

Explainable artificial intelligence models for enhancing classification reliability of ground weapon systems

Bae, Gimin* · Lee, Janghyong**

ABSTRACT

This study focused on the development of a reliable artificial intelligence (AI) model to enhance the classification reliability of ground weapon systems for surveillance and reconnaissance applications. The proposed AI model overcomes the limited data availability of military objects such as tanks, canons, and multiple-launch rockets by leveraging transfer learning and fine-tuning techniques. A comprehensive evaluation of 35 deep learning models using the publicly available Military-Vehicles dataset on Kaggle identified MobileNet as the most suitable model for ground weapon system classification. The selected MobileNet model achieved an average F1 score of 92% when tested on a dataset comprising five types of ground-weapon systems. In addition, the application of the explainable AI technique Grad-CAM provided insights into the decision-making process of the proposed model and verified its reliability. Real-world evaluations using frames extracted from training videos demonstrated promising accuracy for tanks, canons, and multiple-launch rockets. However, challenges related to object occlusion and the absence of target objects in the images were observed, which resulted in misclassifications. Overall, this study contributes to the development of explainable and reliable AI models for enhancing the performance of ground surveillance and reconnaissance systems.

Keywords : classification of ground weapon systems, explainable artificial intelligence, transfer learning, MobileNet, Grad-CAM

* (First Author) Korea Army Research Center for Future and Innovation, Corporal, rlals3501@gmail.com, <https://orcid.org/0000-0001-6587-4209>.

** (Corresponding Author) Korea Army Research Center for Future and Innovation, Lieutenant Colonel, leejangh@gmail.com, <https://orcid.org/0009-0003-2751-7560>.

I. 서론

미래 전의 양상이 다영역작전(Joo, J. Y., 2020), 모자이크전(Nam et. al, 2020), 유무인복합전(Lee, J. Y., 2018)으로 변모하면서 감시정찰체계가 핵심적인 필수 요소로 부상하고 있다. 왜냐하면 감시정찰체계¹⁾는 목표지점이나 대상 물체를 장기간 지속적으로 관찰하는 행위인 감시와 단기간에 정세나 지형을 살피는 정찰을 수행하는 체계로 전장에서 눈과 귀의 역할을 담당하면서 수집·전송된 정보를 기반으로 작전실행 결심을 지원하기 때문이다(Choi, 2021). 해당 체계는 레이더 무기체계, SAR 무기체계, EO/IR 무기체계, 전자전 무기체계로 구분된다.²⁾ 특히, 레이더(RADAR : Radio Detecting And Ranging), 합성개구레이더(SAR : Synthetic Aperture Radar), 전자광학/적외선(EO/IR : Electro-Optics/InfraRed) 등의 센서는 독립적으로 운용되거나 차량, 항공기, 함정, 위성 등에 탑재되어 운용된다. 한편 4차 산업혁명의 핵심 신기술인 인공지능, 사물인터넷, 클라우드, 빅데이터, 모바일 등의 발전으로 물체를 탐지하고 식별하는 기술이 발전하고 있으나 대량의 데이터 확보가 어려운 군사용 물체 탐지 및 식별에 어려움이 있다(e.g., Hwang, Choi, Lee, & Bae, 2021).

이에 본 논문에서 제안하는 인공지능 모델은 대량의 데이터 확보가 어려운 군사용 물체의 한계를 극복하기 위해 부족한 데이터로도 높은 성능 달성이 가능한 방법인 전이학습과 미세조정 기법을 적용하였다. 구체적으로 최적의 인공지능 모델은 Kaggle에 공개된 데이터를 사용하여 35개 딥러닝 모델 성능을 비교 평가를 통해 선정한다. 또한, 적국 무기체계의 수집 한계로 아군의 5종의 지상 무기체계 데이터를 인터넷에서 수집하여 학습에 사용하고, Grad-CAM 기법과 훈련 영상 기반 데이터를 사용하여 모델의 의사결정 과정과 신뢰성 및 실전 적용 가능성을 평가하고자 한다.

본 논문은 2장에서 전이학습 기반의 무기체계 분류 및 탐지에 적용한 선행연구 고찰을 통해 실험에 적용할 전이학습과 딥러닝 모델, 설명가능한 인공지능기법에 관해 정리하였다. 3장은 실험에 사용된 데이터 세트와 전이학습을 적용한 딥러닝 모델의 성능 결과를 확인하고 그 결과를 통해 모델을 선정하였다. 그리고 선정된 모델을 이용하여 신뢰성 및 실전 적용 가능성을 검증하고, 끝으로 4장에서 결론 및 발전사항 등을 제시한다.

II. 본론

2.1 선행 연구 고찰

국방영역에서 감시정찰체계의 인공지능 연구는 현재 활발하게 이루어지고 있으며 해상, 육상, 공

1) 국방기술품질원(2010). 2010 국방과학기술조사서 : 제3권 감시·정찰

2) 국방기술품질원(2019). 2019 국방과학기술조사서 : 국방과학기술 개발동향 제3권 감시·정찰 무기체계

중, 우주와 같은 작전 영역과 E/O, IR, SAR, LiDAR와 같은 센서 유형으로 구분할 수 있다.

지상 작전 영역에서 E/O 센서 기반의 연구를 살펴보면, Kim et al.(2021)은 무인수색 차량을 위해 주요 표적 탐지 및 식별, 적과 아군 무기체계 피아식별이 가능한 인공지능 모델을 제안하였다. Lim and Kang(2021)은 YOLO 알고리즘을 기반으로 한 전차 국적 식별 방법을 제안하여 인공지능 딥러닝 기술을 적용한 무기체계의 성능을 개선하고, 데이터 규모에 따른 평가 척도인 mAP와 IoU를 분석하였다. Hwang et al.(2021)은 기동 무기체계 분류를 위해 전차와 장갑차 이미지를 수집한 후 전이학습을 통해 도출된 각 모델 예측값의 평균을 구하는 앙상블을 수행하였으며, 소규모 데이터를 이용하여 2개의 이중 모델인 InceptionV3와 DenseNet121 앙상블 조합만으로도 98.5%의 분류 정확도를 도출하였다. Kang(2022)은 전장에서 발생하는 다양한 우발상황에 대처하기 메타러닝 기법의 하나인 퓨샷 러닝을 활용하여 미식별된 전차를 식별하는 방법을 제안하였다. Hwang and Ma(2022)은 데이터 증강 모델을 활용하여 군사 데이터의 특성상 적은 수의 적성 전차 데이터를 기반으로 객체 탐지 성능을 향상하는 방안을 제안하였으며, 다양한 데이터 증강 모델의 조합 방식에 따라 평가 지표인 AP(Average Precision)이 최소 0.0141에서 최대 0.0754의 성능 향상을 확인하였다. Lee and Ma(2023)은 포병 무기체계 분류와 식별에 관한 첫 시도로 견인포, 자주포, 다연장 이미지 데이터를 수집하여 6가지 모델을 이용하여 실험을 진행하였다. 특히, Hwang and Ma(2023)은 데이터 부족 문제를 해결하기 위해 군 SinGAN 모델을 활용해 획득하기 어려운 군사 장애물 데이터 증강이 가능하였고, 증강된 데이터가 분류 모델의 성능을 높일 수 있음을 확인하였다.

