

일반논문 (Regular Paper)

방송공학회논문지 제29권 제2호, 2024년 3월 (JBE Vol.29, No.2, March 2024)

<https://doi.org/10.5909/JBE.2024.29.2.198>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

클래스 활성화 지도를 이용한 딥러닝 기반의 SAR 선박 탐지 기술 연구

김 병 준^{a)}, 문 효 이^{a)}, 이 재 석^{a)}, 김 선 옥^{b)†}

SAR Ship Detection based on Deep Learning using Class Activation Map

Byoungjun Kim^{a)}, Hyoui Moon^{a)}, Jaeseok Lee^{a)}, and Sunok Kim^{b)†}

요 약

원격 감지 이미지에서 선박 탐지는 선박의 복잡한 배경과 작은 크기로 인해 어려운 작업이다. 딥러닝 기반 물체 탐지 알고리즘은 선박 감지에서 유망한 결과를 보여주었지만, 종종 해석 가능성이 부족하다. 본 논문에서는 원격 감지 이미지에서 선박 탐지를 위한 Faster-RCNN과 CAM(Class Activation Map)의 융합 기술을 제안한다. 제안한 기술은 정확한 물체 감지에서 Faster-RCNN의 장점과 해석 가능성에서 CAM의 장점을 결합하는데, Faster-RCNN 알고리즘은 관심 영역을 감지하는 데 사용되고, CAM 기술은 선박 감지에 기여하는 이미지의 영역을 강조하는 히트 맵을 생성하는 데 사용된다. 그런 다음 ROI와 클래스 활성화 맵을 결합하여 재현율을 유지 하면서 더 높은 정밀도를 달성하여 선박 탐지 결과를 개선한다.

Abstract

Ship detection in remote sensing images is a challenging task due to the complex backgrounds and small sizes of ships. Deep learning-based object detection algorithms have shown promising results in ship detection, but they often lack interpretability. In this paper, we propose a fusion technique of Faster-RCNN and Class Activation Map (CAM) for ship detection in remote sensing images. Our proposed technique combines the advantages of Faster-RCNN in accurate object detection and CAM in interpretability. The Faster-RCNN algorithm is used to detect regions of interest, and the CAM technique is used to generate a heat map that highlights the regions of the image that contribute to the ship detection. The ROIs and the class activation map are then combined to refine the ship detection results. Our experiments show that the proposed approach achieves higher Precision while maintaining Recall.

Keyword : Object Detection, Deep Learning, SAR ship detection, Class Activation Map

a) 한화시스템(Hanwha System)

b) 한국항공대학교 인공지능학과(Department of Artificial Intelligence, Korea Aerospace University)

† Corresponding Author : 김선옥(Sunok Kim)

E-mail: sunok.kim@kau.ac.kr

Tel: +82-2-300-0262

ORCID: <https://orcid.org/0000-0002-9665-4214>

※ 이 연구는 2022년 정부(방위사업청)의 재원으로 국방과학연구소의 지원을 받아 수행된 미래도전국방기술 연구개발사업(No.915029201)

· Manuscript 26 February, 2024; Revised 25 March, 2024; Accepted 25 March, 2024.

1. 서론

합성 개구 레이더^[1](Synthetic Aperture Radar, SAR)는 공중에서 지상 및 해양을 관찰하며 지구 표면의 고해상도 이미지를 만드는 데 사용되는 레이더 시스템의 한 종류이다. 햇빛에 의존하는 광학 센서를 활용한 전자광학(Electro-Optical, EO)과 달리, SAR는 지구 표면에서 오는 전파의 반사를 감지하고 측정하기 위해 마이크로파 에너지를 사용한다. SAR 이미지는 고유한 특성으로 구름과 다른 대기 조건에 상관없이 침투할 수 있으며 악천후에서도 이미지를 캡처할 수 있다. 이러한 SAR 영상의 특성을 이용하여 표적 탐지, 표적 분류, 지표면 변화 추적 등을 포함한 다양한 애플리케이션에서 SAR 기술을 활용하고 있다.

그 중에서 SAR 이미지를 활용한 선박 탐지는 공간 해상도가 제한되어 있어 작은 특징이나 변경 사항을 식별하기 어려우며 선박의 복잡한 배경과 작은 크기로 인해 어려운 작업이다. 기존 SAR 이미지 탐지방법은 수작업으로 하는 방식과 CFAR^[2](Constant False Alarm Rate) 방식으로 구분할 수 있다. 수작업 방식은 이미지마다 적용하는 매개변수가 달라서 범용성이 떨어진다는 단점이 있고, CFAR는 표적의 픽셀(pixel) 값이 주변 영역보다 높다는 가정으로 픽셀 값에 대한 클러스터링을 통해 탐지하게 되는데 이는 복잡한 배경을 가져 픽셀 분포가 복잡한 SAR 이미지에 대해서는 선박 탐지 성능이 제한적이며 모든 픽셀을 하나씩 처리하기에 속도가 매우 느리다는 단점이 있다.

이후 딥러닝의 발전으로, 합성곱 신경망^[3](Convolutional Neural Network, CNN)을 기반으로 하는 물체 탐지 알고리즘은 선박 탐지에서 좋은 결과를 보여주었다. 이러한 알고리즘 중 Faster-RCNN^[4]은 객체 탐지 작업에서 최첨단 성능을 달성하는 널리 사용되는 방법이다. 그러나 Faster-RCNN은 예측에 대한 시각적 설명을 제공하지 않으므로 해석 가능성이 제한된다는 단점이 존재한다.

이 문제를 해결하기 위해 딥러닝 모델이 클래스 예측에 대한 시각적 설명을 생성하기 위해 클래스 활성화 지도^[5](Class Activation Map, CAM) 기술이 제안되었다. CAM은 특정 클래스 예측에 기여하는 이미지의 영역을 강조 표시하는 열지도(heat map)를 생성한다. 열지도를 통해 네트워크가 어느 부분을 통해 이미지 분류를 하는지 알 수 있다.

CAM은 딥러닝 모델의 해석 가능성을 향상하기 위해 원격 탐지를 포함한 다양한 영역에 시각화 정보를 제공함으로써 객체의 대략적인 위치를 파악하거나 네트워크가 어느 영역에 집중을 하고 있는지에 대한 정보를 아는데 사용된다.

