

---

저자 (Authors)	임정혁, 최학남, 카카니비제이, 타림, 김학일 Jeonghyeok Im, Haknam Choi, Kakani Vijay, Syed Tehreem, Hakil Kim
출처 (Source)	<a href="#">한국정보과학회 학술발표논문집</a> , 2019.12, 694-696(3 pages)
발행처 (Publisher)	<a href="#">한국정보과학회</a> The Korean Institute of Information Scientists and Engineers
URL	<a href="http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE09301699">http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE09301699</a>
APA Style	임정혁, 최학남, 카카니비제이, 타림, 김학일 (2019). 교통 상황에서의 단일 심층 신경망을 이용한 물체 검출 및 추적 통합 시스템. 한국정보과학회 학술발표논문집, 694-696
이용정보 (Accessed)	을지대학교 119.194.245.*** 2020/02/20 17:37 (KST)

---

### 저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

### Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

# 교통 상황에서의 단일 심층 신경망을 이용한

## 물체 검출 및 추적 통합 시스템

임정혁<sup>○</sup>, 최학남, 카카니비제이, 타림, 김학일

인하대학교 정보통신공학과

dlawjdgur78@inha.edu, xncui@inha.edu, vijaykakanivja@inha.edu, naqvitehreem05@gmail.com, hikim@inha.ac.kr

### Integrated Detection and Tracking System

### using a Single Deep Neural Network in Traffic Scenarios

Jeonghyeok Im<sup>○</sup>, Haknam Choi, Kakani Vijay, Syed Tehreem, Hakil Kim  
Department of Information and Communication Engineering, INHA University

#### 요 약

본 논문은 효과적인 물체 검출 및 추적을 위하여 하나의 심층 신경망을 공용으로 사용하는 프레임워크를 제안한다. 동일 신경망을 물체 검출 및 추적에 사용함으로써 정확도와 처리속도 측면에서 모두 우수한 성능을 보장한다. 신경망 가중치 학습은 2단계 학습으로 이루어지며, 다중 객체 추적에서의 학습을 위해 상관관계 필터를 사용하여 계산된 신뢰도 점수를 도입하였고, 추적 기술을 적용하여 처리시간을 단축할 수 있다. 실험결과로 부터 제안한 방법은 정확도가 KITTI 추적 데이터셋 성능 평가에서 차량 MOTP 81.04%, 보행자 MOTP 72.99% 및 처리속도 40ms로 기타 방법들보다 우수함을 확인할 수 있다.

#### 1. 서 론

주행 보조 시스템(ADAS)은 복잡한 교통 상황에서 도로 정보를 참고하여 사고를 예방하는 것을 목표로 한다. 또한 물체 검출 및 다중 물체 추적은 교통 상황을 이해하기 위한 ADAS의 기본 기능이다. 검출 결과에 따라, 다중 물체 추적은 물체의 위치를 추정하고 행동을 감시하여 교통 위험 상황을 피할 수 있다. 따라서 이 두 기능은 높은 정확도와 빠른 처리시간을 필요로 한다.

최근 다량의 실험 데이터 세트의 생산과 GPU 등 하드웨어의 발달로, 딥러닝 연구가 많이 진행되었고, 그 중 합성곱 신경망(Convolutional Neural Networks) [1]은 물체 검출 및 추적에 사용된다. 그러나 선행 연구들은 서로 다른 합성곱 신경망을 사용하여 두 알고리즘을 수행한다. 하지만 두 개의 신경망 사용에는 상당히 많은 양의 메모리가 필요하기 때문에 ADAS와 같은 임베디드 시스템에는 적합하지 않다.

이러한 한계를 극복하기 위해 물체 검출 및 추적을 위해 단일 합성곱 신경망을 설계한다. 이 모델의 학습을

위해 2 단계 학습 전략도 제안한다. 학습의 첫 단계에서 검출 단계 학습, 두 번째 단계에서는 추적 가중치를 학습한다. 자세한 내용은 2장에서 서술한다.

객체 검출 및 추적을 위한 합성곱 신경망이 높은 정확도를 달성 할 수 있지만 처리 시간은 여전히 문제가 된다. 하지만 본 연구에서 추적된 각 객체에 대해 강력한 외관 모델을 구축하고 유지하기 위해 다중 필터 추적 알고리즘에 사용되는 상관관계 필터(correlation filter) 기반 신뢰도 점수를 소개한다.

