

다중 프레임 레이더 이미지를 활용한 해상 동적 객체 검출 성능 향상

Multi-frame application for RADAR Image based Maritime Object Detection

박진범¹ · 김용진¹ · 강산하¹ · 김한근[†]

Jinbum Park¹, Yongjin Kim¹, Sanha Kang¹, Hanguen Kim[†]

Abstract: Recent deep learning advancements have outperformed traditional methods in tasks like image classification, object detection, and semantic segmentation. In marine settings, various sensors detect environmental features and obstacles: cameras and LiDAR are commonly used for short-range mapping, while RADAR is preferred for long-range detection due to its resilience across large areas and in adverse weather. However, most research and available datasets focus on camera and LiDAR data, with limited data and studies on maritime RADAR. Additionally, RADAR images typically contain high noise and clutter, complicating the development of training datasets. While RADAR is valuable for detecting marine objects, its noise levels demand specialized approaches beyond conventional vision-based detection. Previous work applies semantic segmentation to RADAR images to distinguish objects, noise, and land, but uses only single-frame data. This paper proposes a novel method using sequential RADAR images to enhance object detection in marine environments. By leveraging temporal information and minimizing preprocessing, our approach significantly improves accuracy.

Keywords: Maritime, Radar, Semantic Segmentation, Deep Learning, Computer Vision

1. 서론

최근 딥러닝 기술을 기반으로 한 주변 환경 인식 및 감지에 대한 연구가 활발히 진행되고 있다. 딥러닝은 이미지 분류, 객체 감지, 의미론적 분할 등 다양한 작업에서 기존의 전통적인 기술에 비해 뛰어난 성능을 보여주고 있다. 특히, 자율주행 시장의 성장에 따라 차량 운행 이미지와 라이다(LiDAR)를 활용한 환경 인식에 널리 적용되고 있다. 이와 유사하게, 자율 운항 선박에 대한 수요도 증가하면서 자율 운항 선박 항법 기술 시장이 자율 주행 자동차 시장만큼 성장할 것으로 기대하고 있다.

이에 따라 딥러닝을 자율 운항 선박 항법 기술에 적용하려는 연구가 최근 활발히 진행되고 있다^[1].

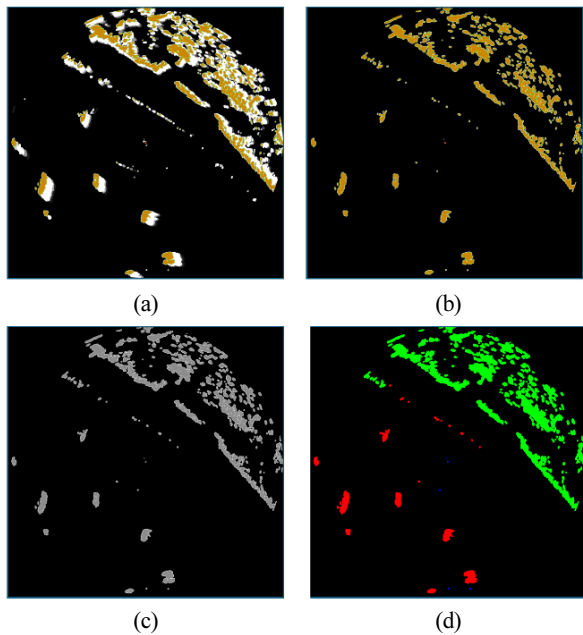
해상 환경에서는 주변 환경과 장애물에 대한 감지 및 인식을 위해 여러 가지 센서를 사용하고 있다. 해상 환경에서 선박의 항법을 보조하고 주변 객체와의 충돌을 회피하기 위해 여러 종류의 센서들이 각각의 목적에 맞게 사용되어 왔다. 카메라와 라이다는 시각 센서로서 근거리 맵핑의 목적으로 사용되며, 레이더(RADAR)는 원거리 객체에 대한 확인을 위해 주로 사용되고 있다^[2]. 특히 레이더 시스템은 넓은 관측 범위를 가지며 카메라나 라이다와 달리 날씨 조건에 영향을 받지 않기 때문에 해상에서 많이 활용된다^[3,4]. 하지만 대부분의 딥러닝 연구는 주로 카메라와 라이다를 활용한 컴퓨터 비전에 초점을 맞추고 있으며, 특히 자율주행 분야에서 연구가 활발히 진행되어 왔기에 이들 센서를 활용한 학습용 데이터셋도 많이 존재한다^[5,6]. 반면 선박에 장착된 레이더 이미지를 기반으로 한 공개 레이더 이미지 데이터셋이 거의 없어 해상 레이더 이미지를 기반으로 한 딥러

Received : Oct. 28. 2024; Revised : Jan. 2. 2025; Accepted : Mar. 28. 2025

※ This work was supported in part by the Development of a Situational Awareness System for Preventing Collisions and Accidents of Autonomous Ships funded by the Ministry of Trade, Industry and Energy (MOTIE), South Korea, under Grant 20011722.

1. Researcher, AI Research Team, Seadronix Corp, Seoul, Korea (jin.park, yongjin, ksana@seadronix.com)

† Associate Researcher, Corresponding author: CTO, Seadronix Corp, Seoul, Korea (hank05@seadronix.com)



[Fig. 1] Sample images of RADAR : (a) Raw, (b) pre-processed input RGB, (c) Grayscale, and (d) ground truth. (a) contains noise, (b) and (c) are filtered images with less noise

닝 기술에 대한 연구는 부족하다^[1].

