Journal of the KIMST, Vol. 25, No. 4, pp. 381-389, 2022
ISSN 1598-9127(Print) · 2636-0640(Online)
DOI https://doi.org/10.9766/KIMST.2022.25.4.381

Research Paper 정보·통신 부문

2단계 부분 어텐션 네트워크를 이용한 가려짐에 강인한 군용 차량 검출

조 선 영*,1)

1) 국방과학연구소 국방인공지능기술센터

Occlusion Robust Military Vehicle Detection using Two-Stage Part Attention Networks

Sunyoung Cho*,1)

1) *Defense AI Technology Center, Agency for Defense Development, Korea*

(Received 4 January 2022 / Revised 7 March 2022 / Accepted 29 June 2022)

Abstract

Detecting partially occluded objects is difficult due to the appearances and shapes of occluders are highly variable. These variabilities lead to challenges of localizing accurate bounding box or classifying objects with visible object parts. To address these problems, we propose a two-stage part-based attention approach for robust object detection under partial occlusion. First, our part attention network(PAN) captures the important object parts and then it is used to generate weighted object features. Based on the weighted features, the re-weighted object features are produced by our reinforced PAN(RPAN). Experiments are performed on our collected military vehicle dataset and synthetic occlusion dataset. Our method outperforms the baselines and demonstrates the robustness of detecting objects under partial occlusion.

Key Words : Military Vehicle Detection(군용 차량 검출), Occlusion Robust(가려짐에 강인한), Part Attention Networks (부분 어텐션 네트워크)

1. 서 론

심층 합성곱 신경망(Deep Convolutional Neural Networks, DCNN)의 발전과 함께 객체 검출기의 성능이 크게 향상되고 있으나, 부분 가려짐을 포함하는 객체를 검출하는 것은 여전히 어려운 문제로 알려져 있다. 객체가 가려지는 영역의 크기나 패턴이 다양하고, 이로 인해 객체의 형태가 바뀌며 객체의 제한적인 일부 정보만을 활용하기 때문이다. Zhu et al.[1]는 DCNN이 사람에 비해 가려짐이 있는 객체를 인식하는 데 강인하지 못하다는 것을 보였으며, Fawzi et al.[2]는 작은 패치를 합성하여 만들어진 부분 가려짐에 DCNN의 인식 능력이 떨어진다는 것을 보였다.

* Corresponding author, E-mail: sycho22@add.re.kr

이러한 부분 가려짐을 포함하는 객체에 대한 검출 성능을 향상하기 위해 다양한 연구들이 시도되었다. 실제 데이터에 가려짐을 적용하여 이를 모델 학습에 활용하는 데이터 증강 기반 방법[3,4], 부분적인 객체의 정보를 활용하는 부분 기반 방법 등이 있다. 특히 부분 기반 방법에 대한 여러 연구들이 진행되어 왔는데, 객체의 부분 검출기 학습[5-7], 가려진 부분 또는 보이는 부분에 대한 검출[8-10], 객체의 부분에 대한 가려짐 유무 정보 추출[11,12] 등의 연구가 있다. 이러한 연구들은 객체의 부분 정보를 활용하여 가려짐에 강인하도록 하였지만, 객체의 각 부분마다 검출기를 학습해야 하거나 보이는 부분에 대한 어노테이션 정보가 필요한 등의 추가적인 계산이나 노력을 필요로 한다. 또한 각 부분에 대한 중요도를 반영하지 않기 때문에 검출 성능이 떨어질 수 있다.

본 논문에서는 부분 어텐션 네트워크를 이용한 가려짐에 강인한 객체 검출 방법을 제안한다. 객체 후보로 추출된 영역에 대한 CNN 특징을 바로 이용하여 객체에 대한 바운딩 박스나 클래스를 찾아내는 것은 가려짐으로 인해 발생하는 객체의 특성을 잘 반영하지 못한다. 제안하는 방법은 객체를 여러 부분들의 집합으로 보고, 각 부분에 대한 어텐션 가중치값을 계산 및 이를 가려짐에 강인한 특징을 추출하는데 사용한다. 2단계의 부분 어텐션 네트워크로 구성함으로써 부분 간의 관계를 고려하며 가려짐에 좀 더 강인한 특징을 추출하도록 하였다. 또한 어텐션 가중치값이 가장 큰 중요한 부분에 대한 특징의 활용성을 높이기 위해, 전체 객체의 특징에 대한 가중치값보다 더 커지도록 손실함수를 설계하였다. 본 논문에서는 이전 연구들처럼 객체의 각 부분마다 검출기를 학습하거나 보이는 부분에 대한 추가적인 어노테이션 정보를 활용하지 않고, 제안하는 2단계 부분 어텐션 네트워크를 통해 각 부분에 대한 중요도를 적응적으로 계산하여 부분 가려짐에 강인한 특징을 추출하여 성능을 향상시킨다.

