

<https://doi.org/10.7236/JIIBC.2022.22.2.63>
JIIBC 2022-2-9

Hellinger 거리 IoU와 Objectron 적용을 기반으로 하는 객체 감지

Object Detection Based on Hellinger Distance IoU and Objectron Application

김용길*, 문경일**

Yong-Gil Kim*, Kyung-Il Moon**

요약 2D 객체 감지 시스템은 최근 몇 년 동안 심층 신경망과 대규모 이미지 데이터셋의 사용으로 크게 개선되었지만, 아직도 범주 내에서 데이터 부족, 다양한 외관 및 객체 형상 때문에 자율 탐색 등과 같은 로봇 공학과 관련된 응용에서 2D 물체 감지 시스템은 적절하지 않다. 최근에 소개되고 있는 구글 Objectron 또한 증강 현실 세션 데이터를 사용하는 새로운 데이터 파이프라인이라는 점에서 도약이라 할 수 있지만, 3D 공간에서 2D 객체 이해라는 측면에서 마찬가지로 한계가 있다. 이에 본 연구에서는 더 성숙한 2D 물체 감지 방법을 Objectron에 도입하는 3D 물체 감지 시스템을 나타낸다. 대부분의 객체 감지 방법은 경계 상자를 사용하여 객체 모양과 위치를 인코딩한다. 본 작업에서는 가우스 분포를 사용하여 객체 영역의 확률적 표현을 탐색하는데, 일종의 확률적 IoU라 할 수 있는 Hellinger 거리를 기반으로 하는 가우스 분포에 대한 유사성 측도를 제시한다. 이러한 2D 표현은 모든 객체 감지기에 원활하게 통합할 수 있으며, 실험 결과 데이터 집합에서 주석이 달린 분할 영역에 더 가까워서 Objectron의 단점이라 할 수 있는 3D 감지 정확도를 높일 수 있다.

Abstract Although 2D Object detection has been largely improved in the past years with the advance of deep learning methods and the use of large labeled image datasets, 3D object detection from 2D imagery is a challenging problem in a variety of applications such as robotics, due to the lack of data and diversity of appearances and shapes of objects within a category. Google has just announced the launch of Objectron that has a novel data pipeline using mobile augmented reality session data. However, it also is corresponding to 2D-driven 3D object detection technique. This study explores more mature 2D object detection method, and applies its 2D projection to Objectron 3D lifting system. Most object detection methods use bounding boxes to encode and represent the object shape and location. In this work, we explore a stochastic representation of object regions using Gaussian distributions. We also present a similarity measure for the Gaussian distributions based on the Hellinger Distance, which can be viewed as a stochastic Intersection-over-Union. Our experimental results show that the proposed Gaussian representations are closer to annotated segmentation masks in available datasets. Thus, less accuracy problem that is one of several limitations of Objectron can be relaxed.

Key Words : 3D Object Recognition, CNN, Gaussian, Hellinger Distance, IoU, Objectron

*정회원, 조선이공대학교 컴퓨터보안과

**정회원, 호남대학교 컴퓨터공학과(교신저자)

접수일자 2022년 2월 10일, 수정완료 2022년 3월 10일

게재확정일자 2022년 4월 8일

Received: 10 February, 2022 / Revised: 10 March, 2022 /

Accepted: 8 April, 2022

*Corresponding Author: kimoon@honam.ac.kr

Department of computer engineering, Honam university, Korea.

I. 서 론

반적으로 장면에서 식별할 지정된 객체 범주의 인스턴스 수를 미리 알 수는 없다. 이 문제를 해결하기 위해 표준 CNN을 구축하면 다음과 같은 상황이 발생한다. 출력 크기가 가변적이라는 것인데, 그것은 찾을 객체의 인스턴스 수와 위치에 따라 다르며 고정 크기의 출력층이 있는 네트워크에서는 불가능하다. 이와 관련하여 R-CNN, YOLO 등의 모델이 사용될 수 있는데, R-CNN은 객체 감지 결과를 기반으로 분할을 수행한다. R-CNN은 초기에 선택적 검색을 사용하여 많은 후보를 추출한 다음 CNN을 통해 그것의 특징들을 계산한다. 마지막으로 특정 선형 분류기(일반적으로 SVM)를 사용하여 각 영역을 분류한다. R-CNN은 객체 감지 및 이미지의 거친 분할과 같은 더 복잡한 작업을 수행할 수 있지만, 두 단계로 진행되는 느린 훈련과 실행 속도가 매우 느리다는 데에 있다. Fast-R-CNN은 객체 감지를 위한 빠른 영역 기반 컨볼루션 네트워크이다. 신경망에 후보 이미지를 제공하는 대신 전체 이미지를 네트워크에 제공하여 CFM(Convolutional Feature Map)을 생성한다. CFM은 RoI 풀링 층을 통해 고정 크기 정사각형 이미지로 다시 형식화된 후보 영역을 검색하는 데 사용된다. 특히, Faster-R-CNN은 관심 영역에 대한 선택적 검색을 제거하여 네트워크를 더욱 빠르게 만드는 솔루션이다. Fast-R-CNN과 유사하게 이미지는 CFM을 생성하기 위한 입력으로 직접 사용된다. 그렇지만, 이 모델에서는 현재 맵에서 선택적 검색을 수행하는 대신에 두 번째 별도의 신경망을 사용하여 후보 영역을 예측한다. 이러한 R-CNN 방식의 단점은 미시적인 접근 방식으로 객체를 식별하기 위해 이미지의 하위 영역에 초점을 두고 있다는 점이다. SSD(Single Shot Detection) CNN은 이미지를 전체적으로 고려하는 접근 방식에 해당한다. 특히, IoU(Intersection Over Union) 개념을 사용하는 YOLO에서 단일 컨볼루션 네트워크는 경계 상자과 감지된 각 객체의 클래스에 속할 확률을 모두 예측한다. YOLO는 아주 빠르다는 장점이 있다. LiDAR(Light Detection and Ranging)와 같은 3D 캡처 센서에 의존하는 자율 주행과 관련하여 물체의 실시간 감지는 매우 중요한 사항이다. 이와 관련하여 모바일 증강 현실 세션 데이터를 사용하는 새로운 데이터 파이프라인이 적용되고 있는데, ARCore 및 ARKit의 개발과 함께, 현재 스마트 폰은 AR 기능들과 카메라 포즈, 희소 3D 포인트 클라우드, 조명 추정 및 표면 평면을 포함하여 AR 세션 동

