



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0048727  
(43) 공개일자 2023년04월12일

(51) 국제특허분류(Int. Cl.)  
G06V 10/40 (2022.01) G06T 3/40 (2006.01)  
(52) CPC특허분류  
G06V 10/40 (2023.01)  
G06T 3/4053 (2013.01)  
(21) 출원번호 10-2021-0131508  
(22) 출원일자 2021년10월05일  
심사청구일자 2021년10월05일

(71) 출원인  
인하대학교 산학협력단  
인천광역시 미추홀구 인하로 100(용현동, 인하대학교)  
(72) 발명자  
배승환  
인천광역시 미추홀구 인하로 100(용현동)  
이성호  
인천광역시 연수구 용담로 14, 103동 1206호(청학동, 연수1차하나라파트)  
(74) 대리인  
양성보

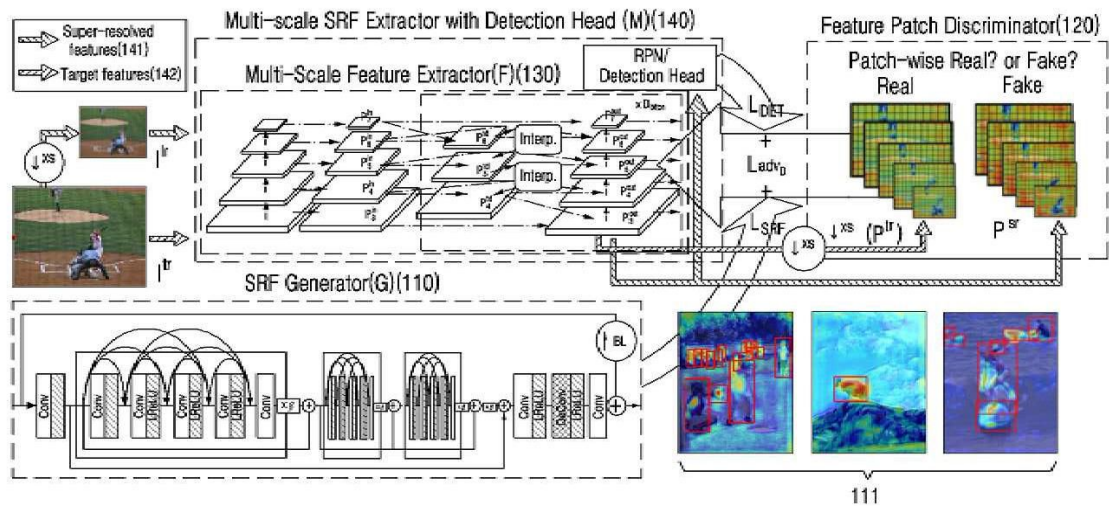
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 적대적 초해상도 다중 스케일 특징맵 학습 및 객체 검출기

(57) 요약

대적 초해상도 다중 스케일 특징맵 학습 및 객체 검출 방법 및 장치가 제시된다. 본 발명에서 제안하는 적대적 초해상도 다중 스케일 특징맵 학습 및 객체 검출 방법은 SRF(Super-Resolved Feature) 생성기를 통해 다중 스케일 특징 학습을 위한 초해상도 특징 맵을 생성하고, 특징 패치 판별기를 통해 상기 초해상도 특징 맵에서 추출된 패치와 대상 특징을 판별하는 단계, 다중 스케일 SRF 특징 추출기를 통해 상기 생성된 초해상도 특징 맵과 검출 정확도를 개선하기 위한 상기 SRF 생성기의 손실을 계산하고, 상기 초해상도 특징 맵에 대한 다중 스케일 특징 맵을 학습하는 단계 및 상기 다중 스케일 추출기의 학습 결과에 따른 다중 스케일 특징을 추출하고, 객체를 검출하는 단계를 포함한다.

대표도



(52) CPC특허분류

G06T 2207/20081 (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

SRF(Super-Resolved Feature) 생성기를 통해 다중 스케일 특징 학습을 위한 초해상도 특징 맵을 생성하고, 특징 패치 판별기를 통해 상기 초해상도 특징 맵에서 추출된 패치와 대상 특징을 판별하는 단계;

다중 스케일 SRF 특징 추출기를 통해 상기 생성된 초해상도 특징 맵과 검출 정확도를 개선하기 위한 상기 SRF 생성기의 손실을 계산하고, 상기 초해상도 특징 맵에 대한 다중 스케일 특징 맵을 학습하는 단계; 및

상기 다중 스케일 추출기의 학습 결과에 따른 다중 스케일 특징을 추출하고, 객체를 검출하는 단계

를 포함하는 다중 스케일 특징맵 학습 및 객체 검출 방법.

#### 청구항 2

제1항에 있어서,

상기 SRF 생성기를 통해 다중 스케일 특징 학습을 위한 초해상도 특징 맵을 생성하고, 특징 패치 판별기를 통해 상기 초해상도 특징 맵에서 추출된 패치와 대상 특징을 판별하는 단계는,

초해상도 특징 맵을 생성하기 위해 SRF 생성기 및 특징 패치 판별기 간의 적대적 훈련을 수행하고, 픽셀 단위 손실과 적대적 손실로 구성된 초해상도 특징을 정의하며, 저해상도 특징과 각 스케일 레벨에서 초해상도의 대응 대상 특징 사이의 불일치를 평가하고, SRF 생성기가 특징 패치 판별기를 속여 초해상도 특징 맵을 생성하도록 하는

다중 스케일 특징맵 학습 및 객체 검출 방법.

#### 청구항 3

제2항에 있어서,

상기 다중 스케일 SRF 특징 추출기를 통해 상기 생성된 초해상도 특징 맵과 검출 정확도를 개선하기 위한 상기 SRF 생성기의 손실을 계산하고, 상기 초해상도 특징 맵에 대한 다중 스케일 특징 맵을 학습하는 단계는,

상기 저해상도 특징을 확장시키기 위해 상기 다중 스케일 SRF 특징 추출기를 통해 추출된 대상 특징을 이용하여 SRF 생성기와 특징 패치 판별기로 구성된 초해상도 특징 생성 적대 네트워크의 적대적 훈련을 수행하고, 사전 훈련된 SRF 생성기로 상기 다중 스케일 SRF 특징 추출기의 모든 보간 모듈을 변경하여 상기 다중 스케일 SRF 특징 추출기를 구축하는

다중 스케일 특징맵 학습 및 객체 검출 방법.

#### 청구항 4

제3항에 있어서,

입력 이미지로부터 추출된 다중 스케일 특징, SRF 생성기와 특징 패치 판별기의 적대적 훈련을 통해 생성된 초해상도 특징 맵을 이용하여 대상 특징 간의 불일치를 평가하여 적대적 최소-최대 손실을 계산하고,

상기 다중 스케일 SRF 특징 추출기를 통해 상기 초해상도 특징 맵에 대한 적대적 훈련을 수행하고, 상기 대상 이미지 및 상기 저해상도 이미지에 대한 다중 스케일 SRF 특징 추출을 수행하여, 대상 특징 패치에서 초해상도 특징 패치를 구별하도록 훈련된 판별기를 속임으로써 각 스케일에서 대상 영상의 특징 분포를 학습하도록 하는

다중 스케일 특징맵 학습 및 객체 검출 방법.

#### 청구항 5

다중 스케일 특징 학습을 위한 초해상도 특징 맵을 생성하는 SRF(Super-Resolved Feature) 생성기;

상기 초해상도 특징 맵에서 추출된 패치와 대상 특징을 판별하는 특징 패치 판별기; 및

상기 생성된 초해상도 특징 맵과 검출 정확도를 개선하기 위한 상기 SRF 생성기의 손실을 계산하고, 상기 초해상도 특징 맵에 대한 다중 스케일 특징 맵을 학습하고, 학습 결과에 따른 다중 스케일 특징을 추출하고, 객체를 검출하는 다중 스케일 SRF 특징 추출기

를 포함하는 다중 스케일 특징맵 학습 및 객체 검출 장치.

## 청구항 6

제5항에 있어서,

상기 SRF 생성기 및 상기 특징 패치 판별기는,

초해상도 특징 맵을 생성하기 위해 SRF 생성기 및 특징 패치 판별기 간의 적대적 훈련을 수행하고, 픽셀 단위 손실과 적대적 손실로 구성된 초해상도 특징을 정의하며, 저해상도 특징과 각 스케일 레벨에서 초해상도의 대응 대상 특징 사이의 불일치를 평가하고, SRF 생성기가 특징 패치 판별기를 속여 초해상도 특징 맵을 생성하도록 하는

다중 스케일 특징맵 학습 및 객체 검출 장치.

## 청구항 7

제6항에 있어서,

상기 다중 스케일 SRF 특징 추출기는,

상기 저해상도 특징을 확장시키기 위해 상기 다중 스케일 SRF 특징 추출기를 통해 추출된 대상 특징을 이용하여 SRF 생성기와 특징 패치 판별기로 구성된 초해상도 특징 생성 적대 네트워크의 적대적 훈련을 수행하고, 사전 훈련된 SRF 생성기로 상기 다중 스케일 SRF 특징 추출기의 모든 보간 모듈을 변경하여 상기 다중 스케일 SRF 특징 추출기를 구축하는

다중 스케일 특징맵 학습 및 객체 검출 장치.

## 청구항 8

제7항에 있어서,

상기 다중 스케일 SRF 특징 추출기는,

상기 다중 스케일 SRF 특징 추출기를 통해 상기 초해상도 특징 맵에 대한 적대적 훈련을 수행하고, 상기 대상 이미지 및 상기 저해상도 이미지에 대한 다중 스케일 SRF 특징 추출을 수행하여, 대상 특징 패치에서 초해상도 특징 패치를 구별하도록 훈련된 판별기를 속임으로써 각 스케일에서 대상 영상의 특징 분포를 학습하도록 하는

다중 스케일 특징맵 학습 및 객체 검출 장치.

## 발명의 설명

### 기술 분야

[0001] 본 발명은 적대적 초해상도 다중 스케일 특징맵 학습 및 객체 검출 방법 및 장치에 관한 것이다.

### 배경 기술

[0002] 심층 컨볼루션 신경망(Convolutional Neural Network; CNN)의 발전으로 인해 컨볼루션 객체 검출기[1]-[3]는 현저한 정확도 향상을 보여주었다. 객체의 스케일 변화에 대한 견고성을 개선하기 위해 최첨단 검출기는 다중 스케일 특징 표현을 기반으로 구성된다. 다중 스케일 객체 검출의 경우, 일부 구조[4]-[6]는 검출기의 기본 네트워크(즉, 백본(backbone))를 위해 설계되고 사용된다. 그 중 특징 피라미드 네트워크(Feature Pyramid Network; FPN)[4]는 하향식 특징 전파를 개발하고 모든 스케일 레벨에서 다중 스케일 특징을 사용할 수 있는 방법을 제공한다. 하위 계층 특징을 높이기 위해 경로 집계 특징 피라미드 네트워크(Path Aggregation Feature Pyramid Network; PAFPN)[5]는 하향식 경로 다음에 추가 상향식 경로를 설계한다.

