

# 어안렌즈 카메라로 획득한 영상에서 차량 인식을 위한 딥러닝 기반 객체 검출기

Tang Quang Hieu<sup>†</sup>, 연승호<sup>\*\*</sup>, 김재민<sup>\*\*\*</sup>

## Deep Learning based Object Detector for Vehicle Recognition on Images Acquired with Fisheye Lens Cameras

Tang Quang Hieu<sup>†</sup>, Sungho Yeon<sup>\*\*</sup>, Jaemin Kim<sup>\*\*\*</sup>

### ABSTRACT

This paper presents a deep learning-based object detection method for recognizing vehicles in images acquired through cameras installed on ceiling of underground parking lot. First, we present an image enhancement method, which improves vehicle detection performance under dark lighting environment. Second, we present a new CNN-based multiscale classifiers for detecting vehicles in images acquired through cameras with fisheye lens. Experiments show that the presented vehicle detector has better performance than the conventional ones.

**Key words:** Deep Learning, Vehicle Detection, Fisheye Lens Cameras, Convolutional Neural Network

### 1. 서 론

딥러닝 학습 기법의 발전과 대용량의 영상 데이터 수집이 가능하면서, 영상에서 객체를 인식하는 기술은 높은 인식 성능을 달성하게 되었다[1]. 딥러닝 기반 객체 검출 방법 중에서 빠른 검출 속도와 높은 성능을 보여주는 방법으로는 Ren 등이 제안한 Faster R-CNN 방법이 있다[2]. 이 방법은 피라미드 구조의 합성곱신경회로망(convolutional neural network)으로 특징을 추출하고, 영상에서 k개의 앵커 박스(anchor box)를 이동시키면서 앵커 박스 안에 객체가 존재하는지를 판단한다. 이때 객체의 존재 여부를 앵커 박스 영역 내의 있는 신경회로망에서 추출한 특징들로 판단한다. 최종적으로 객체를 포함하는 영역의 크기

및 위치, 영역 내에 존재하는 객체의 신뢰도 및 객체가 특정 종류일 확률을 계산한다. 하지만 이 방법은 실시간으로 구현하기에는 연산량이 많다.

Redmon 등은 앵커 박스를 움직여가면서 객체의 영역을 찾는 대신 합성곱신경회로망의 최상층(해상도: 7x7)의 각 셀에서 b개 테두리 상자로 객체 영역의 위치와 크기, 테두리 상자가 객체를 포함하는 신뢰도, 포함된 객체가 특정 종류일 확률을 완전 연결 망(fully connected network)으로 계산하였다[3]. 이 방법은 앵커 박스를 이동시키는 대신 고정된 7x7xb 개의 테두리 상자에서 한번만 객체 검출을 수행하므로 YOLO(You Only Look Once)라고 한다. 이 방법은 매우 적은 연산량을 필요로 하며 실시간 객체 검출 구현이 가능하다. 하지만 작은 객체의 검출 성능이

\* Corresponding Author : Jaemin Kim, Address: (252-401) Yanjaedaero-ro 12180, Songpa-gu, Seoul, Korea, TEL : +82-2-320-1634, FAX : +82-2-320-1193, E-mail : jmkim@hongik.ac.kr

Receipt date : Jan. 14, 2019, Approval date : Jan. 18, 2019  
<sup>†</sup> Department of Electrical Information Control Engineering, Hongik University (E-mail : tqhieu@gmail.com)

<sup>\*\*</sup> Department of Electrical Information Control Engineering, Hongik University (E-mail : yearnsh@hanmail.com)

<sup>\*\*\*</sup> School of Electronic and Electrical Engineering, Hongik University

\* This work was supported by 2016 Hongik University Research Fund.

