

## 두 개의 구분자 기반의 초해상화 기법을 이용한 다중객체 검출 방법

김진서 정영민 황성빈 \*권오설

국립창원대학교 로봇제어계측공학전공

\*osk1@changwon.ac.kr

Multiple Objects Detection using Super-Resolution Method  
with Two Discriminators

Kim, Jin-Seo Jung, Young-Min Hwang, Seong-Bin \*Kwon, Oh-Seol

Dept. of Robot, Control and Instrumentation Eng., Changwon National University

## 요약

최근 자율주행에서 안전한 주행을 위해 영상 기반 다중객체 검출 기술이 활발히 연구되고 있다. 이때, 저해상도 영상은 객체 검출 단계에서 정확도가 떨어지는 한계가 있다. 본 논문에서는 이러한 문제점을 해결하기 위해 초해상화와 객체 검출을 위한 방법을 함께 사용하는 기법을 제안한다. 더 나아가 초해상화 단계에서 하나의 구분자만 사용하는 기존의 방법과 다르게 이미지 생성 과정 중간에서 추가의 구분자를 사용하여 총 두 개의 구분자를 사용하여 성능을 향상하고자 하였다. 본 논문은 한국 고속도로 교통 데이터를 사용하여 실험하였으며, 그 결과 제안된 방법의 성능이 mAP@0.5 및 F1 점수 측면에서 기존 방법보다 우수하다는 것을 확인하였다.

## 1. 서론

ADAS와 같은 자율주행 기술을 실현하기 위해서는 객체 검출 기술이 필수적이다. 일반적으로 자율주행의 ADAS(Advanced Driving Assist System) [1] 를 구현하기 위해 카메라, LIDAR, RADAR와 같은 다양한 센서들이 사용되는데 그 중 카메라를 사용했을 때 다른 방법보다 잘 동작하고 비용적 측면에서도 가장 효율적이다. [2] 또한, 자율주행의 안전 운전을 위해서 경로상의 다중객체를 검출하는 것은 필수적이다. 따라서 우리는 이 연구에서 단일 카메라를 통한 다중객체 검출에 중점을 두었다. 하지만 복잡한 환경에 우리가 찾고자 하는 객체가 있는 경우, 그 경계가 명백하지 않아 배경과 객체를 완벽하게 구분해내기에는 어려움이 있다. 이처럼 저해상도에서의 객체 검출 성능은 저조하므로 우리는 SISR(Single Image Super Resolution) [3] 기술을 활용해 이를 보완한다. 한국형 도로의 초해상화 기술과 객체 검출 기술의 융합은 많이 발전해왔지만, '저해상도의 국내 교통 이미지'에 대해 초해상화 기술과 객체 검출 네트워크를 동시에 다룬 사례는 매우 적다. 따라서 우리는 한국 도로 기반의 저해상도 이미지로부터 고해상도 이미지를 얻어 우수한 성능의 객체 검출을 구현해보려 한다.

초해상화 기술과 객체 검출 기술은 주행 시스템에 사용될 현재의 정보를 제공하는 데 있어 핵심적인 기술이다. 다중객체 검출의 주목표는 카메라의 위치, 크기, 자세, 시점과 관계없이 빠르고 정확하게 클래스를 분류하는 것이고, 이를 구현하는 데는 딥 러닝 기반의 다중객체 검출 기술이 매우 중요하다. 그리고 일반적으로 다단 검출기는 단일 검출기에 비해 정확도는 높지만 낮은 속도를 가지기 때문에 우리는 속도와 정확도에서 모두 좋은 결과를 얻기 위해 단일 검출기를 사용한다.