본 논문에서는 군용 차량에 대한 데이터셋을 수집하여 제안하는 방법의 우수성을 입증한다. 실제적으로 군용 차량은 가려짐에 강인한 검출 방법의 성능을 입증하기에 좋은 데이터이다. 군용 차량의 경우 수풀이나 위장막 등에 의해 가려짐이 많이 발생하며, 일반적인 객체 검출에서 사용되는 객체들과는 달리 서로 비슷한 외형을 지니고 있기 때문에 가려짐이 발생할 경우 객체들 간에 구별이 더욱 어렵다는 특성이 있기 때문이다. 군용 차량에 대한 보안상의 이유로 잘 정리되고 공개된 데이터셋은 찾기가 힘들기 때문에, 본 논문에서는 실제 군용 차량에 대한 데이터 수집 및 바운딩 박스 레이블링 과정을 통해 직접 데이터셋을 구성하였다. 수집한 데이터셋은 다양한 가려짐을 포함하도록 하여 제안하는 방법의 성능을 입증할 수 있도록 하였다. 또한 수집한 데이터셋 중 가려짐이 없는 객체 이미지들에 대해 가려짐 패치를 합성한 데이터셋을 만들어 성능을 평가한다. 다양한 가려짐 정도와 종류를 포함하는 가려짐 합성 데이터셋을 만들어, 가려짐 특성에 따른 검출 성능을 확인하였다. 그 결과 제안하는 방법이 기존 객체 검출기에 비해 우수한 성능을 가지고 있음을 입증하였다.

2. 관련 연구

본 장에서는 다양한 객체 검출 연구들 중에서, 가려짐에 강인한 객체 검출 방법에 대한 관련 연구들을 살펴본다. 또한 어텐션 메커니즘에 대한 소개와 관련 연구들도 소개한다.

2.1 가려짐에 강인한 객체 검출

가려짐에 강인한 객체 검출을 위해 많은 연구들이 부분 기반 모델을 사용해왔다. 이는 부분적으로 객체가 가려질 경우 나머지 보이는 부분을 검출하거나 정보를 활용함으로써 전체 객체 검출 성능을 향상시키려는 것이다. 이러한 연구들의 공통된 방법 중 하나는 객체의 각 부분으로부터 부분 검출기를 학습하는 것이다[6]. 특히 Zhou et al.[5]는 부분 간 연관성을 활용하고 계산량을 줄이기 위해 여러 개의 부분 검출기들을 함께 학습하는 방법을 제안하였다. Tian et al.[7]은 다양한 가려짐 상황을 고려한 부분 프로토타입들로 구성된 부분 풀(Pool)을 구성하여 부분 검출기를 학습하였다.

가려짐 부분 또는 가려지지 않고 보이는 부분에 대한 정보를 활용하여 가려짐 문제를 해결하려는 시도도 있어왔다. Yan et al.[10]은 부분적으로 보이는 객체를 검출하기 위한 Boosted cascade 프레임워크를 제안하였다. Zhou et al.[9]는 보행자 검출 시 전체 보행자와 보행자의 보이는 부분에 대한 바운딩 박스를 각각 검출할 수 있는 네트워크를 제안하였으며, 두 검출 정보를 활용하여 가려진 보행자의 검출 성능을 향상시켰

다. Zhou et al.[8]는 학습 데이터셋에서 자주 나타나는 가려짐 패턴들을 Greedy 알고리즘을 통해 선택하는 방법을 제안하고, 이러한 패턴들을 활용하여 검출 성능을 향상하는 방법을 제안하였다.

객체의 각 부분에 대한 가려짐 유무를 판단하여 이를 통해 검출 성능을 향상하거나 가려짐에 강인한 특징을 추출하는데 활용하기도 한다. Wang et al.[11]은 보행자 템플릿을 여러 개의 블록들로 나누고, 각 블록에 대한 Visibility 상태 정보를 추정함으로써 가려진 영역을 추정한다. 이를 통해 가려짐이 있는 경우 부분 검출기를 적용하여 검출 성능을 향상하였다. Zhang et al.[12]은 보행자를 신체 부위를 기준으로 여러 부분으로 나누고 각 부분에 대해 Visibility 상태 정보를 추정하는 Occlusion process unit을 제안하였다. 이를 통해 추정된 정보는 기존의 특징들에 적용되어 가려짐에 강인한 특징을 추출하도록 하였다.

이외에도, Wang et al.[13]은 가려짐이 있는 상황에서도 객체의 시맨틱 부분들을 검출하기 위해 부분 기반 Voting 방법을 제안하였다. Chi et al.[14]는 보행자의 머리 부분에 대한 정보를 활용한 Mask-guided 모듈을 제안함으로써 가려짐 문제를 해결하고자 하였다.