지상 작전 영역에서 SAR 센서 기반의 연구를 살펴보면, Ryu and Ma(2022)은 기존 SAR ATR 연구에서 고려되지 않았던 초해상화를 적용한 이미지 분류 모델을 통한 실험으로 성능 향상을 확인하였다. 지상 작전 영역에서 IR 센서 기반의 연구를 살펴보면, Yang(2020)은 열 영상 기반 표적의 분할 신뢰도에 따라 형태 특징과 지역 특징의 분류 결과를 융합하는 계층적 분류기법을 제안하였으며, 연속 영상 기반으로 분류 결과를 갱신하는 기법을 새롭게 제안하여 차량 표적 분류 정확도를 개선하였다.

선행 연구 고찰 결과, 대부분의 연구는 지상 작전 영역에서 E/O 센서 기반의 인공지능 적용 부분에 초점을 두고 수행되었다. 상기한 선행 연구는 국방 데이터 활용의 제약으로 웹상에서 자료 수집이나 수집된 데이터의 증강, 소수의 데이터에 기반한 분류 향상을 위한 미터 러닝 기법 등을 제안하고 있다. 하지만 현재 국방 분야의 인공지능 기술 활용에서 국방 인공지능 기술에 대한 신뢰성 부족과 국방 데이터 활용을 제약하는 보안 규제 등(e.g., Lee & Jang, 2022)³⁾의 연구적 제한사항이 여전히 존재하고 있다. 그래서 본 연구는 인공지능 기술에 대한 신뢰성 확보 측면에서 인공지능 모델의 신뢰성 향상에 초점을 두고자 한다.

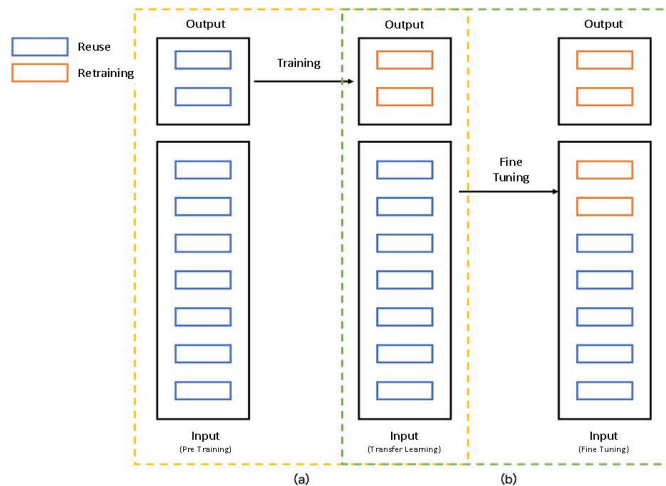
3) 윤정현(2020) 국방 분야의 인공지능 활용성 제고 방안과 시사점. Science & Technology Policy I 과학기술정책 포커스 : 인공지능 기술의 활용 전망과 스마트 강국을 위한 제언.

2.2 관련 기술 고찰

2.2.1 전이학습(Transfer Learning)

전이학습은 한 도메인에서 학습한 모델을 다른 도메인으로 전이하여 사용하는 기법(Pan & Yang, 2019)이며, 데이터의 수가 적거나 모델을 처음부터 학습하는 데에 필요한 시간과 자원이 부족한 경우에 유용하다. 전이학습은 기존에 학습된 모델의 일부 또는 전체를 가져와서 새로운 작업에 맞게 재사용하는 방식으로 동작한다. 기존 모델은 일반적으로 대규모 데이터 세트에서 학습된 고성능 모델이며, 이미지 분류, 객체 감지, 자연어 처리 등 다양한 작업을 수행할 수 있다.

Figure 1은 전이학습(Transfer Learning)(a)과 미세조정(Fine-Tuning)(b)의 과정을 보여준다. 전이학습(a)은 두 가지 주요 단계로 구성된다. 첫 번째로 사전학습 단계는 대규모 데이터 세트에 모델을 사전에 학습시키며 일반적인 특징을 추출하는 데에 중점을 둔다. 예를 들어, 이미지 분류를 위한 사전학습 모델은 주로 ImageNet 데이터 세트를 사용하여 객체의 일반적인 특징을 학습한다. 둘째, 전이학습 단계는 분류하고자 하는 데이터 세트에 맞게 모델의 출력을 수정하며 입력된 이미지의 특징을 추출하는 영역은 동결한 상태로 학습되지 않도록 사전학습 모델을 재사용한다. 즉 모델의 출력 영역만 새롭게 학습된다. 미세 조정(b)의 경우 이미지의 특징을 추출하는 영역의 일부 또는 전체를 동결 해제하여 기존에 사전 학습된 이미지 특징을 추출하는 영역을 추가로 학습한다. 이를 통해 적은 양의 데이터로도 좋은 성능을 얻을 수 있다.



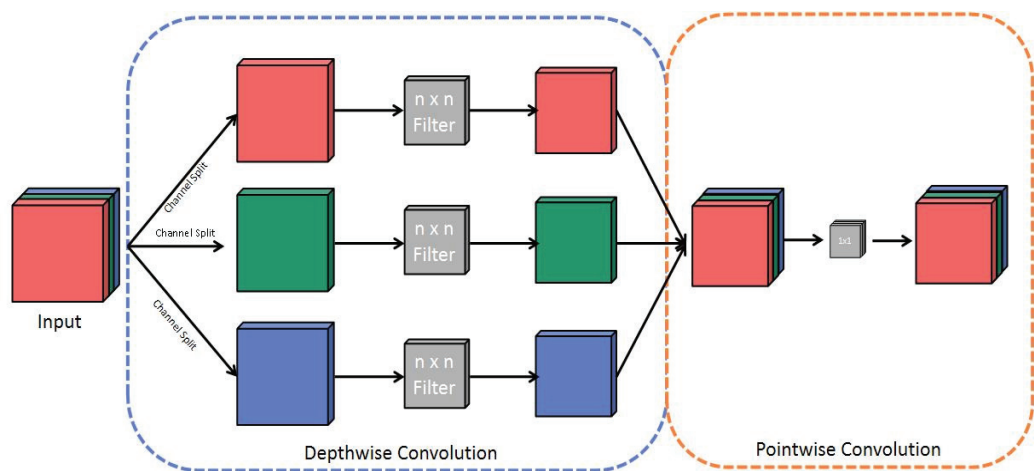
<Figure 1> Transfer Learning (a) and Fine-Tuning (b)

2.2.2 MobileNet

MobileNet은 컴퓨터 비전 분야에서 사용되는 CNN(Convolutional Neural Network) 모델 중 하

나로 깊이 분리형 합성곱(Depthwise Separable Convolution) 기법을 사용하여 컨볼루션 연산을 효율적으로 처리하고, 특징 추출과 분류를 동시에 수행하는 구조로 구성되어 있다(Howard et al., 2017). 기존의 합성곱 연산은 입력 이미지의 모든 채널에 대해 동시에 필터를 적용하는 방식이지만 깊이 분리형 합성곱은 두 단계로 나누어 연산한다.

