위성 SAR 영상의 지상차량 표적 데이터 셋 및 탐지와 객체분할로의 적용

박지훈^{*,1)} · 최여름¹⁾ · 채대영¹⁾ · 임 호¹⁾ · 유지희¹⁾

¹⁾ 국방과학연구소 국방인공지능기술센터

A Dataset of Ground Vehicle Targets from Satellite SAR Images and Its Application to Detection and Instance Segmentation

Ji-Hoon Park^{*,1)} · Yeo-Reum Choi¹⁾ · Dae-Young Chae¹⁾ · Ho Lim¹⁾ · Ji Hee Yoo¹⁾

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

(Received 4 August 2021 / Revised 4 November 2021 / Accepted 13 December 2021)

Abstract

The advent of deep learning-based algorithms has facilitated researches on target detection from synthetic aperture radar(SAR) imagery. While most of them concentrate on detection tasks for ships with open SAR ship datasets and for aircraft from SAR scenes of airports, there is relatively scarce researches on the detection of SAR ground vehicle targets where several adverse factors such as high false alarm rates, low signal-to-clutter ratios, and multiple targets in close proximity are predicted to degrade the performances. In this paper, a dataset of ground vehicle targets acquired from TerraSAR-X(TSX) satellite SAR images is presented. Then, both detection and instance segmentation are simultaneously carried out on this dataset based on the deep learning-based Mask R-CNN. Finally, this paper shows the future research directions to further improve the performances of detecting the SAR ground vehicle targets.

Key Words : Synthetic Aperture Radar(합성 개구면 레이더), Target Detection(표적 탐지), Instance Segmentation(객체 분할), Mask R-CNN(영역 및 객체분할용 합성곱 신경망), Deep Learning(딥러닝)

1. 서 론

전자파에 의한 물체의 전자기 산란현상에 의해 형

성되는 합성 개구면 레이더(Synthetic Aperture Radar, SAR) 영상은 광학영상과는 달리 날씨와 밤낮 관계없이 획득될 수 있다는 장점으로 인하여 군사 및 민간 분야에서 감시정찰의 목적으로 널리 활용되고 있다. 점진적으로 운용되는 SAR 센서의 수가 증가함에 따라서 사람에 의한 영상 판독업무의 경감을 위해 SAR 영상

* Corresponding author, E-mail: dydynoel@add.re.kr
Copyright © The Korea Institute of Military Science and Technology

으로부터의 표적 자동탐지(Automatic Target Detection, ATD)^[1] 및 표적 자동식별(Automatic Target Recognition, ATR)^[2]에 관한 연구가 활발히 진행되고 있다. 이중 표적 자동탐지의 경우, 전통적으로는 SAR 영상의 픽셀 값 및 통계적 분포를 활용하는 CFAR(Constant False Alarm Rate) 기반의 기법이 주류를 이루었다^[3,4]. 그러나 최근 몇 년간 R-CNN(Region with Convolutional Neural Network)^[5]을 필두로 Faster R-CNN^[6], YOLO(You Only Look Once)^[7], SSD(Single-Shot Detector)^[8], FPN(Feature Pyramid Network)^[9] 등과 같은 수많은 딥러닝(deep learning) 알고리즘의 발전에 힘입어, SAR 표적 자동탐지 분야에서도 유의미한 성과를 도출하였다. 이러한 연구사례들은 주로 해상을 촬영한 SAR 영상에서의 선박 탐지^[10-13]와 공항을 주로 촬영한 SAR 영상에서의 항공기 탐지^[14,15]에 집중되어 있으며, 비교적 데이터 획득이 용이한 상용 위성 SAR 영상을 주로 활용한다. 특히 선박 탐지의 경우 TerraSAR-X, RadarSat-2, Sentinel-1, Gaofen-3과 같은 상용 위성 SAR 영상에서 획득된 LS-SSDD^[16], AIR-SARShip^[17], HRSID^[18]와 같은 대용량 데이터 셋이 관련 연구자들을 위해 공개된 바 있다.

이렇듯 딥러닝 알고리즘은 SAR 표적영상 탐지 연구를 촉진시켜왔으나, 상대적으로 지상차량 표적에 대한 연구사례는 적다. 선박이나 항공기에 비해 지상차량 표적을 SAR 영상에서 탐지하는 문제는 다음과 같은 어려움이 예상된다.

- ① 지상의 차량들이 위치한 지면특성의 변동이 선박이나 항공기에 비해 크다. 즉, 동일한 차량이라도 주차장소가 모래인지 시멘트인지, 혹은 평지인지 굴곡이 심한 곳인지의 여부에 따라 SAR 영상에 나타나는 신호 특성이 달라질 수 있다. 일부는 신호 대 클러터 비(Signal-to-Clutter Ratio, SCR)가 매우 낮아서 표적탐지 실패의 원인이 된다.
- ② 선박은 근해나 원해와 무관하게 대부분 해상에 존재하고, 항공기는 공항에 주로 존재하면서 주변 다른 구조물과 달리 정형적인 형태를 갖는다. 반면 지상의 주변 환경은 변동성이 커서 지상차량과 유사한 구조물이 다수 존재할 수 있으므로, 표적 자동탐지 시 상대적으로 오경보(false alarm)가 다수 발생할 수 있다.
- ③ SAR 영상 내의 지상차량들은 밀집하여 주차되어 있는 경우가 많다. 이러한 밀집표적의 문제는 선박

탐지 문제에서도 성능을 하락시키는 요인이 된다. 특히 간격 수십 cm 이내로 밀집된 지상차량의 경우 다중 표적 간 전자파 산란 신호의 간섭 등으로 인해 SAR 영상의 표적 특성을 명확히 특정하지 못할 수 있다.

- ④ 선박은 해상에 위치한다는 특성과 함께 항해의 투명성 확보를 위한 AIS(선박 자동인식 시스템)가 구축되어 있어서 딥러닝 알고리즘 학습을 위해 관심 표적에 별도의 표식을 하는 표지(labeling, annotation)작업이 비교적 수월하다. 항공기의 경우 공항 내에 위치할 수 있는 지점이 한정적이라는 특징이 있다. 반면 지상차량은 임의성이 높아서 정확한 표지작업이 어렵다.

이에 본 논문에서는 위성 SAR 영상의 지상차량 표적 데이터 셋의 구성과정을 제시하고, 이로부터 표적 탐지와 객체 분할을 동시에 수행할 수 있는 Mask R-CNN^[19]을 활용하여 SAR 영상에서의 지상차량 표적 탐지 성능을 산출한다. Mask R-CNN은 탐지결과를 통해 표적종류 및 위치를 특정할 수 있을 뿐만 아니라, 이전 버전인 Faster R-CNN과는 달리 표적 별 객체 분할 결과를 제공하므로 위치뿐만 아니라 윤곽정보, 크기, 방향과 같은 보다 유용한 정보를 제공할 수 있다는 점을 고려하여 차용되었다. 이 때 데이터 셋은 4개 지역에서 TerraSAR-X(TSX)위성으로부터 획득한 18장의 SAR 영상을 이용하여 딥러닝 기반 알고리즘의 학습이 가능하도록 재구성되었다. ‘지상차량 표적’의 의미는 다양하게 해석될 수 있으나, 본 논문에서는 ‘길이 5 m 이상의 트럭/버스/공사용 중장비 등 대형차량’을 의미한다. 이는 승용차와 같은 소형 차량은 해상도가 비교적 낮고, 스펙클(speckle)의 영향을 받는 위성 SAR 영상에서는 표적특성을 명확히 특정하기 어렵기 때문이다. 본 논문의 2장에서는 전술한 위성 SAR 지상차량 표적 데이터셋에 대해 상세히 기술한다. 3장에서는 표적탐지 및 객체분할을 위한 딥러닝 알고리즘 중 하나인 Mask R-CNN에 대해 기술한다. 4장에서는 이러한 Mask R-CNN을 데이터 셋에 적용하여 표적 탐지와 객체 분할에 적용한 사례를 보이고 평균 정밀도(Average Precision, AP) 성능을 산출한다. 5장에서는 논문의 결론을 맺음과 함께 성능 향상을 위한 추후 연구 방향에 대해 기술한다.