2.2 어텐션 메커니즘

어텐션 메커니즘은 데이터의 중요한 부분에 집중하고 나머지 부분은 무시하는 인간의 인지 과정을 모방하는 것이라고 할 수 있다. 원래 자연어 처리 분야의 인코더-디코더 기반 기계 번역 시스템의 성능 향상을 가져오며 주목을 받았다. 그러나 현재는 기계 번역 뿐만 아니라 음성 처리 및 컴퓨터 비전 분야에서도 성공적으로 적용되고 있다. Mnih et al.[15]은 재귀 신경망(Recurrent Neural Network, RNN)에 기반한 비주얼 어텐션 모델을 제안하여 이미지 분류 문제에 적용하였고, CNN 모델보다 성능이 더 우수함을 보였다. Pang et al.[16]은 보행자의 보이는 영역에 대한 특징에 집중하도록 하는 Mask-guided 어텐션 네트워크를 제안하였고, 이는 가려짐에 강인한 보행자 검출에 효과적임을 보였다. 이 논문에서는 어텐션에 대한 Ground Truth (GT)를 생성하기 위해 보행자의 보이는 영역에 대한 바운딩 박스 정보가 필요하나, 이러한 어노테이션을 획득하는 것은 사실상 시간이 오래 걸리고 힘들다. 제안하는 방법은 이러한 추가적인 GT 없이 보이는 영역이 많은 부분에 대한 어텐션 가중치값이 더 커지도록 손실함수를 설계함으로써 검출 성능을 향상한다.

3. 제안하는 방법

본 장에서는 먼저 군용 차량을 검출하기 위해 사용되는 기본 객체 검출기 및 이에 제안하는 방법이 어떻게 적용되는지 소개한다. 다음으로 가려짐에 강인한 특징을 추출하기 위한 부분 기반 어텐션 네트워크에 대해 설명한다. 마지막으로, 제안하는 방법에서 사용하는 손실함수에 대해 설명한다.

3.1 개요

본 논문에서는 기본 객체 검출기로 Faster R-CNN[17]을 사용하며, 기본적인 알고리즘 동작은 다음과 같다. 먼저, 모델에 입력된 이미지에 대해 미리 학습된 ResNet-101[18]과 같은 CNN을 이용하여 특징을 추출하고, 이를 기반으로 영역 제안 네트워크(Region Proposal Network, RPN)를 통해 객체 후보 영역들을 검출한다. 검출된 객체 후보 영역들의 특징들은 특징 맵의 대응되는 관심영역(Region-of-Interest, RoI)으로부터 추출하며, RoI Pooling layer를 통해 고정된 크기의 특징으로 만든다. 이러한 특징들을 이용하여 각 객체 후보 영역에 대한 클래스와 바운딩 박스 좌표를 계산한다. 본 논문에서는 RoI Pooling layer를 통해 추출되는 특징에 대해 부분 기반 어텐션 네트워크를 이용하여 부분 가려짐에 강인한 특징을 생성하는 방법을 제안한다.

3.2 객체의 부분 분할 방법

RPN을 통해 추출된 각 객체의 후보 영역에 대해 부분 영역으로 분할하는 방법은 다양하다. 기존 부분 기반 객체 검출 연구들은 주로 보행자 검출에 관한 연구가 대부분인데, 이들은 일정한 크기의 그리드로 나누거나[5,7,11] 보행자의 신체 부위 특성을 고려하여 경험적으로 비율을 계산[12,19]하여 나누었다. 본 논문에서는 Fig. 1에서 볼 수 있듯이 군용 차량을 5개의 부