레이더 이미지는 [Fig. 1]과 같이 노이즈가 많다는 특성을 가지고 있다. 레이더 시스템은 원거리 물체를 탐지할 수 있다는 장점이 있지만, 많은 노이즈와 함께 다른 물체에 반사되어 발생하는 신호인 클러터를 포함하는 단점이 있다. 특히 비, 눈, 해수면 반사 등 다양한 요인들이 레이더 신호에 혼입되기 때문에 해석에 전문성을 요한다^[1]. 레이더가 표적 탐지에 주로 사용되지만, 노이즈가 많아 기존의 영상 및 라이다 도메인에서 주로 사용되는 객체 검출(object detection) 기법과 다르게 의미론적 분할(semantic segmentation) 알고리즘을 사용하여 객체, 노이즈, 육지 등을 레이더 영상을 구분하는 시도^[1,2]가 되어왔다. 해당 시도들은 레이더 데이터가 가지고 있는 단일 프레임의 정보로만 객체를 탐지하였다. 하지만, 시간 정보의 사용은 vision task에서 추가적인 정보를 활용하는 매우 자연스러운 시도이며 low level vision task^[7-12]에서 부터 high level vision task^[13-15]까지 다양한 computer vision task에서 다중 프레임을 사용하는 방법론이 제안되었고 단일 프레임에 비해 의미 있는 성능 향상을 보여주었다.

이 논문에서는 연속되는 프레임의 레이더 이미지의 사용을 통해 해상 환경에서의 모델의 객체 인식 성능을 향상시키는 새로운 방식을 제안한다. 본 연구는 레이더 이미지의 전처리 과정을 최소화하면서 시간 정보를 함께 활용하는 방식으로 해상 딥러닝 의미론적 분할 모델의 성능 향상 가능성을 제안하고 이를 통해 해상 환경에서 자율 운항 선박의 안전성과 효율성을 크게 향상시킬 수 있을 것으로 기대한다.

2. 선행 연구

2.1 레이더 이미지를 활용한 연구

레이더 센서의 특성상 원거리의 객체에 대한 탐지에는 적합하지만, 다양한 요인들이 레이더 신호에 혼입되어 들어오기 때문에, 노이즈가 많고 해상도가 매우 낮다^[2].

SegNet^[2]은 컨볼루션 신경망(CNN)을 사용한 SegNet 아키텍처를 통해 해상 레이더 이미지의 의미론적 분할을 하였다. 해상 객체의 가시성을 높이고 동시에 훈련 중 클래스 불균형 문제를 줄이기 위해 직교 좌표계로 변환하는 전처리 방식을 제안하고 이를 통해 해상 환경에서 딥러닝 프레임워크의 물체 감지 및 분류 성능을 효과적으로 개선하여 레이더 이미지로 물체감지 가능성을 보여주었다.

DPSE-Net^[1]은 BisenetV2^[16]를 기반으로 하여 해상 레이더 이미지로부터 해상 객체, 육지, 노이즈, 배경을 세분화하였다. Squeeze and Excitation^[17]모듈을 통해서 소형 물표에 대한 인식 성능을 향상시켰고, 실시간으로 세분화를 수행할 수 있는 모델을 제안함으로써 U-Net^[18] 및 BisenetV2^[16]와 같은 기존 의미론적 분할 모델과 비교하여 레이더 이미지 세분화에서 우수한 성능을 입증했다. 추가적으로 레이더 센서 특성을 고려한 추적 알고리즘을 제안함으로써 고속 이동 소형 물표에 대한 객체 추적 성능을 실 환경 데이터를 통해 검증하였다.

위의 두 연구 모두 해상 레이더 이미지의 단일 프레임 정보를 기반으로 의미론적 분할을 통해 해상 객체를 감지하고 감지 성능을 개선하는데 중점을 두었다. 그러나 두 연구 모두 시간적 정보를 활용하지 않고 단일 프레임만 활용했다는 한계점을 가지고 있다.

2.2 다중 프레임을 활용한 연구

2.2.1 Low-level Vision Task에서의 성능향상

다중 프레임의 시간적 정보와 연속된 프레임 간의 상호작용을 활용하는 방법은 정확한 이미지 복원과 개선을 가능하게 한다. 노이즈 제거^[7,9], 초해상도^[8,11,12]에서 다중 프레임을 활용한 방법들이 단일 프레임을 사용하는 기존의 전통적인 방법들보다 우수한 성능을 보여주었다.

MF2F^[9]는 비디오가 가지고 있는 미지의 노이즈에 대해서 기존 단일 이미지를 사용하여 노이즈를 제거하는 F2F (Frame to Frame) 방식을 다중 프레임 네트워크로 확장 노이즈 제거 성능향상을 보였다.

다중 프레임은 장면에 대한 더 많은 정보를 제공하며, 단일 프레임의 정보 뿐만 아니라 프레임 간의 optical-flow와 같은

시공간적 정보도 존재한다. 기존 연구들은 이러한 시공간적 정보를 효과적으로 활용하였다. Liao et al^[20]은 옵티컬 플로우를 활용해 고해상도 후보 이미지를 생성하고 이를 CNN으로 결합하였으며, CVSRnet^[21]은 Druleas 알고리즘^[22]을 통해 프레임 사이의 모션 보상을 처리하고 CNN을 사용해 연속적인 프레임을 입력으로 받아 고해상도 이미지를 예측하였다.