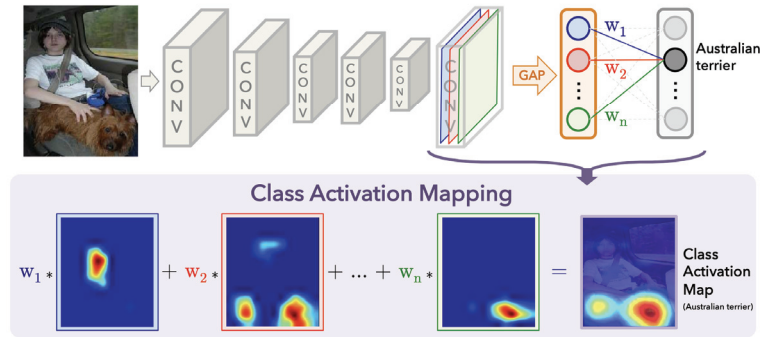
Figure 2는 MobileNet의 연산 과정을 나타낸다. 먼저 입력 데이터의 각 Red, Green, Blue 채널에 대해 컨볼루션 연산을 적용하는 깊이별 합성곱(depthwise convolution)을 수행한다. 이를 통해 필터의 개수가 입력 채널의 개수와 같아지므로 전체 파라미터 수가 크게 감소한다. 이후 채널 간의 선형 결합을 위한 1×1 합성곱(pointwise convolution)을 수행하여 채널마다 독립적으로 특징을 추출한다. 이를 통해 채널 간의 상호작용이 가능해지고, 공간적인 정보와 채널 간의 관계를 함께 고려할 수 있다. 이를 통해 모델의 크기와 연산량을 크게 줄인 경량화된 모델로 기존의 딥러닝 모델과 비교하면 매우 적은 파라미터 수를 가지면서도 높은 분류 성능을 제공하며 이러한 경량화된 특성은 모바일 기기나 임베디드 시스템과 같은 자원이 제한된 환경에서 실시간 응용에 적합하도록 설계되었다.



<Figure 2> Depth-separated convolution and pointwise convolution process

2.2.3 CAM(Class Activation Map) & Grad-CAM(Gradient-weighted Class Activation Map)
 이미지 분류에서 사용되는 CNN(Convolutional Neural Network)은 일반적으로 입력층, 특징 추출 층(Convolutional Layer), 특징 분류 층(Fully Connected Layer)으로 구성되어 있다(e.g., Park, Lee, & Sunwoo, 2021). 입력으로 들어온 이미지에서 특징 추출 층에서 합성곱 연산을 통해 추출된 특징은 특징 분류 층에 의해 구분되는데 이를 위해 평탄화(Flattening)를 사용한다. 이 과정에서 각 픽셀의 위치 정보를 잃어버리게 되며 모델이 이미지의 어떤 영역을 보고 분류하는지를 알 수 없다는 문제가 있다. 그래서 Zhou, Khosla, Lapedriza, Oliva, & Torralba(2016)이 제안한 CAM 기법은 평탄화 대신 GAP(Global Average Pooling)을 사용하여 특징 추출 층에서 추출된 특징 맵(Feature

Map)의 평균값을 계산하여 특정 범주의 출력값과 결합하여 범주별 판별 영역을 강조한다. 특정 범주의 출력값은 확률로 해석되기 전의 값으로 해당 출력값은 소프트맥스 함수에 의해 확률로 변환된다. Figure 3은 GAP가 적용된 CNN 모델을 기반으로 한 CAM 기법 과정을 나타내며 입력 이미지 내의 Australian terrier로 분류된 판별 영역을 나타낸다.

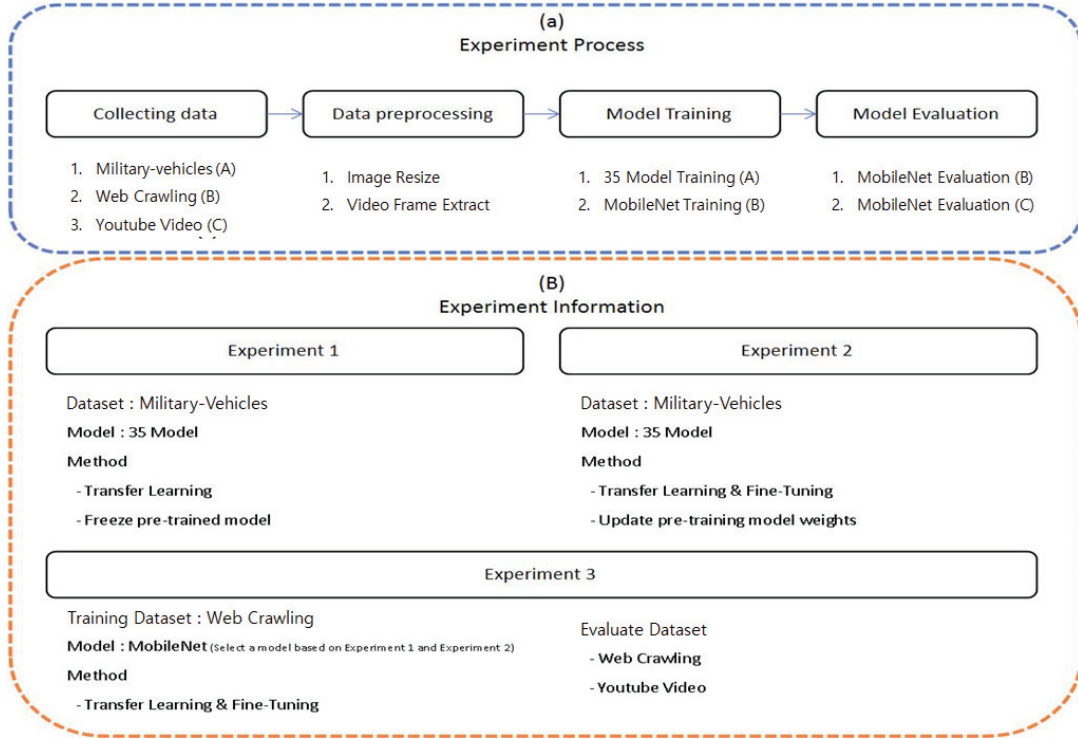


<Figure 3> Class Activation Mapping Process (from Zhou, et al., 2016)

CAM 기법은 GAP을 사용하는 CNN 모델에만 적용이 가능하다는 제한사항이 존재한다. Selvaraju et al.(2017)이 제안한 Grad-CAM 기법은 CAM의 제한 사항을 해결하고 다양한 CNN 구조와 활성화 함수에 적용할 수 있는 확장된 기법이다. Grad-CAM은 특정 범주의 출력값에 대한 중간 계층의 특징 맵의 가중치(Gradient)를 사용하여 CAM을 생성한다. 이를 통해 모델의 분류 기준을 이해하고 모델이 주어진 입력에서 어떤 영역을 중요하게 여기는지 확인할 수 있다.