안 추가 정보를 획득할 수 있는 능력을 지니고 있다. 이러한 맥락에서 Google Objectron은 3D 물체 감지를 위한 거대한 도약이라 할 수 있는데, Objectron은 스마트 기기가 물체의 크기와 방향을 인식할 수 있도록 하는 3D 물체의 실시간 감지를 위한 모바일 기술에 해당한다. 그렇지만, Objectron의 단점으로 3D 공간에서 2D 객체를 이해하는 것은 2D 작업에 비해 실제 데이터 세트가 많지 않기 때문에 여전히 어려운 작업이다.

본 연구와 관련하여 위에서 언급한 YOLO의 경계 상자(BB)와 IoU 개념은 매우 우수한 성능을 보이지만 가장 큰 결함은 이미지의 작은 물체에 대해서는 부정확하다는 것이다. 경계 상자와 관련해서는 기본적으로 HBB(Horizontal Bounding Box) 방식이 사용되고, HBB 방식의 보완과 관련하여 OBB(Oriented Bounding Box)가 HBB와 분할 영역 사이의 대안으로 사용된다. 본 연구에서는 OBB와 관련하여 가우스 분포 모델이 참조되며, 주요 목표는 OBB의 방향 매개변수를 처리하고 Gaussian Wasserstein 거리와 같은 비선형 대응이 지역적 손실 함수로 사용될 수 있도록 IoU 관련 해석학적인 관계를 나타낸다. 이와 관련해서는 두 분포 사이의 확률적 IoU에 해당하는 유사성 함수로 Hellinger 거리를 사용하며, 특정 가우스의 경우에 단순하고 미분 가능한 손실 함수의 정의를 사용한다. 여기에서 가우스 BB를 얻기 위한 기본 표현은 HBB 또는 OBB로 가정하지 않으므로 일반 분할 마스크에도 적용할 수 있다. OBB를 취급하는 검출기 적용과 관련하여 매개변수를 사용한 가우스 BB 표현(예를 들어, 중심, 차원 및 방향의 회귀) 및 가우스 BB의 직접적인 회귀 적용을 위한 평균 벡터와 공분산 행렬 또한 나타낸다. 본 연구에서 제안된 객체 범위 감지 및 인스턴스 분할 방식은 구글 Objectron에서 수행된다.

II. 관련 연구

1. 객체 표현 및 지역적 손실

객체 표현과 관련하여 HBB(Horizontal Bounding Box) 방식은 가장 인기가 있는 표현 방식으로 큰 데이터 집합을 취급할 때 4개의 매개 변수들과 주석을 사용한다. 그렇지만, 물체들은 완전 바이너리 마스크로 표현되기 때문에 상당히 시간 소모적인 주석 작업을 요구한다. 더 나아가서 네트워크들은 의미론적 또는 입체적인 분할 작업 측면에서 바이너리 맵으로 회귀시키는 작업을 요구하기 때문에 결과적으로 더 복잡한 아키텍처를 가지게 된