낮은 문제점이 있다. Liu 등은 작은 객체도 효과적으로 검출할 수 있는 방법을 제안하였다. 이 방법은 테두리상자 안에서 한번에 객체의 위치 및 크기, 객체의 종류별 확률을 모두 한번에 처리한다고 하여 SSD (single-shot detector)라고 한다[4]. 이 방법은 합성곱신경회로망의 높은 해상도 층부터 최저해상도 층까지 다양한 층에서 객체를 검출하는 다중스케일 검출기를 구현하였다. 이 방법은 YOLO와 비슷한 연산량으로 높은 검출율을 보여주고 있다. Redmon 등은 특징 검출을 위한 합성곱신경회로망을 연속된 3x3과 1x1 합성곱(convolution)으로 이루어진 19개의 합성곱 층(convolutional layer)과 5개의 max pooling layers로 구성하고, 최상층의 해상도를 13x13으로 높였다. 이 신경회로망을 darknet-19라고 작명하였다. 객체 검출을 위한 분류기를 완전 연결 망(fully connected network)대신 합성곱신경회로망(CNN)으로 구성된 YOLOv2를 제안하였다[5]. 이 방법은 YOLO보다 적은 연산량으로 SSD보 우수한 검출 성능을 구현하였다. He 등은 수천개의 객체 검출에 필요한 풍부한 특징을 추출할 수 있는 매우 깊은 층으로 구성된 합성곱신경회로망을 제안하였다[6]. 이는 ResNet이라고 한다. 이 신경회로망은 매우 깊은 층(deep layers)을 안정적으로 학습할 수 있는 구조로, 객체 검출 및 인식 등에서 높은 성능을 보여주고 있다. 하지만 층이 깊어지는 만큼 연산량이 늘어난다. Lin 등은 앵커 박스를 사용하여 객체를 검출하는 방법에서 신경회로망을 학습함에 있어 객체 샘플의 수가 배경의 샘플의 수에 비하여 매우 적고 이로 인하여 학습된 검출기의 성능을 저하되는 문제점을 찾아내고 이를 보완하여 학습하는 방법을 제시하였다[7]. 이는 RetinaNet이라고 불린다. Redmon 등은 좀 더 풍부한 특징을 추출하기 위하여 darknet-19과 ResNet의 혼합형태로 53개의 합성곱층(convolutional layer)으로 사용하였다. 이는 darknet-53이라 한다. darknet-53은 darknet-19보다는 많은 연산량을, ResNet에 비하여 적은 연산량을 사용한다[8]. Redmon등은 Darknet-53으로 특징을 추출하고, 다양한 크기의 객체를 검출하기 위하여 최상층, 차상층, 차차상층의 3층에서 객체 검출을 위한 분류기를 생성하였다[8]. 이 방법은 YOLOv3라고 불린다.

Fig. 1은 합성곱신경회로망을 기반으로 하는 다양한 객체 검출기의 연산량 대비 성능을 보여주고 있

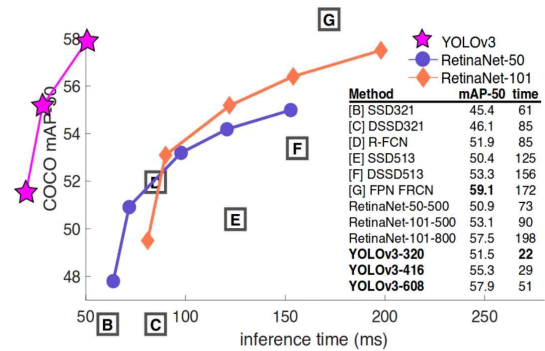


Fig. 1. Speed and accuracy tradeoff on the mAP at 0,5 IOU metric (Source: YOLOv3 paper),

다. YOLOv3가 연산량 대비 검출 성능이 다른 방법에 비하여 월등하게 우수함을 보여주고 있다.

이러한 딥러닝 기반 객체 인식 기술은 다양한 환경에서 다양한 객체를 인식하려고 하면 그 성능이 낮아지는 한계가 있다. 하지만 인식하고자 하는 객체를 특화하고, 객체를 촬영하는 카메라의 시점과 종류를 특화하고, 주변 환경을 특화한다면 높은 인식률의 객체 인식이 가능하다.

본 논문에서는 지하 주차장의 천정에 설치된 어안렌즈 카메라를 통하여 획득된 영상에서 차량을 인식하는데 특화된 딥러닝 기반 실시간 객체 인식 기술을 기술하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서 본 논문의 제안하는 방법에 대하여 기술한다. 2.1에서는 어두운 조명 환경에서 자동차의 검출 성능을 향상 시키기 위한 영상의 화질개선 방법에 대하여 기술한다. 2.2에서 주차장의 천정에 설치된 어안렌즈 카메라로 획득한 영상에서 차량을 검출하는데 특화된 객체 검출기에 대하여 기술한다. 3장에서는 실험을 통하여 제안한 화질 개선 방법이 검출 성능을 향상시키며, 제안한 객체 검출기는 YOLOv3 보다 적은 연산량으로 YOLOv3와 비슷한 성능으로 차량을 검출함을 보여준다. 4장에서는 결론을 맺는다.

