



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0146269
(43) 공개일자 2023년10월19일

(51) 국제특허분류(Int. Cl.)

G06V 20/52 (2022.01) G06N 3/04 (2023.01)
G06V 10/44 (2022.01) G06V 10/774 (2022.01)
G06V 40/10 (2022.01)

(52) CPC특허분류

G06V 20/52 (2023.08)
G06N 3/045 (2023.01)

(21) 출원번호 10-2022-0045032

(22) 출원일자 2022년04월12일

심사청구일자 2022년04월12일

(71) 출원인

인천대학교 산학협력단

인천광역시 연수구 갯벌로 27, 인천대학교 이노베이션센터(송도동)

(72) 발명자

전광길

인천광역시 연수구 해송로 70, 209동 805호(송도동, 송도웹카운티2단지)

(74) 대리인

특허법인충정

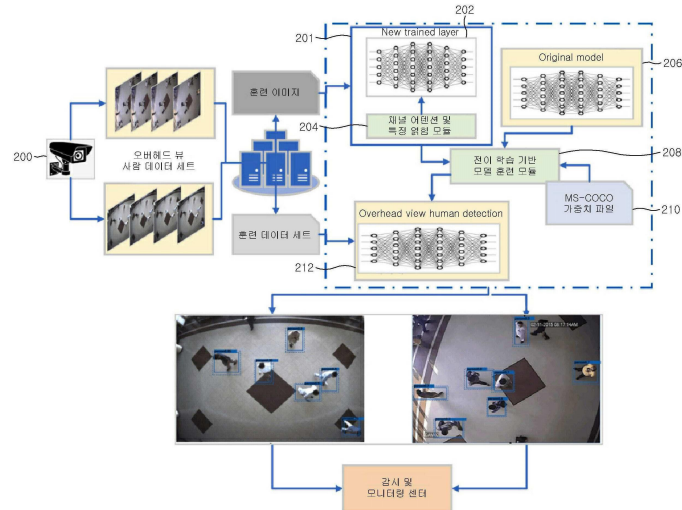
전체 청구항 수 : 총 5 항

(54) 발명의 명칭 감시를 위한 심층 신경망 기반 사람 탐지 시스템

(57) 요약

본 발명의 일 실시예에 의한 감시를 위한 심층 신경망 기반 사람 탐지 시스템은, 오버헤드 뷰 이미지들의 데이터를 훈련 데이터 세트로서 사용하고, 채널 어텐션(channel attention) 및 특징 얽힘 모듈이 결합된 변형된 가우시안(Gaussian) YOLOv3 모델; 일반적인 복수의 데이터 세트로서 사전 훈련된 오리지널 가우시안 YOLOv3 모델; 전이 학습에 기반하여 상기 오리지널 가우시안 YOLOv3 모델과 상기 변형된 가우시안 YOLOv3 모델을 결합하는 전이 학습 기반 모델 훈련 모듈; 및 상기 전이 학습 기반 모델 훈련 모듈에 의해 생성되고, 오버헤드 뷰 이미지들의 데이터를 훈련 데이터 세트로서 사용하여 오버헤드 뷰 이미지에서 사람이 포함된 바운딩 박스를 탐지하도록 다시 훈련되는 오버헤드 뷰 사람 탐지 모델을 포함한다.

대표도



(52) CPC특허분류

G06V 10/449 (2023.08)

G06V 10/774 (2023.08)

G06V 40/10 (2022.01)

G06T 2207/30196 (2013.01)

G06T 2210/12 (2013.01)

명세서

청구범위

청구항 1

오버헤드 뷰 이미지들의 데이터를 훈련 데이터 세트로서 사용하고, 채널 어텐션(channel attention) 및 특징 얹힘 모듈이 결합된 변형된 가우시안(Gaussian) YOLOv3 모델;

일반적인 복수의 데이터 세트에 사전 훈련된 오리지널 가우시안 YOLOv3 모델;

전이 학습에 기반하여 상기 오리지널 가우시안 YOLOv3 모델과 상기 변형된 가우시안 YOLOv3 모델을 결합하는 전이 학습 기반 모델 훈련 모듈; 및

상기 전이 학습 기반 모델 훈련 모듈에 의해 생성되고, 오버헤드 뷰 이미지들의 데이터를 훈련 데이터 세트로서 사용하여 오버헤드 뷰 이미지에서 사람이 포함된 바운딩 박스를 탐지하도록 다시 훈련되는 오버헤드 뷰 사람 탐지 모델을 포함하는, 감시를 위한 심층 신경망 기반 사람 탐지 시스템.

청구항 2

청구항 1에 있어서,

상기 채널 어텐션 및 특징 얹힘 모듈은,

상기 변형된 가우시안 YOLOv3 모델의 입력 특징 맵(F)에 대해 수학식 1에 기반하여 채널 어텐션을 수행하고,

[수학식 1]

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \\ = \sigma(W_1(W_0(F_{\text{avg}}^c)) + W_1(W_0(F_{\text{max}}^c)))$$

F는 입력 특징 맵, $M_c(F)$ 는 채널 어텐션 맵, F_{avg}^c 는 평균-풀(average-pooled) 특징으로서 특징 맵의 각 채널에 대한 전체 평균, F_{max}^c 는 맥스-풀(max-pooled) 특징으로서 특징 맵의 각 채널에서의 최대값, W_0 및 W_1 은 3-레이어 퍼셉트론의 가중치, σ 는 시그모이드 활성화 함수를 나타내는, 감시를 위한 심층 신경망 기반 사람 탐지 시스템.

청구항 3

청구항 2에 있어서,

상기 채널 어텐션 및 특징 얹힘 모듈은,

동일한 특징 맵 계층에서 서로 다른 채널 간 연결을 추가로 수행하는, 감시를 위한 심층 신경망 기반 사람 탐지 시스템.

청구항 4

청구항 1에 있어서,

상기 변형된 가우시안(Gaussian) YOLOv3 모델 및 상기 오리지널 가우시안(Gaussian) YOLOv3 모델 각각은,

예측된 바운딩 박스의 좌표를 가우시안 매개변수로 모델링하여 상기 예측된 바운딩 박스의 불확실성을 추정하는, 감시를 위한 심층 신경망 기반 사람 탐지 시스템.

청구항 5

청구항 4에 있어서,

상기 바운딩 박스의 중심 좌표 값을 (t_x , t_y)라 하고, 상기 바운딩 박스의 너비와 높이를 각각 t_w , t_h 라 할 때,

상기 바운딩 박스의 불확실성은 상기 바운딩 박스의 중심 좌표 중 x축 좌표의 평균(μ_{t_x}), 상기 바운딩 박스의 중심 좌표 중 y축 좌표의 평균(μ_{t_y}), 상기 바운딩 박스의 너비의 평균(μ_{t_w}), 상기 바운딩 박스의 높이의 평균(μ_{t_h}), 상기 바운딩 박스의 중심 좌표 중 x축 좌표의 분산($\sum_{t_x}$), 상기 바운딩 박스의 중심 좌표 중 y축 좌표의 분산($\sum_{t_y}$), 상기 바운딩 박스의 너비의 분산($\sum_{t_w}$) 및 상기 바운딩 박스의 높이의 분산($\sum_{t_h}$)을 모델링하여 예측되는, 감시를 위한 심층 신경망 기반 사람 탐지 시스템.

발명의 설명

기술 분야

[0001] 본 발명은 스마트 도시 및 사회에서 감시를 위한 심층 신경망 기반 사람 탐지 시스템에 관한 것이다.

배경 기술

[0002] 심층 신경망은 지속 가능한 스마트 도시와 사회를 위한 많은 고급 인공 지능 기반 응용 프로그램의 기초로 간주되었다. 이러한 애플리케이션에는 스마트 의료 시스템, 스마트 보안, 스마트 자율주행 또는 자율주행 차량, 로봇 공학, 스마트 감시, 스마트 도시를 포함한다. 심층 신경망은 기존의 기계 학습 기반 기술에 비해 이러한 응용 프로그램의 성능을 최신 수준 이상으로 크게 개선했다. 이 중에서 사람 탐지는 비정상 이벤트 탐지, 군중 모니터링, 사람 계수, 사람 및 보행자 식별, 사람 보행 특성화, 성별 분류 및 노인 낙상 탐지에 국한되지 않고 광범위한 응용 분야로 인해 연구자들의 더 많은 관심을 받고 있다.

[0003] 이러한 응용 프로그램은 모양과 변형 가능한 포즈가 다양한 인체의 특성과 특성으로 인해 여러 가지 문제를 겪을 수 있다. 또한 복잡하고 혼잡한 환경, 조명 변화, 다양한 유형의 폐색 및 다양한 카메라 관점이 사람 탐지 시스템의 성능에 영향을 미칠 수 있다. 이러한 문제를 해결하기 위해 다양한 머신 러닝 및 심층 신경망 기반 방법이 제안되고 개발된다. 이러한 방법은 일반적으로 기존의 수작업 특징과 다양한 기계 학습 분류기를 기반으로 한다. 다른 한편으로, 심층 학습에서의 진전은 효율성 및 계산 속도 측면에서 사람 탐지 작업을 보다 강력하고 효과적으로 만든다.

[0004] 심층 신경망의 주요 이점은 분류 가능성이 더 높은 가장 중요하고 상위 수준 특징을 자동으로 선택하고 추출한다는 것이다. 게다가, 이러한 고급 기술은 일반적으로 다중 클래스 객체 분류 문제에 대해 추가적인 판별력을 갖는다. 개발된 사람 탐지 시스템의 대부분은 일반적으로 정면 또는 비대칭 카메라 관점을 기반으로 한다. 정면 뷰 이미지에서 변화하는 사람의 시각적 외관에 가장 크게 기여하는 것은 인체의 회전, 다양한 자세, 신체 움직임, 몸통 및 신체 관절 등. 이러한 상황에서 관점 변경의 영향은 일반적으로 무시할 수 있다.