## 2. 제안 방법

본 논문은 한국형 도로의 저해상도 이미지에서 초해상화 기법과 객체 인식이 별도로 수행되는 종단 간 다중객체 감지 모델 방식을 제안한다. 제안된 방법은 초해상화 모델을 사용한 저해상도 도로 이미지의 객체 감지율을 증가시키는 것에 중점을 둔다. 제안된 모델의 입력은 초해상화 모델을 통과한 저해상도 도로 이미지이며, 모델의 출력은 훈련된 객체 클래스인 다중객체를 감지하기 위해 다중객체 검출기 네트워크에 주어진다. 제안된 모델의 핵심 아이디어는 고해상도 모델과 YOLOv5 [4] 모델을 별도로 학습시키는 것 대신, 데이터 세트로 SRGAN [5] 모델을 먼저 학습시키고 그다음 초해상화 된 이미지를 YOLOv5의 학습 데이터로 사용하는 것이다. 현존하는 모델과 우리 모델의 차이점은 다음을 따른다. 1) 실생활에서 사용되는 이미지는 720부터 1080의 범위를 가지지만, 기존모델은 128 x 128 이미지를 다루기 때문에 사용 범위가 제한된다. 2) 제안된 모델은 1280 x 672의 이미지로 훈련되었기 때문에 위에서 언급한 기존모델의 이미지보다 더 큰 이미지에 특히 적용된다. 3) 기존모델의 YOLOv3 함수는 더 작은 128 x 128 이미지를 위해 학습된 반면, 제안된 모델은 훨씬 더 크고 현실적인 이미지를 위해 학습되었기 때문에 몇 가지 구조적 및 절차적 차이가 있다. 4) 마지막으로, 구분자가 매우 다르다. 기존모델의 구분자는 512 x 512 초해상도 이미지를 줄여야 하지만, 제안된 모델은 훨씬 큰 초해상도 이미지를 줄여야 하므로 훨씬 더 큰 아키텍처 크기를 가지고 있다. 초해상화 모델에서 도출된 고해상도 이미지는 수식. (1)로 표현된다.

$$I_{SR} = H_{SRGAN}(I_{GSR}, I_{HR}) \quad (1)$$

여기서  $H_{SRGAN}$ 은 SR 모델의 콘볼루션 연산을 나타내고  $I_{SR}$ 은 초해상화 후의 출력 이미지를 나타낸다.  $I_{GSR}$ 은 모델의 생성자 출력 이미지이고  $I_{HR}$ 은  $I_{GSR}$ 과 비교하기 위한 실제 이미지이다. 또한, 생성자로 부터 초해상화 된 이미지는 YOLOv5의 학습데이터로도 사용된다. 우리가 제안한 모델에서 YOLOv5는 오직 초해상화 된 이미지만을 사용하여 학습된다. 검출기 네트워크는 backbone, neck, head의 세 단계로 구성된다. 첫 번째 단계는 backbone 네트워크이며, 일반적으로 CDN(cross-stage spatial network and Darknet)로 구성된다. 두 번째 단계는 neck 네트워크이며, 일반적으로 PANet(path aggregation network)로 구성된다. 마지막 단계는 head 네트워크이며 이는 YOLO 층으로도 불린다. 데이터는 CDN으로 첫 입력 되어 이미지의 특징들을 추출하고, 추출된 특징들은 PANet으로 입력되어 융합된다. 최종적으로 YOLO 층에서 차, 트럭, 속도표지판 등의 분류된 결과를 검출한다. 이는 수식. (2)로 표현된다.

$$PD = H_{fe}(CDN(I_{SR})) + H_{ff}(CDN(I_{SR})) + PAN(I_{SR}) + H_{YOLO}(I_{cls}) \quad (2)$$

여기서  $H_{fe}$ 는 CDN을 통해 추출된  $I_{SR}$  이미지의 특징들에 대한 콘볼루션 연산을 나타내고  $H_{ff}$ 는 CDN과 PAN으로 인한 특징들의 융합을 나타낸다. 그리고  $H_{YOLO}$ 는 초해상화 이미지에서 객체를 분류하기 위한 head 네트워크의 콘볼루션을 의미한다. 제안되는 모델의 최종 출력은 수식. (3)으로 표현된다.

$$I_{pred} = PD_{cls}(I_{SR}) \quad (3)$$

여기서  $PD_{cls}$ 는 초해상화 된 이미지의 객체를 분류하기 위해 제안된 검출기이며,  $I_{PRED}$ 는 초해상화 된 이미지에서 예측된 다중객체에 대한 이미지이다. 단, adversarial loss, content loss, perceptual loss, detector loss의 4가지 손실로 인해 이 네트워크는 완벽하게 학습되지 않는다. Adversarial loss는 생성자가 기존의 이미지로부터 재구성한 이미지를 구분자에서 고해상도 이미지라고 판별할 확률로 정의된다.

$$L_{adv} = \min_G \max_D [E_{I^{HR} \sim P_{train}}(I^{HR})[\log D(I^{HR})]] + E_{I^{LR} \sim P_G}(I^{LR})[[\log(1 - D(G(I^{LR})))] \quad (4)$$

트레이닝에서 사용된 content loss 네트워크는 최적화를 위해 널리 이용되는 방법이며 pixel wise 또는 MSE라 불리기도 한다.