## 2. 제안하는 방법

### 2.1 입력 영상 화질 개선

지하 주차장 내부에 균등하게 조명을 배치하여도 건물의 코너 부분에 주차한 차량은 조명이 어두워 Fig. 2(a)에서와 같이 차량 후방의 윤곽을 알아보기



Fig. 2. Images showing vehicles parked in a corner area: (a) an original input image, (b) the enhanced input image.

힘들다. 합성곱신경회로망 기반 객체 검출기에 객체를 포함하는 영상을 입력으로 넣었을 때, 신경회로망의 중간층에서 출력값을 시각화하여 보면 객체 윤곽의 에지가 강조되어 나타난다. 즉 차량의 검출율을 향상 시키기 위해서는 차량의 윤곽을 뚜렷하게 하는 화질 개선이 필요하다. 물론 차량 검출기를 학습할 때 입력 영상의 밝기(intensity), 색조(hue) 등에 변화를 주는 데이터 증강 방법을 사용하지만, 일반적으로 검출하는 객체의 변화 범위가 크면 객체 검출의 성능이 저하된다.

본 논문에서는 입력 영상의 화질 개선을 위하여 감마 보정(gamma correction)과 에지 향상(edge enhancement)을 적용하였다. 감마 보정은 아래와 같다.

$$V_g = A \left( \frac{V_{in}}{dynamic\ range} \right)^{1/\gamma} \quad (1)$$

$$dynamic\ range = \max V_{in} - \min V_{in}$$

식 (1)에서  $V_{in}$ 는 입력 영상의 밝기,  $dynamic\ range$ 는 입력 영상의 밝기 범위,  $V_g$  감마 보정된 영상의 밝기,  $A$ 는 보정된 영상의 밝기 범위를 각각 나타낸다. R,G,B 각각 독립적으로 보정한다.

에지를 향상함에 있어 에지가 뚜렷한 부분은 향상을 적게 하고, 에지가 약한 부분은 향상을 많이 할

필요가 있다. 본 논문에서 적용한 에지 향상 방법은 다음과 같다.

$$V_e(x,y) = V_g(x,y) + \omega \frac{V_g(x,y) - \mu(x,y)}{\sigma(x,y) + \epsilon} \quad (2)$$

$$\mu(x,y) = \frac{1}{(2L+1)^2} \sum_{i=-L}^L \sum_{j=-L}^L V_g(x+i,y+j)$$

$$\sigma(x,y) = \frac{1}{(2L+1)^2} \sum_{i=-L}^L \sum_{j=-L}^L (V_g(x+i,y+j) - \mu(x,y))^2$$

식 (2)  $\mu(x,y)$ 와  $\sigma(x,y)$  지역 평균과 지역 표준편차이다.

### 2.2 차량 검출에 특화된 객체 검출기

기존의 객체 검출기 중에서 YOLOv3는 연산량 대비 검출 성능이 다른 방법에 비하여 월등하게 우수함을 보여주고 있다. 반면에 YOLOv3보다 연산량이 적은 YOLOv2는 사람같이 작은 객체의 검출에서는 성능이 YOLOv3보다 성능이 많이 낮지만, 버스같은 크기가 큰 객체의 검출의 경우 성능이 YOLOv3에 근접한다[7].

본 논문에서는 지하 주차장의 천정에 설치된 카메라를 통하여 획득된 영상에서 차량만을 인식하는데 특화된 검출기로, YOLOv3보다 연산량이 작으면서 큰 객체의 검출 성능에서는 큰 차이없는 YOLOv2를 고려한다. darknet-19에서 추출된 특징은 darknet-53에서 추출된 특징에 비하여 풍부하지 않지만, 차량만을 검출하는데는 충분하다고 생각된다. 이는 실험을 통하여 입증한다.