[0005] 종래의 한 연구에서는 사람 탐지 응용 분야에 오버헤드 뷰 카메라를 채택했다. 오버헤드 원근법에서 인체의 모습은 지역화 회전, 독특한 신체 움직임 및 이미지 중심에 대한 위치에 따라 정면보다 비정상적이다. 오버헤드 원근법의 오클루전 문제는 장면이 어수선해지면 교차 객체 오클루전이 발생할 수 있는 비대칭 및 정면 원근법보다 상당히 적다. 이러한 특성의 결과 오버헤드 뷰 사람 탐지는 사람 카운팅, 사람 추적, 행동 인식, 군중 분석, 행동 이해, 사람 자세 식별과 같은 다양한 실제 응용 프로그램에서 선호된다. 오클루전 문제를 처리하는 것 외에도 오버헤드 카메라 관점은 계산 비용, 전송 부하, 에너지 소비 및 설치 비용을 극복할 수 있다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) KR 10-2319468 B1

발명의 내용

해결하려는 과제

[0007] 본 발명이 해결하고자 하는 과제는, 오버헤드 뷰 이미지에서의 사람 탐지 정확도를 높일 수 있는, 감시를 위한 심층 신경망 기반 사람 탐지 시스템을 제공하는 것이다.

과제의 해결 수단

[0008] 본 발명의 일 실시예에 의한 감시를 위한 심층 신경망 기반 사람 탐지 시스템은,
 [0009] 오버헤드 뷰 이미지들의 데이터를 훈련 데이터 세트로서 사용하고, 채널 어텐션(channel attention) 및 특징 얹힘 모듈이 결합된 변형된 가우시안(Gaussian) YOLOv3 모델;
 [0010] 일반적인 복수의 데이터 세트에 사전 훈련된 오리지널 가우시안 YOLOv3 모델;
 [0011] 전이 학습에 기반하여 상기 오리지널 가우시안 YOLOv3 모델과 상기 변형된 가우시안 YOLOv3 모델을 결합하는 전이 학습 기반 모델 훈련 모듈; 및
 [0012] 상기 전이 학습 기반 모델 훈련 모듈에 의해 생성되고, 오버헤드 뷰 이미지들의 데이터를 훈련 데이터 세트로서 사용하여 오버헤드 뷰 이미지에서 사람이 포함된 바운딩 박스를 탐지하도록 다시 훈련되는 오버헤드 뷰 사람 탐지 모델을 포함할 수 있다.
 [0013] 본 발명의 일 실시예에 의한 감시를 위한 심층 신경망 기반 사람 탐지 시스템에 있어서,
 [0014] 상기 변형된 가우시안 YOLOv3 모델에 포함된 상기 채널 어텐션 및 특징 얹힘 모듈은,
 [0015] 상기 변형된 가우시안 YOLOv3 모델의 입력 특징 맵(F)에 대해 수학식 1에 기반하여 채널 어텐션을 수행하고,
 [0016] [수학식 1]

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \\ = \sigma(W_1(W_0(F_{\text{avg}}^c)) + W_1(W_0(F_{\text{max}}^c))))$$

[0017]
 [0018] F는 입력 특징 맵, $M_c(F)$ 는 채널 어텐션 맵, F_{avg}^c 는 평균-풀(average-pooled) 특징으로서 특징 맵의 각 채널에 대한 전체 평균, F_{max}^c 는 맥스-풀(max-pooled) 특징으로서 특징 맵의 각 채널에서의 최대값, W_0 및 W_1 은 3-레이어 퍼셉트론의 가중치, σ 는 시그모이드 활성화 함수를 나타낼 수 있다.

[0019] 또한, 본 발명의 일 실시예에 의한 감시를 위한 심층 신경망 기반 사람 탐지 시스템에 있어서, 상기 채널 어텐션 및 특징 얹힘 모듈은,

[0020] 동일한 특징 맵 계층에서 서로 다른 채널 간 연결을 추가로 수행할 수 있다.

[0021] 또한, 본 발명의 일 실시예에 의한 감시를 위한 심층 신경망 기반 사람 탐지 시스템에 있어서, 상기 변형된 가우시안(Gaussian) YOLOv3 모델 및 상기 오리지널 가우시안(Gaussian) YOLOv3 모델 각각은,

[0022] 예측된 바운딩 박스의 좌표를 가우시안 매개변수로 모델링하여 상기 예측된 바운딩 박스의 불확실성을 추정할 수 있다.

[0023] 또한, 본 발명의 일 실시예에 의한 감시를 위한 심층 신경망 기반 사람 탐지 시스템에 있어서,

[0024] 상기 바운딩 박스의 중심 좌표 값을 (t_x , t_y)라 하고, 상기 바운딩 박스의 너비와 높이를 각각 t_w , t_h 라 할 때,

[0025] 상기 바운딩 박스의 불확실성은 상기 바운딩 박스의 중심 좌표 중 x축 좌표의 평균(μ_{t_x}), 상기 바운딩 박스의 중심 좌표 중 y축 좌표의 평균(μ_{t_y}), 상기 바운딩 박스의 너비의 평균(μ_{t_w}), 상기 바운딩 박스의 높이의 평균(μ_{t_h}), 상기 바운딩 박스의 중심 좌표 중 x축 좌표의 분산($\sum_{t_x}$), 상기 바운딩 박스의 중심 좌표 중 y축 좌표의 분산($\sum_{t_y}$), 상기 바운딩 박스의 너비의 분산($\sum_{t_w}$) 및 상기 바운딩 박스의 높이의 분산($\sum_{t_h}$)을 모

텔링하여 예측될 수 있다.

발명의 효과

[0026] 본 발명의 일 실시예에 의한 감시를 위한 심층 신경망 기반 사람 탐지 시스템에 의하면, 오버헤드 뷰 이미지에서의 사람 탐지 정확도를 높일 수 있다.

도면의 간단한 설명

[0027] 도 1은 스마트 도시 및 사회에서 스마트 심층 신경망 기반 오버헤드 뷰 사람 탐지 시스템의 개요를 도시한 도면.

도 2는 본 발명의 일 실시예에 의한 감시를 위한 심층 신경망 기반 사람 탐지 시스템을 도시한 도면으로서, 우선 사전 훈련된 알고리즘이 오버헤드 뷰 사람 데이터 세트에 대해 테스트되고, 추출된 특징 맵을 개선하기 위하여 채널 어텐션 및 특징 얹힘 모듈이 사용되며, 사람 탐지 시스템은 또한 전이 학습을 사용하여 오버헤드 뷰 사람 데이터 세트에 대해 재훈련되고, 새로 훈련된 레이어는 오리지널 모델과 인터페이스됨.

도 3은 오버헤드 뷰 사람 탐지 모듈의 개략적인 구조를 도시한 도면으로서, 상기 구조는 바운딩 박스의 추출을 위해 사용되고, 오버헤드 뷰 이미지에서 사람 탐지를 위해 적용되며, 특징 추출을 위해, 채널 어텐션 및 특징 얹힘 모듈이 컨볼루션 레이어와 함께 사용됨.

도 4는 채널 어텐션 모듈의 개략도로서, 공유 네트워크와 함께 맥스-풀링 및 평균-풀링 출력을 사용함.

도 5의 (a)는 YOLOv3의 예측된 특징 맵 및 바운딩 박스 속성들, (b)는 가우시안 YOLOv3 알고리즘의 가우시안 매개변수들을 사용하여 예측된 바운딩 박스의 속성들 및 요소들을 도시한 도면.

도 6은 오버헤드 뷰 데이터 세트에 대해 테스트된 사전 훈련된 알고리즘의 탐지 결과를 도시한 도면으로서, 테스트 이미지에서, 파란색 직사각형 내의 사람은 신뢰 점수값을 갖는 사람 클래스로 정확하게 분류된 사람이고, 다른 색상의 직사각형에서 사람은 다른 클래스 객체들, 예를 들어 개, 연, 오토바이 등으로 탐지되며, 점선으로 표시된 직사각형 박스들은 바운딩 박스 불확실성을 나타냄.

도 7은 다른 오버헤드 뷰 데이터 세트에 대해 테스트된 사전 훈련된 알고리즘의 탐지 결과를 도시한 도면으로서, 테스트 이미지에서, 파란색 직사각형 내의 사람은 신뢰 점수 값을 갖는 사람 클래스로서 정확하게 분류된 사람이고, 다른 색상의 직사각형에서, 사람은 다른 클래스 객체들, 예를 들어 새, 스케이트보드 등으로 탐지됨.

도 8은 훈련 및 테스트 손실을 도시한 그래프.

도 9는 훈련 및 테스트 정확도를 도시한 그래프.

도 10은 오버헤드 뷰 사람 데이터 세트에 대해 채널 어텐션, 특징 얹힘 및 추가적인 훈련 이후에 가우시안 YOLOv3 탐지 알고리즘의 결과를 도시한 도면으로서, 탐지 모델의 성능이 증가함으로 확인할 수 있고, 테스트 결과들은 본 발명이 오버헤드 뷰 장면의 다양한 위치에서 사람을 효과적으로 분류하고 탐지함을 보여줌.

도 11은 오버헤드 뷰 사람 데이터 세트에 대해 재훈련된 개선된 알고리즘의 결과를 도시한 도면으로서, 탐지 모델 성능이 전이 학습 이후에 증가함을 확인할 수 있고, 테스트 결과는 본 발명에 의한 모델이 이미지에서 사람을 효과적으로 분류하고 다양한 탐부 위치에서 사람을 탐지함을 보여줌.

도 12는 사전 훈련된 그리고 개선되도록 훈련된 알고리즘의 정밀도, 재현율, F1-스코어 및 정확도를 도시한 도면.

발명을 실시하기 위한 구체적인 내용

[0028] 본 발명의 목적, 특정한 장점들 및 신규한 특징들은 첨부된 도면들과 연관되어지는 이하의 상세한 설명과 바람직한 실시예들로부터 더욱 명백해질 것이다.

[0029] 이에 앞서 본 명세서 및 청구범위에 사용된 용어나 단어는 통상적이고 사전적인 의미로 해석되어서는 아니되며, 발명자가 그 자신의 발명을 가장 최선의 방법으로 설명하기 위해 용어의 개념을 적절하게 정의할 수 있는 원칙에 입각하여 본 발명의 기술적 사상에 부합되는 의미와 개념으로 해석되어야 한다.

[0030] 본 명세서에서 각 도면의 구성요소들에 참조번호를 부가함에 있어서, 동일한 구성 요소들에 한해서는 비록 다른

도면상에 표시되더라도 가능한 한 동일한 번호를 가지도록 하고 있음에 유의하여야 한다.

