Journal of KIIT. Vol. 22, No. 2, pp. 9-20, Feb. 28, 2024. pISSN 1598-8619, eISSN 2093-7571 9
http://dx.doi.org/10.14801/jkiit.2024.22.2.9

무인수상정을 위한 개선된 YOLOv7 기반 해양 선박탐지

이준희*, 유지상**, 장경선***

Marine Ship Detection based on Enhanced YOLOv7 for Unmanned Surface Vehicles

Junhee Lee*, Jisang You**, and Kyoungson Jhang***

이 논문은 2023년 정부의 재정으로 수행된 연구 결과임

요 약

무인 기술의 발전과 함께 민간분야뿐만 아니라 방산에서도 무인수상정에 관한 연구가 활발하게 진행되고 있다. 다양한 군사 작전에서 활용 가능한 무인수상정에는 원격제어, 자율운항, 자율제어 등의 첨단기술이 적용된다. 이러한 기술들은 공통으로 선박 자동탐지 기술이 필요하다. 본 논문에서는 무인수상정을 위한 딥러닝 기반의 객체인식 모델을 이용한 해양 선박탐지 방법을 제안한다. 제안하는 방법은 객체인식 모델 YOLOv7에 어텐션 모듈과 업샘플링 기법을 새롭게 적용한다. 그리고 모델의 학습을 위해 무인수상정 환경에 적합한 선박들로 구성된 공개 데이터들을 기반으로 통합 선박 데이터셋을 구축한다. 우리는 통합 선박 데이터셋에 대하여 기존의 모델들과 비교를 통해 성능 향상을 확인하였고, 유·무인수상정에서 획득한 영상에서 선박탐지 확인을 통해 실제 적용 가능성을 확인하였다.

Abstract

With the advancement of unmanned technology, research on USVs(Unmanned Surface Vehicles) is actively progressing not only in civilian sectors but also in defense. Advanced technologies such as remote control, autonomous navigation, and autonomous control are applied to USVs for various military operations. These technologies commonly require ship automatic detection technology. This paper proposes a marine ship detection method for USVs using a deep learning-based object detection model. The proposed is based on the YOLOv7 model, introducing new features such as attention modules and upsampling techniques. Additionally, for model training, a dataset suitable for USVs is created by integrating publicly available datasets of targets. We compare the performance of existing models on the constructed dataset, showing the improved performance of the proposed model.

Keywords
unmanned surface vehicles, ship detection, deep learning, YOLOv7

* 국방과학연구소 선임연구원(교신저자)
- ORCID: http://orcid.org/0009-0003-6388-7091
** 국방과학연구소 책임연구원
- ORCID: http://orcid.org/0000-0001-6650-0137
*** 충남대학교 컴퓨터공학과 교수
- ORCID: http://orcid.org/0000-0001-5659-0503

· Received: Nov. 28, 2023, Revised: Dec. 20, 2023, Accepted: Dec. 23, 2023
· Corresponding Author: Junhee Lee
Agency for Defense Development, Deajeon, YuSeong P.O. Box 35, Daejeon, 34186, Korea
Tel.: +82-54-292-3609, Email: solejun@naver.com

Ⅰ. 서 론

무인 기술은 제조, 물류, 의료 등 산업 전반에 큰 영향을 미치고 있다. 카페에서 커피를 내리는 로봇, 식당에서 음식을 서빙하는 로봇 등 일상생활에서도 쉽게 접할 수 있게 되었다. 최근 인공지능 발전과 함께 무인 기술이 더욱 진화하고 있다. 기존 단순한 루틴에서 인식과 판단을 통한 무인자동차, 무인기, 무인수상정 등의 연구가 진행되고 있다.

무인 기술은 민간분야뿐만 아니라 방산에서도 많은 주목을 받았다. 군 인력 감소, 인명 중시로 인해 무인체계의 필요성이 대두되었고, 이에 따라 우리나라는 2023년 국방혁신 4.0 발표와 함께 미래전장을 대비한 유·무인 복합전투체계를 제시하였다. 해군에서는 해양 유·무인체계를 위한 무인 기술개발이 진행되고 있다[1]. 무인수상정은 주요시설 감시 경계뿐만 아니라 대함전, 대잠전, 대기뢰전 등 해전에서 전장 상황을 파악하고, 적을 식별하는데 중요한 임무를 수행한다. 무인수상정은 모함이나 육상기지에서 원격으로 제어하거나, 자율제어를 통해 운용된다. 무인수상정의 기술 수준이나 완성도에 따라서 자율화 단계가 정의되고, 실전에 활용되는 방법이 달라진다[2]. 자율화 단계 구별은 원격제어, 자율운항, 자율제어 기술 수준에 따라 결정된다. 이러한 핵심 기술들의 고도화를 위해서는 정확한 주변 환경 인식에서 시작된다. 선박 자동탐지 기술은 주변 환경 인식에 필수 기술 중 하나이다.

그림 1은 우리나라에서 개발된 무인수상정 해검 형상이다. 무인수상정의 상단에는 감시를 위한 EOTS(Electro-Optical Target System)과 항해를 위한 카메라가 장착된다.

그림 1. 무인수상정 해검
Fig. 1. Unmanned surface vehicle sea sword

이러한 광학장비들을 통해서 주변 영상을 획득하고, 선내의 신호처리장치로 전달된다. 신호처리장치는 획득 영상에 대한 선박탐지를 수행한 후 결과를 선박 통제장치에 전달하여 자율운항/자율제어에 사용되고, 원격지의 운용자에게 전달하여 선박탐지 알람을 제공한다.

본 논문에서는 무인수상정을 위한 딥러닝 모델 기반의 선박탐지 방법을 제안한다. 딥러닝 모델을 적용하기 위해 공개 데이터셋 기반 대용량 데이터를 구축하였고, 최신 객체인식 모델 YOLOv7[3]을 개선한 모델을 제안하였다. 본 논문의 구성으로 2장에서 고전적인 객체인식 방법들과 딥러닝 기반의 객체인식 모델, 특히 YOLO(You Only Look Once) 계열 모델들을 소개한다. 3장에서는 선박탐지를 위해 사용한 모델, 성능 개선 방법들과 모델 학습을 위해 데이터셋을 구축한 방법에 관해 기술한다. 4장에서는 구축한 데이터셋에 대해 기존 모델들과 제안한 모델의 실험 결과를 분석하고, 5장에서는 결론과 향후 연구에 관해서 기술한다.