다^[1]. HBB의 잘 알려진 문제점은 가늘고 긴 객체 또는 회전된 객체들에 대해 배경의 상당한 부분들을 포함할 수 있다는 것이다. 이러한 이유로 OBB(Oriented Bounding Box)가 HBB와 분할 마스크 사이의 대안으로 사용되었다. 이에 주요 연구로 수평 RoI를 OBB RoI에 적용하기 위해 회전된 RoI 변환기 사용하거나, 회전 객체들을 사용한 새로운 데이터 집합을 통해 객체 탐지를 위한 새로운 IoU 기반 손실 함수로 OBB를 탐색하는 방식을 들 수 있다^[2]. OBB에는 회귀할 추가 매개변수인 각도가 필요하고 더 많은 수의 앵커가 필요할 수 있다(앵커 기반 검출기가 사용되는 경우). 또한 각도 매개변수의 회귀는 OBB 회전이 정의되는 방식에 따라 까다로울 수 있고, IoU를 기반으로 하는 손실 함수는 미분 불가능하여 사용할 수 없다. 그러한 측면에서 Yang 등에 의한 IoU 관련 미분 가능한 손실 함수를 정의는 상당한 의미가 있다^[3]. HBB를 기반으로 하는 대부분의 물체 감지기는 지역화 손실 문제와 관련하여 매개변수 별로 손실 구성요소의 조합을 적용한다(예를 들어, l_1 또는 l_2 -norm 사용). 이러한 norm의 사용은 객체의 크기에 매우 민감한 측면이 있어서 로그 공간에서 HBB 오프셋을 갖는 l_2 -norm을 사용하거나 Huber 손실로 불리는 부드러운 l_1 -norm이 사용될 수 있다^[4]. 매개변수 단위의 손실 함수는 하나의 변수가 다른 매개변수에 영향을 미치지 않는 독립성을 가정하므로 HBB의 가장자리들이 GT(ground-truth) 주석과 일관되지 수렴되지 않을 수 있다. 대안으로 일부 연구자들은 잘 알려진 속성으로 크기 변화의 불변성 및 HBB 매개변수 결합 회귀의 불변성을 가진 IoU 기반 손실 함수를 제안했다. 이와 관련하여 UnitBox는 BB 탐지를 개선하기 위해 IoU 손실을 통합한 것으로 단일 객체 분할에 사용되었다^[5]. 이외에 객체 감지를 위한 경계 IoU 손실 제안^[6] 및 회귀 영역과 GT 상자 영역이 겹치지 않을 때 IoU에 의해 생성되는 기울기가 사라지는 문제 완화를 위해 IoU 일반화에 해당하는 GIoU를 도입했다^[7]. GIoU의 더딘 수렴성 등의 몇 가지 문제점 해결과 관련하여 IoU에 대한 페널티로 중첩 영역, 중심 간에 거리 및 중첩비 제약을 통합하는 DIOU (Distance-IoU)와 CIOU(Complete-IoU)가 사용되었고^[8], l_p -norm과 IoU 기반 손실을 결합하여 각각의 개별 구성요소의 단점을 해결하는 방식이 제안되었으며^[9], 오버랩 영역, 중심점 및 측면 길이로 IoU에 대한 페널티 및 훈련에 대한 초점 손실을 통합하는 방식도 소개되었다^[10]. Chen 등은 OBB를 처리하기 위한 IoU의 확장으로 PIoU(Pixel-IoU)를 제안했는데, 이것은 일반적인 객

체 표현이 가능하다^[11].

2. Objectron 아키텍처

2D 객체 예측은 경계 상자를 제공하여 이미지에서 객체의 위치를 알 수 있지만, 2D 이미지에서 3D 물체 감지 방법으로 이러한 방법을 확장하는 것은 어렵다. 주된 이유는 2D 이미지에 비해 주석이 달린 3D 비디오의 실제 데이터셋이 없기 때문이다. ImageNet을 통해 2D 이미지 작업을 발전시킬 수 있었던 것처럼, Google AI 팀은 3D 물체 감지를 위한 Objectron 데이터셋을 출시했는데, 이 데이터셋은 연구 커뮤니티가 3D 물체 이해를 발전시킬 수 있도록 하는 것을 목표로 하고 있다. Objectron 데이터셋은 현재 15,000개의 주석이 달린 비디오가 포함된 짧은 객체 중심 비디오 클립 모음이다. 동일한 대상을 다른 각도에서 관찰하여 대상 중심의 다시점에 초점을 맞추고 있다. 각 비디오 클립에는 고해상도 이미지, 물체 포즈, 카메라 포즈, 포인트 클라우드 및 표면 평면을 포함한 AR 세션 메타데이터가 수반된다. 데이터셋에는 개체의 위치, 방향 및 치수를 설명하는 각 개체에 대해 수동으로 주석이 달린 3D 경계 상자가 포함되어 있다^[12].