- [0031] 또한, "제1", "제2", "일면", "타면" 등의 용어는, 하나의 구성요소를 다른 구성요소로부터 구별하기 위해 사용되는 것으로, 구성요소가 상기 용어들에 의해 제한되는 것은 아니다.
- [0032] 이하, 본 발명을 설명함에 있어, 본 발명의 요지를 불필요하게 흐릴 수 있는 관련된 공지 기술에 대한 상세한 설명은 생략한다.
- [0033] 이하, 첨부된 도면을 참조하여 본 발명의 바람직한 실시형태를 상세히 설명하기로 한다.
- [0034] 본 발명에서는 심층 신경망 기반 객체 탐지 기술을 활용하여 사람 탐지에 가우시안 YOLOv3 알고리즘(Choi, Chun, Kim, & Lee, 2019, "Gaussian YOLOv3: An Accurate and Fast Object Detector Using Localization Uncertainty for Autonomous Driving")을 사용한다. 이 알고리즘은 좌표를 가우시안 매개변수로 모델링하여 정확도를 개선하고 거짓양성을 줄여 바운딩 박스 불확실성을 결정한다. 가우시안 YOLOv3는 채널 어텐션 및 특징 얹힘 모듈과 결합되어 특정 특징 맵을 개선한다. 채널 어텐션 모듈은 특징 맵과 결합되어 각 채널의 가중치를 자율적으로 학습하고 주요 특징을 개선하며 사람과 배경을 구별하는 네트워크의 능력을 향상시킨다. 동시에, 특징 맵의 다른 채널은 더 대표적인 특징을 얻기 위해 얹혀 있다. 마지막으로, 어텐션 및 특징 얹힘(Intertwin) 모듈에서 얻은 특징이 융합되어 향상된 특징 맵을 형성한다. 또한 사람 탐지 알고리즘의 탐지 정확도를 더욱 높이기 위해 전이 학습을 채택했다. 실험 결과는 훈련이 전체 탐지 정확도 94%로 사람 탐지에 대한 가우시안 YOLOv3 알고리즘의 잠재력을 향상시키는 것으로 나타났다.
- [0035] 비대칭 및 정면 카메라 원근법에 기반한 두 개의 서로 다른 카메라 관점 모두 인체의 시각적 특성(크기, 모양, 스케일, 포즈 및 신체 방향)에 상당한 변화가 있음을 알 수 있다. 오버헤드 뷰 환경에서 사람을 탐지하기 위해 스마트 감시 응용 프로그램을 위한 심층 신경망 기반 시스템이 제공된다.
- [0036] 가우시안(Gaussian) YOLOv3 알고리즘은 일반적으로 YOLOv3(Redmon & Farhadi, 2018, "YOLOv3: An Incremental Improvement")을 기반으로 하며, 각 그리드 셀에 대해 3개의 예측된 바운딩 박스를 결정한다. 여기서 예측된 박스는 바운딩 박스 좌표, 객체 점수 및 클래스 점수로 구성된다. 주요 차이점은 이제 5개의 다른 속성 정보 대신 예측된 바운딩 박스가 8개의 다른 속성 요소로 구성된다는 것이다. 각 바운딩 박스 좌표는 평균 및 분산 값으로 결정된다. 가우시안(Gaussian) YOLOv3는 또한 채널 어텐션(channel attention) 및 특징 얹힘 모듈과 결합된다. 채널 어텐션 모듈은 특징 맵과 연결되어 각 채널의 가중치를 학습한다. 주요 특징 맵을 개선하고 사람과 배경을 구별하는 네트워크의 능력을 향상시킨다. 동시에 특징 맵의 다른 채널은 더 대표적인 특징을 얻기 위해 얹혀 있다. 마지막으로 어텐션 및 특징 얹힘 모듈에서 받은 특징이 융합되어 개선된 특징 맵을 형성한다.
- [0037] 본 발명에서 알고리즘은 먼저 사전 훈련된 아키텍처로 테스트된다. 앞에서 논의한 바와 같이 오버헤드 뷰에서 본 사람의 시각적 특징은 정면과 크게 다르다. 따라서 알고리즘은 탐지되지 않거나 잘못된 결과를 제공한다. 따라서 전이 학습을 활용하여 두 개의 서로 다른 오버헤드 뷰 사람 데이터 세트를 사용하여 아키텍처를 재훈련시키고 새로운 훈련된 레이어를 기존 아키텍처와 결합한다. 추가 훈련은 사람 탐지 시스템의 정확도 결과를 향상시킨다. 일반적으로 본 발명의 주요 기여는 다음과 같다.
- [0038] - 스마트 도시 및 사회에서 스마트 감시 애플리케이션을 위한 심층 신경망 기반 사람 탐지 시스템을 도입한다. 시스템은 오버헤드 뷰 사람 탐지를 위해 가우시안 YOLOv3 알고리즘을 사용한다.
- [0039] - 주요 특징을 개선하고 사람과 배경을 구별하는 네트워크의 능력을 향상시키기 위해 채널 어텐션 모듈이 특징 맵과 결합된다. 특징 맵의 다른 채널도 더 대표적인 특징을 얻기 위해 얹혀 있다. 마지막으로, 어텐션 및 특징 얹힘(Feature Intertwine) 모듈에서 얻은 특징이 융합되어 향상된 특징 맵을 형성한다.
- [0040] - 사람 탐지 결과를 개선하기 위해 전이 학습을 사용하여 두 개의 서로 다른 오버헤드 뷰 데이터 세트에 대한 알고리즘 훈련.
- [0041] - 사전 훈련된 알고리즘과 개선된 훈련 알고리즘의 탐지 결과 비교.
- [0042] 오버헤드 뷰 기반 사람 탐지는 연구원들이 어려운 작업 중 하나로 인식하고 있다. 다양한 오버헤드 뷰 사람 탐지 방법은 다양한 배경 빼기 및 특징 기반 접근 방식을 기반으로 개발되었다.
- [0043] Ahmad, Ahmed, Khan, Qayum 및 Aljuaid(2020)는 오버헤드 뷰 장면에서 사람을 추적하기 위해 컨볼루션 신경망 기반 방법을 채택했다. 딥 러닝 기반 세분화 알고리즘도 오버헤드 뷰 사람 탐지를 위해 탐색된다(Ahmed, Ahmad, Khan, & Asif, 2020). Ahmed, Din, Jeon, Piccialli 및 Fortino(2020)는 탐지에 의한 다중 물체 추적을 위한

감시 시스템을 도입했다. 이 연구에서 저자는 딥 러닝 기반 객체 탐지 기술과 함께 다양한 추적 알고리즘을 사용했다. Ahmed, Ahmad, Ahmad, and Jeon(2020)은 YOLOv3 아키텍처와 함께 심층 SORT 알고리즘을 활용하는 오버헤드 뷰 사람 추적 시스템을 도입했다.

- [0044] 위의 문헌 검토를 기반으로 오버헤드 뷰 사람 탐지에 대해 제시된 대부분의 연구는 주로 특징 기반 방법을 기반으로 하는 것을 알 수 있다.
- [0045] 본 발명에서는 두 가지 다른 오버헤드 뷰 사람 기록 데이터 세트를 사용하여 사람 탐지를 위해 새로 개발된 가우시안 YOLOv3 알고리즘을 채택했다. 알고리즘은 또한 채널 어텐션(channel attention) 및 특징 얽힘(feature intertwine) 모듈과 결합되어 탐지 성능을 개선하고 오버헤드 뷰 데이터 세트에 대해 다시 훈련된다.
- [0046] 본 발명에서는 스마트 감시 응용, 즉 스마트 도시 및 사회에서 사람 탐지를 위한 심층 신경망 기반 시스템을 제안한다. 도 1에는 스마트 감시 시스템의 일반적인 구조가 나와 있다. 시스템은 주로 건물 내부에 설치된 스마트 카메라 장치 또는 센서로 구성된다. 이러한 스마트 카메라 장치는 무선 기술 또는 인터넷 연결을 통해 클라우드를 통해 데이터 센터로 전송되는 비디오 및 이미지를 기록한다. 데이터 센터는 사전 처리를 수행하고 심층 신경망 기반 감시 시스템으로 데이터를 전송하여 추가 분석 및 교육을 수행한다. 사전 처리 및 모델 교육을 수행한 후 분석된 데이터 결과는 다양한 실시간 스마트 감시 애플리케이션을 위해 모니터링 및 감시 센터로 다시 전송된다.
- [0047] 본 발명에서는 앞서 논의한 바와 같이 다양한 감시 애플리케이션에서 중요한 작업 중 하나로 간주되는 사람 탐지에 주로 초점을 맞추었다. 다양한 심층 신경망 기반 객체 탐지 알고리즘이 있지만 대부분 잘못된 지역화를 예상하지 못하고 바운딩 박스의 결정론적 결과를 제공하며 바운딩 박스의 불확실성을 고려하지 않는다. 따라서, 본 발명에서는 바운딩 박스의 좌표를 가우시안 매개변수로 모델링하여 바운딩 박스의 불확실성을 추정하는 가우시안(Gaussian) YOLOv3(Choi et al., 2019)를 활용했다.
- [0048] 알고리즘은 이전에 KITTI 및 버클리 딥 드라이브(BDD) 데이터 세트로 훈련되었으며 자율 주행 애플리케이션용으로 개발되었다. 본 발명에서는 오버헤드 뷰 사람 탐지를 위한 모델을 확장한다. 도 2는 본 발명의 일 실시예에 의한 감시를 위한 심층 신경망 기반 사람 탐지 시스템을 도시한 도면이다.
- [0049] 도 2를 참조하면, 본 발명의 일 실시예에 의한 감시를 위한 심층 신경망 기반 사람 탐지 시스템은, 오버헤드 뷰 카메라(200)로부터 획득된 오버헤드 뷰 이미지들의 데이터를 훈련 데이터 세트로서 사용하고, 채널 어텐션(channel attention) 및 특징 얽힘 모듈(204)을 포함하는 변형된 가우시안(Gaussian) YOLOv3 모델(201); 일반적인 복수의 데이터 세트에서 사전 훈련된 오리지널 가우시안 YOLOv3 모델(206); 전이 학습에 기반하여 상기 오리지널 가우시안 YOLOv3 모델(206)과 상기 변형된 가우시안 YOLOv3 모델(201)을 결합하는 전이 학습 기반 모델 훈련 모듈(208); 및 상기 전이 학습 기반 모델 훈련 모듈(208)에 의해 생성되고, 오버헤드 뷰 이미지들의 데이터를 훈련 데이터 세트로서 사용하여 오버헤드 뷰 이미지에서 사람이 포함된 바운딩 박스를 탐지하도록 훈련되는 오버헤드 뷰 사람 탐지 모델(212)을 포함할 수 있다.
- [0050] 참조번호 202는 가우시안(Gaussian) YOLOv3 모델을 나타내고, 참조번호 210는 MS-COCO 가중치 파일을 나타낸다.
- [0051] 오버헤드 뷰 카메라(200)에 의해 획득된 이미지의 데이터 세트가 훈련 샘플 및 테스트 샘플로 구분되는 것을 관찰할 수 있다. 먼저 MS-COCO(Common Object in Context) 데이터 세트 가중치 파일(210)을 사용하여 사전 훈련된 가우시안(Gaussian) YOLOv3 모델(206)을 훈련시켰다. 또한 채널 어텐션(channel attention) 및 특징 얽힘(feature intertwine) 모듈(204)을 새로 훈련된 가우시안(Gaussian) YOLOv3 모델(202)과 결합하여 특징 맵을 개선할 수 있다.
- [0052] 첫째, 채널 어텐션 모듈은 컨볼루션 신경망에서 채널 기반 어텐션에 사용되며, 이는 주요 특징 맵을 개선하고 전경 객체(사람)와 배경을 인식하는 네트워크의 능력을 향상시킬 수 있다.
- [0053] 둘째, 동시에 특징 맵의 다른 채널이 서로 얽혀 있어 더 설명적인 특징을 얻을 수 있다. 마지막으로, 어텐션 및 특징 얽힘 모듈에 의해 얻은 특징이 융합되어 개선된 특징 맵을 형성할 수 있다.
- [0054] 이전에 훈련된 아키텍처와 마찬가지로 사람/인체는 대부분 정면 또는 측면 관점에서 고려된다. 따라서 오버헤드 뷰 관점에서 사람 탐지에 대한 탐지 알고리즘의 정확도 결과를 향상시키기 위해 추가적인 훈련이 수행된다. 본 발명에서는 오버헤드 뷰 사람 데이터 세트에 기반하여 오버헤드 뷰 사람 탐지 모델(212)을 다시 훈련시켰다. 오버헤드 뷰 사람 탐지 모델(212)은 데이터 세트의 새로운 특징을 학습하고 테스트 데이터 세트를 활용하여 사람 탐지를 수행했다.