III. 실험 방법 및 결과

Figure 4의 (a)는 본 논문의 실험 과정을 나타내며 (b)는 실험 정보를 확인할 수 있다. 실험 과정은 데이터 수집, 데이터 전처리, 모델 학습, 모델 평가순으로 진행하였다. 데이터 수집은 Kaggle의 Military-Vehicles(A)와 웹 크롤링(B), 유튜브(C)를 통해 수집되었다. 수집된 데이터는 224x224 크기로 이미지 크기를 조정하였다. 이후 35개의 CNN 모델에 해당 데이터 세트를 사용하여 실험 1과 같이 전이학습을 실험 2와 같이 미세조정을 적용하여 최종 모델로 MobileNet을 선정하였다. 최종 선정된 MobileNet 모델을 웹 크롤링으로 수집된 대한민국 육군의 5종 무기체계 데이터를 이용하여 실험 3과 같이 전이학습 및 미세조정을 적용하여 학습 및 평가하였다. 또한 유튜브를 통해 수집된 영상 데이터를 이용하여 MobileNet 모델을 평가하였으며 이를 통해 실제 전력화 가능성을 확인하였다.



<Figure 4> Experiment process and information

3.1 표본 선정과 자료수집

첫 번째로 CNN 모델의 성능을 비교 평가 및 재현성 확보를 위한 벤치마크 데이터 세트로 Kaggle⁴⁾에 공개된 Military-Vehicles 데이터⁵⁾를 사용하였다. Table 1은 해당 데이터 세트의 범주 명과 범주별 이미지 데이터 개수를 나타낸다.

4) Kaggle은 데이터 분석 및 머신러닝에 대한 학습과 대회가 열리는 플랫폼으로 개인, 기관 또는 특정 사용자가 데이터를 첨부해서 문제를 출제하면 Kaggle 사용자 누구나 문제에 대한 답을 제출할 수 있다.

5) <https://www.kaggle.com/datasets/amanrajbose/Military-Vehicles>

<Table 1> Number of data per category in the Military-Vehicles dataset

| Class Name | Train Image | Test Image | Validation Image |
|---------------------------------|-------------|------------|------------------|
| Anti-aircraft | 942 | 337 | 404 |
| Armored Combat Support Vehicles | 937 | 335 | 401 |
| Armored Personnel Carriers | 944 | 337 | 405 |
| Infantry Fighting Vehicles | 947 | 338 | 406 |
| Light Armored Vehicles | 951 | 340 | 407 |
| Mine-Protected Vehicles | 1179 | 421 | 506 |
| Prime Movers and Trucks | 1170 | 418 | 502 |
| Self-Propelled Artillery | 1188 | 424 | 509 |
| Tanks | 967 | 345 | 414 |

두 번째로 Military-Vehicles 데이터 세트의 범주는 대한민국 육군의 무기체계와는 일치하지 않는다는 문제점이 있다.⁶⁾ 이러한 문제를 해결하기 위해 대한민국 육군의 무기체계에서 5종을 선정하여 인터넷에서 해당 무기체계의 이미지 데이터를 수집하였다. Figure 5는 인터넷에서 수집한 데이터로 전차 범주의 경우 K-2로 검색한 결과를 수집하였으며, 자주포 범주의 경우 K-9, 다연장로켓 범주의 경우 K-239, 자주대공포 범주의 경우 K-30, 보병전투차량 범주의 경우 K-21로 검색된 결과를 각각 300장씩 수집하였다.

6) 국방기술진흥연구소 보고서의 보병전투장갑차 개발동향 및 발전방향에 따르면 장갑차는 주된 운용개념에 따라 크게 보병전투장갑차와 병력수송장갑차 두 가지로 구분하고 있다. (https://dtims.krit.re.kr/vps/vpsFileView.do?attclDcn=BKC000220211028173624611&attcOrd=2&tabsGubn=VPS_DSP_STMG). 하지만 Kaggle에서 공개된 데이터를 보면 ACSV, LAV, MPV등 상세하게 분류하고 있음을 확인할 수 있어 이에 따라 기존 라벨링을 사용하는 경우에 하나의 무기체계가 여러 라벨에 속하는 문제가 나타날 수 있어 본 연구에서는 기존 Kaggle 데이터 셋의 라벨을 따라가지 않고 연구자가 직접 라벨을 구성하였다.



<Figure 5> Image datasets for ROK(Republic of Korea)'s representative ground weapon systems

세 번째는 전력화 가능성을 평가하기 위해 유튜브 국방 NEWS, 국방 TV 채널에 공개된 훈련 영상에서 이미지를 추출하여 데이터 세트를 구축하였다. 실제 군 훈련 영상을 사용하는 데는 많은 보안 및 제도상의 문제로 인해 유튜브에 공개된 훈련 영상을 사용하였으며 파이썬 기반의 오픈소스 소프트웨어인 yt-dlp을 사용하여 해당 영상을 다운로드하였다.⁷⁾ 다운로드된 영상은 1280x720 크기를 가지며, 파이썬 기반의 OpenCV 라이브러리를 통해 영상에서 프레임별 이미지를 추출하였다. 추출된 이미지를 직접 검수하여 분류하고자 하는 객체가 없는 경우 해당 이미지는 제외하였으며, 영상별 60장의 이미지를 무작위로 선정하여 총 180장의 이미지를 평가하였다. Figure 6는 평가에 사용된 훈련 영상의 일부이다.

7) [최초공개] K-2전차, 최초 정규훈련 현장(동영상), https://www.youtube.com/watch?v=2T_tH_8vERk
 육군, '천무' 유도탄 360여 발 실 사격 현장K239, <https://www.youtube.com/watch?v=YN9MLcTqR9g>
 육군 8군단 K-9자주포 실사격 훈련 드론영상, <https://www.youtube.com/watch?v=5JYZ0FyS39Y>



<Figure 6> Datasets based on real-world training video

3.2 실험 환경 및 조건 설정

Tensorflow에서 제공하는 ImagetNet 데이터로 사전학습이 완료된 모델 35개를 기반으로 Military-Vehicles 데이터를 사용하여 전이학습을 적용하였으며 35개의 모델 모두 동일한 조건에서 실험을 진행하여 성능이 우수한 모델 1개를 선정하였다. 선정된 모델을 다시 인터넷에서 수집한 5종 무기체계 이미지 데이터로 학습하여 유튜브에 공개된 훈련 영상 데이터 세트로 평가를 진행하였다.

Table 2의 경우 실험에 사용된 환경 및 조건으로 입력 이미지의 넓이와 높이는 224픽셀로 지정하였으며 컬러 이미지를 입력으로 받아 3개의 채널(Red, Green, Blue)이 존재한다. 연산 한 번에 들어가는 데이터의 크기인 배치 사이즈(Batch Size)는 32, 손실 함수의 최솟값을 향해 이동하면서 각 반복에서 단계 크기를 결정하는 매개변수인 학습률(Learning Rate)은 $1e-5$, 최적화 알고리즘은 Adam, 학습 횟수(Epoch)는 최대 1,000회를 지정하였으며 검증 데이터 성능이 가장 좋은 모델을 사용하였다. 또한 평가 데이터 세트가 없는 경우 데이터를 분할하여 80%는 학습용으로 20%는 검증용으로 사용하였으며 학습 과정에서 검증용 데이터의 정확도가 가장 좋은 모델을 저장하여 평가에 사용하였다. 학습 과정에서 과대 적합을 예방하기 위해 수집된 데이터의 40도 범위 회전 및 기울이기, 20% 내 수평/수직 평행 이동, 무작위 확대 및 축소를 통해 데이터를 증식하여 학습에 사용하였다.

