

개방형 객체 탐지를 활용한 딥러닝 기반 중국어선 분류 모델

조동휘*, 강상길**

Deep Learning Model for Chinese Fishing Vessel Classification using Open-Set Object Detection

Dong-Hwi Cho*, Sang-Gil Kang**

본 논문은 국방부 재원으로 정보통신기획평가원의 지원을 받아 수행된(고유번호 22900000035)
이동형 모바일 환경 인공지능을 활용한 경계감시 시스템 기술개발(A) 연구입니다

요 약

최근 수년간 한국의 배타적 경제수역(EEZ) 내에서 불법 조업을 일삼는 중국어선으로 인해 심각한 해양 수산 피해와 자원 고갈 문제가 발생한다. 수동으로 사람이 일일이 CCTV를 통해서 중국어선을 탐지하는데 시간과 비용이 크다. 이러한 문제를 해결하기 위해서는 선박 이미지를 바탕으로 중국어선을 검출 및 탐지하는 자동화 시스템이 필요하다. 본 논문에서는 개방형 객체 탐지 모델인 Grounding DINO와 이미지 분류 및 특징 추출에 효율적인 성능을 보이는 EfficientNet과 One Class SVM을 결합한 Two-Step 중국어선 분류 모델을 제안한다. 계층적 모델을 통해 Ship Feature Detection, Nationality Classification의 두 단계로 나누어서 어선의 특징을 추출하고 특징에 종류에 따른 어선 분류를 수행하게 된다. 성능 평가를 진행하여 기존의 단일 Resnet50 보다 약 9.1%, EfficientNet-B0보다 약 3.3%의 성능 개선하였음을 보인다.

Abstract

Illegal fishing by Chinese vessels within South Korea's Exclusive Economic Zone (EEZ) has caused serious damage to marine fisheries and led to resource depletion. Manually detecting these vessels by monitoring CCTV is time-consuming and costly. To address this, an automated system for detecting and identifying Chinese fishing vessels based on ship imagery is necessary. This paper proposes a two-step Chinese fishing vessel classification model that combines Grounding DINO, an open-set object detection model, with EfficientNet and a One-Class SVM. Through a hierarchical model, the process is divided into two stages: Ship Feature Detection and Nationality Classification. This involves extracting the features of the fishing vessels and then classifying them based on those features. The proposed model outperforms a conventional ResNet50 by approximately 9.1% and EfficientNet-B0 by 3.3%.

Keywords

image processing, classification, grounding DINO, EfficientNet, one class SVM, vessel

* 인하대학교 전기컴퓨터공학과
- ORCID: <https://orcid.org/0009-0009-5730-4877>
** 인하대학교 컴퓨터공학과 교수(교신저자)
- ORCID: <https://orcid.org/0000-0001-5886-0407>

• Received: Sep. 15, 2025, Revised: Nov. 04, 2025, Accepted: Nov. 07, 2025
• Corresponding Author: Sang-Gil Kang
Dept. of Computer Engineering 100 Inha-ro, Michuhol-Gu, Incheon, 22212
Republic of Korea, HighTech Center 1204
Tel.: +82-32-860-8377, Email: sgkang@inha.ac.kr

I. 서 론

최근 수년간 한국의 배타적 경제수역(EEZ) 내에서 불법 조업을 일삼는 중국어선으로 인해 심각한 해양 수산 피해와 자원 고갈 문제가 발생한다[1][2]. 이는 한국 어업 종사자들의 생계를 직접적으로 위협할 뿐만 아니라, 해양 생태계 전반에 부정적인 영향을 미친다. 해양경찰청의 단속 강화에도 불구하고, 불법 조업을 시도하는 중국어선들은 여전히 끊임없이 출몰하고 있으며, 그 규모는 매년 수만 척에 달하는 것으로 추정된다[3]. 특히, 야간이나 악천후 등 단속이 어려운 상황에서 불법 조업이 빈번하게 발생하여 효과적인 감시와 단속에 어려움을 겪고 있다[4]. 기존의 중국어선 단속은 주로 해양 경찰의 육안 관찰이나 레이더, 적외선 감시장비 등에 의존한다[5]. 그러나 이러한 방식은 넓은 해상에서 실시간으로 대규모의 어선들을 효율적으로 감시하는 데 한계가 있다[6].