- [0055] 전이 학습 기반 모델 훈련 모듈(208)에서 전이 학습을 통해 새로 훈련된 모델(202)은 사전 훈련된 모델(206)과 결합되어 오버 헤드 뷰 사람 탐지 모델(212)이 생성되고, 오버 헤드 뷰 사람 탐지 모델(212)은 도 2에서와 같이 동일한 테스트 데이터 세트에 대해 테스트된다. 이 알고리즘은 오버헤드 뷰 이미지에 대한 사람 탐지 결과를 개선한다. 훈련되고 사전 훈련된 알고리즘 결과는 모니터링 애플리케이션을 위해 평가되고 추가로 처리된다. 이하, 오버헤드 뷰 사람 탐지에 사용되는 방법에 대해 상세히 설명하기로 한다.
- [0056] 가우시안 YOLOv3는 일반적으로 YOLOv3을 기반으로 한다. 그것은 YOLO(You Only Look Once)의 객체 탐지 알고리즘의 확장 버전이다. 이 버전에서는 알고리즘의 이전 버전에 비해 특히 작은 객체의 정확도가 크게 향상되었다. YOLOv3의 주요 변경 사항에는 백본 네트워크 조정, 객체 탐지를 위한 다중 스케일 특징 활용, 객체 분류를 위해 Softmax를 Logistic 레이어로 교체 등이 있다.
- [0057] 본 발명의 일 실시예에 의한 감시를 위한 심층 신경망 기반 사람 탐지 시스템의 네트워크 아키텍처가 도 3에 나와 있다. Darknet-53은 모든 잔여 모듈이 바로 가기 연결과 두 개의 컨볼루션 레이어로 구성된 ResNet 아키텍처의 아이디어를 적용한 백본 네트워크로 활용된다. YOLOv3는 잔여 스킵 연결을 활용하여 심층 네트워크의 그라디언트 소실 문제를 결정한다. YOLOv3는 업샘플링 레이어와 캐스캐이딩 시스템을 적용하여 작은 크기의 객체에 대한 세밀한 특징을 유지한다. 모든 컨볼루션 계층은 ReLU 활성화 및 배치 정규화 계층에서 지원된다.
- [0058] 도 3에 도시된 아키텍처는 특징 추출기와 탐지기의 두 가지 주요 부분으로 나뉜다. 특징 추출기는 입력 이미지의 특징 맵을 3가지 다른 스케일로 추출한다. 다음으로, 1×1 탐지 커널을 적용하여 이러한 맵에서 예측이 이루어진다. 예측 특징 맵의 모든 셀은 3개의 바운딩 박스를 예측한다. 각 바운딩 박스에는 중심 좌표, 높이, 너비, 객체 점수 및 클래스 신뢰도를 포함하는 속성 정보가 있다. 입력 이미지는 최종 특징 맵과 동일하게 예측을 담당하는 특징 맵의 셀을 결정하기 위해 $S \times S$ 차원의 그리드 셀로 분할된다. 이미지에서 탐지된 각 객체에 대해 그리드 셀은 예측을 담당하는 대상 객체/사람의 실측 정답(ground truth) 바운딩 박스의 중심으로 구성된다. 여기서 실측 정답(ground truth)은 학습하고자 하는 데이터의 원본 혹은 실제 값을 의미한다.
- [0059] 특징 맵의 모든 셀은 검출된 바운딩 박스 정보 중 하나를 통해 객체 예측에 활용된다. 특징 맵의 각 그리드 셀은 도 5a와 같이 3개의 바운딩 박스를 예측한다. YOLOv3는 바운딩 박스의 너비와 높이를 예측하는 대신 앵커라고도 하는 미리 정의된 상자에 대한 오프셋을 예측한다. 앵커 박스(Anchor box)는 일반적으로 훈련 데이터 세트에서와 같이 실측 정답(ground truth) 바운딩 박스(bounding box)를 매핑하기 위해 결정된 미리 정의된 상자/모양이며, 대부분의 객체는 평균 너비와 높이 비율을 갖는다. 따라서 이러한 오프셋은 앵커 박스에서 사용하여 객체에 대한 예측을 얻을 수 있다.
- [0060] 결과적으로 물체의 탐지를 담당하는 바운딩 박스(본 발명에서는 사람)는 정답 바운딩 박스가 있는 앵커의 결합에 대한 교차 비율이 가장 높다. 또한, 본 발명의 일 실시예에서는 도 4에 도시된 채널 어텐션 모듈을 특징의 채널 간 관계를 사용하여 아키텍처와 결합했다. 각 채널의 특징 맵이 특징 탐지기로 인식되므로 채널 어텐션은 주어진 입력 이미지에 대해 의미 있는 특징을 추출할 수 있다.
- [0061] 도 4에서 $F \in R^{H \times W \times C}$ 는 입력 특징 맵이고, $F' \in R^{H \times W \times C}$ 는 채널 어텐션을 나타낸다. 특징 맵에서 W, H, C는 너비, 높이 및 채널 수를 의미하고 + 표시가 있는 원(oplus)은 요소별 곱셈을 의미한다. 채널 어텐션 모듈에 대한 계산은 다음 수학적식을 사용하여 완료된다.

수학적식 1

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \\ = \sigma(W_1(W_0(F_{\text{avg}})) + W_1(W_0(F_{\text{max}}))))$$

[0062]

- [0063] 위의 수학적식에서 $M_c(F)$ 는 F 채널 어텐션 맵을 생성함을 나타낸다. 여기서, F_{avg}^c 는 평균-풀(average-pooled) 특징으로서 특징 맵의 각 채널에 대한 전체 평균이고, F_{max}^c 는 맥스-풀(max-pooled) 특징으로서 특징 맵의 각 채널에서의 최대 값이다. 위의 각 특징 값에는 이미지 배경 정보를 효과적으로 학습할 수 있는 피드백이 있다.

맥스 풀링(MaxPool)은 두드러진 특징 값 정보를 수집할 수 있다.

[0064] 멀티레이어 퍼셉트론(MLP: Multilayer Perceptron)이 사용되는데, 여기서 $(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c)))$ 는 3-레이어 퍼셉트론의 가중치, σ 는 시그모이드 활성화 함수이다. W_0 및 W_1 은 멀티레이어 퍼셉트론(MLP: Multilayer Perceptron)의 가중치이고, 네트워크의 비선형성을 향상시키기 위하여, W_0 이후 ReLu(Rectified Linear Unit) 활성화 함수가 뒤따를 수 있다.

[0065] 특징 추출을 위해 컨볼루션 신경망에서 주로 사용하는 방법은 계층 구조를 기반으로 하며 각 계층의 각 채널이 동일한 방식으로 추출된다. 상기 네트워크는 특징들을 추출하기 위한 각 계층에 대한 다중 스케일 컨볼루션 커널을 사용하거나, 융합 기반 방법을 사용할 수 있다.

[0066] 본 발명에서 특징 얹힘 모듈은 동일한 특징 맵 계층에서 서로 다른 채널 간의 연결을 구축하여 채널 어텐션을 결합한다. 이 방법의 특성은 전송 정보의 손실 문제를 해결하고 작은 객체에 대한 네트워크의 특징 추출 능력을 향상시키는 것이다.

[0067] YOLOv3의 예측된 특징 맵은 도 5a에 나와 있다. 3개의 예측 박스들이 각 그리드 셀에 대해 획득됨을 알 수 있다. 탐지된 각 예측 바운딩 박스는 다음 좌표 값(예: t_x , t_y , t_w , t_h), 객체성(객체 점수) 및 클래스 점수로 구성된다. 객체 점수는 객체가 이미 바운딩 박스에 있는지 여부를 결정한다. 각 객체의 범주 또는 클래스는 클래스 점수로 결정되며 그 값은 0에서 1 사이이다. 이 모델은 객체 탐지를 위해 이 두 값의 곱을 사용한다. 바운딩 박스 좌표 값은 객체 및 클래스 정보와 달리 결정적 좌표 값으로 출력에서 정의된다. 따라서 탐지된 바운딩 박스 신뢰도를 알 수 없다.

