



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0039043
(43) 공개일자 2020년04월16일

(51) 국제특허분류(Int. Cl.)
G06K 9/20 (2006.01) G06N 3/04 (2006.01)
G06T 7/70 (2017.01)
(52) CPC특허분류
G06K 9/20 (2013.01)
G06N 3/04 (2013.01)
(21) 출원번호 10-2018-0116571
(22) 출원일자 2018년09월28일
심사청구일자 없음
기술이전 희망 : 기술양도

(71) 출원인
한국전자통신연구원
대전광역시 유성구 가정로 218 (가정동)
(72) 발명자
김진규
세종특별자치시 누리로 27, 614동 603호 (한솔동, 첫마을6단지)
김병조
세종특별자치시 누리로 27, 614동 1703호(한솔동, 첫마을아파트)
(뒷면에 계속)
(74) 대리인
팬코리아특허법인

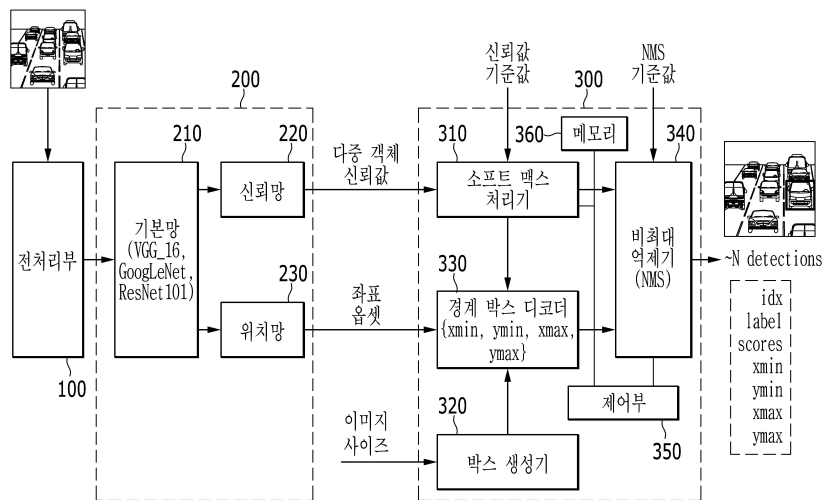
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 객체 인식 장치 및 그 동작 방법

(57) 요약

뉴럴넷(Neural Nets)을 이용하여 입력 영상 내의 객체를 검출하는 객체 인식 장치에서, 컨볼루션 뉴럴 네트워크(Convolution Neural Networks, CNN)를 사용하여 수행되는 뉴럴넷 학습에 기반하여, 객체의 신뢰값을 출력하는 학습부, 그리고 신뢰값을 객체가 특정 객체 종류에 해당할 확률을 나타내는 확률값으로 변환하고, 확률값과 기저장된 기준값의 비교결과에 기반하여 객체에 대응하는 박스를 생성하는 객체 검출부를 포함하는 객체 인식 장치가 제공된다.

대표도 - 도1



(52) CPC특허분류

G06N 3/08 (2013.01)

G06T 7/70 (2017.01)

G06T 2210/12 (2013.01)

(72) 발명자

김성민

세종특별자치시 새롬남로 18, 103동 2802호 (새롬동, 새뜸마을1단지)

김주엽

대전광역시 서구 청사로 281, 215동 2001호 (둔산동, 샘머리아파트2단지)

이미영

대전광역시 유성구 죽동로 39, 211동 702호 (죽동, 칸타빌)

이주현

대전광역시 유성구 상대남로 26, 906동 1404호 (상대동, 트리폴시티아파트)

이 발명을 지원한 국가연구개발사업

과제고유번호 2016-0-00098

부처명 과학기술정보통신부

연구관리전문기관 정보통신기술진흥센터(IITP)

연구사업명 정보통신방송기술개발사업

연구과제명 신경모사 인지형 모바일 컴퓨팅 지능형반도체 기술개발

기 여 율 1/1

주관기관 한국전자통신연구원

연구기간 2017.01.01~2017.12.31

명세서

청구범위

청구항 1

뉴럴넷(Neural Nets)을 이용하여 입력 영상 내의 객체를 검출하는 객체 인식 장치에서,

컨볼루션 뉴럴 네트워크(Convolution Neural Networks, CNN)을 사용하여 수행되는 뉴럴넷 학습에 기반하여, 상기 객체의 신뢰값을 출력하는 학습부, 그리고

상기 신뢰값을 상기 객체가 특정 객체 종류에 해당할 확률을 나타내는 확률값으로 변환하고, 상기 확률값과 기 저장된 기준값의 비교결과에 기반하여 상기 객체에 대응하는 박스를 생성하는 객체 검출부

를 포함하는 객체 인식 장치.

청구항 2

제1항에서,

상기 객체 검출부는,

상기 확률값이 상기 기준값 보다 크면 인덱스를 생성하고, 상기 인덱스가 생성된 객체에 대해서 박스를 생성하는, 객체 인식 장치.

청구항 3

제2항에서,

상기 객체 검출부는,

상기 뉴럴넷으로부터 출력된 옵션에 기반하여 상기 박스의 위치 정보를 출력하는, 객체 인식 장치.

청구항 4

제3항에서,

상기 객체 검출부는,

각 확률값에 대응하는 상기 출력된 위치 정보에서 서로 겹치는 위치 정보를 제거하는, 객체 인식 장치.

청구항 5

뉴럴넷(Neural Nets)을 이용하여 객체를 검출하는 객체 인식 장치의 동작 방법에서,

컨볼루션 뉴럴 네트워크(Convolution Neural Networks, CNN)을 사용하여 수행되는 뉴럴넷 학습에 기반하여, 상기 객체의 신뢰값을 출력하는 단계,

상기 신뢰값을 상기 객체가 특정 객체 종류에 해당할 확률을 나타내는 확률값으로 변환하는 단계, 그리고

상기 확률값과 기 저장된 기준값의 비교결과에 기반하여 상기 객체에 대응하는 박스를 생성하는 단계,

를 포함하는 객체 인식 장치의 동작 방법.