최근에는 Domain adaptive 방식이나 Transformer를 활용한 SAR 선박 탐지 방식도 연구되고 있다. Yanrui^[6]가 제안한 Domain adaptive 방식은 CycleGAN 구조에 SCA attention module을 추가하여 소스 도메인과 타겟 도메인 간의 간격을 줄이는 pseudo 이미지를 만든다. 이후 self attention feature extraction(SFE)를 통해 feature representation을 만들어 SAR 영상의 복잡한 배경에서 받는 영향을 줄인다. 이후 업샘플링과 context aggregation module을 활용한 lightweight feature fusion and feature enhancement(LFE)를 통해 전체적인 공간 정보를 학습하여 소형 선박의 낮은 해상도와 feature를 해결하여 소형 선박 표적의 탐지 효과를 향상한다. Zhuhao^[7]가 제안한 Transformer를 활용한 방식은 Swin Transformer와 YOLOv5를 결합한 추출기를 통해 local과 global 정보를 인코딩하여 복잡한 배경과 타겟의 차이점을 강조하였고, Information Enhancement Module(IEM)과 Feature Refinement Module(FRM)을 통해 공간 컨텍스트 정보를 학습하고 다중 스케일의 정보를 합쳐 소형 및 다중 스케일 선박의 feature를 개선하여 탐지 효과를 향상한다.

본 논문에서는 SAR 이미지에서 선박 탐지를 위한 Faster-RCNN과 CAM의 융합 기술을 제안한다. 제안된 기술은 정확한 객체 탐지에서의 Faster-RCNN의 장점과 시각적 해석 가능성에서 CAM의 장점을 결합하는데 Faster-RCNN 알고리즘은 관심 영역(Region of Interest, ROI)을 탐지하는 데 사용되며, CAM 기법은 선박 탐지에 기여한 이미지의 영역을 강조하는 CAM을 생성하는 데 사용된다. 우리는 ROI와 CAM을 결합하여 선박 탐지 결과를 개선한다. 제안된 기술은 기존 Faster-RCNN을 사용하였을 때보다 CAM을 통해 잘못 예측한 ROI를 제거함으로써 재현율^[8](Recall)은 유지하면서 정밀도(Precision)를 올려 성능을 높인다. 이는 탐지할 선박들은 그대로 탐지하면서 오탐지된 선박을 효과적으로 제거함을 나타낸다. 본 논문에서 제안하는 방식이 기존의 방식보다 좋은 정확도를 가진 SAR 선박 탐지 모델이 되는 성과를 보여주었다. 본 논문은 서론

에 이어 “II. 연구 배경”에서 객체 탐지와 CAM에 대한 설명을 하고 “III. 제안하는 방법”에서 기존 검출기와 제안하는 방식의 비교 및 사용한 네트워크의 세부 동작 과정을 말하고 “IV. 실험결과 및 고찰”에서 실험에 사용된 데이터 셋과 실험환경 세부사항 및 실험 결과를 나타낸다. 이후 “V. 결론”에서 본 논문이 제안하는 방식에 대한 정리를 한다.

II. 연구 배경

1. 객체 탐지(Object Detection)

객체 탐지^[9]기술은 이미지와 비디오에서 물체를 자동으로 탐지하고 위치를 찾는 것을 목표로 하는 중요한 어플리케이션 중 하나이다. 객체를 정확하게 식별하고 현지화하는 기능은 자율 주행 차량, 감시 시스템 및 의료 영상을 포함한 많은 실제 애플리케이션^{[10][11]}에서 중요한 작업이다. 이는 일반적으로 새로운 보이지 않는 이미지에서 객체를 정확하게 식별하고 위치를 찾기 위해 라벨이 있는 이미지의 대규모 데이터 세트에 대해 훈련된 딥러닝 알고리즘을 사용한다. 세부적으로 객체 탐지에는 이미지 또는 비디오에서 객체의 존재를 식별하고 위치를 확인하는 두 가지 주요 작업이 포함된다. 이를 위해 객체 탐지 알고리즘은 일반적으로 영역 제안(Region proposal), 특징 추출(Feature Extraction) 및 분류(Classification)와 같은 기술을 조합하여 사용한다.

영역 제안은 이미지에서 후보 객체 위치 집합을 생성하는 기술이다. 그것은 객체를 포함할 가능성이 있는 이미지의 영역을 식별하여 객체 탐지를 위한 검색 공간을 줄이는 것을 목표로 한다. 영역 제안이 없으면 객체 탐지 알고리즘은 이미지의 모든 픽셀을 검사해야 하며, 이는 계산 비용이 많이 들고 실제 애플리케이션에서 적용하게 되면 실용적이지 않다.

영역 추출을 위해 선택적 검색^[12](Selective Search), 지역 기반 컨볼루션 신경망^[13](R-CNN)이 있다. 선택적 검색은 색상, 질감 및 가장자리와 같은 낮은 수준의 이미지 기능을 결합하고 이미지를 더 작은 영역으로 나눈 다음 유사한 영역을 점차 병합하여 더 큰 영역을 형성하여 객체와의 유사

성을 기준으로 순위가 매겨지고 최상위 영역은 후보 객체 위치로 사용하는 방식이다. R-CNN은 CNN을 사용하여 일련의 후보 객체 위치를 생성하는 딥러닝 기반 영역 제안 알고리즘으로, 선택적 검색으로 객체 후보를 생성한 다음 CNN을 사용하여 각 제안에서 특징을 추출함으로써 작동하고 분류 알고리즘에 입력되어 객체의 존재와 위치를 결정한다.

특징 추출 과정에서는 CNN을 통해 원시 이미지 픽셀을 객체 탐지에 사용할 수 있는 특징 집합으로 변환하는 작업이 포함되는데 이 과정에서 사용하는 CNN은 일련의 컨볼루션(Convolution) 및 풀링 레이어(Layer)를 통해 이미지에서 관련 기능을 추출하는 방법을 배운 후 객체 클래스와 위치를 결정하는 분류 알고리즘에 입력되어 객체 탐지의 마지막 단계이며 이미지에서 객체 클래스와 해당 위치를 결정한다.