2. 위성 SAR 지상차량 표적 데이터 셋

2.1 SAR 영상 기본정보

지상차량 표적 데이터 셋을 구축하기 위해 이용된 TSX 위성 SAR 영상은 빔의 방위(azimuth)방향 조향을 최대로 집중시킨 ST(Staring spotlight)모드에서 획득되었으며, 송신신호의 대역폭은 300 MHz으로서 경사거리(slant-range) 및 방위방향 해상도는 각각 약 0.59 m 및 0.23 m이다. 주파수는 X대역의 9.65 GHz를 사용하며, HH 편파이다. SAR 영상의 처리레벨은 SSC(Single-look Slant-range Complex, Level1-A)로서, 경사거리와 방위방향 픽셀간격은 각각 0.455 m와 0.167 m이다. 위성 SAR 영상의 가독성을 높이기 위하여 다음의 작업을 수행하였다.

- ① 경사거리 영상을 지면거리(ground range) 영상으로 변환한다. 거리방향 길이가 수 km에 달하는 위성 SAR 영상의 특성상, 근거리와 원거리에서의 SAR 신호 입사각도(내림각)에는 다소 차이가 발생한다. 따라서 지면거리에서의 근거리 및 원거리 픽셀간격은 미세한 차이가 발생한다.
- ② 변환은 삼각함수 관계에 의해 수행되며, 간격이 각기 다른 지면거리 방향의 픽셀간격과 방위방향의 픽셀간격을 등간격으로 맞추기 위해 이중선형 보간법(bilinear interpolation)을 이용한다. 이러한 과정에서 신호의 평탄화(smoothing)가 발생하여 공간상의 해상도는 다소 저하되지만 스펙클 잡음이 개선되고, 방사(radiometric) 해상도는 높아지는 효과가 나타난다.
- ③ 등간격 값은 0.2 m로 설정하였으며, 이는 거리방향 해상도는 물론 방위방향 해상도보다도 작은 값이다.

Table 1에는 데이터 셋을 구축하기 위해 사용된 TSX 위성 SAR 영상의 정보를 나타내었다. Table 1의 정보는 다음과 같이 정리할 수 있다.

- ① ‘ID’의 경우 지역 코드와 촬영날짜가 ‘yymmdd’ 형식으로 표기되었다. JC, SJ, SW는 각각 체첸, 세종, 수원을 의미하며 A는 국외지역이다.
- ② ‘내림각’의 경우 SAR 신호의 LOS(Line-Of-Sight)와 SAR 영상의 중심에 있는 지면이 이루는 각도를 의미하며 ‘depression’ 또는 ‘grazing’ 각도로도 불린다.

Table 1. Information on TSX SAR images

ID	내림각	길이(km)	해상도(m)	패치/표적수
JC200614	53.93°	6.13/3.05	0.99/0.23	8 / 165
JC200830	53.93°	6.13/3.05	0.99/0.23	11 / 152
JC200926	40.65°	5.12/3.20	0.78/0.23	14 / 188
JC201024	53.93°	6.13/3.05	0.99/0.23	11 / 163
JC201104	53.93°	6.13/3.05	0.99/0.23	7 / 138
JC201109	40.65°	5.12/3.20	0.78/0.23	9 / 132
A151009	53.62°	6.09/3.05	0.99/0.23	9 / 193
A191207	50.08°	5.59/3.05	0.92/0.23	6 / 47
SJ191209	35.45°	5.89/3.53	0.72/0.23	8 / 127
SJ191210	62.60°	6.87/2.86	1.28/0.23	3 / 62
SJ200430	35.45°	5.89/3.53	0.72/0.23	12 / 130
SJ201103	35.45°	5.89/3.53	0.72/0.23	12 / 120
SJ201103	35.45°	5.89/3.53	0.72/0.23	13 / 120
SJ201114	35.45°	5.89/3.53	0.72/0.23	10 / 120
SW141010	47.48°	5.40/3.05	0.87/0.23	26 / 247
SW141226	47.47°	5.40/3.05	0.87/0.23	17 / 198
SW150106	47.52°	5.38/3.05	0.87/0.23	27 / 357
SW150128	47.46°	5.40/3.05	0.87/0.23	19 / 273

- ③ ‘길이(km)’와 ‘해상도(m)’는 각각 지면거리와 방위 방향에 대한 값을 나타내었으며, 모든 픽셀이 0.2 m 간격을 갖도록 처리되었으므로 영상을 구성하는 픽셀의 전체 개수를 추정할 수 있다. 경사거리 해상도는 전술한 바와 같이 0.59 m로 고정되지만 지면거리 해상도는 이의 $1/\cos(\text{내림각})$ 배에 해당한다.
- ④ 위성 SAR 영상의 모든 지점에 지상차량 표적이 존재하지는 않는다. ‘패치/표적수’의 경우 SAR 영상에서 표적이 있는 곳을 선별하여 800×800 크기의 패치(patch)를 추출한 개수 및 각 영상 별 추출된 모든 패치에 존재하는 표적의 개수를 의미한다. 이 때 패치마다 5개 내외의 표적이 있을 수도 있고 수십 개의 표적이 밀집하여 위치하고 있을 수도 있다. HRSID와 같은 공개 데이터 셋도 큰 SAR 영상으로부터 표적이 존재하는 일부를 패치로 추출한 형태이다.
- ⑤ Table 1에서 흰색 바탕으로 표시된 ID의 SAR 영상은 후술할 Mask R-CNN을 학습시키는데 활용되고, 연두색 바탕으로 표시된 SAR 영상은 성능을 확인하기 위한 시험영상으로 활용된다. 이 때 체첸, 세

중, 수원의 경우에는 학습용 영상과 촬영시기가 다른 영상을 시험영상으로 활용한다. SAR 영상이 신호의 내림각에 민감하기 때문에 시기가 다르더라도 동일한 내림각에 의해 촬영된 영상이 학습 및 시험에 모두 포함되도록 하였다. 다만 지역 A 영상 2장은 내림각 차이는 5° 이내이지만 촬영시기가 매우 상이하다.

- ⑥ TSX 위성의 촬영 방문시각은 궤도에 따라 다소 다르나 18시10분~18시40분이다. 실제 촬영이 이루어지는 시각은 약 10초이다.

(a)

(b)

(c)

Fig. 1. Annotation examples for SAR patches from (a) ID:JC200614, (b) ID:SJ191209, (c) ID:SJ201125

2.2 위성 SAR 영상 패치에 대한 표지작업

앞서 추출된 800×800 크기의 패치들에 대해서 지상차량 표적 별로 위치 박스(bounding box) 및 객체 마스크를 표시한다. Fig. 1에는 이러한 표지작업을 수행한 결과의 예시를 나타내었다. 좌측의 영상은 SAR 영상에서 추출한 800×800 패치에 표적탐지를 위한 박스 구획결과를, 우측의 영상은 표적 객체분할을 위한 마스크 구획결과를 나타내었다.