천정에 설치된 어안렌즈카메라로 하단 주차면의 차량을 촬영하면, 영상의 중심에 놓인 차량은 크기가 크지만 영상의 외곽에 놓인 차량은 크기가 작다. 다양한 크기의 차량을 검출하기 위해서는 YOLOv3에

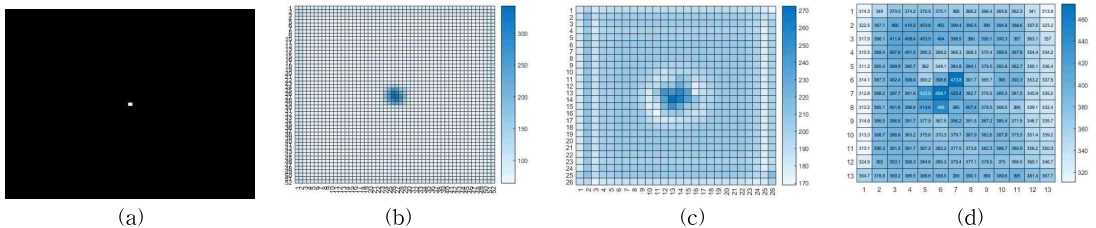


Fig. 3. The larger the layer value of the network, the larger the area in which the center pixel affects: (a) the impulse image, (b) the spread of the impulse image at the layer 10, (c) the spread of the impulse image at the layer 16, (d) the spread of the impulse image at the layer 29.

서와 같이 다중스케일로 객체를 검출하는 것이 필요하다. 때문에 제안하는 검출기는 darknet-19 합성곱 신경회로망을 기반으로 다중스케일에서 객체 검출하는 구조를 가지게 된다.

2.2.1 차량 검출에 특화된 특징 추출기

제안하는 검출기에서 특징을 추출하는 합성곱 신경회로망은 darknet-19를 채택한다. 이는 Fig. 4에서 보여주고 있으며 아래와 같이 요약된다.

- 연속된 3×3 과 1×1 합성곱(convolution)으로 이루어진 19개의 합성곱층과 5개의 max pooling 층
- 각각의 합성곱층의 입력에 배치정규화(batch normalization) 적용
- 활성화 함수로 Leaky ReLU 적용

2.2.2 차량 검출에 특화된 다중스케일 분류기

darknet-19의 입력 영상은 416×416이며 차상층(the next higher layer, the next lower resolution layer)은 26×26 셀들로 구성되며, 최상층(the highest layer, the lowest resolution layer)의 13×13인 셀들

로 구성된다.

제안하는 다중스케일 검출기에는 YOLOv2에서와 같이 최상층의 각 13×13개의 셀 각각에 5개의 객체를 검출하는 분류기가 있다. 분류기에서는 검출한 객체들을 테두리 상자(bounding box)로 나타내는데, 테두리 상자의 중심좌표 (x,y), 넓이, 높이, 신뢰도 그리고 테두리에 포함된 객체의 클래스확률 등을 추정한다. 이들 분류기의 입력으로 사용되는 특징 벡터는 다음과 같이 생성한다. 우선 최상층의 해당 셀에서 추출하는 1024개의 특징과 최상층의 셀에 해당하는 차상층의 4셀에서 추출하는 4×64개의 특징을 병합한다. 다음으로 병합한 1280개의 특징을 1024개의 1×1×1280 합성곱 필터를 통과시켜 구한 1024개의 특징을 사용한다. 이는 Fig. 4에서 보여주고 있다.

제안하는 다중스케일 검출기는 최상층의 13×13 셀에서 5개의 객체를 검출하는 것에 추가하여, 차상층의 26×26 개의 각 셀에서도 5개의 객체를 검출하는 분류기가 있다. 이들 분류기의 입력으로 사용되는 특징 벡터는 다음과 같이 생성한다. 우선 차상층의 해당 셀에서 추출하는 1024개의 특징과 차상층의 셀에