본문은 다음과 같이 구성되어 있다. 2장에서는 제안 방법을, 3장은 실험 결과를 제시하고 특정 비교 분석에 대한 논의를 포함한다. 마지막 4장에서는 결론을 서술한다.

#### 2. 제안 방법

##### 2.1 물체 검출 및 추적 구조

제안 방법은 단일 심층 신경망을 사용하여 물체 검출 및 추적을 한다(그림 1). 물체 검출에서는 Single Shot multi-box Detector(SSD) [2]를 이용하여 객체의 종류 및 위치를 예측한다. 그 후, 물체의 최종 위치를 얻기 위해 Non-Maximum Suppression(NMS)가 수행된다. 추출된 각 객체(경계 상자)의 특징점은 추적 계층에서 사용된다. 그리고 tracklet들은 객체의 외관 및 위치 정보를 포함한

\* 이 (성과물)은 산업통상자원부 ‘산업전문인력역량강화사업’의 재원으로 한국산업기술진흥원(KIAT)의 지원을 받아 수행된 연구임. (2019년 임베디드SW 전문인력양성사업, 과제번호 : N0001884)

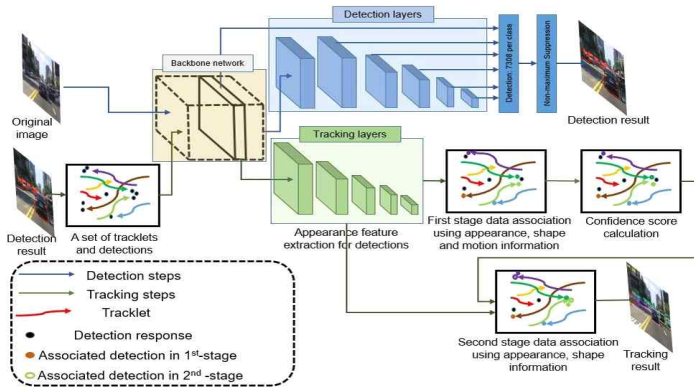


그림 1 단일 합성곱 신경망을 이용한 물체 검출 및 추적 구조도

양상블 정보에 의해 계산되며, Hungarian 알고리즘 [3]을 사용하여 최근 프레임의 검출에 사용된다. 각 tracklet을 단일 객체 추적 알고리즘에 이용되며, 상관관계 필터 [4-6]를 통해 외관 모델을 갱신한다.

## 2.2. 신경망 설계

본 연구의 신경망에서는 물체의 위치를 예측하고 다수의 물체 추적을 위한 외관 특징 추출을 수행한다. 메모리 절약을 통한 처리시간 단축을 위해 본 논문에서는 SSD기반 단일 신경망을 제안한다. 구조는 백본 신경망과 백본에서 나온 특징 지도를 이용하는 층으로 구성된다. 또한 생성된 특징 지도를 물체 검출과 추적에 모두 이용하며, 검출에는 6개의 특징 지도를 이용하며 추적에는 38 x 38 특징 지도를 제외한 총 5개의 지도를 이용한다. 이 신경망은 검출과 추적 계층의 손실 함수가 다르며, 2.3절에서 설명한다.

## 2.3 신경망 학습 방법

물체 검출 및 추적 알고리즘은 요구 사항이 다르기 때문에 두 가지 손실 함수를 사용한다. 따라서 단일 신경망을 사용하여도 2 단계의 학습 방법이 필요하다.

먼저, 백본 신경망 및 검출 계층을 학습한다. 손실 함수 및 교육 방법은 기존 SSD와 동일하다. 추적 계층은 사람 재식별 데이터 세트인 MARS [7] 또는 차량 재식별 데이터 세트인 VeRi [8]를 사용하여 학습한다. 제 2 단계에서, 학습 데이터가 백본 신경망으로부터 전달되기 때문에, 백본 신경망 및 추적 계층은 순방향 전파 동안 함께 작동하는 반면, 역 전파(Backpropagation)동안 추적 계층의 가중치만 갱신되고 백본 신경망의 가중치는 변경되지 않는다.