다중프레임의 초해상도가 단일프레임에 비해 초해상도에 높은 성능향상을 보인다는 점은 해당 방법론이 안면인식^[23], 위성 이미지^[24], 의료 이미지^[25] 등의 다양한 분야에서 적용되면서 그 효과를 다시 한번 검증하였다. 단일 프레임 모델에 비해 다중 프레임 모델의 대표적이 단점은 추가적인 정보를 줄 수 있는 시간 연속적인 학습데이터가 있어야 한다는 점인데, 이러한 단점을 해결하기 위해서 Yanan et al^[12]은 이미지 생성 모델을 활용해서 다중 프레임 이미지를 생성, 초해상도에 활용하였다.

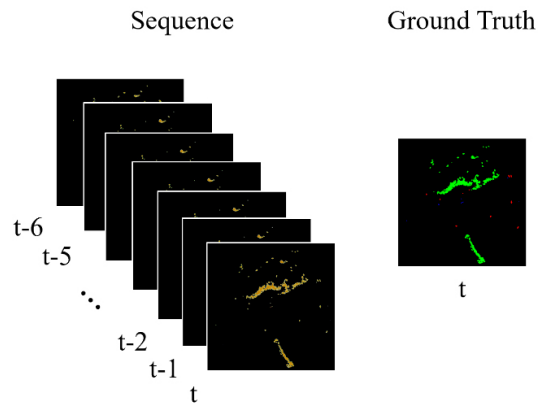
2.2.2 High-level Vision Task에서의 성능 향상

딥러닝을 활용한 high level 비전 작업에서, 다중 프레임 접근법은 시간적 연속성을 통해 얻어지는 맥락 정보를 통합하여 검출^[13,14], 의미론적 분할^[15] 등의 작업에서 보다 정밀한 결과를 도출할 수 있다. 특히, 비디오 기반 객체 추적, 동작 예측과 같은 응용 분야에서 다중 프레임을 활용한 방법들이 단일 프레임 기반 방법들보다 우수한 성능을 보이는 것으로 보고되고 있고, 이러한 선행 연구들은 딥러닝 기반의 high level 비전 시스템에서 다중 프레임 활용의 중요성을 강조하고 있다.

MFSSD^[13]는 비디오 데이터의 시간적 정보를 통합하여 물체 감지 정확도를 향상시킴으로써 기존의 SSD^[26]를 개선하였다. Recurrent 레이어를 추가함으로써 여러 프레임에 걸쳐 특징을 융합하여 추가적인 시공간 정보를 활용할 수 있도록 하였고, 벤치마크 데이터 세트에서 단일프레임 모델에 비해 평균 정밀도를 크게 향상시켰다. 추가적인 프레임을 사용해도 모델의 연산속도를 유지하여 단일 프레임 SSD^[26]의 속도를 유지, 초당 50프레임의 속도로 실시간 물체 감지가 가능하였다.

T2-YOLOv5^[14]는 작은 객체에 대한 검출 성능을 향상시키기 위한 방법으로 다중프레임을 사용하는 방식을 도입하였다. 시간 정보를 다양하게 포함하는 두가지 프레임워크를 제안하였습니다. 하나의 이미지의 RGB 색상 이미지를 사용하는 방식 대신 세계의 연속적인 프레임의 흑백 이미지를 사용하는 방식과 추가적으로 프레임 간의 차이를 사용하는 방식을 통해서 시간 정보를 추가하였고, 하나의 프레임을 사용하는 방식보다 좋은 검출 성능을 보여주었다.

Accel^[27]는 현재 프레임의 주변 프레임을 reference branch를 통해서 연결하고, update branch를 통해서 현재 프레임의 정보와 주변 프레임의 정보를 교정하고 고정하는 방식으로 의미론적 분할 마스크를 예측했다. Accel은 단일 프레임 기반의 DeepLab-101^[28]



[Fig. 2] Configuration of sequential data. Single sequence contains seven sequential input raw images and one ground truth labeled image

보다 더 높은 정확도를 달성했으며, 이전의 단일프레임 기반의 모델보다 크게 향상된 정확도를 보여주었다. 효율적으로 다중 프레임을 의미론적 분할에 활용하는 방법론을 보여주었다.

또한, J. Hur 과 S. Roth^[29]는 이전 시간 단계에서의 광학 흐름 추정치를 활용하는 다중 프레임 방식을 통해서 광학 흐름 추정에서 더 좋은 성능을 보이는 모델을 소개하였다. 두 프레임의 정보보다 이전의 연속 프레임을 해당 분야에서 처음으로 적용하였으며 자율 주행 벤치마크 데이터셋인 KITTI^[5]에서 기존의 방식보다 더 높은 성능을 보였다.