[0068] 또한, 바운딩 박스의 객체 점수는 검출된 바운딩 박스의 신뢰도를 나타내지 않고 바운딩 박스 불확실성 문제를 야기하는데, 이것은 검출된 바운딩 박스의 결과가 얼마나 불확실한지를 인식하지 못한다는 것을 의미한다. 따라서 이 문제를 해결하기 위해 가우시안(Gaussian) YOLOv3가 채택되었다.

[0069] YOLOv3는 도 5a에 도시된 바와 같이 $(t_x \& t_y)$ 로 표현되는 바운딩 박스 중심 정보를 추출하기 위해 바운딩 박스 리그레션(regression)을 사용하고, $(t_w \& t_h)$ 로 바운딩 박스 크기 정보를 나타낸다. 탐지된 각 경계에 대해 실측 정답인 정확한 답이 하나 있다. 따라서 지역화 불확실성의 예측을 위해 복잡한 모델링이 필요하지 않다.

[0070] 바운딩 박스의 불확실성은 $t_x, t_y, t_w, \& t_h$ 의 단일 가우시안 모델을 사용하여 모델링될 수 있다. 주어진 입력 이미지 x 에 대해 가우시안 모델 출력 y 는 다음과 같이 제공된다(Choi et al., 2019).

수학식 2

$$p(y|x) = N(y; \mu(x), \Sigma(x))$$

[0072] 상기 수학식에서, $\mu(x)$ 는 평균이고 $\Sigma(x)$ 는 분산 함수로 사용된다. 바운딩 박스의 불확실성은 도 5b와 같이 예측 특징에서 탐지된 각 바운딩 박스 좌표의 μ (평균) 및 Σ (분산)을 모델링하여 예측된다. 상기 바운딩 박스의 중심 좌표 값을 (t_x, t_y) 라 하고, 상기 바운딩 박스의 너비와 높이를 각각 t_w, t_h 라 할 때, 상기 바운딩 박스의 불확실성은 상기 바운딩 박스의 중심 좌표 중 x축 좌표의 평균(μ_{t_x}), 상기 바운딩 박스의 중심 좌표 중 y축 좌표의 평균(μ_{t_y}), 상기 바운딩 박스의 너비의 평균(μ_{t_w}), 상기 바운딩 박스의 높이의 평균(μ_{t_h}), 상기 바운딩 박스의 중심 좌표 중 x축 좌표의 분산(Σ_{t_x}), 상기 바운딩 박스의 중심 좌표 중 y축 좌표의 분산(Σ_{t_y}), 상기 바운딩 박스의 너비의 분산(Σ_{t_w}) 및 상기 바운딩 박스의 높이의 분산(Σ_{t_h})을 모델링하여 예측될 수

있다.

[0073] 바운딩 박스의 출력은 μ_{t_x} , μ_{t_y} , μ_{t_w} , μ_{t_h} , $\sum_{t_x}$, $\sum_{t_y}$, $\sum_{t_w}$ 및 $\sum_{t_h}$ 로 표시된다. 탐지 계층에서 모든 좌표의 평균값은 바운딩 박스의 예측 좌표이며 각 분산 값은 각 좌표의 불확실성을 의미한다. 바운딩 박스의 중심 좌표는 그리드 셀 내부에 μ_{t_x} 및 μ_{t_y} 로 표시된다. 각 평균과 분산 값은 0과 1 사이에서 처리되며, 수학 식 3에 주어진 시그모이드 함수를 적용한다.

수학식 3

$$\sigma(x) = \frac{1}{(1 + \exp(-x))}$$

[0074]

[0075] YOLOv3에서 바운딩 박스의 높이 및 너비 정보는 탐지된 바운딩 박스 및 지수 함수 이전에 t_w 와 t_h 를 사용하여 처리된다. 이는 도 5에서 μ_{t_w} 와 μ_{t_h} 가 YOLOv3의 t_w 와 t_h 를 나타내며, 둘 다 양수 값과 음수 값을 가질 수 있으므로 시그모이드 함수로 처리되지 않음을 의미한다. YOLOv3 검출 계층에서 얻은 바운딩 박스 좌표의 불확실성을 예측하기 위해 구조 변경 및 전체 계산 복잡성 증가 없이 가우시안 모델링이 적용된다.

[0076] 훈련 중에 YOLOv3는 클래스 및 객체 예측을 위해 이진 교차 엔트로피 손실 함수를 사용한다. 바운딩 박스의 경우 YOLOv3은 바운딩 박스에 대한 제곱 오차 손실의 조합을 적용한다. 바운딩 박스의 좌표 정보는 가우시안 모델링을 통해 가우시안 매개변수로 출력된다. 따라서 바운딩 박스의 손실 함수는 음수 로그 가능성(NLL: negative log-likelihood) 손실로 재정의된다. 여기서 클래스 및 객체 손실에 대한 함수는 변경되지 않는다. 바운딩 박스의 예측에 사용되는 손실 함수는 수학적으로 다음과 같이 주어진다.

수학식 4

$$L_x = - \sum_{i=1}^W \sum_{j=1}^H \sum_{k=1}^K \gamma_{ijk} \log(N(x_{ijk}^G | \mu_{t_x}(x_{ijk}), \sum_{t_x}(x_{ijk}))) + \varepsilon$$

[0077]

수학식 5

$$L_y = - \sum_{i=1}^W \sum_{j=1}^H \sum_{k=1}^K \gamma_{ijk} \log(N(y_{ijk}^G | \mu_{t_y}(y_{ijk}), \sum_{t_y}(y_{ijk}))) + \varepsilon$$

[0078]

수학식 6

$$L_w = - \sum_{i=1}^W \sum_{j=1}^H \sum_{k=1}^K \gamma_{ijk} \log(N(w_{ijk}^G | \mu_{t_w}(w_{ijk}), \sum_{t_w}(w_{ijk}))) + \varepsilon$$

[0079]

수학식 7

$$L_h = - \sum_{i=1}^W \sum_{j=1}^H \sum_{k=1}^K \gamma_{ijk} \log(N(h_{ijk}^G | \mu_{t_h}(h_{ijk}), \sum_{t_h} (h_{ijk}))) + \epsilon)$$

[0080]

[0081]

수학식 4 내지 7에서, NLL 손실은 L_x, L_y, L_w, L_h 로 표시된다. 각 높이와 너비에 대한 격자 셀의 수는 H와 W로 표시되고 K는 앵커의 수를 나타낸다. 또한, $\sum_{t_x} (x_{ijk})$ 는 탐지 계층의 출력에서 tx 좌표의 불확실성을 나타내고, tx의 실측 정답(ground truth) 좌표는 x_{ijk}^G 로 나타낸다(자세한 내용은 Choi et al.(2019) 참조).

[0082]

가우시안(Gaussian) YOLOv3는 또한 이미지의 각 탐지에 대한 바운딩 박스의 불확실성을 얻는다. 객체 탐지 동안 YOLOv3는 클래스 점수 및 객체 점수만 추정하고 바운딩 박스 좌표 점수 정보를 알 수 없기 때문에 바운딩 박스 점수를 추정하지 않는다. 그럼에도 불구하고 가우시안(Gaussian) YOLOv3의 출력에는 바운딩 박스의 점수라고도 하는 지역화 불확실성이 포함되어 있다. 따라서 탐지 과정에서 클래스 및 객체 점수와 함께 지역화 불확실성도 추정된다. 최종적으로 클래스, 객체 및 바운딩 박스 점수에 대한 신뢰도가 높은 예측이 결정된다. 결과적으로 가우시안(Gaussian) YOLOv3는 거짓 양성을 줄이고 참 양성을 개선하여 탐지 정확도를 높인다. 지역화 불확실성을 분석하여 탐지 기준은 다음과 같이 제공된다.

수학식 8

$$Cr = \sigma(\text{Object}) \times \sigma(\text{Class}) \times (1 - \text{Uncertainty}_{avg})$$

[0083]

[0084]

위의 수학식에서 Cr은 가우시안(Gaussian) YOLOv3에 대한 탐지 기준을 반영하며 i번째 클래스에 대한 객체 점수 및 클래스 점수는 각각 $\sigma(\text{Object})$ 및 $\sigma(\text{Class})$ 로 표시된다. 지역화 불확실성(Uncertainty_{avg})은 예측된 바운딩 박스 좌표의 평균 불확실성을 나타낸다.

[0086]

실험 결과

[0087]

데이터 세트에 대한 자세한 설명과 다양한 실험 결과가 여기에서 설명된다. 실험은 Intel Core i7-875, 64GB RAM, 2.2GHz CPU, 8GB 디스플레이 카드 NVIDIA GEFORCE GTX 및 64비트 Windows 10 운영 체제와 함께 Python 프로그래밍 언어를 사용하여 실행된다.

[0088]

전체 테스트 결과는 하기에 설명된다. 먼저 사전 훈련된 알고리즘의 결과 성능이 설명된다. 둘째, 채널 어텐션과 특징이 얹히고 전이 학습을 사용한 후의 탐지 알고리즘 결과가 설명된다. 더 나은 비교를 위해 동일한 데이터 세트 이미지를 사용하여 알고리즘을 테스트한다.

[0089]

알고리즘 성능 평가는 다양한 정량적 측정을 사용하여 수행된다. 또한 다른 심층 신경망 기반 알고리즘과도 비교한다.

[0091]

데이터 세트 설명

[0092]

알고리즘의 훈련 및 테스트를 위해 두 가지 다른 오버헤드 뷰 사람 데이터 세트가 활용된다. 두 데이터 세트는 동일한 환경에서 캡처되지만 다른 카메라 렌즈를 사용한다. 첫 번째 데이터 세트에서는 일반 렌즈 카메라를 사용하여 사람의 이미지를 캡처한 반면, 두 번째 데이터 세트에서는, 이미지가 광각 렌즈로 캡처된다.

[0093]

이미지는 라이트닝 조건이 다양한 통제되지 않은 상황에서 오버헤드 카메라 관점에서 캡처된다. 사람들은 제한 없이 현장에서 자유롭게 움직인다. 따라서 크기, 규모, 방향, 포즈 및 카메라 렌즈 측면에서 사람의 시각적 특

징에 차이가 있다. 이미지는 5미터 높이에 위치한 카메라를 사용하여 640×480 픽셀의 해상도로 캡처되었다. 총 2000개의 샘플 이미지를 사용했으며 1000개는 학습에, 1000개는 테스트에 사용되었다.