최신 Objectron 버전은 2단계 아키텍처를 사용한다. 첫 번째 단계에서는 TensorFlow 객체 감지 모델을 활용하여 각 프레임에서 객체의 위치를 찾는다. 그런 다음 두 번째 단계에서는 절단된 2D 이미지를 사용하여 3D 경계 상자를 추정하는 동시에 다음 프레임에 대한 객체의 2D 자르기를 계산하므로 객체 감지가 모든 프레임을 실행할 필요가 없다. 그림 1은 2단계 파이프라인으로 이전 1단계 파이프라인보다 3배 더 빠르다. 3D 물체 감지 작업과 관련하여 3D 지향 상자에 대한 정확한 3D IoU 값을 계산하기 위한 알고리즘은 세 부분으로 구성된다. 먼저 Sutherland-Hodgman Polygon 절단 알고리즘을 사용하여 두 상자의 면 사이의 교차점을 계산한다. 교차의 볼륨은 잘린 모든 다각형의 볼륨 겹침에 의해 계산된다. 그런 다음 IoU는 교차의 볼륨과 두 상자의 합집합 볼륨에서 계산된다. 그림 2는 Polygon 절단 알고리즘에 의한 3D 교차 계산을 나타내는데, 좌측은 상자에 대해 다각형을 잘라서 각 면의 교차점을 계산한 것이고, 우측은 모든 교차점(녹색)의 볼륨 겹침을 계산하여 교차 볼륨을 계산하는 과정을 나타낸다. 일반적인 객체 감지 모델은 2D 경계 상자를 계산하지만, Objectron의 새 모델은 깊이 정보가 있는 경계 상자를 반환한다.

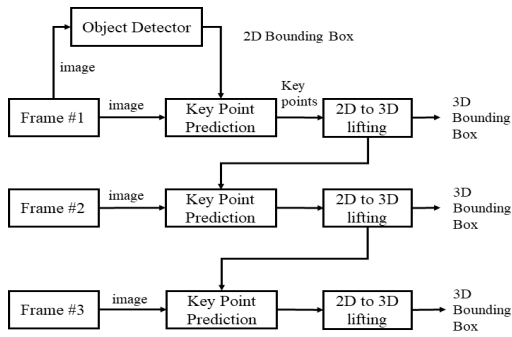


그림 1. 2단계 파이프라인
Fig. 1. two-stage pipeline

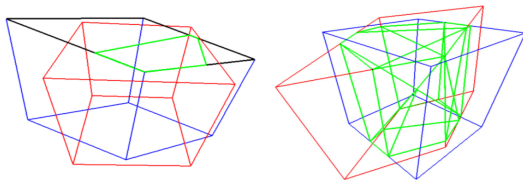


그림 2. polygon 절단 알고리즘에 의한 교차 계산
Fig. 2. Intersection calculation by polygon cutting algorithm

Objectron은 스마트 기기가 물체의 크기와 방향을 인식할 수 있도록 하는 3D 물체의 실시간 감지를 위한 모바일 기술로 AR 세션 데이터에 대한 새로운 주석 도구를 통해 실측 데이터에 라벨을 지정하여 주석자가 객체에 대한 3D 경계 상자에 빠르게 라벨을 지정할 수 있도록 했다. 기본 구조 수준에서 모델은 분류 및 시각적 감지를 위한 구글 신경망인 MobileNetV2를 기반으로 하는 인코더-디코더 아키텍처를 가진다. 이러한 아키텍처, 이미지 2D 투영 및 3D 추정 알고리즘을 사용하여 모델은 해당 물체의 3D 출력을 처리할 수 있다. 이미지의 가능한 변동을 완화하기 위해 2D 객체에 대해 동일한 감지 및 추적 체계에서 작업이 진행되는데, 이것은 이미지 분할 작업에 해당한다. 후속 객체 감지를 위해 경계 상자가 적용되며, 객체는 가우스 분포로 감지되고, 분할 마스크가 적용된다.

III. 가우스 BB 표현 및 Objectron 적용

이미지의 2D 투영과 관련하여 본 연구의 핵심은 2차원의 회전 가능한 가우스 분포를 사용하여 객체를 나타내는 것이다. 가우스 분포는 평균 벡터 $\mu = (x_0, y_0)^T$ 와 공분산 행렬 Σ 를 특징으로 하며 원하는 2D 영역의 표현

을 유도한다. 영역을 확률 분포로 나타내면 두 분포 간의 거리 또는 유사성을 비교하기 위한 다양한 통계 도구를 사용할 수 있다. 가우스 분포를 사용하는 Bhattacharyya 거리 및 Kullback-Leiber divergence와 같은 측도는 미분 가능한 식으로서 딥 러닝 전략의 매력적인 옵션으로 사용되고 있다^[13]. 여기에서는 HBB 및 OBB의 가우스 BB 표현과 Hellinger 거리에 의한 확률적 IoU 개념을 나타낸다.

