[별첨 5]

경량 객체 및 장면 인식 기술

ETRI 한국전자통신연구원
www.etri.re.kr
통신미디어연구소 차세대콘텐츠연구본부

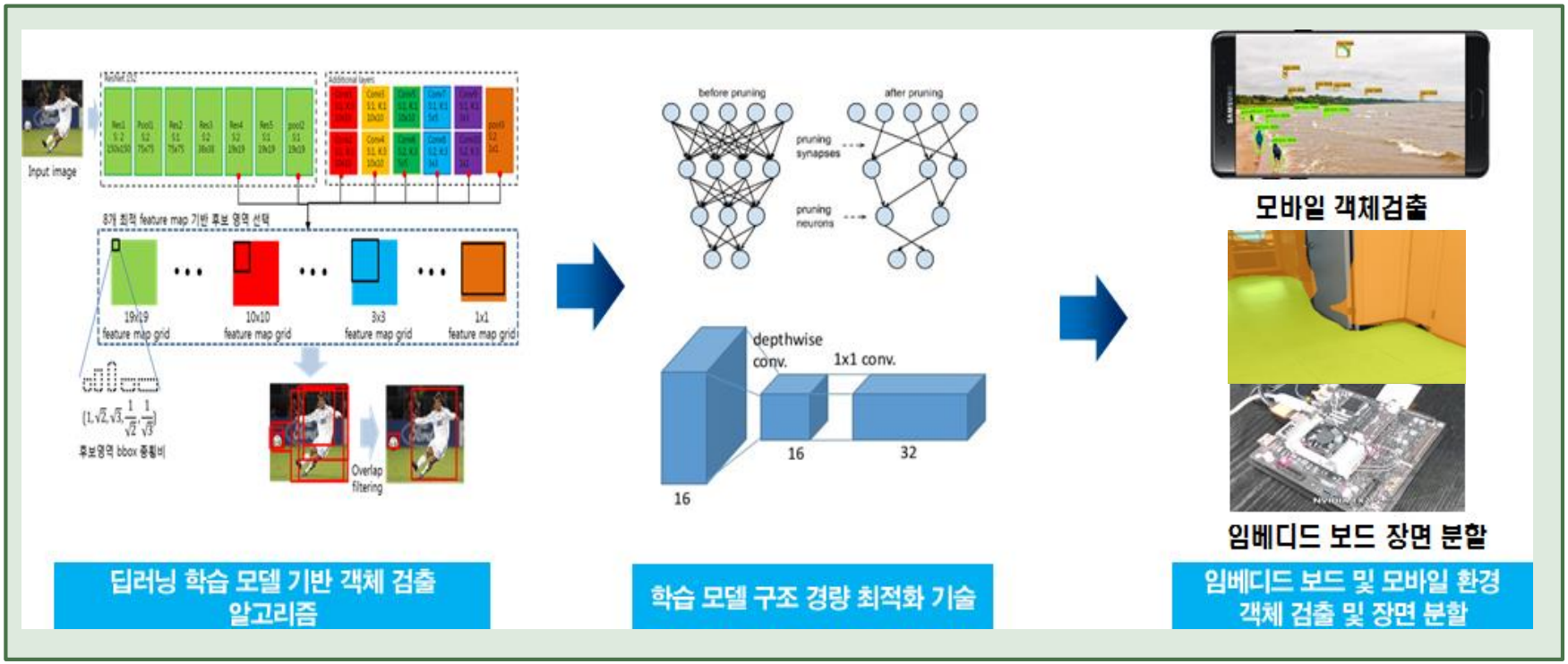
목 차

1. 기술의 개요
2. 기술이전 내용 및 범위
3. 경쟁기술과 비교
4. 기술의 사업성
5. 국내외 시장 동향

1. 기술의 개요

▣ 경량 객체 장면 인식 기술

❖ 모바일 및 임베디드 보드 환경에서 객체를 검출하고 장면영역을 인식하도록 하는 경량화된 객체/장면 인식 기술 기술



<얼굴 학습 이미지 데이터 생성 기술 개념도>

1. 기술의 개요

▣ 기술 특징 및 장점

- ❖ 딥러닝을 기반객체 검출 및 장면영역 분할 기술로 모바일 및 임베디드 보드 환경에서 동작 운용이 가능한 형태로 인식 모델을 경량 최적화
- ❖ 저사향의 환경에서 객체를 탐지하고 장면 영역을 분할을 가능하게 하는 기술
- ❖ 모바일/임베디드 환경 객체 인식 지원
  - ❖ 객체(80개 class) 검출 지원
- ❖ 모바일/임베디드 환경 장면 인식 지원
  - ❖ 실내(바닥, 벽, 천정 등) 장면 영역 인식 지원

2. 기술이전 내용 및 범위

▣ 기술이전 내용

- ❖ 경량 객체인식 기술
- ❖ 경량 장면 인식 기술

▣ 기술이전 범위

- ❖ 경량 객체인식 모델(model(바이너리))
- ❖ 경량 장면인식 모델(model(바이너리))
- ❖ 요구사항 정의서 및 설계서
- ❖ 경량 객체/장면인식 시험절차서 및 결과서

2. 기술이전 내용 및 범위

▣ 기술 개발 현황

❖ 기술성숙도(TRL : Technology Readiness Level) 단계 : (5)단계

구 분	단계	정 의	세 부 설 명
기초 연구 단계	1	기초 이론/실험	∘기초이론 정립 단계
	2	실용목적의아이디어, 특허 등 개념정립	∘기술개발개념정립및아이디어에대한특허출원단계
실험 단계	3	실험실 규모의 기본성능 검증	∘실험실 환경에서 실험 또는 전산 시뮬레이션을 통해 기본성능이 검증될 수 있는 단계 ∘개발하려는부품/시스템의기본설계도면을확보하는단계
	4	실험실 규모의 소재/부품/시스템 핵심성능 평가	∘시험샘플을제작하여핵심성능에대한평가가완료된단계 ∘3단계에서 도출된 다양한 결과 중에서 최적의 결과를 선택하려는 단계 ∘컴퓨터 모사가 가능한 경우 최적화를 완료하는 단계
시작품 단계	5	확정된 소재/부품/시스템시작품제작 및 성능 평가	∘확정된 소재/부품/시스템의 실험실 시작품 제작 및 성능 평가가 완료된 단계 ∘개발 대상의 생산을 고려하여 설계하나 실제 제작한 시작품 샘플은 1~수개 미만인 단계 ∘경제성을 고려하지 않고 기술의 핵심성능으로만 볼 때, 실제로 판매가 될 수 있는 정도로 목표 성능을 달성한 단계
	6	파일롯 규모 시작품 제작 및 성능 평가	∘파일롯 규모(복수 개~양산규모의 1/10정도)의 시작품 제작 및 평가가 완료된 단계 ∘파일롯규모생산품에대해생산량,생산용량,불량률등제시 ∘파일롯 생산을 위한 대규모 투자가 동반되는 단계 ∘생산기업이 수요기업 적용환경에 유사하게 자체 현장테스트를 실시하여 목표 성능을 만족시킨 단계 ∘성능평가결과에대해가능하면공인인증기관의성적서확보
실용화 단계	7	신뢰성평가 및 수요기업 평가	∘실제 환경에서 성능 검증이 이루어지는 단계 ∘부품 및 소재개발의 경우 수요업체에서 직접 파일롯 시작품을 현장 평가(성능 및 신뢰성 평가) ∘가능하면 인증기관의 신뢰성 평가 결과 제출
	8	시제품 인증 및 표준화	∘표준화 및 인허가 취득 단계
사업화	9	사업화	∘본격적인 양산 및 사업화 단계 ∘6-시그마 등 품질관리가 중요한 단계

3. 경쟁기술과 비교

▣ 휴대폰 및 임베디드 보드 환경의 객체 및 장면 인식 기술

❖ 기술의 특징

- 서버환경의 딥러닝 객체 및 장면 인식기술에서 더 나아가 응용 확장성이 더 높은 모바일 환경에서의 객체 및 장면 인식기술의 요구 증가
- 기존 서버 환경의 딥러닝 기반 객체 및 장면인식 기술은 모바일 환경에 동작 운용이 어려움
- 딥러닝 기반 모델 구조 경량 및 최적화를 통해 휴대폰 및 임베디드 보드 환경에서 객체 인식 및 장면 영역 인식 기능을 제공

4. 기술의 사업성

▣ 경량 객체 및 장면 인식 기술

적용분야	적용 분야 세부내용
모바일 AR 서비스	❖ 객체정보 및 장면정보가 융합되는 실가상 정합의 모바일 AR 서비스에 활용 가능
스마트 드론 서비스	❖ 드론에서 카메라를 통해 영상정보 뿐만아니라 검출되는 객체 및 장면 정보를 제공하는 스마트 드론 서비스 활용 가능
지능형 감시 서비스	❖ CCTV 에지 카메라에서 동작 운용 가능한 객체 탐지 및 장면 분할을 통해 지능형 감시 에지 서비스에 활용 가능함
로봇 자율주행 응용 서비스	❖ 실내외 로봇의 자율주행에 있어서 주변 객체 정보 및 장면 정보 분석을 통한 응용서비스에 활용 가능함

5. 국내외 시장 동향

▣ 국내외 시장 동향

❖ 국외

- 구글은 모바일/임베디드 기기용 딥러닝 Tensorflow 경량화 버전인 Tensorflow Lite공개 하였으며 페이스북은 모바일용 Pytorch 개발 서비스 환경을 제공함
- Qualcomm은 Snapdragon 칩을 사용하는 모바일 디바이스용 기계학습 SDK인 Snapdragon NPE (Neural Processing Engine) SDK를 공개하였으며, Caffe, Caffe2, Tensorflow 모델 사용을 지원
- 구글은 안면인식이 제공되는 스마트 AI 스피커 제품을 선보이며 서버가 아닌 단말 형태의 AI 영상 분석 기술 시장에 관심을 증대하고 있음

❖ 국내

- 모바일/임베디드 디바이스 기반 국제저전력영상인식대회 (LPIRC 2018)에서 한국전자통신연구원이 1위를 기록함
- 기존 서버 기반의 영상감시 서비스에서 에지컴퓨팅 기반의 카메라에서의 영상분석 시장의 요구와 자동차 및 로봇 주행에서의 경량 영상분석 시장 요구가 증가하고 있음

감사합니다.

www.etri.re.kr

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0106104
(43) 공개일자 2020년09월11일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2006.01) *G06K 9/00* (2006.01)
G06N 3/04 (2006.01)
(52) CPC특허분류
G06N 3/082 (2013.01)
G06K 9/00362 (2013.01)
(21) 출원번호 10-2019-0021310
(22) 출원일자 2019년02월22일
심사청구일자 없음

(71) 출원인
한국전자통신연구원
대전광역시 유성구 가정로 218 (가정동)
(72) 발명자
이승재
대전광역시 유성구 상대로 17, 309동 701호
고종국
대전광역시 유성구 상대로 17, 306동 1201호
(뒷면에 계속)
(74) 대리인
특허법인이상

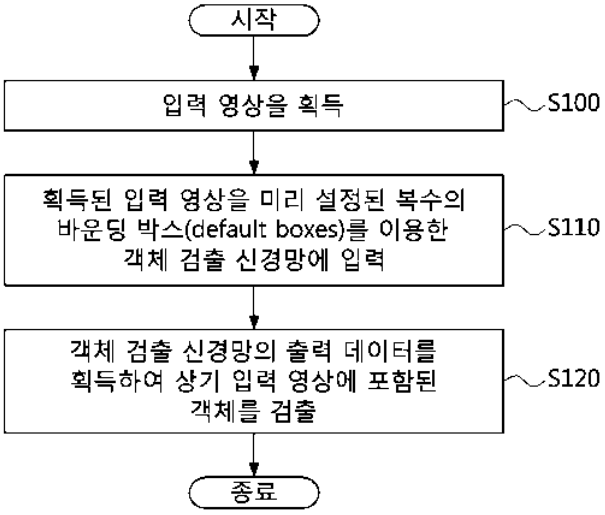
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 **인공 신경망을 이용한 고속 객체 검출 방법 및 장치**

(57) 요 약

인공 신경망을 이용한 고속 객체 검출 방법 및 장치가 개시된다. 인공 신경망을 이용한 고속 객체 검출 방법은, 입력 영상을 획득하는 단계, 획득된 입력 영상을 미리 설정된 복수의 바운딩 박스(default boxes)를 이용한 객체 검출 신경망에 입력하는 단계 및 상기 객체 검출 신경망의 출력 데이터를 획득하여 상기 입력 영상에 포함된 객체를 검출하는 단계를 포함한다.