[0095] 사전 훈련된 알고리즘의 테스트 결과

[0096] 사전 훈련된 알고리즘(MS-COCO 데이터 세트로 훈련됨)을 사용하여 오버헤드 뷰 사람 탐지에 대한 테스트 결과는 도 6 및 도 7에 나와 있다. 결과는 데이터 세트의 다른 이미지를 사용하여 평가된다. 현장의 사람들은 자유롭게 걷고 있다. 두 오버헤드 관점에서 사람의 시각적 특징이 정면 뷰의 경우와 동일하지 않다는 것을 이미지에서 볼 수 있다. 인체의 시각적 특징은 머리 위에서부터 도 6 및 도 7과 같이 다양한 위치에서 크게 변화하고 있다. 알고리즘은 다양한 클래스 객체에 대해 사전 훈련되기 때문에, 대부분의 경우 인체는 다른 물체로 탐지된다.

[0097] 예를 들어, 도 6a에서 딥러닝 알고리즘을 위해 4명의 사람은 시각적인 외모가 사람처럼 보이기 때문에 정확히 사람 등급으로 분류되고, 1명은 개로 분류되고 수동으로 적십자로 표시된 바와 같이 3명은 알고리즘에 의해 탐지되지 않는다. 도 6b에서 한 사람만 탐지되고 다른 사람들은 적십자 표시와 같이 탐지되지 않는다. 도 6c에서 장면에 있는 사람의 수는 증가하고 있지만 사전 훈련된 알고리즘은 두 사람만 탐지할 수 있다. 하나는 새(bird)로 분류되고 다른 사람들은 사전 훈련된 알고리즘에 의해 탐지되지 않는다. 유사하게, 도 6d에서 이미지 하단에서 사람의 발만 사람 등급으로 분류되고 신체의 다른 부분은 오토바이로 분류된다. 아마도 그의 어두운 색 옷 때문일 것이다. 샘플 이미지에서 1명은 소로 분류되고 3명은 탐지되지 않는다. 그 결과, 도 6a에서는 샘플 이미지의 상단에 흰색 옷을 입은 사람이 검출된 반면, 나머지 3개의 샘플 이미지에서는 동일인이 검출되지 않았음을 알 수 있다.

[0098] 또한 사람들의 이미지가 광각 렌즈 카메라를 사용하여 캡처되는 또 다른 오버헤드 뷰 사람 데이터 세트에 대해 사전 훈련된 알고리즘이 테스트되었다. 그 결과를 도 7에 나타내었다. 두 가지 오버헤드 뷰 모두에서 사람의 외모와 포즈가 변경되었음을 알 수 있다. 사전 훈련된 알고리즘은 훈련 없이도 좋은 결과를 제공한다. 시각화 결과에서 볼 수 있듯이 대부분의 경우 사람이 해당 클래스로 올바르게 분류된다.

[0099] 도 7a에서 가우시안 YOLOv3 알고리즘에 의해 장면에 진입한 4명의 사람이 올바르게 탐지되었다. 한 경우, 사람의 발은 스케이트 보드로 분류되지만 적십자 표시가 있는 두 사람은 탐지되지 않았다. 도 7b에서는 이미지에서 5명의 사람이 알고리즘에 의해 올바르게 분류되었지만 서로 반대 방향으로 움직이는 2명은 탐지되지 않았다. 도 7c에서 카메라 바로 아래에 있는 사람은 딥러닝 알고리즘에 의해 올바르게 분류되지만 동시에 거의 같은 위치에 있는 한 명의 여성은 모델에 의해 탐지되지 않는다. 현장을 떠나는 세 사람도 적십자로 표시된 바와 같이 알고리즘에 의해 탐지되지 않는다. 도 7d에서 알고리즘은 단일 탐지 결과만 제공하고 다른 사람은 탐지하지 않음을 관찰했다. 한 사람은 새(bird)로 분류되고 한 사람은 사람으로 탐지된다.

[0101] 개선된 알고리즘의 탐지 결과

[0102] 알고리즘은 탐지 결과를 개선하기 위해 오버헤드 뷰 사람 데이터 세트에 대해 훈련된다. 훈련 후, 알고리즘의 탐지 정확도는 향상된다. 전체 실험은 Darknet 프레임워크를 기반으로 한다. Gaussian_yolov3_BDD.weights는 초기화 매개변수로 사용된다. 또한 MS-COCO 가중치 파일도 사용했다. 배치 크기는 512로 설정되고 훈련 중에 Epoch 수 = 100이 사용되며 SGD(stochastic gradient descent)가 최적화기로 사용된다. 네트워크 계층은 45,000번의 반복을 위해 훈련되었다. 훈련 정확도와 손실은 19번째 에포크(end)에 달성된다. 알고리즘의 정확도와 손실은 각각 도 8 및 도 9에 도시된다. 도 8에서 손실 수렴이 쉽게 나타나고 오차 값은 20번째 에포크 이후에 0.2 또는 0.3에 가깝다. 유사하게, 도 9에서 훈련 정확도는 알고리즘의 훈련 정확도가 30번째 에포크 이후에 거의 95%임을 보여준다.

[0103] 훈련 후 알고리즘은 이전에 설명한 것과 동일한 테스트 이미지에 대해 다시 테스트된다. 실험 결과는 도 10 및 도 11과 같이 추가 훈련 및 전이 학습 후에 탐지 결과가 크게 향상되었음을 보여준다. 탐지 알고리즘은 오버헤드 장면의 다른 위치에 있는 사람을 탐지한다. 도 10의 샘플 이미지에서, 상이한 시각적 특징을 갖는 상이한 위치에 있는 사람들이 정확하게 검출된다. 도 10a에서 볼 수 있듯이, 도 6a에서 사전 훈련된 알고리즘으로 탐지되지 않은 모든 사람이 이제 올바르게 탐지됨을 알 수 있다.

[0104] 도 6b에서 이전에 탐지되지 않아 적십자로 표시되었던 이미지 중앙에 있는 세 사람이 도 10b에서 딥러닝 알고리즘에 의해 효과적으로 탐지되었다. 도 10c에서, 훈련된 알고리즘은 도 10a 및 도 10b의 사람 수에 비해 많은

사람에 대해 테스트되었으며 알고리즘은 효율적으로 여러 사람을 탐지한다. 도 6d와 도 10d의 결과를 비교하면, 훈련된 모델이 사람을 정확하게 탐지하는 것을 알 수 있다.

[0105] 또한 도 11과 같이 다른 오버헤드 뷰 사람 데이터 세트 이미지에 대해 가우시안 YOLOv3 훈련된 모델을 테스트하였다. 도 7 및 도 11을 참조하면, 전이 학습 후 결과가 크게 개선되었음을 알 수 있다. 예를 들어, 도 7a와 도 11a를 비교하면, 이전에는 중앙에서 사람이 탐지되지 않은 것으로 적십자로 표시되었던 것처럼 여러 오버헤드 위치에 있는 사람들이 올바르게 탐지되었음을 알 수 있다. 유사하게, 도 11b에서 흰색 옷을 입은 두 사람이 반대 방향으로 걷는 것을 탐지했지만, 사전 훈련된 모델은 탐지하지 못했다. 도 11c에서, 딥러닝 알고리즘에 의해 이미지 중앙에 있는 여러 사람이 인식된다. 대조적으로, 동일한 결과가 도 11d의 경우 얻어지는데, 도 11d의 경우 훈련된 알고리즘이 오버헤드 카메라 바로 아래에 있는 다른 사람들을 정확하게 탐지한다. 도 11d에서 같은 위치에 있는 사람들은 사전 훈련된 알고리즘에 의해 탐지되지 않는다. 또한 탐지되지 않거나 다른 물체로 분류된 사람도 정확하게 탐지된다.

[0107] 평가 결과

[0108] tp, fp, fn, tn과 같은 다양한 정량적 매개변수를 사용하여 탐지 알고리즘의 성능이 평가된다. tp(참양성)는 사람 또는 사람 클래스로 정확하게 탐지된 바운딩 박스를 정의한다. fp(거짓양성)는 잘못 분류된 바운딩 박스를 사람 클래스로 결정하고, tn(참음성)은 배경으로 정확하게 분류된 바운딩 박스를 나타낸다. 그리고 fn(거짓음성)은 다른 객체로 분류되어 부적절하게 인식된 바운딩 박스를 나타낸다. 정밀도(Precision), 재현율(Recall), F1-스코어(F1-score) 및 정확도(Accuracy)가 알고리즘에 대해 계산된다. 도 12에서 오버헤드 뷰 데이터 세트를 사용하여 사전 훈련 및 개선된 알고리즘의 결과를 조사할 수 있다. 이러한 값은 수학적으로 다음과 같이 주어진다.

수학식 9

$$\text{Precision} = \frac{tp}{tp + fp} = \frac{tp}{\text{all detections}}$$

[0109]

수학식 10

$$\text{Recall} = \frac{tp}{tp + fn} = \frac{tp}{\text{all ground truths}}$$

[0110]

수학식 11

$$F1 - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

[0111]

수학식 12

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn}$$

[0112]

[0113] 표준 오차 방법은 정밀도, 재현율, F1 점수 및 정확도를 계산하는 데 사용된다. 도 12에서, 오버헤드 뷰 데이터

세트를 활용하여 사전 훈련되고 훈련된 알고리즘의 결과를 조사할 수 있다. 결과로부터 학습과 전이학습을 적용한 후에 탐지 모델의 정확도가 높아짐을 알 수 있다. 도 12에서 평균값을 보여주고 있으며 본 발명의 일 실시예에 의한 감시를 위한 심층 신경망 기반 사람 탐지 시스템의 전체 탐지 정확도가 94%까지 향상되었음을 알 수 있다.

또한 알고리즘의 성능을 조사하기 위해 평균 정밀도인 mAP 값이 계산되었다. mAP 값은 정밀도 및 재현율 값을 사용하여 탐지 모델의 정확도를 추정하는 데 사용된다. 재현율에서 정밀도를 사용하면, N 클래스에 대한 mAP 평균 정밀도가 다음과 같이 주어진다.