또한 다중 객체 추적의 정확도와 처리 시간의 균형을

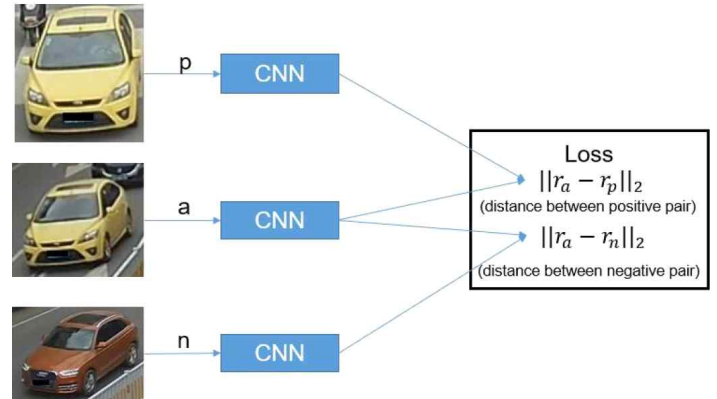


그림 2 삼중항 손실 함수 및 영상 구조도

유지하기 위해 별도의 분류 학습이 필요하다. 이를 위해 삼중 항 손실 함수 [9]를 사용하여 신경망을 학습시킨다. 그림 2에 도시된 바와 같이, 정답 영상, 앵커 영상, 오답 영상을 이용한다.

## 3. 실험 결과

### 3.1 데이터 세트 및 구현 세부 사항

검출 및 추적 평가를 포함하는 KITTI 데이터 세트 [10]의 제안 구조를 평가한다. 학습은 다양한 데이터 세트에 대해 이루어지며, 첫째, 백본 신경망 및 검출 계층은 7418개의 영상을 포함하는 KITTI 교육 데이터 세트에 대해 학습된다. 둘째로, 백본 신경망 및 추적 계층은 MARS 및 VeRi 데이터 세트에 대해 학습하면서 백본 신경망의 가중치는 변경되지 않는다.

### 3.2 평가 지표

시스템의 최종 출력은 추적기에서 나오므로 MOT 평가 지표는 시스템을 평가하는 데 사용된다. CLEAR MOT 매트릭스 [11]를 사용하여 평가한다. 여기에는 다중 객체 정확도(MOTA), 다중 객체 추적 정밀도(MOTP), Mostly Lost targets(ML), Mostly Track targets(MT) 및 총 ID 스위치 수(IDS)가 포함된다. 또한 처리 시간도 평가 지표에 포함한다.

### 3.3 시스템 평가

KITTI 추적 데이터 세트를 사용하여 시스템 평가를 진행한다. 이 데이터 세트에는 자동차 및 보행자를 포함한 29개의 영상 분류 정보가 있다. 결과는 표 1(차량) 및 표 2(보행자)와 같다. 제안 추적기는 자동차 및 보행자 추적 모두에서 합리적인 성능을 얻는다. 추적 방법은 이전 프레임의 정보만 사용하는 다중 객체 추적이며, 모든 프레

임 정보를 사용하는 일부 방법과 비교할 때 성능은 일부 요소에서 훨씬 우수하다.

표 1 KITTI 추적 데이터 세트 차량 평가

Method	MOTA ↑	MOTP ↑	MT ↑	ML ↓	IDS ↓	Time
Ours	76.82%	81.04%	48.46%	9.85%	396	0.04s
JCSTD [11]	80.57%	81.81%	56.77%	7.38%	643	0.07s
SCEA [12]	75.58%	73.39%	53.08%	11.54%	448	0.06s
MDP [13]	76.59%	82.10%	52.15%	13.38%	130	0.9s
RMOT [14]	65.83%	75.42%	40.15%	9.69%	209	0.02s
NOMT-HM [15]	61.17%	78.45%	33.85%	28.00%	241	0.09s
mbodSSP [16]	56.03%	78.84%	23.23%	27.23%	699	0.01s

표 2 KITTI 추적 데이터 세트 보행자 평가

Method	MOTA ↑	MOTP ↑	MT ↑	ML ↓	IDS ↓	Time
Ours	54.09%	72.99%	27.49%	28.52%	171	0.04s
TuSimple [17]	58.12%	71.93%	30.58%	24.05%	138	0.6s
JCSTD [11]	44.20%	72.09%	16.49%	33.68%	53	0.07s
SCEA [12]	43.91%	71.86%	16.15%	43.00%	56	0.06s
MDP [13]	47.22%	70.36%	26.12%	27.84%	87	0.9s
RMOT [14]	34.54%	58.06%	14.43%	47.42%	81	0.02s