Fig. 1과 같은 annotation을 위하여, SAR 영상에 존재하는 지상차량의 실 정보(Ground Truth, GT)를 확인하는 방법에는 다음의 3가지가 있다.

- ① 방법 1 : 특정 지역의 좌표가 주어지면, TSX 위성의 촬영 가능한 스케줄을 사전에 확인할 수 있다. 이러한 점을 이용하여, 제천의 일부 장소에 대해 Fig. 2와 같은 대형 차량표적을 섭외하여 예정된 TSX 위성 방문시각에 배치해두는 방법을 활용하였다. 전술한 바와 같이 통상 지상차량 표적은 위치의 임의성이 높지만, 상기 섭외한 표적은 통제 가능하기 때문에 위성 SAR 영상에서의 표지작업 정확도가 매우 높다.

Fig. 2. Ground vehicles cast for TSX measurement

- ② 방법 2 : 방법 1은 정확성은 높으나, 차량표적 섭외를 위한 높은 비용 및 배치가능한 장소의 가용여부 등으로 어려움이 따를 수 있다. 방법 2는 표지작업의 정확도를 다소 희생시키면서 비용을 크게 줄일 수 있는 방법이다. 제천/세종 일부 구역에 대하여, Fig. 3과 같이 지상차량 표적이 집중적으로 배치된 곳의 현장사진을 위성 SAR 촬영시각 즈음에 촬영한다. 이를 획득한 위성 SAR 영상과 비교하여 표지작업을 수행한다. 이를 통해 비교적 높은 정확도의 표지작업이 가능하나, 지상차량 표적을 통제할 수 없기 때문에 현장사진 촬영 시점과 실제 위성 SAR 촬영시각 사이에 표적 위치 차이가 발생할 수 있다.

Fig. 3. Site photo and its corresponding SAR image

Fig. 4. Examples of indirect information acquisition from Google Earth for annotation by method 3

- ③ 방법 3 : TSX 위성 SAR 영상은 수 km^2 면적을 포함하므로 표적이 존재하는 것으로 이미 알고 있거나 직접 접근하여 사진 촬영이 가능한 구역 외에도 지상차량 표적들이 존재할 수 있다. 특히 수원

지역은 촬영 시기가 이미 5년이 넘었으며 국외 지역은 접근이 어렵기 때문에 TSX 촬영 당시의 현장상황을 직접 파악할 수 없다. 방법 3은 제천과 세종 일부구역, 수원 및 국외지역에서의 전 구역에 대한 표지작업을 위해 사용된 간접정보 기반의 방법이다. 전체 표지작업의 90 % 이상이 방법 3에 의해 수행되었다. 먼저 SAR 영상을 분석한 경험을 토대로, 지상차량 표적이 존재하는 것으로 예상되는 후보 구역을 선정한다. 그 후에는 Google Earth와 대조하여 해당 구역이 실제로 표적이 존재하는 구역인지 확인한다. Fig. 4에는 방법 3을 위한 간접정보 획득의 예시를 나타내었다.

Fig. 4(a)의 경우, ID SJ-191210인 SAR 영상의 $160 \text{ m} \times 160 \text{ m}$ 구역에 있는 여러 SAR 표적신호가 우측의 Google Earth(190403 촬영) 광학영상을 토대로 버스임을 추정할 수 있는 사례를 나타낸 것이다. 수직방향으로 위치한 차량들과는 달리 다른 방향으로 위치한 차량들의 표적특성은 밀집배치로 인해 비교적 약하다. Fig. 4(b)의 경우, ID SW-150128인 SAR 영상의 $80 \text{ m} \times 80 \text{ m}$ 구역 내에 있는 ‘A’ 신호와 ‘B’의 위쪽에 있는 여러 신호들에 대해 우측의 Google Earth(150208 촬영) 광학영상을 토대로 트럭 또는 콘크리트 펌프카임을 추정할 수 있다. Fig. 4(c)의 경우, ID SW-150128인 SAR 영상의 $240 \text{ m} \times 240 \text{ m}$ 구역에 지상표적이 다수 위치하고 있다. Google Earth 대조결과 ‘A’에는 버스 차고지로, ‘B’는 화물트럭과 같은 대형차량들의 주차장으로 확인되며, ‘C’는 도로변의 안전지대에 버스와 같은 차량들이 정차한 것을 나타낸다. 이러한 방법 3은 데이터 구축 측면에서는 매우 효율적이지만 정확도의 한계가 뚜렷하다. 추후에는 지속적으로 영상 데이터를 확보하여 방법 1과 2를 기반으로 데이터 셋의 표지 정확도를 높여나갈 것이다.

2.3 딥러닝 알고리즘 학습 및 시험용 영상 준비

18개의 SAR 영상에서 총 222개의 800×800 크기 패치를 추출하였고, 표적의 개수는 3318개이다. 그러나 Mask R-CNN을 이 크기의 영상으로 학습시키기에는 연산 자원(MATLAB Deep Learning Toolbox 및 NVIDIA GTX 1080 GPU)의 측면도 있으나 무엇보다 수량 자체가 매우 적다. 이에 GPU 메모리 절감 및 학습영상 수량을 증대하기 위하여 Fig. 5와 같이 보다 작은 패치의 중복 추출 및 데이터 증대(data augmentation)를 병

행하였다. 먼저 좌측의 800×800 SAR 패치로부터 수작업으로 N 개의 450×450크기 패치를 추출한다. 패치의 크기가 줄어들기 때문에 기존 패치의 모든 표적을 포함할 수 없게 되므로 작업자의 판단에 의해 추출 개수 N 을 조정할 수 있다. Fig. 5는 총 4번 추출을 수행한 예시를 나타낸 것이다. 이 때 공개된 SAR 선박 탐지용 데이터 셋을 구축하던 방식과 마찬가지로 본 논문의 데이터에 대해서도 일부 면적의 중복 추출을 허용한다. 이렇게 추출된 패치로부터 더 작은 400×400 크기의, 좌상단의 가로, 세로 좌표가 25 픽셀 씩 차이나는 패치 9개를 자동 추출한다. 또한 400×400 패치 별로 영상 값 동적범위(dynamic range)를 35 dB, 40 dB, 45 dB, 50 dB로 달리하여 4개를 더 추출한다. 정리하면 하나의 800×800 SAR 패치로부터 $N \times 9 \times 4$ 개의 학습용 SAR 영상(400×400 크기)이 만들어지게 된다. 여기서 N 은 1~5 사이의 값이며 800×800 패치 내의 표적 수에 따라 달라질 수 있다. 탐지 위치박스와 마스크와 같은 GT 정보는 800×800 패치에서 400×400 패치에

Fig. 5. Processing to 800×800 SAR patch for obtaining the images for Mask R-CNN training

Table 2. Number of training/test TSX SAR images

지역	학습영상 수	시험영상 수
JC	4,104	486
A	720	81
SJ	2,448	414
SW	5,184	567
합계	12,456	1,548

맞는 값으로 변환되므로 별도의 표지작업은 불필요하다. 시험용으로 분류된(Table 1의 연두색 바탕) 패치의 경우에는, 동적범위를 45 dB로 고정하여 $N \times 9$ 개의 영상만을 생성한다. Table 2에는 이렇게 준비된 지역 별 Mask R-CNN 딥러닝 알고리즘 학습용/시험용 400×400 크기 영상의 수량을 나타내었다.