Ⅱ. 관련 연구

고전적인 객체인식 방법은 탐지할 객체에 대해서 특징을 추출하고, 학습하여 영상에서 찾는다. 특징 추출을 위한 대표적인 방법은 코너점 이용하는 SIFT(Scale Invariant Feature Transform), 에지 픽셀의 방향을 이용하는 HOG(Histogram of Oriented Gradients)과 헤시안 행렬을 사용하는 SURF(Speed-Up Robust Feature) 등이 있다. 특징 학습을 위해서는 RF(Random Forest), SVM(Support Vector Machine) 등이 주로 사용되었고, SVM 이후에는 배깅(Bagging), 부스팅(Boosting), 캐스케이딩(Cascading) 기술을 접목한 앙상블 학습을 사용하였다. 이러한 방법은 영상 전체에 슬라이딩 윈도우 방식으로 영역들을 비교하기 때문에 속도가 매우 느리다.

2012년 컴퓨터 비전의 영상 분류(Image classification)에서 AlexNet 선보인 이후 객체인식 분야도 딥러닝 기반 연구가 시작되었다. 2013년에 처음으로 CNN을 사용하여 객체 위치 영역을 찾고(Localization), 제안된 영역을 분류(Classification)하는 R-CNN이 제안되었다[4].

위치추정과 분류가 분리된 two-stage 구조를 갖는다. 2016년에는 학습과 추론 속도 개선을 위해 위치추정과 분류를 동시에 학습을 수행하는 one-stage 구조의 YOLOv1 모델이 제안되었다[5]. Two-stage는 정확도가 높지만 학습과 추론 속도가 느리고, one-stage는 정확도는 낮지만 속도가 빠르다는 장단점이 있다. 무인 플랫폼에 관한 연구가 늘어남에 따라 제한된 연산속도와 용량을 가진 소형 장치에 적용하기 위해 one-stage 계열 모델의 연구가 지속해서 이루어졌다. 특히, YOLO 모델은 실시간 객체인식 분야의 대표적인 모델로 현재까지 개선되고 있다.

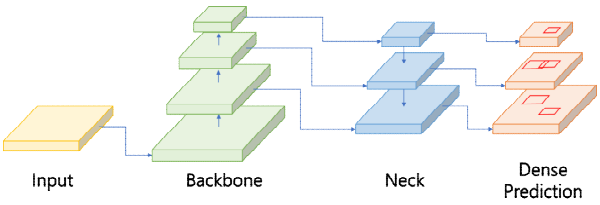


그림 2. YOLO 모델의 기본 구조
Fig. 2. Basic structure of the YOLO model

그림 2는 YOLO 모델의 기본 구조로 영상이 입력되면, 백본(Backbone)과 넥(Neck)을 통해서 특징을 추출하고, 각 픽셀에 대해서 예측(Dense prediction) 수행한다. YOLOv1에서는 영상을 격자로 나누어 격자마다 2개의 바운딩 박스와 분류 결과를 생성한다. 중첩된 결과에 대해서는 NMS(Non-Max Suppression)를 통해 제거한다. 손실함수는 바운딩 박스의 좌표, 분류, 신뢰도 등을 개별 손실에 대해 합산하여 사용한다. YOLOv2에서는 중첩 객체를 탐지 개선을 위해 anchor box를 도입, 학습 시간을 줄이기 위해 예측 좌표 변경, 다양한 크기의 영상을 위해 fully connected layer 제거하고, 성능 개선을 위해 skip connection과 배치 정규화 도입하였다[6]. YOLOv3에서는 기존에 다른 객체인식 모델에서 사용하던 기법들을 적용하여 성능을 개선하였다[7]. 다중 라벨이 가능하도록 클래스별 이진 분류를 적용하였고, 바운딩 박스의 정확도를 높이기 위해 예측 좌표 변경하였다. 그리고 예측에 사용된 특징맵을 3개로 늘려서 작은 객체인식 성능을 높였다. YOLOv4에서는 실시간 동작과 하나의 GPU를 이용한 학습에 중점을 두었으며, 다양한 기법들을 실험을 통해 최적으로 조합하였다[8]. 학습 데이터 증강을 위한 mosaic 기법, 특징맵을 효과적으로 융합하기 위한 PAN(Path Aggregation Network), 수용 범위 증가를 위한 SPP(Spatial Pyramid Pooling) 등을 사용하였다. YOLOv5에서는 YOLOv3를 기반으로 한 모델로 계층을 더 깊게 쌓기 위해 C3 모듈을 추가하였고, 학습을 위한 hyper parameter를 최적화하였다[9]. YOLOx에서는 YOLOv3을 기반으로 예측 결과 바운딩 박스와 분류를 분리하고, 데이터 증강을 위한 mosaic, mixup 기법, anchor free를 통한 성능 개선을 제안하였다[10]. YOLOv7에서는 효율적으로 특징을 추출하기 위한 ELAN(Efficient Layer Aggregation Network) 모듈, 추론 비용을 늘리지 않고 정확도를 개선을 위한 재매개변수화 기법과 입출력의 너비를 유지하면서 깊이를 확장하는 방법들을 제안하였다[3]. YOLO 객체 인식을 활용한 연구가 연구됐으며[11]-[13], 최근에는 YOLOv7을 기반으로 모델을 개선하는 연구가 진행 중이다[14][15].

III. 제안하는 방법

3.1 YOLOv7 모델

본 논문에서 제안하는 선박탐지 모델은 그림 3와 같이 YOLOv7을 기반으로 어텐션 모듈과 업샘플링 기법을 추가하였다. YOLOv7의 구조는 특징을 추출하는 백본, 특징을 융합하는 넥, 바운딩 박스와 분류를 예측하는 헤드(Head) 세 부분으로 나뉜다. 영상은 정규화와 크기를 변환하여 백본에 입력된다. 백본은 CBS, ELAN, MPConv 모듈을 단위로 특징을 계산한다. CBS 모듈은 컨볼루션과 배치 정규화, SiLU(Sigmoid Linear Unit) 활성화 함수를 순차적으로 수행한다. ELAN 모듈은 확장(Expand), 섞기(Shuffle)와 병합 카디널리티(Merge cardinality)를 사용하여 원래 기울기 경로를 파괴하지 않고, 네트워크 학습 능력을 지속해서 향상한다. 그룹 컨볼루션은 계산 블록의 채널과 카디널리티를 확장하는 데 사용되고, 각 특징맵의 채널 수가 원래 아키텍처의 채널 수와 동일하게 유지되도록 보장한다. ELAN 결과 채널의 수는 입력의 두 배가 된다.