선박 탐지 및 식별에 관한 기존의 연구들은 주로 위성이나 CCTV 이미지로부터 선박의 길이, 폭, 선체 형태와 같은 일반적이고 거시적인 외형 특징을 추출하여 화물선, 유조선, 어선 등 광범위한 유형으로 분류하는 데 중점을 두고 있다[7]. 그러나 이러한 접근 방식은 특정 국적의 어선만을 정밀하게 식별해야 하는 본 연구의 목적에는 한계가 있다. 일반적인 어선 분류 모델은 한국어선과 중국어선 간의 미묘한 외형적 차이를 구분하기 어렵기 때문이다.

이러한 문제를 해결하기 위해 본 논문에서는 개방형 객체 탐지 모델인 grounding DINO와 이미지 분류 및 특징 추출에 효율적인 성능을 보이는 EfficientNet과 SVM(Support Vector Machine)을 결합한 Two-Step 중국어선 분류 모델을 제안한다. 중국어선의 데이터를 충분히 확보하는 데에는 근본적인 제약이 있기 때문에 별도의 학습이 필요하지 않은 개방형 객체 탐지 모델을 사용한다. 여기에 EfficientNet과 SVM의 특징 추출 능력을 결합하여 해상 환경에서 중국어선을 정확하게 식별할 수 있도록 모델의 성능을 개선하고자 한다.

II. 관련 연구

2.1 Grounding DINO

Grounding DINO는 임의의 텍스트 프롬프트(Text prompt)를 활용하여 개방형 객체 탐지(Open-set object detection)를 수행하는 멀티모달 모델이다. 사전에 정의된 객체만 인식할 수 있었던 폐쇄형(Closed-set) 탐지 방식의 한계를 극복하기 위해 개발되었다. 이미지의 시각적 정보와 사용자가 입력한 자연어 정보를 그라운드링(Grounding)을 통해 텍스트와 이미지와의 관계를 연결한다. 이를 통해 훈련 데이터에서 한 번도 본 적 없는 새로운 종류의 객체까지도 제로샷(Zero-shot)으로 정확하게 탐지하고 해당 위치를 특정할 수 있게 된다[8]. 본 논문에서는 grounding DINO를 어선에서 볼 수 있는 특징 중에서 선박 자체와 정보를 담고 있는 깃발 영역을 추출하기 위해서 해당 모델을 사용한다.

2.2 EfficientNet

EfficientNet은 합성곱 신경망의 효율적인 연산을 위해 개발된 모델이다. 기존 방식이 네트워크의 깊이, 너비, 해상도 중 한 가지만 키워서 성능을 높이는 것이 아닌, 세 가지 차원의 균형을 맞추는 복잡 스케일링 방법을 사용한다. 적은 파라미터와 연산량으로도 높은 정확도를 보여주며, 기존의 대규모 모델보다 효율적인 성능을 보여준다[9]. 본 논문에서는 국적별 어선에서 보이는 깃발의 특징을 학습하여 어선의 국적을 구분하는 모델로 사용한다.

2.3 One class SVM

One class SVM(Support vector machine)은 비지도 학습 기반 분류 알고리즘으로, 주어진 데이터를 고차원 특징 공간(Feature space)으로 대치한 뒤 최적의 초평면(Hyperplane)을 찾아 데이터 간의 경계를 결정한다. 특히, 입력 데이터를 비선형적으로 변환하는 커널 함수(Kernel function)를 활용함으로써 선형적으로 구분되지 않는 문제 또한 해결할 수 있다[10]. 이러한 특성으로 패턴 인식, 문서 분류, 영상 처리 등 다양한 분야에서 활용하고 있다. 본 논문에서

서는 깃발이 검출되지 않은 어선을 분류하기 위해 사용한다. 수집된 데이터 중 한국어선의 비중이 중국어선보다 높아서, 한국어선 데이터를 기준으로 국적 분류를 진행하고자 한다.

III. 연구 방법

본 논문에서 제안하는 우리의 시스템은 중국어선을 분류하기 위해서 그림 1과 같이 계층적 모델을 구성한다. 전체 과정은 ship feature detection과 nationality classification의 두 단계로 구분되어 있다.

3.1 Ship feature detection

Ship feature detection 단계에서는, 그림 2와 그림 3에서 보듯이 원본 어선 이미지가 들어오면 입력된 어선 이미지를 대상으로 어선과 깃발의 위치를

grounding DINO를 통해서 object detection을 진행한다. 원본 이미지와 해당 이미지에서 추출할 대상인 선박과 깃발을 text prompt에 'a ship', 'a flag'로 지정하여 입력으로 사용하면, 모델은 입력 이미지 내에서 어선과 깃발 후보 영역을 탐지하여 해당 좌표와 그에 대한 레이블 정보를 제공한다. 이때 다음 단계인 nationality classification을 위해 그림 2와 같이 원본 이미지와 grounding DINO가 탐지한 annotation 정보를 같이 제공한다.