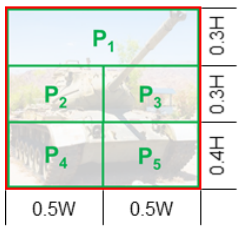

Fig. 1. We divide the military vehicle object region into five parts

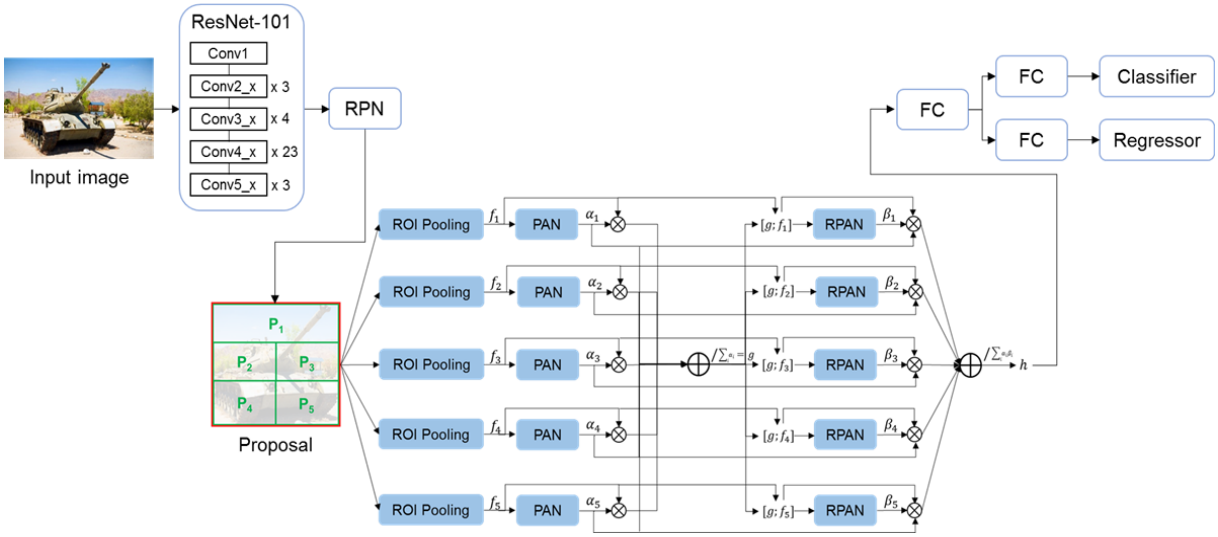

Fig. 2. The overall network architecture of our part-based attention network. For each proposal, we divide it into 5 parts (P_1, ⋯, P_5) and extract the features (f_1, ⋯, f_5) using RoI Pooling. Each feature is fed into PAN to predict the attention weights (α_1, ⋯, α_5), and the new object feature g is generated by applying the element-wise summation of re-weighted features f_i with α_i. After combining g and each re-weighted part feature, we again feed the features into RPAN to predict the attention weights (β_1, ⋯, β_5) and generate the final object feature h.

분(P_1, P_2, P_3, P_4, P_5)으로 분할하였다. 군용 차량을 촬영한 각도에 따라 다르겠지만, 대략적으로 P_1에는 상부나 포신, P_2와 P_3에는 선체, P_4와 P_5에는 궤도 바퀴 부분을 포함할 확률이 높게 된다. 부분을 다양한 개수로 분할하고 이에 따른 객체 검출 성능을 평가한 결과는 4.5.1절에 나타내었다.

3.3 부분 기반 어텐션 네트워크

부분 기반 어텐션 네트워크는 RPN을 통해 추출된 각 객체 후보 영역을 여러 개의 부분으로 나누고, 각 부분에 대해 2단계로 특징을 추출한다(Fig. 2 참고). 먼저, RoI Pooling layer를 통해 추출된 부분 특징 f_i는 부분 어텐션 네트워크(Part Attention Network, PAN)에 입력된다. PAN은 2개의 Fully Connected Layer (FCL)와 Sigmoid 함수로 구성되어 있으며, 각 부분에 대한 어텐션 가중치 값 α_i를 계산하게 된다. 이러한 부분 특징과 어텐션 가중치 값을 통해, 객체 후보 영역에 대한 새로운 특징 g를 아래의 식 (1)과 같이 생성한다:

$$g = \frac{1}{\sum_i \alpha_i} \sum_i \alpha_i f_i \tag{1}$$

새로운 객체 특징 g는 각 부분의 특징 f_i와 결합(Concatenation)하여 강화된 부분 어텐션 네트워크(Reinforced Part Attention Network, RPAN)에 입력된다. RPAN도 역시 2개의 FCL와 Sigmoid 함수로 구성되어 있으며, 각 부분에 대한 어텐션 가중치 값 β_i를 계산하게 된다. 최종적으로 객체에 대한 특징 h는 아래의 식 (2)와 같다:

$$h = \frac{1}{\sum_i \alpha_i \beta_i} \sum_i \alpha_i \beta_i [g; f_i] \tag{2}$$

위 식에서, $[g;f_i]$는 특징 g와 f_i 간 결합을 의미한다. 특징 h는 객체의 클래스와 바운딩 박스 좌표를 계산하는 데 사용되며, 기존의 Fast R-CNN 모델[20]과 동작 방법은 같다. 특징 h가 FCL에 입력되어 출력된 특징이 객체 분류기와 바운딩 박스 Regressor에 입력되어

각각 클래스와 바운딩 박스 좌표에 대한 값을 계산하게 된다.

3.4 손실 함수

본 논문에서는 기본 객체 검출기의 RoI Pooling layer 다음에 제안하는 부분 기반 어텐션 네트워크를 추가함으로써 가려짐에 강인한 특징을 추출하고, 이를 객체의 클래스와 바운딩 박스 좌표를 계산하는데 사용한다. 따라서 손실 함수는 아래와 같이 부분 기반 어텐션 네트워크에 의한 손실 값이 추가되어 아래의 식 (3)과 같다:

$$L = L_0 + \gamma L_{AN} \quad (3)$$

위 식에서 L_0는 Faster R-CNN에서 사용하는 손실 값으로, 아래의 식 (4)와 같이 RPN과 Fast R-CNN[20]에 대한 손실값으로 구성된다.

$$L_0 = L_{RPN} + L_{Fast\,RCNN} \quad (4)$$

γ는 실험을 통해 1을 사용하였으며, L_{AN}은 아래의 식 (5)와 같이 제안하는 부분 기반 어텐션 네트워크를 위한 손실 값이다.