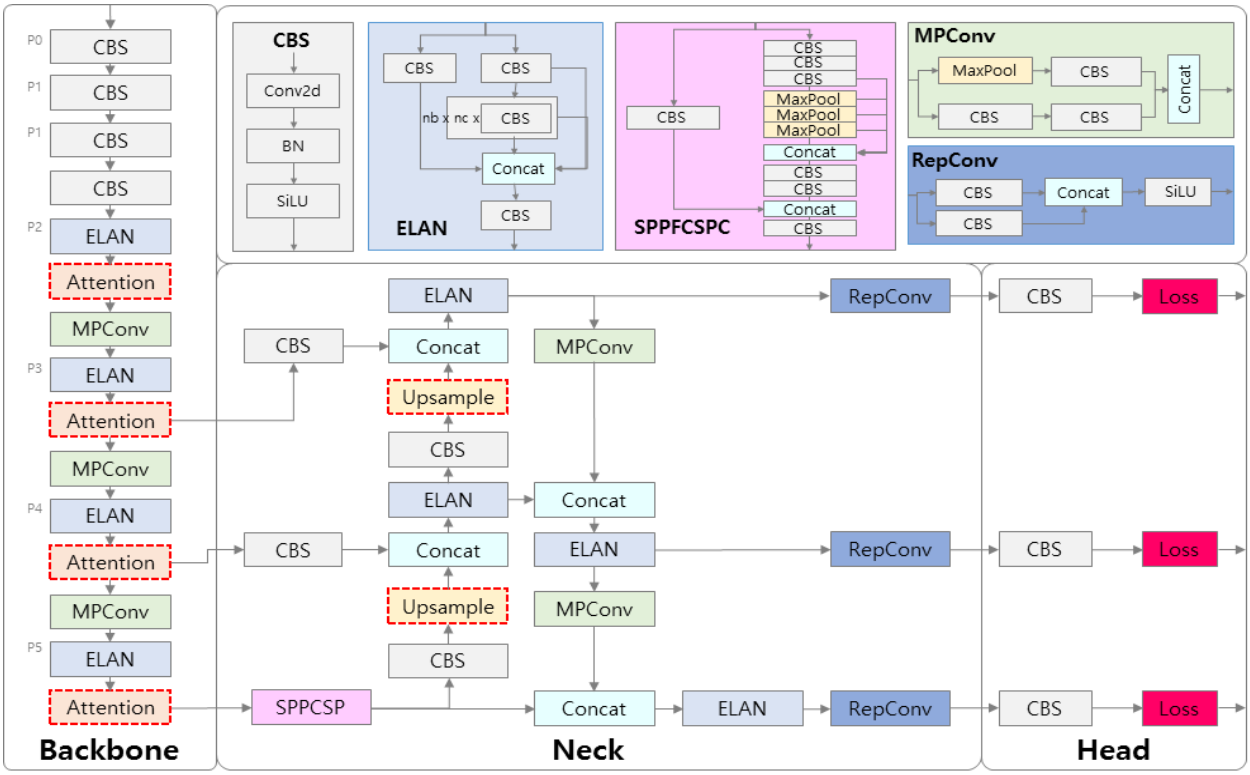


그림 3. 제안하는 선박탐지 모델 구조
Fig. 3. Proposed ship detection model

MPConv 모듈은 상위 분기에서 맥스 풀링(Max pooling)을 통해 특징맵의 길이와 너비를 절반으로 줄이고, 컨볼루션에 통해 채널이 절반으로 줄어든다. 하위 분기에서는 첫 번째 컨볼루션에 의해 특징맵의 길이와 너비가 절반으로 줄어든다. 상위와 하위 분기가 결합하여 반으로 줄어든 길이와 너비, 그리고 입력과 출력 채널이 동일한 출력 특징맵을 얻게 된다.

넥에서는 백본에서 3개의 특징맵(P3, P4, P5)을 하향식과 상향식으로 융합하는 PAN 구조를 갖는다. 백본에서 사용한 모듈과 더불어 SPPCSP, RepConv, 업샘플링 모듈을 사용한다. SPPCSP 모듈은 이전 특징맵으로부터의 수용 영역을 증가시킨다. 백본의 마지막 특징맵(P5)에 대해서 세 번의 컨볼루션 연산을 적용하고, 5, 9, 13의 커널 크기를 갖는 맥스 풀링 연산을 각각 세 번 수행한다. 이를 통해 특징맵의 크기를 유지하면서 다중 스케일의 객체 특징을 얻을 수 있다. RepConv 구조는 학습과 추론의 구조가 다르게 구성된다. 학습에서는 CBS와 BN의 조합하고, 결과를 결합하는 형태이다. 추론에는 세 부분이 다시 매개변수화 되어, 3×3 컨볼루션 출력을 사용한다. 즉, 매개변수를 다른 매개변수 집합으로 등등하게 변환한다. 이렇게 배포된 모델은 학습 때보다 속도가 더 빠르게 된다. Head에서는 YOLOv3의 구조와 동일하게 바운딩 박스 위치와 박스에 대한 선박 종류별 확률과 선박이 존재할 확률이 계산된다. 이러한 예측에 대한 손실함수는 바운딩 박스의 손실은 CIoU(Complete Intersection over Union), 확률의 손실은 BCE(Binary Cross Entropy)으로 가중치 합산을 사용한다.

3.2 어텐션 모듈 : SimAM

본 논문에서는 백본에서 특징을 효율적으로 추출하기 위해 입력에서 중요한 정보에 집중하는 어텐션 모듈을 그림2와 같이 백본에 삽입하였다. 어텐션 모듈은 1차원, 2차원 형태의 추가적인 파라미터를 요구하는 기법과 달리 추가 파라미터가 없는 3차원 정보를 이용하는 SimAM(Simple Attention Module)을 적용하였다[16].