이러한 객체 탐지 알고리즘으로 많은 딥러닝 기반 접근 방식들이 연구되어왔다. 크게 1단계 객체 탐지 방식과 2단계 객체 탐지 방식으로 나뉜다. 1단계 객체 탐지 방식은 객체의 검출과 분류, 그리고 바운딩 박스 회귀를 동시에 수행하는 방식으로, 대표적인 예시 알고리즘으로 YOLO^[14](You Only Look Once)와 SSD^[15](Single Shot Detector)가 있다. YOLO는 이미지를 그리드(grid)로 나누고, 합성곱 연산을 이용한 FCN^[16](Fully Convolutional Network)을 통해 grid 별로 바운딩 박스를 얻어낸 뒤 NMS^[17](Non Maximum Suppression)를 적용한다. YOLO는 박스 별로 분류를 진행하기 때문에 작은 물체나 겹쳐있는 물체를 잘 탐지하지 못하는 단점이 있다. SSD는 YOLO의 작은 객체 탐지의 어려움을 해결하기 위해 다양한 크기의 특징 맵을 활용하고 NMS를 적용한다. 하지만 SSD는 작은 객체를 앞쪽 layer에서 생성된 피쳐 맵을 이용하여 객체 탐지를 하기에 깊이가 충분하지 않기 때문에 YOLO보다 개선되었으나 여전히 작은 크기의 객체 탐지에 대한 어려움이 있다.

2단계 객체 탐지 방식은 분류와 바운딩 박스 회귀를 따로 진행하는 방식으로, 대표적인 예시 알고리즘으로 Faster-RCNN과 FPN^[18](Feature Pyramid Network)이 있다. Faster-RCNN는 기존 Fast R-CNN이 가진 속도 문제를 해결하기 위해 선택적 검색 부분을 영역 제안 네트워크(Region Proposal Network, RPN)으로 대체한 방식이다. FPN은 Top-

down 방식과 Bottom-up 방식을 둘 다 사용하면서 각 방식에서 추출된 피쳐 맵을 lateral connection을 통해 각 단계마다 예측을 하는 방식이다.

2. Class Activation Map, CAM

CNN 기반의 분류 네트워크들은 일반적으로 마지막 FC(Fully Connected) layer를 통해 피쳐 맵을 1차원으로 만들고 이를 소프트맥스(Softmax) 함수를 통해서 각 이미지의 클래스를 분류한다. 하지만 FC layer에 의해서 피쳐 맵이 1차원으로 바뀌게 되면서 각 pixel들에 대한 위치 정보를 잃게 되어 네트워크가 어느 영역을 보고 각 이미지의 클래스를 판별하였는지 알 수가 없다. 따라서 네트워크가 위치 정보를 잃는 것을 막기 위해 FC layer를 GAP(Global Average Pooling)로 변경하여 피쳐 맵에 대한 새로운 가중치를 만들고 마지막에 softmax 함수를 통해 학습을 시킨다. 이를 통해 네트워크는 특정 객체의 전체적인 분포에 대한 정보를 학습할 수가 있다. 이후 생성된 가중치를 convolution layer들과 가중치 합을 하여 이미지 클래스를 어느 영역을 보고 분류하였는지 알 수 있는 열지도를 생성한다.

CAM은 객체 탐지, 이미지 분류와 같은 다양한 연구들에 적용되고 있다. 객체 탐지에서 CAM은 탐지된 물체에 해당하는 이미지의 영역을 강조 표시하는 데 사용된다. 이를 이용해서 라벨이 없는 이미지에도 객체의 위치를 간접적으로 파악하여 약지도학습 기반의 객체 탐지^{[19][20]}를 할 수 있다. 이미지 분류에서 CAM은 분류 결과에 대한 시각적 설명을 생성하는 데 사용되어 모델의 해석 가능성을 향상하는 데 도움이 될 수 있다.

CAM 이후에도 많은 관련 연구가 이어졌다. Grad-CAM^[21]은 CNN에서 피쳐 맵에 대한 미분 값을 이용해 클래스에 대한 분류 기준을 시각화한다. 기존에 네트워크 구조가 맞지 않으면 구조를 바꾸고 나서 발생하는 재학습을 해야 하는 단점을 해결한다. GradCAM++^[22]는 Grad-CAM과 마찬가지로 가중치를 역전파하여 구하지만 그대로 사용하지 않고 수식에 변화를 주어 마지막 피쳐 맵의 양의 편미분 값에 대한 가중 합을 사용하여 하나의 이미지에서 같은 분류의 객체가 반복하여 나와도 구별할 수 있으며 Grad-CAM보다 객체의 전체적인 부분을 커버할 수 있다.

Ablation-CAM^[23]은 Grad-CAM와 달리 미분 값을 사용하지 않고 피쳐 맵을 제거하여 변하는 클래스 점수의 변화를 바탕으로 가중치를 구해 CAM을 생성한다. 이를 통해 미분 값이 0이 되는 현상을 해결한다. LayerCAM^[24]은 클래스 분류 기준을 추출하는 데에 CNN 모델의 각 layer를 이용한다. LayerCAM은 CNN 모델의 모든 layer에서 클래스 분류 기준을 추출하여 이를 시각화한다. 이를 통해 LayerCAM은 모델의 모든 layer에서 클래스 분류에 기여하는 정도를 파악할 수 있어 더 세부적인 클래스 분류 기준을 추출할 수 있다.

III. 제안하는 방법

본 논문에서는 딥러닝 기반 SAR 선박 탐지 네트워크를 통해 선박을 탐지하는 과정 중 Fig. 1 (a)의 Faster-RCNN 네트워크가 예측에 대한 시각적 설명을 제공하지 않는 문제를 Fig. 1 (b)와 같이 입력한 이미지에 CAM을 투영시켜 두 데이터의 특징을 반영하였고 이후 활성화 되지 않는 지역의 ROI를 제거함으로써 향상된 탐지 성능을 가지는 것을 목표로 하는 프레임워크를 제안한다.