<Table 2> Hyperparameters used in training

| Hyperparameters | Value |
|-------------------------------------|-------------------|
| Input Size (width, height, channel) | 224x224x3 (pixel) |
| Batch Size | 32 |
| Learning Rate | $1e-5$ |
| Epoch | 1000 (Step) |
| Optimizer | Adam |

IV. 실험 결과 및 해석

4.1 실험 결과(1) : Military-Vehicles 데이터 세트

이 데이터 세트를 사용하여 2가지의 실험을 진행하였으며 실험 결과는 Table 3과 같다. 실험 1(Experiment 1)은 사전 학습만 적용하여 실험하였으며 해당 실험은 ImageNet 데이터 세트로 사전 학습된 CNN 모델을 기반으로 해당 데이터 세트에 맞게 모델의 출력을 수정하였다. 또한 학습 과정에서 입력 이미지의 특징을 추출하는 특징 추출 층은 학습되지 않도록 동결하여 특징분류 층만 학습하였다. 실험2(Experiment 2)은 사전학습과 미세조정을 모두 적용하여 실험하였으며 ImageNet 데이터 세트로 사전 학습된 CNN 모델을 기반으로 특징 추출 층과 특징 분류 층을 모두 학습하였다. 이는 일반화된 특징 추출 및 Military-Vehicles 데이터 세트 분류를 하는 첫 번째 실험과는 다르게 상기한 데이터 세트에 맞는 특징을 추출하고 분류하는 기준을 학습하게 된다.

실험1의 경우 ResNet50V2, ResNet101V2, MobileNet, DenseNet121, DenseNet169, DenseNet201 모델의 정확도가 90%~92%임을 확인하였으며, 실험2는 Xception, ResNet50, DenseNet121, DenseNet169, DenseNet201 모델의 정확도가 90%~98%로 나타났다(Table 3). 실험1과 실험2에서 성능이 급격하게 향상된 모델은 ResNet50, ResNet101, ResNet152, EfficientNetV2B0으로 정확도가 55%p에서 71%p까지 증가하였다. 이는 ImageNet 데이터에서 학습된 범주별 일반화된 특징과 Military-Vehicles 데이터의 특징이 차이가 있음을 확인할 수 있으며 특징 추출 층을 학습하는 것이 더 좋음을 알 수 있다.

ResNet, DenseNet과 같은 모델은 인공지능 모델에서 경사 소실(Vanishing Gradient) 문제를 해결하기 위해 스킵 연결(Skip Connection)이 적용되어 입력값이 네트워크를 직접 통과하여 더 깊은 레이어로 전달되어도 원래의 정보를 보존할 수 있지만 VGG16과 VGG19는 해당 방법이 적용되어 있지 않아 미세 조정을 적용한 결과 과대 적합으로 인해 정확도가 각각 75%p, 70%p 감소한 것을 확인하였다. 실험1과 실험2 모두 80% 이상의 정확도를 보이며 모델 성능이 개선되거나 성능 감소 폭이 낮은 Xception, ResNet50V2, MobileNet, DenseNet121, DenseNet169, DenseNet201 모델을 선정하였다.

<Table 3> Experiment results on military-Vehicles dataset

| | Experiment 1 | Experiment 2 | Change |
|-------------------|--------------|--------------|--------|
| Xception | 83 | 95 | +12%p |
| VGG16 | 86 | 11 | -75%p |
| VGG19 | 81 | 11 | -70%p |
| ResNet50 | 28 | 90 | +62%p |
| ResNet101 | 26 | 87 | +61%p |
| ResNet152 | 29 | 84 | +55%p |
| ResNet50V2 | 90 | 89 | -1%p |
| ResNet101V2 | 90 | 84 | -6%p |
| ResNet152V2 | 87 | 83 | +4%p |
| InceptionV3 | 73 | 84 | +11%p |
| InceptionResNetV2 | 77 | 86 | +9%p |
| MobileNet | 93 | 93 | - |
| MobileNetV2 | 90 | 85 | -5%p |
| MobileNetV3 Small | 11 | 9 | -2%p |
| MobileNetV3 Large | 26 | 12 | -14%p |
| DenseNet121 | 91 | 98 | +7%p |
| DenseNet169 | 91 | 96 | +5%p |
| DenseNet201 | 92 | 96 | +4%p |
| NasNetMobile | 79 | 70 | -9%p |
| NasNetLarge | 85 | 75 | -10%p |
| EfficientNetB0 | 11 | 11 | - |
| EfficientNetB1 | 11 | 9 | -2%p |
| EfficientNetB2 | 16 | 9 | +7%p |
| EfficientNetB3 | 11 | 8 | -3%p |
| EfficientNetB4 | 12 | 9 | -3%p |
| EfficientNetB5 | 15 | 9 | -6%p |
| EfficientNetB6 | 11 | 9 | -3%p |
| EfficientNetB7 | 11 | 9 | -3%p |
| EfficientNetV2B0 | 11 | 82 | +71%p |
| EfficientNetV2B1 | 11 | 12 | +1%p |
| EfficientNetV2B2 | 16 | 11 | -5%p |
| EfficientNetV2B3 | 16 | 65 | +49%p |
| EfficientNetV2S | 42 | 72 | +30%p |
| EfficientNetV2M | 12 | 28 | +16%p |
| EfficientNetV2L | 19 | 27 | +8%p |

Table 4는 해당 모델 파일 크기(Size), 모델의 파라미터 수(Parameter), 추론시간(CPU, GPU Time)을 비교한 것으로 CPU는 AMD EPYC Processor (with IBPB) (92 core), RAM은 1.7T GPU는 Tesla A100 환경을 사용했으며 추론시간의 경우 추론 단계당 시간은 30개의 배치와 10회의 반복을 평균한 시간 값이다.⁸⁾

<Table 4> Model-specific information

| | Size (MB) | Parameters (M) | Time (ms, cpu) | Time (ms, gpu) |
|-------------|-----------|----------------|----------------|----------------|
| Xception | 88 | 22.9 | 109.4 | 8.1 |
| ResNet50V2 | 98 | 25.6 | 45.6 | 4.4 |
| MobileNet | 16 | 4.3 | 22.6 | 3.4 |
| DenseNet121 | 33 | 8.1 | 77.1 | 5.4 |
| DenseNet169 | 57 | 14.3 | 96.4 | 6.3 |
| DenseNet201 | 80 | 20.2 | 127.2 | 6.7 |

감시정찰체계에 탑재될 모델을 선정함에 있어 연산량의 제한으로 인해 모델의 크기가 크고 파라미터의 개수가 많은 Xception, ResNet50V2, DenseNet169, DenseNet201 모델은 감시정찰 체계 적용이 어렵다고 판단된다. 또한, 실시간 분류 작업에서는 모델의 추론 속도도 중요한 요소이며 GPU 환경에서는 추론 시간이 비슷하였으나 추론을 위한 하드웨어를 감시정찰체계에 추가해야 한다는 어려움이 존재한다. 이에 따라 Xception, ResNet50V2, DenseNet121 모델은 높은 성능에도 불구하고 추론 속도가 상대적으로 MobileNet 모델에 비해 느리기 때문에 적합하지 않다고 판단하였다. 이러한 고려 사항을 종합하여 MobileNet 모델이 감시정찰체계에 탑재되어 지상 무기체계 분류를 위한 인공지능 모델로 가장 적합하다고 판단된다.