대 표 도 - 도10



(52) CPC특허분류
G06N 3/04 (2013.01)

(72) 발명자
이근동
대전광역시 유성구 은구비남로 55, 709동 2101호
이수웅
세종특별자치시 새롬중앙로 90, 1007동 302호
유원영
대전광역시 유성구 어은로 57, 138동 1301호

이 발명을 지원한 국가연구개발사업

과제고유번호	2018-0-00198
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원(IITP)
연구사업명	ETRI연구개발지원사업
연구과제명	객체추출 및 실-가상 정합 지원 모바일 AR 기술 개발
기 여 율	1/1
과제수행기관명	한국전자통신연구원
연구기간	2018.02.01 ~ 2018.12.31

명 세 서

청구범위

청구항 1

인공 신경망을 이용한 고속 객체 검출 방법에서,

입력 영상을 획득하는 단계;

획득된 입력 영상을 미리 설정된 복수의 바운딩 박스(default boxes)를 이용한 객체 검출 신경망에 입력하는 단계; 및

상기 객체 검출 신경망의 출력 데이터를 획득하여 상기 입력 영상에 포함된 객체를 검출하는 단계를 포함하고,

상기 객체 검출 신경망은,

기본 네트워크(Base Network);

상기 기본 네트워크의 출력을 수신하는 추가 계층들(Additional Layers); 및

상기 추가 계층들 또는 상기 기본 네트워크의 출력을 수신하는 검출 계층들(Detection Layers)을 포함하는, 인공 신경망을 이용한 고속 객체 검출 방법.

청구항 2

청구항 1에서,

상기 기본 네트워크는,

VGG, ResNet, ResNext, Mobilenet 중 하나인, 인공 신경망을 이용한 고속 객체 검출 방법.

청구항 3

청구항 1에서,

상기 기본 네트워크는,

3×3 컨볼루셔널 계층(Convolutional Layer); 및

2×2 맥스풀링 계층(Maxpooling Layer)을 포함하는, 인공 신경망을 이용한 고속 객체 검출 방법.

청구항 4

청구항 3에서,

상기 기본 네트워크는,

상기 2×2 맥스풀링 계층의 출력을 수신하는 depthwise 기반의 3×3 컨볼루셔널 계층; 및

상기 depthwise 기반의 3×3 컨볼루셔널 계층의 출력을 수신하는 1×1 컨볼루셔널 계층을 더 포함하는, 인공 신경망을 이용한 고속 객체 검출 방법.

청구항 5

청구항 1에서,

상기 추가 계층들은,

depthwise 기반의 3×3 컨볼루셔널 계층; 및

상기 depthwise 기반의 3×3 컨볼루셔널 계층의 출력에 연결된 1×1 컨볼루셔널 계층을 포함하는, 인공 신경망을 이용한 고속 객체 검출 방법.

청구항 6

청구항 1에서,

상기 검출 계층들은,

상기 객체의 위치를 검출하는 1×1 컨볼루셔널 계층; 및

상기 객체의 종류를 식별하는 1×1 컨볼루셔널 계층을 포함하는, 인공 신경망을 이용한 고속 객체 검출 방법.

청구항 7

청구항 5에서,

상기 depthwise 기반의 3×3 컨볼루셔널 계층은,

합성곱을 수행하는 간격으로서, 스트라이드(stride) 값이 2로 설정되는, 인공 신경망을 이용한 고속 객체 검출 방법.

청구항 8

청구항 1에서,

상기 추가 계층들은,

상기 기본 네트워크의 출력을 수신하여 특징을 추출하는 제1 추가 계층; 및

상기 제1 추가 레이어의 출력을 수신하여 특징을 추출하는 제2 추가 계층을 포함하는, 인공 신경망을 이용한 고속 객체 검출 방법.

청구항 9

청구항 8에서,

상기 검출 계층들은,

상기 기본 네트워크의 출력을 수신하여 상기 객체의 위치와 종류를 식별하는 제1 검출 계층; 및

상기 제1 추가 계층의 출력을 수신하여 상기 객체의 위치와 종류를 식별하는 제2 검출 계층을 포함하는, 인공 신경망을 이용한 고속 객체 검출 방법.

청구항 10

청구항 8에서,

상기 제1 추가 계층은,

depthwise 기반의 3×3 컨볼루셔널 계층 및 상기 depthwise 기반의 3×3 컨볼루셔널 계층의 출력에 연결된 1×1 컨볼루셔널 계층으로 구성되는 구조를 적어도 두 개 포함하는, 인공 신경망을 이용한 고속 객체 검출 방법.

청구항 11

인공 신경망을 이용한 고속 객체 검출 장치로서,

적어도 하나의 프로세서(processor); 및

상기 적어도 하나의 프로세서가 적어도 하나의 단계를 수행하도록 지시하는 명령어들(instructions)을 저장하는 메모리(memory)를 포함하고,

상기 적어도 하나의 단계는,

입력 영상을 획득하는 단계;

획득된 입력 영상을 미리 설정된 복수의 바운딩 박스(default boxes)를 이용한 객체 검출 신경망에 입력하는 단계; 및

상기 객체 검출 신경망의 출력 데이터를 획득하여 상기 입력 영상에 포함된 객체를 검출하는 단계를 포함하고,

상기 객체 검출 신경망은,

기본 네트워크(Base Network);

상기 기본 네트워크의 출력을 수신하는 추가 계층들(Additional Layers); 및

상기 추가 계층들 또는 상기 기본 네트워크의 출력을 수신하는 검출 계층들(Detection Layers)을 포함하는, 인공 신경망을 이용한 고속 객체 검출 장치.

청구항 12

청구항 11에서,

상기 기본 네트워크는,

VGG, ResNet, ResNext, Mobilenet 중 하나인, 인공 신경망을 이용한 고속 객체 검출 장치.

청구항 13

청구항 11에서,

상기 기본 네트워크는,

3×3 컨볼루셔널 계층(Convolutional Layer); 및

2×2 맥스풀링 계층(Maxpooling Layer)을 포함하고, 인공 신경망을 이용한 고속 객체 검출 장치.

청구항 14

청구항 13에서,

상기 기본 네트워크는,

상기 2×2 맥스풀링 계층의 출력을 수신하는 depthwise 기반의 3×3 컨볼루셔널 계층; 및

상기 depthwise 기반의 3×3 컨볼루셔널 계층의 출력을 수신하는 1×1 컨볼루셔널 계층을 더 포함하는, 인공 신경망을 이용한 고속 객체 검출 장치.

청구항 15

청구항 11에서,

상기 추가 계층들은,

depthwise 기반의 3×3 컨볼루셔널 계층; 및

상기 depthwise 기반의 3×3 컨볼루셔널 계층의 출력에 연결된 1×1 컨볼루셔널 계층을 포함하는, 인공 신경망을 이용한 고속 객체 검출 장치.

청구항 16

청구항 11에서,

상기 검출 계층들은,

상기 객체의 위치를 검출하는 1×1 컨볼루셔널 계층; 및

상기 객체의 종류를 식별하는 1×1 컨볼루셔널 계층을 포함하는, 인공 신경망을 이용한 고속 객체 검출 장치.

청구항 17

청구항 15에서,

상기 depthwise 기반의 3×3 컨볼루셔널 계층은,

합성곱을 수행하는 간격으로서, 스트라이드(stride) 값이 2로 설정되는, 인공 신경망을 이용한 고속 객체 검출 장치.

청구항 18

청구항 11에서,

상기 추가 계층들은,

상기 기본 네트워크의 출력을 수신하여 특징을 추출하는 제1 추가 계층; 및

상기 제1 추가 레이어의 출력을 수신하여 특징을 추출하는 제2 추가 계층을 포함하는, 인공 신경망을 이용한 고속 객체 검출 장치.

청구항 19

청구항 18에서,

상기 검출 계층들은,

상기 기본 네트워크의 출력을 수신하여 상기 객체의 위치와 종류를 식별하는 제1 검출 계층; 및

상기 제1 추가 계층의 출력을 수신하여 상기 객체의 위치와 종류를 식별하는 제2 검출 계층을 포함하는, 인공 신경망을 이용한 고속 객체 검출 장치.

청구항 20

청구항 18에서,

상기 제1 추가 계층은,

depthwise 기반의 3×3 컨볼루셔널 계층 및 상기 depthwise 기반의 3×3 컨볼루셔널 계층의 출력에 연결된 1×1 컨볼루셔널 계층으로 구성되는 구조를 적어도 두 개 포함하는, 인공 신경망을 이용한 고속 객체 검출 장치.

발명의 설명

기 술 분 야

[0001] 본 발명은 인공 신경망을 이용한 고속 객체 검출 방법 및 장치에 관한 것으로, 더욱 상세하게는 미리 설정된 복수의 바운딩 박스를 이용하는 단일 단계 기반의 객체 검출 네트워크에서 연산속도를 더욱 향상시킬 수 있는 인공 신경망 구조를 이용하여 객체를 검출하는 기술에 관한 것이다.

배 경 기 술

[0002] 객체 검출(Object Detection) 기술은 로봇, 비디오 감시, 자동차 안전 등과 같은 여러 응용 분야에서 널리 사용되고 있는 핵심 기술이다. 최근에는, 객체 검출 기술에 인공 신경망(Artificial Neural Network) 또는 합성 곱 신경망(convolutional neural network, CNN)을 사용하는 방식이 알려짐에 따라, 단일 영상을 이용한 객체 검출 기술은 비약적으로 발전하였다.

[0003] 객체 검출 기술은 영상 내에서 특정 위치에 따른 객체를 식별하는 기술로서, 객체 분류(Object Classification) 기술과는 달리 객체의 위치와 식별을 동시에 추정해야 하고 영상 내에서 검출하고자 하는 모든 객체를 식별해야 한다.

[0004] 합성 곱 신경망을 이용한 객체 검출 방법은 영역 추출(RoI pooling, region of interest pooling)을 기반으로 하는 기법과 격자 공간(Grid Cell)을 기반으로 하는 기법으로 분류할 수 있다. 이때, 영역 추출을 기반으로 하는 방법은 입력된 영상에 대하여 합성 곱 신경망을 이용하여 영상 전체에 대한 특징 맵을 추출하고, 추출된 특징 맵에서 후보 영역(Region proposal)을 추출한 후, 후보 영역에 대한 특징 또는 픽셀을 리샘플링(resampling)하고, 분류(classification)함으로써 객체를 식별한다.

[0005] 한편, 기존의 영역 추출을 기반으로 하는 기법은 하나의 특징 맵에서 서로 다른 크기를 갖는 객체를 검출할 수 있도록 바운딩 박스(bounding box)의 크기 추정, 위치 추정 등이 필수적이고 이 때문에 많은 속도 저하가 발생

한다.

[0006] 따라서, 합성 곱 신경망을 이용한 객체 검출 방법에서 고속으로 객체를 검출할 수 있는 방법이 필요한 실정이다.

발명의 내용

해결하려는 과제

[0007] 상기와 같은 문제점을 해결하기 위한 본 발명의 목적은, 인공 신경망을 이용한 고속 객체 검출 방법을 제공하는 데 있다.

[0008] 상기와 같은 문제점을 해결하기 위한 본 발명의 다른 목적은, 인공 신경망을 이용한 고속 객체 검출 장치를 제공하는 데 있다.

과제의 해결 수단

[0009] 상기 목적을 달성하기 위한 본 발명은, 인공 신경망을 이용한 고속 객체 검출 방법을 제공한다.

[0010] 인공 신경망을 이용한 고속 객체 검출 방법은, 입력 영상을 획득하는 단계, 획득된 입력 영상을 미리 설정된 복수의 바운딩 박스(default boxes)를 이용한 객체 검출 신경망에 입력하는 단계 및 상기 객체 검출 신경망의 출력 데이터를 획득하여 상기 입력 영상에 포함된 객체를 검출하는 단계를 포함할 수 있다.

[0011] 상기 객체 검출 신경망은, 기본 네트워크(Base Network), 상기 기본 네트워크의 출력을 수신하는 추가 계층들(Additional Layers) 및 상기 추가 계층들 또는 상기 기본 네트워크의 출력을 수신하는 검출 계층들(Detection Layers)을 포함할 수 있다.

[0012] 상기 기본 네트워크는, VGG, ResNet, ResNext, Mobilenet 중 하나일 수 있다.

