR, pages 779-788, 2016. 1, 2
- [0119] [35] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. Faster R-CNN: towards real-time object detection with region proposal networks. In NIPS, pages 91-99, 2015. 1, 2
- [0120] [36] Evan Shelhamer, Jonathan Long, and Trevor Darrell. Fully convolutional networks for semantic segmentation. TPAMI, 39(4):640-651, 2017. 4
- [0121] [37] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556, 2014. 1
- [0122] [38] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In CVPR, 2015. 1
- [0123] [39] Mingxing Tan, Ruoming Pang, and Quoc V. Le. Efficientdet: Scalable and efficient object detection. In CVPR, June 2020. 1, 2, 3
- [0124] [40] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. Fcos: Fully convolutional one-stage object detection. In ICCV, pages 9627-9636, 2019. 1, 2
- [0125] [41] Thang Vu, Hyunjun Jang, Trung X. Pham, and Chang Dong Yoo. Cascade RPN: delving into high-quality region proposal network with adaptive convolution. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alche-Buc, Emily B. Fox, and Roman Garnett, editors, ' NIPS, pages 1430-1440, 2019. 2
- [0126] [42] Li Wan, David Eigen, and Rob Fergus. End-to-end integration of a convolutional network, deformable parts model and non-maximum suppression. In CVPR, pages 851-859, 2015. 1, 2, 6
- [0127] [43] Tong Yang, Xiangyu Zhang, Zeming Li, Wenqiang Zhang, and Jian Sun. Metaanchor: Learning to detect objects with customized anchors. In NIPS, pages 318-328, 2018. 2, 6
- [0128] [44] S. Zagoruyko, A. Lerer, T.-Y. Lin, P. O. Pinheiro, S. Gross, S. Chintala, and P. Dollar. A multipath network for object ' detection. In BMVC. 2
- [0129] [45] Hongkai Zhang, Hong Chang, Bingpeng Ma, Shiguang Shan, and Xilin Chen. Cascade retinanet: Maintaining consistency for single-stage object detection. In BMVC, page 227, 2019. 2
- [0130] [46] Shifeng Zhang, Longyin Wen, Xiao Bian, Zhen Lei, and Stan Z. Li. Occlusion-aware R-CNN: detecting pedestrians in a crowd. In ECCV, pages 657-674, 2018. 6
- [0131] [47] Qijie Zhao, Tao Sheng, Yongtao Wang, Zhi Tang, Ying Chen, Ling Cai, and Haibing Ling. M2det: A single-shot object detector based on multi-level feature pyramid network. In AAAI, 2019. 1, 2 8
- [0132] [48] Chunluan Zhou and Junsong Yuan. Bi-box regression for pedestrian detection and occlusion estimation. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, ECCV, pages 138-154, 2018. 6
- [0133] [49] Tianfei Zhou, Wenguan Wang, Siyuan Qi, Haibin Ling, and Jianbing Shen. Cascaded human-object interaction recognition. In CVPR, pages 4262-4271. IEEE, 2020. 2

도면

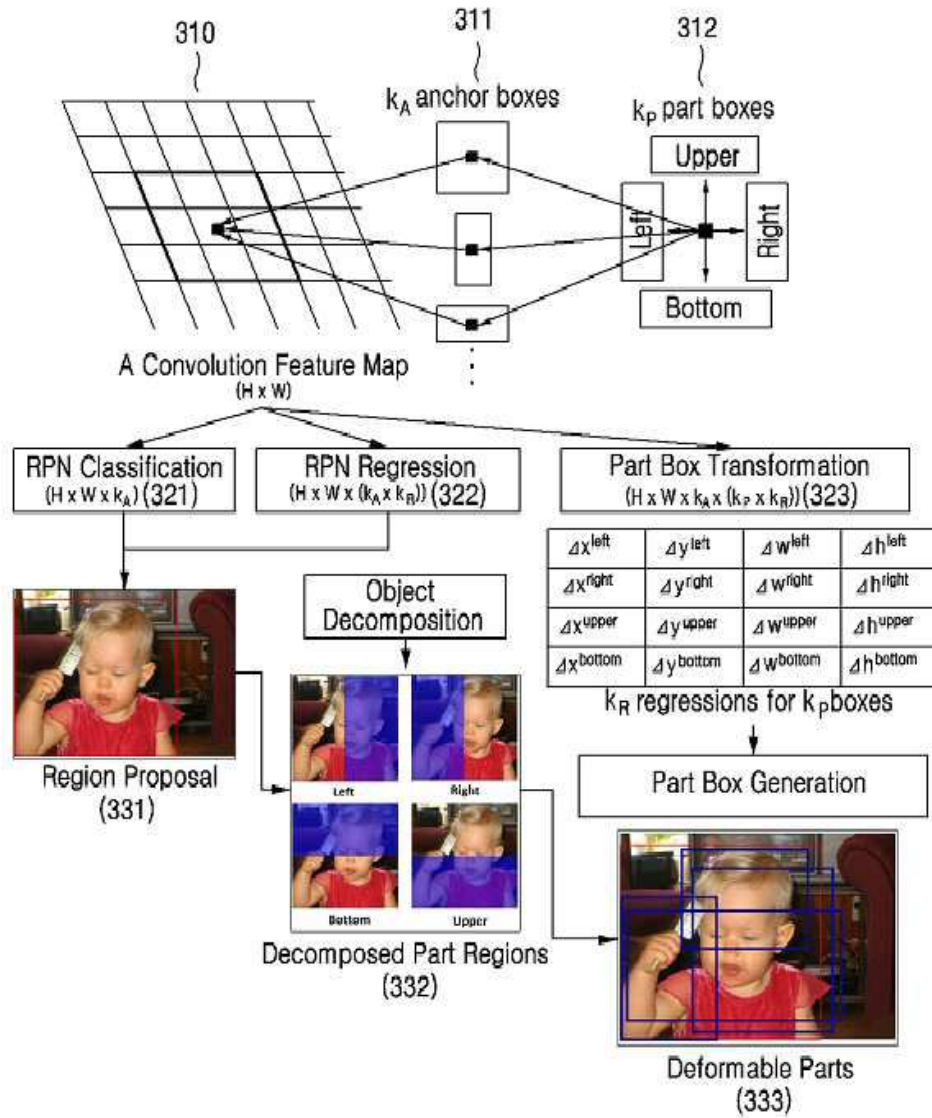
도면1



도면2



도면3



도면4