청구항 6

제5항에서,

상기 박스를 생성하는 단계는,

상기 확률값이 상기 기준값 보다 크면 인덱스를 생성하고, 상기 인덱스가 생성된 객체에 대해서 박스를 생성하는, 객체 인식 장치의 동작 방법.

청구항 7

제6항에서,

상기 박스를 생성하는 단계 이후에는,

상기 뉴럴넷으로부터 출력된 옵션에 기반하여 상기 박스의 위치 정보를 도출하는 단계를 포함하는, 객체 인식 장치의 동작 방법.

청구항 8

제7항에서,

상기 위치 정보를 도출하는 단계 이후에는,

각 확률값에 대응하는 상기 도출된 위치 정보에서 서로 겹치는 위치 정보를 제거하는 단계를 포함하는, 객체 인식 장치의 동작 방법.

발명의 설명

기술 분야

[0001] 본 기재는 뉴럴넷(Neural Nets)을 이용하여 객체를 검출하는 객체 인식 장치 및 그 동작 방법에 관한 것이다.

배경 기술

[0002] 최근 들어 인공 지능과 관련하여, 딥러닝 기술이 영상에서 사람의 얼굴이나 사물을 인식하고 위치를 검출하는 분야에 활용되고 있다.

[0003] 딥러닝 기술을 활용하기 위해서는 뉴럴넷의 구조를 정의하고, 데이터 셋(data set)을 이용하여 학습을 해야 한다. 데이터 셋은 학습에 사용되는 다양한 사진 데이터, 객체의 종류 및 위치에 대한 정확한 정보로 구성된다. 학습에 사용되는 데이터 셋의 종류에는 1000개의 객체 클래스 정보를 제공하는 이미지넷(ImageNets), 20개의 사물에 관한 종류와 위치 정보를 갖고 있는 PASCAL VOC 데이터 셋(Pattern Analysis, Statistical Modeling and Computational Learning Visual Object Classes Data set), 80개의 사물에 관한 종류와 위치 정보를 나타내는 MS-COCO 데이터 셋 등이 있다. 데이터 셋은 목적에 따라 학습용 DB와 테스트용 DB로 나뉘어진다.

[0004] 학습용 DB는 매우 많은 양의 이미지 정보 및 정답 정보를 가지고 있으며, 학습 파라미터를 유도하기 위해 사용된다. 테스트용 DB는 학습을 마친 신경망의 파라미터인 가중치(weight) 및 바이어스(bias)의 효과를 판단하기 위해 사용된다. 테스트용 DB는 백분율로 표시된 검출율(accuracy)나 mAP(mean average precision) 값을 가진다.

[0005] 종래의 딥러닝을 이용하여 객체의 종류와 위치를 검출하는 기술은 매우 높은 연산량을 수반하기 때문에, 인식 속도가 저하되는 문제점이 있다.

발명의 내용

해결하려는 과제

[0006] 한 실시예는 뉴럴넷을 이용하여 객체 인식 성능을 확보하고, 인식 속도를 향상시키는 객체 인식 장치 및 그 동작 방법을 제공한다.

과제의 해결 수단

[0007] 한 실시예에 따르면, 뉴럴넷(Neural Nets)을 이용하여 입력 영상 내의 객체를 검출하는 객체 인식 장치는, 컨볼루션 뉴럴 네트워크(Convolution Neural Networks, CNN)를 사용하여 수행되는 뉴럴넷 학습에 기반하여, 상기 객체의 신뢰값을 출력하는 학습부, 그리고 상기 신뢰값을 상기 객체가 특정 객체 종류에 해당할 확률을 나타내는 확률값으로 변환하고, 상기 확률값과 기저장된 기준값의 비교결과에 기반하여 상기 객체에 대응하는 박스를 생성하는 객체 검출부를 포함한다.

[0008] 상기 객체 검출부는, 상기 확률값이 상기 기준값 보다 크면 인덱스를 생성하고, 상기 인덱스가 생성된 객체에

대해서 박스를 생성할 수 있다.

- [0009] 상기 객체 검출부는, 상기 뉴럴넷으로부터 출력된 옵션에 기반하여 상기 박스의 위치 정보를 출력할 수 있다.
- [0010] 상기 객체 검출부는, 각 확률값에 대응하는 상기 출력된 위치 정보에서 서로 겹치는 위치 정보를 제거할 수 있다.
- [0011] 다른 실시예에 따르면, 뉴럴넷(Neural Nets)을 이용하여 객체를 검출하는 객체 인식 장치의 동작 방법은, 컨볼루션 뉴럴 네트워크(Convolution Neural Networks, CNN)를 사용하여 수행되는 뉴럴넷 학습에 기반하여, 상기 객체의 신뢰값을 출력하는 단계, 상기 신뢰값을 상기 객체가 특정 객체 종류에 해당할 확률을 나타내는 확률값으로 변환하는 단계, 그리고 상기 확률값과 기저장된 기준값의 비교결과에 기반하여 상기 객체에 대응하는 박스를 생성하는 단계를 포함한다.
- [0012] 상기 박스를 생성하는 단계는, 상기 확률값이 상기 기준값 보다 크면 인덱스를 생성하고, 상기 인덱스가 생성된 객체에 대해서 박스를 생성할 수 있다.
- [0013] 상기 박스를 생성하는 단계 이후에는, 상기 뉴럴넷으로부터 출력된 옵션에 기반하여 상기 박스의 위치 정보를 도출하는 단계를 포함할 수 있다.
- [0014] 상기 위치 정보를 도출하는 단계 이후에는, 각 확률값에 대응하는 상기 도출된 위치 정보에서 서로 겹치는 위치 정보를 제거하는 단계를 포함할 수 있다.