수학식 13

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i$$

표 1에서는 탐지 모델과 다른 딥 러닝 알고리즘을 비교한 결과를 보여준다. 본 모델이 다른 탐지 알고리즘에 비해 더 나은 성능을 보이는 것을 알 수 있다. 본 발명의 일 실시예에 의한 감시를 위한 심층 신경망 기반 사람 탐지 시스템의 mAP 값은 0.90으로 다른 물체 탐지 모델보다 높다.

표 1

Results of different deep learning algorithms used for overhead view human images.

S. no.	Detection algorithm	mAP
1	Faster-RCNN (Ren, He, Girshick, & Sun, 2015)	0.83
2	Mask-RCNN (He, Gkioxari, Dollár, & Girshick, 2017)	0.83
3	YOLOv3 (Redmon et al., 2016)	0.86
4	SSD (Liu et al., 2016)	0.86
5	Gaussian YOLOv3 (Choi et al., 2019) (pre-trained)	0.87
6	Gaussian YOLOv3 (Choi et al., 2019) (improved-trained)	0.90

본 발명은 스마트 도시와 사회에서 사람 탐지를 위한 심층 신경망의 적용을 제시했다. 채널 어텐션 및 특징 얹힘 모듈이 포함된 다양한 오버헤드 뷰에서 사람 탐지를 위해 가우시안 YOLOv3 알고리즘을 적용했다. 이를 통해 특징 맵은 각 채널의 가중치를 학습하고 이러한 채널을 엮어 더 대표적인 특징을 얻을 수 있으므로 탐지가 향상된다. 알고리즘의 가중치는 사전 훈련된 알고리즘을 사용하여 초기화되었다. 알고리즘은 탐지 결과를 향상시키기 위해 오버헤드 뷰 사람 데이터 세트에 대해 추가로 재훈련된다. 또한 알고리즘의 거짓양성 비율을 더욱 줄이고 오버헤드 뷰에서 사람 탐지의 정확도를 높이기 위해 전이 학습이 채택되었다. 알고리즘의 주요 이점은 바운딩 박스의 좌표를 가우시안 매개변수로 모델링하여 바운딩 박스의 불확실성을 결정하므로 정확도가 향상되고 거짓양성이 감소한다는 것이다. 실험 결과는 채널 어텐션, 특징 얹힘 모듈 및 추가적인 훈련을 결합하면 오버헤드 이미지에 대한 mAP가 0.90인 본 발명의 일 실시예에 의한 감시를 위한 심층 신경망 기반 사람 탐지 시스템의 전체 정확도가 향상됨을 보여준다.

이상 본 발명을 구체적인 실시예를 통하여 상세하게 설명하였으나, 이는 본 발명을 구체적으로 설명하기 위한 것으로, 본 발명은 이에 한정되지 않으며, 본 발명의 기술적 사상 내에서 당 분야의 통상의 지식을 가진 자에 의해 그 변형이나 개량이 가능함은 명백하다고 할 것이다.

본 발명의 단순한 변형 내지 변경은 모두 본 발명의 영역에 속하는 것으로, 본 발명의 구체적인 보호 범위는 첨부된 청구범위에 의하여 명확해질 것이다.