[0013] 상기 기본 네트워크는, 3×3 컨볼루셔널 계층(Convolutional Layer) 및 2×2 맥스풀링 계층(Maxpooling Layer)을 포함할 수 있다.

[0014] 상기 기본 네트워크는, 상기 2×2 맥스풀링 계층의 출력을 수신하는 depthwise 기반의 3×3 컨볼루셔널 계층 및 상기 depthwise 기반의 3×3 컨볼루셔널 계층의 출력을 수신하는 1×1 컨볼루셔널 계층을 더 포함할 수 있다.

[0015] 상기 추가 계층들은, depthwise 기반의 3×3 컨볼루셔널 계층 및 상기 depthwise 기반의 3×3 컨볼루셔널 계층의 출력에 연결된 1×1 컨볼루셔널 계층을 포함할 수 있다.

[0016] 상기 검출 계층들은, 상기 객체의 위치를 검출하는 1×1 컨볼루셔널 계층 및 상기 객체의 종류를 식별하는 1×1 컨볼루셔널 계층을 포함할 수 있다.

[0017] 상기 depthwise 기반의 3×3 컨볼루셔널 계층은, 합성곱을 수행하는 간격으로서, 스트라이드(stride) 값이 2로 설정될 수 있다.

[0018] 상기 추가 계층들은, 상기 기본 네트워크의 출력을 수신하여 특징을 추출하는 제1 추가 계층 및 상기 제1 추가 레이어의 출력을 수신하여 특징을 추출하는 제2 추가 계층을 포함할 수 있다.

[0019] 상기 검출 계층들은, 상기 기본 네트워크의 출력을 수신하여 상기 객체의 위치와 종류를 식별하는 제1 검출 계층 및 상기 제1 추가 계층의 출력을 수신하여 상기 객체의 위치와 종류를 식별하는 제2 검출 계층을 포함할 수 있다.

[0020] 상기 제1 추가 계층은, depthwise 기반의 3×3 컨볼루셔널 계층 및 상기 depthwise 기반의 3×3 컨볼루셔널 계층의 출력에 연결된 1×1 컨볼루셔널 계층으로 구성되는 구조를 적어도 두 개 포함할 수 있다.

[0021] 상기 목적을 달성하기 위한 본 발명의 다른 측면은, 인공 신경망을 이용한 고속 객체 검출 장치를 제공한다.

[0022] 인공 신경망을 이용한 고속 객체 검출 장치는, 적어도 하나의 프로세서(processor) 및 상기 적어도 하나의 프로세서가 적어도 하나의 단계를 수행하도록 지시하는 명령어들(instructions)을 저장하는 메모리(memory)를 포함할 수 있다.

[0023] 상기 적어도 하나의 단계는, 입력 영상을 획득하는 단계, 획득된 입력 영상을 미리 설정된 복수의 바운딩 박스(default boxes)를 이용한 객체 검출 신경망에 입력하는 단계 및 상기 객체 검출 신경망의 출력 데이터를 획득

하여 상기 입력 영상에 포함된 객체를 검출하는 단계를 포함할 수 있다.

[0024] 상기 객체 검출 신경망은, 기본 네트워크(Base Network), 상기 기본 네트워크의 출력을 수신하는 추가 계층들(Additional Layers) 및 상기 추가 계층들 또는 상기 기본 네트워크의 출력을 수신하는 검출 계층들(Detection Layers)을 포함할 수 있다.

[0025] 상기 기본 네트워크는, VGG, ResNet, ResNext, Mobilenet 중 하나일 수 있다.

[0026] 상기 기본 네트워크는, 3×3 컨볼루셔널 계층(Convolutional Layer) 및 2×2 맥스풀링 계층(Maxpooling Layer)을 포함할 수 있다.

[0027] 상기 기본 네트워크는, 상기 2×2 맥스풀링 계층의 출력을 수신하는 depthwise 기반의 3×3 컨볼루셔널 계층 및 상기 depthwise 기반의 3×3 컨볼루셔널 계층의 출력을 수신하는 1×1 컨볼루셔널 계층을 더 포함할 수 있다.

[0028] 상기 추가 계층들은, depthwise 기반의 3×3 컨볼루셔널 계층 및 상기 depthwise 기반의 3×3 컨볼루셔널 계층의 출력에 연결된 1×1 컨볼루셔널 계층을 포함할 수 있다.

[0029] 상기 검출 계층들은, 상기 객체의 위치를 검출하는 1×1 컨볼루셔널 계층 및 상기 객체의 종류를 식별하는 1×1 컨볼루셔널 계층을 포함할 수 있다.

[0030] 상기 depthwise 기반의 3×3 컨볼루셔널 계층은, 합성곱을 수행하는 간격으로서, 스트라이드(stride) 값이 2로 설정될 수 있다.

[0031] 상기 추가 계층들은, 상기 기본 네트워크의 출력을 수신하여 특징을 추출하는 제1 추가 계층 및 상기 제1 추가 레이어의 출력을 수신하여 특징을 추출하는 제2 추가 계층을 포함할 수 있다.

[0032] 상기 검출 계층들은, 상기 기본 네트워크의 출력을 수신하여 상기 객체의 위치와 종류를 식별하는 제1 검출 계층 및 상기 제1 추가 계층의 출력을 수신하여 상기 객체의 위치와 종류를 식별하는 제2 검출 계층을 포함할 수 있다.

[0033] 상기 제1 추가 계층은, depthwise 기반의 3×3 컨볼루셔널 계층 및 상기 depthwise 기반의 3×3 컨볼루셔널 계층의 출력에 연결된 1×1 컨볼루셔널 계층으로 구성되는 구조를 적어도 두 개 포함할 수 있다.

발명의 효과

[0034] 상기와 같은 본 발명에 따른 인공 신경망을 이용한 고속 객체 검출 방법 및 장치를 이용할 경우에는 메모리와 연산량에서 효율적인 객체 검출이 가능할 수 있다.

[0035] 또한, 검출하려는 객체의 개수와 분류하는 후보의 개수가 많을수록 종래의 방식보다 고속으로 객체를 검출할 수 있는 장점이 있다.

[0036] 또한, 메모리와 연산능력이 낮은 하드웨어 장비에서도 빠르게 객체를 검출할 수 있다.

도면의 간단한 설명

[0037] 도 1은 객체 검출을 위해 사용되는 기본 네트워크 구조를 설명하기 위한 예시도이다.

도 2는 본 발명의 일 실시예에 따른 객체 검출을 위해 추가 계층을 결합한 SSD(Single shot multibox detector)의 원리를 설명하기 위한 예시도이다.

도 3은 단일 단계 방식에 따른 객체 검출 방법에 적용되는 네트워크 구조를 설명하기 위한 예시도이다.

도 4는 도 2에 따른 SSD(Single shot multibox detector)의 네트워크 구현례이다.

도 5는 본 발명의 일 실시예에 따른 인공 신경망을 이용한 고속 객체 검출 방법에서 사용할 수 있는 추가 계층의 구성도이다.

도 6은 본 발명의 일 실시예에 따른 인공 신경망을 이용한 고속 객체 검출 방법에서 사용될 수 있는 검출 계층의 구성도이다.

도 7은 본 발명의 일 실시예에 따른 인공 신경망을 이용한 고속 객체 검출 방법에 따른 추가 계층과 검출 계층을 포함하는 객체 검출 네트워크의 구성도이다.

도 8 내지 도 9는 일 실시예에 따른 인공 신경망을 이용한 고속 객체 검출 방법의 메모리와 계산량을 다른 방식들과 비교한 그래프이다.

도 10은 본 발명의 일 실시예에 따른 인공 신경망을 이용한 고속 객체 검출 방법에 대한 흐름도이다.

도 11은 본 발명의 일 실시예에 따른 인공 신경망을 이용한 고속 객체 검출 장치에 대한 구성도이다.

발명을 실시하기 위한 구체적인 내용

[0038] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 각 도면을 설명하면서 유사한 참조부호를 유사한 구성요소에 대해 사용하였다.

[0039] 제1, 제2, A, B 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. 및/또는 이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.

[0040] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다.

[0041] 본 출원에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0042] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥 상 가지는 의미와 일치하는 의미를 가지는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0043] 이하, 본 발명에 따른 바람직한 실시예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0045] 도 1은 객체 검출을 위해 사용되는 기본 네트워크 구조를 설명하기 위한 예시도이다.

[0046] 도 1을 참조하면, 객체 검출을 위해 보편적으로 사용되는 기본 네트워크의 구조를 확인할 수 있다. 구체적으로 객체 검출을 위한 기본 네트워크(base network)은, 미리 설정된 크기의 이미지를 입력 이미지로 입력받아, 특징 맵을 추출하는 컨볼루셔널 계층, 추출된 특징에 대해 샘플링을 수행하는 풀링 계층 등을 포함할 수 있다.

[0047] 컨볼루셔널 계층은 입력 영상의 특징을 추출하는 필터, 필터의 값을 비선형 값으로 바꾸는 활성화 함수(activation function) 및 풀링 계층(pooling layer) 중 적어도 하나를 포함할 수 있다. 필터는 일종의 행렬로 표현되는 입력 영상의 특징적 부분을 검출하는 함수로서, 일반적으로 행렬로 표현된다. 여기서 행렬로 표현된 입력 영상과 필터를 서로 합성 곱함으로써, 객체의 특징을 추출할 수 있는데, 여기서 추출된 특징은 특징 맵(feature map)으로 지칭될 수도 있다. 또한, 합성곱을 수행하는 간격 값을 스트라이드(stride)라고 지칭할 수 있는데, 스트라이드 값에 따라 다른 크기의 특징 맵이 추출될 수 있다. 이때, 특징 맵은 필터의 크기가 입력 영상보다 작으면, 기존의 입력 영상보다 더 작은 크기를 갖게 되는데, 여러 단계를 거쳐 특징이 소실되는 것을 방지하기 위하여 패딩 과정이 추가로 수행될 수 있다. 이때, 패딩 과정은 생성된 특징 맵의 외곽에 미리 설정된 값(예를 들면 0)을 추가함으로써 입력 영상의 크기와 특징 맵의 크기를 동일하게 유지하는 과정일 수 있다.

[0048] 활성화 함수는 어떠한 값(또는 행렬)으로 추출된 특징을 비선형 값으로 바꾸는 함수로서, 시그모이드(sigmoid)

함수, ReLU 함수, 소프트맥스(softmax) 함수 등이 사용될 수 있다. 예를 들어, 소프트맥스 함수는 입력된 값을 0~1사이의 값으로 모두 정규화하며 출력 값들의 총합은 항상 1이 되는 특성을 가진 함수일 수 있다.

[0049] 풀링 계층은 추출된 특징맵에 대하여 서브 샘플링(subsampling) 또는 풀링(pooling)을 수행하여 특징맵을 대표하는 특징을 선정하는 계층으로서, 특징맵의 일정 영역에 대하여 가장 큰 값을 추출하는 맥스 풀링(max pooling), 평균값을 추출하는 애버리지 풀링(average pooling) 등이 수행될 수 있다. 이때, 풀링 계층은 활성화 함수 이후에 반드시 수행되는 것이 아니라 선택적으로 수행될 수 있다.

[0050] 도 1을 참조하면, 보편적으로 사용되는 VGG16 의 신경망 구조를 확인할 수 있다. VGG16은, 224×224 크기를 갖는 입력 이미지를 이용하여 컨볼루셔널 계층과와 ReLU 함수를 이용해 특징 맵을 추출하고, 추출된 특징맵에 대해 맥스풀링을 수행하여 더 작은 크기의 특징 맵으로 변환하는 과정이 반복하여 수행될 수 있고, 완전 접속 계층(fully connected layer) 및 소프트맥스 함수를 거쳐 최종적으로 1000개의 라벨에 대한 확률을 획득할 할 수 있다.

[0051] 여기서는 기존에 공개된 VGG 를 예로 들었으나, 기본 네트워크에는 ResNet, ResNext, Mobilenet 등이 사용될 수도 있다. 또한, 그 밖에 기본 네트워크의 구성과 동작에 대해서는 본 발명이 속하는 기술분야에서 통상의 기술자가 용이하게 이해할 수 있을 것이므로 구체적인 설명은 생략한다.

[0052] 또한, 본 발명에서 입력 데이터 사용되는 영상은 정지 영상, 픽처, 동영상, 이미지 등으로 혼용하여 지칭될 수 있다.