1. HBB 및 OBB의 가우스 BB 표현

가우스 분포의 2D 공분산 행렬 Σ 는 양정치 대칭행렬로 임의의 회전 행렬 $R(\theta)$ 에 관해 다음과 같이 표현될 수 있다.

$$\Sigma = R(\theta) \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} R(\theta)^T, \quad R(\theta) = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \quad (1)$$

$(\sigma_1, \sigma_2, \theta)$ 의 회귀를 갖는 네트워크 설계는 OBB에 대한 기존 물체 감지의 출력에 더 정렬되고 GBB에 더 쉽게 적용할 수 있다. 이러한 매개변수 표현에서 $\sigma_1, \sigma_2 > 0, \theta \in R$ 이고, 이는 공분산 행렬의 모호한 표현으로 이어지는데, 매개변수 $(\sigma_1, \sigma_2, \theta)$ 및 $(\sigma_1, \sigma_2, \theta + \pi/2)$ 는 같은 행렬을 생성한다. 이러한 문제를 완화하기 위해 $\theta \in [-\pi/4, \pi/4]$ 을 선택하여 $(\sigma_1, \sigma_2, \theta)$ 의 식으로 공분산 행렬 Σ 의 고유한 표현을 생성할 수 있다. $\sigma_1 = \sigma_2$ 일 때 등방성 가우스 분포가 되고, 이 경우 회전 R 은 관련이 없게 된다. 반면에 정사각형 모양의 OBB는 여전히 연관된 방향을 나타내기 때문에 정사각형 모양의 OBB에서 가우스 BB로의 대응은 가역적이지 않다.

객체 감지를 위한 대부분의 기존 데이터 집합은 주석을 HBB, OBB 또는 분할 마스크로 표시하기 때문에 이러한 표현에서 몇 가지 GBB를 얻을 수 있다. 객체 영역이 연속적인 2차원 이진 영역으로 표시되고, Ω 가 균일한 확률 밀도 함수를 나타낸다고 가정하면 분포의 평균과 공분산 행렬은 다음과 같다.

$$\mu = \frac{1}{N} \int_{x \in \Omega} x dx, \quad \Sigma = \frac{1}{N} \int_{x \in \Omega} (x - \mu)(x - \mu)^T dx \quad (2)$$

여기서 N 은 Ω 의 면적이다. HBB에 관해 Ω 은 중심이 (x_0, y_0) 이고, 너비 W , 높이 H 를 갖는 직사각형 영역이다. 이 경우에 $\mu = (x_0, y_0)^T$ 은 단순히 직사각형 영역의 중심이고, 공분산 행렬은 다음과 같이 계산될 수 있다(Ω 의 각 점을 평균에서 빼는 방식으로).

$$\Sigma = \frac{1}{WH} \int_{-H/2}^{H/2} \int_{-W/2}^{W/2} \begin{pmatrix} x^2 & xy \\ xy & y^2 \end{pmatrix} dx dy = \frac{1}{12} \begin{pmatrix} W^2 & 0 \\ 0 & H^2 \end{pmatrix} \quad (3)$$

따라서 $\sigma_1^2 = W^2/12, \sigma_2^2 = H^2/12$ 가 된다. OBB가 있는 경우에는 위와 같이 해당 축 정렬 HBB의 측면을 기반으로 상관되지 않은 분산들을 정의하고 회전 각도 θ 를 계산할 수 있다.

$$\Sigma = \begin{pmatrix} \sigma_1^2 \cos^2(\theta) + \sigma_2^2 \sin^2(\theta) & (\sigma_1^2 - \sigma_2^2) \sin(\theta) \cos(\theta) \\ (\sigma_1^2 - \sigma_2^2) \sin(\theta) \cos(\theta) & \sigma_1^2 \sin^2(\theta) + \sigma_2^2 \cos^2(\theta) \end{pmatrix} \quad (4)$$

주석이 매개변수 표현의 2D 모양(예를 들어, 다각형 표현)으로 제공되는 경우에는 식(2)를 사용하여 분석적으로 가우스 매개변수를 계산할 수 있다. 이진 분할 마스크를 사용하는 일반 표현의 경우 경험적 평균과 공분산 행렬을 추정하는데, 이것은 본질적으로 식(2)의 이산화된 버전에 해당한다. 영역을 결정론적 마스크로 시각화하고 두 지역 간의 기존 IoU를 계산하기 위해 GBB에서 제공하는 표현에서 이진 마스크를 생성하는 전략을 제공하는 것도 중요하다. 응용 프로그램에 HBB 또는 OBB 표현이 필요한 경우 식(3)의 역행렬을 사용하여 대각 공분산 행렬에서 상자 차원과 각도를 추출할 수 있다.

2. 확률적 IoU와 손실 함수

GBB로 표현되는 두 개의 객체(GT 주석 및 예측)가 주어지면 다음 단계는 이들 간의 유사성을 계산하는 방법을 정의해야 한다. 이와 관련하여 두 확률 분포 사이의 중첩(유사성) 또는 차이를 계산하기 위해 여러 통계 도구를 사용할 수 있다. 여기에서는 실제 거리 측정법을 얻는데 사용할 수 있는 Bhattacharyya 계수/거리에 초점을 맞춘다. 두 개의 2차원 확률밀도함수 $p(x)$ 와 $q(x)$ 사이의 Bhattacharyya 계수 B_C 는 다음과 같이 정의되며, 분포 간의 겹침 정도, 즉 이들 간에 유사도를 측정한다.

