

접수번호

※ 해당 없음

「해양수산 정보서비스·비즈니스 아이디어 공모전」 기 획 서(빅데이터 분석 부문)

* 해당란에 ☒ 표시

1. 참가자 정보

개인·팀·기업명	서울과학기술대학교 (김재호, 이정언), 인천대학교 (김민성)	
서비스(제품)명	유사어종의 이미지 분류를 통한 수산물 거래 시장의 정보비대칭 개선 (능성어와 다금바리 등 바리와 품종의 분류를 중심으로)	
공모 분야	<input type="checkbox"/> 해양 환경·기후·안전 <input type="checkbox"/> 해운·항만 물류 <input type="checkbox"/> 해양·어촌 관광·체험·레저 <input type="checkbox"/> 낚시·귀어귀촌·양식 <input checked="" type="checkbox"/> 수산물 유통·안전·품질인증	
서비스 유형	<input checked="" type="checkbox"/> 모바일(태블릿pc) 앱·웹 서비스 <input type="checkbox"/> 제품제작 <input type="checkbox"/> 온라인 웹 서비스 <input type="checkbox"/> 영상 <input type="checkbox"/> 웨어러블디바이스 <input type="checkbox"/> 기타 ()	

2. 기획서 작성

1. 개요

1.1) 빅데이터 분석 핵심내용

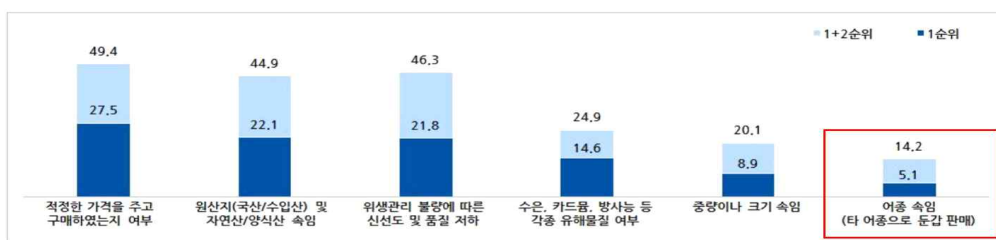
1. 배경

- 「KMI 수산시장의 소비자 신뢰 제고를 위한 정책 지원방안 연구보고서 - 2017년 제 4 호」에 따르면, 수산시장에서 수산물 거래시 위생, 안전, 적정가격, 원산지, 중량(크기), 어종 속임 등을 불안 요소로 파악
- 특히 가격, 중량, 어종 속임과 같이 비용과 직결되는 요소들은 수산물 판매자와 소비자 사이에 정보가 비대칭하기 때문에 발생 될 수 있음
- 이에 소비자들이 안심하고 수산물을 구매하기 위해서는 구매하고자 하는 수산물에 대하여 양질의 정보가 제공되어야 하며, 관련하여 다양한 빅데이터 분석과 이를 활용 할 수 있는 서비스 개발이 필요할 것으로 판단함

2. 목적

- 이와 같은 배경에서 수산물거래 판매자와 소비자의 정보비대칭 개선과 신뢰 제고를 위한 빅데이터 분석이 목적임
- 특히, 수산물 구매와 관련하여 조사 인원의 14.2%가 어종 속임(타 어종으로 둔갑 판매)에 대하여 불안해 하고 있는데, 어종 속임은 소비자를 기만하는 행위로 다른 불안 요소들보다 비율은 낮지만, 절대 발생되어서는 안되는 중대한 요인이라 판단함

수산시장에서 수산물 구매 시 불안 요소



자료 : 수산시장에 대한 대국민 설문조사 결과(2017년 4월 12~18일, 국민 1,000명 대상)

그림1. KMI 연구보고서 수산시장의 소비자 신뢰 제고를 위한 정책 지원방안 연구 <2017년 제 4호>

- 결론적으로, 진위 구별이 어려운 점을 악용해 저가의 수산물을 고가로 둔갑해 판매하는 사례가 있는 유사어종을 선정하여 이미지 분류(Image Classification) 빅데이터 분석을 진행하고자 하였음