[0003] 이러한 연구로부터 객체 검출을 위한 여러 가지 다중 스케일 특징 방법[6]-[12] 또한 제시되었다. 구체적으로 [6], [7], [11], [12]는 특징 표현을 개선하기 위한 추가 특징 전파 경로를 설계한다. 또한 검출을 개선하기 위해 다중 스케일 특징을 효과적으로 사용하기 위한 검출 방법 [8]-[10]을 제안한다.

[0004] 다중 스케일 특징 표현을 기반으로 한 작업에서 주요 프로세스는 특징 맵을 다음 스케일 레벨로 전파하기 전에 특징 맵의 크기를 조정하는 것이다. 일반적으로 상향식 및 하향식 특징 맵은 각각 다운샘플링 및 업샘플링된다. 결과적으로, 이전 레벨의 특징 해상도는 동일한 경로의 다음 레벨에서 특징 해상도에 적합할 수 있지만 다른 경로에서 전달된 특징과도 결합될 수 있다. 그러나 특징 해상도를 높일 때 간단한 보간법(예를 들어, 가장 가까운 이웃 및 이중선형)을 여전히 이용한다. 종래기술[13]에 표시된 것처럼 이러한 보간법은 노이즈가 많고 흐릿한 특징 맵을 유발한다. 이러한 특징을 검출기의 입력으로 사용하면 검출 정확도가 저하된다.

## 발명의 내용

### 해결하려는 과제

[0005] 본 발명이 이루고자 하는 기술적 과제는 다중 스케일 특징 학습에 사용되는 초해상도 특징을 생성할 수 있는 새로운 특징 생성기를 이용한 적대적 초해상도 다중 스케일 특징맵 학습 및 객체 검출 방법 및 장치를 제공하는데 있다. 제안하는 SRF 생성기와 특징 패치 판별기로 구성된 초해상도 특징 생성 적대 네트워크(Super-Resolved Feature Generative Adversarial Network; SRF-GAN)를 제안하고, 또한 SRF-GAN을 다중 작업 학습에서 다중 스케일 객체 검출 및 분할에 더 적합하게 만들 수 있는 새로운 통합 손실을 제안한다.

### 과제의 해결 수단

[0006] 일 측면에 있어서, 본 발명에서 제안하는 적대적 초해상도 다중 스케일 특징맵 학습 및 객체 검출 방법은 SRF(Super-Resolved Feature) 생성기를 통해 다중 스케일 특징 학습을 위한 초해상도 특징 맵을 생성하고, 특징 패치 판별기를 통해 상기 초해상도 특징 맵에서 추출된 패치와 대상 특징을 판별하는 단계, 다중 스케일 SRF 특징 추출기를 통해 상기 생성된 초해상도 특징 맵과 검출 정확도를 개선하기 위한 상기 SRF 생성기의 손실을 계산하고, 상기 초해상도 특징 맵에 대한 다중 스케일 특징 맵을 학습하는 단계 및 상기 다중 스케일 추출기의 학습 결과에 따른 다중 스케일 특징을 추출하고, 객체를 검출하는 단계를 포함한다.

[0007] 상기 SRF 생성기를 통해 다중 스케일 특징 학습을 위한 초해상도 특징 맵을 생성하고, 특징 패치 판별기를 통해 상기 초해상도 특징 맵에서 추출된 패치와 대상 특징을 판별하는 단계는 초해상도 특징 맵을 생성하기 위해 SRF 생성기 및 특징 패치 판별기 간의 적대적 훈련을 수행하고, 픽셀 단위 손실과 적대적 손실로 구성된 초해상도 특징을 정의하며, 저해상도 특징과 각 스케일 레벨에서 초해상도의 대응 대상 특징 사이의 불일치를 평가하고, SRF 생성기가 특징 패치 판별기를 속여 초해상도 특징 맵을 생성하도록 한다.

[0008] 상기 다중 스케일 SRF 특징 추출기를 통해 상기 생성된 초해상도 특징 맵과 검출 정확도를 개선하기 위한 상기 SRF 생성기의 손실을 계산하고, 상기 초해상도 특징 맵에 대한 다중 스케일 특징 맵을 학습하는 단계는 상기 저해상도 특징을 확장시키기 위해 상기 다중 스케일 SRF 특징 추출기를 통해 추출된 대상 특징을 이용하여 SRF 생성기와 특징 패치 판별기로 구성된 초해상도 특징 생성 적대 네트워크의 적대적 훈련을 수행하고, 사전 훈련된 SRF 생성기로 상기 다중 스케일 SRF 특징 추출기의 모든 보간 모듈을 변경하여 상기 다중 스케일 SRF 특징 추출기를 구축한다.

[0009] 본 발명의 실시예에 따르면, 입력 이미지로부터 추출된 다중 스케일 특징, SRF 생성기와 특징 패치 판별기의 적대적 훈련을 통해 생성된 초해상도 특징 맵을 이용하여 대상 특징 간의 불일치를 평가하여 적대적 최소-최대 손실을 계산하고, 상기 다중 스케일 SRF 특징 추출기를 통해 상기 초해상도 특징 맵에 대한 적대적 훈련을 수행하고, 상기 대상 이미지 및 상기 저해상도 이미지에 대한 다중 스케일 SRF 특징 추출을 수행하여, 대상 특징 패치에서 초해상도 특징 패치를 구별하도록 훈련된 판별기를 속임으로써 각 스케일에서 대상 영상의 특징 분포를 학습하도록 한다.

[0010] 또 다른 일 측면에 있어서, 본 발명에서 제안하는 적대적 초해상도 다중 스케일 특징맵 학습 및 객체 검출 장치는 다중 스케일 특징 학습을 위한 초해상도 특징 맵을 생성하는 SRF(Super-Resolved Feature) 생성기, 상기 초해상도 특징 맵에서 추출된 패치와 대상 특징을 판별하는 특징 패치 판별기 및 상기 생성된 초해상도 특징 맵과 검출 정확도를 개선하기 위한 상기 SRF 생성기의 손실을 계산하고, 상기 초해상도 특징 맵에 대한 다중 스케일 특징 맵을 학습하고, 학습 결과에 따른 다중 스케일 특징을 추출하고, 객체를 검출하는 다중 스케일 SRF 특징

추출기를 포함한다.

## 발명의 효과

[0011] 본 발명의 실시예들에 따르면, 다중 스케일 특징 학습을 위한 초해상도 특징을 생성하고 다른 컨볼루션 검출기에 쉽게 적용할 수 있는 새로운 SRF-GAN을 제안한다. 고품질의 초해상도 특징을 생성하고 검출 정확도를 개선하기 위한 SRF-GAN 손실을 계산하고, 초해상도 특징의 실측 정보 없이 안정적인 SRF-GAN 훈련을 위한 다중 스케일 특징 학습을 수행할 수 있다. 최근의 다중 스케일 특징 추출기 및 최신 검출 네트워크로 다양한 버전의 SRF 검출기를 구현하여 광범위한 평가 및 일관된 mAP 개선할 수 있다.

## 도면의 간단한 설명

[0012] 도 1은 본 발명의 일 실시예에 따른 적대적 초해상도 다중 스케일 특징맵 학습 및 객체 검출 장치의 구성을 나타내는 도면이다.

도 2는 본 발명의 일 실시예에 따른 적대적 초해상도 다중 스케일 특징맵 학습 및 객체 검출 방법을 설명하기 위한 흐름도이다.

도 3은 본 발명의 일 실시예에 따른 다중 스케일 SRF 특징 추출기를 설명하기 위한 도면이다.

도 4는 본 발명의 일 실시예에 따른 SRFGAN 훈련 알고리즘이다.

도 5는 본 발명의 일 실시예에 따른 SRF-GAN의 검출 결과를 비교한 도면이다.

도 6은 본 발명의 일 실시예에 따른 다른 5개의 잔여 블록이 있는 SRF 생성기GAN을 나타내는 도면이다.

## 발명을 실시하기 위한 구체적인 내용

[0013] 최근의 컨볼루션 객체 검출기는 다양한 규모의 객체를 검출하고 보다 강력한 의미론적 특징 응답을 학습하기 위해 하향식 경로와 함께 추가된 다중 스케일 특징 표현을 활용한다. 일반적으로 하향식 특징 전파 중에 거친 특징 맵은 상향식 경로에서 전달된 형상과 결합되도록 업샘플링되며 결합된 보다 강력한 의미론적 특징은 검출기 헤더 입력이다. 그러나 간단한 보간법(예를 들어, 가장 가까운 이웃 및 이중선형)은 종종 노이즈와 흐릿한 특징이 발생하지만 특징 해상도를 증가시키기 위해 여전히 사용된다.

[0014] 본 발명에서는 기존의 보간법을 쉽게 대체할 수 있는 새로운 초해상도 특징(Super-Resolved Feature; SRF) 생성기를 제안한다. 구체적으로는, SRF 생성기와 특징 패치 판별기로 구성된 초해상도 특징 적대적 생성 네트워크(Super-Resolved Feature Generative Adversarial Network; SRF-GAN)를 설계한다. 또한 초해상도 특징의 고품질 생성 및 검출 정확도 향상을 위한 SRF-GAN 손실을 제안하고, 초해상도 특징과 대상 특징 간의 불일치 및 대상 검출 작업에 대한 손실을 최소화하여 SRF 생성기를 훈련시킨다. SRF 생성기는 기존의 보간법을 대체할 수 있으며, 다른 기존 검출기와 쉽게 결합되어 미세 조정된다. 어블레이션(ablation) 연구를 통해 서로 다른 검출 헤드와 다중 스케일 특징 네트워크를 가진 SRF 생성기의 효과와 유연성을 입증하고 COCO 테스트 디바이스 세트의 최첨단 검출기에 비해 평균 1.7%, 1.1% 더 나은 박스 및 마스크 AP를 달성하는 것을 확인하였다. 이하, 본 발명의 실시 예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0016] 먼저, 본 발명의 실시예에 따른 심층 객체 검출, 다중 스케일 표현 및 검출을 위한 초해상도 GAN에 대해 설명한다.