$$B_C(p, q) = \int_{R^2} \sqrt{p(x)q(x)} dx \quad (5)$$

두 분포가 같은 경우에만 $B_C(p, q) = 1$ 이다. 두 분포 p 와 q 사이의 Bhattacharyya 차이 B_D 는 $B_D(p, q) = -\ln B_C(p, q)$ 와 같이 주어진다. B_D 가 증가하면 B_C 가 감소하고 그 반대도 마찬가지이다. p 가 평균 μ_1 , 공분산 행렬이 Σ_1 을 갖는 가우스 분포이고, q 가 평균 μ_2 , 공분산 행렬이 Σ_2 을 갖는 가우스 분포를 갖는 경우에, B_D 는 다음과 같이 나타낼 수 있다.

$$\mu_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \quad \Sigma_1 = \begin{pmatrix} a_1 & c_1 \\ c_1 & b_1 \end{pmatrix}, \quad \mu_2 = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} a_2 & c_2 \\ c_2 & b_2 \end{pmatrix},$$

$$\Sigma = (\Sigma_1 + \Sigma_2)/2$$

$$B_D = \frac{1}{8}(\mu_1 - \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2) + \frac{1}{2} \ln(|\Sigma| / \sqrt{|\Sigma_1| |\Sigma_2|}) \quad (6)$$

B_D 의 우측 식의 첫 번째 항을 B_1 이라 하고, 두 번째 항을 B_2 라 할 때, B_2 는 형상 매개변수인 Σ_1 과 Σ_2 만 포함하고, 평균들에 종속되지 않으므로 중심점 접촉 또는 형상 일관성을 우선할 수 있는 대체 식은 B_1 과 B_2 의 가중치를 다르게 하여 설계할 수 있다. Bhattacharyya 거리는 삼각형 부등식을 만족하지 않기 때문에 실제 거리가 아니다. 그렇지만, Hellinger 거리는 $H_D = \sqrt{1 - B_C}$ 을 만족하기 때문에 가우스 매개변수 함수로 표현될 수 있다. $0 \leq H_D(p, q) \leq 1$ 을 만족하고, 가우스 분포의 경우에 p 와 q 가 같지 않으면 항상 $H_D(p, q) > 0$ 을 만족한다. 따라서 본 연구에서는 두 가우스 분포 간에 유사성 측도로 $1 - H_D(p, q)$ 를 사용한다. 이 측도는 대응하는 가우스 경계 상자 간에 확률적 IoU라 할 수 있다. 이에 반해 가우스 Wasserstein 거리는 지역적 손실 탐색에 있어서 분명한 관계를 나타내지 못한다. 객체 탐지기의 훈련과 관련하여 $p = (x_1, y_1, a_1, b_1, c_1)$ 이 네트워크에 의해 회귀되는 가우스 BB 매개변수의 집합이고, $q = (x_2, y_2, a_2, b_2, c_2)$ 가 원하는 가우스 BB에 관한 GT라 할 때, 객체 탐지기에서 지역적 손실 함수로 $1 - H_D(p, q)$ 를 사용한다. p 에 관해 해석학적인 미분이 가능하고, $p = q$ 인 경우에 이상적인 최소 점에 도달하게 된다.

IV. 실험 결과

Objectron의 기본 파이프라인에서 모델 백본은 MobileNetV2로 내장된 인코더-디코더 아키텍처를 가진다. 다중 작업 학습 접근 방식을 사용하여 탐지 및 회귀를 통해 물체의 모양을 예측한다. 모양 작업은 가능한 기본 정보 주석에 따라 물체의 모양 신호를 예측한다. 훈련 데이터에 모양 주석이 없는 경우에는 선택 사항에 해당한다. 탐지 작업에서는 주석이 달린 경계 상자를 사용하고, 상자 중심을 평균으로 하고, 표준 편차는 상자 크기에 비례하는 가우스 분포를 사용한다. 물체의 중앙 위치를 나타내는 peak로 분포를 예측한다. 회귀 작업은 8개의 경계 상자 정점의 2D 투영을 추정한다. 본 연구에서 제안된 IoU 개념의 적용 실험 결과는 SOTA IoU 기반 회귀 손실과 비교할 때 경쟁력이 있는 것으로 평가되었다. 회전된 객체의 검출과 관련해서는 최근 제안된 SOTA 손실 함수인 GWD와 유사한 결과를 나타냈다. 현

재 Objectron의 주된 문제점은 신발과 의자의 두 가지 등급만 인식할 수 있으며, 신발과 의자의 두 가지 등급에서도 정확도가 떨어진다는 점이다.