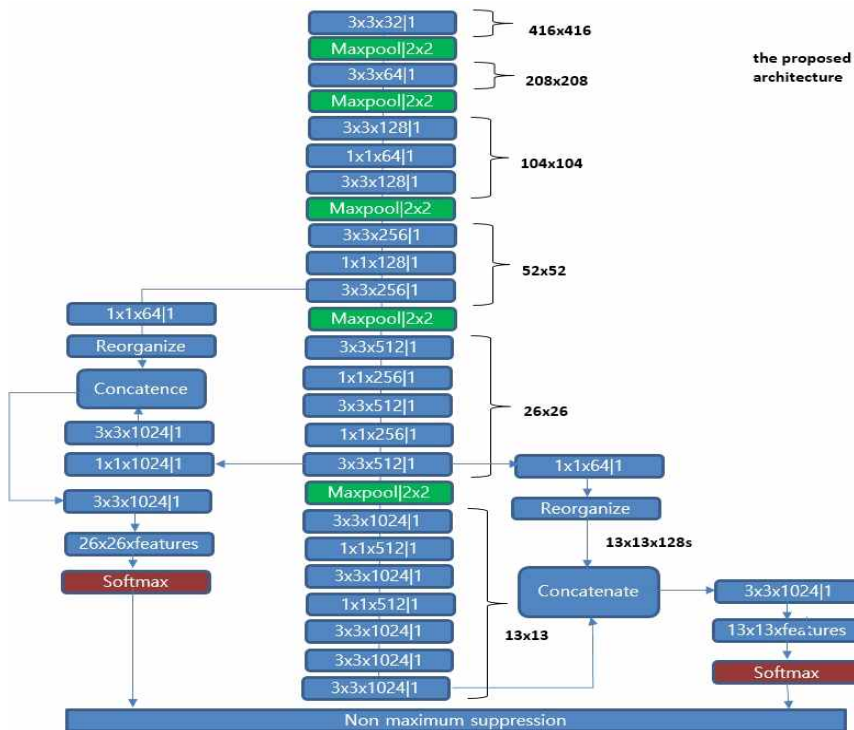


Fig. 4. The proposed CNN based multiscale object detector.

해당하는 차차상층의 4셀에서 추출하는 4×64개의 특징을 병합한다. 다음으로 병합한 1280개의 특징을 1024개의 1×1×1280 합성곱 필터를 통과시켜 구한 1024개의 특징을 사용한다. 이는 Fig. 4에서 보여주고 있다.

최상층과 차상층에서 중복되어 검출된 객체들은 non maximum suppression을 과정을 통하여 정리한다.

YOLOv3에서는 최상층, 차상층, 차차상층에서 각각 분류기로 객체를 검출을 하는데, 각 층의 객체 검출기는 최상층에서 추출한 특징을 업샘플링(upsampling)하여 병합한 후 합성곱 필터를 거쳐 분류기에 사용할 특징을 추출한다. Fig. 3에서 보여준 것과 같이 최상층의 특징은 광범위한 영역에서 추출한 특징이기 때문에 제한된 차량 영역 밖의 배경이 특징 값에 영향을 줄 수 있다. 물론 학습과정에서 최상위층에 추출한 특징 중에서 필요한 부분만 선택할 수도 있지만, 본 논문에서 득보다 실이 많다고 판단하여 최상위 층에서 구한 특징을 차상위 층의 분류기에 사용하지 않았다. 제안한 방법에서 이러한 접근의 타당성은 실험을 통하여 입증한다.

또한 차량의 크기는 일정 크기 이상이기 때문에 YOLOv3와 달리 최상층과 차상층 두 층에서만 객체를 검출한다. 본 논문에서 제안하는 다중스케일 검출기의 최종 구조는 Fig. 4에서 보여주고 있다.

합성곱신경회로망은 학습 시 최소화할 손실 함수(loss function)로는 YOLOv2에서 사용한 함수를 적용하였다. 그 형태는 다음과 같다.

$$\begin{aligned} & \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \\ & + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} [(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2] \\ & + \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} (C_i - \hat{C}_i)^2 + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{noobj} (C_i - \hat{C}_i)^2 \\ & + \sum_{i=0}^{S^2} \sum_{c \in classes} 1_{ij}^{obj} (p_i(c) - \hat{p}_i(c))^2 \end{aligned} \quad (3)$$

식 (3)에서  $x, y, w, h$  객체를 포함하는 테두리 상자의 중심 좌표, 넓이, 높이를 나타내며,  $C_{ij}$ 는  $i$ 번째 셀의  $j$ 번째 테두리상자에 포함된 객체의 신뢰도,  $p_i(c)$   $i$ 번째 셀에 존재할 객체가 특정 클래스일 확률을 나타낸다.  $1_{ij}^{obj}$ 는  $i$ 번째 셀의  $j$ 번째 테두리상자에 객체가 존재하면 1 아니면 0 값을 가진다.  $1_i^{obj}$ 는  $i$ 번째

셀에 객체의 중심이 놓여지면 1 아니면 0 값을 가진다. 본 논문에서는 총 객체의 학습 데이터에서 객체를 포함하는 셀의 갯수와 배경만 포함하는 셀의 개수의 비를 바탕으로  $\lambda_{noobj}$  값을 설정하였다.