초해상화 모델의 출력 이미지는 고해상도의 디테일이 추가됨에도 불구하고 감지를 위한 몇몇 정보들이 손실되고 흐리게 보일 것이다. 이를 최적화하기 위해 우리는 perceptual loss를 사용한다.

기존의 단일 구분자를 사용한 GAN 모델은 형태 왜곡, 중요 세부 이미지 정보 손실과 같은 문제점이 있었다. 따라서 우리는  $N \times N$  이미지를  $4N \times 4N$ 으로 초해상화 시키는 과정에서  $2N \times 2N$  이미지를 추출하여 해당 이미지에 대한 구분자를 추가하여 총 두 개의 구분자를 사용하는 방법으로 이를 보완하였다. 그림1은 기존의 하나의 구분자를 사용하던 방식과는 다른 두 개의 구분자를 사용하는 SRGAN의 구조를 나타낸다.

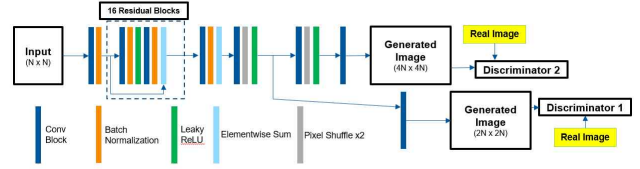


그림 1. 두 개의 구분자 기반의 SRGAN 구조

### 3. 실험 및 결과

제안한 방법의 검증은 위해서 국내 데이터 세트를 기반으로 학습시키기 위해 프로그램을 설계하였고 모든 모델은 CUDA 11.6이 설치된 단일 NVIDIA GeForce RTX 3090 딥 러닝 컴퓨터에서 테스트 및 교육되었다. 대부분의 데이터 세트는 실외 영상에서 자율주행을 위한 LR 영상의 물체 감지를 위한 외국 데이터 세트이기 때문에 연구 목적으로 데이터 세트를 구축하였다. 이 데이터 세트는 한국 고속도로 교통 장면의 15개의 비디오 연속화면 모음이며 이 연속화면을 3,168개의 이미지로 변환하였다. 그런 다음 데이터 세트를 저해상도(LR) 및 고해상도(HR) 데이터 세트 두 부분으로 나누었으며, 각각 KoHT\_LR 및 KoHT\_HR로 표시하였다. KoHT\_HR 데이터 세트는 훈련용 이미지 2,000개와 테스트용 이미지 1,068개로 구성되어 있으며, 고속도로의 데이터 세트이기 때문에 인식을 목표로 하는 물체는 ‘차’, ‘트럭’, ‘속도표지판’ 총 3개이다. 그림2는 세 가지 클래스(자동차, 트럭, 속도표지판)에 대한 정밀도(Precision) vs 재현율(Recall) 곡선이다. 정밀도는 bounding box의 예측이 정확한 정도를 측정한다. 학습 결과, 모든 클래스의 값은 0.928이었고 평균 정밀도(mAP)는 0.5임을 확인할 수 있었다. 결론적으로 우리의 탐지 모델의 성능은 YOLOv3의 성능보다 더 향상되었음을 확인하였다.

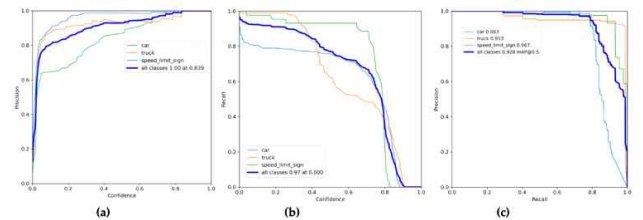


그림 2. PR-curves, (a). P-curve: precision vs confidence, (b). R-curve: recall vs confidence, (c). PR-curve: precision vs recall.