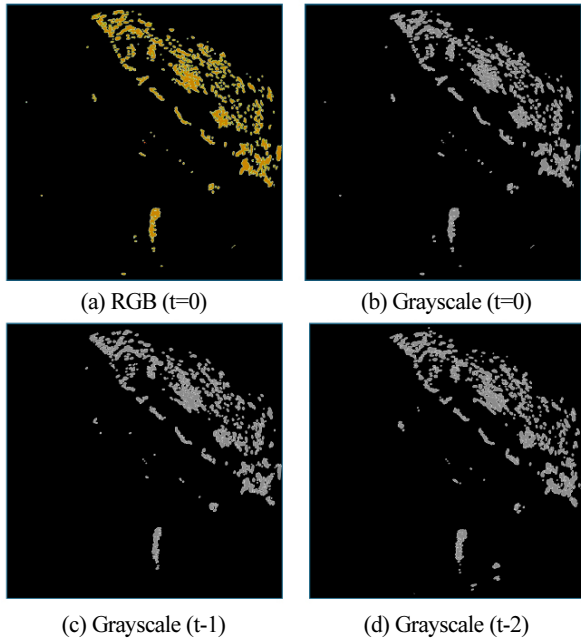
본 논문은 레이더 이미지를 활용한 의미론적 분할 검출 분야에서 기존 연구의 한계를 보완하기 위해 시간적 정보를 적극적으로 활용하는 접근법을 제안한다. 레이더 다중 프레임 데이터를 활용하여 시간 변화에 따른 물체 이동 및 지형과 같은 고정 물체에 대한 검출 성능 향상을 기대하고자 한다.

3. 방법론

기존의 해상 세분화 모델에 단일 프레임 대신 시퀀스 레이더 이미지를 입력으로 활용함으로써, 다중 프레임의 적용이 모델 성능에 미치는 영향을 분석한다. 실험에 사용된 VaDA^[30]와 DPSE-Net^[11]은 실시간 처리 성능을 갖춘 해상 세분화 모델들로, Short-Term Dense Concatenate (STDC)^[31] 모듈을 동일한 backbone으로 포함하고 있고 적은 수의 파라미터로도 우수한 세분화 성능을 보여주었다. 방법론에서는 이와 같은 모델을 바탕으로, 레이더 이미지 시퀀스 처리 방법과 다중 프레임 적용 방식을 설명한다.

3.1 레이더 이미지에 대한 시퀀스처리

레이더 이미지를 활용한 의미론적 분할 모델 학습을 위해서



[Fig. 3] Model input image samples (a) single frame model input RGB, (b)-(d), multi frame model input grayscales from t to $t-2$. Instead of using (a) single frame, temporal information is utilized by using (b), (c), and (d)

[Table 1] Label Class and Color Configuration

Class	Color
Background	Black [0, 0, 0]
Land	Green [0, 255, 0]
Objects	Red [0, 0, 255]
Noise	Blue [255, 0, 0]

DPSE-Net^[1]과 같이 각각의 픽셀들을 라벨링 하였으며, [Table 1]과 같이 레이더 영상의 픽셀들은 객체, 지형, 소음, 배경의 4 가지 클래스로 분류하였다. 레이더의 원시 이미지는 라벨작업이 힘들기 때문에, 선박에 설치된 GPS 및 IMU 센서 정보를 이용하여 선박의 회전 움직임과 관계없이 입력 레이더 영상을 진북으로 고정하는 방식으로 정렬하고, 지도 및 AIS센서를 통해 추정되는 객체 정보를 이용하여 라벨링을 하였다^[1].

1 Hz의 주기로 설정되어 출력되어진 레이더 이미지를 누적하여 레이더 시퀀스 데이터로 처리하였다. 각 시퀀스는 총 7장의 연속되는 레이더 원시 이미지 $T = \{t, t-1, t-2, t-3, \dots, t-6\}$ 를 포함하며 시퀀스 당 총 7개의 이미지를 포함하고 있다. 현재 프레임 (t)의 라벨 이미지를 해당 시퀀스의 Ground-Truth [Fig. 2]로 활용하였다.

3.2 다중프레임 의미론적 분할

연속적인 다중 프레임을 모델에 적용하는 방법은 입력 단계, 모델 중간 단계, 출력 단계에서의 결합을 통해 추가적인 정보의

효과를 검증할 수 있다. 본 연구에서는 식 (1)의 단일 프레임의 color 이미지와 연속적인 grayscale 이미지를 입력으로 식 (2)와 같이 사용하여 다중프레임 시간 정보를 입력단계에서 결합하는 효과를 확인하였다 [Fig. 3]. 이를 통해 추가적인 모듈의 도입 없이 객체의 검출 성능을 향상 시키는 방식으로 다중 프레임의 효과를 탐구한다.

$$Pred = Seg(I_t^{RGB}) \tag{1}$$

$$Pred = Seg(I_t^{Grayscale}, I_{t-1}^{Grayscale}, I_{t-2}^{Grayscale}) \tag{2}$$