1. CAM 생성 네트워크

CAM 생성 네트워크를 위해 심층적인 네트워크 구조를 가지고 있어 이미지 인식 분야에서 매우 효과적인 딥러닝 모델 중 하나인 Resnet^[25]101을 사용한다. 그리고 학습된 Resnet101을 이용하여 LayerCAM을 생성한다. LayerCAM은 피쳐 맵에서 각 공간 위치에 대해 별도의 가중치를 생성하기 위해 역방향 미분 값을 사용한다. 피쳐 맵의 위치에 해당하는 양의 기울기는 가중치로, 음의 기울기를 가진 위치에는 0을 할당한다. f 는 이미지 분류기이고, θ 는 이미지 분류기의 매개변수이고 이미지 I 가 있을 때 객체 클래스가 c 인 객체 분류 예측 점수 y^c 을 (1)과 같이 나타낸다.

$$y^c = f^c(I, \theta) \quad (1)$$

A 를 CNN의 마지막 convolution layer의 피쳐 맵이고 A^k 를 A 의 k 번째 피쳐 맵이라 할 때, A^k 의 각 공간 위치 (i, j)

에 대한 예측 점수 y^c 의 기울기 g_{ij}^{kc} 를 (2)와 같이 구한다.

$$g_{ij}^{kc} = \frac{\partial y^c}{\partial A_{ij}^k} \quad (2)$$

기울기 g_{ij}^{kc} 를 통해 각 공간 위치 (i, j) 에 대한 k 번째 피쳐 맵의 가중치를 (3)과 같이 구한다.

$$w_{ij}^{kc} = ReLU(g_{ij}^{kc}) \quad (3)$$

이 때 양의 기울기만 사용하고 음의 기울기를 가진 위치는 0을 할당해서 ReLU 활성화 함수를 사용한다. 이후 특정 layer의 CAM을 (4)와 같이 피쳐 맵의 각 공간의 활성화 값을 가중치와 곱한다.

$$\hat{A}_{ij}^k = w_{ij}^{kc} \cdot A_{ij}^k \quad (4)$$

최종적으로 \hat{A}^k 는 (5)와 같이 선형적으로 채널 차원을 따라 결합 되어 CAM인 M^c 을 얻고 생성한 CAM과 입력 이미지를 가중치를 두어 서로 결합하는 방안을 제안한다.

$$M^c = ReLU(\sum_k \hat{A}^k) \quad (5)$$

이후 생성한 CAM과 입력 이미지를 결합함으로써 CAM에서 시각적 설명에 대한 정보를 제공한다는 장점과 SAR 데이터 셋 특성상 타겟이 배경들보다 더 두드러져서 이미 지 자체로도 시각적 설명에 대한 정보를 제공할 수 있다는 장점을 더한다.

2. SAR 선박 탐지 네트워크

SAR 선박 탐지 네트워크는 Fig. 1 (a)와 같이 Faster-RCNN을 기본 네트워크 구조로 사용한다. Faster-RCNN은 RPN과 탐지 네트워크의 두 단계로 구성된다. RPN 단계에서 네트워크는 관심 대상을 포함할 가능성이 있는 ROI를 생성한다. RPN은 각 지역 제안에 대한 객관성 점수와 회귀 오프셋을 예측하도록 훈련된다. 객체성 점수는 객체를 포함하는 영역 제안의 가능성을 나타내며 회귀 오프셋은 영역 제안의 좌표를 구체화하는 데 사용된다.

이후 SAR 선박 탐지 네트워크의 정확도를 높이기 위해 CAM을 사용하여 ROI를 제한시키는 방법을 제안한다. CAM은 특정 클래스와 가장 관련이 있는 이미지 영역을 강

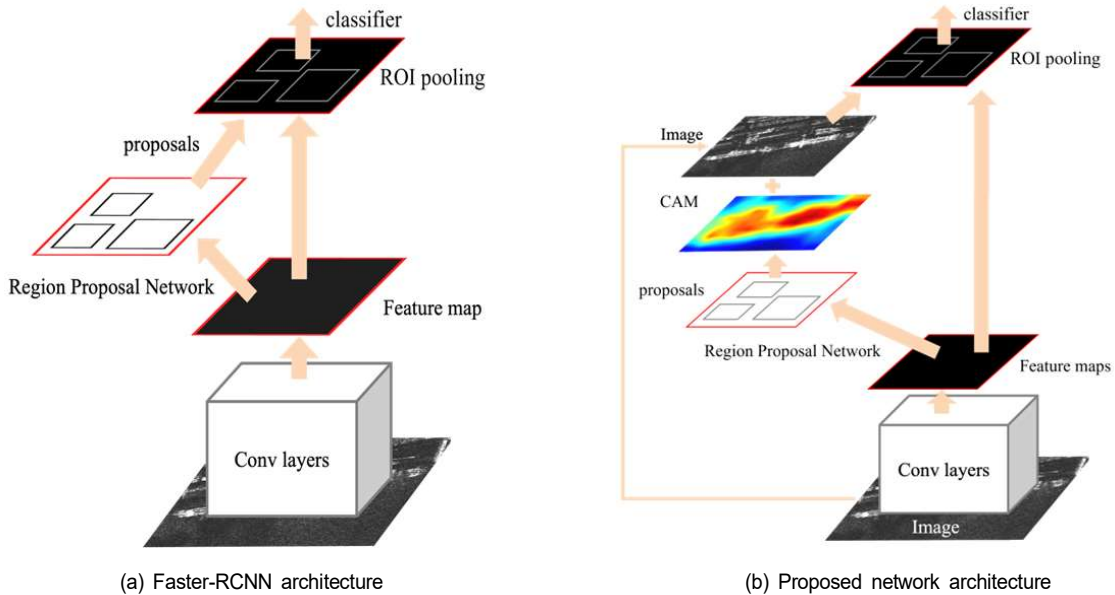


그림 1. 기존 방식과 제안하는 SAR 선박 탐지 네트워크 구조
Fig. 1. The Proposed SAR Ship detection network architecture

조 표시하는 열 지도이기에 Faster R-CNN의 객체 탐지 네트워크 단계에서 CAM은 대상 객체 클래스에 대해 활성화가 높은 영역과 중복되는 제안의 우선순위를 지정하여 ROI 제안의 선택을 세분화하는 데 사용된다. 특히, CAM은 해당 클래스의 가장 차별적인 이미지 영역에 초점을 맞춘 클래스별 ROI 폴링 작업을 계산하는 데 사용된다. 이는 여러 개의 예측된 ROI들 중에서 잘못 예측된 ROI를 줄이고 객체 탐지 네트워크의 정확도를 향상하는 데 도움이 된다.