발명의 효과

- [0015] 본 기재의 객체 인식 장치 및 그 동작 방법에 따르면, 높은 신뢰값을 갖는 위치 기반 객체에 대해서만 기본 박스 생성 및 박스 디코딩 동작을 수행함으로써, 고속으로 객체를 검출할 수 있다.
- [0016] 또한, 불필요한 박스에 대한 처리를 사전에 미리 방지하고, 정렬(sorting)이나 NMS 처리 시간을 단축시킬 수 있다.

도면의 간단한 설명

- [0017] 도 1은 한 일 실시예에 따른 객체 인식 장치의 블록도이다.
- 도 2는 한 실시예에 따른 기본망의 구조를 나타내는 도면이다.
- 도 3 및 도 4는 한 실시예에 따른 사물 위치에 관한 정보를 나타내는 도면이다.
- 도 5는 한 실시예에 따른 박스 생성기에 의한 기본 박스 생성을 나타내는 도면이다.
- 도 6 및 도 7은 한 실시예에 따른 경계 박스 디코더의 동작을 나타내는 도면이다.
- 도 8은 한 실시예에 따른 자카드 오버랩을 나타내는 도면이다.
- 도 9는 한 일 실시예에 따른 객체 인식 장치의 동작 방법의 순서도이다.

발명을 실시하기 위한 구체적인 내용

- [0018] 아래에서는 첨부한 도면을 참고로 하여 본 발명의 실시예에 대하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.
- [0019] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성 요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.
- [0020] 도 1은 한 일 실시예에 따른 객체 인식 장치의 블록도이다.
- [0021] 도 1을 참조하면, 한 실시예에 따른 객체 인식 장치(1)는 전처리부(100), 학습부(200), 객체 검출부(300)를 포함한다.
- [0022] 전처리부(100)는 스케일러(scaler), 이미지 크롭퍼(cropper), 그리고 노멀라이저(normalizer)를 포함할 수

있다. 전처리부(100)는 영상이 입력되면, 사전 처리(pre-processing)를 수행한다. 구체적으로, 전처리부(100)에 의한 사전 처리 과정은 스케일러에 의해 입력 영상의 해상도를 뉴럴넷의 입력에 맞도록 변경하는 과정, 이미지 크롭퍼에 의해 특정 영역만 선택하여 잘라내는 과정, 그리고 노멀라이저에 의해 영상의 RGB 채널에 따라 평균값을 빼고 스케일을 조정하는 과정을 포함한다. 이러한 사전 처리 동작은 실제 인식물에 큰 영향을 미치기 때문에, 학습을 진행하기 전에 실제 응용 어플리케이션을 고려하여 수행되어야 한다.

[0023] 예를 들어, 640x480 해상도를 갖는 영상이 입력되고 기본망(BaseNet)(210)의 입력이 300x300의 크기를 가지면, 640x480의 입력 영상은 스케일러에 의해 300x300으로 축소된다. 이때 이미지 크롭퍼에 의해 적절한 영역으로 절삭될 수도 있고, 노멀라이저에 의해 절삭 없이 원본 동영상의 크기만 조정될 수도 있다. RGB 데이터는 1바이트의 크기를 가지며, 0~255까지의 값으로 분포된다. 신경망 데이터가 0을 중심으로 분포되는 것이 학습에 유리하다. 이를 위해, 상수 값 128을 빼서 -128~127까지의 범위 값을 만들 수도 있고, 수백 만장의 학습 이미지의 해당 위치 픽셀의 평균값을 모두 구한 다음 이를 뺄 수도 있다. 전처리부(100)에 의한 사전 처리 동작이 완료되면, 컨볼루션 뉴럴 네트워크(Convolution Neural Networks, CNN)를 사용하여 뉴럴넷 학습을 수행하는 학습부(200)가 동작한다. 학습부(200)는 기본망(BaseNet)(210), 신뢰망(Confidence Net)(220), 그리고 위치망(Location Net)(230)을 포함한다. 전처리된 데이터는 기본망(210)에 먼저 입력되고, 기본망(210)의 출력은 신뢰망(220)과 위치망(230)에 각각 입력된다.

[0024] 기본망(210)은 VGG-16, Google inception, ResNet 등을 활용할 수도 있고, 새로운 형태의 CNN 구조를 적용할 수도 있다. 기본망(210)은 커널의 크기에 따라 이미지의 피쳐(feature)를 검출하는 컨볼루션 레이어(convolutional layer), 정류된 선형 활성화 함수(non-linear activation function) 처리를 수행하는 ReLU 레이어, 공간 영역(spatial domain)을 축소하는 풀링레이어(pooling layer), 그리고 특징 데이터(feature data)의 평균과 분산값을 이용하여 데이터의 범위를 조정하는 정규화 레이어(normalize layer) 등을 포함할 수 있다.

[0025] 도 2는 한 실시예에 따른 기본망의 구조를 나타내는 도면이다.

[0026] 도 2를 참조하면, 카메라로부터 입력된 영상은 첫 번째 컨볼루션 레이어에 입력되기 전 전처리부(100)에 의해 사전 처리 작업이 수행된다. 채널은 색상별로 Red, Green, Blue 3채널이며, 채널의 크기는 동일하다. 공간 영역(spatial domain)에 해당하는 크기는 신경망의 입력에 맞게 300x300으로 축소(scale down)된다. 첫 번째 컨볼루션 레이어에서, 3x3의 커널 크기를 가지며 채널이 3인 필터를 적용하여 한 개의 출력을 연산한다. 총 64개의 서로 다른 3차원 필터(3x3x3)가 사용될 수 있다. 3차원 필터의 수가 총 64개이므로 컨볼루션 레이어 출력의 채널 크기는 64가 된다.