4. 실험

4.1 실험 설정

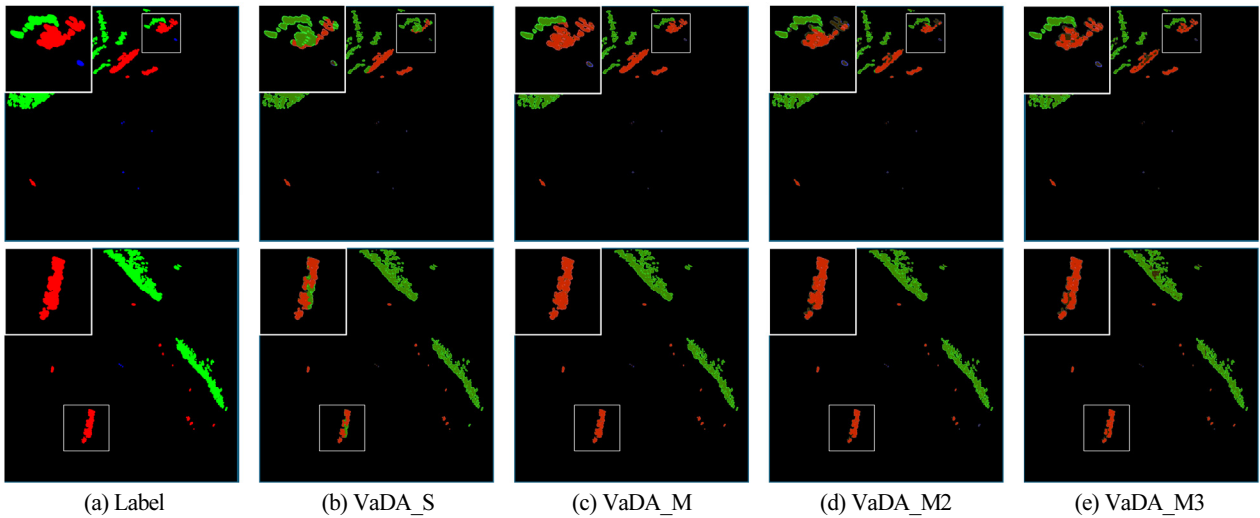
레이더 출력 영상 크기는 2048×2048 픽셀이며, 기존 모델 간의 성능 비교를 위해 입력 영상을 1024×1024 로 크기로 조정하였다. 레이더 영상 획득과 함께 라이다, 카메라, AIS 및 전자해도 센서 데이터를 수신하여 픽셀 단위의 레이블링에 사용하였다. 또한 선박에 설치된 IMU 및 GPS를 사용하여 영상의 방향을 진북방향으로 수정하였다. 개별적인 레이더 센서에 대한 튜닝을 할 수 없기 때문에 일반적인 영상처리 기법을 활용해서 센서에서 입력되는 원시데이터를 전처리 하였다. 전체 학습 데이터셋은 1694개의 시퀀스, 테스트 데이터셋은 350개의 시퀀스로 정하여 실험에 활용을 하였다.

본 논문에서는 레이더 영상 분할에서 다중프레임의 효과를 검증하기 위해 DPSE-Net^[1]과 VaDA^[28]을 활용하여 그 효과를 증명한다. 4.2에서는 단일 프레임과 다중프레임의 성능, 그리고 다중프레임의 시간 간격에 대한 비교 실험을 진행하고, 4.3에서는 비교모델에서의 다중프레임의 적용이 레이더 영상 분할 성능 향상에 도움을 줄 수 있다는 것을 확인한다.

4.2 다중프레임 시간 간격에 대한 실험

다중 프레임의 입력은 프레임 간의 간격에 따라서 VaDA_M, VaDA_M2, VaDA_M3로 정의하였다[Table 2]. 단일 프레임으로 렌더링 된 3채널의 컬러 이미지 대신 3개의 연속적인 프레임의 이미지를 grayscale로 변환하여 모델의 입력으로 넣는 방법을 활용, 모델의 연산 복잡도를 유지하면서 시간 정보를 사용하였다.

첫번째 실험에서는 모두 동일한 시퀀스의 연속적인 프레임을 사용하지만 시간 간격을 다르게 하는 3개의 모델을 통해서 가장 효율적인 프레임 간격을 찾았다. 현재 프레임($t=0$) 기준으로 가장 근접한 두개의 프레임($t-1, t-2$)을 활용한 VaDA_M 모델에서 객체와 지형 클래스의 의미론적 분할 성능이 가장 좋



[Fig. 4] Qualitative segmentation results of the different input frame interval results. (a) ground truth, (b) single frame results, (c) multi frame results ($t, t-1, t-2$), (d) multi frame results ($t, t-2, t-4$) and (e) multi frame results ($t, t-3, t-6$). (b) While there is a fragmentation result in (c), (d), and (e), the segmentation result is rarely fragmented. (c) showed the highest segmentation performance

[Table 2] Definition of experiment options

Model	Single/ Multi	Input Frames
VaDA_S	Single	Image (t , RGB)
VaDA_M	Multi	Image (t , Grayscale), Image ($t-1$, Grayscale), Image ($t-2$, Grayscale)
VaDA_M2	Multi	Image (t , Grayscale), Image ($t-2$, Grayscale), Image ($t-4$, Grayscale)
VaDA_M3	Multi	Image (t , Grayscale), Image ($t-3$, Grayscale), Image ($t-6$, Grayscale)

았다. 동적 객체와 지형의 Intersection over Union (IoU)가 단일 프레임만을 사용한 모델에 비해 약 21% 향상되는 것을 보여주었다. VaDA_M2($t, t-2, t-4$), VaDA_M3($t, t-3, t-6$)의 순서로 프레임 간의 간격이 커질수록 IoU의 향상은 줄어들었다 [Table 3]. 또한, 상대적으로 정적인 신호인 지형(Land)의 경우 단일프레임 대비 프레임 시간 간격의 차이에 따른 성능 향상이 크지 않았지만, 동적인 신호를 갖는 해상 객체(Objects)의 경우 프레임 차이가 적은 VADA_M에서 성능의 향상이 단일프레임의 성능에 비해서 가장 컸다. [Fig. 4]와 같이 동적객체에서 다중프레임을 사용한 모델이 더 높은 세분화 결과를 보여주는 것을 정성적인 결과에서도 확인할 수 있었다.