[0054] 도 2는 본 발명의 일 실시예에 따른 객체 검출을 위해 추가 계층을 결합한 SSD(Single shot multibox detector)의 원리를 설명하기 위한 예시도이다.

[0055] 앞선 도 1에 따른 네트워크는 하나의 이미지에 대해 특징을 추출하는 데 사용되는 데, 입력 이미지에서 객체를 검출하기 위해서는 별도의 크기를 갖는 입력 영상을 입력하거나 추출된 특징에 대하여 후보 영역(또는 바운딩 박스)을 검출할 수 있는 모듈을 별도로 학습해야 한다. 이러한 문제를 해결하는 방안으로서 바운딩 박스(또는 후보 영역)를 검출하는 모듈을 학습하는 과정을 생략하고 미리 정해진 후보 영역을 사용함으로써 검출 속도를 향상시키는 방법을 고려할 수 있다.

[0056] 도 2의 도면기호 20을 참조하면, 하나의 정지 영상 내에서 서로 크기가 다른 두개의 객체가 있는 것을 확인할 수 있다. 이때, 하나의 특징 맵에서 두개의 객체를 각각 검출할 경우, 바운딩 박스(또는 후보 영역)의 크기를 객체에 따라 달리 설정해야 하므로 바운딩 박스의 크기를 추정하고, 위치를 추정하는 과정이 추가로 필요할 수 있다. 이때, 특징 맵을 서로 다른 크기를 갖도록 복수개 생성하고, 각 셀(또는 블록)의 크기가 작게 분할된 특징 맵에서는 작은 객체를 검출하고, 각 셀의 크기가 크게 분할된 특징 맵에서 큰 객체를 검출함으로써 검색 속도를 향상시킬 수 있다.

[0057] 도면 기호 21을 참조하면, 8X8의 크기를 갖는 특징 맵을 확인할 수 있고, 도면 기호 22를 참조하면 4×4의 크기를 갖는 특징 맵을 확인할 수 있다. 이때, 도면 기호 22를 참조하면, 모든 객체의 카테고리들(c1, c2, c3, 쩐, cp)에 대한 신뢰도(conf)와 바운딩 박스가 각 셀에 대하여 가질 수 있는 상대적 위치 및 크기를 나타내는 오프셋 값(offset, △로 표시)을 예측함으로써 객체를 식별할 수 있다. 이때, SSD는 바운딩 박스가 가질 수 있는 크기 비율(aspect ratio)를 미리 복수개 설정해두고 설정된 복수개의 크기 비율을 갖는 바운딩 박스들(default boxes) 중에서 객체와 상응하는 크기의 바운딩 박스를 추정할 수 있다.

[0058] 여기서, SSD의 훈련 방법을 비롯한 객체 검출 방식은 선행 논문(W. Liu, D. Anguelov, D. Erhan, C. Szegedy, and S. E. Reed. SSD: single shot multibox detector. CoRR, abs/1512.02325,2015.)을 통해 본 발명이 속하는 기술분야에서 통상의 기술자가 용이하게 이해할 수 있으므로, 자세한 설명은 생략한다.

[0059] SSD는 후보 영역을 학습, 검출하는 과정을 생략하고, 미리 정의된 바운딩 박스들을 사용하여 객체의 종류와 위치를 식별하는 방식이기 때문에, 단일 단계 방식의 객체 검출 기법에 대한 예시일 수 있다. 하지만 기존의 단일 단계 방식의 객체 검출 기법 또한 많은 바운딩 박스를 사용함에 따른 속도 저하, 메모리 부족 문제가 발생할 수 있다.

[0060] 따라서, 이하에서는 단일 단계 방식에 따른 검출 속도를 향상시키고 메모리를 적게 사용하기 위하여 본 발명의 일 실시예에 따른 인공 신경망을 이용한 고속 객체 검출 방법을 설명한다.

[0062] 도 3은 단일 단계 방식에 따른 객체 검출 방법에 적용되는 네트워크 구조를 설명하기 위한 예시도이다. 도 4는 도 2에 따른 SSD(Single shot multibox detector)의 네트워크 구현례이다.

[0063] 도 3을 참조하면, 단일 단계 방식에 따른 객체 검출 방법에 적용되는 네트워크 구조는 기본 네트워크(31), 기본 네트워크의 출력을 수신하는 추가 계층들(additional layers, 32), 및 추가계층들(32)의 출력과 기본 네트워크(31)의 출력을 각각 수신하는 두개 이상의 검출 계층들(detection layers, 33)을 포함할 수 있다.

[0064] 따라서, 단일 단계 방식에 따른 객체 검출 방법은 기본 네트워크(31)를 통과시켜 얻은 특징 맵을 이용하여 서로 다른 크기의 특징 맵을 추가로 추출하는 추가 계층들(additional layers, 32)에 통과시킨 후, 추가 계층들과 기본 네트워크로부터 전달받은 특징을 이용하여 객체의 위치와 분류를 결정하는 복수의 검출 계층들(detection layers, 33)에 통과시켜 최종적으로 객체를 검출할 수 있다.

[0065] 한편, 도 4를 참조하면, 도 2에서 설명한 SSD는, 도 1에서 설명한 VGG16을 기본 네트워크로 가지며, 기본 네트워크의 후단에 서로 다른 크기를 갖는 복수의 추가 계층들(Extra Feature Layers) 및 검출 계층들(Detections)을 포함할 수 있다. 이때, 컨볼루셔널 계층의 표기 형식에서 3×3X1024를 예로 들면, 크기가 3×3이고 채널이 1024임을 의미할 수 있다.

[0066] 구체적으로, 도 4에 따른 SSD는, 기본 네트워크와 연결된 추가 계층들(Extra Feature Layers)로서, 1×1 컨볼루셔널 계층과 3×3 컨볼루셔널 계층을 순차로 연결하여 사용하고 있고, 검출 계층으로는 3×3 컨볼루셔널 계층을 사용하고 있는 것을 확인할 수 있다.

[0067] 한편, 일반적인 컨볼루셔널 계층은 맵 전체를 훑어가면서 컨볼루션 연산을 수행하므로 계산량이 상당히 많다. 이러한 문제를 해결하기 위하여 depthwise Separable Convolution을 사용할 수 있다. Depthwise Separable Convolution은 depthwise 방식의 컨볼루션으로 입력층 채널마다 컨볼루션을 수행하여 보관하고, 보관된 채널마다의 출력값에 1×1 pointwise 컨볼루션을 수행하여 각 채널의 데이터를 종합함으로써 일반적인 컨볼루셔널 계층의 연산과 유사한 출력을 만들어낼 수 있다.

[0068] 이하에서, SSDLite는 위와 같은 depthwise Separable Convolution을 이용하여, 1×1 컨볼루셔널 계층, depthwise 기반의 3×3 컨볼루셔널 계층 및 pointwise 기반의 1×1 컨볼루셔널 계층을 순차로 연결함으로써 추가 계층들을 구성함으로써 SSD를 변경한 방식일 수 있다.

[0069] 여기서 depthwise 기반의 3×3 컨볼루셔널 계층은 채널을 제외하고, 공간 방향으로 3×3 크기의 커널(또는 필터)로 컨볼루션을 수행하는 계층을 의미할 수 있고, pointwise 기반의 컨볼루셔널 계층은 1×1의 크기를 갖는 커널(또는 필터)로 채널 방향에 대하여 컨볼루션을 수행하는 계층을 의미할 수 있으며 간략히 1×1 컨볼루셔널 계층으로 지칭될 수 있다.

[0071] 도 5는 본 발명의 일 실시예에 따른 인공 신경망을 이용한 고속 객체 검출 방법에서 사용할 수 있는 추가 계층의 구성도이다.

[0072] 도 5를 참조하면, 본 발명의 일 실시예에서는 단일 단계 방식의 객체 검출 방법에서 검출 속도를 더욱 향상시키기위하여 도 3의 추가 계층들(32)을 기존과 달리 구성할 수 있다. 즉, 본 발명의 일 실시예에 따른 추가 계층들(32)은 depthwise 기반의 3×3 컨볼루셔널 계층(32a) 및 상기 depthwise 기반의 3×3 컨볼루셔널 계층의 출력에 연결된 1×1 컨볼루셔널 계층(32b)을 포함할 수 있다.

[0073] 추가 레이어로 추출된 영상의 특징은 추가 레이어의 크기를 고려하여 후보 영역을 기존의 방법과 같이 regular grid로 구성하여 후보 영역을 구성 후 검출 레이어를 통해 학습할 수도 있고, 추가 레이어 내의 학습 데이터의 위치 정보의 중심값(k-means) 등을 활용하여 후보 영역을 구성할 수도 있다.

[0074]

[0075] 도 6은 본 발명의 일 실시예에 따른 인공 신경망을 이용한 고속 객체 검출 방법에서 사용될 수 있는 검출 계층의 구성도이다.

[0076] 본 발명의 일 실시예에서는 단일 단계 방식의 객체 검출 방법의 검출 속도를 향상시키기위하여 위치(location)와 분류(class)를 추정하는 계층으로, 1×1 컨볼루셔널 계층을 사용할 수 있다.

[0077] 구체적으로 도 3에 따른 검출 계층들(33)에서, 검색한 객체(또는 바운딩 박스)의 위치를 검출하는 계층으로, 1×1 컨볼루셔널 계층(33a)을 사용하고, 검색된 객체를 식별(또는 분류)하여 검출하는 계층으로, 1×1 컨볼루셔널 계층(33b)을 사용한 것을 확인할 수 있다. 즉, 객체의 위치를 검출하는 계층과 객체를 식별하는 계층은 서로 병렬로 구성될 수 있다.

[0078] 각 추가 계층과 검출 계층을 통한 결과는 분류 정확도와 객체의 검출 정보 (BBOX)를 활용하여 기존에 활용하는 NMS, soft-NMS등의 방법으로 다시 정제될 수 있다.

[0079] 도 5 및 도 6에서 추가 계층들과 검출 계층들의 개수는 복수개가 사용될 수 있으며, 추가 레이어들과 검출 레이어들의 개수에 따라 바운딩 박스(또는 후보 영역)의 크기와 개수는 변동될 수 있다.

[0080] 이하에서는 도 5에 따른 추가 계층 및 도 6에 따른 검출 계층 구조를 갖는 네트워크를 eSSD로 지칭하고, 상세히 설명한다.

[0082] 도 7은 본 발명의 일 실시예에 따른 인공 신경망을 이용한 고속 객체 검출 방법에 따른 추가 계층과 검출 계층을 포함하는 객체 검출 네트워크의 구성도이다.

[0083] 도 7을 참조하면, 도 5와 도 6에 따른 추가 계층 및 검출 계층을 갖는 객체 검출 네트워크(eSSD)를 확인할 수 있다.

[0084] 도 7에 따른 객체 검출 네트워크에서는, 입력 이미지(Input Image)가 기본 네트워크(Base Network)에 입력된 이후에 추가 계층들(Additional Layers)과 검출 계층들(Detection Layers)로 각각 전달될 수 있다.

[0085] 먼저 기본 네트워크에서 채널이 N이고 크기가 B×B인 출력 이미지가 도출되면 기본 네트워크에서 도출된 출력 이미지는 제1 추가 계층(eSSD Additional Layer1) 및 제1 검출 계층(eSSD Detection Layer1)으로 입력될 수 있다. 제1 검출 계층(eSSD Detection Layer1)은 기본 네트워크의 출력에 대하여 1×1 컨볼루셔널 계층을 이용하여 객체 위치와 객체 분류 결과를 도출할 수 있다. 즉, 제1 검출 계층은 기본 네트워크에 의해 추출된 특징을 이용하여 객체 위치와 객체 분류를 수행하는 계층일 수 있다.

[0086] 제1 추가 계층(eSSD Additional Layer1)에서 채널이 2N이고 크기가 (B/2)×(B/2)인 출력 이미지가 도출되면 여기서 도출된 출력 이미지는 제2 추가 계층(eSSD Additional Layer2) 및 제2 검출 계층(eSSD Detection Layer2)으로 입력될 수 있다. 제2 검출 계층(eSSD Detection Layer2)은 제1 추가 계층(eSSD Additional Layer1)의 출력에 대하여 1×1 컨볼루셔널 계층을 이용하여 객체 위치와 객체 분류 결과를 도출할 수 있다. 즉, 제2 검출 계층(eSSD Detection Layer2)은 제1 추가 계층(eSSD Additional Layer1)에 의해 추출된 특징을 이용하여 객체 위치와 객체 분류를 수행하는 계층일 수 있다.