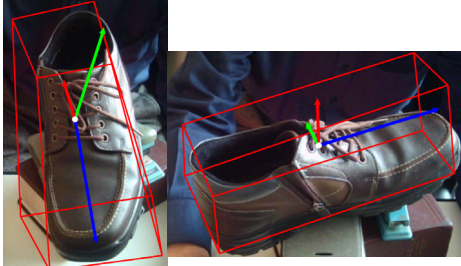


그림 3. Objectron 파이프라인을 사용한 감지
Fig. 3. Detection using the Objectron pipeline

그림 3은 기본 Objectron 파이프라인을 사용한 신발 감지 결과이고, 그림 4는 확률적 IoU 개념을 적용한 결과로 회전된 물체의 검출과 관련하여 기존 Objectron 검출기의 정확도가 떨어지지만, 제안된 IoU 개념을 적용할 때 정확도가 향상됨을 알 수 있다. 그림 5는 Objectron의 기본 파이프라인에 제안된 Hellinger 거리의 IoU 개념을 적용한 파이프라인을 나타낸다. Objectron은 경계 상자에 대한 최종 3D 좌표를 얻기 위해 견고한 포즈 추정 알고리즘(EPnP)을 사용한다^[14]. 알고리즘은 물체 차원에 대한 사전 분포가 없어도 물체의 3D 경계 상자를 복구할 수 있다. 탐지 및 추적과 관련하여 장치에서 획득한 모든 프레임에 모델을 적용하는 경우 각 프레임에서 추정되는 3D 경계 상자의 모호성으로 인해 jitter가 발생할 수 있는데, 이를 완화하기 위해 2D 물체 탐지와 MediaPipe의 Box Tracking에서 동일한 감지 및 추적 방식을 채택하고 있다.

Objectron의 3D 개체 감지 및 추적은 내부적으로 탐지 하위 그래프와 추적 하위 그래프를 사용하는 MediaPipe 그래프로 구현된다. 탐지 하위 그래프는 계산 부하를 줄이기 위해 몇 프레임마다 한 번만 ML 추론을 수행하고 출력 텐서를 9개의 키 점들이(3D 경계 상자 중심과 8개 정점) 포함된 FrameAnnotation으로 디코딩된다. 추적 하위 그래프는 Box Tracking의 상자 추적기를 사용하여 모든 프레임을 실행하여 3D 경계 상자의 투영을 단단히 둘러싸는 2D 상자를 추적하고 추적된 2D 키 점들에 대해 EPnP 알고리즘을 사용하여 3D로 생성한다. 탐지 하위 그래프에서 새로운 탐지를 사용할 수 있게 되면 추적 하위 그래프는 중첩 영역을 기반으로 탐지와 추적 결과 간의 통합도 담당한다.

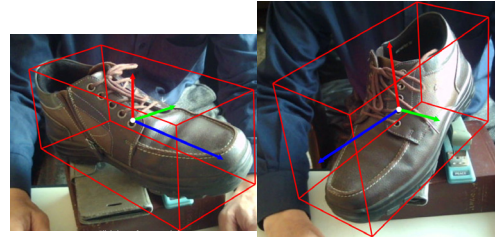


그림 4. 개선된 IoU를 적용한 신발 감지
Fig. 4. Shoe detection with improved IoU

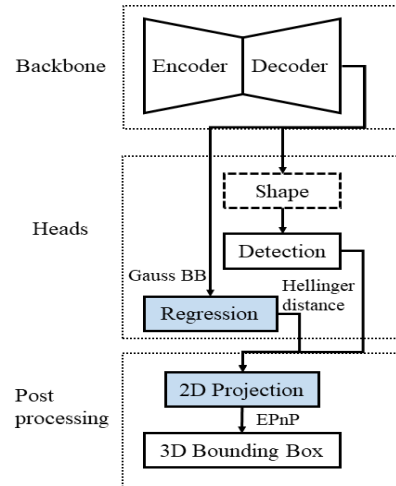


그림 5. 처리 파이프라인
Fig. 5. processing pipeline

V. 결 론

본 연구의 OBB와 관련하여 Yang 등에 의한 가우스 분포 모델이 참조되며, Hellinger 거리를 기반으로 미분 가능한 손실 함수가 사용된다. 주요 목표는 OBB의 방향 매개변수를 처리하고 Hellinger 거리를 중간 표현으로 탐구하는 것에 해당한다. 추가적인 초매개변수를 갖는 경험적인 비선형 대응이 지역적 손실 함수로 사용되기 위해서는 Hellinger 거리에 적용되어야 하며, IoU 관련 해석학적인 관계가 제시될 필요가 있다. Yang 등의 작업과 마찬가지로 객체 표현 인코딩과 관련하여 가우스 BB를 탐색하지만, 가우스 BB를 얻기 위해 기본 표현을 HBB 또는 OBB로 가정하지 않으므로 일반 분할 마스크에도 적용할 수 있다. 특정 가우스 경계 상자와 관련해서는 단순하고 미분 가능한 손실 함수로 나타낼 수 있다. 본 연구에서 제안된 IoU 개념의 적용 실험 결과는 SOTA IoU 기반 회귀 손실과 비교할 때 경쟁력이 있다.

이에 3D 객체 검출 관련 현재 우수한 파이프라인으로 소개되고 있는 구글 Objectron에 적용한 결과 Objectron의 단점이라 할 수 있는 정확도가 높아지는 것으로 나타났다.