4.3 범용 모델에서의 다중프레임 활용

다중 프레임의 효과를 검증하기 위해 VaDA^[28]와 함께 DPSE-Net^[1]에서도 그 효과를 확인하였다. 이전 실험에서 가장 높은

[Table 3] Performance comparison with different input frame intervals

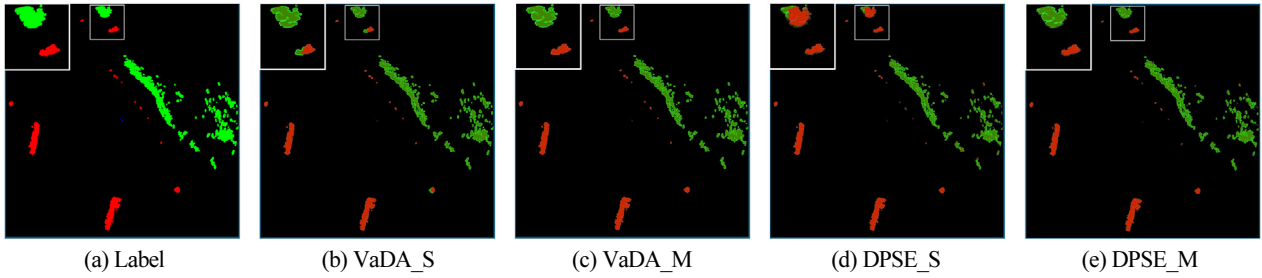
Model	Param	IoU (Objects)	IoU (Land)
VaDA_S	Single	0.6673	0.7208
VaDA_M	Multi	0.8106	0.8764
VaDA_M2	Multi	0.7090	0.8716
VaDA_M3	Multi	0.6824	0.8452

성능 향상을 보였던 VaDA_M의 입력과 같은 연속 프레임을 적용, 단일 프레임 컬러 이미지에서의 DPSE에서의 성능을 비교, 동적 객체와 지형 객체에서 각각 9%와 7%의 성능 향상을 보였다 [Table 4]. 특히 단일 프레임에서 동적객체의 세분화가 부분적으로 다른 클래스로 추론이 되면서 세분화의 결과가 객체를 오인식 하는 결과가 있었는데 다중 프레임을 활용하였을 때 이러한 현상이 줄어들고 원시데이터의 신호가 보존되는 것을 볼 수 있었다[Fig. 5].

5. 결론

본 논문에서는 레이더 이미지 기반 객체 검출의 성능을 향상시키기 위해 레이더 이미지를 시퀀스방식으로 처리하고 연속적인 다중 프레임을 입력으로 활용한 딥러닝 네트워크를 제안하였다. 단일 프레임 레이더 이미지 의미론적 세분화 모델과의 성능 비교를 통해 다중프레임 레이더 이미지의 적용 효과를 검증하였다.

1 Hz 주기로 수집된 연속적인 레이더 이미지를 시퀀스로 처리하여 학습하였으며, 동일 시퀀스 내에서 프레임 간 간격을 조



[Fig. 5] Qualitative comparison between single frame and multi frame input results (a) ground truth, (b) single frame VaDA (c) multi-frame VaDA ($t, t-1, t-2$), (d) single frame DPSE-Net and (e) multi frame DPSE-Net ($t, t-1, t-2$) results. In (b), (d), there was a misrecognition result in the segmentation of dynamic objects, and in (c), (d), it was seen that raw signals were conserved and showed better segmentation results

[Table 4] Performance comparison with multiple baseline radar segmentation models

Model (Param)	Input Frame	IoU (Objects)	IoU (Land)
VaDA S	Image (t , RGB)	0.6673	0.7208
VaDA M	Image (t , Grayscale), Image ($t-1$, Grayscale), Image ($t-2$, Grayscale)	0.8106	0.8764
DPSE S	Image (t , RGB)	0.5316	0.7309
DPSE M	Image (t , Grayscale), Image ($t-1$, Grayscale), Image ($t-2$, Grayscale)	0.5836	0.7826

정하여 최적의 시간 간격을 탐색하였다. 제안된 다중프레임 모델은 기존 단일 프레임 모델에 과거 프레임을 grayscale 변환하여 추가하는 방식을 통해 연산 복잡도를 유지하면서 시간 정보를 효과적으로 활용하였다. 실험 결과, 모든 클래스에서 단일 프레임 모델 대비 다중프레임 모델(VaDA_M)의 성능이 전반적으로 향상되는 것을 확인하였다. 특히, 시간 차이에 따른 레이더 신호의 변화가 적은 “Land” 클래스는 모델간의 성능 차이가 크지 않았지만, 신호 변화가 큰 “Object” 클래스에서는 시간 차이에 따른 성능 변화가 두드러졌다. 이러한 결과는 시간 간격이 짧은 연속 프레임의 활용이 동적인 객체 검출에서 더 효과적임을 보여주며, 다중프레임 기반 접근법이 해상 동적 객체의 세분화 성능 향상에 기여함을 입증하였다.