Algorithm 1 Ship Feature Detection

```

1: function SHIPFEATUREDETECTION(original_image)
2:   text_prompts ← {"a ship", "a flag"}
3:   detectedObjects ← GroundingDINO_Model.detect(
     original_image, text_prompts)
4:   annotations ← detectedObjects
5:   return original_image, annotations
6: end function

```

그림 3. Ship feature detection 의사결정코드
Fig. 3. Ship feature detection pseudo code

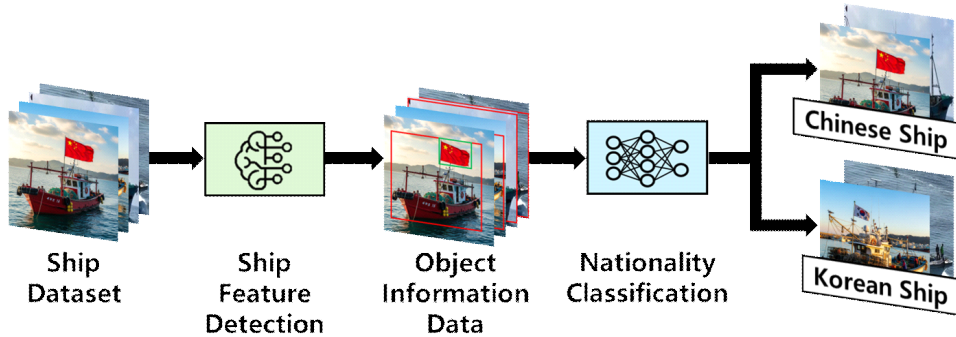


그림 1. 모델 흐름도
Fig. 1. Model process

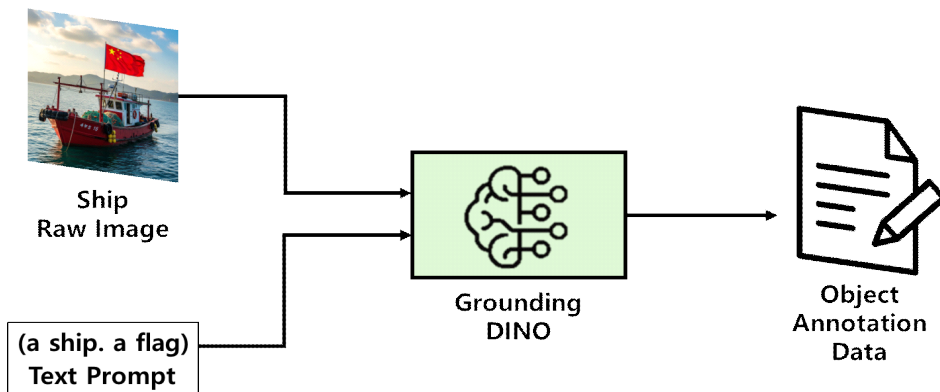


그림 2. Ship feature detection 구조도
Fig. 2. Ship feature detection architecture

3.2 Nationality classification

Ship feature detection 단계에서 얻은 원본 어선 이미지와 Annotation 정보를 기반으로 그림 4 및 그림 5와 같이 nationality classification 단계를 수행한다. 먼저 annotation 정보를 통해 깃발 검출 여부를 확인한다. 깃발은 국적 식별에 직접적인 단서를 제공하기 때문에 별도의 특징 추출 과정 없이 깃발의 ROI(Region of Interest)를 EfficientNet에 입력으로 사용하여 분류를 진행한다. 반면, 깃발이 미검출 된 경우에는 선박 영역의 ROI를 사용한다.

Algorithm 2 Nationality Classification

```

1: function NATIONALITYCLASSIFICATION(image, detected_annotations)
2:   flag_annotation ← FindAnnotationByLabel(
     detected_annotations, "a flag")
3:   if flag_annotation exists then then
4:     flag_roi ← ExtractRoi(
       image, flag_annotation.coordinates)
5:     classification_result ← EfficientNet_Model.classify(flag_roi)
6:   else
7:     ship_annotation ← FindAnnotationByLabel(
       detected_annotations, "a ship")
8:     ship_roi ← ExtractRoi(
       image, ship_annotation.coordinates)
9:     classification_result ← OneClassSVM_Model.classify(ship_roi)
10:  end if
11:  final_result ← classification_result
12:  return final_result
13: end function

```

그림 5. Nationality classification 의사결정코드

Fig. 5. Nationality classification pseudo code

이때 데이터 수집의 한계로 인해 한국어선의 특징만을 반영한 one class SVM 모델을 사용하여 분

류를 진행한다. 최종적으로 EfficientNet과 SVM에서 분류된 결과를 병합하여 중국어선 분류 결과를 도출한다.