이 방법은 시각 신경과학 이론을 바탕으로 정보력 있는 뉴런은 주변 다른 뉴런들과 다른 발화 패턴을 가진다는 점을 이용하여 선형 분리를 통해 뉴런을 구별하였다. 먼저, 뉴런에 대한 에너지 함수를 다음 수식과 같이 정의한다.

$$e_t(w_t, b_t, y, x_i) = \frac{1}{M-1}\sum_{i=1}^{M-1}(-1-(x_i w_t + b_t))^2 + (1-(w_t t + b_t))^2 + \lambda w_t^2 \quad (1)$$

t는 대상 뉴런, x_i는 채널 내 다른 뉴런을 나타낸다. w_t와 b_t는 선형 분리의 가중치와 편향을 의미한다. i는 공간 차원의 인덱스, λ는 계수이다. M은 하나의 채널에서 모든 뉴런의 수이다. 변환 가중치와 편향은 다음 수식과 같다.

$$w_t = -\frac{2(t-\mu_t)}{(t-\mu_t)^2 + 2\sigma_t^2 + 2\lambda} \quad (2)$$

$$b_t = -\frac{1}{2}(t+\mu_t)w_t \quad (3)$$

u_t와 σ_t^2는 t을 제외한 해당 채널의 모든 뉴런의 평균과 분산이다. 가중치 w_t와 b_t, 해당 채널 내 모든 뉴런의 평균과 분산에 대한 해를 계산함으로써 최소 에너지는 다음과 같아진다.

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t-\hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (4)$$

$\hat{\mu}$와 $\hat{\sigma}^2$는 모든 뉴런에 대한 평균과 분산이다. 식 (4)에서 보듯이 에너지 함숫값이 작아질수록 뉴런 t와 다른 뉴런 간의 선형 분리가 커지기 때문에 뉴런 t에 대한 중요성은 $1/e_t$으로 계산된다. 특징맵에서는 다음 수식과 같이 계산한다.

$$\tilde{X} = X \bullet sigmoid\left(\frac{1}{E}\right) \quad (5)$$

E는 채널과 공간 차원에서 $1/e_t$을 그룹화 한 것이고, 시그모이드를 통해 범위를 제한한다. $\bullet$는 요소별 곱셈(Element-wise multiplication)을 의미한다. 이 값은 3차원 가중치로 그림 4과 같이 적용된다.

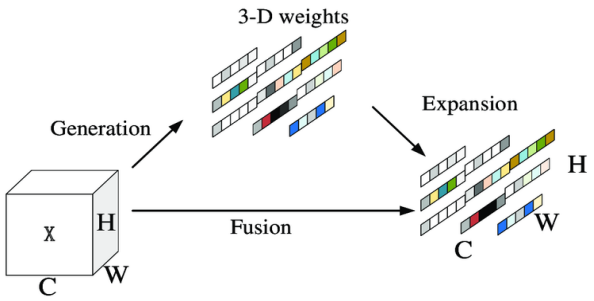


그림 4. SimAM 모듈 구조
Fig. 4. SimAM modules structure

3.3 업샘플링 모듈 : DySample

특징 업샘플링은 특징맵 해상도를 복원하기 위한 중요한 요소로 YOLOv7의 넥에서 사용되는 업샘플링은 최근접 이웃 보간법(Nearest Neighbor Interpolation)을 이용한다[3]. 이는 입력 특징맵에서 공간 정보만 활용하고 의미 정보를 무시하여 정보 손실이 발생한다. 본 논문에서는 업샘플링 과정에서 의미 정보가 동적으로 포함되도록 DySample을 사용하였다[17]. 기존의 동적 컨볼루션이나 커널을 사용하는 방식은 추가적인 파라미터나 많은 부하를 발생시키지만, DySample는 그림 5와 같이 포인트 샘플링을 통해 부하를 줄였다.

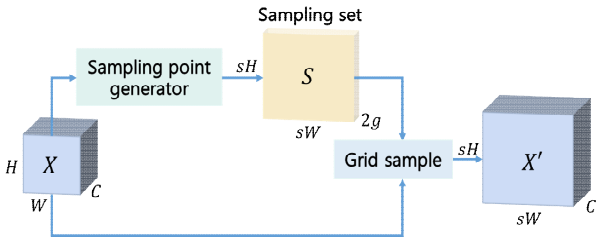


그림 5. 동적 업샘플링 기반 포인트 샘플링
Fig. 5. Point sampling based dynamic upsampling

특징맵에서 샘플 포인트 생성기를 통해서 동적으로 포인트 위치에 대한 샘플링 셋을 만들고, 다시 특징맵에 격자 샘플(Grid sample) 방식으로 업샘플링을 수행한다. 이때 격자의 위치는 샘플링 셋을 사용한다. 격자 샘플 방식은 다음 수식과 같다.

$$X'(x', y') = \sum_{i,j} X(x_i, y_j) \bullet W(x', y', i, j) \quad (6)$$

X'는 업샘플링된 특징맵, X는 원래의 특징맵을 나타낸다. W는 격자 샘플링 가중치로 변환 매트릭스에 따라 입력과 출력 좌표 간의 관계를 나타내고, 다음과 같이 양선형 보간(Bilinear interpolation) 사용한다.

$$W(x', y', i, j) = relu(1 - |x' - i|) \cdot relu(1 - |y' - j|) \quad (7)$$

relu는 ReLU(Rectified Linear Unit) 함수를 나타내고, 이 가중치 함수는 입력과 출력 좌표 간 거리에 따라 가중치를 할당한다.

샘플 포인트 생성기는 그림 6와 같이 격자 샘플링에서의 위치를 오프셋 O을 생성한다. 먼저, 특징맵에서 선형 레이어를 사용하여 $2s^2 \times H \times W$를 생성한다. 업스케일링의 경우 위치에 대한 중첩으로 인한 성능 저하 현상을 방지하기 위해 0.25를 곱하여 중첩과 비중첩 간의 마진을 둔다. 특징맵을 확장하기 위해서는 다음과 같이 픽셀 셔플링을 통해 $2 \times sH \times sH$ 크기의 O을 생성한다.