부호의 설명

200 : 오버헤드 뷰 카메라

201 : 변형된 가우시안(Gaussian) YOLOv3 모델

202 : 가우시안(Gaussian) YOLOv3 모델

204 : 채널 어텐션(channel attention) 및 특징 얽힘(feature intertwine) 모듈

206 : 오리지널 가우시안 YOLOv3 모델

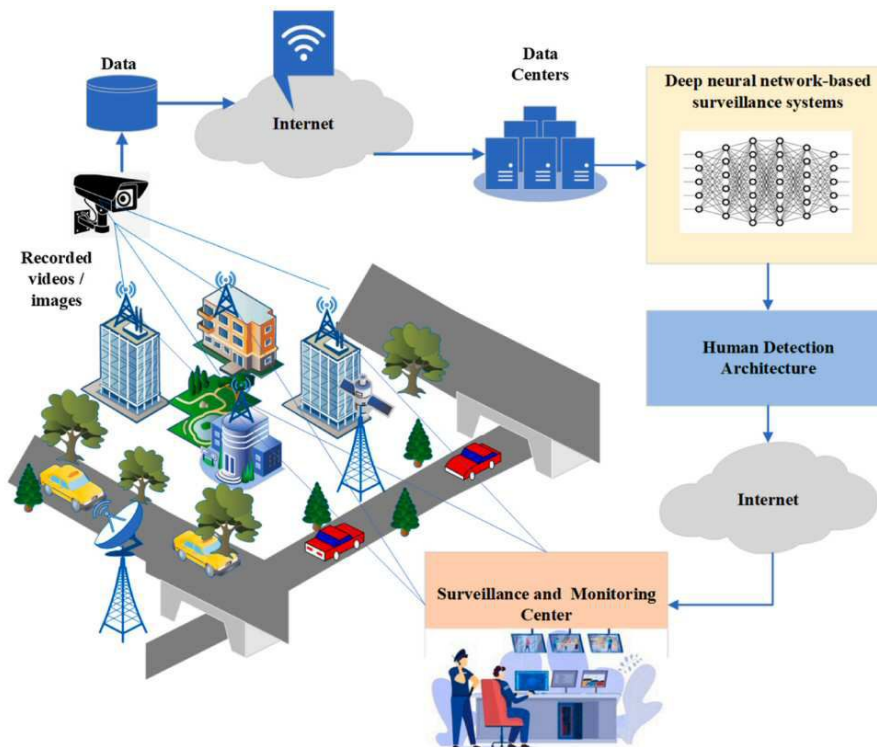
208 : 전이 학습 기반 모델 훈련 모듈

210 : MS-COCO 가중치 파일

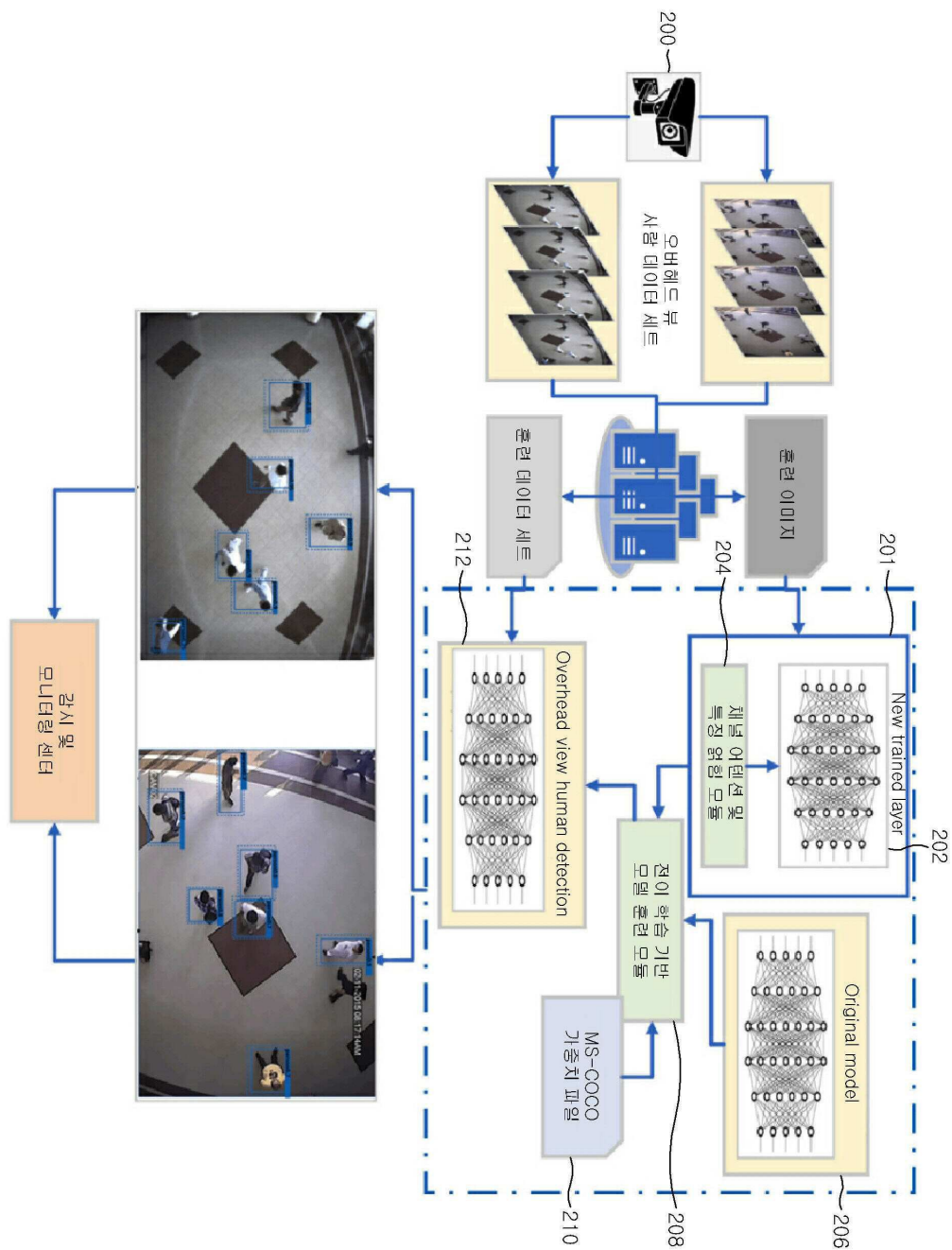
212 : 오버 헤드 뷰 사람 탐지 모델

도면

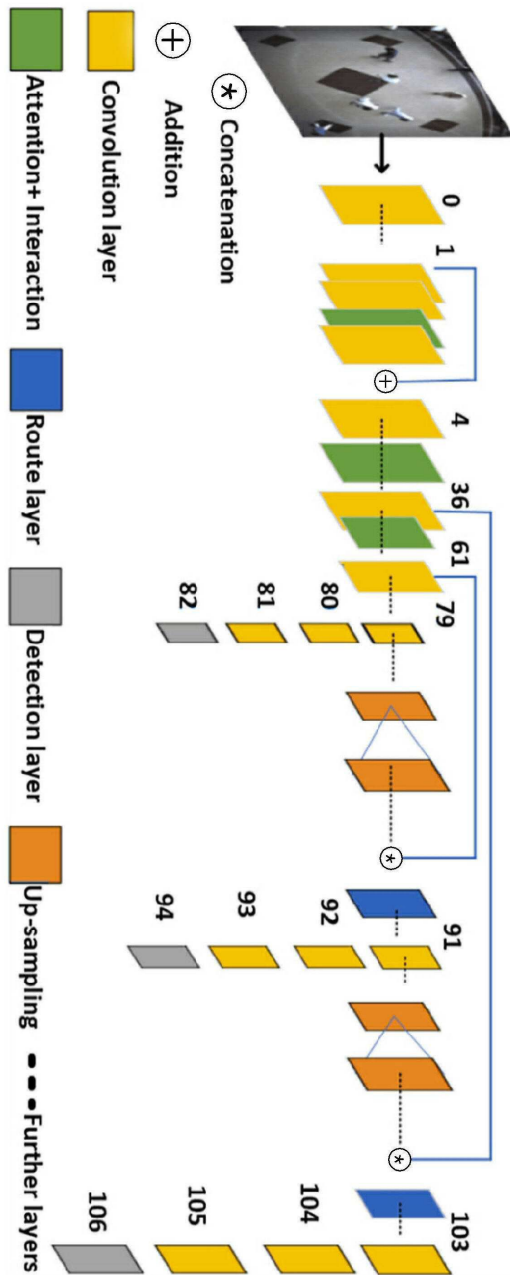
도면1



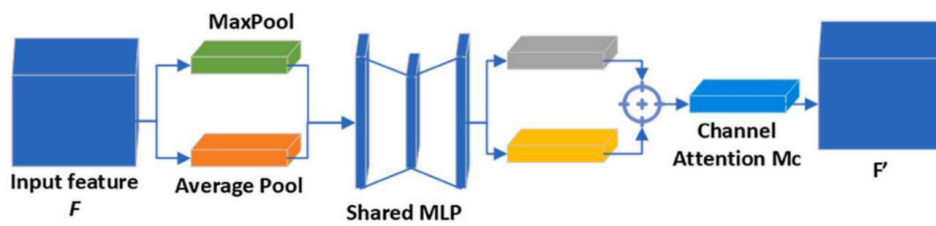
도면2



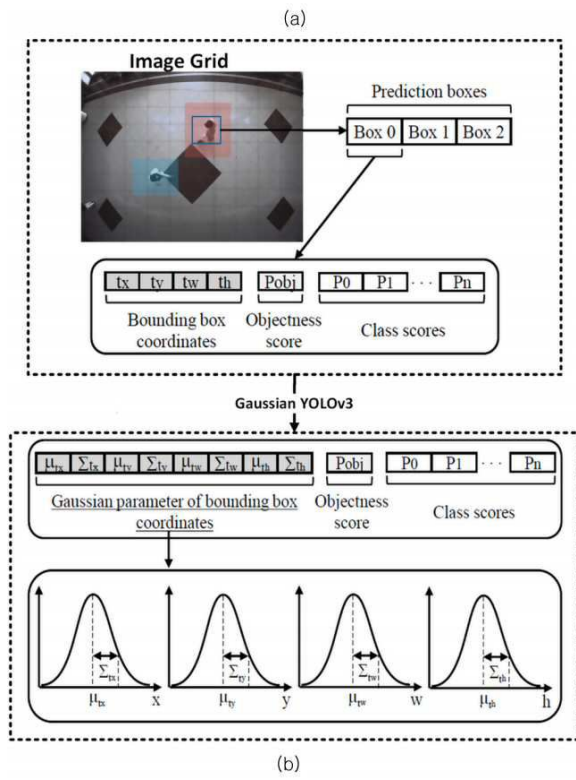
도면3



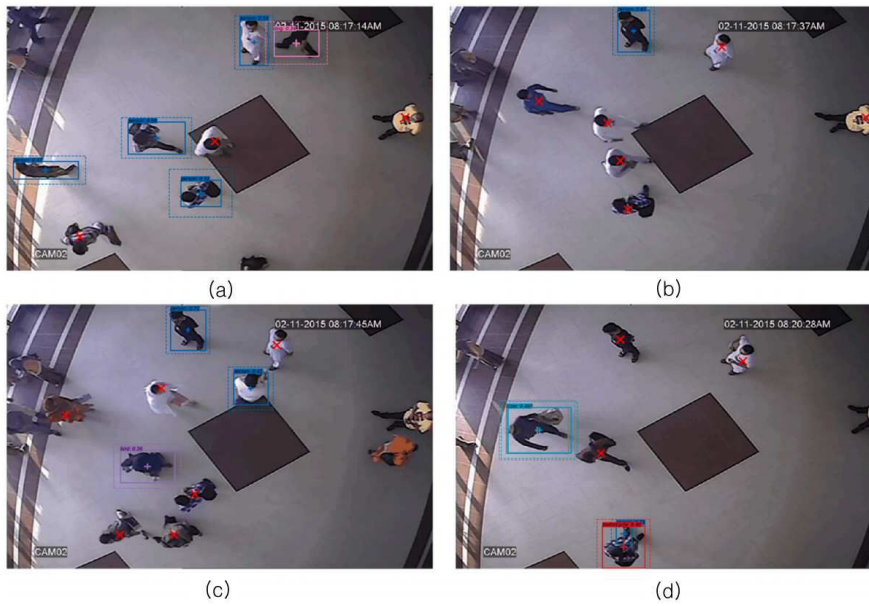
도면4



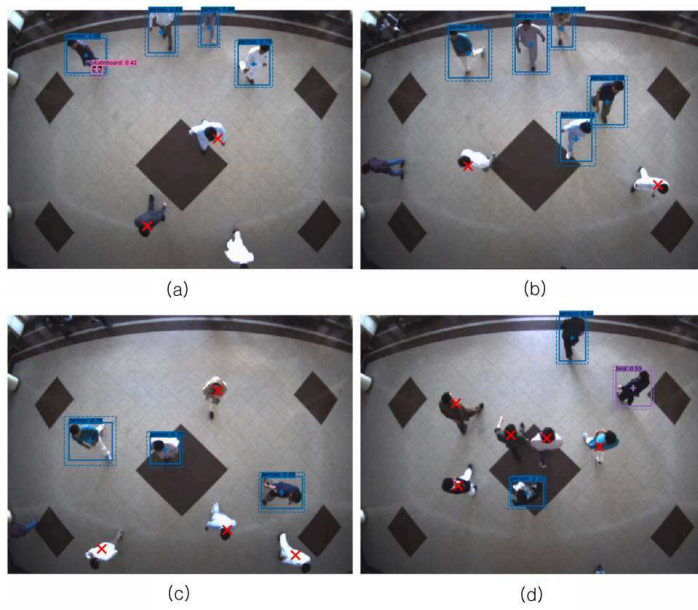
도면5



도면6



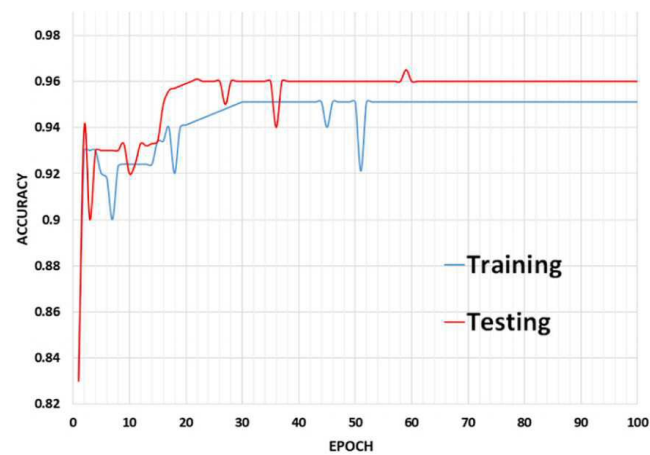
도면7



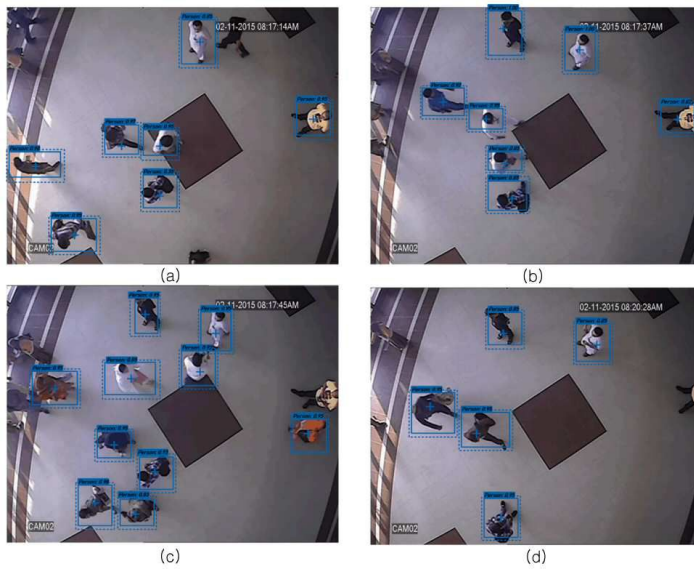
도면8



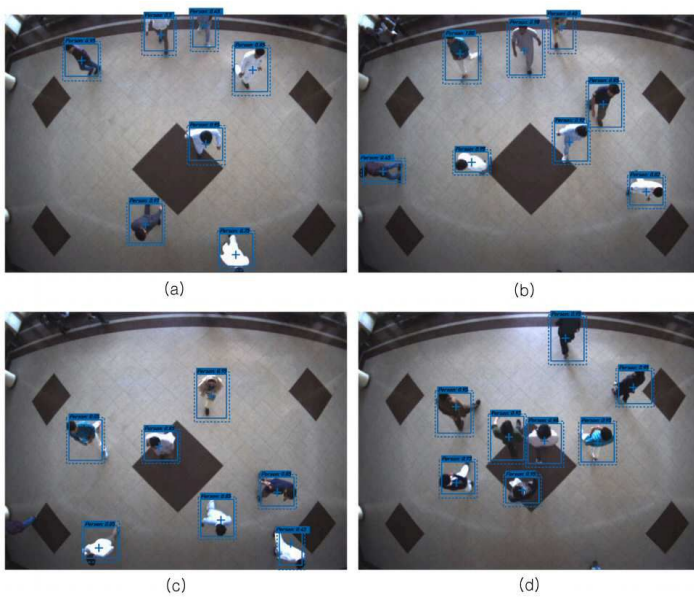
도면9



도면10



도면11



도면12