[0017] 최근의 심층 객체 검출에는 앵커 기반 및 앵커 프리 객체 검출이라는 두 가지 주요 접근 방식이 있다. 앵커 [17]-[21]를 사용하는 심층 컨볼루션 검출기가 여러 검출 벤치마크[22]-[24]에서 현저한 향상을 보인 이후 앵커 기반 객체 검출이 많이 이용되고 있다. 또한 앵커 기반 검출은 투-스테이지 및 원-스테이지 방법으로 나눌 수 있다. 투-스테이지 검출 방법은 먼저 영역 제안 네트워크로 관심 영역(Region of Interest; RoI)을 생성한 다음 후속 R-CNN으로 RoI를 세분화한다. Mask R-CNN[15]은 정확한 픽셀 단위 분할을 위해 투-스테이지 검출기[20]에 마스크 헤드를 부착한다. 다중 스테이지 검출 방법 [2], [25]은 계단식 방법으로 RoI를 반복적으로 개선할 수 있다. Hier R-CNN [26]과 Ellipse R-CNN [27]은 각각 인체 부분 검출 및 타원형 객체 검출을 위한 Mask R-CNN [15]의 확장 버전이다. 반면 원-스테이지 검출기는 영역 제안 없이 검출을 직접 예측한다. SSD[28]는 미리 정의된 앵커를 사용하여 다양한 스케일의 특징 맵에서 다양한 스케일의 예측을 생성한다. RetinaNet[14]은 초점 손

실을 도입하여 클래스 불균형 문제를 해결한다. YOLACT [29]와 RetinaMask [30]는 각각 인스턴스 분할 분기를 [31]과 [14]로 통합한다.

- [0018] 앵커 프리 객체 검출은 앵커 생성을 위한 계산 복잡성과 하이퍼 파라미터를 줄일 수 있다. CornerNet[32] 공격이 키포인트를 쌍으로 결합했다. Grid R-CNN[33]은 Faster R-CNN 헤드[20]를 그리드 포인트를 사용하여 박스를 회귀시키도록 변경한다. FoveaBox [34]는 참조 앵커 없이 객체의 기존 확률과 경계 박스 좌표를 공동으로 예측한다. FCOS[16]는 박스의 중심 영역을 개선하기 위해 중심 분기를 도입한다. CenterMask[3]는 FCOS에 공간 주의 유도 마스크 분기를 추가한다. ATSS[35]는 FCOS[16]와 RetinaNet[14]의 정확도를 향상시키기 위해 객체에 대한 통계적 접근법에 기초한 자동 양 및 음의 샘플 선택 방법을 도입한다.
- [0019] 객체 스케일 변동에 대한 견고성과 더 나은 검출을 달성하기 위해 직관적인 접근 방식은 다양한 규모의 특징을 사용하는 것이다. [36]은 훈련 중 객체 스케일 변동을 줄이기 위한 이미지 피라미드 체계에 대한 스케일 정규화를 제시한다. Autofocus[37]는 다중 스케일 추론 중에 각 스케일에 더 초점을 맞출 영역을 결정할 수 있다. 이러한 작업은 또한 스케일 불변 검출기를 훈련시킬 수 있는 좋은 직관을 제공한다.
- [0020] 다중 스케일 영상 표현을 사용하는 대신, 특징 피라미드(또는 다중 스케일 특징)를 스케일 불변 검출에 사용할 수도 있다. 종래기술[28], [38]-[40]은 상향식 경로에서 추출한 다중 스케일 특징을 결합한다. 한편, 최근 종래 기술[4]-[7], [11], [12]는 여러 다른 경로에서 다중 스케일 특징을 학습한다. FPN[4]은 먼저 상향식 및 하향식 특징을 사용하는 것이 스케일 불변 검출에 효과적이라는 것을 보여준다. PANet[5]은 추가적인 상향식 경로를 추가로 도입한다. 이러한 작업으로부터, 교차 스케일 [6], [7], [11], [12] 및 다중 스케일 특징 융합[8]을 사용하는 많은 네트워크 구조가 제시되었다. NAS-FPN[7]은 신경 구조 검색 알고리즘을 사용하여 특징 피라미드에 적합한 구조를 발견한다[41]. AugFPN[8]은 FPN[4]을 더욱 향상시킨다. EfficientDet[6]은 양방향 기능과 하향식 및 상향식 특징 융합을 반복적으로 적용한다. 그러나 이러한 모든 방법은 특징 해상도를 증가시킬 때 순수한 보간법을 이용한다. 따라서 특징 피라미드를 보다 정확하게 학습하기 위한 특징 확대 방법을 개발하는 데 중점을 둔다. 반면에 본 발명에서는 보간법을 제안하는 기법으로 대체함으로써 이전의 모든 방법에 쉽게 적용할 수 있다.
- [0021] GAN[42], [43]의 발전과 함께 GAN[44]-[46]을 사용하는 초해상도 방법이 제안되었다. 또한 객체 검출 개선을 위해 SR을 적용하기 위한 몇 가지 종래기술들[47]-[50]이 있다. EESRGAN[47]은 초해상도 영상을 직접 이용하여 소규모의 물체를 검출한다. SOD-MTGAN[48]은 객체 제안에 대한 SR 손실과 함께 다중 작업 손실을 설계한다. 초해상도 ROI 특징[49], [50]은 소형 객체 검출 개선을 위해 학습된다. 이러한 종래기술과 비교하여, 제안하는 기법은 특징 맵 자체를 업샘플링할 수 있다.
- [0023] 도 1은 본 발명의 일 실시예에 따른 적대적 초해상도 다중 스케일 특징맵 학습 및 객체 검출 장치의 구성을 나타내는 도면이다.
- [0024] 본 발명에서는 종래기술의 문제를 해결하기 위해 다중 스케일 특징 학습에 사용되는 초해상도 특징을 생성할 수 있는 새로운 특징 생성기를 개발하는 것을 목표로 한다. 이러한 생성기를 학습하기 위해 SRF 생성기와 특징 패치 판별기로 구성된 초해상도 특징 생성 적대 네트워크(Super-Resolved Feature Generative Adversarial Network; SRF-GAN)를 제안한다. 또한 SRF-GAN을 다중 작업 학습에서 다중 스케일 객체 검출 및 분할에 더 적합하게 만들 수 있는 새로운 통합 손실을 제안한다.
- [0025] SRF-GAN 학습을 위해 초해상도 특징 GAN 손실을 가진 특징 패치 판별기와 SRF 생성기 간에 적대적 훈련을 수행한다. 따라서 SRF-GAN은 입력 특징에서 일반적인 초해상도 특징 표현을 학습할 수 있다. 이후 보간 모듈을 SRF 생성기로 교체하여 SRF-GAN을 다중 스케일 특징 추출기와 통합한다. 그런 다음 다중 스케일 SRF 추출기를 학습하기 위해 객체 검출 및 초해상도 특징 GAN 손실을 포함한 통합 손실을 사용하여 다중 스케일 SRF 추출기와 특징 패치 판별기 사이의 적대적 훈련을 따른다.
- [0026] SRF-GAN 훈련의 또 다른 문제점은 초해상도 기능의 실측 정보를 사용할 수 없다는 데서 비롯된다는 점에 유의한다. 이를 해결하기 위해, 본 발명의 핵심 아이디어는 이러한 SRGAN 훈련 중에 다중 스케일 특징 네트워크(예를 들어, FPN [4], PAFPN [5], BiFPN [6])를 대상 특징 생성기로 활용하는 것이다. 이를 위해 특징 네트워크에 원본 및 다운샘플링된 영상을 공급한 다음 각 레벨에서 특징을 추출한다. 그런 다음 다운샘플링된 영상의 다중 스케일 특징이 SRF 생성기로 전달되고, 초해상도 특징이 도 1에 도시된 특징 패치 판별기에 의해 원본 영상에서 추출된 해당 특징과 비교된다.

- [0027] SRF-GAN이 훈련되면 보간에 직접 사용할 수 있다. 하지만, SRF 생성기를 사용한 대상 검출기 훈련은 파라미터를 특정 작업에 대해 함께 조정할 수 있기 때문에 더 나은 검출을 보여준다. 실제로 본 발명에서는 다른 백본과 함께 훈련된 SRF-GAN의 재사용이 가능하기 때문에 도 1에 표시된 SRF-GAN의 사전 훈련을 생략할 수 있다. 이는 SRFGAN이 일반화된 초해상도 특징을 학습할 수 있고 다양한 백본에 대해 높은 유연성을 가질 수 있음을 나타낸다. SRFGAN을 증명하기 위해 앵커 기반 원-스테이지 또는 투-스테이지 검출기와 앵커 프리 검출기를 나타내는 RetinaNet [14], Mask R-CNN [15], FCOS [16], CenterMask [3]를 사용하여 여러 버전의 대상 검출기를 구현했다. COCO 데이터셋의 최근 검출기에 비해 검출 정확도가 크게 향상되었음을 보여주었다. 또한 다양한 백본과 검출기를 사용하여 광범위한 어블레이션(ablation) 연구를 수행했다.
- [0028] 본 발명의 실시예에 따르면, 다중 스케일 특징 학습을 위한 초해상도 특징을 생성하고 다른 컨볼루션 검출기에 쉽게 적용할 수 있는 새로운 SRF-GAN을 제안한다. 고품질의 초해상도 특징을 생성하고 검출 정확도를 개선하기 위한 SRF-GAN 손실을 계산하고, 초해상도 특징의 실측 정보 없이 안정적인 SRF-GAN 훈련을 위한 다중 스케일 특징 학습을 수행한다. 최근의 다중 스케일 특징 추출기(즉, FPN [4], PAFPN [5], BiFPN [6]) 및 최신 검출 네트워크(예를 들어, RetinaNet [14], Mask R-CNN [15], FCOS [16], CenterMask [3])로 다양한 버전의 SRF 검출기를 구현하여 광범위한 평가 및 일관된 mAP 개선할 수 있다.
- [0029] 도 1을 참조하면, 다중 스케일 특징 학습에 적용할 수 있는 모든 규모의 초해상도 특징을 생성하기 위해, 본 발명의 실시예에 따른 SRF 생성기(110)와 특징 패치 판별기(120)로 구성된 SRF-GAN(Super-Resolved Feature Generative Adversarial Network)을 설계한다. 하지만, 초해상도 특징의 실측 정보를 일반적으로 사용할 수 없기 때문에 직접적인 지도가 어렵다. 따라서 본 발명에서는 기존의 다중 스케일 특징 네트워크를 대상 특징 생성기로 사용하고 SRF 생성기의 초해상도 형상과 대상 특징 생성기의 해당 해상도의 특징을 일치시킨다. SRF-GAN을 처음부터 훈련하기 위해 과도하게 장착되지 않도록 점진적인 학습을 제안한다. 하지만 다중 스케일 특징 네트워크를 사용하여 사전 훈련된 SRF 생성기가 제공되는 경우 SRF 생성기에 내장된 대상 검출기를 즉시 훈련시킬 수 있다. 따라서 SRF-GAN을 위한 일부 사전 훈련 단계를 실제로 생략할 수 있다.
- [0030] 도 1과 같이, SRF 생성기(G)(110)는 저해상도 특징  $P^{lr}$ 에 대해 입력 특징 맵에 대한 초해상도 특징 맵  $P^{sr}$ 을 생성한다. 한편, 특징 패치 판별기(D)(120)는 초해상도 특징 맵  $P^{sr}$ 에서 추출된 패치와 대상 특징  $P^{tr}$ 을 식별한다. 초해상도 특징(Super-Resolved Feature)(141) 및 대상 특징(Target Feature)(142) 경로를 도시하였다.
- [0031] SRF 생성기(G)(110)의 경우  $3 \times 3$  컨볼루션 및 Leaky ReLU 활성화(LReLU) [51] 계층( $\alpha = 0.2$ )에 해상도의  $P^{lr}$ 을 공급한다(111). 이후, 3개의 연속 잔여 밀도 블록[45]과 이들을 우회하는 바로 가기 연결을 추가하여 초해상도에 대한 보다 개선된 표현을 활용한다. 구체적으로는, 각 잔여 밀도 블록은 성장률이 32인 밀집하게 연결된  $3 \times 3$  컨볼루션 층 5개, Leaky ReLU 층 4개, 바로가기 연결 1개를 포함하고 있다. 훈련 안정화를 위해 식별 맵 작성 전에 잔여 스케일링 [52]( $\beta = 0.2$ )을 적용한다. 그런 다음 하나의 컨볼루션과 하나의 디컨볼루션 블록을 따라 특징 해상도를 2의 배수로 확대한다.
- [0032] 마지막으로, 이중선 보간법에 의해 디컨볼루션 특징과 업샘플링된 입력 특징 사이에  $3 \times 3$  컨볼루션 계층과 바로가기 연결을 추가한다.
- [0033] 특징 패치 판별기(D)(120)는 512, 1024, 1024 채널로 세 개의 컨볼루션 블록을 사용하여  $P^{sr}$  또는  $P^{tr}$ 를 컨볼루션한다. 여기서 각 블록에는  $3 \times 3$  컨볼루션, 배치 정규화[54] 및 Leaky ReLU 활성화 계층( $\alpha = 0.2$ )가 포함된다. 그런 다음 특징 맵 픽셀당 클래스는  $3 \times 3$  컨볼루션 및 시그모이드 활성화 함수에 의해 예측된다.
- [0034] 본 발명의 실시예에 따른 다중 스케일 SRF 특징 추출기(140)를 통해 생성된 초해상도 특징 맵과 검출 정확도를 개선하기 위한 SRF 생성기(110)의 손실을 계산하고, 초해상도 특징 맵에 대한 다중 스케일 특징 맵을 학습한다.
- [0035] 이미지 I에서 다중 스케일 특징  $F(I) = \{P_i | n_s \leq i \leq n_e\}$ 를 나타낸다. 여기서 F는 다중 스케일 특징 추출기(F)(130),  $P_i$ 는 레벨 i의 특징 맵,  $n_s$ 와  $n_e$ 는 더 미세한 해상도에서 더 거친 해상도로의 하향식 특징 맵의 첫 번째 및 마지막 스케일 레벨이다. 대상 이미지  $I^{tr}$  및 저해상도  $I^{lr}$ 이 주어지면, 적대적 최소-최대 손실  $L_{adv}$ 를 사용하여 F로 G 및 D를 학습하는 문제를 정의한다.