3. Mask R-CNN 기반 딥러닝 알고리즘

3.1 네트워크 구조

Mask R-CNN^[19]은 딥러닝 기반의 표적탐지(detection)와 객체분할(instance segmentation)을 동시에 수행하는 네트워크로 구성된다. Mask R-CNN은 기존의 Faster R-CNN^[6]에 객체분할을 위한 지류를 추가한 형태로써, 입력영상의 특징추출을 위한 백본(backbone) 네트워크와 탐지 후보영역 제안 네트워크(Region Proposal Network, RPN), 최종 탐지 및 객체분할 결과를 도출하는 헤드(head) 등 3개의 요소로 구성된다. 각 부분의 세부설계는 설계자나 적용하는 문제에 따라 달라질 수 있다.

① 백본 네트워크 : Fig. 6에는 본 논문에서 지상 차량 표적 탐지/객체분할 학습 및 시험을 위해 사용되는 백본 네트워크의 구조를 나타내었다. 기본적인 형태는 ResNet18^[20]과 유사하며 각 단계에서의 특징추출을 위한 합성곱 계층(convolutional layer)의 개수가 다소 변경되었다. ‘stage 1’에서는 400×400 크기의 단일 채널을 갖는 SAR 영상을 입력으로 받아서 7×7 크기의 커널과 스트라이드(stride) 2, 영-패딩(zero-padding)수 3을 갖는 합성곱 계층(conv7×7@64,s2p3) 및 3×3 크기의 커널과 스트라이드 2, 영-패딩 수 1을 갖는 최댓값 풀링(max-pooling) 계층(max-pool3×3,s2p1)을 거쳐 100×100 크기의 채널수 64를 갖는 특징맵(feature map)을 생성한다. 이후 ‘stage 2~5’에서는 2개의 합성곱 계층 및 입력 특징맵을 출력 특징맵에 더하는 스킵(skip)연결과 덧셈 계층으로 구성된 단위 네트워크가 1~3개 위치하며, ‘stage 3~5’의 첫 번째 합성곱 계층에서 스트라이드가 2로 주어지므로 특징맵의 가로, 세로크기가 1/2배씩 감소한다. 각 ‘stage’ 별 도출되는 특징맵의 크기는 다음과 같다. 여기서 ‘stride x’로 표기한 것은 특징맵의 가로, 세로 크기 대비 입력영상 크기의 비율이다.

- stage 1 : $100 \times 100 \times 64$ (stride 4)
- stage 2 : $100 \times 100 \times 64$ (stride 4), 2 unit networks
- stage 3 : $50 \times 50 \times 128$ (stride 8), 3 unit networks
- stage 4 : $25 \times 25 \times 256$ (stride 16), 3 unit networks
- stage 5 : $13 \times 13 \times 512$ (stride 32), 1 unit network

‘stage 5’ 이후에는 특징맵의 채널 정보만 남기는 전역 평균값 풀링(global average pooling)계층과 노드 2개를 갖는 완전연결(fully connected, fc)계층, 소프트맥스 분류기가 추가된다. 모든 합성곱 계층 이후에는 배치정규화(Batch Normalization, BN) 계층과 ReLU (Rectified Linear Unit) 활성화 함수 계층이 포함되어 있으며, Fig. 6에서는 공간제약으로 인해 표시가 생략되었다. 탐지 및 객체분할을 위한 특징추출의 정확도 및 효율성을 높이기 위해, 통상적으로 Fig. 6과 같은 백본 네트워크를 수량이 풍부한 데이터 셋으로 기학습(pre-training)시킨다. 상당수의 SAR 선박 탐지 연구 사례^{[11-13]}}에서는 광학영상인 ImageNET 데이터로 기학습된 백본 네트워크를 사용한다. 이는 SAR 선박 탐지 알고리즘 개발용 공개 데이터셋 자체의 수량도 비교적 많기 때문에, 광학 영상에 의해 기학습된 네트워크의 가중치를 충분히 변화시킬 수 있기 때문이다. 그러나 본 논문의 지상차량 표적 데이터 셋은 수량이 상대적으로 적기 때문에 Fig. 7에 간략히 나타낸 별도의 데이터 셋으로 기학습시킨다. 해당 데이터 셋은 TSX 영상에서 중심의 지상차량 표적과 클러터를 별도로 추출한 ‘SAR 칩(chip)’ 데이터 셋으로서, 총 2655개의 표적 칩과 2795개의 클러터 칩으로 구성된다. Fig. 7 좌측의 표적 칩을 보면 중심의 표적뿐만 아니라 인접한 곳에 다수의 표적이 위치하고 있는 경우가 많으며, 우측의 클러터 칩을 보면 지형 및 숲에 의한 자연 클러터(natural clutter)보다는 강한 전자파 산란이 발생하는 건물, 구조물 등의 인공 클러터(cultural clutter)가 대다수임을 알 수 있다. 상기 표적 및 클러터 칩의 약 70 %에 해당하는 1858개의 칩으로 및 데이터 증대를 적용하여 백본 네트워크를 학습시키며, 나머지는 검증용 데이터로 활용한다. 본래 칩의 크기는 128×128 이나, 탐지 및 객체분할의 용도에 맞게 400×400 크기로 변환되었다. 797개의 검증용 표적 칩과 937개의 검증용 클러터 칩에 대하여, 백본 네트워크는 98.90 %의 정확도로 표적과 클러터를 구분하였다. 이후 기학습이 완료된 백본 네트워크에서 ‘stage 5’ 이후의 계층은 제거하고 RPN 및 head 구조를 추가한다.

Fig. 6. ResNet-type backbone network structure

Fig. 7. Dataset for pretraining backbone network (left : targets, right : clutters)

② RPN : Fig. 8에는 RPN 및 주변의 하부 네트워크 구조를 나타내었다. 백본 네트워크의 'stage 4'에서 도출된 $25 \times 25 \times 256$ 크기의 특징맵을 입력받아서 3×3 합성곱 계층과 ReLU를 거쳐서 관심영역(Region of Interest, ROI) 분류를 위한 1×1 합성곱 계층과 ROI 박스 크기/위치 추정(regression)을 위한 1×1 합성곱 계층에서 각각 출력 특징맵을 도출한다. 이어서 RPN에서는 상기 두 개의 특징맵을 입력 받아서 특징맵 상의 유력한 ROI R 개의 위치 및 크기 값을 산출한다. 끝으로 'ROI Align' 계층에서는 'stage 4'에서 도출된 특징맵으로부터 RPN의 ROI 정보를 받아서 해당 부분만 추출한 후 일괄적으로 8×8 가로, 세로크기로 변환한다. 이 '추출'과정에서 Faster R-CNN의 경우에는 ROI가 특징맵의 25×25 그리드에 정확히 걸치지 않는 경우 반올림을 수행하지만 (ROI pooling), Mask R-CNN의 ROI align에서는 주변 그리드 값과의 보간을 수행하여 ROI 추출의 정확도를 높였다. 상기의 과정에서 추출된 $8 \times 8 \times 256$ 크기의 특징맵 R 개(= Fig. 8의 'ROIs')가 백본 네트워크의 'stage 5'에 입력된다.