3. 데이터 선정

○ 유사 어종 선정

- 유사어종의 이미지 분류를 위하여 방어-부시리, 참조기-부세, 민어-큰민어, 다금바리(자바리)-능성어 등 형태 또는 명칭이 유사하고, 육안으로 구분하기 어려워 악용되고 있는 품목 중 가격 차이가 크고, 활어 중심으로 유통되고 있는 다금바리(자바리)와 능성어를 모델 학습 및 테스트 어종으로 선정

종류	방어	참조기	민어	다금바리(자바리)
유사어종	- 부시리	- 부세 - 긴가이석태 - 보구치	- 큰민어 - 영상가이석태 - 점성어(홍민어)	- 능성어 - 대왕자바리 - 대왕범바리 등

표1. 유사어종의 종류

○ 다금바리와 능성어

- 우리가 아는 다금바리는 제주도 방언으로 도감상의 표준명은 ‘자바리’이며, 표준명으로 기술되고 있는 ‘다금바리’는 국내에서는 발견하기 어려운 품종임.
(이하 ‘다금바리’는 ‘자바리’와 같은 생선으로 분류)



종류	다금바리	자바리	능성어
사진			
가격	확인불가	kg당 150,000원 이상	kg당 70,000원 이상

표2. 다금바리, 자바리, 능성어 비교

- 자바리와 능성어는 외형은 비슷하지만 체형은 능성어가 더 통통하고, 자바리는 날씬한 편. 자바리는 불규칙한 호피무늬가 나타나며 능성어는 일곱줄의 가로 줄무늬로 구분 단 둘다 성어로 자라면 이러한 무늬들이 흐릿해지면서 구별이 어려워질 수 있음



표3. 블로그 ‘입질의 추억’ - 제주 다금바리와 능성어의 구별법 참고

- 방송사 연예인과 유명 수산물 유투버도 다금바리(자바리)를 구매하였으나, 실제로는 능성어이거나 교잡종이 둔갑된 상태로 판매가 되어서 논란이 되었던 사례도 있었음



그림2. 매일경제(좌측), 인플루언서닷컴(우측) ‘다금바리’ 관련 기사 중

- 원산지 표시가 시행되기 이전에는 능성어 품종이 다금바리로 둔갑된 채 판매되는 것이 성행하였고, 최근에는 다금바리를 사칭하는 어종으로 중국산 교잡종인 하이퍼 그루퍼들이 대거 수입되면서 육안으로만 자바리를 식별하는 것이 더욱 어려워졌음

종류	대왕바리	대왕범바리	대왕자바리	갈색등근바리	홍기흑점바리	붉바리
바리와 유사어종						

표4. 바리와 유사어종

○ 기타 어종 (바리와 유사어종 및 AI-Hub 어류 개체 촬영 이미지 5종 등)

- 위와 같은 문제로 실제 자바리(다금바리)를 구매하려고 했을 때, 능성어 외 바리와 유사어종은 아닌지 식별도 필요하고, 또한 활어를 판매하는 수산시장의 수조에는 다른 종류의 어종들도 혼합하여 있으므로, 주 분류 대상인 자바리와 능성어 외 바리와 유사어종 7종과 AI-Hub에서 제공하는 있는 넙치, 조피볼락 등 5종의 어류 개체 촬영 이미지 도 추가 활용하여 학습 모델의 신뢰성을 확보하고자 하였음

4. 데이터 수집

○ 공공데이터 수집

- 국립수산물과학원 수산생명자원정보센터, 해양과학기술원 사진관리 DB시스템 등 수산자원에 대한 정보를 제공해주는 기관 서비스를 통하여 분석 대상에 대한 이미지를 간헐적으로 수집 (각 기관별 1~2장 내외의 이미지를 제공)