$$L_{AN} = \max\{0, \delta - (\alpha_{\max} - \alpha_0)\} \quad (5)$$

위 식에서 $\alpha_{\max}$는 PAN을 통해 출력된 각 부분에 대한 어텐션 가중치 값 중에서 최대값이며, α_0는 객체 후보 영역 전체에 대한 RoI Pooling layer를 통해 추출된 특징에 대한 어텐션 가중치 값이다. 각 가중치 값은 0과 1사이의 값을 갖는다. δ는 $\alpha_{\max}$와 α_0 간 margin 값을 나타내는 하이퍼 파라미터로써, 실험을 통해 0.02를 사용하였다. 이는 특정 객체 부분에 대한 어텐션 가중치 값은 전체 객체에 대한 가중치 값에 비해 margin을 가지고 더 커서 구별 가능한 특징을 갖고 있게 만든다.

4. 실험 결과 및 분석

4.1 데이터셋

4.1.1 실제 군용 차량 데이터셋

실제 군용 차량에 대한 데이터셋을 수집하여 제안하는 방법을 평가하였다. 차량의 지붕 및 드론에 카메라를 부착하여 1920×1080 해상도의 동영상을 촬영하였고, 총 7개 종류의 군용 차량에 대한 데이터를 수집하였다.(각 클래스는 보안상의 이유로 A, B, C, D, E, F, G로 표기한다.) 데이터 수집 시, 각 클래스 별로 수풀, 건물, 위장막, 발연기 등의 다양한 가려짐이 있는 상황을 연출하여 촬영하였다. 촬영된 동영상으로부터 프레임을 추출하여 총 10,345개의 이미지를 포함하는 데이터셋을 구성하였고, 8,239개의 학습용 및 2,106개의 테스트용 이미지로 나누어 실험하였다. 데이터셋 중 60 %의 데이터에 부분 가려짐이 있는 객체를 포함하고 있으며, 각 데이터의 가려짐의 정도는 다양하다. 데이터셋의 각 이미지들은 군용 차량에 대한 바운딩 박스와 클래스 정보를 포함하는 레이블링이 수행되었다.

4.1.2 가려짐 합성 데이터셋

다양한 가려짐 데이터에서의 성능 평가를 위해, 4.1.1 절에서 수집한 실제 군용 차량 데이터셋 중 가려짐이 없는 946개의 테스트용 데이터를 이용하여 가려짐을 합성한 데이터를 생성하였다. 3가지의 가려짐 레벨(L1, L2, L3) 및 3종류의 가려짐 패치에 따라 총 9가지의 가려짐 합성 데이터를 생성하였다. 가려짐 레벨 L1은 객체의 바운딩 박스 영역의 0-10 %, L2는

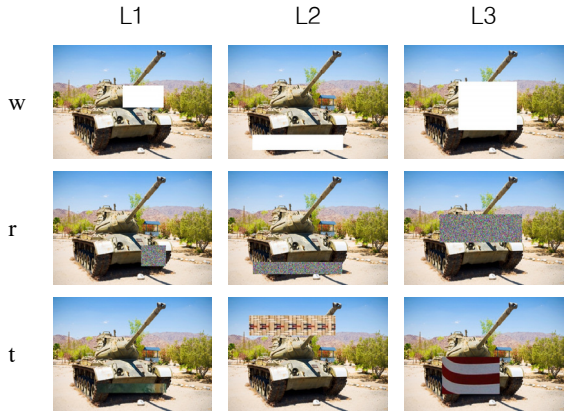

Fig. 3. Example of images in our occlusion synthesis dataset. Each row shows samples of different types of occlusion patches. *w*, *r*, and *t* indicate white, random noise, and texture box patches. Each column shows samples with increasing amount of partial occlusion: 0–10 % (L1), 10–20 % (L2), 20–30 % (L3)

10-20 %, L3는 20-30 %의 가려짐을 갖는다. 가려짐 종류는 흰색, 랜덤 노이즈, 텍스쳐 사각형 패치가 있다. 텍스쳐 패치는 다양한 텍스쳐 이미지들을 포함하는 데이터셋[21]으로부터 패치의 크기만큼 잘라와서 사용하였다. 가려짐 패치는 가려짐 레벨 내 무작위로 선택된 크기를 갖도록 하여, 객체의 바운딩 박스 영역 내 무작위로 선택된 위치에 합성하였다. Fig. 3은 임의의 탱크 이미지에 대해 가려짐 합성 데이터를 생성한 예이며, 가려짐 합성 데이터셋은 실제 수집한 군용 차량을 이용하여 구성되었다.

4.2 학습 환경 및 평가 방법

객체 검출 프레임워크는 ResNet-101[18]에 기반한 Faster R-CNN 모델[17]을 사용하였고, ImageNet 분류 모델을 이용하여 초기화되었다. 모델 학습은 30 K iteration 동안 학습율 0.01의 SGD를 적용하였다. Fig. 4는 모델 학습 시 측정한 학습 및 검증 손실값에 대한 그래프를 보여준다. 모델의 성능을 평가하기 위해서는 평균 평균 정밀도(mean Average Precision, mAP)@[.5:.95]를 사용하였다.