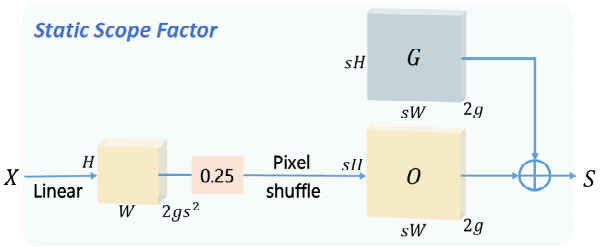


그림 6. 샘플링 포인터 생성기
Fig. 6. Sampling pointer generator

$$pixel_shuffle(X)_{i,j,k} = X_{i, \lfloor k/s \rfloor, \lfloor j/s \rfloor} \quad (8)$$

X는 입력 텐서, i, j, k는 출력 텐서의 채널, 높이, 너비의 인덱스를 의미한다. 이 수식은 픽셀의 재배치하여 스케일에 따라 입력 텐서의 공간적인 해상도를 높인다.

3.4 통합 선박 데이터셋 구축

기존 선박의 선박 관련 공개 데이터들은 동영상에서 프레임을 추출하여 영상과 레이블을 구성한다. 이러한 방식은 프레임 간격을 두어 영상 간 중복을 줄이기 위해 노력하였지만, 획득 환경 자체가 같으므로 특정한 배경, 촬영 구도와 제한된 선박 종류 등으로 유사성을 갖게 되어, 새로운 환경에서 획득한 데이터에 대한 성능이 떨어지게 된다. 본 논문에서는 이러한 문제를 해결하기 위해 무인수상정에서 관측할 수 있는 선박과 배경 등을 포함하는 다수 공개 데이터셋을 통합한 데이터셋을 구축하였다.

기존 연구에서 획득한 SMD, SeaShips, WSODD, Pascal VOC 등의 데이터셋을 활용하고[18], 새로운 데이터셋 2개를 추가하였다. ABOShips 데이터셋은 핀란드에서 구축된 데이터셋으로 선박의 전방에 카메라를 장착하여 동영상을 획득하였다[19]. 강을 따라서 운항하며 수로 좌우에 정박한 선박들이 대부분을 자치한다. 선박 종류는 부표, 모터보트, 세일보트, 여객선, 화물선, 군함, 페리, 기타 보트, 부유물 등으로 구성된다.

Maritime Ports 데이터셋은 크로아티아에서 구축된 데이터셋으로 항구 입구에서 카메라를 고정한 환경에서 입출항하는 선박을 획득하였다[20]. 지나가는 선박이기 때문에 선박 측면이 대부분이다. 선박 종류는 소형 선박, 소형 어선, 소형 여객선, 어

로선, 대형 여객선, 세일보트, 스피드 보트, 모터보트, 요트, 중·대형 페리, 고속정 등 12가지가 있다.

공개 데이터셋은 각 데이터셋 마다 선박 종류가 다르므로 여객선, 화물선, 보트, 특수선, 부표 등 5개 선박으로 종류를 병합하였다. 그리고 데이터셋 내부의 유사성을 줄이기 위해 PASCAL VOC를 제외한 나머지에 대해 프레임 간격을 추가하였다. 그림 7은 6개 데이터셋을 통합한 데이터셋의 선박 개수이다. 화물선과 보트의 종류가 많고, 여객선, 특수선, 부표 등은 상대적으로 작다. 이는 실제 바다에서 관측된 영상을 바탕으로 만들어졌기 때문이다. 표 1은 학습으로 사용된 각 데이터셋에 대한 영상 개수로 무인수상정 환경을 고려하여 구성하였다. 그림 8은 각 데이터셋에 대한 영상으로 각 데이터셋마다 배경과 선박 형태에 대한 특성이 있다. 데이터셋 통합을 통해 선박탐지 모델이 다양한 배경과 선박을 학습하도록 하였다.

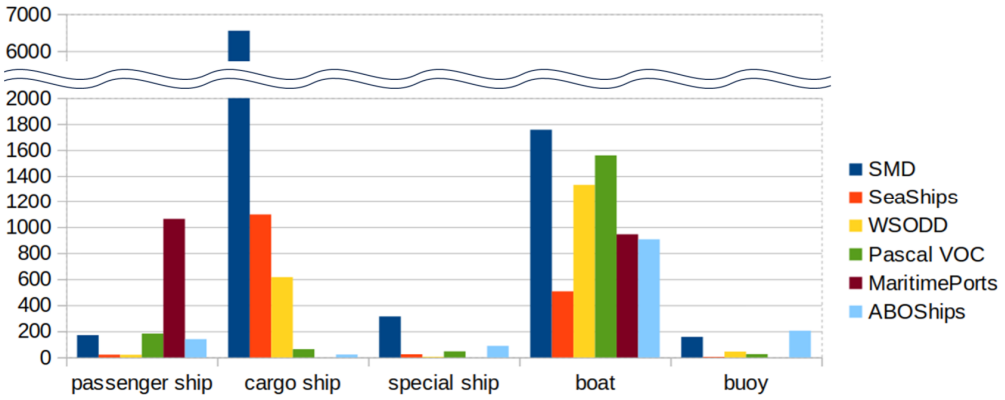


그림 7. 무인수상정을 위한 선박 데이터셋 구성
Fig. 7. Ship dataset configuration for USVs

그림 8. 각 데이터셋의 이미지 예제
Fig. 8. Image examples from each dataset

표 1. 학습 데이터셋의 영상 개수
Table 1. Number of images in training dataset

Dataset	Images
SMD	1,134
SeaShips	1,148
WSODD	996
Pascal VOC	901
Maritime ports	618
ABOShips	263

Ⅳ. 실험 및 분석

선박탐지 성능평가는 객체인식에서 일반적으로 사용하는 mAP(mean Average Precision)을 사용한다. mAP는 영상 분류의 성능 지표로 사용되는 정밀도(Precision)와 재현율(Recall)을 이용한다. 실제값과 예측값에 따른 혼돈행렬은 다음과 같다.