[0027] 3x3 커널은 입력 데이터의 좌측 상단에서부터 래스터 주사(raster-scan) 형태로 스트라이드(stride)에 따라 건너뛰면서 수행된다. 컨볼루션 연산이 완료되면 특징 데이터(feature data)들에 대해 엘리먼트 와이즈(element wise) 형태로 배치 정규화(batch normalization) 및 ReLU 연산이 수행된다. 이때 신경망의 학습을 통해 얻어진 평균(mean), 분산(variance), 스케일(scale) 및 바이어스(bias) 값이 사용된다. ReLU 레이어는 0보다 작은 값은 0으로 출력하고 0보다 큰 값은 입력 그대로 출력하는 레이어이다.

[0028] 풀링 레이어에서는 채널이 64로 유지되면서, 커널의 크기가 2x2이고 스트라이드(stride)가 2인 경우에는 이미지의 크기가 150x150에서 75x75로 축소된다. 기본망(210)의 전체 구조는 도 2에 도시된 구조가 체인 형태로 반복되는 구조이며, 데이터의 크기 및 필터의 크기는 파라미터에 따라 변화될 수 있다. 풀링 레이어는 체인 구조에 따라 생략될 수도 있다.

[0029] 객체의 종류와 위치까지 추출하기 위해서는 기본망(210) 이외에 추가적으로 신뢰망(220)과 위치망(230)이 필요하다. 신뢰망(220)은 각각의 영역에서 어떤 종류의 객체가 가장 높은 확률 값을 갖는가를 판단한다. 위치망(230)은 각각의 영역에서 실제 객체의 위치에 맞게 해당 영역(박스 형태의 사각형)의 중심점과 크기를 조정한다. 신뢰망(220)과 위치망(230)은 컨볼루션 레이어로만 구성될 수 있으며, 신뢰망(220)과 위치망(230)에서 정규화 레이어, ReLU, 풀링 레이어 등은 사용되지 않을 수 있다.

[0030] 신뢰망(220)은 객체의 클래스 종류에 따라 신뢰값을 계산한다. 계산된 신뢰값은 소프트맥스(softmax) 처리기(310)를 통해 확률 값으로 변환된다. 예를 들어, 20개의 객체 종류를 판별하는 PASCAL VOC 데이터 셋을 이용하는 뉴럴넷이 있을 때, 다중 객체(multi-class)의 종류는 총 20가지가 되고, 각 객체는 객체 종류별로 20개의 확률값을 갖게 된다. 확률값은 객체가 특정 객체 종류에 해당할 확률을 나타내는 값이다.

[0031] 객체의 위치가 배경에 해당될 수도 있기 때문에 객체의 종류는 배경(background)이 추가되어 총 21개가 될 수 있다. 신뢰망(220)의 입력은 기본망(210)의 마지막 출력값이 된다. 예를 들어, 기본망(210)의 마지막 출력 공간

영역(spatial domain)이 5x5라고 가정하고, 픽셀당 1개의 기본 박스가 사용되며, 총 20개의 물체를 검출하는 시스템이라고 가정하면, 총 525개(=25x21)의 출력이 발생한다.

[0032] 위치망(230)은 객체의 위치에 대한 정보를 학습한다. 객체의 위치는 영상의 객체를 포함하는 사각형 박스의 위치로 표현되는데, 위치망은 그 위치에 대한 좌표 정보를 이용하여 학습을 수행한다.

[0033] 도 3 및 도 4는 한 실시예에 따른 사물 위치에 관한 정보를 나타내는 도면이다.

[0034] 사각형의 좌표 정보를 표현하는 방법 중 다음 2가지 방법이 사용된다. 2가지 방법 중 하나는 객체 위치의 좌측 상단의 꼭지점을 알려주는 {xmin, ymin}와 우측 하단의 꼭지점을 알려주는 {xmax, ymax}로 표현하는 방법이고 (도 3 참조), 다른 하나는 사각형의 중심점인 {cx, cy} 및 사각형의 높이와 폭을 알려주는 h, w로 표현하는 방법이다(도 4 참조). 2가지 방법은 4개의 정보량을 요구하며, 입력 이미지는 0~1로 정규화된 값으로 표현될 수 있다. 도 3 및 도 4를 참조하면, 640x480 크기의 입력 이미지에서 좌측 상단은 {0,0}로 표현되고, 우측 하단은 {1,1}로 표현된다.

[0035] 위치망(230)은 기본 박스의 위치 정보에서 어느 정도 오프셋(offset)을 갖는가에 대해 학습을 진행한다.

[0036] 객체 검출부(object detector)(300)는 소프트맥스 처리기(softmax processor)(310), 박스 생성기(default box generator)(320), 경계 박스 디코더(bounding box decoder)(330), 비최대 억제기(non-maximum suppressor, NMS)(340), 제어부(350), 그리고 메모리(360)를 포함한다.

[0037] 소프트맥스 처리기(310)는 신뢰망(220)으로부터 출력된 다중 객체(multi-class)의 신뢰값을 확률값으로 변환한다. 다중 객체의 개수는 학습에 사용되는 데이터 셋의 종류에 따라 달라질 수 있다. 예를 들어, 데이터 셋의 종류가 PASCAL VOC인 경우에는 총 20개의 객체에 비물체(background)까지 더해지므로, 다중 객체의 개수는 21개가 된다. 이때, 소프트맥스 처리기(310)는 다중 객체 각각의 위치 정보에 따라 총 21개의 확률값을 출력할 수 있다. 배경에 대한 확률값이 검출되는 경우에는 소프트맥스 처리기(310)는 해당 결과가 필요 없기 때문에 이를 삭제하고, 다음 단계부터 사용하지 않는다.