탐지 네트워크 단계에서 네트워크는 RPN 단계에 의해 생성된 영역 제안을 입력으로 받아들이고 각 영역 제안에 대한 클래스 확률과 회귀 오프셋을 출력한다. 클래스 확률은 특정 클래스의 개체를 포함하는 각 제안의 가능성을 나타내며 회귀 오프셋은 제안의 좌표를 구체화하는 데 사용된다.

IV. 실험결과 및 고찰

본 장에서는 제안한 방법을 통해 NMS 작업 이후 잘못 예측된 RoI를 제거하여 recall을 유지하며 precision을 높인 실험결과를 제시한다. 실험환경에 관한 설명과 사용한 2가지 데이터 셋에 대한 설명 후, Faster-RCNN만 사용하였을

때와 CAM 특징을 반영할 수 있는 이미지와 element-wise 하는 방식과 concat 하는 방식의 실험을 진행하여 본 논문에서 제시한 방법을 적용하였을 때와 비교해 좋은 결과를 얻음을 보인다.

1. 데이터 셋

SAR 데이터 셋으로 선박에 대해서 pascal voc^[26] 데이터 형태로 데이터 크기 가로, 세로 길이와 객체 박스 좌표를 제공하는 HRSID^[27] 데이터 셋과 SSDD^[28] 데이터 셋을 사용하였다. 데이터 셋에 대한 자세한 정보는 Table 1에 작성하였다. 추가로 Fig. 2처럼 CAM 생성 네트워크에서 선박 데이터는 각 이미지에서 GT의 라벨을 이용해 해당 위치의 선박을 기준으로 잘라서 증강하였고, 배경 데이터는 마찬가지로 GT의 라벨을 이용해 해당 위치의 선박을 없애서 바다와 같은 검은색 픽셀로 바꿔서 사용하였다.

2. 실험 환경 및 구현 세부사항

본 연구에서는 실험과정에 2.80GHz CPU와 32GB RAM, NVIDIA GeForce RTX 3080이 장착된 PC와 딥러

표 1. HRSID^[27]와 SSDD^[28] 데이터 셋 세부 정보
 Table 1. Dataset detail for HRSID^[27] and SSDD^[28]

Dataset	Date	Source	Resolution	Image Size	Images/Ships	Train	Test
HRSID	2020.06	Sentinel-1 TerraSAR	0.5m, 1m, 3m	800×800	5,604/16,951	4,483	1,121
SSDD	2017.12	RadarSat-2 TerraSAR	1m - 15m	190 - 668	1,160/2,456	927	232

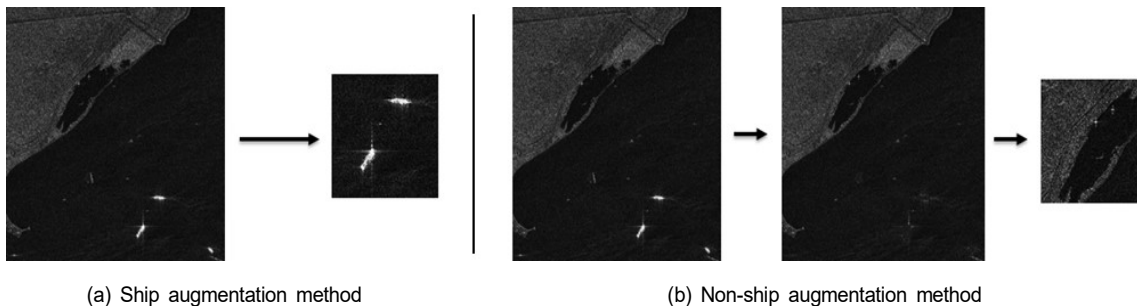


그림 2. 데이터 증강 방식
 Fig. 2. Data augmentation method

닝 라이브러리는 Pytorch^[29] 1.10.2 및 python 3.7 버전으로 실험 환경을 구성하였다.

CAM 생성 네트워크 실험은 타겟이 선박만 있으므로 resnet101의 출력층의 FC layer 채널을 1000에서 2로 변경하여 선박과 배경 클래스로 분류하였다. 이후 모든 데이터를 224x224 크기로 조정하였고, 배치 크기는 32, 시작 학습률을 0.001, SGD (Stochastic Gradient Descent, SGD) 옵티마이저를 기반으로 네트워크를 최적화하였으며 크로스 엔트로피^[30](Cross Entropy) 손실 함수를 사용하여 총 20 에폭(epoch)로 학습을 진행하였다.

SAR 선박 탐지 네트워크 실험은 각 데이터 셋 모두 800 × 800 크기로 조정하였고, 배치 크기는 2로, 시작 학습률은 0.001로, 5 epoch마다 학습률 감소 비율은 0.1로 설정하였다. SAR 선박 탐지기의 백본 네트워크는 resnet101의 파라미터로 초기화를 하였으며 SGD 옵티마이저를 기반으로 네트워크를 최적화하였으며 회귀 오프셋에 대한 한계값과 NMS에 대한 한계값을 0.5로 사용하였다. 데이터 불균형에 강한 포컬(Focal) 손실 함수^[31]를 사용하여 총 20 epoch로

학습을 진행하였다.

3. CAM 생성 네트워크를 활용한 생성 결과

제한한 CAM 생성 네트워크를 통해 얻은 CAM의 정확도를 확인하기 위해 Fig. 3에서 실제 정답(Ground Truth, GT)인 선박 위치를 빨간색 박스로 표현하였고 CAM이 GT 위치에 잘 활성화 되었는지를 비교하였을 때 잘 생성됐는지 확인하였다.

4. SAR 선박 탐지 네트워크를 활용한 선박 탐지 결과

제한한 SAR 선박 탐지 네트워크의 효율성을 보이기 위해 선박 탐지를 Faster-RCNN만 이용하였을 경우, CAM과 이미지를 element-wise하는 방식, CAM과 이미지를 concat하는 방식, NMS 단계에서 생성된 RoI들을 CAM과 이미지를 0.5 가중치를 통해 결합한 출력물로 융합한 경우로 총 4가지 방식으로 구성하여 실험을 진행하였다.