4.2 실험 결과(2) : 5종 무기체계 데이터 세트

Military-Vehicles 데이터 세트와 동일한 환경과 설정값을 기반으로 MobileNet 모델을 5종 무기체계 데이터 세트를 사용하여 학습과 평가를 진행하였다. 제안한 모델의 성능을 평가한 결과(Table 5), 평균 F1-Score는 92%이며 범주에 따라 86%에서 96%의 성능을 확인하였다.

8) <https://keras.io/api/applications/>

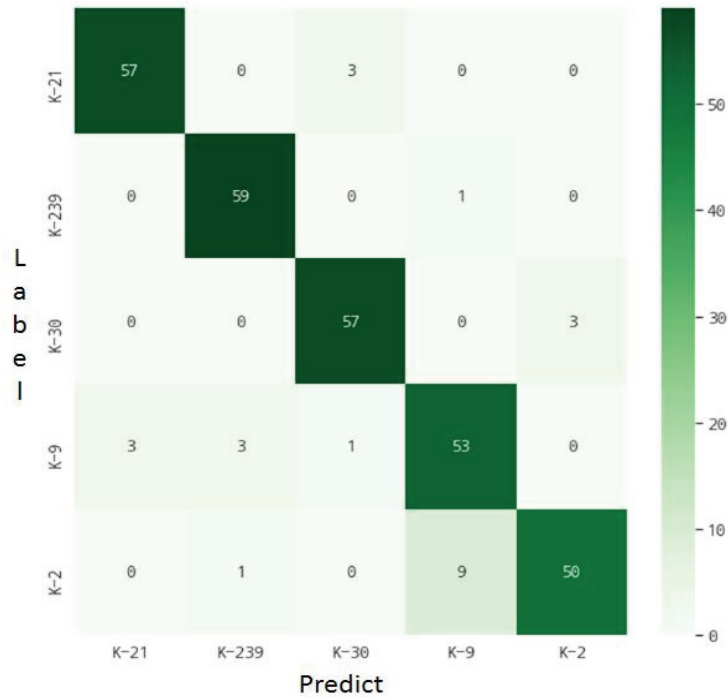
<Table 5> Experimental results of the MobileNet model on a dataset of 5 weapon systems

| | Precision | Recall | F1-score | Count |
|---------|-----------|--------|----------|-------|
| K-21 | 95 | 95 | 95 | 60 |
| K-239 | 94 | 98 | 96 | 60 |
| K-30 | 93 | 95 | 94 | 60 |
| K-9 | 84 | 88 | 86 | 60 |
| K-2 | 94 | 83 | 88 | 60 |
| Average | 92 | 92 | 92 | - |

정밀도(Precision)는 모델이 특정 범주로 예측한 데이터 중에서 실제로 해당 범주인 데이터의 비율로 모델이 얼마나 정확하게 예측하는지를 나타내는 지표이다. 재현율(Recall)은 실제로 특정 범주인 데이터 중에서 모델이 얼마나 많은 데이터를 해당 범주로 예측하였는지에 대한 비율로 모델이 실제 데이터를 놓치지 않고 잘 찾아내는지를 보여주는 지표이다. 두 지표의 조화평균인 F1-Score는 모델이 얼마나 정확하게 예측하면서도 실제 해당 범주를 놓치지 않는지를 함께 고려하는 지표로 수식 1과 같이 계산된다. 정밀도, 재현율, F1 Score는 분류 문제에서 범주별 데이터의 개수가 불균형한 경우 정확도를 대체하여 사용된다.

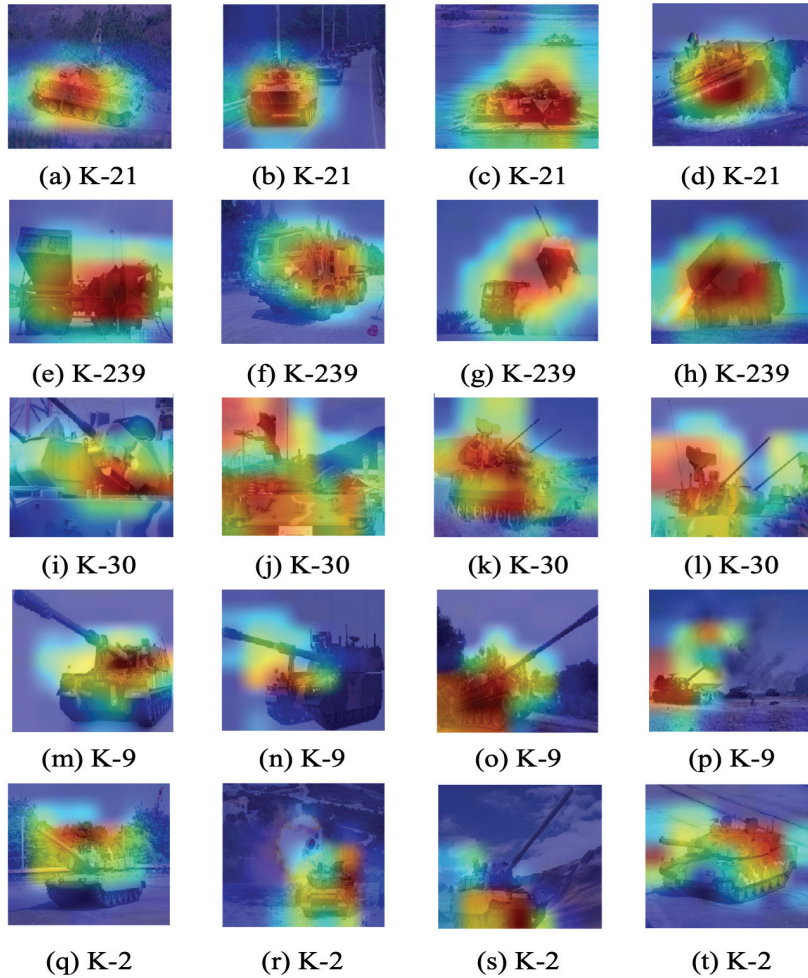
$$F1\text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (1)$$

Figure 7은 MobileNet 모델의 범주별 예측 결과와 실제 결과가 일치하는지를 나타내는 혼동 행렬 시각화 결과로 보병전투차량, 다연장로켓, 자주대공포의 경우 오분류가 3건 이하로 정밀도와 재현율 모두 높은 성능을 확인할 수 있었으나 자주포와 전차의 경우 오분류가 각각 7건, 10건으로 재현율이 다른 범주에 비해 낮게 나타났다. 자주포의 경우 6건을 전차로 1건을 보병전투차량으로 오분류했으며, 전차의 경우 9건을 자주포로 1건을 다연장로켓으로 오분류함을 확인하였다. 이는 전차와 자주포의 외형상 유사성이 높아 발생한 문제로 예상된다.