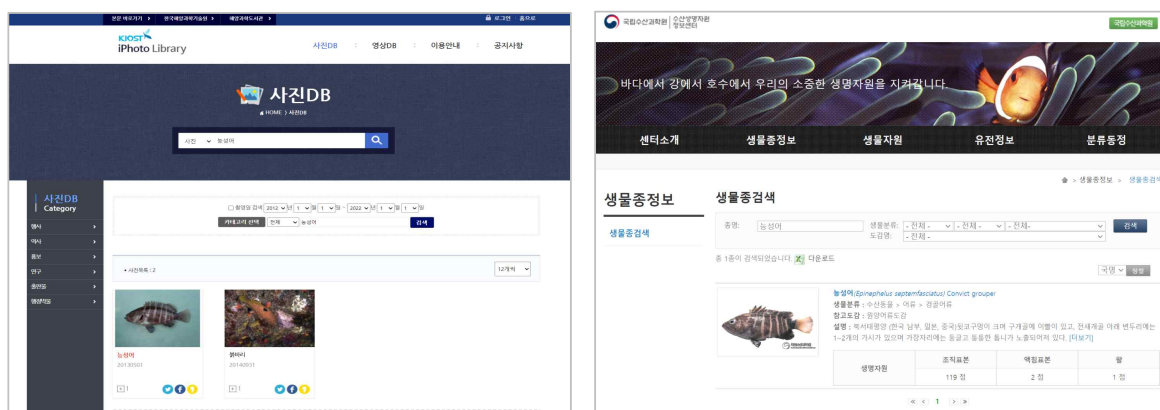


그림3. 한국해양과학기술원 사진관리DB (좌측), 국립수산물과학원 수산생명자원정보센터 (우측)

- 넙치, 조피볼락 등 5종의 기타 어종은 AI-Hub에서 제공해주고 있는 어류 개체 촬영 이미지 데이터셋에서 자바리, 능성어 학습데이터 셋에 비례하여 랜덤으로 추출하였음

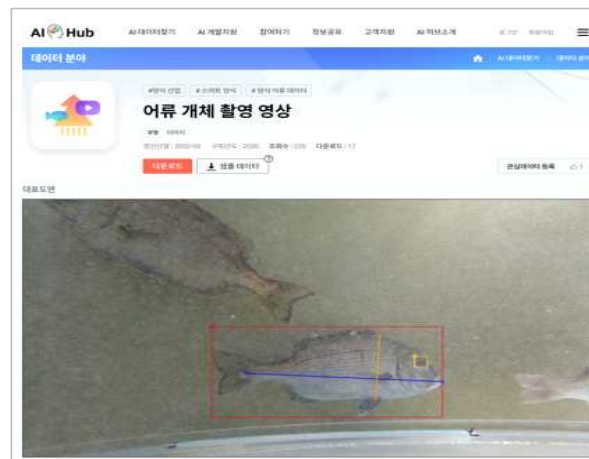


그림4. AI-Hub 어류 개체 촬영 영상 데이터

○ 민간데이터 수집 (웹 크롤링 및 레이블 식별)

- 다금바리(자바리) 및 능성어와 바리과 유사어종에 대하여 공공데이터로 수집할 수 있는 이미지가 지극히 제한적인 관계로 온라인(구글, 네이버, 등), 유튜브 검색 등을 통하여 웹 크롤링 및 레이블 식별을 진행하면서 대부분의 학습데이터를 수집하였음

○ 최종 학습 데이터

- 공공데이터와 민간데이터를 포함하여 최종 학습 데이터는 다음과 같이 구성
- 유사어종(바리과)는 어종별 이미지가 많지 않은 관계로 하나의 Class를 병합하였음
- 능성어 : 공공데이터 및 민간데이터 포함 160장 구성
- 다금바리(자바리) : 공공데이터 및 민간데이터 포함 160장 구성
- 유사어종(바리과) : 갈색둥근바리, 다금바리(농어과), 대왕바리, 대왕범바리, 대왕자바리, 붉바리, 홍기흑점바리 등 총 7종에 대하여 각 30장씩 수집하여 구성
- 기타 (AI-Hub) : 넙치, 조피볼락, 참돔, 감성돔, 돌돔 등 총 5종의 어류 개체 촬영 영상 데이터에서 198개 이미지를 추출하여 구성