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_{I^{tr} \sim p_{train}(I^{tr})} \left[ \sum_{i=n_s}^{n_e} \frac{1}{W_i H_i} \sum_{x=1}^{W_i} \sum_{y=1}^{H_i} \log \left( D_{\theta_D} (P_i^{tr})_{x,y} \right) \right] \\ + \mathbb{E}_{I^{tr} \sim p_G(I^{tr})} \left[ \sum_{i=n_s}^{n_e} \frac{1}{W_i H_i} \sum_{x=1}^{W_i} \sum_{y=1}^{H_i} \log \left( 1 - D_{\theta_D} (G_{\theta_G} (P_i^{tr}))_{x,y} \right) \right] \quad (1)$$

이를 해결하기 위해  $I^{tr}$  를  $I^{lr} = \downarrow^{\times s} (I^{tr})$  로 축소한다. 여기서  $\downarrow^{\times s}$  는 축소 요인  $s (< 1)$  에 의한 축소를 의미한다. 그런 다음  $I^{tr}$  및  $I^{lr}$  을 각각 F에 공급하여 다중 스케일 특징  $\mathbf{P}^{tr} = \{P_{n_s}^{tr}, \dots, P_i^{tr}, \dots, P_{n_e}^{tr}\}$  및  $\mathbf{P}^{lr} = \{P_{n_s}^{lr}, \dots, P_i^{lr}, \dots, P_{n_e}^{lr}\}$  을 추출한다. 따라서 이 공식은 G(110)가 대상 특징 패치에서 초해상도 특징 패치를 구별하도록 훈련된 D(120)를 속임으로써 각 스케일에서 대상 영상의 특징 분포를 학습하도록 하는 것이다. 다중 스케일 특징의 경우  $W_i$ 와  $H_i$ 는 각각 스케일 레벨  $i$ 에 따른 다중 스케일 특징의 폭과 높이이다.  $x$ 와  $y$ 는 특징 픽셀 좌표의 인덱스이다.  $\theta_D$ 와  $\theta_G$ 는 각각 D(120)와 G(110)의 파라미터이다. 또한  $P_i^{tr}$ 와  $D(P_i^{tr})$ 의 해상도는 동일하다. 도 1과 같이 BiFPN[6]을 F로 활용한다. 그러나 다른 다중 스케일 특징 추출기(예를 들어, PAFPN [5] 및 FPN [4])로 교체할 수 있다. 본 발명의 실시예에 따른 F(130)에 대해 BiFPN [6], PAFPN [5] 및 FPN [4]를 사용한다. 또한 BiFPN에서  $P_{i-1}$ 의 해상도가  $P_i$ 의 해상도보다 2배 더 높기 때문에  $s$ 를 0.5로 설정했다( $n_s < i \leq n_e$ ).

도 2는 본 발명의 일 실시예에 따른 적대적 초해상도 다중 스케일 특징맵 학습 및 객체 검출 방법을 설명하기 위한 흐름도이다.

제안하는 적대적 초해상도 다중 스케일 특징맵 학습 및 객체 검출 방법은 SRF(Super-Resolved Feature) 생성기를 통해 다중 스케일 특징 학습을 위한 초해상도 특징 맵을 생성하고, 특징 패치 판별기를 통해 상기 초해상도 특징 맵에서 추출된 패치와 대상 특징을 판별하는 단계(210), 다중 스케일 SRF 특징 추출기를 통해 상기 생성된 초해상도 특징 맵과 검출 정확도를 개선하기 위한 상기 SRF 생성기의 손실을 계산하고, 상기 초해상도 특징 맵에 대한 다중 스케일 특징 맵을 학습하는 단계(220) 및 상기 다중 스케일 추출기의 학습 결과에 따른 다중 스케일 특징을 추출하고, 객체를 검출하는 단계(230)를 포함한다.

단계(210)에서, SRF 생성기를 통해 다중 스케일 특징 학습을 위한 초해상도 특징 맵을 생성하고, 특징 패치 판별기를 통해 상기 초해상도 특징 맵에서 추출된 패치와 대상 특징을 판별한다. 초해상도 특징 맵을 생성하기 위해 SRF 생성기 및 특징 패치 판별기 간의 적대적 훈련을 수행한다. 픽셀 단위 손실과 적대적 손실로 구성된 초해상도 특징을 정의하고, 저해상도 특징과 각 스케일 레벨에서 초해상도의 대응 대상 특징 사이의 불일치를 평가하여, SRF 생성기가 특징 패치 판별기를 속여 초해상도 특징 맵을 생성하도록 한다.

단계(220)에서, 다중 스케일 SRF 특징 추출기를 통해 상기 생성된 초해상도 특징 맵과 검출 정확도를 개선하기 위한 상기 SRF 생성기의 손실을 계산하고, 상기 초해상도 특징 맵에 대한 다중 스케일 특징 맵을 학습한다.

저해상도 특징을 확장시키기 위해 상기 다중 스케일 SRF 특징 추출기를 통해 추출된 대상 특징을 이용하여 SRF 생성기와 특징 패치 판별기로 구성된 초해상도 특징 생성 적대 네트워크의 적대적 훈련을 수행한다. 사전 훈련된 SRF 생성기로 상기 다중 스케일 SRF 특징 추출기의 모든 보간 모듈을 변경하여 상기 다중 스케일 SRF 특징 추출기를 구축할 수 있다.

더욱 상세하게는 입력 이미지로부터 추출된 다중 스케일 특징, SRF 생성기와 특징 패치 판별기의 적대적 훈련을 통해 생성된 초해상도 특징 맵을 이용하여 대상 특징 간의 불일치를 평가하여 적대적 최소-최대 손실을 계산한다.

상기 다중 스케일 SRF 특징 추출기를 통해 상기 초해상도 특징 맵에 대한 적대적 훈련을 수행하고, 상기 대상 이미지 및 상기 저해상도 이미지에 대한 다중 스케일 SRF 특징 추출을 수행하여, 대상 특징 패치에서 초해상도 특징 패치를 구별하도록 훈련된 판별기를 속임으로써 각 스케일에서 대상 영상의 특징 분포를 학습하도록 한다.

단계(230)에서, 다중 스케일 추출기의 학습 결과에 따른 다중 스케일 특징을 추출하고, 객체를 검출한다. 도 3

을 참조하여 본 발명의 실시예에 따른 다중 스케일 SRF 특징 추출 과정을 더욱 상세히 설명한다.

[0048] 도 3은 본 발명의 일 실시예에 따른 다중 스케일 SRF 특징 추출기를 설명하기 위한 도면이다.

[0049] 도 3을 참조하여 SRF 생성기와 결합된 다중 스케일 특징 추출기를 도시하였다. 도 3(a) 내지 도 3(c)는 FPN(Feature Pyramid Network)[4], PAFPN(Path Aggregation Network)[5], BiFPN(Bi-directional Feature Pyramid Network)[6]의 구조를 나타낸다. 각기, 각 네트워크의 보간 모듈(Interp.)에서 가장 가까운 이웃 보간을 SRF 생성기로 대체하고, 본 발명의 실시예에서  $D_{bifpn}$ 을 3으로 설정하였다.

[0050] 도 3을 참조하여, FPN, PAFPN, BiFPN과 같은 다양한 다중 스케일 특징 추출기를 설명한다. 제안하는 다중 스케일 SRF 특징 추출기는 F로서 모든 다중 스케일 특징 추출기를 적용할 수 있다. BiFPN과 PAFPN의 경우 SRF-GAN을 훈련할 때  $P_i^{out}$ 와  $N_i$ 를 각각 대상 특징으로 활용한다. 이는  $P_i^{out}$ 와  $N_i$ 가 도 1과 도 3에서와 같이 다중 스케일 특징 추출기에서 상향식 및 하향식 전파를 반복한 후의 최종 출력 특징이기 때문이다. BiFPN을 기본 특징 추출기 F로 사용하는 방법에 대해 아래에서 설명하고, 단순성을 위해 I번째 레벨 다중 스케일 특징 맵을  $P_i$ 로 표시한다.