[0038] 확률값은 아래의 수학적 식 1을 통해 계산된다.

수학적 식 1

$$P(x_i) = \frac{\exp(x_i - M)}{\sum_{j=1}^N \exp(x_j - M)}, \quad M = \max_j(x_j)$$

[0039]

[0040] 수학적 식 1에서 N은 배경(background)까지 포함하는 값으로 21이 되고, 입력 및 출력 수 또한 21이 된다. M은 모든 입력 값 중에서 최대값을 나타내며, 최대값을 구하지 않고 0으로 적용해도 수식적으로 동일한 값이 출력된다. 수의 정밀도가 한정된 고정 소수점으로 연산을 수행할 경우, 최대값(M)을 적용하면 정확한 확률값을 계산할 수 있다.

[0041] 소프트맥스 처리기(310)에 의한 소프트맥스 처리는 박스 별로 하나씩 처리될 수 있으며, 처리 순서는 임의로 정해질 수 있다. 소프트맥스 처리기(310)에 의해 출력된 총 21개의 확률값 중에서 배경의 확률값이 가장 크게 나오면, 기본 박스 생성기(320)에 의한 기본 박스 생성이나 경계 박스 디코더(330)에 의한 박스 디코딩 과정 없이 비최대 억제기(340)에 의한 박스 처리가 수행될 수 있다. 이때, 소프트맥스 처리기(310)는 신뢰값(x_j)이 가장 큰 값에 대해서만 확률값을 구한다. 그 이유는 한 화면 내에서 해당 영역을 차지하는 객체는 하나만 존재하기 때문이다.

[0042] 박스 생성기(320)는 위치 정보를 출력한다. 위치 정보는 각 레이어의 출력 데이터의 픽셀 위치에 대응될 수 있다. 위치 정보의 개수는 기본 박스의 개수에 따라 결정될 수 있다. 예를 들어, 기본망(210)의 출력 데이터가 256채널 및 5x5의 해상도를 갖는다면, 25개의 위치 정보가 존재할 수 있다. 이때 픽셀당 3개의 기본 박스를 갖는다면, 총 75개의 위치 정보가 존재할 수 있다. 각각의 위치 정보는 4개의 데이터로 구성된다. 각각의 위치 정보에 따라 모든 객체가 후보가 될 수 있기 때문에 신뢰망(220)의 출력수는 총 1575개(박스 75개x다중 객체

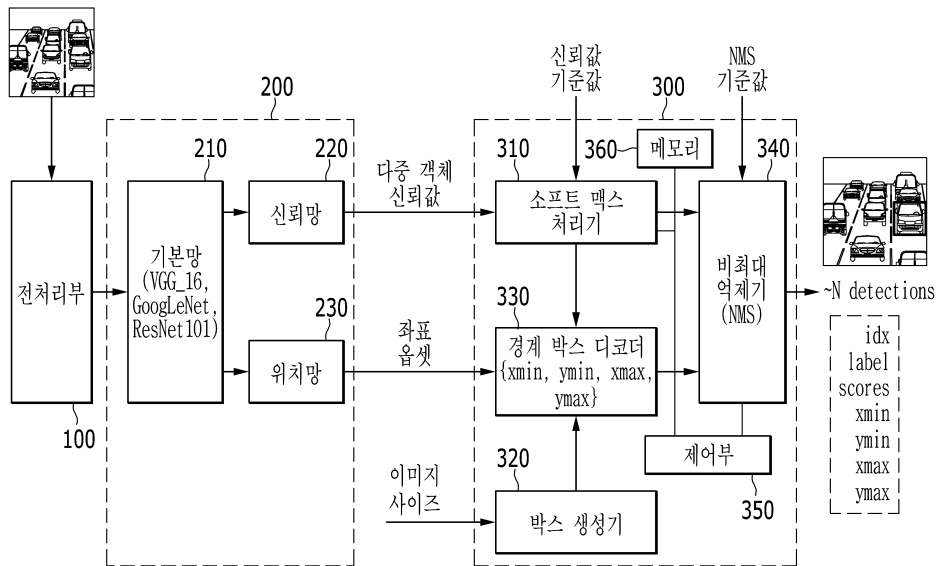
21개)가 된다. 이는 75개의 박스 별로 총 21개의 확률값을 출력함을 의미한다.