하지만 레이더 이미지의 낮은 해상도로 인해 여전히 일반 이미지와 같이 높은 세분화 정확도 달성하지 못했다. 과거프레임을 현재의 프레임과 융합하는 방식은 다양하게 존재하지만 본 연구에서는 레이더와 같이 해상도가 낮은 영상에서의 다중프레임의 효과에 집중하였다. 향후 연구에서는 다중프레임을 더 효과적으로 융합할 수 있는 모듈을 제안하고, 융합 과정에서 발생할 수 있는 시간적 동기화 문제를 해결하는 방안을 모색할 예정이다. 또한, 더 나아가 카메라 이미지를 레이더 이미지와 퓨

전 하여 레이더 이미지에서 부족한 공간적 정보를 보완하는 방법 또한 하나의 연구 방향으로 가능하다. 레이더의 거리 정보와 카메라의 해상도 정보를 결합함으로써, 해상에서의 원거리의 작은 객체에 대한 검출 성능을 크게 향상시킬 수 있을 것으로 기대된다. 이를 위해 카메라 이미지와 레이더 이미지가 동시에 취득될 수 있는 센서융합셋을 구성하여, 더욱 정교한 객체 검출 네트워크를 개발할 계획이다.

References

- [1] H. Kim, D. Kim, and S.-M. Lee, “Marine object segmentation and tracking by learning marine radar images for autonomous surface vehicles,” *IEEE Sensors journal*, vol. 23, no. 9, pp. 10062-10070, May, 2023, DOI: 10.1109/JSEN.2023.3259471.
- [2] K. Kim and J. Kim, “Semantic segmentation of marine radar images using convolutional neural networks,” *OCEANS 2019 - Marseille*, Marseille, France, pp. 1-6, 2019, DOI: 10.1109/OCEANSE.2019.8867504.
- [3] R. Vicen-Bueno, R. Carrasco-Álvarez, M. Rosa-Zurera, and J. C. Nieto-Borge, “Sea clutter reduction and target enhancement by neural networks in a marine radar system,” *Sensors*, vol. 9, no. 3, pp. 1913-1936, Mar., 2009, DOI: 10.3390/s90301913.
- [4] Y. Zhou, T. Wang, R. Hu, H. Su, Y. Liu, X. Liu, J. Suo, and H. Snoussi, “Multiple kernelized correlation filters (MKCF) for extended object tracking using X-band marine radar data,” *IEEE transactions on signal processing*, vol. 67, no. 14, pp. 3676-3688, Jul., 2019, DOI: 10.1109/TSP.2019.2917812.
- [5] Y. Liao, J. Xie, and A. Geiger, “KITTI-360: A novel dataset and benchmarks for urban scene understanding in 2D and 3D,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 45, no. 3, pp. 3292-3310, Mar., 2022, DOI: 10.1109/TPAMI.2022.3179507.
- [6] M. Cordts, O. Mohamed, R. Sebastian, R. Timo, E. Markus, B. Rodrigo, F. Uwe, R. Stefan, and S. Bernt, “The cityscapes dataset for semantic urban scene understanding,” *2016 IEEE conference on computer vision and pattern recognition (CVPR)*, Las Vegas, NV, USA, pp. 3213-3223, 2016, DOI: 10.1109/CVPR.2016.350.