## 4. 실험 결과 및 고찰

### 3.1 학습

실험에 사용하는 데이터는 주차장의 천정에 설치한 16대의 어안렌즈 카메라를 통하여 획득한 영상이다. 학습에 사용한 영상 데이터의 해상도는 640×480이며, 데이터 수량은 250,000장이다. 학습에 앞서 데이터 량을 늘리기 위하여 데이터 확대(data augmentation)을 적용하였다. 학습 시 신경회로망의 초기값은 Redmon[9] 등이 제공하는 이미지넷(ImageNet) 데이터로 사전학습한 결과를 사용하였다.

화질 개선 성능평가를 위한 실험은 배경의 조명이 어두운 특정 카메라에서 획득한 영상을 사용하였고, 제안하는 객체 검출기의 성능평가를 위해 VOC 데이터를 사용하였다. 최종적으로 지하주차장의 16대 모든 카메라에 대한 성능평가를 하였다. 화질을 개선한 영상으로 학습한 검출기의 성능과 원 영상으로 학습한 검출기의 성능의 비교는 16대의 모든 카메라에서 획득한 영상이 아니라 배경의 조명이 어두운 특정 카메라에서 획득한 영상으로 검출기를 학습하고 테스트하였다. 이 때 특정 카메라에서 획득한 영상 데이터가 충분하지 않아, 학습 시 신경회로망의 초기값은 16개 카메라에서 획득한 모든 영상으로 학습한 결과를 초기 값으로 사용하였다.

획득한 250,000장의 영상은 240,000 개의 학습 데이터와 10,000 개의 테스트 데이터로 나누었다. 학습과 테스트를 매 반복 학습마다 평가를 하면서, 100,000번 반복 수행 하였다. 학습과 테스트 결과는 50,000번을 넘어서면서 큰 변화가 없었고, 100,000번 이전에 최대 테스트 성능을 가진 학습결과를 선택하였다. 선택한 검출기는 실제 필드에서 테스트하면서 오버피팅(overfitting) 여부를 확인하였다. 이러한 학습, 테스트, 검증(validation)을 거쳐 최종적으로 검출기를 선택하였다.

### 3.2 입력 영상 화질 개선에 따른 검출 성능 향상

Table 1은 화질을 개선한 영상으로 학습한 검출기

Table 1. Detection performance of parked vehicles in underground parking lot; Training data: 40,000 images acquired from the 9th camera installed on parking lot ceiling. Test data: 2,000 images acquired from the 9th camera

|                                    | true positive | false positive | false negative |
|------------------------------------|---------------|----------------|----------------|
| Trained with the original data set | 10153         | <b>19</b>      | 682            |
| Trained with enhanced data set     | <b>10819</b>  | 122            | <b>16</b>      |

에서는 검출하지 못하는 차량의 숫자가 현저하게 줄어들음을 보여주고 있다. 이는 화질이 개선된 영상에서 추출한 특징이 차량의 검출이 효과적임을 보여주고 있다. 반면에 화질을 개선한 영상으로 학습한 검출기는 복잡한 배경을 차량으로 잘못 검출하는 경우가 많아지는 단점이 있다.

어두운 조명 하에서 주차한 차량의 후방 부분은 배경과 잘 구분이 되지 않고 차량의 후방 윤곽도 알아보기 어렵다. 어두운 부분의 명암 대비(contrast)를 향상 시키는 식 (1)의 감마 보정과 약한 에지를 개선시키는 식 (2)의 영상 선명화는 차량의 전반적인 윤곽을 뚜렷하게 하여 차량 검출의 성능을 개선한다.

반면에 감마 보정은 밝은 부분의 명암 대비를 저하시키고, 영상 선명화는 배경 노이즈로 인하여 생기는 에지를 증폭하여, 배경을 차량으로 잘못 검출하는 단점이 있다. 전체적으로 화질을 개선한 영상으로 학습한 검출기는 원 영상으로 학습한 검출기에 비하여 오검출의 총 개수는 701개에서 138개로 줄어들음을 보여주고 있다.