- [0043] 도 5는 한 실시예에 따른 박스 생성기에 의한 기본 박스 생성을 나타내는 도면이다.
- [0044] 도 5를 참조하면, 640x480 해상도의 영상이 입력된 후에 기본망(210)의 공간 영역(partial domain)이 7x7 어레이(array)로 출력되고, 가운데 회색으로 표시된 중심 박스에 대해 처리하면, 중심점은 $\{cx, cy\} = \{320, 240\}$ 이 된다. 중심점을 기준으로 박스의 크기, 박스의 높이, 및 폭의 비율(aspect ratio)에 따라 다양한 기본 박스가 정의되어 사용될 수 있다. 박스의 종류가 많아질수록 연산해야 하는 박스의 수가 늘어나기 때문에, 목표 검출 성능에 따라 적절한 양의 후보 박스 수가 정의되어야 한다. 어플리케이션에 따라 검출될 최소 객체의 사이즈 및 객체의 종류의 특징 등을 반영하여 박스의 크기와 비율이 선정될 수 있다.
- [0045] 경계 박스 디코더(330)는 박스 생성기(320)로부터 출력된 기본 위치 정보에 위치망(230)으로부터 출력된 위치 정보에 대한 오프셋(offset)을 반영하여, 객체의 정확한 위치 정보를 출력할 수 있다. 기본 박스의 위치 정보는 도 3에 도시된 바와 같이, 꼭지점, 중심값, 높이나 폭 데이터로 표현될 수 있다. 단일 박스를 표현하는 데 필요한 데이터의 수는 4가 된다. 박스 생성기(320)에서 설명한 바와 같이, 100개의 위치 정보가 있을 때 총 400개의 데이터가 발생할 수 있다. 위치 정보는 전체 화면의 가로 및 세로의 크기의 값을 1로 정규화하여 0~1 사이의 값을 갖는 데이터이다. 위치 정보는 입력 영상의 해상도가 바뀌어도 동일하게 동작하게끔 해준다.
- [0046] 도 6은 한 실시예에 따른 경계 박스 디코더의 동작을 나타내는 도면이다.
- [0047] 도 6을 참조하면, 박스의 표현 방식은 2개의 꼭지점을 사용하는 방식이며, 오프셋은 기본 박스에서 실제 박스까지 움직이는 벡터로 표현된다. 벡터 값은 학습 과정을 통해 도출될 수 있다.
- [0048] 비최대 억제기(340)는 소프트맥스 처리기(310)로부터 출력된 확률들을 높은 순으로 정렬한 후, 각 확률값에 대응하는 박스 정보에서 겹치는 박스 정보를 제거한다. 이를 통해, 동일한 객체에 여러 개의 박스가 중첩되어 나타나는 현상을 제거할 수 있다. 2개의 박스가 겹치는 양은 자카드 오버랩(jaccard overlap) 정보를 통해 계산된다. 2개의 박스가 동일하게 겹치는 경우 자카드 오버랩 값은 '1'이 되고, 전혀 겹치지 않는 경우 '0'이 된다. 자카드 오버랩(jaccard overlap) 정보는 0~1의 값의 범위를 갖는다.
- [0049] 도 7은 한 실시예에 따른 경계 박스 디코더의 동작을 나타내는 도면이고, 도 8은 한 실시예에 따른 자카드 오버랩을 나타내는 도면이다.
- [0050] 도 7 및 도 8을 참조하면, 경계 박스 디코딩 전 박스의 수가 많으면, 경계 박스 디코더(330)에 의한 경계 박스 디코딩 과정에서 많은 수의 박스가 검출될 수 있다. 비최대 억제기(340)는 많은 박스들 중에서 가장 확률값이 큰 박스만을 검출한다. 이때, 비최대 억제기(340)는 도 8에 도시된 것과 같이 자카드 오버랩을 이용하여 비슷한 박스를 걸러낼 수 있다. 단일 화면상에 유사한 객체가 겹쳐서 존재할 수도 있기 때문에, 겹침 정도는 NMS 기준 값을 통해 조정될 수 있다.
- [0051] 객체 검출부(300)는 비최대 억제기(340)에 의한 겹치는 박스 정보 제거 후, 최종적으로 객체의 클래스 번호를 나타내는 인덱스 정보, 객체의 이름을 나타내는 라벨 정보, 객체의 확률값을 나타내는 스코어 정보, 그리고 객체의 위치를 나타내는 박스 정보(xmin, ymin, xmax, ymax)를 출력한다.
- [0052] 도 1을 참조하면, 한 실시예에 따른 객체 인식 장치(1)는 컨볼루션 뉴럴 네트워크(Convolution Neural Networks, CNN)를 사용하여 수행되는 뉴럴넷 학습에 기반하여 객체의 신뢰값을 출력하는 학습부(200), 그리고 신뢰값을 객체가 특정 객체 종류에 해당할 확률을 나타내는 확률값으로 변환하고, 확률값과 기저장된 기준값의 비교결과에 기반하여 객체에 대응하는 박스를 생성하는 객체 검출부(300)를 포함한다.
- [0053] 학습부(200)는 입력 영상 객체의 신뢰값을 출력하는 신뢰망(220)을 포함한다. 객체는 복수의 객체를 포함하는 다중 객체이며, 학습부(200)는 다중 객체에 대한 복수의 신뢰값을 출력한다.
- [0054] 객체 검출부(300)는 신뢰값을 확률값으로 변환하는 소프트맥스 처리기(310), 박스 생성기(320), 경계 박스 디코더(330), 비최대 억제기(340), 및 제어부(350)를 포함할 수 있다.
- [0055] 소프트맥스 처리기(310)는 학습부(200)로부터 출력된 복수의 신뢰값을 객체가 특정 객체 종류에 해당할 확률을 나타내는 복수의 확률값으로 변환한다. 소프트맥스 처리기(310)는 복수의 확률값 중에서 가장 큰 확률값을 갖는 객체의 확률값이 기저장된 기준값인 신뢰도 문턱값(confidence threshold) 보다 큰 경우에만, 그 객체에 대해 인덱스를 생성하고, 생성된 인덱스와 확률값을 저장한다. 기준값은 설계자의 설정에 따라 달라질 수 있는 값이며, 미리 설정되어 메모리(360)에 저장된 값이다.