[0052] 도 4는 본 발명의 일 실시예에 따른 SRFGAN 훈련 알고리즘이다.

[0053] 본 발명의 실시예에 따른 SRF-GAN 훈련의 목표는 대상 검출기에 대한 다중 스케일 SRF 추출기를 생성하는 것이다. 먼저 저해상도 특징을 2배 확장할 수 있는 일반화된 SRF-GAN을 훈련한다. 이를 훈련시키기 위해 PAFPN의 다중 스케일 특징을 대상 특징으로 활용하여 G와 D 사이의 적대적 훈련을 수행한다. 그런 다음 사전 훈련된 SRF 생성기로 PAFPN의 모든 보간 모듈을 변경하여 다중 스케일 SRF 추출기를 구축한다. 다중 스케일 SRF 추출기와 D는 대안적인 방법으로 적대적으로 훈련될 수 있다. 기본적으로, 본 발명에서는 식(1)을 이용하여 훈련시킬 수 있다. 하지만, 픽셀 단위 L1과 검출 손실이 추가된다. 결과적으로, 본 발명에서는 객체 검출을 위한 스케일 당 초해상도 특징과 다중 스케일 표현 품질을 향상시킬 수 있다. 마지막으로, 적대적 훈련 없이 검출 손실을 최소화하여 훈련된 SRF 생성기로 여러 FPN 기반 Mask R-CNN, RetinaNet, CenterMask 검출기를 훈련시킬 수 있다. 보다 명확히 하기 위해 SRFGAN 훈련 알고리즘 표 1을 제시한다.

[0054] 먼저, 본 발명의 실시예에 따른 SRF-GAN에 있어서, 초해상도 특징 생성을 위해 SRF 생성기 G와 특징 패치 판별기 D 사이에서 적대적 훈련을 수행한다. 먼저 픽셀 단위 L1 손실과 식(1)의 적대적 손실로 구성된 초해상도 특징  $L_{SRF}(G, D, F) = L_{L1}(G, F) + \lambda L_{adv}(G, D, F)$ 를 정의한다.  $L_{L1}$ 은 저해상도 특징과 각 스케일 레벨 i에서 초해상도의 대응 대상 특징 사이의 불일치를 평가한다. 반면,  $L_{advG}$ 는 G가 D를 속여 초해상도 특징을 만들도록 한다.  $\theta_G$ 에 관한  $L_{SRF}$ 를 최소화함으로써, G를 다음과 같이 훈련시킬 수 있다.

$$\min_{\theta_G} \sum_{i=n_s}^{n_e} \frac{1}{CW_i H_i} \sum_{c=1}^C \sum_{x=1}^{W_i} \sum_{y=1}^{H_i} \left| (P_i^{tr})_{x,y}^c - G_{\theta_G}(P_i^{tr})_{x,y}^c \right| + \lambda \sum_{i=n_s}^{n_e} \frac{1}{W_i H_i} \sum_{x=1}^{W_i} \sum_{y=1}^{H_i} -\log \left( D_{\theta_D}(G_{\theta_G}(P_i^{tr}))_{x,y} \right) \quad (2)$$

[0056]  $\lambda$ 는 D의 피드백을 제어하기 위한 하이퍼 파라미터로 0.001로 조정된다. C는 특징 맵의 채널이며 256으로 설정된다.

[0057] 반면, 훈련 D에서는 식(1)에 설명된 일반적인 GAN 손실을 활용한다. D는 G의 지정된 대상 및 초해상도 특징 패치에 대한 정확한 레이블을 식별할 확률을 최대화하려고 한다. 이 적대적 훈련에서 SRF-GAN은 입력 특징에 대한 일반화된 초해상도 특징 표현을 학습할 수 있다.

[0058] 다시 도 1을 참조하면, 본 발명의 실시예에 따른 다중 스케일 SRF 추출기는 훈련된 SRF 생성기(110)를 BiFPN에 내장하여 설계한다. 간단히, SRF 생성기를 사용하여 BiFPN의 모든 보간 모듈을 변경한다. 도 1(140)과 같이 추출기(130)에 박스와 마스크 헤드를 부착하였고, 제안하는 다중 스케일 SRF 추출기를 M(140)으로 표시하였다. 적

대적 훈련을 위해, 또한 특징 패치 판별기 D(120)를 사용하고 D의 훈련된 파라미터는 재사용된다. 검출 정확도 및 초해상도 특징의 품질을 고려하여  $L_{INT}(M, D, F) = L_{SRF}(M, D, F) + L_{DET}(M)$ 로 통합 손실을 정의한다. 여기서,  $L_{DET}(M)$ 은 검출 헤드에 따른 전체적인 검출 손실이다. 본 발명의 실시예의 경우 다중 스케일 SRF 추출기(140)에 헤드를 연결할 때 Mask R-CNN [15], RetinaNet [14], FCOS [16], CenterMask [3], CascadeR-CNN [25]의 손실을 사용한다. 식(2)와 유사하게  $L_{SRF}(M, D, F)$ 는 초해상도 특징과 대상 특징 간의 불일치를 평가하고 G가 D를 속여서 초해상도 특징을 생성하도록 권장하는 초해상도 특징 손실이다.  $L_{INT}(M, D, F)$ 를 최소화하여 M을 훈련하는 동안 F는 훈련되지 않고 D에 대상 특징만 제공한다.

[0059]  $L_{SRF}$  평가의 경우 먼저  $I^{tr}$ 를 축소하여  $I^{lr} = \downarrow^{0.5}(I^{tr})$ 를 재샘플링한 다음 M과 F에  $I^{lr}$  및  $I^{tr}$ 를 제공하여 각각 그림과 같이  $\mathbf{P}^{lr} = \{P_3^{lr}, \dots, P_7^{lr}\}$  및  $\mathbf{P}^{tr} = \{P_3^{tr}, \dots, P_7^{tr}\}$ 을 추출한다. 그런 다음 일련의 초해상도 특징  $\mathbf{P}^{sr} = \{G_{\theta_G}(P_3^{lr}), \dots, G_{\theta_G}(P_7^{lr})\}$ 를 추출하는 한편  $P^{tr}$ 을  $\downarrow^{0.5}(\mathbf{P}^{tr})$ 로 축소한다. 이러한 이유는, 의미론적 정보 레벨이 [4], [5]에서 논한 특징 피라미드 레벨 간에 다르기 때문에 동일한 스케일(또는 피라미드) 레벨 i에서 초해상도 및 대상 특징을 비교하기 위함이다. 예를 들어, 본 발명에서는 각각  $\{P_4^{sr}, P_5^{sr}, P_6^{sr}, P_7^{sr}\}$ 와  $\{P_3^{tr}, P_4^{tr}, P_5^{tr}, P_6^{tr}\}$ 을 비교하기 위해 M과 F에 동일한  $I^{tr}$ 을 공급할 수 있다. 그러나  $L_{SRF}$ 를 평가할 때 특징 의미론적 레벨의 불일치가 mAP를 약 1.6%로 저하시킨다. 또한,  $P^{tr}$ 을 축소하는 것은 GPU 사용량을 감소시키기 위함이다. 그렇지 않으면 원래의  $I^{tr}$ 과  $\uparrow^{2}(I^{tr})$ 를 M과 F에 공급하여 다운샘플링 없이  $\{G_{\theta_G}(P_3^{tr}), \dots, G_{\theta_G}(P_7^{tr})\}$ 와  $F(\uparrow^{2}(I^{tr}))$  사이의 레벨별 특징을 비교할 수 있다. 그러나 GPU 메모리에는 매우 많은 비용이 든다. 그 대가로  $\downarrow^{0.5}(I^{tr})$ 의 입력으로  $L_{DET}(M)$ 을 평가하기 위해서는 박스 위치와 마스크 영역의 실측 정보를  $\downarrow^{0.5}(I^{tr})$  해상도에 맞춰야 한다. 다중 스케일 SRF 추출기의 파라미터 M을 훈련시키기 위해 다음  $L_{SRF}(M, D, F)$ 를 최소화한다.

$$\begin{aligned} \min_{\theta_M} & \sum_{i=n_s}^{n_e} \frac{1}{C[sW_i][sH_i]} \\ & \times \sum_{c=1}^C \sum_{x'=1}^{[sW_i]} \sum_{y'=1}^{[sH_i]} \left| (\downarrow^{0.5}(P_i^{tr}))_{x',y'}^c - (P_i^{sr})_{x',y'}^c \right| \\ & + \lambda \sum_{i=n_s}^{n_e} \frac{1}{[sW_i][sH_i]} \sum_{x'=1}^{[sW_i]} \sum_{y'=1}^{[sH_i]} -\log(D_{\theta_D}(P_i^{sr})_{x',y'}) \end{aligned} \quad (3)$$

[0061] 여기서  $[sW_i]$  및  $[sH_i]$ 는 레벨 i에서  $s(= 0.5)$ 의 요인으로 다운샘플링된 대상 특징의 폭과 높이이다. 식(2)와 동일한  $\lambda$ 를 사용한다. 식(2)와 비교하여, 이전 레벨의 초해상도 특징은 다음 스케일 레벨의 입력으로 사용되며, 여기서 초해상도 특징은  $P_i^{sr} = M_{\theta_M}(P_{i+1}^{sr})$ ,  $n_s \leq i < n_e$ 로 추출된다. 따라서  $L_{SRF}(M, D, F)$ 는 M을 다중 스케일 표현에 더 적합하게 만든다.

[0062] D의 적대적 훈련의 경우 식(1)에서  $L_{advD}$ 를 D의 적대적 손실로 표시한다.  $\downarrow^{0.5}(\mathbf{P}^{tr})$  및  $\mathbf{P}^{sr}$ 을 실제 및 가짜 입력 특징으로 사용한다. 유사한 방법으로,  $L_{advD}$ 를 극대화함으로써, 본 발명의 실시예에 따른 D를 훈련시킬 수 있고, 훈련 M에 대해 생성된  $\mathbf{P}^{sr}$ 에 대하여 예측을 활용할 수 있다.