IV. 실험 결과

본 연구에서는 Ubuntu 20.04 Local, AMD EPYC 7413 24-Core Processor, 64GB RAM, NVIDIA RTX A6000, PyTorch 2.6.0 환경에서 진행하였다.

4.1 Dataset construction

표 1, 표 2 및 그림 6에서 보듯이 여러 기상 상황에서 발생할 수 있는 데이터를 자체적으로 수집하여 깃발의 유무, 해상에 환경을 고려한 데이터셋 구축을 진행하였다. 학습 및 평가에 사용한 데이터는 자체적으로 대천항에서 고정형 CCTV를 이용하여 수집한 한국어선 이미지 3,000개와 구글 크롤링을 활용하여 수집한 중국어선 이미지 100개를 사용하였다. 표 3에서 보듯이 데이터셋 분할은 8:1:1의 비율로 나누어서 진행하였다.

표 1. 깃발 유무에 따른 수집된 이미지 개수

Table 1. Number of images collected with or without flag

Type	Chinese	Korean
No Flag	27	1,387
Flag	73	1,613
Total	100	3,000

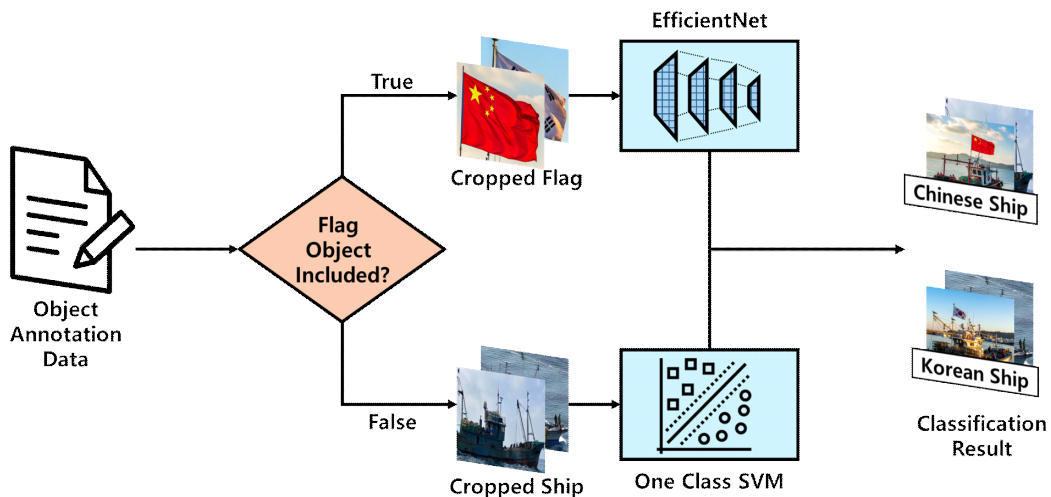


그림 4. Nationality classification 구조도

Fig. 4. Nationality classification architecture

표 2. 환경에 따른 수집된 이미지 개수

Table 2. Number of images collected depending on the environment

Type	Chinese	Korean
Normal	64	2,674
Wave	21	189
Fog	15	137
Total	100	3,000

표 3. 데이터셋 분할 비율에 따른 이미지 개수

Table 3. Number of images according to dataset split ratio

Type	Train	Valid	Test
Korean	2,400	300	300
Chinese	80	10	10



그림 6. 환경별 선박 데이터 예시

Fig. 6. Example fishing ship data by enviroment

평가를 진행하기 위한 별도의 레이블링 작업은 그림 7에서 보듯이 label-studio를 사용하여 작업, 선박 영역과 깃발 영역을 bounding box를 사용하여 표시하였다. Ship feature detection 단계에서 object detection의 결과를 확인하기 위해서 COCO dataset의 JSON 형식으로 저장하여 총 4,000개의 데이터를 구축하여 진행하였다.