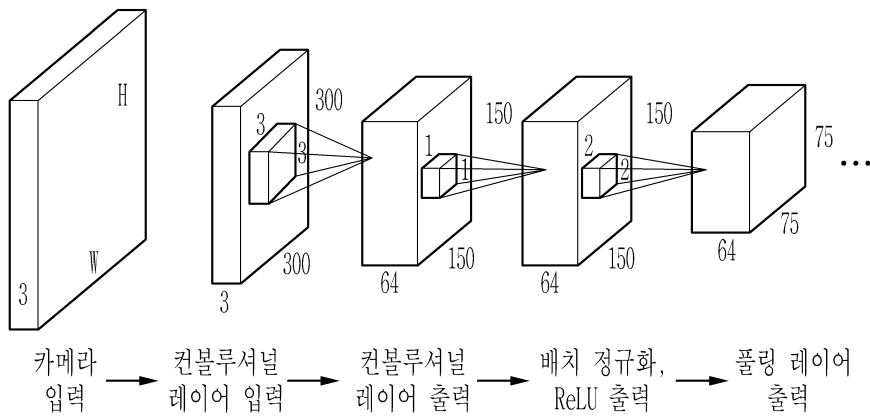
- [0056] 박스 생성기(320)는 인덱스가 생성된 객체에 대해서만 박스를 생성한다. 경계 박스 디코더(330)는 뉴럴넷으로부터 출력된 읍셋을 반영하여 박스 생성기(320)로부터 생성된 박스의 정확한 위치 정보를 출력한다.
- [0057] 비최대 억제기(340)는 각 확률값에 대응하는 출력된 위치 정보에서 서로 겹치는 위치 정보를 제거한다.
- [0058] 이를 통해, 불필요한 박스 생성에 소요되는 시간을 절약할 수 있고, 생성된 박스 정보를 이용하여 정확한 박스 정보를 생성하는 경계 박스 디코더(330)의 처리시간도 절약할 수 있다. 예를 들어, 지원 가능한 박스의 수가 전체 레이어를 통틀어 10,000개라고 가정하면, 객체 검출부(300)는 박스 생성이나 디코딩 동작을 10,000번 수행하므로 많은 동작 시간이 소요될 수 있다. 그러나, 본 기재에 따르면 객체 검출부(300)는 높은 확률값을 갖는 객체에 대해서만 박스 디코딩 및 NMS 처리를 수행하므로, 동작 시간을 크게 단축시킬 수 있다.
- [0059] 도 9는 한 일 실시예에 따른 객체 인식 장치의 동작 방법의 순서도이다.
- [0060] 도 9를 참조하면, 한 실시예에 따른 객체 인식 방법은, 컨볼루션 뉴럴 네트워크(Convolution Neural Networks, CNN)을 사용하여 수행되는 뉴럴넷 학습에 기반하여, 객체의 신뢰값을 출력하는 단계(S30), 신뢰값을 객체가 특정 객체 종류에 해당할 확률을 나타내는 확률값으로 변환하는 단계(S40), 그리고 확률값과 기저장된 기준값의 비교결과에 기반하여 객체에 대응하는 박스를 생성하는 단계(S90)를 포함한다.
- [0061] 객체는 복수의 객체를 포함하는 다중 객체이며, 객체의 신뢰값을 출력하는 단계는 다중 객체에 대한 복수의 신뢰값을 출력하며(S30),
- [0062] 확률값으로 변환하는 단계는 복수의 신뢰값을 복수의 확률값으로 변환하고(S40),
- [0063] 박스를 생성하는 단계는 복수의 확률값 중에서 가장 큰 확률값을 갖는 객체의 확률값이 기준값 보다 크면(S50), 복수의 확률값 중에서 가장 큰 확률값을 갖는 객체에 대해 인덱스를 생성하고(S70), 인덱스가 생성된 객체에 대해서 박스를 생성한다(S90).
- [0064] 박스를 생성하는 단계 이후에는, 뉴럴넷으로부터 출력된 읍셋에 기반하여 박스의 위치 정보를 도출한다(S110).
- [0065] 위치 정보를 도출하는 단계 이후에는, 각 확률값에 대응하는 도출된 위치 정보에서 서로 겹치는 위치 정보를 제거한다(S150).
- [0066] 객체 인식 장치의 동작 방법은 위에서 설명한 학습부(200) 및 객체 검출부(300)의 동작 방법과 동일하므로, 상세한 설명은 생략한다.
- [0067] 이상에서 본 발명의 실시예에 대하여 상세하게 설명하였지만 본 발명의 권리범위는 이에 한정되는 것은 아니고 다음의 청구범위에서 정의하고 있는 본 발명의 기본 개념을 이용한 당업자의 여러 변형 및 개량 형태 또한 본 발명의 권리범위에 속하는 것이다.