표 2. 혼돈행렬
Table 2. Confusion matrix

Ground truth	Predict result	
	Positive	Negative
Positive	TP (True Positive)	FN (False Negative)
Negative	FP (False Positive)	TN (True Negative)

TP는 모델이 어떤 객체를 예측했는데 GT 상에 존재하는 경우를 의미한다. FP는 모델이 어떤 객체가 있다고 예측했는데 틀린 경우이다.

FN은 모델이 객체가 없다고 예측했는데 틀린 경우로, 실제 존재하는 물체를 아예 검출하지 못한 경우이다. TN은 모델도 물체가 없다고 예측했고, 실제로도 없는 경우이다. 정밀도는 모델이 검출한 모든 결과 중 옳게 검출한 비율로 다음과 같다.

$$\mathrm{Precision} = \frac{TP}{TP+FP} \tag{9}$$

재현율은 실제 모든 객체 중 모델이 옳게 검출한 비율로 다음과 같다.

$$Recall = \frac{TP}{TP+FN} \tag{10}$$

AP는 정밀도와 재현율에 대한 곡선에서 다음과 같이 면적을 계산한 값이다.

$$AP = \int_0^1 P(R)dR \tag{11}$$

mAP는 각 선박 종류마다 AP를 다음과 같이 평균한 결과이다.

$$mAP = \frac{1}{c}\sum_{i=1}^{c} AP_i \tag{12}$$

본 논문에서 선박탐지 모델 학습을 위해 사용된 학습 파라미터는 표 3과 같다.

표 3. 학습 파라미터
Table 3. Training parameters

Parameters	Value
Momentum	0.937
Weight decay	0.0005
Batch size	16
Learning rate	0.01
Image size	640×640
Epoch	100
Optimizer	SGD

학습에 사용된 하드웨어 사양은 AMD 라이젠 3970X CPU, NVIDIA 지포스 RTX 3090 2개, 256GB 램으로 구성된다. 모델 학습과 검증 데이터는 통합 선박 데이터셋을 7:3 비율로 생성하였다. 운영체제는 Ubuntu 20.04, 개발언어는 Python 3.8, 프레임워크는 Pytorch 2.0.1에서 구현하였다. 학습 영상은 3,530장, 검증 영상은 1,540장을 사용하였다.

표 4는 본 논문에서 구축한 통합 선박 데이터셋에 대한 YOLO 모델의 학습 결과이다. 모델에 관한 결과로 mAP와 파라미터 개수, 연산량, 실제 추론 속도 등을 측정하였다. mAP에서 0.5는 IoU를 의미하고, 그 이상 값이어야 TP로 판단한다. Pascal VOC 경진대회에서 모델 평가를 위해 사용되었던 지표로 위치보다는 분류에 집중되어 있다. 0.5:0.95는 IoU를 0.5 단위로 측정하고 평균한 값으로 MS COCO 경진대회에서 사용되었던 지표로 분류보다 위치에 집중되어 있다. 무인수상정에서 선박탐지에는 정확한 위치보다는 신속하게 탐지 후 결과를 전달하는 것이 중요하기 때문에 mAP@0.5 기준으로 모델을 평가하는 것이 타당하다.

표 4. 선박 데이터셋에 대한 YOLO 모델 결과
Table 4. Results of YOLO models for ship dataset

Class / Models	mAP@0.5 (%)	mAP@0.5:0.95 (%)	Params (M)	GFLOPS	FPS (Frame/s)
YOLOv4	71.9	34.8	63.9	141.9	30.8
YOLOv5	71.8	41.3	46.6	114.6	42.7
YOLOv7	72.3	41.7	37.2	105.1	42.6

표 5. YOLOv7과 제안한 방법의 결과
Table 5. Results of YOLOv7 and proposed methods

Models	mAP@0.5 (%)	mAP@0.5:0.95 (%)	Params (M)	GFLOPS	FPS (Frame/s)
YOLOv7	72.3	41.7	37.216	105.182	42.6
+DySample	73.4	42.0	37.229	105.199	42.4
+DySample+SimAM (Proposed)	76.2	44.0	37.229	105.199	40.9

성능 측정 결과를 보면 mAP@0.5 기준으로 YOLOv4와 YOLOv5는 비슷하고, YOLOv7가 0.5% 정도로 높다. mAP0.5:0.95 결과를 보면, YOLOv4에 비해 후속 모델들이 6% 이상 개선된 것을 확인할 수 있다. 바운딩 박스의 결과가 개선되었다는 것을 확인할 수 있다. YOLOv5, YOLOv7 간에는 1% 미만으로 크게 개선되지는 않았다. 모델 경량화에서는 많이 개선되었다. 모델 자체의 파라미터 개수를 보 모델이 개선됨에 따라서 매우 감소하였고, YOLOv7은 YOLOv4와 비교하여 50%나 감소하였다. 또한, 그래픽카드의 초당 부동소수점 연산을 의미하는 GFLOPS의 경우 YOLOv5가 YOLOv4 보다 17M 정도 감소하였고, YOLOv7은 YOLOv5 보다 9M 정도 감소하였다. 마지막으로 FPS는 실제 모델이 초당 선박탐지를 수행하는 프레임 수를 의미하고, 30프레임 이상으로 처리되기 실시간 선박탐지가 가능하다. YOLOv5와 YOLOv7이 YOLOv4 보다 10 FPS 이상 빨라진 것을 확인할 수 있다. 최신 모델일수록 mAP가 개선되진 않았지만, 속도에서 많이 개선된 것을 확인할 수 있다.

YOLOv7 모델에 본 논문에서 제안한 성능 개선 기법들을 적용한 결과는 표 5와 같다. YOLOv7에서 넥의 업샘플링 2곳을 DySample으로 변경한 결과 mAP@0.5 기준에서 1.1% 증가하였고, mAP@0.5:0.95에서는 0.3%가 증가하였다. 경량의 구조로 되어 있어서 파라미터는 13K, 연산량은 0.017 GFLOPS으로 전체 1% 미만 수준으로 증가하였다. FPS에서도 성능 저하가 거의 없다. DySample과 함께 백본의 4곳에 SimAM 어텐션 모듈을 적용한 결과 mAP@0.5에서는 2.8% 증가, mAP@0.5:0.95에서는 2%가 증가하였다. 파라미터가 별도로 없고, 부동소수점 연산하지 않기 때문에 GFLOPS에서도 증가는 없다. FPS 측정 결과 SimAM 추가에 따라 초당 1.5 프레임 정도 떨어지지만, 40프레임 이상으로 사람이 체감하기 불가능하다.