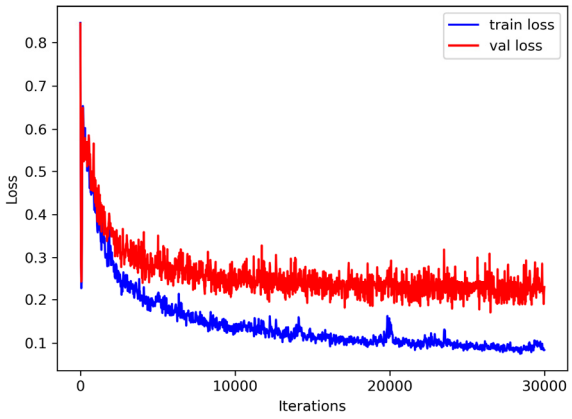

Fig. 4. Training and validation loss

4.3 가려짐 데이터셋에서의 검출 성능

4.1.1절에서 언급한 실제 수집한 가려짐을 포함하는 데이터셋에 대해, Faster R-CNN과 제안하는 방법 간 검출 성능을 Table 1에 나타내었다. 평균적으로 제안하는 방법이 Faster R-CNN에 비해 2.91 % 높은 것을 확인할 수 있다. 클래스 별로 보면 7개 클래스 중에서

Table 1. Detection results on our military vehicle dataset

Method \ Class	A	B	C	D	E	F	G	mAP (%)
Faster R-CNN	**78.92**	65.54	65.75	72.75	**67.17**	65.79	74.54	70.07
Proposed	78.54	**69.12**	**75.67**	**74.80**	67.11	**68.09**	**77.53**	**72.98**

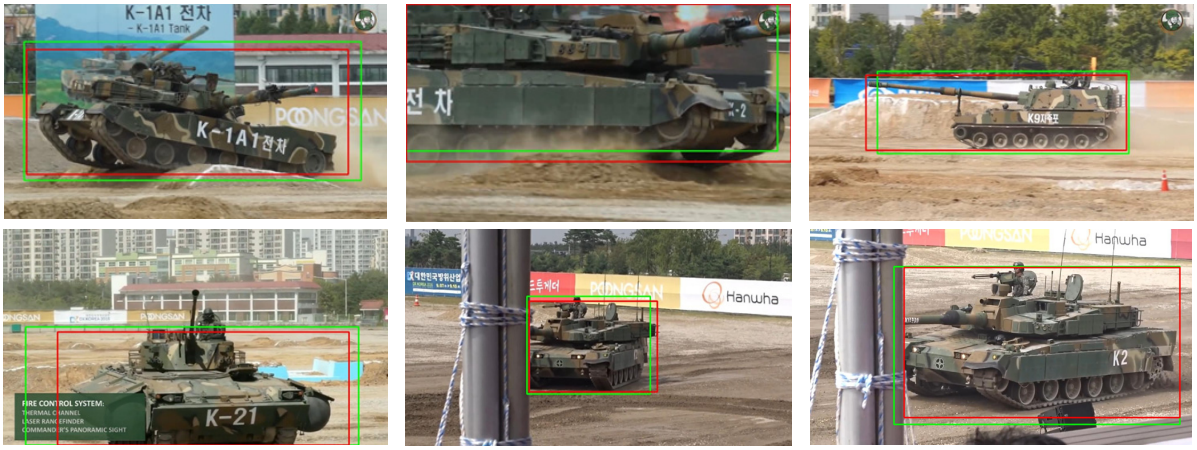

Fig. 5. Selected examples of localization results on images from YouTube. Green box: Faster R-CNN, red box: Proposed method

5개 클래스에 대해 제안하는 방법의 성능이 더 좋았으며, 나머지 2개 클래스 역시 성능 차이가 크지 않았다. 보안상의 이유로 수집한 데이터셋에 대한 검출 결과를 보여주는 것은 어렵기 때문에, YouTube에서 다운로드 받은 군용 차량 관련 동영상으로부터 객체에 가려짐이 있는 프레임들을 일부 추출하여 검출한 결과를 Fig. 5에 나타내었다. 클래스 이름은 공개하기 어렵기 때문에 군용 차량에 대한 바운딩 박스 지역화 성능만을 확인해보면, 제안하는 방법이 Faster R-CNN에 비해 가려짐이 있는 상황에서 객체 영역을 더 잘 포함하여 지역화하는 것을 확인할 수 있다.