그림 7. Label-studio를 이용한 바운딩 박스 레이블링

Fig. 7. Bounding box labeling with label-studio

4.2 Ship feature detection

Ship feature detection 단계에서는 HuggingFace 라이브러리를 사용하여 grounding DINO를 구축하여 진행하였다. 자체적으로 구축한 데이터셋의 Valid 데이터셋을 바탕으로 해당 이미지에 나와 있는 선박과 깃발의 검출 여부를 통해서 mAP@.50 지표를 기준으로 측정을 진행하였다. 모델의 threshold는 grid search를 통해서 가장 최적의 값을 탐지하였으며, 표 4에서 보이는 것처럼 0.1~0.2에서의 모델의 성능이 가장 우수하게 나왔다. 하지만 threshold가 너무 낮으면 노이즈 부분에 대해서도 바운딩 박스 결과를 도출하기 때문에 ship feature detection 모델에서는 0.2의 값을 사용하여 진행한다. 이외의 threshold에서는 grounding DINO에서 낮게 나온 결과들이 검사 대상에 포함이 되지 않아서 최종 mAP@.50에 영향을 준 것을 확인하였다.

표 4. Threshold별 grounding DINO 정확도

Table 4. Grounding DINO accuracy with threshold

Threshold	mAP@.50
0.1	83.1%
0.15	82.9%
0.2	82.8%
0.25	82.1%
0.3	80.4%

4.3 Nationality classification

Nationality classification 단계에서는 HuggingFace에서 제공하는 EfficientNet-B0 모델을 사용하여 학습에 사용하였으며, one class SVM은 scikit-learn에서

제공하는 모델과 torchvision에서 제공하는 ResNet50 모델을 사용하여 진행하였다. 학습 정확도는 valid 데이터셋과 F1-Score를 사용하여 평가하였다. EfficientNet 모델은 성능과 효율성을 고려해서 파라미터의 수가 적은 Efficient-B0 모델을 사용하여 진행하였으며, one class SVM의 특징은 ResNet을 사용하여 특징을 추출하여 진행하였으며, grid search를 통해 최적의 파라미터를 탐색하였다. 표 5에서 보이는 것처럼 ‘poly’ Kernel, ‘0.05’ Nu, ‘0.009’ Gamma의 파라미터를 사용했을 때 가장 좋은 성능을 보여주었다. 특히 동일 파라미터 조건에서 ‘poly’ Kernel이 ‘linear’ Kernel 보다 약 0.4%의 좋은 성능을 보여주고 있다.

표 5. 파라미터 별 one class SVM 정확도
Table 5. One class SVM accuracy with parameters

Kernel	Nu	Gamma	F1-Score
poly	0.05	0.009	90.3%
linear	0.05	0.009	89.9%
linear	0.1	0.009	88.0%
poly	0.1	0.009	87.3%
sigmoid	0.15	0.009	87.0%
sigmoid	0.15	0.019	86.8%

4.4 실험 결과

Ship feature detection에서 찾은 파라미터 값과 nationality classification에서 찾은 파라미터 값을 적용해서 모델의 성능을 평가하였으며, 단일 ResNet, 단일 EfficientNet, 단일 one class SVM 모델과 비교한 결과를 표 6과 같이 얻어 낼 수 있었다.

기존 one class SVM을 사용했을 때보다 18.2%의 F1-Score 상승을 보여주었으며, 단일 ResNet50 대비 9.1%, 단일 EfficientNet-B0 대비 3.3%의 F1-Score 상승률을 확인 할 수 있었다.

표 6. 분류 모델 결과 비교
Table 6. Compare result with classification models

Model	Precision	Recall	F1-Score
ResNet50	86.0%	84.8%	85.4%
EfficientNetB0	91.5%	90.9%	91.2%
One class SVM	77.1%	75.5%	76.3%
Ours	94.6%	94.4%	94.5%

일부 데이터에서 오분류가 발생한 경우는 그림 8과 같이 한국어선에서 식별용으로 사용하는 깃발이 중국어선의 오성홍기와 유사한 색상으로 적용이 되어 있어서 오인하는 경우가 있었다.



그림 8. 오분류가 발생한 한국어선 이미지
Fig. 8. Korean ship image with misclassification

V. 결론 및 향후 과제

본 연구에서는 개방형 객체 탐지를 적용하여 중국어선 분류 모델을 개발하고, 기존 분류 모델 대비 성능의 우수성을 입증하였다. 특히, 대용량의 데이터만을 이용하지 않더라도 중국어선과 한국어선을 분류할 가능성을 보여주었다.