[0063] 본 발명의 실시예에 따른 대상 검출기에 있어서, 다중 스케일 특징 추출기를 백본으로 활용하는 대상 검출기 T를 훈련하기 위해 훈련된 SRF 생성기를 적용한다. 보다 구체적으로, SRF 생성기로만 T의 모든 보간 모듈을 변경하지만 특징 추출기의 다른 훈련된 파라미터는 재사용하지 않는다. T를 훈련시키기 위해 앞서 설명된 바와 같이

T의 헤드 유형에 의해 정의된 전체 검출 손실  $L_{DET}$ 를 최소화한다. 본 발명의 실시예에 따른 다중 스케일 SRF 특징 추출기에 대한 이전 훈련과의 주요 차이점은 다운샘플링 없이 원본 이미지 자체를 T에 공급한다는 점이다. 따라서 SRF 생성기는 본 훈련을 통해 고해상도 영상에서 검출에 더 적합하도록 미세 조정할 수 있다. 본 발명의 실시예에 따르면, 다중 스케일 특징 추출기를 백본으로 활용하는 대상 검출기 T를 훈련하기 위해 훈련된 SRF 생성기를 적용한다. 보다 구체적으로, SRF 생성기로만 T의 모든 보간 모듈을 변경하지만 특징 추출기의 다른 훈련된 파라미터는 재사용하지 않는다. T를 훈련시키기 위해 앞서 설명한 대로 T의 헤드 유형에 의해 정의된 전체 검출 손실  $L_{DET}$ 를 최소화한다. 다중 스케일 SRF 특징 추출기에 대한 이전 훈련과의 주요 차이점은 다운샘플링 없이 원본 이미지 자체를 T에 공급한다는 점이다. 따라서 SRF 생성기는 본 훈련을 통해 고해상도 영상에서 검출에 더 적합하도록 미세 조정할 수 있다.

[0064] 또한 T를 훈련하는 동안 SRF 생성기의 파라미터를 미세 조정한다. 실제로, 훈련된 SRF 생성기를 제공할 경우 SRF-GAN 및 다중 스케일 SRF 특징 추출기의 훈련 없이 T를 직접 훈련할 수 있다. 이는 SRF 생성기를 사용하는 T의 훈련 복잡성이 크게 감소할 수 있음을 나타낸다. 또한, SRF 생성기만 사용하는 대신 T에 대해 전체 다중 스케일 SRF 특징 추출기를 재사용할 수도 있다. 이 경우 T의 mAP는 더욱 개선될 수 있다.

[0065] 본 발명의 실시예에 따르면, 다양한 최신 검출 헤드를 SRF 구조에 결합하고 다른 검출 헤드 T를 가진 다양한 버전의 SRF 기반 검출기를 구현한다. 각 검출 헤드의 세부 정보를 다음과 같이 나타낼 수 있다:

[0066] 본 발명의 실시예에 따른 Mask R-CNN는 앵커 기반 투-스테이지 객체 검출 방법[15]에 관한 것으로, 영역 제안 및 개선을 위한 투-스테이지로 구성된다. 원-스테이지에서 검출기는 다중 스케일 특징 및 참조 박스(다시 말해, 앵커)를 RPN[20]에 공급하여 영역 제안을 생성한다. 전경 분류 점수 및 IOU 점수가 있는 제안에 최대값이 아닌 억제(Non-Maximum Suppression; NMS)를 적용한 후 RoI 정렬 연산에 의해 제안의 특징이 잘린다. 투-스테이지에서는 박스 위치를 세분화하고 각 제안에 대한 클래스 신뢰도 및 객체 마스크 영역을 예측하기 위해 제안의 특징들이 3개의 병렬 헤드 분기(다시 말해, 박스 회귀, 분류, 클래스 무관 마스크 예측 헤드)를 통해 전달된다. 그런 다음 최종 결과를 산출하기 위해 후보 박스에 클래스별 NMS를 적용한다. 마스크 R-CNN  $L_{DET}^M$ 의 손실 함수는 다음과 같이 정의된다.

$$L_{DET}^M = L_{cls}^M + L_{loc}^M + L_{mask}^M + L_{rpn}^M \quad (4)$$

[0068] 여기서  $L_{cls}^M$ 와  $L_{loc}^M$ 은 로그(즉, 교차 엔트로피) 및 평활화 L1 손실이다 [17].  $L_{mask}^M$ 는 평균 이진 교차 엔트로피 손실이다. 영역 제안 네트워크의 손실  $L_{rpn}^M$ 은 분류 및 박스 회귀를 위한 이진 교차 엔트로피와 평활화 L1 손실로 구성된다.

[0069] 본 발명의 실시예에 따른 RetinaNet은 앵커 기반 원-스테이지 객체 검출 방법[14]으로, 클래스 불균형 문제를 해결하기 위한 초점 손실을 제시한다. 검출기는 분류 및 박스 회귀를 위한 백본 네트워크와 두 개의 서브넷으로 구성된다. 백본 네트워크의 경우, 다중 스케일 특징을 추출한다. 그런 다음 분류 서브넷은 사전 정의된 9개의 앵커와 80개의 객체 클래스 각각에 대해 각 공간 위치에서 분류 확률을 예측한다. 박스 회귀 서브넷은 각 객체의 공간 위치를 예측하기 위한  $4 \times 9$  선형 출력을 추론한다. 그런 다음 RetinaNet은 검출기 신뢰도 0.05를 임계값으로 설정한 후 다중 스케일 특징 추출기 레벨당 상위 1k개 후보로부터 박스 예측을 디코딩한다. 모든 레벨에서 상위 예측을 병합한 후 클래스당 NMS를 적용하여 최종 결과를 얻는다. 본 RetinaNet 훈련을 위해 손실 함수  $L_{DET}^R$ 는 다음과 같이 정의된다.

$$L_{DET}^R = L_{cls}^R + L_{loc}^R \quad (5)$$

[0071] 여기서  $L_{cls}^R$ 와  $L_{loc}^R$ 은 각각 초점 및 평활화 L1 손실이다.

[0072] 본 발명의 실시예에 따른 FCOS는 앵커 프리 원-스테이지 객체 검출 방법[16]으로, 박스의 중심 영역을 정제하는 중심 분기를 도입한다. 검출기는 백본과 3개의 헤드 분기(다시 말해, 분류, 회귀 및 중심)로 구성된다. 검출기는 다중 스케일 특징 추출기를 기반으로 한다. 박스 위치를 결정하기 위해 회귀 분기는 4차원 정규화된 간격 띄우기(다시 말해, 왼쪽, 오른쪽, 위쪽, 아래쪽)를 예측하여 중앙 지점의 박스 길이를 조정한다. 분류 분기는 80개의 다른 객체 범주를 분류할 때 박스당 80차원 벡터를 추론한다. 중심 분기는 예측된 중심 지점이 해당 GT 지

점에 얼마나 가까운지를 측정한다. 이 점수를 예측하기 위해 단일 계층 분기가 추가된다. 박스의 분류 점수에 중심 지점을 곱하면 박스 점수를 다시 가중할 수 있다. 그 결과, NMS 중에는 저품질 박스를 더 잘 억제할 수 있다. FCOS  $L_{DET}^F$ 의 손실은 다음과 같다.

$$L_{DET}^F = L_{cls}^F + L_{reg}^F + L_{ctr}^F \quad (6)$$

여기서  $L_{cls}^F$ 는 [14]에서와 같이 초점 손실이며  $L_{reg}^F$ 는 [55]에서와 같이 IOU 손실이다.  $L_{ctr}^F$ 은 0에서 1 사이의 중심 점수를 예측하기 위한 이진 교차 엔트로피 손실이다.

본 발명의 실시예에 따른 CenterMask[3]는 분할을 위해 공간 주의 유도 마스크(Spatial Attention-Guided Mask SAG-Mask) 분기를 연결한 FCOS[16]의 개선된 버전이다. 그러므로, 앞서 설명한 FCOS와 유사한 구조를 가지고 있다. 그러나 CenterMask는 잘린 영역 내의 분할 마스크를 픽셀 단위로 예측하기 위해 공간 주의 유도 마스크 헤드로 명명된 마스크 헤드를 추가한다. 또한 예상 마스크 품질을 고려하여 분류 점수를 재보정하기 위해 마스크 점수 헤드[56]를 부착한다. 이를 훈련하기 위해 다음과 같은 검출 손실  $L_{DET}^C$ 가 사용된다.

$$L_{DET}^C = L_{cls}^C + L_{reg}^C + L_{ctr}^C + L_{mask}^C + L_{maskiou}^C \quad (7)$$

여기서  $L_{cls}^C, L_{reg}^C, L_{ctr}^C$ 은 앞서 설명된 바와 같이 FCOS에서와 동일하다.  $L_{mask}^C$ 는 앞서 설명된 바와 같이 Mask R-CNN에서와 동일한 평균 이진 교차 엔트로피 손실이다.  $L_{maskiou}^C$ 는 [56]에서와 같이 MaskIoU를 회귀시키기 위한 L2 손실이다.

일반적으로 다중 스케일 특징의 GT를 사용할 수 없기 때문에 본 발명의 실시예에 따른 SRF-GAN에서는 검출기에 대해 사전 훈련된 다중 스케일 특징 추출기(예를 들어, BiFPN, PAFPN, FPN)를 안내 특징 네트워크로 활용하는 것이다. 그런 다음 SRF 생성기가 내장된 BiFPN과 사전 훈련된 BiFPN 사이의 각 레벨에서 특징 불일치를 최소화한다. 이 다중 스케일 특징 학습을 통해 SRF 생성기가 주어진 입력 특징을 확대할 때 특징 로컬리티와 판별성을 보존하고 증가시킬 수 있다. 반면 기존 GAN 및 SR 방법은 입력 영상을 확대한 후 데이터 분포를 보존하는 데 더 중점을 둔다.

하지만, SRF 생성기가 특정 특징 스케일 레벨에 쉽게 과적합되는 경향이 있기 때문에 일반화된 SRF 생성기를 학습하는 것은 매우 어렵다. 그 결과 최종 검출 결과도 저하된다. 또한 효율성이 검출기의 핵심 요소이기 때문에 SRF 생성기는 더 경량 네트워크가 되도록 설계되어야 한다. 이러한 과제를 해결하기 위해 본 발명에서는 효율성을 위한 몇 가지 컨볼루션으로 새로운 SRF 구조를 제안한다. 또한, 과적합 방지를 위한 도 4의 알고리즘에 나타난 점진적 학습 체계를 제안하고, 다중 작업 학습을 위한 SRF-GAN 손실 식(1), 식(2), 식(3)을 제안한다. 본 발명의 실시예에 따른 사전 훈련 모델이 제공되면 SRF 생성기 훈련 절차를 생략할 수 있다.

도 5는 본 발명의 일 실시예에 따른 SRF-GAN의 검출 결과를 비교한 도면이다.

도 5(a)는 NN 보간, 도 5(b)는 제안하는 SRF-GAN의 검출 결과를 나타낸다. 도 5(a) 및 도 5(b) 모드 12 COCO 에포크(epoch) 동안 R-50-FPN이 있는 Mask R-CNN을 기반으로 훈련되었다. 모든 이미지는 COCO 테스트-개발 세트를 이용하였다.

도 6은 본 발명의 일 실시예에 따른 5개의 잔여 블록이 있는 SRF 생성기GAN를 나타내는 도면이다.