### 3.3 제안한 검출기의 성능 평가

Table 2는 VOC 데이터로 제안한 검출기와 YOLOv2,

YOLOv3와 성능을 비교한 결과를 보여주고 있다. 제안한 검출기와 YOLOv3는 YOLOv2에 비하여 모든 객체의 검출에서 큰 차이로 우수한 성능을 보여주고 있다. 객체의 크기가 작은 person, bicycle의 경우에는 YOLOv3가 제안한 검출기에 비하여 우수한 성능을 보여 주고 있으며, 상대적으로 크기가 큰 bus, motorbike의 경우 제안한 검출기가 YOLOv3에 비하여 우수한 성능을 보여주고 있다. YOLOv3는 합성곱 신경회로망의 최상층, 차상층, 차차상층 등 3개의 층에서 객체 검출을 위한 분류기를 생성하였지만, 제안한 합성곱신경회로망은 차상층, 차차상층 2개의 층에서 분류기를 생성하였다. 때문에 작은 크기의 객체 검출의 경우 YOLOv3의 성능이 우수하다.

Table 3는 지하주차장의 천정에 설치된 16대의 어안렌즈 카메라로 하단의 주차면에 주차한 차량을 촬영한 영상에서 차량 검출 성능을 평가하였다. 제안한 검출기와 YOLOv3는 YOLOv2에 비하여 주차 차량의 검출에서 큰 차이로 우수한 성능을 보여주고 있다. 제안한 검출기는 YOLO v3에 비하여 false negative의 개수가 조금 적지만, false positive의 개수는 조금 많다. IOU는 YOLOv3가 조금 높다. 연산량의 경우 제안한 검출기는 YOLOv3의 연산량의 72

Table 2. mAP performance in VOC data; Training data: 8000 images of VOC 2007 and VOC 2012, Test data: 5000 images of VOC 2007

|                  | car          | bus          | motorbike    | person       | bicycle      |
|------------------|--------------|--------------|--------------|--------------|--------------|
| YOLOv2           | 69.55        | 67.35        | 66.39        | 60.33        | 65.65        |
| YOLOv3           | <b>76.85</b> | 71.80        | 72.54        | <b>72.45</b> | <b>72.32</b> |
| the proposed one | 75.31        | <b>72.38</b> | <b>74.88</b> | 71.43        | 70.68        |

Table 3. Detection performance of parked vehicles in underground parking lot; Training data: 240,000 images acquired from 16 cameras installed on parking lot ceiling. Test data: 10,000 images acquired from 16 cameras

|                  | true positive | false positive | false negative | IOU          | frames per second |
|------------------|---------------|----------------|----------------|--------------|-------------------|
| YOLOv2           | 77073         | 2830           | 3374           | 83.65        | <b>60.15</b>      |
| YOLOv3           | 78073         | <b>1412</b>    | 2374           | <b>87.73</b> | 37.04             |
| the proposed one | <b>78101</b>  | 2071           | <b>2346</b>    | <b>86.97</b> | 51.28             |