Fig. 8. RPN and sub-network structures

③ head : Fig. 9에는 head의 구조를 나타내었다. 앞서 백본 네트워크의 'stage 5' 부분에서 $4 \times 4 \times 512$ 크기의 ROI R 개에 대한 특징맵을 도출하면, 먼저 전역 평균 풀링을 거쳐 탐지를 위한 분류(classification)와 ROI의 정확도를 높이기 위한 추가적인 위치/크기값 추정으로 들어간다. 다른 한편으로는 부류와 무관하게(class agnostic) 탐지된 표적의 마스크를 추정하는 부분으로 들어가게 된다. 분류와 박스 값

추정 부분에서는 각각 노드개수 2와 4를 갖는 완전 연결 계층에 의해 RPN에서 도출된 ROI 개수만큼의 부류(표적인지 배경인지) 점수와 단일 부류의 표적에 한하여 탐지 박스 값 4개(좌상단 x 좌표, 좌상단 y 좌표, 너비 w , 높이 h)가 ROI 개수만큼 도출된다. 마스크 추정부분에서는 전치 합성곱 등을 거쳐 ROI의 개수만큼의 $8 \times 8 \times 1$ 마스크를 추정한다. 이들은 차후 객체분할 결과 표시 과정에서 보간법을 통해 원 영상대비 탐지 박스의 크기에 대응되도록 크기가 변경된다.

Fig. 9. head for Mask R-CNN structure

3.2 Mask R-CNN의 순방향(forward) 연산

3.2.1 RPN 및 하부 네트워크의 순방향 연산

3.2장에서 Mask R-CNN의 RPN 및 하부 네트워크, head를 중심으로 순방향 연산에 대해 기술한다.

① 대부분의 탐지용 딥러닝 네트워크에서는 처음부터 탐지 박스 값을 추정하지 않고, 특정 위치에 여러 개의 앵커(anchor)라는 후보 박스(초기값)를 두어 이로부터 실제 GT 박스에 근접하게 값을 변화시키는 방식을 차용한다. 후보 박스의 생성 위치는 Fig. 10(a)와 같이 RPN 하부 네트워크에 입력되는 특징맵의 25×25 그리드이며, 이는 1/16배로 줄어든 입력 영상의 위치와 정사영 관계를 갖는다. 각 그리드에 생성되는 앵커의 개수는 설계자 또는 적용되는 문제에 따라 달라질 수 있으며, 본 논문에서는 9개를 설정하였다. 앵커의 크기(너비 w 및 높이 h)는 다수의 GT 박스로부터 k -평균 군집화(k -means clustering)에 의해 가장 대표적인 9개의 크기를 Fig. 10(b)과 같이 결정하였다. 이 때 앵커의 크기 값은 입력영상에 대한 값으로 설정된다. 이렇게 초기에 생성되는 앵커의 수는 영상 하나 당 $25 \times 25 \times 9 = 5625$ 개이다.

- ② Fig. 8과 같이 RPN에는 분류용 1×1 conv(Cls) 및 박스 추정용 1×1 conv(Box)로부터의 $25 \times 25 \times 18$, $25 \times 25 \times 36$ 크기의 특징맵이 입력된다. 먼저 conv(Box)의 출력 특징맵을 $25 \times 25 \times 9 \times 4$ 의 형태로 변환한다. 이때 4는 앵커박스의 (x, y) 중심, 너비 w , 크기 h 를 변화시키는 목표치의 개수를 의미하며, 이러한 값이 특징맵의 각 그리드 및 앵커 별로 출력된다. 따라서 다음과 같이 앵커 박스를 conv(Box)의 출력을 이용하여 조정한다.

Fig. 10. (a) anchors defined in each grid of feature map, (b) anchor size decision via k-means clustering

$$x = w_a t_x + x_a, y = h_a t_y + y_a, w = w_a e^{t_w}, h = h_a e^{t_h} \quad (1)$$

식 (1)에서 (x_a, y_a, w_a, h_a) 는 초기 앵커의 4개 값을

의미하고, (t_x, t_y, t_w, t_h) 는 conv(Box)의 출력으로서 앵커의 값을 움직이는 양이다. (x, y, w, h) 는 조정된 앵커박스의 값이다.

- ③ 이어서 conv(Cls)의 출력 특징맵을 $25 \times 25 \times 9 \times 2$ 로 변환하면 그리드 및 앵커 별로 전경(표적, foreground)에 해당하는 점수와 배경(background)에 해당하는 점수를 알 수 있다. 여기서 관심사항은 표적에 대한 점수이므로, 조정된 앵커박스 중에서 상위 2000개를 추리고 이후 비 최대억제(Non-Maximum Suppression, NMS)로 겹치는 박스를 추가 제거하여 영상 하나당 500개의 박스만을 남겨놓는다.
- ④ 이렇게 RPN에서는 하나의 영상에 대해 500개, 즉 Fig. 8의 'proposal number'만큼 ROI를 생성하며, 각 ROI는 4개의 특징 값(attributes, x, y, w, h)을 갖는다. 학습 반복 단위인 미니배치(mini-batch)내 영상 수는 후술하는 바와 같이 8개로 지정하였으므로 반복 회수 1회당 $4 \times (8 \times 500)$ 개의 특징 값이 RPN에서 도출된다.

3.2.2 head의 순방향 연산

- ① Fig. 9와 같이 완전 연결 계층 fc(Cls)에서는 ROI 별 표적/배경 점수(2 노드)가 산출되고, fc(Box)에서는 조정된 ROI를 보다 정확한 탐지결과에 근접시키기 위한 추정 목표 값 4개가 ROI 별로 산출된다. 이로부터 식 (1)을 이용하여 앞서 RPN에서 조정된 ROI가 추가 조정된다. 여기서 (x_a, y_a, w_a, h_a) 는 조정된 앵커의 값이고, (t_x, t_y, t_w, t_h) 는 fc(Box)의 출력이며 (x, y, w, h) 는 최종 탐지결과 박스의 값이다.
- ② 동시에 ROI 별로 8×8 크기 마스크가 산출된다.
- ③ 여기까지가 사실상 Mask R-CNN의 최종 출력단인 head에서 도출되는 것이고 이후에는 fc(Cls)의 점수에 의한 선별 및 추가적인 NMS와 같은 후처리에 의해 원하는 탐지결과만 남기게 된다. 마스크는 후처리 결과에 대응되어 선별된다.

3.3 Mask R-CNN의 역방향(reverse) 연산

3.3장에서는 Mask R-CNN의 RPN 및 하부 네트워크, head를 중심으로 역방향 연산에 대해 기술한다. 딥러닝 네트워크의 학습은 항상 실제 값(또는 목표 값) 대비 출력 값의 손실(loss)을 구하고, 손실에 대한 네트워크의 가중치 변화량(gradient)을 구함으로써 반복적으로 수행된다.

3.3.1 head의 역방향 연산

- ① 분류 손실 : RPN에서 도출된 ROI(500개×미니배치)들과 표지작업에 의해 생성된 영상 내 GT 박스와의 IOU(Intersection Of Union)를 일괄 산출한 결과를 토대로, ‘표적’과 ‘배경’을 구분한다. 구분을 위한 IOU 값은 설계자에 의해 결정되는 변수이며, ‘표적’인 경우 [10], ‘배경’인 경우 [01], 어느 것에도 해당되지 않는 경우 [00]으로 이진수 인코딩(one-hot encoding)하여 $2 \times \text{ROI}$ 형태의 목표 값 텐서(tensor)로 둔다. 이를 완전 연결 계층 fc(Cls)의 동일크기 출력 텐서와 다음과 같이 이진 교차 엔트로피 손실(binary cross-entropy loss)을 구한다.

$$L_{ds,RCNN} = \frac{1}{N} \sum T \log X \quad (2)$$

여기서 T 는 목표 텐서, X 는 head의 출력 값으로서 소프트맥스(softmax) 연산이 수행된 상태이다.