도면

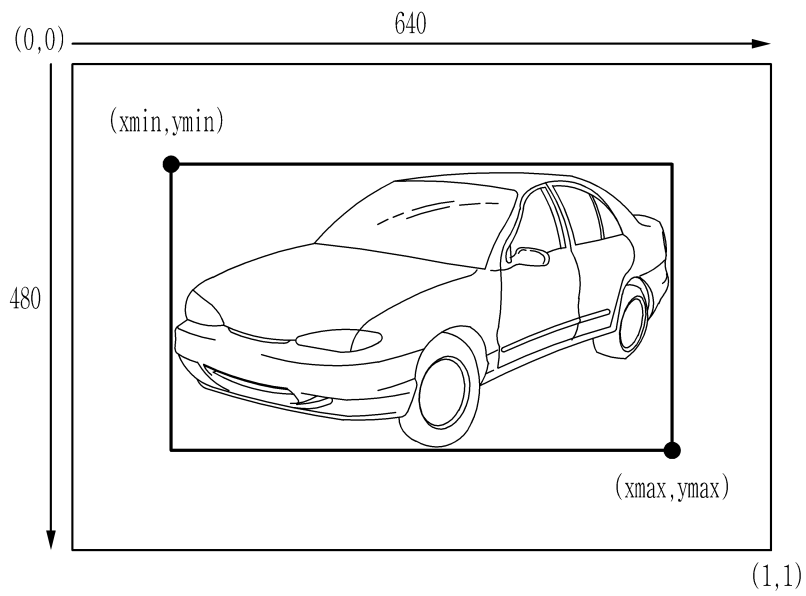
도면1



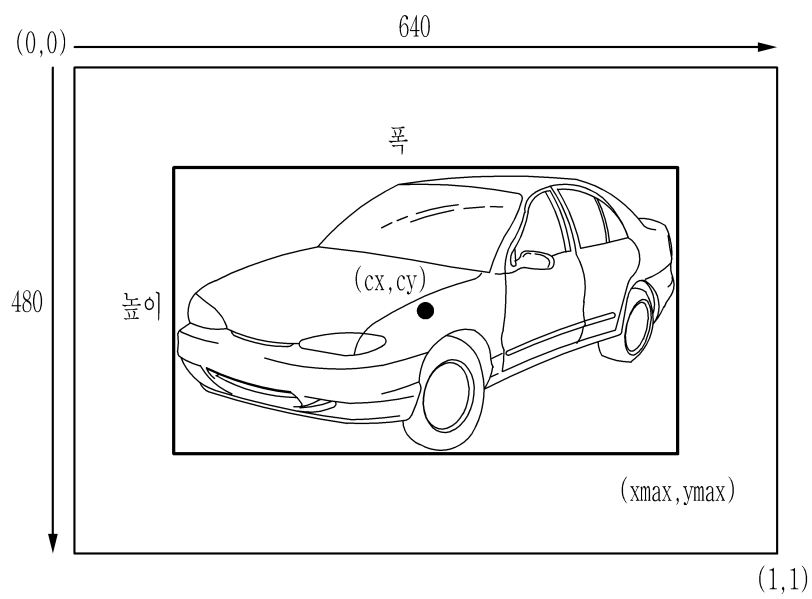
도면2



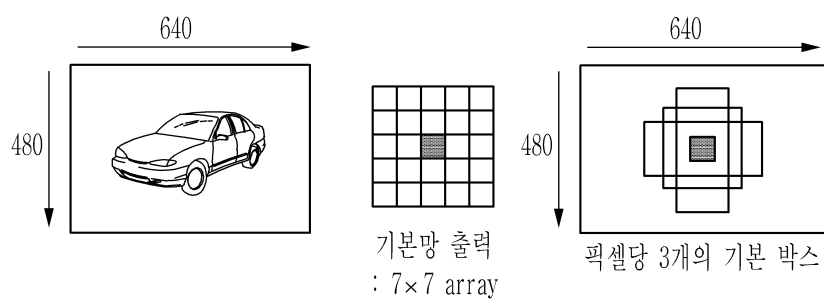
도면3



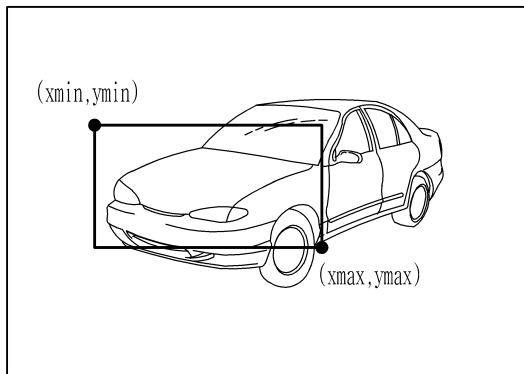
도면4



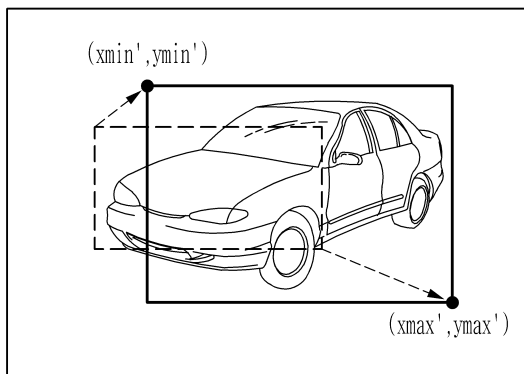
도면5



도면6

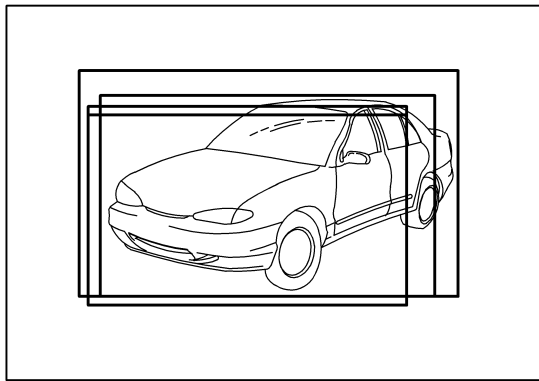


기본 박스

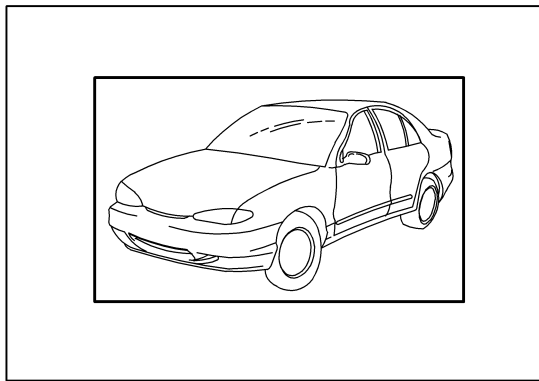


오프셋 적용 후

도면7

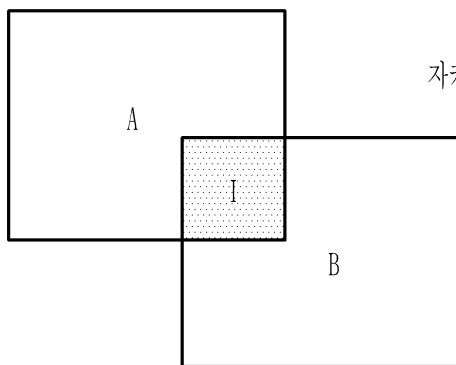


NMS 전



NMS 후

도면8



자카드 오버랩: $I/(A+B-I)$

도면9