Element-wise하는 방식에서는 이미지와 CAM 모두 3

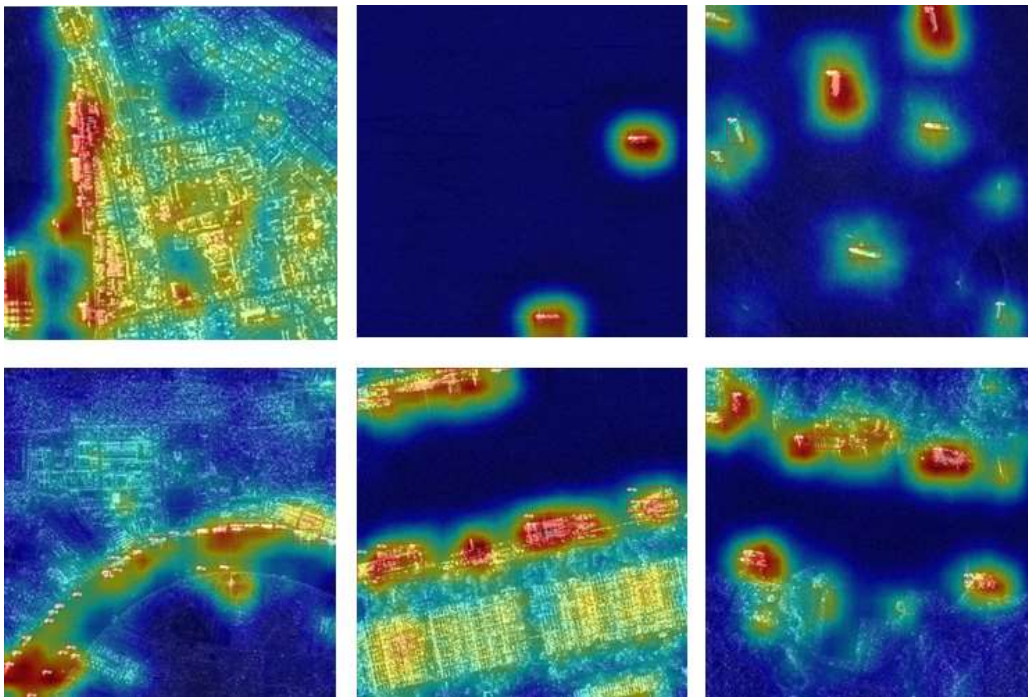


그림 3. CAM 생성 결과
Fig. 3. CAM result

channel을 가졌기에 두 개의 feature를 element-wise하여 각 위치에 있는 pixel의 값을 서로 곱하여 학습을 진행하였다. Concat 하는 방식에서는 이미지와 CAM을 concat하여 3 channel인 이미지 뒤에 3 channel의 CAM을 연결하여 총 6 channel을 가지게 하였고 resnet101의 첫 번째 convolution block의 input channel을 3에서 6으로 변경하여 학습을 진행하였다.

그리고 융합할 때는 표적의 크기를 Small($0 < \text{area} < 32^2$), Medium($32^2 \leq \text{area} < 96^2$), Large($96^2 \leq \text{area}$)로 나눈 후 Table 2와 Table 3처럼 각 크기마다 다른 한계점을 정해 그 값을 넘기지 못하면 box를 제거하였다.

표 2. HRSID^[27] 데이터 셋 CAM 가중치
 Table 2. CAM weight for HRSID^[27] dataset

	Small	Medium	Large
Threshold	0.842	0.882	0.941

표 3. SSDD^[28] 데이터 셋 CAM 가중치
 Table 3. CAM weight for SSDD^[28] dataset

	Small	Medium	Large
Threshold	0.826	0.875	0.937

선박 탐지 성능은 Average Precision^[32](Ap)를 사용하였

고 세부 척도로 precision과 recall을 비교하였다. precision이란, 모델이 참이라고 분류한 것 중에서 실제 참인 비율을 의미한다. Recall은 실제 참인 것 중에서 모델이 참이라고 예측한 것의 비율을 의미한다. 특히 precision이 높을수록 실제 타겟만 타겟으로 예측할 확률이 높아지기에 본 논문에서는 precision을 높이며 선박 탐지 네트워크가 보다 더 정확해질 수 있도록 한다. Table 4,5는 각각 HRSID와 SSDD에 대한 여러 실험 선박 탐지 성능을 나타내며, 두 데이터 셋 모두 Faster-RCNN만 사용한 baseline 방식과 이미지와 CAM의 element-wise 방식 그리고 concat 방식과 비교하였을 때 CAM과 이미지를 융합하여 사용하여 pixel의 최댓값 한계점을 높였을 때 recall은 소폭 하락하면서 precision이 상승하면서 ap가 상승하였다. 또한 Fig. 4의 (c), (d)를 비교하였을 때 첫 번째 예시는 107개, 76개로 줄어들었고, 두 번째 예시는 113개, 80개로 줄어드는 것으로 보아 CAM과 이미지 융합 방식을 통해 네트워크가 잘못된 ROI를 잘 제거하는 결과를 볼 수 있다. 또한 NMS 단계 이후 CAM을 통해 잘못 예측된 ROI를 줄임으로써, NMS 이후 소요시간이 HRSID 데이터 셋에 대해서는 3.99초에서 1.732초로, SSDD 데이터 셋에 대해서는 0.84초에서 0.38초로 약 45퍼센트 감소하는 효과를 얻었다.