[0087] 제2 추가 계층(eSSD Additional Layer2)에서 채널이 N이고 크기가 (B/4)×(B/4)인 출력 이미지가 도출되면 여기서 도출된 출력 이미지는 제3 추가 계층(eSSD Additional Layer3) 및 제3 검출 계층(eSSD Detection Layer3)으로 입력될 수 있다. 제3 검출 계층(eSSD Detection Layer3)은 제2 추가 계층(eSSD Additional Layer2)의 출력에 대하여 1×1 컨볼루셔널 계층을 이용하여 객체 위치와 객체 분류 결과를 도출할 수 있다. 즉, 제3 검출 계층(eSSD Detection Layer3)은 제2 추가 계층(eSSD Additional Layer2)에 의해 추출된 특징을 이용하여 객체 위치와 객체 분류를 수행하는 계층일 수 있다.

[0088] 제3 추가 계층(eSSD Additional Layer3)에서 채널이 N/2이고 크기가 (B/8)×(B/8)인 출력 이미지가 도출되면 여기서 도출된 출력 이미지는 제4 추가 계층(eSSD Additional Layer4) 및 제4 검출 계층(eSSD Detection Layer4)으로 입력될 수 있다. 제4 검출 계층(eSSD Detection Layer4)은 제3 추가 계층(eSSD Additional Layer3)의 출력에 대하여 1×1 컨볼루셔널 계층을 이용하여 객체 위치와 객체 분류 결과를 도출할 수 있다. 즉, 제4 검출 계층(eSSD Detection Layer4)은 제3 추가 계층(eSSD Additional Layer3)에 의해 추출된 특징을 이용하여 객체 위치와 객체 분류를 수행하는 계층일 수 있다.

[0089] 제4 추가 계층(eSSD Additional Layer4)에서 채널이 N/2이고 크기가 (B/16)×(B/16)인 출력 이미지가 도출되면 여기서 도출된 출력 이미지는 제5 검출 계층(eSSD Detection Layer5)로 입력될 수 있다. 제5 검출 계층(eSSD Detection Layer5)은 제4 추가 계층(eSSD Additional Layer1)의 출력에 대하여 1×1 컨볼루셔널 계층을 이용하여 객체 위치와 객체 분류 결과를 도출할 수 있다. 즉, 제5 검출 계층(eSSD Detection Layer5)은 제4 추가 계층(eSSD Additional Layer4)에 의해 추출된 특징을 이용하여 객체 위치와 객체 분류를 수행하는 계층일 수 있다.

[0090] 제1 추가 계층(eSSD Additional Layer1), 제2 추가 계층(eSSD Additional Layer2), 제3 추가 계층(eSSD Additional Layer3) 및 제4 추가 계층(eSSD Additional Layer4)은 각각 Depthwise 기반의 3×3컨볼루셔널 계층(이때 스트라이드-stride는 2) 및 1×1 컨볼루셔널 계층이 순서대로 연결됨으로써 구성될 수 있다. 즉, 추가계층들(Additional Layers)은 각각 Depthwise 기반의 3×3컨볼루셔널 계층(이때 스트라이드-stride는 2) 및 1×1 컨볼루셔널 계층이 순서대로 연결된 구조를 하나 이상 가질 수 있다.

[0091] 또한, 제1 검출 계층(eSSD Detection Layer1), 제2 검출 계층(eSSD Detection Layer2), 제3 검출 계층(eSSD Detection Layer3), 제4 검출 계층(eSSD Detection Layer4) 및 제5 검출 계층(eSSD Detection Layer5)은 1×1 컨볼루셔널 계층으로 이루어진 객체 분류(Box class) 계층 및 1×1 컨볼루셔널 계층으로 이루어진 객체 위치 검출(Box Coordinate) 계층이 병렬로 연결됨으로써 구성될 수 있다.

[0092] 제1 검출 계층(eSSD Detection Layer1), 제2 검출 계층(eSSD Detection Layer2), 제3 검출 계층(eSSD Detection Layer3), 제4 검출 계층(eSSD Detection Layer4) 및 제5 검출 계층(eSSD Detection Layer5)으로부터 객체 분류 결과(Box Classes, 또는 객체 식별 결과) 및 객체 위치 검출 결과(Box Coordinate, 또는 바운딩 박스의 위치 식별 결과)가 각각 도출될 수 있다.

[0093] 제1 검출 계층(eSSD Detection Layer1), 제2 검출 계층(eSSD Detection Layer2), 제3 검출 계층(eSSD Detection Layer3), 제4 검출 계층(eSSD Detection Layer4) 및 제5 검출 계층(eSSD Detection Layer5)으로부터 도출된 객체 분류 결과와 객체 위치 검출 결과는 비최대 억제 계층(Non-Max Suppression)으로 입력될 수 있다. 여기서 비최대 억제 계층은 객체 분류 결과와 객체 위치 검출 결과에 대하여 확률값이 최대가 아닌 결과를 제외함으로써 최종 결과(Detection Results)를 도출할 수 있다. 예를 들어 비최대 억제 계층은 영상(또는 이미지)에서 객체와 중첩되는 바운딩 박스들 중에서 확률값이 가장 큰 바운딩 박스를 선택하고 나머지 바운딩 박스를 제외할 수 있다. 여기서 최종 결과는 검출된 객체를 표시하는 바운딩 박스의 위치값, 바운딩 박스 내의 객체에 대한 식별값을 포함할 수 있다.

[0094] 이때, 도 7에 따른 각 계층에서 도출되는 결과에서 크기와 채널의 수는 처리할 데이터에 따라 변동될 수 있는 것으로 해석되어야 하며 도면에 도시된 값으로 한정하여 해석되어서는 안된다. 또한, 추가 계층의 개수와 검출 계층의 개수 역시 예시적인 것으로 해석되어야 하며 구현례에 따라 변동될 수 있다.

[0095] 아래 표 1은 도 7에 따른 객체 검출 네트워크를 구현한 경우에 대한 상세 예시이다.

표 1

[0096]

Layer No.	Type / Stride	Filters	Input Size	output Size	Prediction filters
1	Conv / 2	3x3x3x16	224x224x3	112x112x16	
2	Maxpool /2	2x2x16x16	112x112x16	56x56x16	
3	Conv(dw) / 1	3x3x16dw	56x56x16	56x56x16	
4	Conv / 1	1x1x16x48	56x56x16	56x56x48	
5	Conv(dw) / 2	3x3x48w	56x56x48	28x28x48	
6	Conv / 1	1x1x48x128	28x28x48	28x28x128	
7	Conv(dw) / 1	3x3x128dw	28x28x128	28x28x128	
8	Conv / 1	1x1x128x192	28x28x128	28x28x192	
9	Conv(dw) / 2	3x3x192dw	28x28x192	14x14x192	
10	Conv / 1	1x1x192x384	14x14x192	14x14x384	
11-15	Conv(dw) / 1	3x3x384dw	14x14x384	14x14x384	
	Conv / 1	1x1x384	14x14x384	14x14x384	1x1x384x603 (cls) / 1x1x384x12 (loc)
16	Conv(dw) / 2	3x3x384dw	14x14x384	7x7x384	
17	Conv / 1	1x1x384x786	7x7x384	7x7x768	
18	Conv(dw) / 1	3x3x768dw	7x7x768	7x7x768	
19	Conv / 1	1x1x768x768	7x7x768	7x7x768	1x1x768x1206 (cls) / 1x1x768x24 (loc)
20	Conv(dw) / 2	3x3x768dw	7x7x768	7x7x768	
21	Conv / 1	1x1x768x512	4x4x768	4x4x512	1x1x512x1206 (cls) / 1x1x512x24 (loc)
22	Conv(dw) / 2	3x3x512dw	4x4x512	4x4x512	
23	Conv / 1	1x1x512x256	2x2x512	2x2x256	1x1x256x1206 (cls) / 1x1x256x24 (loc)
24	Conv(dw) / 2	3x3x256dw	2x2x256	1x1x256	
25	Conv / 1	1x1x256x256	1x1x256	1x1x256	1x1x256x1206 (cls) / 1x1x256x24 (loc)

[0097] 상기 표 1에 따라 구현된 전체 네트워크는 각각의 계층 번호(Layer No)에 따라 순서대로 연결될 수 있으며, 계층의 종류(Type)로서 일반적인 컨볼루셔널 레이어를 Conv로, depthwise 기반의 컨볼루셔널 레이어를

Conv(dw)로, 맥스풀링을 수행하는 레이어를 Maxpool로 표기하였다. 또한, 스트라이드 값(Stride)으로 2 또는 1이 적용되었으며, 각 계층에 따른 필터는 BXBXCXN 으로 표현되었을 때, BXB의 크기를 갖고, C의 채널수를 갖는 N개의 필터를 갖는 것을 의미한다. 또한, 입력 이미지의 크기(Input Size)와 출력 이미지의 크기(Output Size)는 BXBXC로 표현되었을 때 BXB의 크기를 갖고, C의 채널수를 갖는 것을 의미할 수 있다.

[0098] 표 1을 참조하면, 본 발명의 일 실시예에 따른 인공 신경망을 이용한 고속 객체 검출 방법에서의 기본 네트워크(Base Network)는, 1번부터 15번까지에 따른 계층을 서로 연결하여 구성될 수 있다. 표 1에 따른 기본 네트워크는 기존의 MobileNetV1을 변형하여 구현한 것으로 기본 네트워크의 초기 계층을 컨볼루셔널 계층과 맥스 풀링 계층(표 1의 2번 계층)으로 대체하는 점에서 차이가 있을 수 있다. 표1과 같이 변경할 경우 채널 개수가 줄어들고 전체적인 연산량 감소 효과가 있는 장점이 있다.

[0099] 즉, 표 1에 따른 기본 네트워크는 3×3 컨볼루셔널 계층 및 2×2 맥스풀링 계층을 포함하고, 상기 2×2 맥스풀링 계층의 후단으로, depthwise 기반의 3×3 컨볼루셔널 계층 및 1×1 컨볼루셔널 계층이 서로 순차적으로 연결된 구조를 더 포함할 수 있다. 이때, depthwise 기반의 3×3 컨볼루셔널 계층 및 1×1 컨볼루셔널 계층이 서로 순차적으로 연결된 구조는, 9개(표 1의 3번부터 15번)일 수 있다. 이때, 표 1에서 11-15번 계층은 depthwise 기반의 3×3 컨볼루셔널 계층 및 1×1 컨볼루셔널 계층이 순차적으로 연결된 구조를 하나의 번호로 표기하였다.

[0100] 표 1에서 기본 네트워크의 출력은 제1 추가 계층과 제1 검출 계층으로 전달될 수 있다. 제1 검출 계층은 표 1에서와 같이 1×1 컨볼루셔널 계층으로 이루어진, 객체 분류(1×1X384X603, cls) 계층 및 객체 위치 검출(1×1X384X12, loc) 계층으로 구성될 수 있다.

[0101] 또한, 제1 추가 계층은 표 1의 16~19번 계층이 서로 순차적으로 연결될 수 있다. 즉, 제1 추가 계층은 depthwise 기반의 3×3 컨볼루셔널 계층과 1×1 컨볼루셔널 계층이 순차로 연결된 구조가 2개 포함될 수 있다. 제1 추가 계층의 출력은 제2 추가 계층 및 제2 검출 계층으로 전달될 수 있다. 표 1을 참조하면, 제2 검출 계층은 1×1 컨볼루셔널 레이어로 이루어진, 객체 분류(1×1X768X1206) 계층 및 객체 위치 검출(1×1X768X24) 계층으로 구성될 수 있다.

[0102] 또한, 제2 추가 계층은 표 1의 20~21번 계층이 서로 순차적으로 연결될 수 있다. 즉, 제2 추가 계층은 depthwise 기반의 3×3 컨볼루셔널 계층과 1×1 컨볼루셔널 계층이 순차로 연결된 구조가 1개 포함될 수 있다. 제2 추가 계층의 출력은 제3 추가 계층 및 제3 검출 계층으로 전달될 수 있다. 표 1을 참조하면, 제3 검출 계층은 1×1 컨볼루셔널 계층으로 이루어진, 객체 분류(1×1X512X1206) 계층 및 객체 위치 검출(1×1X512×24) 계층으로 구성될 수 있다.