Class	능성어	다금바리 (자바리)	유사어종 (바리과 7종)	기타 (AI-hub)	합계
데이터 개수	160	160	210	198	728

표5. 최종 학습데이터 구성

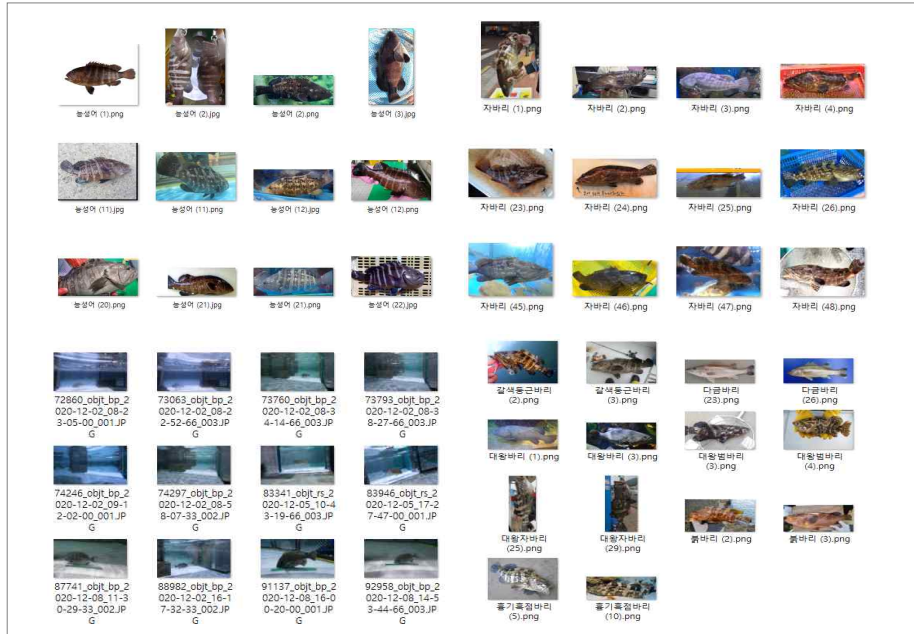


그림5. 최종 학습 데이터 Samples

5. 데이터 분석

○ Data Augmentation

- 총 6가지 Augmentation을 이용하여 학습을 진행. 모든 이미지는 300x300 크기로 자른 후 $[-1, 1]$ 사이값으로 normalization을 하여 전처리하여 사용함

1. Resize and Center Crop
2. RGBShift
3. Blur
4. Affine transform
5. Rotate
6. Horizontal Flip



그림6. Data Augmentation Samples

○ 사용 모델 및 학습 하이퍼 파라미터

- ImageNet 데이터로부터 사전학습된(pretrained) EfficientNet-B7 모델을 이용하여 학습을 진행하였으며, 학습을 위한 하이퍼 파라미터는 아래 표 6와 같음

하이퍼파라미터	epoch	Batch Size	Learning rate	optimizer	Loss function
값	30	4	0.00005	AdamW	Cross Entropy loss

표6. EfficientNet-B7 학습 파라미터

○ 실험방법

- 실험은 아래와 같이 총 2가지 방법으로 진행하였음
 - 실험1 : 능성어, 다금바리 분류만을 목적으로한 Binary Class 이미지 분류
 - 실험2 : 4가지 어종 모두를 사용한 Multi Class 이미지 분류
- 테스트는 총 5번의 fold를 나눠서 평가지표 측정

6. 실험 결과

○ 실험1 : 능성어, 다금바리 2개 Class 분류

Fold	Accuracy	f1	precision	recall
1-fold	0.89	0.482	0.5	0.117
2-fold	0.95	0.482	0.5	0.117
3-fold	0.92	0.4825	0.5	0.117
4-fold	0.89	0.4702	0.5	0.1132
5-fold	0.9	0.482	0.5	0.117
평균	0.91	0.480	0.5	0.116

표7. 능성어, 다금바리 2개 Class 분류 모델 성능표

○ 실험2 : 능성어, 다금바리, 유사어종, 기타 4개 Multi Class 분류

Fold	Accuracy
1-fold	0.92
2-fold	0.94
3-fold	0.87
4-fold	0.88
5-fold	0.91
평균	0.90