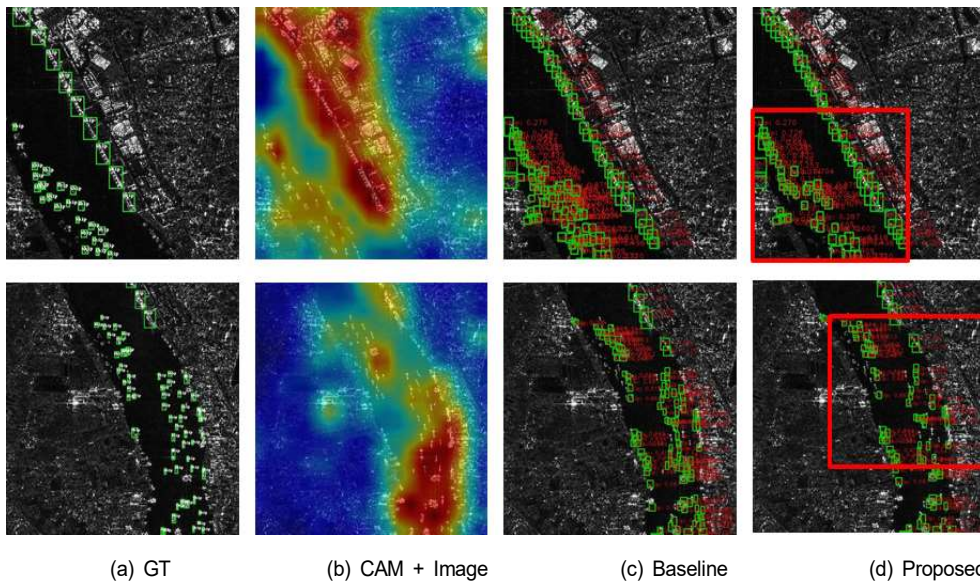


그림 4. SAR 선박 탐지 결과 비교
 Fig. 4. Comparison about SAR ship detection result

표 4. HRSID 데이터 셋 선박 탐지 결과

Table 4. Ship detection result for HRSID dataset

	Ap	Precision	Recall
Baseline	0.7706	0.1121	0.8093
Element-wise	0.5704	0.0924	0.6503
Concat	0.5796	0.0930	0.6537
Proposed	0.7712	0.2303	0.7906

표 5. SSDD 데이터 셋 선박 탐지 결과

Table 5. Ship detection result for SSDD dataset

	Ap	Precision	Recall
Baseline	0.8775	0.1109	0.9495
Element-wise	0.4294	0.0707	0.6044
Concat	0.1720	0.0467	0.4428
Proposed	0.8808	0.2383	0.9265

V. 결 론

본 논문에서는 Faster-RCNN 기반 SAR 영상 선박 탐지 네트워크에 CAM을 융합하여 성능을 개선하는 기법을 제안하였다. HRSID와 SSDD 두 데이터 셋에 대해서 Resnet101을 기본 네트워크로 학습하여 LayerCAM 방식을 이용해 더 나은 CAM을 생성하였고, Faster-RCNN의 RPN 단계에서 추출한 ROI에 대해서 생성한 CAM을 적용하여 ROI 내 pixel 값이 표적의 크기에 따라 정한 한계값을 넘어가지 못하면 제거하였다. 이를 통해 잘못 예측된 ROI를 제거함으로써 recall이 조금 감소하였으나 그에 반해 precision이 상승하여 SAR 영상 선박 탐지기의 성능이 향상됨을 확인하였다. 복잡한 구조를 추가하는 것 없이 NMS 단계 뒤에 CAM을 적용만 하는 것으로 성능 상승을 얻을 수 있었다. 따라서 본 논문에서 제안하는 방식은 NMS 단계를 사용하는 일반적인 다른 detector들에도 모두 활용될 수 있다.

참 고 문 헌 (References)

- [1] J. C. Curlander and R. N. McDonough, Synthetic aperture radar, Vol. 11. Wiley, New York, 1991.
- [2] A. Farina and F. A. Studer, "A review of CFAR detection techniques in radar systems," *Microwave Journal*, Vol. 29, pp. 115-118, 1986.
- [3] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," arXiv preprint arXiv:1511.08458, 2015. doi: <https://doi.org/10.48550/arXiv.1511.08458>
- [4] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Proceeding of Advances in neural information processing systems*, 2015.
- [5] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," *Proceedings of the IEEE conference on computer vision and pattern recognition.*, 2016. doi: <https://doi.org/10.1109/CVPR.2016.319>
- [6] Y. Yang, J. Chen, L. Sun, Z. Zhou, Z. Huang, and B. Wu et al, "Unsupervised Domain-Adaptive SAR Ship Detection Based on Cross-Domain Feature Interaction and Data Contribution Balance," *Remote Sensing*, Vol. 16, No. 2, pp. 420, 2024. doi: <https://doi.org/10.3390/rs16020420>
- [7] Z. Lu, P. Wang, Y. Li, and B. Ding, "A New Deep Neural Network Based on SwinT-FRM-ShipNet for SAR Ship Detection in Complex Near-Shore and Offshore Environments," *Remote Sensing*, Vol. 15, No. 24, pp. 5780, 2023. doi: <https://doi.org/10.3390/rs15245780>
- [8] M. Buckland and F. Gey, "The relationship between recall and precision," *Journal of the American society for information science*, Vol. 45, No. 1, pp. 12-19, 1994. doi: [https://doi.org/10.1002/\(SICI\)1097-4571\(199401\)45:1<12::AID-ASIS>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1097-4571(199401)45:1<12::AID-ASIS>3.0.CO;2-L)
- [9] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proceedings of the IEEE*, Vol. 111, No. 3, pp. 257-276, 2023. doi: <https://doi.org/10.1109/JPROC.2023.3238524>
- [10] E. Arnold, O. Y. Al-Jarrah, M. Dianati, S. Fallah, D. Oxtoby, A. Mouzakitis, "A survey on 3d object detection methods for autonomous driving applications," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 20, No. 10, pp. 3782-3795, 2019. 3782-3795. doi: <https://doi.org/10.1109/TITS.2019.2892405>
- [11] Y. Liu, Z. Ma, X. Liu, S. Ma, and K. Ren, "Privacy-preserving object detection for medical images with faster R-CNN," *IEEE Transactions on Information Forensics and Security*, Vol. 17, pp. 69-84, 2019. doi: <https://doi.org/10.1109/TIFS.2019.2946476>
- [12] J. R. Uijlings, K. E. Van De Sande, T. Gevers, and A. W. Smeulders., "Selective search for object recognition," *International journal of computer vision*, Vol. 104, pp. 154-161, 2013. doi: <https://doi.org/10.1007/s11263-013-0620-5>
- [13] R. Girshick, J. Donahue, T. Darrell, J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014. doi: <https://doi.org/10.1109/CVPR.2014.81>
- [14] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016. doi: <https://doi.org/10.1109/CVPR.2016.91>