도 6의 잔여블록 아키텍처에 대한 세부 정보를 위해  $3 \times 3$  해상도 및 Leaky ReLU 활성화 계층( $\alpha = 0.2$ )에 해상도의 특징 맵을 제공한다. 이후, 2개의  $3 \times 3$  컨볼루션, 2개의 배치 정규화 및 1개의 Leaky ReLU 계층으로 구성된 5개의 연속 잔여 블록을 추가하여 초해상도 표현을 더 자세히 알아본다. 그런 다음 하나의 컨볼루션과 하나의 디컨볼루션 블록을 따라 특징 해상도를 2의 배수로 확대한다. 채널 치수를 입력과 동일하게 만들기 위해  $3 \times 3$  컨볼루션 계층을 추가한다. 잔여 학습의 경우 이중선 보간법에 의해 디컨볼루션 특징과 확대된 입력 특징 사이에 바로가기 연결이 추가된다.

- [0087] 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.
- [0088] 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(embodiment)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.
- [0089] 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.
- [0090] 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.
- [0091] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.
- [0093] <참고자료>
- [0094] [1] S.-H. Bae, "Object detection based on region decomposition and assembly," in AAAI, vol. 33, 2019, pp. 8094-8101.
- [0095] [2] K. Chen, J. Pang, J. Wang, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Shi, W. Ouyang et al., "Hybrid task cascade for instance segmentation," in CVPR, 2019, pp. 4974-4983.
- [0096] [3] Y. Lee and J. Park, "Centermask: Real-time anchor-free instance segmentation," in CVPR, 2020, pp. 13 906-13 915.
- [0097] [4] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks

for object detection," in CVPR, 2017, pp. 2117-2125.

- [0098] [5] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in CVPR, 2018, pp. 8759-8768.
- [0099] [6] M. Tan, R. Pang, and Q. V. Le, "Efficientdet: Scalable and efficient object detection," in CVPR, 2020, pp. 10 781-10 790.
- [0100] [7] G. Ghiasi, T.-Y. Lin, and Q. V. Le, "Nas-fpn: Learning scalable feature pyramid architecture for object detection," in CVPR, 2019, pp. 7036- 7045.
- [0101] [8] C. Guo, B. Fan, Q. Zhang, S. Xiang, and C. Pan, "Augfpn: Improving multi-scale feature learning for object detection," in CVPR, 2020, pp. 12 595-12 604.
- [0102] [9] Y. Li, Y. Chen, N. Wang, and Z. Zhang, "Scale-aware trident networks for object detection," in ICCV, 2019, pp. 6054-6063.
- [0103] [10] J. Pang, K. Chen, J. Shi, H. Feng, W. Ouyang, and D. Lin, "Libra rcnn: Towards balanced learning for object detection," in CVPR, 2019, pp. 821-830.
- [0104] [11] S. Qiao, L.-C. Chen, and A. Yuille, "Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution," arXiv preprint arXiv:2006.02334, 2020.
- [0105] [12] H. Xu, L. Yao, W. Zhang, X. Liang, and Z. Li, "Auto-fpn: Automatic network architecture adaptation for object detection beyond classification," in ICCV, 2019, pp. 6649-6658.
- [0106] [13] Z. Wang, J. Chen, and S. C. Hoi, "Deep learning for image super-resolution: A survey," IEEE TPAMI, 2020.
- [0107] [14] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in ICCV, 2017, pp. 2980-2988.
- [0108] [15] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask r-cnn," in ICCV, 2017, pp. 2961-2969.
- [16] Z. Tian, C. Shen, H. Chen, and T. He, "Fcos: Fully convolutional one-stage object detection," in ICCV, 2019, pp. 9627-9636.
- [0109] [17] R. Girshick, "Fast r-cnn," in ICCV, 2015, pp. 1440-1448.
- [0110] [18] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in CVPR, 2014, pp. 580-587.
- [0111] [19] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," IEEE TPAMI, vol. 37, no. 9, pp. 1904-1916, 2015.
- [0112] [20] S. Ren, K. He, R. Girshick, and J. Sun, "Faster-cnn: Towards real-time object detection with region proposal networks," in NIPS, 2015, pp. 91-99.
- [0113] [21] X. Wang, T. E. Huang, J. Gonzalez, T. Darrell, and F. Yu, "Frustratingly simple few-shot object detection," in ICML, 2020, pp. 9919-9928.
- [0114] [22] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," IJCV, vol. 88, no. 2, pp. 303-338, 2010.
- [0115] [23] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein et al., "Imagenet large scale visual recognition challenge," IJCV, vol. 115, no. 3, pp. 211-252, 2015.
- [0116] [24] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L. Zitnick, "Microsoft coco: Common objects in context," in ECCV. Springer, 2014, pp. 740-755.
- [0117] [25] Z. Cai and N. Vasconcelos, "Cascade r-cnn: Delving into high quality object detection," in CVPR, 2018, pp. 6154-6162.

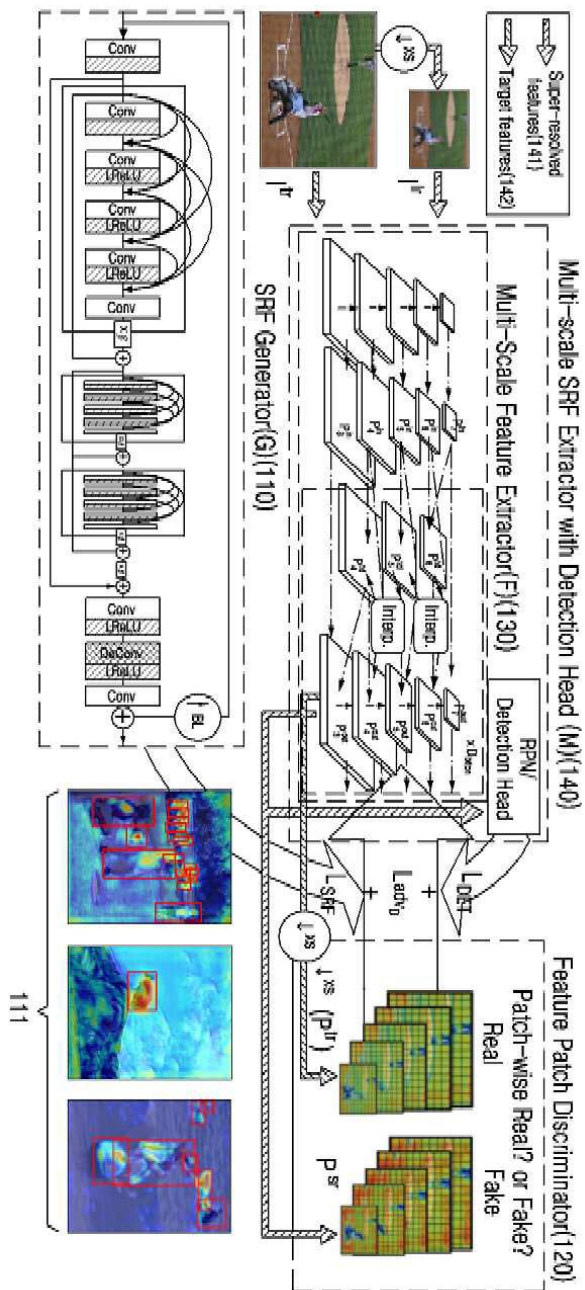
- [0118] [26] L. Yang, Q. Song, Z. Wang, M. Hu, and C. Liu, "Hier r-cnn: Instancelevel human parts detection and a new benchmark," IEEE Transactions on Image Processing, vol. 30, pp. 39-54, 2020.
- [0119] [27] W. Dong, P. Roy, C. Peng, and V. Isler, "Ellipse r-cnn: Learning to infer elliptical object from clustering and occlusion," IEEE Transactions on Image Processing, vol. 30, pp. 2193-2206, 2021.
- [0120] [28] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in ECCV. Springer, 2016, pp. 21-37.
- [0121] [29] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "Yolact: Real-time instance segmentation," in ICCV, 2019, pp. 9157-9166.
- [0122] [30] C.-Y. Fu, M. Shvets, and A. C. Berg, "Retinamask: Learning to predict masks improves state-of-the-art single-shot detection for free," arXiv preprint arXiv:1901.03353, 2019.
- [0123] [31] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in CVPR, 2017, pp. 7263-7271.
- [0124] [32] H. Law and J. Deng, "Cornersnet: Detecting objects as paired keypoints," in ECCV. Springer, 2018, pp. 734-750.
- [0125] [33] X. Lu, B. Li, Y. Yue, Q. Li, and J. Yan, "Grid r-cnn," in CVPR, 2019, pp. 7363-7372.
- [0126] [34] T. Kong, F. Sun, H. Liu, Y. Jiang, L. Li, and J. Shi, "Foveabox: Beyond anchor-based object detection," IEEE Transactions on Image Processing, vol. 29, pp. 7389-7398, 2020.
- [0127] [35] S. Zhang, C. Chi, Y. Yao, Z. Lei, and S. Z. Li, "Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection," in CVPR, 2020, pp. 9759-9768.
- [0128] [36] B. Singh and L. S. Davis, "An analysis of scale invariance in object detection snip," in CVPR, 2018, pp. 3578-3587.
- [0129] [37] M. Najibi, B. Singh, and L. S. Davis, "Autofocus: Efficient multi-scale inference," in ICCV, 2019, pp. 9745-9755.
- [0130] [38] T. Kong, A. Yao, Y. Chen, and F. Sun, "Hypernet: Towards accurate region proposal generation and joint object detection," in CVPR, 2016, pp. 845-853.
- [0131] [39] S. Bell, C. Lawrence Zitnick, K. Bala, and R. Girshick, "Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks," in CVPR, 2016, pp. 2874-2883.
- [0132] [40] B. Hariharan, P. Arbelaez, R. Girshick, and J. Malik, "Object instance segmentation and fine-grained localization using hypercolumns," IEEE TPAMI, vol. 39, no. 4, pp. 627-639, 2016.
- [0133] [41] B. Zoph and Q. V. Le, "Neural architecture search with reinforcement learning," in ICLR, 2017.
- [0134] [42] W. Nie, T. Karras, A. Garg, S. Debnath, A. Patney, A. B. Patel, and A. Anandkumar, "Semi-supervised stylegan for disentanglement learning," in ICML, 2020, pp. 7360-7369.
- [0135] [43] U. Hwang, D. Jung, and S. Yoon, "Hexagan: Generative adversarial nets for real world classification," in ICML, 2019, pp. 2921-2930.
- [0136] [44] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang et al., "Photo-realistic single image super-resolution using a generative adversarial network," in CVPR, 2017, pp. 4681-4690.
- [0137] [45] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in ECCV Workshops, 2018.
- [0138] [46] T. R. Shaham, T. Dekel, and T. Michaeli, "Singan: Learning a generative model from a single natural image," in ICCV, 2019, pp. 4570-4580.
- [0139] [47] J. Rabbi, N. Ray, M. Schubert, S. Chowdhury, and D. Chao, "Small-object detection in remote sensing images with end-to-end edge-enhanced gan and object detector network," Remote Sensing, vol.