[0103] 또한, 제3 추가 계층은 표 1의 22~23번 계층이 서로 순차적으로 연결될 수 있다. 즉, 제3 추가 계층은 depthwise 기반의 3×3 컨볼루셔널 계층과 1×1 컨볼루셔널 계층이 순차로 연결된 구조가 1개 포함될 수 있다. 제3 추가 계층의 출력은 제4 추가 계층 및 제4 검출 계층으로 전달될 수 있다. 표 1을 참조하면, 제4 검출 계층은 1×1 컨볼루셔널 계층으로 이루어진, 객체 분류(1×1X256X1206) 계층 및 객체 위치 검출(1×1X256X24) 계층으로 구성될 수 있다.

[0104] 또한, 제4 추가 계층은 표 1의 24~25번 계층이 서로 순차적으로 연결될 수 있다. 즉, 제4 추가 계층은 depthwise 기반의 3×3 컨볼루셔널 계층과 1×1 컨볼루셔널 계층이 순차로 연결된 구조가 1개 포함될 수 있다. 제4 추가 계층의 출력은 제5 검출 계층으로 전달될 수 있다. 표 1을 참조하면, 제5 검출 계층은 1×1 컨볼루셔널 계층으로 이루어진, 객체 분류(1×1X256X1206) 계층 및 객체 위치 검출(1×1X256X24) 계층으로 구성될 수 있다.

[0105] 표 1의 제1 검출 계층 내지 제5 검출 계층에 따른 출력 데이터들은 도 7에 따른 Non-Max Suppression 계층으로 입력되어 최종 검출 결과가 도출될 수 있다.

[0107] 도 8 내지 도 9는 일 실시예에 따른 인공 신경망을 이용한 고속 객체 검출 방법의 메모리와 계산량을 다른 방식들과 비교한 그래프이다.

[0108] 단일 단계 방식의 객체 검출 방법에서, 도 5 내지 도 7에 따른 추가 계층들 및 검출 계층들을 구성하여 객체를 검출하는 방식을 eSSD로 지칭하고, eSSD를 앞에서 설명한 SSD 및 SSDLite와 비교하면 다음의 표 2와 같다.

표 2

[0109]

	SSD	SSDLite	제안 방법(eSSD)
Additional layer 구성	1 by 1 3 by 3(stride 2)	1 by 1 depthwise 3 by 3(stride 2) 1 by 1	depthwise 3 by 3 (stride 2) 1 by 1
메모리	NM+9MK	NM+9M+MK	9N+NK
연산량	$NMB^2+\frac{9MKB^2}{4}$	$NMB^2+\frac{9MB^2}{4}+\frac{MKB^2}{4}$	$\frac{9NB^2}{4}+\frac{NKB^2}{4}$

[0110] 표 2를 참조하면, SSD 는 추가 계층들로, 1×1 컨볼루셔널 계층 및 3×3 컨볼루셔널 계층을 연결하여 사용하였고, SSDLite는 SSD의 추가 계층에 포함된 3×3 컨볼루셔널 계층을 depthwise 기반의 3×3 컨볼루셔널 계층 및 1×1 컨볼루셔널 계층으로 대체하여 사용하였다. 또한, 도 5 내지 도 7에 따라 제안된 방식(eSSD)은 추가 계층들로서 depthwise 기반의 3×3 컨볼루셔널 계층 및 1×1 컨볼루셔널 계층을 사용하였다.

[0111] 이때, 표 2의 메모리와 연산량 계산식에서, 추가 계층에서 입력 채널의 수를 N, 중간 출력 채널의 수를 M, 최종 채널의 수를 K로 지칭하고, 입력 영상의 크기가 B×B이며, 바이어스(Bias)는 편의상 제외한 것이다.

[0112] 표 2의 메모리 크기에 따르면, 도 5 내지 도 7에서 제안한 eSSD의 메모리 크기가 가장 작을 것이 예상되고, 연산량에서도 가장 작은 것이 확인될 수 있다.

[0113] 추가 레이어(Additional Layer)에 대해서 B=10,N=1024,M=512,K=256로 설정하고, 필터의 메모리 사용량, 컨볼루션(Convolution)의 계산량을 도 4에 따른 SSD를 기준으로한 다른 방식들(SSDLite, eSSD)의 비율을 계산해 보면 도 8의 그래프와 같다. 또한, B=10,N=1024, M=256, K=512로 설정하고, bottleneck case 에 대한 비교 그래프는 도 9과 같다. 도 8과 도 9에서 메모리 및 계산량 비는 각각의 비를 0.5로 가중치를 두어 계산하였다.

[0114] 도 8과 도 9를 참조하면, 제안된 방법 eSSD가 메모리 및 계산량에 있어서 기존의 SSD 및 SSDLite보다 상당 부분 개선된 것을 확인할 수 있다. 구체적으로 도 8과 같이 일반적인 경우, eSSD는 SSD 대비 약 1/10 정도 메모리 및 계산량이 더 적고 도 9와 같은 bottle neck의 경우, eSSD는 SSD 대비 약 1/3 수준으로 메모리 및 계산량이 작은 것을 확인할 수 있다.

[0115] 추가 계층 외에도 검출 계층에서 SSD의 3×3 컨볼루셔널 계층 또는 SSDLite의 depthwise 3×3 컨볼루셔널 계층과 1×1 컨볼루셔널 계층을 대신하여 1×1 컨볼루셔널 계층을 사용하면, 메모리 크기를 감소시키는 효과를 가져올 수 있다. 이러한 효과를 확인하기 위해서 MobileNetv1를 기본 네트워크로 사용하고, VOC 0712 데이터를 학습 데이터로 이용하여 학습한 모델의 mAP 및 모델 크기, inference time을 비교한 결과는 다음의 표 3과 같다.

표 3

[0116]

	Baseline (300 by300)	추가레이어	검출레이어	성능(mAP)	모델사이즈	속도(fps)
1	MobileNet	SSD	3*3	**68.6**	34.8M	122.5490196
2	MobileNet	SSD	1*1	67.8	23.1M	128.2051282
3	MobileNet	SSDLite	1*1	67.8	**16.8M**	137.9310345
4	MobileNet	eSSD(제안방법)	1*1	67.9	17.4M	**141.843972**

[0117] 표 3에서 속도는 Titan X에서 측정하였고, batch size는 1이며, 추가 레이어는 앞선 도 5(또는 더 상세하게는 표 1)에 따라 구성되었다. 표 3을 참조하면, 기본 네트워크(또는 Baseline)로 사용한 MobileNet에서, eSSD는 기존의 SSD 및 SSDLite 와 비교할 때, 유사한 성능을 유지하면서 적은 메모리와 빠른 검출 속도를 갖는 것을 확인할 수 있다.

[0119] 도 10은 본 발명의 일 실시예에 따른 인공 신경망을 이용한 고속 객체 검출 방법에 대한 흐름도이다.

[0120] 도 10을 참조하면, 인공 신경망을 이용한 고속 객체 검출 방법은, 입력 영상을 획득하는 단계(S100), 획득된 입력 영상을 미리 설정된 복수의 바운딩 박스(default boxes)를 이용한 객체 검출 신경망에 입력하는 단계(S110) 및 상기 객체 검출 신경망의 출력 데이터를 획득하여 상기 입력 영상에 포함된 객체를 검출하는 단계(S120)를 포함할 수 있다.

[0121] 상기 객체 검출 신경망은, 기본 네트워크(Base Network), 상기 기본 네트워크의 출력을 수신하는 추가 계층들(Additional Layers) 및 상기 추가 계층들 또는 상기 기본 네트워크의 출력을 수신하는 검출 계층들(Detection Layers)을 포함할 수 있다.

[0122] 상기 기본 네트워크는, VGG, ResNet, ResNext, Mobilenet 중 하나일 수 있다.

[0123] 상기 기본 네트워크는, 3×3 컨볼루셔널 계층(Convolutional Layer) 및 2×2 맥스풀링 계층(Maxpooling Layer)을 포함할 수 있다.

[0124] 상기 기본 네트워크는, 상기 2×2 맥스풀링 계층의 출력을 수신하는 depthwise 기반의 3×3 컨볼루셔널 계층 및 상기 depthwise 기반의 3×3 컨볼루셔널 계층의 출력을 수신하는 1×1 컨볼루셔널 계층을 더 포함할 수 있다.

[0125] 상기 추가 계층들은, depthwise 기반의 3×3 컨볼루셔널 계층 및 상기 depthwise 기반의 3×3 컨볼루셔널 계층의 출력에 연결된 1×1 컨볼루셔널 계층을 포함할 수 있다.

[0126] 상기 검출 계층들은, 상기 객체의 위치를 검출하는 1×1 컨볼루셔널 계층 및 상기 객체의 종류를 식별하는 1×1 컨볼루셔널 계층을 포함할 수 있다.

[0127] 상기 depthwise 기반의 3×3 컨볼루셔널 계층은, 합성곱을 수행하는 간격으로서, 스트라이드(stride) 값이 2로 설정될 수 있다.

[0128] 상기 추가 계층들은, 상기 기본 네트워크의 출력을 수신하여 특징을 추출하는 제1 추가 계층 및 상기 제1 추가 레이어의 출력을 수신하여 특징을 추출하는 제2 추가 계층을 포함할 수 있다.

[0129] 상기 검출 계층들은, 상기 기본 네트워크의 출력을 수신하여 상기 객체의 위치와 종류를 식별하는 제1 검출 계층 및 상기 제1 추가 계층의 출력을 수신하여 상기 객체의 위치와 종류를 식별하는 제2 검출 계층을 포함할 수 있다.

[0130] 상기 제1 추가 계층은, depthwise 기반의 3×3 컨볼루셔널 계층 및 상기 depthwise 기반의 3×3 컨볼루셔널 계층의 출력에 연결된 1×1 컨볼루셔널 계층으로 구성되는 구조를 적어도 두 개 포함할 수 있다.

[0132] 도 11은 본 발명의 일 실시예에 따른 인공 신경망을 이용한 고속 객체 검출 장치에 대한 구성도이다.

[0133] 도 11을 참조하면, 인공 신경망을 이용한 고속 객체 검출 장치(100)는, 적어도 하나의 프로세서(processor) 및 상기 적어도 하나의 프로세서가 적어도 하나의 단계를 수행하도록 지시하는 명령어들(instructions)을 저장하는 메모리(memory)를 포함할 수 있다.

[0134] 상기 적어도 하나의 프로세서(110)는 중앙 처리 장치(central processing unit, CPU), 그래픽 처리 장치(graphics processing unit, GPU), 또는 본 발명의 실시예들에 따른 방법들이 수행되는 전용의 프로세서를 의미할 수 있다. 메모리(120) 및 저장 장치(160) 각각은 휘발성 저장 매체 및 비휘발성 저장 매체 중에서 적어도 하나로 구성될 수 있다. 예를 들어, 메모리(120)는 읽기 전용 메모리(read only memory, ROM) 및 랜덤 액세스 메모리(random access memory, RAM) 중에서 적어도 하나로 구성될 수 있다.

[0135] 또한, 인공 신경망을 이용한 고속 객체 검출 장치(100)는, 무선 네트워크를 통해 통신을 수행하는 송수신 장치(transceiver, 130)를 포함할 수 있다. 또한, 인공 신경망을 이용한 고속 객체 검출 장치(100)는 입력 인터페이스 장치(140), 출력 인터페이스 장치(150), 저장 장치(160) 등을 더 포함할 수 있다. 인공 신경망을 이용한 고속 객체 검출 장치(100)에 포함된 각각의 구성 요소들은 버스(bus)(170)에 의해 연결되어 서로 통신을 수행할 수 있다.

[0136] 상기 적어도 하나의 단계는, 입력 영상을 획득하는 단계, 획득된 입력 영상을 미리 설정된 복수의 바운딩 박스

(default boxes)를 이용한 객체 검출 신경망에 입력하는 단계 및 상기 객체 검출 신경망의 출력 데이터를 획득하여 상기 입력 영상에 포함된 객체를 검출하는 단계를 포함할 수 있다.

[0137] 상기 객체 검출 신경망은, 기본 네트워크(Base Network), 상기 기본 네트워크의 출력을 수신하는 추가 계층들(Additional Layers) 및 상기 추가 계층들 또는 상기 기본 네트워크의 출력을 수신하는 검출 계층들(Detection Layers)을 포함할 수 있다.

[0138] 상기 기본 네트워크는, VGG, ResNet, ResNext, Mobilenet 중 하나일 수 있다.

[0139] 상기 기본 네트워크는, 3×3 컨볼루셔널 계층(Convolutional Layer) 및 2×2 맥스풀링 계층(Maxpooling Layer)을 포함할 수 있다.