4.4 가려짐 합성 데이터셋에서의 검출 성능

4.1.2절의 가려짐 합성 테스트 데이터셋을 이용한 검출 결과를 Table 2에 나타내었다. 3가지의 가려짐 레벨에 대해 각 가려짐 종류 별로 Faster R-CNN과 제안하는 방법 간 검출 성능을 나타내었다. 각 레벨에 대해 가려짐 종류들의 평균 검출 성능을 계산한 결과, 제안하는 방법이 L1, L2, L3에 대해서 각각 1.11 %, 1.69 %, 0.72 %의 검출 성능 향상을 보였다. 각 가려짐 레벨 및 종류 별로 보면 L3의 텍스쳐 가려짐 패치를 제외한 모든 가려짐 데이터에 대해 제안하는 방법이 더 높은 검출 성능을 획득하였다.

Table 2. Detection results on our occlusion synthesis dataset, measured by mAP (%).

		Faster R-CNN	Proposed
L1	w	71.92	**72.75**
	r	77.32	**78.37**
	t	75.97	**77.43**
	Avg	75.07	**76.18**
L2	w	43.38	**44.70**
	r	59.28	**61.69**
	t	54.12	**55.47**
	Avg	52.26	**53.95**
L3	w	19.57	**20.00**
	r	37.03	**40.16**
	t	**29.21**	27.79
	Avg	28.60	**29.32**

4.5 Ablation study

4.5.1 객체의 부분 분할 방법 평가

RPN을 통해 추출된 객체 후보 영역을 여러 개의 부분으로 나눌 때, 본 논문에서는 군용 차량의 부분 특성을 고려하여 5개 부분으로 나누었다. 이러한 분할 방법의 효과를 입증하기 위해, 다양한 개수(2, 3, 4, 6)로 나누어 제안하는 방법의 성능을 비교하였다. 각 분할 방법 별 비율은 Fig. 6과 같으며, 이를 기반으로 한 검출 성능은 Table 3에 나타내었다. 전반적으로 부분 분할 개수에 상관없이 제안하는 방법이 Faster R-CNN의 검출 성능(70.07 %)보다 더 높은 것을 확인할 수 있다. 또한 다양한 부분 분할 방법들 중에서는 제안하는 5개 부분으로 객체 영역을 분할하는 것이 가장 높은 성능을 획득하였다. 이러한 실험 결과로부터, 객체 영역을 분할하는 부분의 개수 및 비율 선택이 검출 성능을 결정하는 중요한 요소가 된다는 것을 알 수 있다.

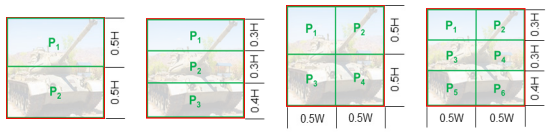

Fig. 6. Partition schemes for region proposals

4.5.2 부분 기반 어텐션 네트워크의 단계 별 추출되는 특징의 성능 평가

본 논문에서는 2단계의 부분 기반 어텐션 네트워크를 이용하여 가려짐에 강인한 특징을 추출하였다. PAN을 통해 객체의 각 부분에 어텐션 가중치 값을 반영한 특징 g를 추출하고, 이를 다시 각 부분 특징과 결합하여 RPAN을 통해 어텐션 가중치 값을 재계산 및 이를 활용하여 새로운 객체에 대한 특징 h를 추출하였다. 이렇게 2단계로 특징을 추출하는 것이 효과적임을 보이기 위해, Table 4에 PAN을 통해 추출한 특징 g를 사용한 검출 성능과 RPAN을 통해 추출한 특징 h를 사용한 검출 성능을 비교하였다. 그 결과 PAN을 통해 추출한 특징만을 이용한 검출 결과는 RPAN을 통해 추출한 특징을 이용한 검출 결과에 비해 3.84 % mAP의 성능이 떨어짐을 확인하였다. 이러한 실험 결과로부터 2단계로 특징을 추출함으로써 부분 간 관계를 고려한 특징 추출을 통해 어텐션 가중치 값의 정확도를 높이고 성능을 향상시키는 것을 알 수 있다.

Table 3. Performance comparison of different partition schemes of object proposals

# of partition	mAP (%)
2	71.99
3	71.85
4	71.74
5 (Proposed)	**72.98**
6	71.33

Table 4. Comparison of detection results between two features from PAN and RPAN

Feature	mAP (%)
g from PAN	69.14
h from RPAN	**72.98**

4.6 검출 시간 평가

제안하는 방법은 기존의 Faster R-CNN과 비교하여, 객체의 부분에 대한 어텐션 가중치 값을 계산하는 2단계 어텐션 네트워크를 추가적으로 사용하기 때문에 검출 시간에 대한 평가를 수행하였다. Nvidia GeForce RTX 2080 Ti GPU 및 3.3 GHz Intel(R) Xeon(R) Gold 6234 CPU 환경에서 추론 시간을 평가하였다. 1920×1080 이미지에 대해 평가한 결과, Faster R-CNN은 0.19초, 제안하는 방법은 0.24초가 걸려서 평균적으로 0.05초가 더 걸렸다. 현재는 코드 최적화가 안된 상태에서 측정한 추론 시간이기는 하지만, 실시간 검출은 어려운 상황이다. 그러나 현재 사용된 Backbone 네트워크인 ResNet-101 보다 경량화된 네트워크 사용 및 입력 이미지의 해상도를 줄이고 코드 최적화를 수행한다면 추론 속도를 향상시킬 수 있을 것이다.