- ② 박스 값 추정 손실 : ‘표적’인 ROI의 (x_a, y_a, w_a, h_a) 값과 그와 일대일로 매칭되는 GT 박스 값 (x, y, w, h) 을 대조하여 추정의 목표값을 산출한다. 이는 ROI가 GT 박스에 가까워지기 위한 이동량으로서 식 (1)의 (t_x, t_y, t_w, t_h) 를 역으로 계산하면 된다. 이러한 4개의 값을 ROI 별로 계산하여 $4 \times \text{ROI}$ 형태의 목표 값 텐서로 둔다. GT와 매칭되지 않는 ROI는 0으로 채워진다. 이를 fc(Box)의 출력 $4 \times \text{ROI}$ 텐서와 다음과 같이 평탄화 L1 손실(smooth L1 loss)을 구한다.

$$L_{reg,RCNN} = \frac{1}{M} \sum_{i \in x, y, w, h} \text{smooth}_{L_1}(t_i - x_i) \quad (3)$$

$$\text{smooth}_{L_1}(u) = \begin{cases} 0.5u^2 & \text{if } |u| \leq 1 \\ |u| - 0.5 & \text{otherwise} \end{cases} \quad (4)$$

여기서 $t_i = (t_x, t_y, t_w, t_h)$ 이고, x_i 는 네트워크의 출력 값 4개, M 은 미니배치 내 ‘표적’ ROI의 개수에 독립적인 목표값 개수인 4를 곱한 수로서 손실의 정규화(normalization)역할을 한다.

- ③ 마스크(mask) 손실 : ‘표적’으로 분류된 ROI에 한하여, ROI와 매칭되는 GT의 마스크를 8×8 크기로 변경한 것을 목표 값으로 삼는다. Fig. 11과 같이 좌측의 ROI 개수만큼의 목표 마스크에서 ‘표적’

ROI에 해당하는 인덱스에 대해서는 그와 매칭되는 GT의 마스크가 채워진 것을 알 수 있고, 그 외에는 0으로 채워진다. 이를 네트워크의 ‘sigmoid’ 출력 값인 동일 크기의 텐서와 다음과 같은 픽셀 단위의 교차 엔트로피 손실을 산출한다.

$$L_{mask,RCNN} = \frac{1}{K} \sum T_{mask} \log X_{mask} \quad (5)$$

여기서 T_{mask} 는 Fig. 11 좌측과 같은 목표 마스크, X_{mask} 는 Fig. 11 우측과 같은 출력 마스크이며 K 는 하나의 마스크 내의 픽셀 수($=8 \times 8=64$)에 해당한다.

Fig. 11. Target (left) and output (right) masks

3.3.2 RPN 및 하부 네트워크의 역방향 연산

- ① head의 손실 계산 과정이 RPN의 ROI와 GT 간 관계에서 출발하였다면, RPN 학습을 위한 역방향 연산에서는 특징맵 25×25 그리드에 배치된 모든 앵커와 GT 간의 관계에서 출발한다.
- ② 모든 앵커 중에서 GT와의 IOU를 기준으로 양성(positive)으로 매칭되는 것과 음성(negative)로 매칭되는 것을 각각 100개씩 임의로 선정한다. 통상 양성 샘플 수가 100개에 미치지 못하므로 모자라는 부분은 음성 샘플로 채운다. 이때 양성이란 표적을 나타내는 전경, 음성이란 표적 외의 배경을 의미한다.
- ③ RPN 분류 손실 : 선정된 양/음성 샘플 앵커에 대하여 앵커의 그리드 내 위치 및 이진수 인코딩을 통해 $25 \times 25 \times 18$ 크기 텐서로 매핑한 것을 목표 값으로 삼는다. 즉 18은 그리드 ($1 \leq m, n \leq 25$) 상의 앵커 1의 양/음성 인덱스, 앵커 2의 인덱스, ... 등의 순서를 의미한다. 이러한 매핑 결과를 conv(Cls) 출력의 소프트맥스 값과 대조하여 교차 엔트로피 손실을 산출한다.

④ RPN 박스 값 추정 손실 : 양성 앵커 샘플에 한하여, 이와 일대일로 매칭되는 GT 박스와의 (t_x, t_y, t_w, t_h) 값을 식 (1)에 의해 계산하여, 앵커의 그리드 상 위치를 고려한 결과를 바탕으로 $25 \times 25 \times 36$ 크기 텐서로 매핑한다. 즉 36은 그리드 ($1 \leq m, n \leq 25$) 상의 앵커 1의 t 값, 앵커 2의 t 값, ... 등의 순서를 의미하며, 이 결과를 $\text{conv}(\text{Box})$ 출력과 평탄화 L1 손실을 계산한다.

이렇게 구한 총 5개의 손실을 더한 통합 손실 값의 변화량으로 네트워크의 가중치를 업데이트한다.

4. 실험결과

2장에서 언급한 데이터셋을 3장에 소개된 Mask R-CNN의 학습 및 성능산출을 위한 시험에 사용하였다. Table 2에 나타낸 12,456개의 학습영상으로 Mask R-CNN을 학습시키기 위한 변수들은 다음과 같다.

- 미니배치 내 영상의 개수 : 8(mini-batch size)
- 매 에폭(epoch, 모든 영상이 1번씩 네트워크 학습에 활용되는 주기)마다 학습영상이 섞임(shuffle)
- 최적화 기법 : SGDM(Stochastic Gradient Descend with Momentum)
- 학습률(learning rate) : 초기 0.01, 매 반복 회수 i 마다 $1/(1+\text{decay} \times i)$ 배 감소, $\text{decay} = 0.0001$
- 모멘텀(momentum) : 0.9
- 최대 에폭 수 : 100 = $(12456/8) \times 100$ 반복 회수
- 양성(background) 및 음성(background) ROI 판정을 위한 IOU 범위 : [0.7, 1.0] 및 [0.0, 0.5]

학습이 완료된 Mask R-CNN에 대해, 시험용 영상을 입력하여 탐지 결과 박스 및 표적 개별영역에 대한 객체분할 결과를 획득한다. 이때 입력영상에 대하여 Mask R-CNN 내부에서 수행되는 과정은 백본 네트워크에 의한 특징맵 형성 이후 3.2장의 순방향 연산과 동일하다. Fig. 12에는 Table 1의 시험용 SAR 영상에서 추출한 400×400 크기 영상에 대해 표적 탐지와 객체분할을 적용한 사례를 나타내었다. 지면관계로 8가지 경우에 대해서만 나타내었으며, 좌측에는 GT를 흰 박스로, 우측에는 탐지결과를 노란 박스 및 푸른색 마

스크로 나타내었다. 결과의 표시를 위해, 후처리 과정에서는 head의 완전 연결 계층 $\text{fc}(\text{Cls})$ 에서 도출되는 표적에 대한 신뢰도 점수(confidence score)의 문턱치(threshold)보다 높은 결과만 남긴다. 본 실험에서는 실제 탐지되는 표적의 개수 및 탐지 자체의 정확도 간 절충을 고려하여 문턱치를 0.01로 설정하였다. NMS 문턱치는 0.7이다.