단계별로 grid search를 통해 최적의 파라미터를 설정하고 기존의 단일 모델과의 결과 비교를 통해 단일 EfficientNet-B0의 정확도 보다 3.3%의 정확도가 향상되었다.

향후 연구에서는 grounding DINO에서 모델 경량화 기법인 양자화 기법을 적용하여 실제 해상 환경에서 사용하는 CCTV의 해상도 및 주사율 기준인 30FPS@2160p에서 운영할 수 있는 수준(약 50ms의 반응속도)으로 모델을 고도화할 것이다. 또한 데이터 불균형을 추가로 고려해서 다양한 데이터 증강 기법을 고려하여 모델의 신뢰성을 향상할 것이다.

References

- [1] B. K. Jung, "A Study on the Improvement of Control Effectiveness against Illegal Chinese Fishing Boat of Maritime Enforcement Organization", The Journal of Fisheries and Marine Sciences Education, Vol. 26, No. 6, pp.

- 1695-1706, Dec. 2017. <https://doi.org/10.13000/JFMSE.2017.29.6.1695>.
- [2] M. S. Ko, "A Study on the Legal Issues of Illegal Chinese Fishing Vessels", Korean Association of Maritime Police Science, Vol. 14, No. 4, pp. 195-215, Dec. 2024. <https://doi.org/10.30887/jkmps.2024.14.4.195>.
- [3] H. R. Roh, "Illegal Fishing of Chinese Fishing Boat and Countermeasure of Korean Maritime Police", Korean Police Studies, Vol. 9, No. 2, pp. 29-58, Jan. 2010.
- [4] Ministry of Oceans and Fisheries and Korea Coast Guard Join Forces to Combat Illegal Fishing by Chinese Vessels, <https://www.mof.go.kr/doc/ko/selectDoc.do?docSeq=34877&menuSeq=971&bbsSeq=10>. [accessed: Sep. 15, 2025]
- [5] The Issue of Illegal Chinese Fishing Vessels Before and After COVID-19: What Has Changed and What is the Outlook?, <https://kims.or.kr/issubrief/kims-periscope/peri357/>. [accessed: Sep. 15, 2025]
- [6] C. H. Lim, "A Study on the Law Enforcement of Korea Coast Guard against the Illegal Chinese Fishing Vessels", Journal of The Korean Society of Marine Environment & Safety, Vol. 20, No. 1, 49-58. Jan. 2014. <https://doi.org/10.7837/kosomes.2014.20.1.049>.
- [7] R. D. Kwon, G. R. Nam, J. Tak, J. Shin, H. Cha, and S. W. Lee, "Enhancing Ship Classification in Optical Satellite Imagery: Integrating Convolutional Block Attention Module with ResNet for Improved Performance", IEEE Access, Aug. 2024. <https://doi.org/10.48550/arXiv.2404.02135>
- [8] S. Liu, Z. Zeng, T. Ren, F. Li, H. Zhang, J. Yang, and L. Zhang, "Grounding dino: Marrying dino with grounded pre-training for open-set object detection", European Conference on Computer Vision., Milan, Italy, pp. 38-55, Sep. 2024. <https://doi.org/10.48550/arXiv.2303.05499>.
- [9] M. Tan and Q. Le, "Efficientnet: Rethinking

model scaling for convolutional neural networks", International Conference on Machine Learning., Los Angeles, United States of America, pp. 6105-6114, May 2019. <https://doi.org/10.48550/arXiv.1905.11946>

- [10] C. Cortes and V. Vapnik, "Support-vector networks", Machine Learning, Vol. 20, No. 3, pp. 273-297, Sep. 1995.

저자소개

조 동 휘 (Dong-Hwi Cho)



2025년 2월 : 인하대학교
소프트웨어융합공학과 (공학사)
2025년 3월 ~ 현재 : 인하대학교
대학원 전기컴퓨터공학과
석사과정
관심분야 : 머신러닝, 딥러닝,
영상처리

강 상 길 (Sang-Gil Kang)



1989년 2월 : 성균관대학교
전기공학과(공학사)
1995년 5월 : Columbia University
전기공학과 전자공학과
(공학석사)
2002년 12월 : Syracuse University
전자공학과(공학박사)
2006년 9월 ~ 현재 : 인하대학교 컴퓨터공학과 교수
관심분야 : 모바일 컴퓨팅, 인공지능 시스템, 정보검색,
신경 회로망, 멀티미디어 시스템, 유비쿼터스 시스템,
신호처리