References

- [1] Kirillov, Alexander, He, Kaiming, Girshick, Ross, Rother, Carsten, & Dollár, Piotr. "Panoptic segmentation", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 9404-9413, 2019.
DOI: <https://doi.org/10.1109/CVPR.2019.00963>
- [2] Chen Zhiming, Chen Kean, Lin Weiyao, See John, Yu Hui, Ke Yan, & Yang Cong. "PIoU Loss: Towards Accurate Oriented Object Detection in Complex Environments", Computer Vision - ECCV, 195-211, 2020.
DOI: https://doi.org/10.1007/978-3-030-58558-7_12
- [3] Yang, Xue, Yan, Junchi, Qi, Ming, Wang, Wentao, Xiaopeng, Zhang, & Qi, Tian. "Rethinking Rotated Object Detection with Gaussian Wasserstein Distance Loss", International Conference on Machine Learning (ICML), 2021.
- [4] Girshick, R., Donahue, J., Darrell, T., & Malik, J. "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(1), 142-158, 2016.
DOI: <https://doi.org/10.1109/tpami.2015.2437384>
- [5] Yu, Jiahui, Jiang, Yuning, Wang, Zhangyang, Cao, Zhimin, & Huang, Thomas. UnitBox: "An Advanced Object Detection Network. Proceedings of the 24th ACM International Conference on Multimedia", MM '16. New York, NY, USA: Association for Computing Machinery, 516-520, 2016.
DOI: <https://doi.org/10.1145/2964284.2967274>
- [6] Tytsen-Smith, L., & Petersson, L. 2018. "Improving Object Localization with Fitness NMS and Bounded IoU Loss", IEEE/CVF Conference on Computer Vision and Pattern Recognition, 6877-6885, 2018.
DOI: <https://doi.org/10.1109/CVPR.2018.00719>
- [7] Rezatofighi Hamid, Tsoi Nathan, Gwak JunYoung, Sadeghian Amir, Reid Ian & Savarese Silvio. "Generalized intersection over union: A metric and a loss for bounding box regression", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 658-666, 2019.
DOI: <https://doi.org/10.1109/CVPR.2019.00075>
- [8] Zheng Zhaohui, Wang Ping, Liu Wei, Li Jinze, Ye Rongguang & Ren Dongwei. "DistanceIoU loss: Faster and better learning for bounding box regression", Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, 12993-13000, 2020.
DOI: <https://doi.org/10.1609/aaai.v34i07.6999>
- [9] Sun, D., Yang, Y., Li, M., Yang, J., Meng, B., Bai, R., Li, L., & Ren, J. "A Scale Balanced Loss for Bounding Box Regression", IEEE Access, 8, 108438-108448, 2020.
DOI: <https://doi.org/10.1109/ACCESS.2020.3001234>
- [10] Zhang, Yi-Fan, Ren, Weiqiang, Zhang, Zhang, Jia, Zhen, Wang, Liang, & Tan, Tieniu. "Focal and Efficient IOU Loss for Accurate Bounding Box Regression", 2021.
- [11] Ding, J., Xue, N., Long, Y., Xia, G., & Lu, Q. "Learning RoI Transformer for Oriented Object Detection in Aerial Images", 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2844-2853, 2019.
DOI: <https://doi.org/10.1109/CVPR.2019.00296>
- [12] Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M. G., & Lee, J. Mediapipe: "A framework for building perception pipelines", 2019.
- [13] Pieter Van Molle, Tim Verbelen, Bert Vankeirsbilck, Jonas De Vyllder, Bart Diricx, Tom Kimpes, Pieter Simoons & Bart Dhoedt. "Leveraging the Bhattacharyya coefficient for uncertainty quantification in deep neural networks", Neural Computing and Applications volume 33, 10259-10275, 2021.
DOI: <https://doi.org/10.1007/s00521-021-05789-y>
- [14] Vincent Lepetit, Francesc Moreno-Noguer & Pascal Fua. EPnP: "An Accurate O(n) Solution to the PnP Problem", International Journal of Computer Vision volume 81, Article number: 155, 2009.
DOI: <https://doi.org/10.1007/s11263-008-0152-6>

저 자 소 개

김 용 길(정회원)



- 1990년 : 호남대학교 전산통계학과 졸업(이학사)
 - 1992년 : 광주대학교 대학원 컴퓨터공학과 졸업(공학석사)
 - 2021년 : 호남대학교 대학원 컴퓨터공학과 졸업(공학박사)
 - 2014년 ~ 현재 : 조선이공대학교 컴퓨터보안과 부교수
- 관심분야 : 네트워크보안, 지능시스템, 정보보호

문 경 일(정회원)



- 1982년 : 서울대학교 계산통계학과 졸업(이학사)
 - 1980년 : 서울대학교 대학원 계산 통계학과 졸업(이학석사)
 - 1991년 : 서울대학교 대학원 계산 통계학과 졸업(이학박사)
 - 1987년 ~ 현재 : 호남대학교 컴퓨터공학과 교수
- 관심분야 : 지능 시스템, 복잡성 과학