- ① Fig. 12(a)에서는 3개의 큰 표적이 매우 높은 신뢰도 점수를 가지며 탐지된 것을 확인할 수 있다. 이는 해당 표적들을 동일지역(JC) 다른 시점(200614)에서 유사한 방향으로 촬영한 영상이 네트워크 학습에 포함되었기 때문으로 추정된다.
- ② Fig. 12(b)는 공사용 차량 차고지로서, 배치된 차량 내역은 달라질 수 있으나 동일지역(JC) 및 다른 시점에서 촬영된 영상이 다수 학습에 포함되어 있었으므로, 대체적으로 높은 신뢰도 값을 가지고 지상 차량 표적을 탐지하였다. 그러나 일부 크기가 작은 표적이거나 SAR 영상 내 표적특성이 불분명한 것은 신뢰도 점수가 낮음을 알 수 있다. 이러한 표적까지 탐지하려면 문턱치를 0.01 수준으로 내려야하는데, 이 과정에서 0.01보다 높은 점수를 갖는 오경보도 탐지된다.
- ③ Fig. 12(c)는 밀집하여 배치된 표적의 예시로서, 2개를 제외한 대부분의 표적을 성공적으로 탐지하였다.
- ④ Fig. 12(d)는 지면이 거친 개활지로서, 클러스터 신호가 비교적 높은 편이다. 일부 크기가 큰 표적에 대해서는 탐지 및 객체분할이 가능하나, 표적특성이 미약한 경우에는 실패하였다.
- ⑤ Fig. 12(e) 및 Fig. 12(f)는 세종 차량 주차장 내 대부분의 표적을 잘 탐지하고 있으나, 일부 신뢰도 점수가 낮은 오경보 및 탐지하지 못한 표적이 관측된다.
- ⑥ Fig. 12(g) 및 Fig. 12(h)는 수원의 차량 주차장에 대한 결과를 나타내었다. 여기서도 다수의 표적을 효과적으로 탐지하고 있으나, 일부 지상차량 표적과 유사한 SAR 신호 패턴에 의해 오경보가 발생하며 특히 작은 차량에 대해서는 탐지 박스와 GT 박스 간 IOU가 적거나 아예 검출에 실패한 경우가 존재한다.

Fig. 12. GTs and Results of target detection and instance segmentation (a) JC201024, (b) JC201109, (c) A191207, (d) SJ200430, (e) SJ200430, (f) SJ201114, (g) SW150106, (h) SW150106

성능산출을 위해 범용적으로 사용되는 정밀도 및 재현율(precision & recall), 그리고 AP를 사용한다. 정밀도는 모든 탐지 결과 중에서 정확하게 탐지한 비율을 의미하며, ‘true positive / all detections’로 계산된다. 재현율은 반드시 탐지하여야 하는 표적들 중에서 실제 탐지한 비율을 의미하며, ‘true positive / all GT boxes’로 계산된다. 여기서 ‘정확하게 탐지’하거나 ‘잘못 탐지’한 것의 기준은 실제 탐지 박스와 GT 박스 간의 IOU에 의해 결정된다. IOU 기준이 낮을수록 탐지 박스가 GT 박스에 약간만 겹쳐있어도 정답으로 간주한다는 것이고, IOU 기준이 높으면 탐지 박스가 GT 박스가 ‘상당히’ 유사하여야만 정답으로 간주한다는 의미가 된다. Fig. 13에는 IOU 기준이 0.5인 상태에서 신뢰도 점수의 문턱치에 따른 정밀도와 재현율 간의 곡선(PR 곡선)을 나타내었다. 곡선은 모든 시험 대상 지역 및 지역 별로 함께 나타내었다. 신뢰도 점수의 문턱치가 높으면 탐지되는 표적의 수(재현율)는 적지만 정밀도는 높음을 알 수 있고, 문턱치가 낮아짐에 따라 반비례하게 변화함을 알 수 있다. 이러한 PR 곡선 아래의 그래프 넓이로 정의한 값이 AP이며, 탐지 성능 산출 시 가장 널리 쓰이는 지표이다. A지역 영상의 경우 네트워크 학습에 사용된 동일지역 영상과의 시간 차이가 4년 이상 차이 남에도 불구하고, 비교적 단조로운 표적 SAR 영상의 형태로 0.76의 높은 AP 값을 보인다. 그 다음으로는 제천 0.73, 세종 0.71의 순이며 밀집표적이 다수 위치한 주차장과 차고지 등이 많은 수원 지역의 경우 다소 낮은 0.64의 AP를 갖는다. 또한 수원에서는 표지작업이 전적으로 방법 3에 의해 수행되었으므로 낮은 표지작업 정확도가 성능에 영향을 미친 것으로 추정된다. 전체 지역에 대해서는 0.69의 AP가 산출된다.

Table 3에는 상기 IOU 기준에 따른 모든 지역(All) 및 각 지역에 대한 AP 성능 값을 나타내었다. Table 3의 아래쪽 두 줄에는 두 가지 성능 값(성능 1, 성능 2)을 나타내었다. 전자는 사물검출용 사진 데이터 셋인 PASCAL VOC에서 쓰이는 성능 지표로서 IOU 기준이 0.5인 경우에 대한 AP값이다. 후자는 보다 수량이 많은 사진 데이터 셋인 MS-COCO에서 쓰이는 성능 지표로서 IOU 기준 0.5부터 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95 인경우의 AP를 평균한 값이다. 성능지표는 SAR 영상의 실정에 맞게 설계자가 별도로 고안할 수 있으며, 여기서는 범용적인 지표를 차용하여 산출한 결과를 나타낸 것이다.

Fig. 13. PR curves of test SAR images

Table 3. AP metrics for different IOU standards

IOU	All	JC	A	SJ	SW
0.10	0.7518	0.7911	0.7890	0.7414	0.7117
0.20	0.7460	0.7849	0.7890	0.7395	0.7031
0.30	0.7365	0.7743	0.7865	0.7342	0.6910
0.40	0.7227	0.7568	0.7796	0.7268	0.6762
0.50	0.6972	0.7340	0.7604	0.7126	0.6404
0.55	0.6744	0.7155	0.7436	0.6916	0.6119
0.60	0.6243	0.6565	0.7010	0.6699	0.5563
0.65	0.5541	0.5727	0.6405	0.6264	0.4849
0.70	0.4544	0.4616	0.4555	0.5847	0.3961
0.75	0.3215	0.3208	0.2704	0.4216	0.2764
0.80	0.1708	0.1688	0.1081	0.2435	0.1441
0.85	0.0559	0.0504	0.0348	0.0865	0.0527
0.90	0.0070	0.0062	0.0045	0.0095	0.0090
0.95	0.0002	0.0002	0.0001	0.0002	0.0002
성능 1	0.6972	0.7340	0.7604	0.7126	0.6404
성능 2	0.3560	0.3687	0.3719	0.4047	0.3172

Table 4에는 상기 IOU 기준을 0.5로 고정된 상태에서 앞서 0.01로 고정시켰던 신뢰도 점수의 문턱치를 다양하게 조절하면서 모든 지역(All) 및 각 지역에 대한 AP 성능 값을 나타내었다. IOU 기준이 고정된 상

태에서 문턱치를 높이면 Fig. 13과 같은 PR 곡선의 길이가 짧아지며, 이는 표적으로 판정하는 기준이 높아짐에 따라 재현율의 상한선이 낮아짐을 의미한다. 앞서 언급한 바와 같이 본 논문에서는 탐지 수와 탐지 정확도의 절충을 나타내는 AP 값을 고려하여 0.01의 문턱치를 선정하였다.