[0140] 상기 기본 네트워크는, 상기 2×2 맥스풀링 계층의 출력을 수신하는 depthwise 기반의 3×3 컨볼루셔널 계층 및 상기 depthwise 기반의 3×3 컨볼루셔널 계층의 출력을 수신하는 1×1 컨볼루셔널 계층을 더 포함할 수 있다.

[0141] 상기 추가 계층들은, depthwise 기반의 3×3 컨볼루셔널 계층 및 상기 depthwise 기반의 3×3 컨볼루셔널 계층의 출력에 연결된 1×1 컨볼루셔널 계층을 포함할 수 있다.

[0142] 상기 검출 계층들은, 상기 객체의 위치를 검출하는 1×1 컨볼루셔널 계층 및 상기 객체의 종류를 식별하는 1×1 컨볼루셔널 계층을 포함할 수 있다.

[0143] 상기 depthwise 기반의 3×3 컨볼루셔널 계층은, 합성곱을 수행하는 간격으로서, 스트라이드(stride) 값이 2로 설정될 수 있다.

[0144] 상기 추가 계층들은, 상기 기본 네트워크의 출력을 수신하여 특징을 추출하는 제1 추가 계층 및 상기 제1 추가 레이어의 출력을 수신하여 특징을 추출하는 제2 추가 계층을 포함할 수 있다.

[0145] 상기 검출 계층들은, 상기 기본 네트워크의 출력을 수신하여 상기 객체의 위치와 종류를 식별하는 제1 검출 계층 및 상기 제1 추가 계층의 출력을 수신하여 상기 객체의 위치와 종류를 식별하는 제2 검출 계층을 포함할 수 있다.

[0146] 상기 제1 추가 계층은, depthwise 기반의 3×3 컨볼루셔널 계층 및 상기 depthwise 기반의 3×3 컨볼루셔널 계층의 출력에 연결된 1×1 컨볼루셔널 계층으로 구성되는 구조를 적어도 두 개 포함할 수 있다.

[0147] 인공 신경망을 이용한 고속 객체 검출 장치(100)의 예를 들면, 통신 가능한 데스크탑 컴퓨터(desktop computer), 랩탑 컴퓨터(laptop computer), 노트북(notebook), 스마트폰(smart phone), 태블릿 PC(tablet PC), 모바일폰(mobile phone), 스마트 워치(smart watch), 스마트 글래스(smart glass), e-book 리더기, PMP(portable multimedia player), 휴대용 게임기, 네비게이션(navigation) 장치, 디지털 카메라(digital camera), DMB(digital multimedia broadcasting) 재생기, 디지털 음성 녹음기(digital audio recorder), 디지털 음성 재생기(digital audio player), 디지털 동영상 녹화기(digital video recorder), 디지털 동영상 재생기(digital video player), PDA(Personal Digital Assistant) 등일 수 있다.

[0149] 본 발명에 따른 방법들은 다양한 컴퓨터 수단을 통해 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 컴퓨터 판독 가능 매체에 기록되는 프로그램 명령은 본 발명을 위해 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다.

[0150] 컴퓨터 판독 가능 매체의 예에는 롬(ROM), 램(RAM), 플래시 메모리(flash memory) 등과 같이 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함될 수 있다. 프로그램 명령의 예에는 컴파일러(compiler)에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터(interpreter) 등을 사용해서 컴퓨터에 의해 실행될 수 있는 고급 언어 코드를 포함할 수 있다. 상술한 하드웨어 장치는 본 발명의 동작을 수행하기 위해 적어도 하나의 소프트웨어 모듈로 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

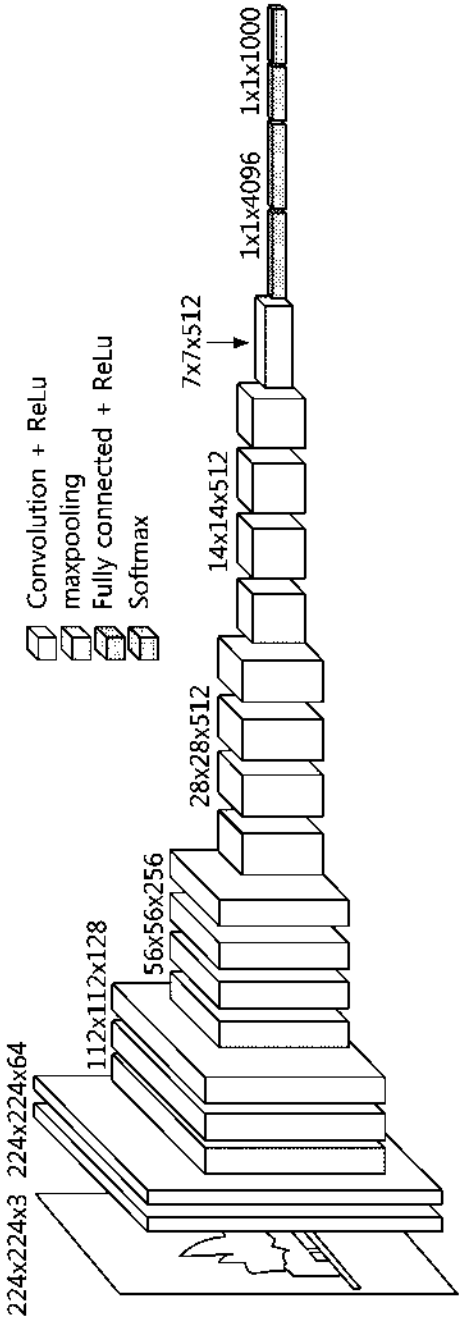
[0151] 또한, 상술한 방법 또는 장치는 그 구성이나 기능의 전부 또는 일부가 결합되어 구현되거나, 분리되어 구현될 수 있다.

[0152] 상기에서는 본 발명의 바람직한 실시예를 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구의 범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정

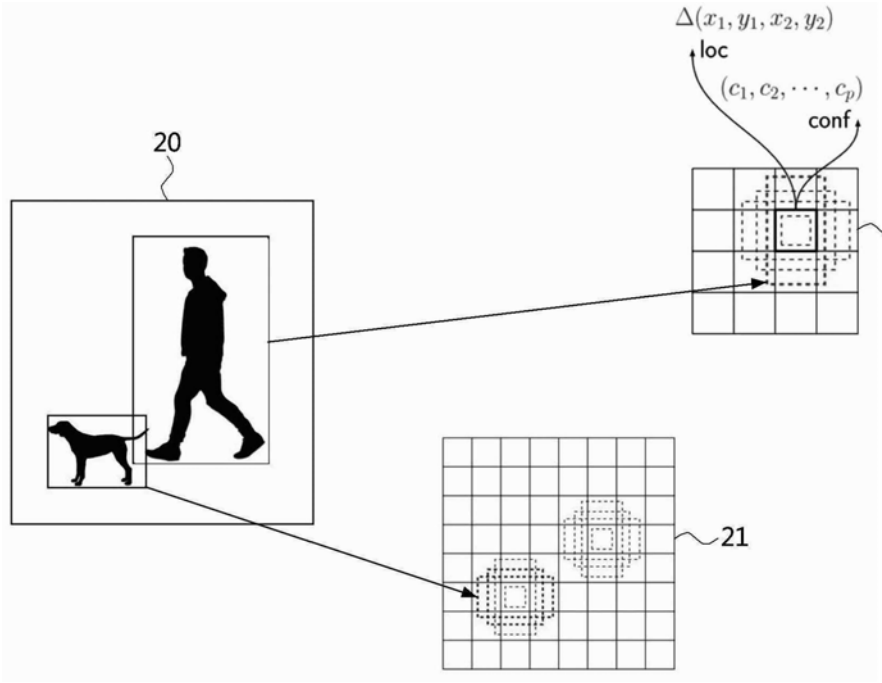
및 변경시킬 수 있음을 이해할 수 있을 것이다.

도면

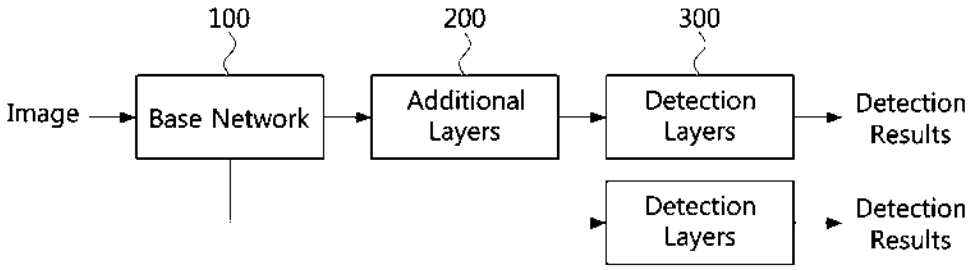
도면1



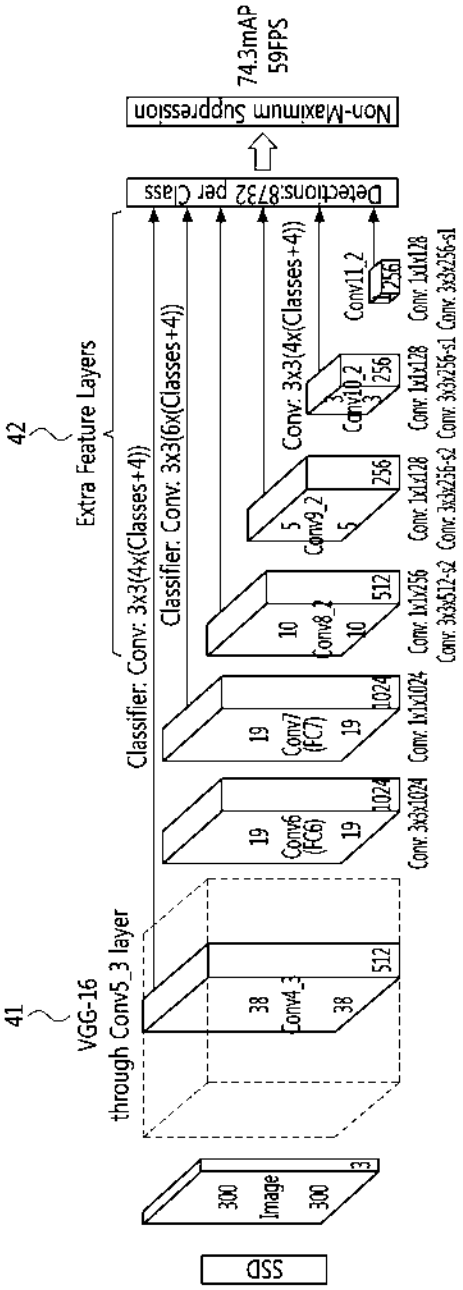
도면2


$\Delta(x_1, y_1, x_2, y_2)$
loc
$(c_1, c_2, \cdots, c_p)$
conf
20
22
21

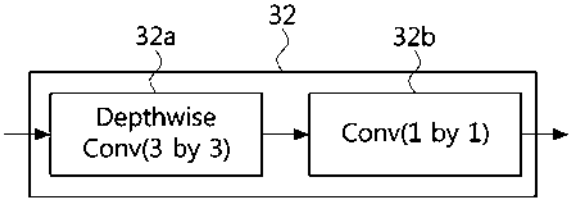
도면3


100
200
300
Image
Base Network
Additional Layers
Detection Layers
Detection Results
Detection Layers
Detection Results

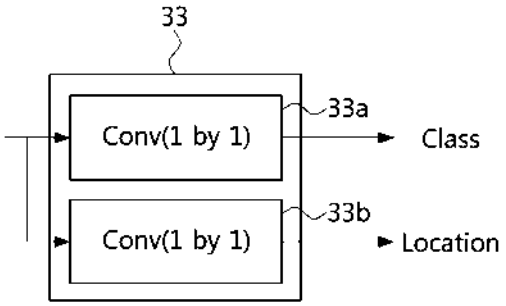
도면4


SSD
300
Image
300
3
41
VGG-16
through Conv5_3 layer
38
Conv4_3
38
512
19
Conv6
(FC6)
19
1024
Conv: 3x3x1024
19
Conv7
(FC7)
19
1024
Conv: 1x1x1024
10
Conv8_2
10
512
Conv: 1x1x256
Conv: 3x3x512-s2
5
Conv9_2
5
256
Conv: 1x1x128
Conv: 3x3x256-s2
3
Conv10_2
3
256
Conv: 1x1x128
Conv: 3x3x256-s1
1
Conv11_2
1
256
Conv: 1x1x128
Conv: 3x3x256-s1
42
Extra Feature Layers
Classifier: Conv: 3x3(4x(Classes+4))
Classifier: Conv: 3x3(6x(Classes+4))
Conv: 3x3(4x(Classes+4))
Detections:8732 per Class
Non-Maximum Suppression
74.3mAP
59FPS