12, no. 9, p. 1432, 2020.

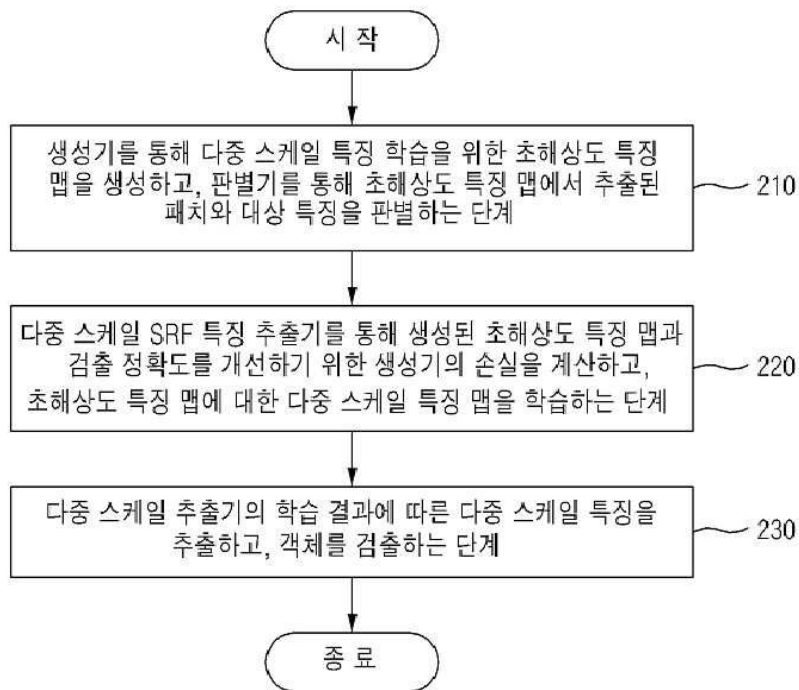
- [0140] [48] Y. Bai, Y. Zhang, M. Ding, and B. Ghanem, "Sod-mtgan: Small object detection via multi-task generative adversarial network," in ECCV. Springer, 2018, pp. 206-221.
- [0141] [49] J. Li, X. Liang, Y. Wei, T. Xu, J. Feng, and S. Yan, "Perceptual generative adversarial networks for small object detection," in CVPR, 2017, pp. 1222-1230.
- [0142] [50] J. Noh, W. Bae, W. Lee, J. Seo, and G. Kim, "Better to follow, follow to be better: towards precise supervision of feature super-resolution for small object detection," in ICCV, 2019, pp. 9725-9734.
- [0143] [51] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in ICML, vol. 30, no. 1, 2013, p. 3.
- [0144] [52] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in AAAI, vol. 31, no. 1, 2017.
- [0145] [53] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in CVPR, 2017, pp. 1125-1134.
- [0146] [54] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in ICML, 2015, pp. 448-456.
- [0147] [55] J. Yu, Y. Jiang, Z. Wang, Z. Cao, and T. Huang, "Unitbox: An advanced object detection network," in Proceedings of the 24th ACM international conference on Multimedia, 2016, pp. 516-520.
- [0148] [56] Z. Huang, L. Huang, Y. Gong, C. Huang, and X. Wang, "Mask scoring r-cnn," in CVPR, 2019, pp. 6409-6418.
- [0149] [57] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in CVPR, 2016, pp. 770-778.
- [0150] [58] H. Zhang, C. Wu, Z. Zhang, Y. Zhu, Z. Zhang, H. Lin, Y. Sun, T. He, J. Mueller, R. Manmatha et al., "Resnest: Split-attention networks," arXiv preprint arXiv:2004.08955, 2020.
- [0151] [59] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in CVPR, 2017, pp. 1492- 1500.
- [0152] [60] W. Ke, T. Zhang, Z. Huang, Q. Ye, J. Liu, and D. Huang, "Multiple anchor learning for visual object detection," in CVPR, 2020, pp. 10 206- 10 215.
- [0153] [61] Z. Dong, G. Li, Y. Liao, F. Wang, P. Ren, and C. Qian, "Centripetalnet: Pursuing high-quality keypoint pairs for object detection," in CVPR, 2020, pp. 10 519-10 528.
- [0154] [62] G. Song, Y. Liu, and X. Wang, "Revisiting the sibling head in object detector," in CVPR, 2020, pp. 11 563-11 572.
- [0155] [63] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, "Detectron2," <https://github.com/facebookresearch/detectron2>, 2019.
- [0156] [64] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in ICCV, 2017, pp. 618-626.

도면

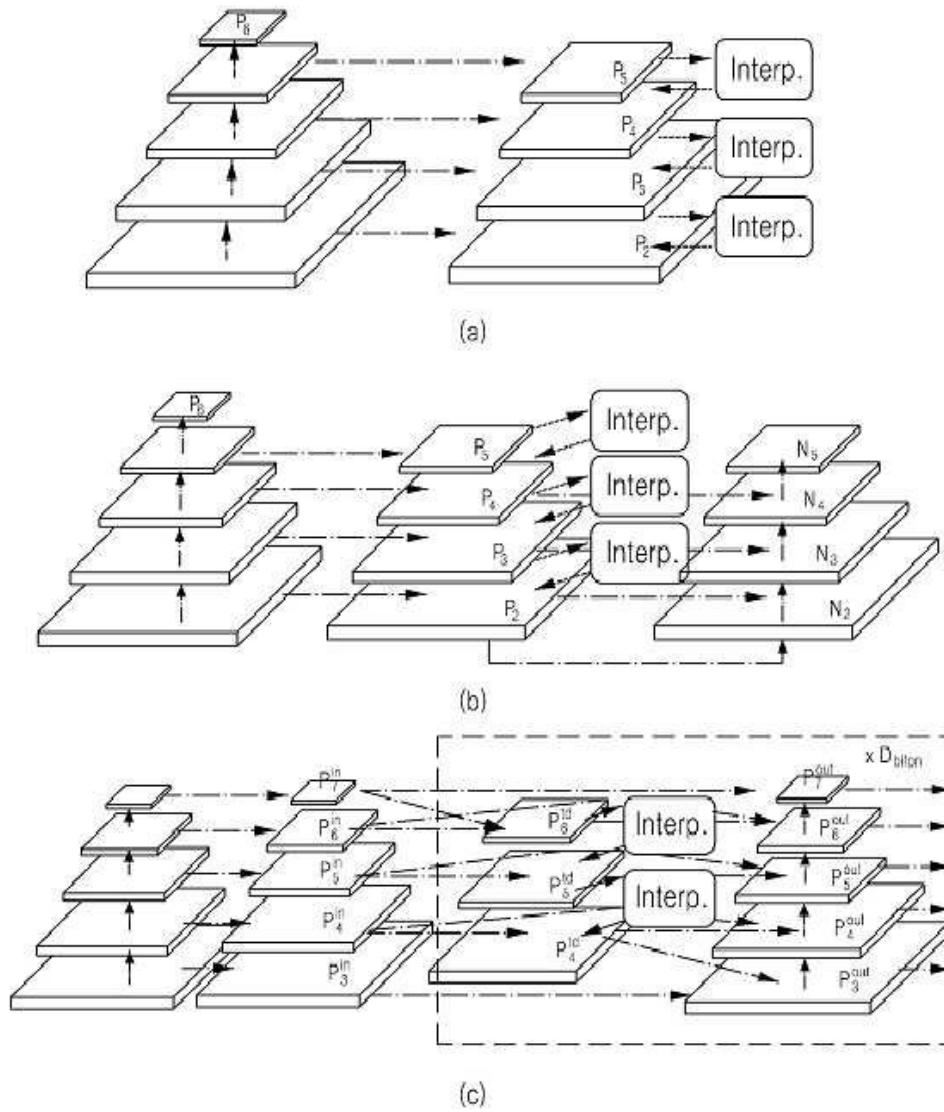
도면1



도면2



도면3



도면4

---

|                                                                                                      |                                                                                                      |
|------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------|
| <b>Algorithm 1:</b> Proposed SRF-GAN training                                                        |                                                                                                      |
| <b>Input</b> : Mini-batch of image set $I = \{I_i^{tr}, I_i^{lr}\}_{i=1}^N$<br>and pre-trained $F$ . |                                                                                                      |
| <b>Output:</b> Trained SRF generator $G$ and detector $T$                                            |                                                                                                      |
| 1                                                                                                    | // Sec IV-A : Learning $G$ and $D$ while freezing $F$                                                |
| 2                                                                                                    | <b>for</b> $k_1 = 1, \dots, K_1$ <b>do</b>                                                           |
| 3                                                                                                    | Extract $P_{k_1}^{sr}, P_{k_1}^{tr} \leftarrow G(F(I_{k_1}^{lr})), F(I_{k_1}^{tr})$                  |
| 4                                                                                                    | Train $D$ by maximizing $L_{adv_D}(P_{k_1}^{sr}, P_{k_1}^{tr})$ in Eq. (1)                           |
| 5                                                                                                    | Train $G$ by minimizing $L_{SRF}(P_{k_1}^{sr}, P_{k_1}^{tr})$ in Eq. (2)                             |
| 6                                                                                                    | <b>end</b>                                                                                           |
| 7                                                                                                    | // Sec IV-B: Learning $M$ embedded $G$ and $D$ while<br>freezing $F$                                 |
| 8                                                                                                    | <b>for</b> $k_2 = 1, \dots, K_2$ <b>do</b>                                                           |
| 9                                                                                                    | Extract $P_{k_2}^{sr}, P_{k_2}^{tr} \leftarrow M(G(I_{k_2}^{lr})), F(I_{k_2}^{tr})$                  |
| 10                                                                                                   | Train $D$ by maximizing $L_{adv_D}(P_{k_2}^{sr}, \downarrow^{\times s}(P_{k_2}^{tr}))$ by<br>Eq. (1) |
| 11                                                                                                   | Train $M$ by minimizing $L_{INT}(P_{k_2}^{sr}, \downarrow^{\times s}(P_{k_2}^{tr}))$ by<br>Eq. (3)   |
| 12                                                                                                   | <b>end</b>                                                                                           |
| 13                                                                                                   | // Sec IV-C: Learning $T$ with $G$                                                                   |
| 14                                                                                                   | <b>for</b> $k_3 = 1, \dots, K_3$ <b>do</b>                                                           |
| 15                                                                                                   | Train $T$ by minimizing $L_{DET}(T(G(I_{k_3}^{tr})))$ by Eq.<br>(4), Eq. (5), Eq. (6), or Eq. (7)    |
| 16                                                                                                   | <b>end</b>                                                                                           |

---



도면6