<Figure 7> Number of actual and predicted categories

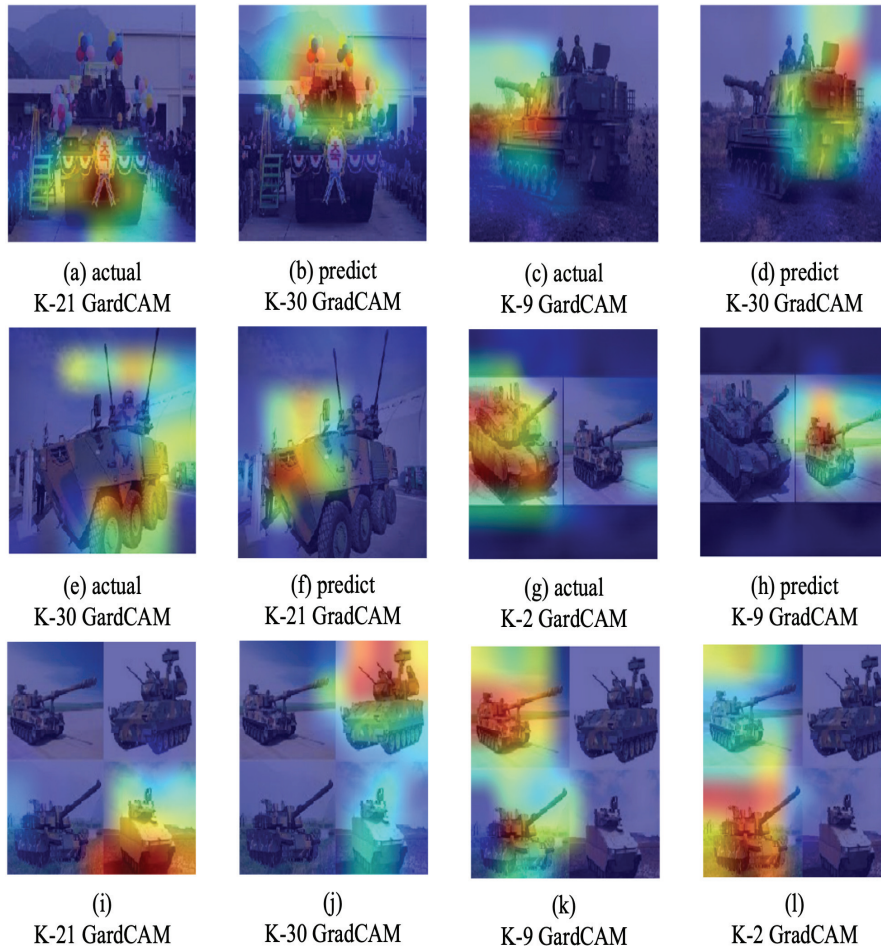
Figure 8은 5종 무기체계 평가 데이터에서 올바르게 분류된 데이터에 Grad-CAM을 적용한 결과로 학습된 모델이 해당 범주를 분류하는 데 있어 입력 이미지의 어떤 부분에 집중했는지를 시각적으로 확인할 수 있다. 이를 통해 모델의 의사결정 과정을 확인하며 사람의 판단 기준에서 벗어나는지를 확인한다. (a)~(d)는 보병전투차량의 결과로 기동 환경, 도하 환경에서도 이미지 내의 보병전투차량의 영역에 집중되어 올바르게 분류하고 있는 것을 확인할 수 있다. (e)~(h)는 다연장로켓의 결과로 포대 전개 전, 전개 후, 발사 등 다양한 상황에서도 이미지 내의 보병전투차량의 영역에 집중되어 올바르게 분류되었다. (i)~(l)은 자주대공포의 결과로 주로 레이더 안테나, 기관포의 영역을 집중하고, (m)~(p)는 자주포의 결과로 주로 포신, 포신과 차체의 연결부에 집중하는 것으로 나타났다. (q)~(t)는 전차의 분류 결과로 포탑과 차체에 집중하는 것을 확인할 수 있다.



<Figure 8> Grad-CAM application results (classification)

Figure 9는 5종 무기체계 평가 데이터에서 잘못 분류된 데이터에 Grad-CAM을 적용한 결과이다. (a)는 보병전투차량의 Grad-CAM 결과이며 (b)는 자주대공포의 Grad-CAM 결과이다. 포탑 및 풍선으로 인해 보병전투차량을 자주대공포로 잘못 예측한 것을 확인할 수 있다. (c)는 자주포의 Grad-CAM 결과이며 (d)는 자주대공포의 Grad-CAM 결과이다. 앞서 (b)와 동일하게 포탑 해치로 인해 자주대공포로 잘못 예측한 것을 확인할 수 있다. (e)는 자주대공포의 Grad-CAM 결과이며 (f)는 보병전투차량의 Grad-CAM 결과이다. 자주대공포의 차체에 집중되어 오분류가 일어난 것을 확인할 수 있다. (g)와 (h)는 한 이미지에 자주포와 전차가 같이 있는 형태로 (g)는 전차의 Grad-CAM 결과이며 (h) 자주포의 결과이다. 해당 데이터는 전차로 라벨링이 되어 있으며 전차로 올바르게 예측된 결과이다. 입력데이터에 서로 다른 범주가 같이 존재해도 모델이 잘 구분하고 있음을

확인할 수 있다. (i)~(l)은 한 이미지에 시계방향으로 자주포, 자주대공포, 전차, 보병전투장갑차로 구성된 이미지이다. 보병전투장갑차는 보병전투차량의 차체를 기반으로 만들어졌는데 모델 또한 차체에 집중하여 보병전투차량으로 분류된 것을 (i)를 통해 알 수 있다. (j)는 자주대공포, (k)는 자주포, (l)은 전차의 Grad-CAM 결과이다.



<Figure 9> Grad-CAM application results (Misclassification)

학습 모델이 5종의 무기체계를 대부분 올바르게 분류하였으나(Figure 8 to 9), 하나의 이미지에 여러 개의 객체가 있는 경우에 특정 영역이 이미지에서 많은 영역을 차지하여 분류 기준에 혼동을 주는 문제점이 발견되었다. 이러한 문제점을 개선하기 위해서 추후 다양한 환경에서의 고품질의 데이터를 추가로 확보가 필요하며, 모델의 입력 크기를 맞추기 위한 이미지 크기 조절 과정에서의 방법 개선, 모델의 입력 이미지 크기 증가와 같은 방법을 적용한다면 모델의 오분류 사례를 줄일 수 있을 것이다.

4.3 실험 결과(3) : 훈련 영상 데이터 세트

실제 환경 테스트를 위해 유튜브 국방TV에서 공개한 전차, 자주포, 다연장로켓의 훈련 영상에서 각각 60장의 이미지를 추출하였으며 추출된 이미지를 제안한 모델로 평가를 진행하였다. 평가 결과 전차의 경우 92%를, 자주포의 경우 95%를, 다연장 로켓의 경우 98%의 정확도를 확인하였다. 이를 통해 제안한 모델은 실제 훈련 및 작전 환경에서도 단순한 예측 범주가 아닌 Grad-CAM 기법을 적용하여 인공 지능 모델의 의사 결정 과정을 설명할 수 있어 신뢰성이 높은 모델임을 확인하였다.