- [7] X. Li, Y. Hu, X. Gao, D. Tao, and B. Ning, "A multi-frame image super-resolution method," *Signal Processing*, vol. 90, no. 2, pp. 405-414, Feb., 2010, DOI: 10.1016/j.sigpro.2009.05.028.
- [8] Y. Huang, W. Wang, and L. Wang, "Video Super-Resolution via Bidirectional Recurrent Convolutional Networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 1015-1028, Apr., 2018, DOI: 10.1109/TPAMI.2017.2701380.
- [9] V. Dewil, J. Anger, A. Davy, T. Ehret, P. Arias, and G. Facciolo, "Self-supervised training for blind multi-frame video denoising," *arXiv:2004.06957*, 2021, pp. 1559-1568, DOI: 10.48550/arXiv.2004.06957.
- [10] J. Cai, S. Gu, and L. Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE transactions on image processing*, vol. 27, no. 4, pp. 2049-2062, Apr., 2018, DOI: 10.1109/TIP.2018.2794218.
- [11] M. S. M. Sajjadi, R. Vemulapalli, and M. Brown, "Frame-recurrent video super-resolution," 2018 *IEEE/CVF conference on computer vision and pattern recognition*, Salt Lake City, UT, USA, pp. 6626-6634, 2018, DOI: 10.1109/CVPR.2018.00693.
- [12] Y. Wu and Y. Xu, "Multi-frame super-resolution via generative image model," 2022 *6th International Conference on Video and Image Processing*, pp. 1-5, 2022, DOI: 10.1145/3579109.3579115.
- [13] A. Broad, M. Jones, and T.-Y. Lee, "Recurrent multi-frame single shot detector for video object detection," *British machine vision conference (BMVC)*, pp. 1-12, 2018, [Online], <http://bmvc2018.org/contents/papers/0309.pdf>.
- [14] C. W. Corsel, M. van Lier, L. Kampmeijer, N. Boehrer, and E. M. Bakker, "Exploiting temporal context for tiny object detection," 2023 *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, HI, USA, pp. 1234-1243, 2023, DOI: 10.1109/WACVW58289.2023.00013.
- [15] P. W. Patil, D. Akshay, and M. Subrahmanyam, "Multi-frame recurrent adversarial network for moving object segmentation," 2021 *IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, HI, USA, pp. 567-576, 2021, DOI: 10.1109/WACV48630.2021.00235.
- [16] C. Yu, C. Gao, J. Wang, G. Yu, S. Chunhua, and N. Sang, "Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation," *International journal of computer vision*, vol. 129, pp. 3051-3068, 2021, DOI: 10.1007/s11263-021-01515-2.
- [17] J. Hu, S. Li, and G. Sun, "Squeeze-and-excitation networks," 2018 *IEEE conference on computer vision and pattern recognition*, Salt Lake City, UT, USA, pp. 7132-7141, 2018, DOI: 10.1109/CVPR.2018.00745.
- [18] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *Medical image computing and computer-assisted intervention-MICCAI 2015*, vol. 9351, pp. 234-241, 2015, DOI: 10.1007/978-3-319-24574-4_28.
- [19] T. Ehret, A. Davy, J.-M. Morel, G. Facciolo, and P. Arias, "Model-blind video denoising via frame-to-frame training," *arXiv:1811.12766*, 2018, DOI: 10.48550/arXiv.1811.12766
- [20] R. Liao, X. Tao, R. Li, Z. Ma, and J. Jia, "Video super-resolution via deep draft-ensemble learning," 2015 *IEEE international conference on computer vision (ICCV)*, Santiago, Chile, pp. 531-539, 2015, DOI: 10.1109/ICCV.2015.68.
- [21] A. Kappeler, S. Yoo, Q. Dai, and K. Katsaggelos, "Super-resolution of compressed videos using convolutional neural networks," 2016 *IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA, pp. 1150-1154, 2016, DOI: 10.1109/ICIP.2016.7532538.
- [22] M. Drulea and S. Nedevschi, "Total variation regularization of local-global optical flow," 2011 *14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, Washington, DC, USA, pp. 318-323, 2011, DOI: 10.1109/ITSC.2011.6082986.
- [23] E. Ustinova and V. Lempitsky, "Deep multi-frame face super-resolution," *arXiv:1709.03196*, 2017, DOI: 10.48550/arXiv.1709.03196.
- [24] H. Shen, Z. Qiu, L. Yue, and L. Zhang, "Deep-Learning-Based Super-Resolution of Video Satellite Imagery by the Coupling of Multiframe and Single-Frame Models," *IEEE transactions on geoscience and remote sensing*, vol. 60, pp. 1-14, 2022, DOI: 10.1109/TGRS.2021.3121303.
- [25] E. M. Masutani, N. Bahrami, and A. Hsiao, "Deep learning single-frame and multiframe super-resolution for cardiac MRI," *Radiology*, vol. 295, no. 3, pp. 552-561, Apr., 2020, DOI: 10.1148/radiol.2020192173.
- [26] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," *Computer Vision-ECCV 2016*, pp. 21-37, 2016, DOI: 10.1007/978-3-319-46448-0_2.
- [27] S. Jain, X. Wang, and J. Gonzalez, "ACCEL: A corrective fusion network for efficient semantic segmentation on video," *arXiv:1807.06667*, 2019, pp. 8866-8875, DOI: 10.48550/arXiv.1807.06667.
- [28] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834-848, Apr., 2018, DOI: 10.1109/TPAMI.2017.2699184.
- [29] J. Hur and S. Roth, "Self-supervised multi-frame monocular scene flow," *arXiv:2105.02216*, 2021, pp. 7226-7236, DOI: 10.48550/arXiv.2105.02216.
- [30] Y. Kim, J. Park, S. Kang, and H. Kim, "Introducing VaDA: Novel image segmentation model for maritime object segmentation using new dataset," *arXiv:2407.09005*, 2024, DOI: 10.48550/arXiv.2407.09005.
- [31] M. Fan, S. Lai, J. Huang, X. Wei, Z. Chai, J. Luo, and X. Wei, "Rethinking bisenet for real-time semantic Segmentation," 2021 *IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, Nashville, TN, USA, pp. 9716-9725, 2021, DOI: 10.1109/CVPR46437.2021.00959.



박진범

2018 UC SanDiego 전자공학과(학사)
2018 LG전자 연구원
2024 고려대학교 전기공학과(석사)
2024~현재 Seadronix Corp 인공지능 연구원

관심분야: 컴퓨터 비전, 딥러닝, 영상 및 신호처리 이미지 센서



강산하

2023 광주 SW마이스터 고등학교
2023~현재 Seadronix Corp 인공지능 연구원

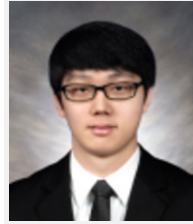
관심분야: 컴퓨터 비전, 딥러닝, 객체 인식



김용진

2016 전북대학교 전자공학과(학사)
2018 전북대학교 전자공학과(석사)
2022.7~현재 Seadronix Corp 인공지능 연구원

관심분야: 컴퓨터 비전, 딥러닝, 데이터 분석



김한근

2016 한국과학기술원 건설 및환경공학과
(공학박사)
2016 삼성 S1 책임 연구원
2017~현재 Seadronix Corp. 기술이사

관심분야: 딥러닝, 영상처리, 센서 및 로봇항법