종합적으로 보면 모델에 대한 성능 비교 결과는 그림 9과 같다. 제안한 방법은 YOLOv4 보다 mAP@0.5는 4.3%, mAP@0.5:0.95는 9.2% 높고, YOLOv7 보다는 mAP@0.5는 3.9%, mAP@0.5:0.95는 2.3% 높다. 반면에 모델 성능 향상에도 불구하고, Params, GFLOPS와 FPS에 대한 손해는 매우 작다.

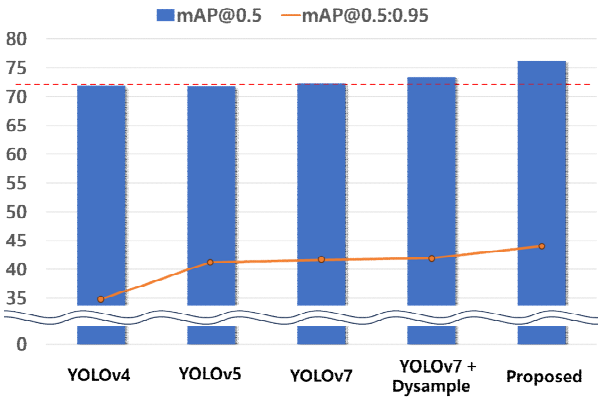
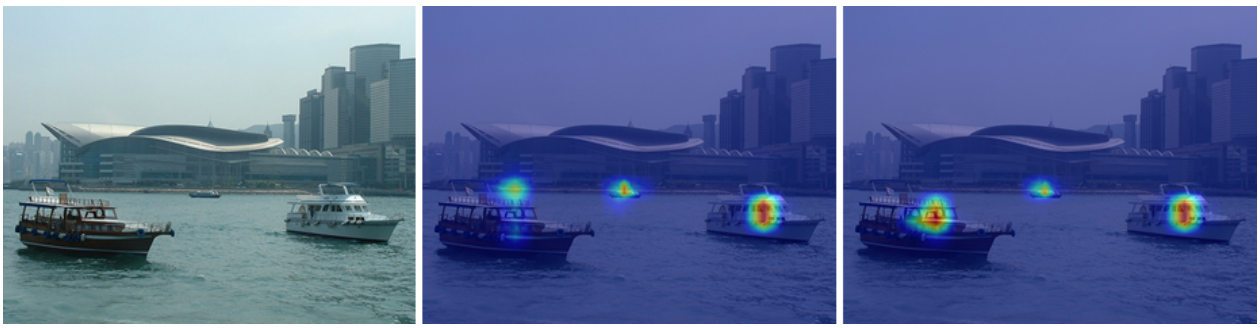


그림 9. 모델 성능 비교 결과
Fig. 9. Model performance comparison results

그림 10는 YOLOv7 모델과 제안한 방법으로 선박 영상에 대해 히트맵을 생성한 결과이다. 히트맵은 탐지 결과에 기여한 부분을 가시적으로 확인할 수 있다. 그림 10(b)의 YOLOv7의 히트맵은 좌측 선박이 제대로 히트 강도가 약한 것을 볼 수 있다. 반면에 그림 10(c)의 제안한 방법의 경우에는 좌측 선박에서도 높은 강도로 히트된 것을 확인할 수 있다.



(a) 원본 영상
(a) original image

(b) YOLOv7의 히트맵
(b) heat map by YOLOv7

(c) 제안한 방법의 히트맵
(c) heat map by proposed method

그림 10. YOLOv7과 제안한 방법에서 히트맵 결과
Fig. 10. Results of heat maps by YOLOv7 and proposed method

그림 11은 현재 개발 중인 유·무인수상정에서 획득한 동영상에 대해서 선박탐지를 수행한 결과이다. 왼쪽 영상은 제안한 방법의 결과이며, 오른쪽은 YOLOv7의 탐지 결과이다. 상단 영상을 보면 바다를 배경으로 중앙에서 예인선이 화물선을 끌고 있는 모습이다. 제안한 방법의 경우 화물선과 보트를 모두 탐지하였다. YOLOv7의 경우 화물선은 비슷한 확률로 탐지하였지만, 보트를 탐지하지 못했다. 하단 영상을 보면 산을 배경으로 어선이 위치하여 배경과 구분이 모호하다. 제안한 방법에서는 보트로 탐지하였지만, YOLOv7은 아무것도 탐지하지 못했다. 이를 통해서 제안한 방법이 실제 유·무인수상정에서 획득한 영상에서 선박탐지 가능성을 확인하였고, 기존 모델보다 우수함을 확인하였다.

Ⅴ. 결론 및 향후 과제

무인 기술의 발달에 따라 미래전장을 대비한 무인수상정에 관한 연구가 활발하게 진행되고 있다. 무인수상정에서 원격제어, 자율운항, 자율제어 등은 핵심 기술이며, 이러한 기술에는 공통으로 주변 환경 인식을 위한 선박탐지 기술이 요구된다. 무인수상정에서의 선박탐지는 탐지 속도가 빨라야 하고, 선박의 움직임, 형상 변화에 강인해야 한다. 이를 위해서는 딥러닝 기반의 방법이 필요하다.

본 논문에서는 무인수상정에 적용할 수 있는 실시간 객체인식 모델 YOLOv7 기반 선박탐지 방법을 제안하였다. 모델의 성능 개선을 위해 어텐션 모듈 SimAM와 동적 업샘플링 기법 DySample을 적용하였다. 그리고 제안한 모델 학습을 위해 무인수상정에서 관측할 수 있는 선박들로 구성된 공개 데이터셋들을 통합한 선박 데이터셋을 구축하였다. 구축한 데이터셋에 대한 모델의 학습과 검증을 통해 개선 모델의 성능 향상을 확인할 수 있었다. 또한, 현재 개발 중인 유·무인수상정에서 획득한 동영상에 대해 선박탐지 결과 확인을 통해 실제 적용 가능성을 확인하였다. 하지만, 일련의 입력 영상에서 선박의 형상 변화나 영상 자체의 일시적인 밝기 변화로 인해 일시적으로 탐지가 실패하는 경우가 발생하였다. 이러한 문제를 해결하기 위해서 향후에는 기존 탐지 결과를 이용하는 탐지에 의한 추적(Track by detection)을 적용할 예정이다.