표8. 능성어, 다금바리, 유사어종, 기타 4개 Multi Class 분류 모델 성능표

○ Confusion matrix, ROC-AUC curve 시각화

- 실험 1, 2에서 각 fold의 결과가 비슷하기 때문에 1-fold에 대해서만 Confusion matrix 시각화 진행

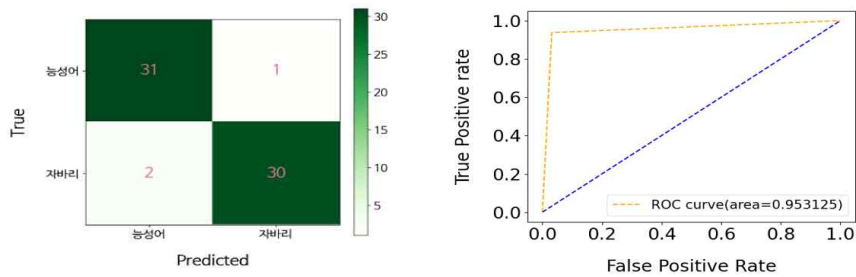


그림7. 실험1 - Confusion matrix & Roc-Auc Curve 시각화

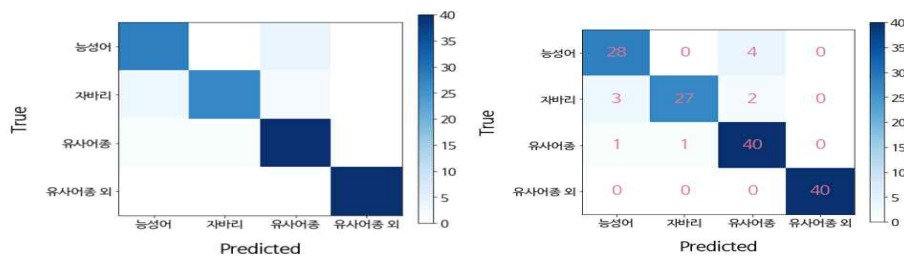


그림8. 실험2 - Confusion matrix 시각화

6. 결론 및 고찰

- 수산물 시장에서 판매자와 소비자의 정보비대칭 개선과 신뢰 제고를 위하여 능성어와 다금바리 등 유사어종에 대한 이미지 분류를 진행
- 학습모델은 ImageNet 데이터로부터 사전학습된(pretrained) EfficientNet-B7 모델을 이용하여 총 5번의 fold를 나눠서 테스트 하였으며, 능성어-다금바리의 이진분류와 바리과 유사어종과 기타어종 모두를 포함한 Multi Class 분류 2가지로 진행하였음
- 두 실험 모두 평균 Accuracy가 0.90 내외로 우수한 이미지 분류 성능을 보여주었음

○ 한계 및 시사점

- 공공데이터를 통하여 수집할 수 있는 유사어종의 이미지가 제한적이었음
- 민간데이터 또한 어종별 이미지가 충분하지 않았던 관계로 학습데이터를 준비하는데 어려움이 있었으며, 현재보다 더 많은 데이터를 수집하지 못한 것에 아쉬움이 있었음
- 이번 실험을 통하여 유사어종 분류에 대한 필요성을 제시하였고, 데이터는 적었지만 프로토타입 형태로 사업의 타당성을 검토하는데는 의미가 있다고 판단함
- 민어, 참조기 등 다른 어종들과 생태환경(자연-양식), 원산지별 이미지 분류도 기대
- 수산생물 수출입검역이나 유사어종 단속 등에 활용도 기대

1.2) 분석배경 및 데이터의 활용 적정성

- 분석 배경 1.1 개요 페이지 작성 내용 참고
- 데이터 출처 (공공데이터 + 민간데이터)
- 활용빈도
 - 공공데이터 : 아래 1~7번까지 각 기관별로 제공하는 이미지가 1~2장 내외로
자바리-능성어 학습 데이터의 10% 미만 수집
 - 민간데이터 : 웹 크롤링 및 라벨링을 통하여 90% 이상 수집
(유튜브는 직접 영상을 시청하면서 스크린샷 이미지를 수집)
- 따라서 유사어종을 이미지 분석할 수 있는 공공데이터 확보가 필요함