Table 4. AP metrics for different confidence thresholds

conf. thresh.	All	JC	A	SJ	SW
0.5	0.6049	0.6462	0.6840	0.6251	0.5393
0.4	0.6155	0.6559	0.6936	0.6364	0.5506
0.3	0.6268	0.6676	0.7031	0.6473	0.5618
0.2	0.6396	0.6779	0.7138	0.6592	0.5777
0.1	0.6577	0.6958	0.7308	0.6747	0.5978
0.01	0.6972	0.7340	0.7604	0.7126	0.6404
0.001	0.7250	0.7547	0.7744	0.7448	0.6742
0.0001	0.7415	0.7679	0.7841	0.7673	0.6913
0.00001	0.7519	0.7777	0.7917	0.7811	0.7007

5. 결 론

광학영상에서 시작된 딥러닝 알고리즘의 출현 및 발전은 SAR 영상에서의 해상표적 및 항공표적에 대한 자동탐지 연구의 성숙을 추동하였으나, 상대적으로 지상표적에 대한 연구사례는 적었다. 이러한 한계점을 극복하고자 본 논문에서는 위성 SAR 영상의 지상차량 표적 데이터 셋을 구성하여, Mask R-CNN 기반의 딥러닝 알고리즘을 적용하여 표적 탐지 및 객체분할을 수행하였다. 수행결과를 바탕으로 관련 성능을 도출함으로써 데이터 셋과 적용 알고리즘의 유효성을 보였다. 본 논문의 목적은 최신 딥러닝 알고리즘을 적용하여 최적의 성능을 도출하기보다, 추후 연구를 위한 기준(baseline)성능을 제시하기 위함이며, 연구방향 및 보완사항은 다음과 같이 요약될 수 있다.

① 2장에서 언급한 바와 같이, 표적 종류 및 위치에 관한 실제정보가 면밀히 파악된 데이터 셋을 추가로 확보하여 데이터 셋의 표지 정확도를 보완해야 한다.

- ② 현재는 SAR 영상에서 추출한 패치 단위로 독립적인 탐지 및 객체분할을 수행하였으나, 실용성을 높이기 위해서는 전체 영상과의 연계가 필요하다.
- ③ 본 논문에서 사용된 Mask R-CNN은 가장 기본적인 백본 네트워크 및 RPN 및 head 구조를 가지고 있다. 실제 적용결과를 통해 살펴본 성능의 향상방안을 크게 3가지로 생각할 수 있다.
 - 특징 일반화를 위해 보다 깊은 백본 네트워크 사용
 - 작은 지상차량 표적 탐지성능을 높이기 위한 다중 스케일 구조 사용(feature pyramid network 등)
 - 일부 표적신호 특성이 약한 지상차량 표적 탐지성능을 높이기 위한 공간 및 채널방향의 네트워크 집중 메커니즘(spatial & channel attention mechanism)
- ④ 양성/음성 박스 샘플을 결정하기 위한 GT와의 IOU 범위 및 RPN 학습 샘플 수, 평탄화 L1 손실 대신 IOU 손실 적용과 같은 네트워크 학습을 위한 내부변수를 조정한다.
- ⑤ 전체 SAR 영상 적용 시, 딥러닝 네트워크만으로는 오경보 제거가 어려울 수 있다. 이에 별도의 표적과 클러스터 구분을 위한 분류기 등을 부가할 수 있다.
- ⑥ 본 논문에서는 마스크 추출 결과에 대해서는 구체적인 언급을 하지 않았으나, 추후 표적영역 자동분할이나 스펙클의 영향을 받지 않는 표적영상 분류 알고리즘 등에 활용가능하다.

References

- [1] K. El-Darymli, et. al., "Target Detection in Synthetic Aperture Radar Imagery: A State-of-the-Art Survey," J. Appl. Remote Sens., Vol. 7, No. 071598, pp. 1-36, 2013.
- [2] K. El-Darymli, et. al., "Automatic Target Recognition in Synthetic Aperture Radar Imagery: A State-of-the-Art Review," IEEE Access, Vol. 4, pp. 6014-6058, 2016.
- [3] G. Gao, et. al., "An Adaptive and Fast CFAR Algorithm Based on Automatic Censoring for Target Detection in High-Resolution SAR Images," IEEE Trans. Geosci. Remote Sens., Vol. 47, No. 6, pp. 1685-1697, 2009.

- [4] X. Qin, et. al., "A CFAR Detection Algorithm for Generalized Gamma Distributed Background in High-Resolution SAR Images," IEEE Geosci. Remote Sens. Lett., Vol. 10, No. 4, pp. 806-810, 2013.
- [5] R. Girshick, et. al., "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 580-587, 2014.
- [6] S. Ren, et. al., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE PAMI, Vol. 39, No. 6, pp. 1137-1149, 2017.
- [7] J. Redmon, et. al., "You Only Look Once: Unified, Real-Time Object Detection," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 779-788, 2016.
- [8] W. Liu, et. al., "SSD: Single Shot MultiBox Detector," Proceedings of the European Conference on Computer Vision(ECCV), pp. 21-37, 2016.
- [9] T. Lin, et. al., "Feature Pyramid Networks for Object Detection," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 2117-2125, 2017.
- [10] C. Chen, et. al., "A Deep Neural Network Based on an Attention Mechanism for SAR Ship Detection in Multiscale and Complex Scenarios," IEEE Access, Vol. 7, pp. 104848-104863, 2019.
- [11] Y. Chang, et. al., "Ship Detection Based on YOLOv2 for SAR Imagery," Remote Sens., Vol. 11, No. 786, pp. 1-14, 2017.
- [12] J. Jiao, et. al., "A Densely Connected End-to-End Neural Network for Multiscale and Multiscene SAR Ship Detection," IEEE Access, Vol. 6, pp. 20881-20892, 2018.
- [13] Y. Zhao, et. al., "Attention Receptive Pyramid Network for Ship Detection in SAR Image," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens, Vol. 13, pp. 2738-2756, 2020.
- [14] Y. Zhao, et. al., "Pyramid Attention Dilated Network for Aircraft Detection in SAR Images," IEEE Geosci. Remote Sens. Lett., Vol. 18, No. 4, pp. 662-666, 2021.
- [15] Q. Guo, et. al., "Scattering Enhanced Attention Pyramid Network for Aircraft Detection in SAR Images," IEEE Trans. Geosci. Remote Sens., Early Access(DOI: 10.1109/TGRS.2020.3027762), pp. 1-18, 2021.
- [16] T. Zhang, et. al., "LS-SSDD-v1.0: A Deep Learning Dataset Dedicated to Small Ship Detection from Large-Scale Sentinel-1 SAR Images," Remote Sens., Vol. 12, No. 18, pp. 1-37, 2020.
- [17] S. Xian, et. al., "AIR-SARShip-1.0: High-Resolution SAR Ship Detection Dataset," J. Radars, Vol. 8, No. 6, pp. 852-862, 2019.
- [18] S. Wei, et. al., "A High-Resolution SAR Images Dataset for Ship Detection and Instance Segmentation," IEEE Access, Vol. 8, pp. 120234-120254, 2020.
- [19] K. He, et. al., "Mask R-CNN," Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2961-2969, 2017.
- [20] K. He, et. al., "Deep Residual Learning for Image Recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 770-778, 2016.