- [15] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," *Proceeding of the European Conference on Computer Vision*, 2016.
doi: https://doi.org/10.1007/978-3-319-46448-0_2
- [16] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015.
doi: <https://doi.org/10.1109/CVPR.2015.7298965>
- [17] A. Neubeck and L. Van Gool, "Efficient non-maximum suppression," *Proceedings of the international conference on pattern recognition*, 2006.
doi: <https://doi.org/10.1109/ICPR.2006.479>
- [18] T. Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017.
doi: <https://doi.org/10.1109/CVPR.2017.106>
- [19] Y. Yang, Z. Pan, Y. Hu, and C. Ding, "PistonNet: Object separating from background by attention for weakly supervised ship detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 15, pp. 5190-5202, 2022.
doi: <https://doi.org/10.1109/JSTARS.2022.3184637>
- [20] F. Gu, H. Zhang, C. Wang, and B. Zhang, "Weakly supervised ship detection from SAR images based on a three-component CNN-CAM-CRF model," *Journal of Applied Remote Sensing*, Vol. 14, No. 2, pp. 26506, 2020.
doi: <https://doi.org/10.1117/1.JRS.14.026506>
- [21] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," *Proceedings of the IEEE international conference on computer vision*, 2017.
doi: <https://doi.org/10.1109/ICCV.2017.74>
- [22] A. Chattopadhyay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, "Grad-cam++: Generalized gradient-based visual explanations for deep convolutional network," *Proceedings of the IEEE winter conference on applications of computer vision*, 2018.
doi: <https://doi.org/10.1109/WACV.2018.00097>
- [23] S. Desai and H. G. Ramaswamy, "Ablation-cam: Visual explanations for deep convolutional network via gradient-free localization." *Proceedings of the IEEE winter conference on applications of computer vision*, 2020.
doi: <https://doi.org/10.1109/WACV45572.2020.9093360>
- [24] P. T. Jiang, C. B. Zhang, Q. Hou, M. M. Cheng, and Y. Wei, "Layercam: Exploring hierarchical class activation maps for localization," *IEEE Transactions on Image Processing*, Vol. 30, pp. 5875-5888, 2021.
doi: <https://doi.org/10.1109/TIP.2021.3089943>
- [25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.
doi: <https://doi.org/10.1109/CVPR.2016.90>
- [26] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision*, vol. 88, pp. 303-338, 2010.
doi: <https://doi.org/10.1007/s11263-009-0275-4>
- [27] S. Wei, X. Zeng, Q. Qu, M. Wang, H. Su, and J. Shi, "HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation," *Ieee Access*, Vol. 8, pp. 120234-120254, 2020.
doi: <https://doi.org/10.1109/ACCESS.2020.3005861>
- [28] T. Zhang, X. Zhang, J. Li, X. Xu, B. Wang, X. Zhan, Q. Xu, X. Ke, T. Zeng, H. Su, I. Ahmad, D. Pan, C. Liu, Y. Zhou, J. Shi, and S. Wei, "SAR ship detection dataset (SSDD): Official release and comprehensive data analysis," *Remote Sensing*, Vol. 13, No. 18, pp. 3690, 2021.
doi: <https://doi.org/10.3390/rs13183690>
- [29] Paszke, Adam, et al. "Pytorch: An imperative style, high-performance deep learning library," *Proceedings of Advances in neural information processing systems*, 2019.
- [30] Z. Zhang and S. Mert, "Generalized cross entropy loss for training deep neural networks with noisy labels," *Proceedings of Advances in neural information processing systems*, 2018.
- [31] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," *Proceedings of the IEEE international conference on computer vision*, 2017.
doi: <https://doi.org/10.1109/ICCV.2017.324>
- [32] P. Henderson and V. Ferrari, "End-to-end training of object class detectors for mean average precision," *Proceedings of Asian Conference on Computer Vision*, 2016.
doi: https://doi.org/10.1007/978-3-319-54193-8_13

저 자 소 개



김 병 준

- 2016년 ~ 2022년 : 한국항공대학교 소프트웨어학과 학사
- 2022년 ~ 2024년 : 한국항공대학교 인공지능학과 석사
- 2024년 ~ 현재 : 한화시스템 우주연구소 연구원
- ORCID : <https://orcid.org/0009-0002-8906-1004>
- 주관심분야 : 컴퓨터비전, 객체 탐지, 위성 영상 처리

저 자 소 개



문 호 이

- 2014년 : 이화여자대학교 전자공학과 학사
- 2014년 ~ 2016년 : 한국과학기술원 전기 및 전자공학과 석사
- 2016년 ~ 현재 : 한화시스템 레이더연구소 전문연구원
- ORCID : <https://orcid.org/0000-0003-4804-0880>
- 주관심분야 : 레이더 신호처리, 합성 개구 레이더



이 재 석

- 2008년 : 고려대학교 전파통신공학과 학사
- 2008년 ~ 2015년 : 고려대학교 컴퓨터-전파통신공학과 박사
- 2015년 ~ 2015년 : 서울대학교 뉴미디어통신공동연구소 선임연구원
- 2015년 ~ 2017년 : 대구경북과학기술원 정보통신융합공학전공 리서치펠로우
- 2017년 9월 ~ 현재 : 한화시스템 레이더연구소 전문연구원
- ORCID : <https://orcid.org/0000-0002-4337-8711>
- 주관심분야 : 레이더 신호처리, 레이더 시스템, 압축 센싱, 딥러닝



김 선 옥

- 2009년 ~ 2014년 : 연세대학교 전기전자공학과 학사
- 2014년 ~ 2019년 : 연세대학교 전기전자공학과 박사
- 2019년 ~ 2021년 : 연세대학교 박사후연구원
- 2021년 ~ 현재 : 한국항공대학교 인공지능학과 조교수
- ORCID : <https://orcid.org/0000-0002-9665-4214>
- 주관심분야 : 컴퓨터비전, 인공지능, 위성 영상처리