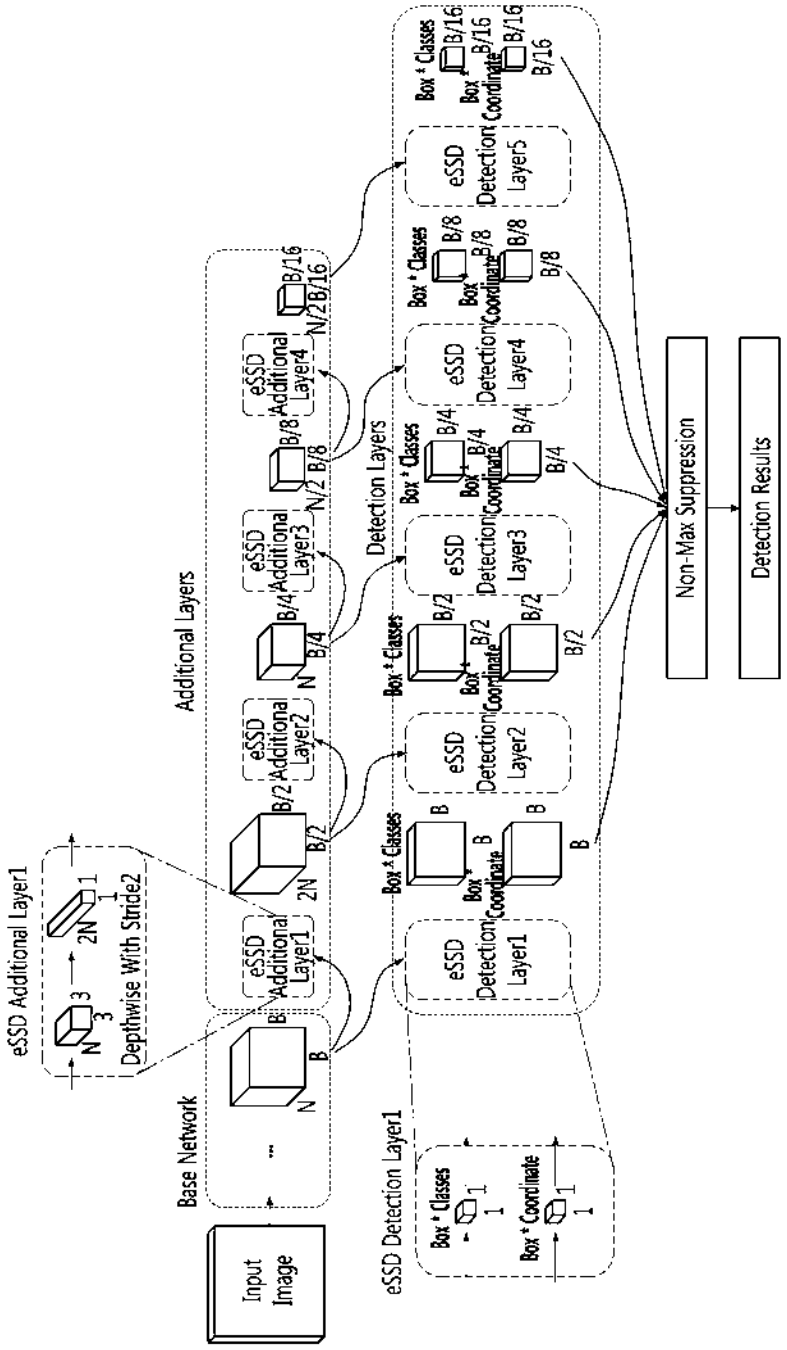
도면5


32a
32
32b
Depthwise
Conv(3 by 3)
Conv(1 by 1)

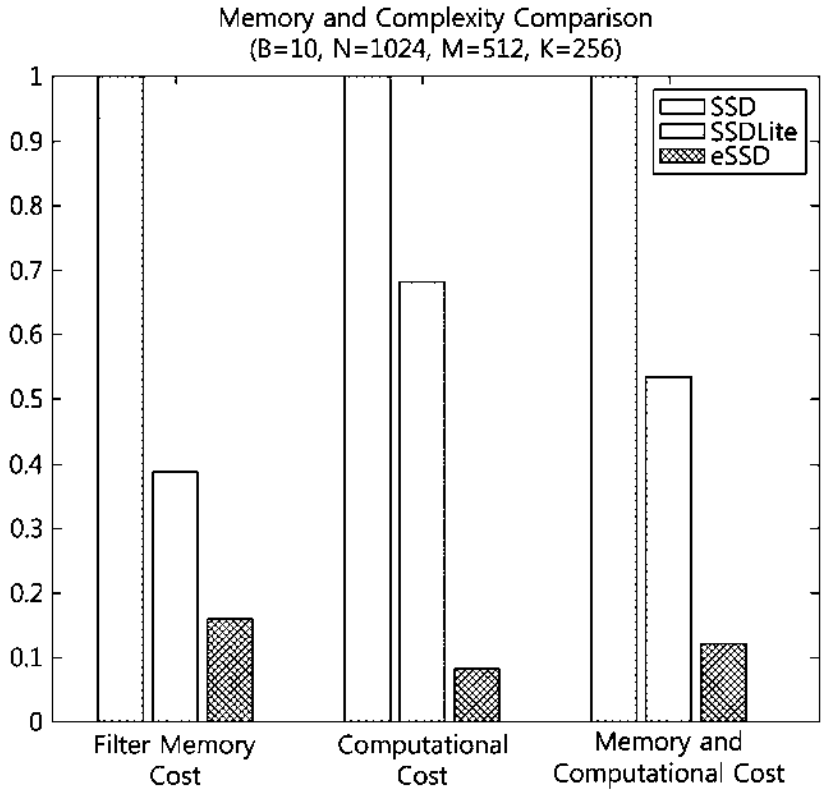
도면6


33
Conv(1 by 1)
33a
Class
Conv(1 by 1)
33b
Location

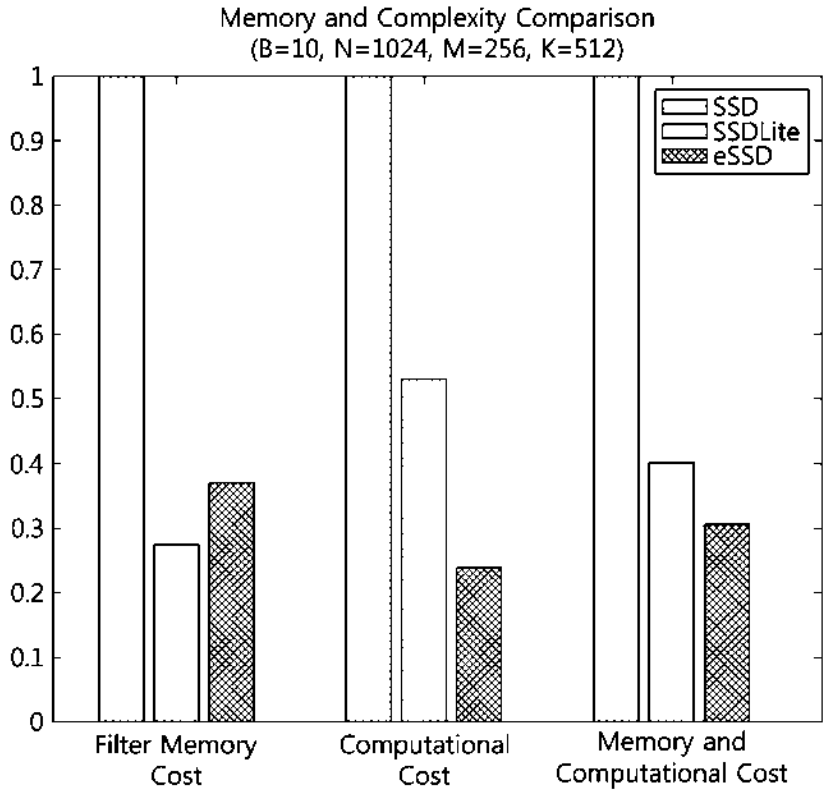
도면7


eSSD Additional Layer1
Depthwise With Stride2
Input Image
Base Network
Additional Layers
eSSD Additional Layer1
eSSD Additional Layer2
eSSD Additional Layer3
eSSD Additional Layer4
Detection Layers
eSSD Detection Layer1
eSSD Detection Layer2
eSSD Detection Layer3
eSSD Detection Layer4
eSSD Detection Layer5
Box * Classes
Box * Coordinate
Non-Max Suppression
Detection Results

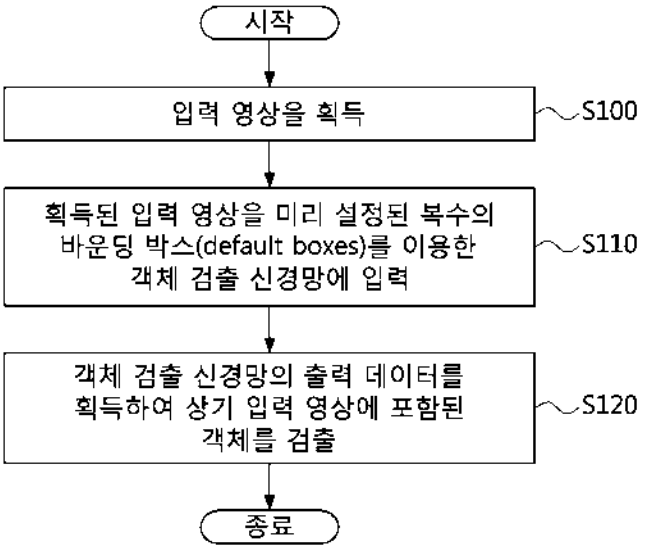
도면8


Memory and Complexity Comparison
(B=10, N=1024, M=512, K=256)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
SSD
SSDLite
eSSD
Filter Memory Cost
Computational Cost
Memory and Computational Cost

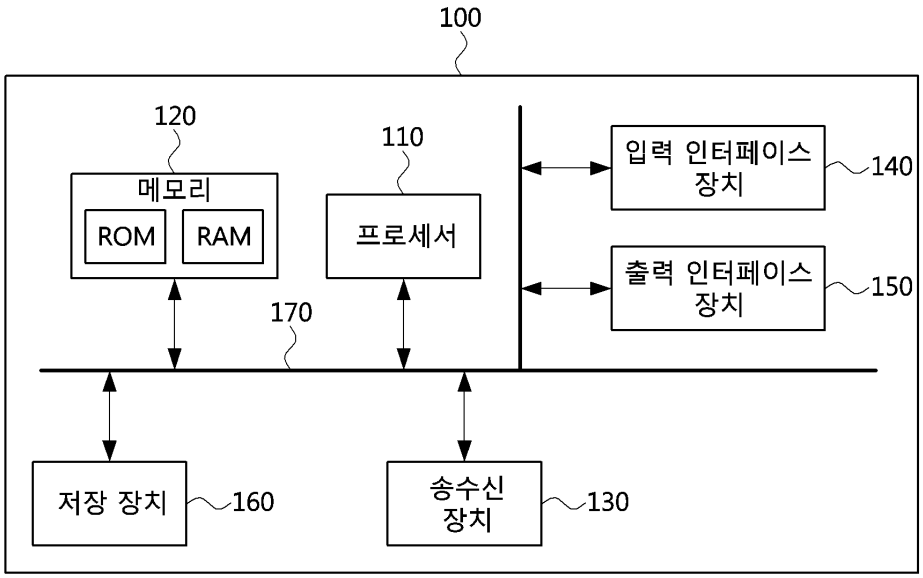
도면9


Memory and Complexity Comparison
(B=10, N=1024, M=256, K=512)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
SSD
SSDLite
eSSD
Filter Memory Cost
Computational Cost
Memory and Computational Cost

도면10


시작
입력 영상을 획득
S100
획득된 입력 영상을 미리 설정된 복수의 바운딩 박스(default boxes)를 이용한 객체 검출 신경망에 입력
S110
객체 검출 신경망의 출력 데이터를 획득하여 상기 입력 영상에 포함된 객체를 검출
S120
종료

도면11


100
120
메모리
ROM
RAM
110
프로세서
입력 인터페이스 장치
140
출력 인터페이스 장치
150
170
저장 장치
160
송수신 장치
130