구분	데이터 출처		주소
1	국립수산물연구원	수산생명자원정보센터 (어종 검색 이미지)	https://www.nifs.go.kr/frcenter/
2	한국해양과학기술원	사진관리 DB시스템 (어종 검색 이미지)	https://ipphoto.kiost.ac.kr/www/index.do
3	한국지능정보사회진흥원	AI-Hub (어류 개체 촬영 영상)	https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=154
4	농림수산식품교육문화정보원	생명자원정보서비스 (어종 검색 이미지)	https://www.bris.go.kr/portal/main/main.do
5	한국생명공학연구원 KOBIS	국가생명연구자원 통합정보시스템 (어종 검색 이미지)	https://www.kobic.re.kr/kobis/main.do
6	환경부	국립생물자원관 (어종 검색 이미지)	https://species.nibr.go.kr/index.do
7	식품의약품안전처	수산식품원료의 형태학적 및 유전학적 분류 매뉴얼(III) (능성어, 자바리)	https://impfood.mfds.go.kr/CFBDD06F02/getCntntsDetail?cntntsSn=456200
8	일반 검색 (구글, 네이버, 웹페이지, 블로그, 페이스북, 인스타그램 등)	어류도감	http://fishillust.com/home
		추라우미생물도감	https://churaumi.okinawa.kr/fishbook/
		:	:
		인어교주해적단	https://www.tpirates.com/
9	유튜브	입질의추억TV jiminTV	https://youtu.be/IRd1BepgLzM
		수빙수tv sooBingsoo	https://youtu.be/P5pB4aS2K8U
		:	:
		거제도 수아아빠TV_생선회 sashimi 이야기	https://youtu.be/q3pYLakF-0g
		순천영일수산tv(sashimi)	https://youtu.be/2_iipkT-bWY

표9. 데이터 출처

1.3) 기존 서비스와의 차별성 및 독창성

○ 차별성 및 독창성

- 대다수의 어종 이미지 분석은 국내 생태계를 위협하는 외래종 식별과 낚시 또는 조업 활동으로 한 어획물에 대하여 식별 및 품목 분류를 목적으로 하고 있었음
- 우리 실험은 수산물 거래시장에서 소비자들을 대상으로 안심하고 수산물을 구입할 수 있도록 하는 것을 목적으로 하고 있음
- 그동안 인공지능을 통하여 이미지 분류 대상이 되었던 어종들 중 다금바리-능성어 비교 연구는 유일한 것으로 확인이 되며, 교잡종까지 포함된 연구는 더욱 찾기 어려웠음

구분	기존 연구들	본 연구
대상	- 어업인 - 조업인	- 수산물 거래시장 판매자 - 수산물 거래시장 소비자
목적	- 외래종 확인 및 퇴치 - 어획물 품종 확인 및 분류	- (판매자) 진품 인증 - (소비자) 유사어종 식별
어종 종류	- 큰입배스, 블루길 잉어, 붕어, 메기, 쏘가리, 누치 - 연어, 고등어, 방어 - 참돔, 우럭, 배스, 쏘가리	- 다금바리, 능성어 외 대왕범바리, 갈생등근바리 등 교잡종이 포함

표10. 기존 연구 차별성 비교

2. 사업화

2.1) 분석된 정보를 활용한 사업계획

- 수산물 유통관련 앱에 판매자가 진품, 유사어종을 인증하는 서비스
- 수산물 유통관련 앱에 구매자가 진품, 유사어종을 확인할 수 있도록 하는 서비스
- 낚시, 조업활동 관련 앱에 어종 인증을 진행할 수 있는 서비스
- 수산물 식품관리를 하는 공공기관 등에 검사 시스템으로 활용할 수 있는 서비스 등 고려

2.2) 자유타이틀 기재

-

3. 기타

3.1) 추가로 제공이 필요한 데이터(필요 시 작성)

-