5. 결 론

본 논문에서는 부분적으로 가려진 객체의 검출에 강인한 2단계 부분 어텐션 네트워크를 제안하였다. 제안하는 방법은 객체를 여러 부분들로 나눈 뒤에, 각 부분에 대한 어텐션 가중치 값을 계산하고 이를 통해 가려짐에 강인한 특징을 추출한다. 추출된 특징에 대해 부분 어텐션 네트워크를 한번 더 적용함으로써 가려짐에 더욱 강인한 특징을 추출하도록 하였다. 실제 수집한 군용 차량 데이터셋 및 이에 기반한 다양한 가려짐 종류 및 레벨로 구성된 가려짐 합성 데이터셋에 대해, 제안하는 방법의 성능을 평가하였다. 그 결과 제안하는 방법은 기존 객체 검출기보다 가려짐에 강인하며 더 우수한 성능을 가지고 있음을 입증하였다.

References

[1] H. Zhu, P. Tang, J. Park, S. Park, A. Yuille, “Robustness of Object Recognition Under Extreme Occlusion in Humans and Computational Models,” CoRR, Vol. abs/1905.04598, 2019.

[2] A. Fawzi, P. Frossard, “Measuring the Effect of Nuisance Variables on Classifiers,” BMVC, 2016.

[3] T. DeVries, G. W. Taylor, “Improved Regularization of Convolutional Neural Networks with Cutout,” arXiv preprint arXiv:1708.04552, 2017.

[4] S. Yun, D. Han, S. J. Oh, S. Chun, Y. Yoo, “Cutmix: Regularization Strategy to Train Strong Classifiers with Localizable Features,” ICCV, pp. 6023-6032, 2019.

[5] C. Zhou, J. Yuan, “Multi-Label Learning of Part Detectors for Heavily Occluded Pedestrian Detection,” ICCV, pp. 3486-3495, 2017.

[6] C. Zhou, J. Yuan, “Non-Rectangular Part Discovery for Object Detection,” BMCV, 2014.

[7] Y. Tian, P. Luo, X. Wang, X. Tang, “Deep Learning Strong Parts for Pedestrian Detection,” ICCV, pp. 1904-1912, 2015.

[8] C. Zhou, J. Yuan, “Occlusion Pattern Discovery for Object Detection and Occlusion Reasoning,” TCSVT, Vol. 30, No. 7, pp. 2067-2080, 2020.

[9] C. Zhou, J. Yuan, “Bi-Box Regression for Pedestrian Detection and Occlusion Estimation,” ECCV, pp. 135-151, 2018.

[10] S. Yan, Q. Liu, “Inferring Occluded Features for Fast Object Detection,” Signal Processing, Vol. 110, pp. 188-198, 2015.

[11] X. Wang, T. X. Han, S. Yan, “An HOG-LBP Human Detector with Partial Occlusion Handling,”

ICCV, pp. 32-39, 2009.

[12] S. Zhang, L. Wen, X. Bian, Z. Lei, S. Z. Li, “Occlusion-Aware R-CNN: Detecting Pedestrians in a Crowd,” ECCV, pp. 657-674, 2018.

[13] J. Wang, L. Xie, A.L. Yuille, Z. Zhang, C. Xie, “Deepvoting: A Robust and Explainable Deep Network for Semantic Part Detection Under Partial Occlusion,” CVPR, pp. 1372-1380, 2018.

[14] C. Chi, S. Zhang, J. Xing, Z. Lei, S. Z. Li, X. Zou, “PedHunter: Occlusion Robust Pedestrian Detector in Crowded Scenes,” AAAI, pp. 10639-10646, 2020.

[15] V. Mnih, N. Heess, A. Graves, K. Kavukcuoglu, “Recurrent Models of Visual Attention,” NIPS, pp. 2204-2212, 2014.

[16] Y. Pang, J. Xie, M. H. Khan, R. M. Anwer, F. S. Khan, L. Shao, “Mask-Guided Attention Network for Occluded Pedestrian Detection,” ICCV, pp. 4967-4975, 2018.

[17] S. Ren, K. He, R. Girshick, J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” NIPS, pp. 91-99, 2015.

[18] K. He, X. Zhang, S. Ren, J. Sun, “Deep Residual Learning for Image Recognition,” CVPR, pp. 770-778, 2016.

[19] P. F. Felzenszwalb, R. B. Girshick, D. A. McAllester, D. Ramanan, “Object Detection with Discriminatively Trained Part-based Models,” TPAMI, Vol. 32, No. 9, pp. 1627-1645, 2010.

[20] R. Girshick, “Fast R-CNN,” ICCV, pp. 1440-1448, 2015.

[21] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, A. Vedaldi, “Describing Textures in the Wild,” CVPR, pp. 3606-3613, 2016.