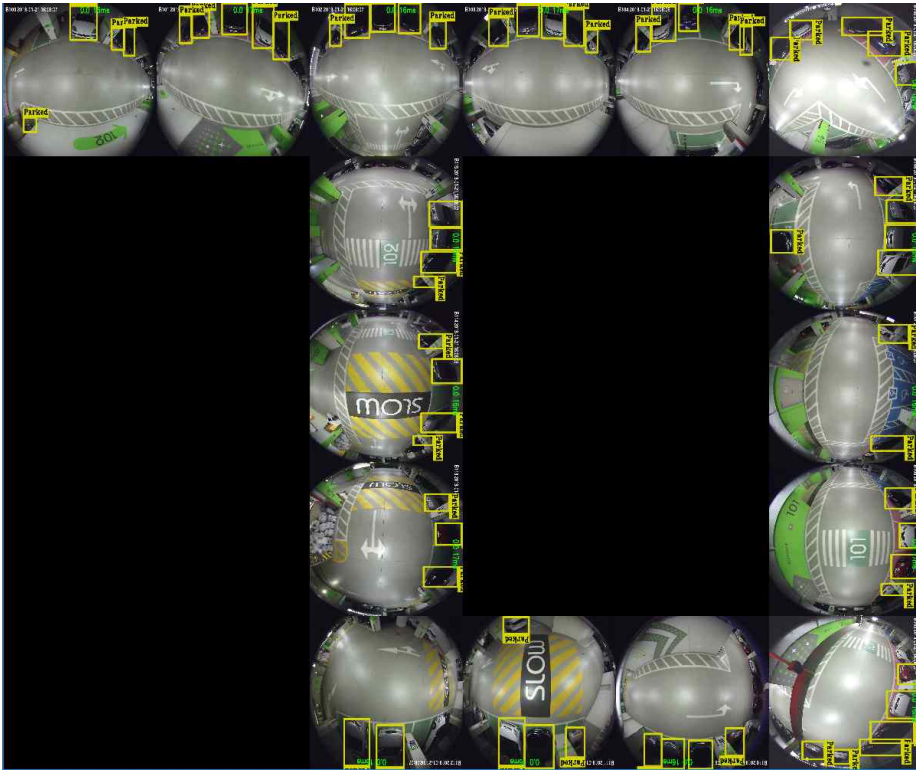


Fig. 5. The Detection results of the proposed CNN based multiscale object detector; Vehicles detected from 16 camera images with overlapped surveillance area.

% 정도이고, YOLOv2의 연산량의 1.2배 정도이다. 제안하는 검출기는 주차차량 검출에 있어 YOLOv3의 약 72% 정도의 연산량으로 YOLOv3와 비슷한 검출 성능을 보여 주고 있다.

Fig. 5는 감시영역이 중복된 16대의 카메라 영상에 검출된 차량을 붉은색 테두리 상자로 표시한 결과를 보여주고 있다. 감시 영역이 중복되기 때문에 한 카메라의 감시 구역에서 작은 크기의 차량은 인접한 카메라의 감시 구역에서는 상대적으로 큰 크기를 가지고 있다. 때문에 각 카메라 영상의 외곽 부근을 제외한 중앙 부근의 일정 크기 이상의 차량을 검출하면 주차한 모든 차량을 검출할 수 있다.

### 5. 결 론

본 논문에서는 지하주차장에 주차한 차량의 검출에 최적화된 객체 검출기 알고리즘을 제안하였다. 제안한 알고리즘은 합성곱신경회로망 기반의 특징 추출과 객체 검출을 위한 다중스케일 분류기로 구성된

다. 검출기는 주차차량 검출에 있어 YOLOv3의 약 72% 정도의 연산량으로 YOLOv3와 비슷한 검출 성능을 보여 주고 있다. 또한 어두운 조명 환경에 주차한 차량을 효과적으로 검출하기 위하여 영상 화질 개선 방법을 제안하였다. 화질 개선은 감마 보정과 적외선 에지 향상으로 구성된다.

### REFERENCE

[1] V.P. Than, T.B. Nguyen, and S.T. Chung, "Accurate Human Localization for Automatic Labelling of Human from Fisheye Images," *Journal of Korea Multimedia Society*, Vol. 20, No. 5, pp. 769-781, 2017.

[2] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 6, pp. 1137-1149,

2017.

[ 3 ] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only Look Once: Unified, Real-time Object Detection," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779-788, 2016.

[ 4 ] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, et al., "Ssd: Single Shot Multibox Detector," *Proceeding of European Conference on Computer Vision*, pp. 21-37, 2016.

[ 5 ] J. Redmon and A. Farhadi, "Yolo9000: Better, Faster, Stronger," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6517-6525, 2017.

[ 6 ] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016.

[ 7 ] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2999-3007, 2017.

[ 8 ] YOLOv3 : <https://pjreddie.com/publications/> (accessed Jan., 11, 2019)

[ 9 ] J. Redmon. Darknet: Open Source Neural Networks in c, <http://pjreddie.com/darknet> (accessed Jan., 11, 2019).



Tang Quang Hieu

2012년~2017년 Electronics Engineering, HCMUT  
 2018년~현재 홍익대학교 전기정보 제어공학과 석사  
 관심분야 : deep learning, object detection, computer vision



연 승 호

2007년~2011년 홍익대학교 전자 전기공학부 학사  
 2011년~2013년 홍익대학교 전기 정보제어공학과 석사  
 2013년~현재 홍익대학교 전기정보 제어공학과 박사

관심분야 : 화질개선, 객체검출, 객체인식



김 재 민

1980년~1984년 서울대학교 전기 공학과 학사  
 1984년~1986년 서울대학교 전기 공학과 석사  
 1989년~1994년 Rensselaer Polytechnic Institute 전기공학과 박사

관심분야 : 영상처리, 패턴인식, 컴퓨터비전, 딥러닝