#### 4. 결 론

일반적으로 메모리 용량이 제한된 임베디드 시스템인 지능형 차량 시스템의 메모리를 절약하기 위해 제안한 신경망을 물체 검출 및 추적에 사용할 수 있다. 객체의 위치를 예측할 수 있을 뿐만 아니라 여러 객체 추적에 사용할 수 있는 외관 특징점도 추출한다. 또한 2 단계 신경망 학습 전략을 사용하며 정확도와 처리 시간 사이의 균형을 유지하기 위해 다중 객체 추적에서 외관 모델 갱신에 신뢰 점수를 사용한다.

제안 시스템은 다른 최신 방법과 비교해 KITTI 추적 데이터셋 실험결과에서 차량 MOTP 81.04%, 보행자 MOTP 72.99% 등 제안 추적기가 많은 난제를 잘 수행하며, 처리시간 0.04s로 임베디드 시스템에 요구사항에 만족한다. 하지만 기존 SSD의 소형 물체 검출에 대한 문제 및 비교적 수치가 좋지 않은 ML, IDS 정확도 해결을 위한 연구가 진행되어야 할 것이다.

#### 참 고 문 헌

- [1] LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324.
- [2] Liu, Wei, et al. "Ssd: Single shot multibox detector." European conference on computer vision. Springer, Cham, 2016.
- [3] Ahuja, Ravindra K., Thomas L. Magnanti, and James B. Orlin. "Network flows." (1988).
- [4] Henriques, João F., et al. "High-speed tracking with kernelized correlation filters." IEEE transactions on pattern analysis and machine intelligence 37.3 (2014): 583-596.
- [5] Ma, Chao, et al. "Hierarchical convolutional features for visual tracking." Proceedings of the IEEE international conference on computer vision. 2015.
- [6] Liu, Mingjie, et al. "Occlusion-robust object tracking based on the confidence of online selected hierarchical features." IET Image Processing 12.11 (2018): 2023-2029.
- [7] Zheng, Liang, et al. "Mars: A video benchmark for large-scale person re-identification." European Conference on Computer Vision. Springer, Cham, 2016.
- [8] Liu, Xinchun, et al. "Large-scale vehicle re-identification in urban surveillance videos." 2016 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2016.
- [9] Baltas, Vassileios, et al. "PN-Net: Conjoined triple deep network for learning local image descriptors." arXiv preprint arXiv:1601.05030 (2016).
- [10] Geiger, Andreas, et al. "Vision meets robotics: The KITTI dataset." The International Journal of Robotics Research 32.11 (2013): 1231-1237.
- [11] Bernardin, Keni, and Rainer Stiefelhausen. "Evaluating multiple object tracking performance: the CLEAR MOT metrics." Journal on Image and Video Processing 2008 (2008): 1.
- [12] Tian, Wei, Martin Lauer, and Long Chen. "Online Multi-Object Tracking Using Joint Domain Information in Traffic Scenarios." IEEE Transactions on Intelligent Transportation Systems (2019).
- [13] Hong Yoon, Ju, et al. "Online multi-object tracking via structural constraint event aggregation." Proceedings of the IEEE Conference on computer vision and pattern recognition. 2016.
- [14] Xiang, Yu, Alexandre Alahi, and Silvio Savarese. "Learning to track: Online multi-object tracking by decision making." Proceedings of the IEEE international conference on computer vision. 2015.
- [15] Yoon, Ju Hong, et al. "Bayesian multi-object tracking using motion context from multiple objects." 2015 IEEE Winter Conference on Applications of Computer Vision. IEEE, 2015.
- [16] Choi, Wongun. "Near-online multi-target tracking with aggregated local flow descriptor." Proceedings of the IEEE international conference on computer vision. 2015.
- [17] Lenz, Philip, Andreas Geiger, and Raquel Urtasun. "Followme: Efficient online min-cost flow tracking with bounded memory and computation." Proceedings of the IEEE International Conference on Computer Vision. 2015.