그림 11. 유무인수상정에서 획득한 영상의 선박탐지 결과 : 제안한 방법(좌), YOLOv7(우)
Fig. 11. Ship detection results for videos from manned/unmanned surface vehicle : Proposed(left), YOLOv7(right)

References

[1] J. Lee and J. Na, "Required Capabilities for Maritime Manned-Unmanned Teaming", Journal of the KNST, Vol. 6, No. 3, pp. 308-313, Sep. 2023. https://doi.org/10.31818/JKNST.2023.09.6.3.308.

[2] S.-K. An and D.-S. Park, "A Study on the Autonomous Level of the Maritime Manned and Unmanned Combined Combat System", Journal of the KNST, Vol. 6, No. 3, pp. 286-292, Sep. 2023. https://doi.org/10.31818/JKNST.2023.09.6.3.286.

[3] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7476-7475, 2023.

[4] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580-587, Jun. 2014. https://doi.org/10.1109/CVPR.2014.81

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once : Unified, Real-Time Object Detection", Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, pp. 779-788, Jun. 2016.

[6] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, pp. 7263-7271, Jul. 2017.

[7] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement", arXiv preprint arXiv:1804.02767, Apr. 2018. https://doi.org/10.48550/arXiv.1804.02767.

[8] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection", arXiv preprint arXiv:2004.10934, Apr. 2020. https://doi.org/10.48550/arXiv.2004.10934.

[9] G. Jocher, "YOLOv5 by Ultralytics", 2020. https://github.com/ultralytics/yolov5 [accessed: Nov. 1, 2023]

[10] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "Yolox: Exceeding yolo series in 2021", arXiv preprint arXiv:2107.08430, Jul. 2021. https://doi.org/10.48550/arXiv.2107.08430.

[11] H. Kim and S. Ok, "Vessel Collision Warning Method using Deep Learning-based Vessel Detection and AIS", Journal of KIIT, Vol. 21, No. 11, pp. 125-132, Nov. 2023. https://doi.org/10.14801/jkiit.2023.21.11.125.

[12] J. Si, G. Kim, J. Kim, and S. Kim, "Enhanced Location-based Facility Management in Mobile Environments using Object Recognition and Augmented Reality", Journal of KIIT, Vol. 21, No. 11, pp. 183-192, Nov. 2023. https://doi.org/10.14801/jkiit.2023.21.11.183.

[13] H. Bang, J. Lee, and J. Choi, "Implementation of Fire Detection with Color Filters and Deep Learning", Journal of KIIT, Vol. 21, No. 10, pp. 137-145, Oct. 2023. https://doi.org/10.14801/jkiit.2023.21.10.137.

[14] W. Yan, W. Liu, H. Bi, C. Jiang, Q. Zhang, T. Wang, T. Dong, X. Ye, and Y. Sun, "YOLO-PD: Abnormal Signal Detection in Gas Pipelines Based on Improved YOLOv7", IEEE Sensors Journal, Vol. 23, pp. 19737-19746, Jul. 2023. https://doi.org/10.1109/JSEN.2023.3296131.

[15] Q. Zhang and X. Hu , "MSFFA-YOLO Network: Multiclass Object detection for Traffic Investigation in Foggy Weather", IEEE Transactions on Instrumentation and Measurement, Vol. 72, Sep. 2023. https://doi.org/10.1109/TIM.2023.3318671.

[16] L. Yang, R.-Y. Zhang, L. Li, and X. Xie, "SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks", Proceedings of Machine Learning Research, Vol. 139, pp. 11863-11874, Jul. 2021.

[17] W. Liu, H. Lu, H. Fu, and Z. Cao, "Learning to Upsample by Learning to Sample", Proceedings of the IEEE/CVF Conference on Computer Vision, pp. 6027-6037, 2023.

[18] J. Lee and J. You, "Ship Detection Dataset Construction for Unmanned Surface Vehicle based on Public Ship-related Dataset", Proceedings of the KIMST Autumn Conference, Jeju, Korea, pp. 273-274, Nov. 2022.

[19] B. Iancu, V. Soloviev, L. Zelioli, and J. Lilius, "ABOships-An Inshore and Offshore Maritime Vessel Detection Dataset with Precise Annotations", Remote Sensing, Vol. 13, No. 5, pp. 988. Mar. 2021. https://doi.org/10.3390/rs13050988.

[20] M. Petkovic, I. Vujovie, Z. Lusic, and J. Soda, "Image Dataset for Neural Network Performance Estimation with Application to Maritime Ports", Jouranal of Marine Science and Engineering, Vol. 11, No. 3, pp. 578, Mar. 2023. https://dio.org/10.3390/ jmse11030578.

저자소개

이 준 희 (Junhee Lee)

2011년 2월 : 금오공과대학교 컴퓨터공학부(공학사)
2013년 2월 : 한국과학기술원 전산학과(공학석사)
2013년 4월 ~ 현재 : 국방과학연구소 선임연구원
관심분야 : 딥러닝, 영상처리

유 지 상 (Jisang You)

2003년 2월 : 인하대학교 전자공학과(공학사)
2005년 2월 : 광주과학기술원 정보통신공학과(공학석사)
2020년 8월 : 한국과학기술원 전기 및 전자공학부(공학박사)
2005년 2월 ~ 현재 : 국방과학연구소 책임연구원
관심분야 : 영상/통신 신호처리, 플랫폼 원격제어

장 경 선 (Kyoungson Jhang)

1986년 2월 : 서울대학교 전자계산기공학과(공학사)
1988년 2월 : 서울대학교 대학원 컴퓨터공학과(공학석사)
1995년 2월 : 서울대학교 대학원 컴퓨터공학과(공학박사)
2001년 9월 ~ 현재 : 충남대학교 컴퓨터공학과 교수
관심분야 : 영상처리, 컴퓨터비전, 딥러닝