표1은 이전의 방법들과 SRMODNet으로 학습한 결과를 F1 점수와 mAP를 통해 나타난 표이다. F1 점수는 정밀도와 재현율을 사용하여 평가되며 mAP(mean Average Precision)는 훈련 결과값 검출률(recall), 정확도(precision)로 구한다. 표에 제시된 결과로부터 우리가 제안한 모델이 다른 방법에 비해 최고의 성능을 보여주고 KoHT에 대해 2번째로 우수한 mAP 및 F1 점수값을 나타낸다는 것을 알 수 있다.

표 1. mean average precision (mAP) 와 F1 점수를 통한 객체 검출 성능 비교

Dataset	KoHT	
	mAP@0.5	F1 score
Joint-SRVDNet [18]	72.46	0.702
YOLOV3_HR [35]	85.33	0.826
SRMODNet (ours)	92.20	0.872
YOLOV5s_HR [39]	99.40	0.990

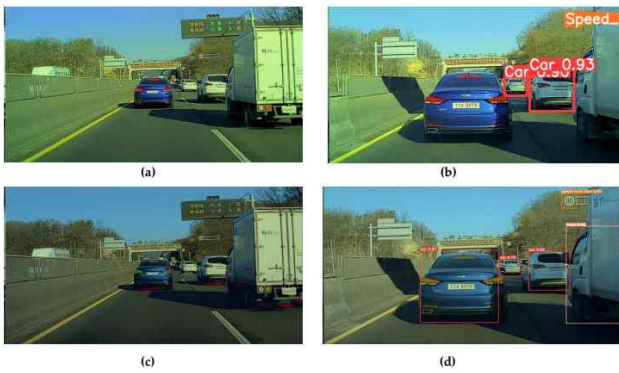


그림 3. KoHT 데이터 세트: (a) 저해상도, (b) 저해상도에 대한 예측 (c) 초해상화 이미지 (d) 초해상화 이미지에 대한 예측.



그림 4. 서로 다른 상황의 한국 고속도로에서 도출된 다중객체 검출의 결과들

그림 3과 4는 교통 환경에서 초해상화 전후의 물체 감지 성능을 비교한 사진이다. 초해상화 기법 전의 저해상도 이미지를 통해 객체 검출했을 때보다 초해상화 기법 후의 이미지로 객체 검출했을 때 정확도가 더 향상됨을 확인할 수 있다.

#### 4. 결론

본 논문은 한국 고속도로 환경에서 저해상도 영상의 다중객체를 인식에 대한 방법을 제안하였다. 제안한 모델은 저해상도 이미지로부터 두 개의 구분자를 통해 고해상도 이미지를 생성함으로써 정보의 복원을 높였으며, 객체 인식을 위해 Yolov5s를 사용해 간편하면서도 인식 정확성을 높이고자 하였다. 다양한 실험 결과를 통해 새로 제안한 모델의 성능은 기존의 방식들의 성능보다 우수함을 확인하였다.

#### 감사의 글

본 성과물은 중소벤처기업부에서 지원하는 2022년도 산학연 플랫폼 협력 기술개발 사업(S3311946) 및 교육부와 한국연구재단의 재원으로 지원을 받아 수행된 3단계 산학연협력 선도대학 육성사업(LINC 3.0)의 연구결과임.(1345356160)

#### 5. 참고문헌

[1] Jaswanth, N.; Chinmayi, N.; Rolf, A.; Hrishikesh, V. A Progressive Review - Emerging Technologies for ADAS Driven Solutions. IEEE Transactions on Intelligent Vehicles., 2021, 1, 1-18  
 [2] Kim, J.; Hong, S.; Kim, E. Novel On-Road Vehicle Detection

System Using Multi-Stage Convolutional Neural Network. IEEE Access., 2021, 9, 94371-94385.

[3] 15. Saeed, A.; Salman, K.; Nick, B. A Deep Journey into Super-resolution: A Survey. ACM Comput.Surv.2020, 53, 1-21.  
 [4] Yingfeng, C.; Tianyu, L.; Hongbo, G.; Hai, W.; Long, C.; Yicheng, L.; Miguel, S.; Zhixiong, L. YOLOv4-5D: An Effective and Efficient Object Detector for Autonomous Driving. IEEE Transactions on Instrumentation and Measurement., 2021, 70, 4503613-4503613.  
 [5] Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z. Photo-realistic single image super-resolution using a generative adversarial network.