V. 결론

본 연구는 무인수색차량, 견마형 로봇 등 지상형 감시정찰체계에 적용가능한 신뢰도 높은 인공 지능 모델을 제안하였다. 이를 위해 부족한 데이터로도 높은 성능을 달성할 수 있는 방법인 전이학습을 적용하여 5종의 무기체계 분류 모델을 개발하였다. 특히, 지상형 감시정찰체계에 적용 가능한 최적의 인공지능 모델을 선정하기 위해 Kaggle에 공개된 Military-Vehicles 데이터를 사용하여 35개의 모델을 비교 평가하였다. 본 연구는 실험 분석을 통해 Xception, ResNet50V2, MobileNet, DenseNet121, DenseNet169, DenseNet201 모델이 좋은 성능을 보임을 확인하였으며 연산량과 추론 속도를 고려하여 MobileNet이 지상 무기체계 분류를 위한 딥러닝 모델로 가장 적합하다고 판단하였다.

선정된 MobileNet 모델을 기반으로 5종 무기체계 데이터 세트를 실험한 결과, 평균 92%의 F1-Score 확인하였으며 자주포와 전차에서 오분류 가능성을 제시하였다. 또한 설명 가능한 인공지능 기법인 Grad-CAM을 적용하여 모델이 어떤 부분에 집중했는지를 시각적으로 확인할 수 있었으며 이를 통해 모델의 의사 결정 과정을 이해하고 딥러닝 모델의 신뢰성을 확인하였다. 특히, 본 연구의 제안 모델이 실제 환경에서 적용 가능성을 확인하기 위해 인터넷에 공개된 전차, 자주포, 다연장로켓의 훈련 영상에서 프레임을 추출하고, 모델평가를 진행하여 전차(92%), 자주포(95%), 다연장로켓(98%)의 정확도를 달성하였다. 하지만 분석 객체가 가려지거나 작게 나타나는 경우, 이미지 내 해당 객체 없음의 이유로 오분류가 발생함을 확인하였다. 따라서 본 연구에서 제안한 모델은 실제 훈련 및 작전 환경에서도 단순한 예측 범주가 아닌 Grad-CAM 기법을 적용하여 인공지능 모델의 의사 결정 과정을 설명할 수 있어 신뢰성이 높은 모델이라고 볼 수 있다. 이런 실무적 시사점에도 불구하고 후속 연구는 향후 다양한 무기체계의 범주와 정밀 분류를 위해 더 많은 데이터를 확보가 필요하며, 현실적인 데이터 제약의 문제를 극복하기 위해 메타러닝 기반의 인공지능 모델 적용이 중요할 것이다.

Acknowledgements

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Author contributions

Conceptualization and Literature review: BG, Resources and Data curation, Investigation and Methodology: BG and LJ, Writing (Original Draft): BG, Project administration and Supervision: LJ.

Reference

- Choi, J. H. (2021). *Trends and Prospects of Small Satellite Constellation for ISR Mission*. In Proceedings of the KSAS 2021 Fall conference (pp. 1552-1553). Retrieved from https://www.dbpia.co.kr/pdf/pdfView.do?nodeId=NODE11031991&googleIPSandBox=false&mark=0&ipRange=false&b2cLoginYN=false&isPDFSizeAllowed=true&nodeHistoryTotalCnt=2&accessgl=Y&language=ko_KR&hasTopBanner=true
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.
- Hwang, J. M. & Ma, J. M. (2023). An Empirical Study on the Problem of Lack of Military Obstacle Data Using SinGAN. *Journal of the Korea Academia-Industrial cooperation Society*, 24(1), 310-317. <https://doi.org/10.5762/KAIS.2023.24.1.310>
- Hwang, J. Y., Choi, B. A., Lee, J. H., & Bae, J. H. (2021). A Study on Transfer Learning-based Ensemble Model for Classification of Maneuver Weapon System. *Journal of Korean Institute of Information Technology*, 19(12), 1-10. <https://doi.org/10.14801/jkiit.2021.19.12.1>
- Hwang, K. S & Ma, J. M. (2022). Improvement of Object Detection Performance for Enemy Tanks Using the Combination of Data Augmentation Models. *Korean Journal of Computational Design and Engineering*, 27(2), 148-159. <https://doi.org/10.7315/CDE.2022.148>
- Joo, J. Y. (2020). A Study on Multi Domain Operations in U.S. Army. *The Quarterly Journal of Defense Policy Studies*, 36(1), 9-41. <https://doi.org/10.22883/jdps.2020.36.1.001>
- Kim, J. H., Kim, T. W., Kim, S. C., & Ryu, J. Y. (2021). Transfer Learning based Intelligent Target Detection and Identification Model Analysis for Unmanned Recon Vehicle. *Journal of the Military Operations Research Society of Korea*, 47(2), 26-35. <https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE11062013>
- Lee, J. Y. (2018). Manned-Unmanned Collaborative Combat System(MUM-CCS) Operation Concept and Development Direction. *Korea Research Institute for National Strategy KRINS QUARTERLY*, 3(3). 193 - 216. <https://doi.org/10.46322/krinsq.3.3.6>
- Lee, S. G., & Jang, S. G. (2022). Necessity of establishing an open source military R&D platform to promote AI development in defense. *Journal of Advances in Military Studies*, 5(3), 23-42. <https://doi.org/10.37944/jams.v5i3.177>
- Lee, Y. G., & Ma, J. M. (2023). A Comparative Study on the Performance of CNN Models for

- Classification of Artillery Weapon Systems. *Journal of Korea Academia-Industrial cooperation Society*, 24(1), 344-350. <https://doi.org/10.5762/KAIS.2023.24.1.344>
- Lim, S. G. & Kang, D. S. (2021) Identifications and Evaluation of Tank Nationality using YOLO Algorithm. *KIISE Transactions on Computing Practices*, 27(12), 555-562. <https://doi.org/10.5626/KTCP.2021.27.12.555>
- Nam, D. H., Im, T. H., Lee, D. J., & Cho, S. G. (2020). Mosaic Warfare of the Fourth Industrial Revolution. *Journal of National Defense Studies*, 63(3), 141-170. <https://doi.org/10.23011/jnds.2020.63.3.006>
- Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10), 1345-1359. <https://doi.org/10.1109/TKDE.2009.191>
- Park, J. H., Lee, B. D., & Sunwoo, M. H. (2021). MaskSLIC based CNN Classification Model for Mammogram Feature Extraction. *Journal of the Institute of Electronics and Information Engineers*, 58(10), 59-67. <https://doi.org/10.5573/ieie.2021.58.10.59>
- Ryu, J. M & Ma, J. M. (2022). Automatic Target Recognition of Military SAR Images Using Super-Resolution based Convolutional Neural Network. *Journal of Institute of Control, Robotics and Systems*, 28(1), 22-27. <https://doi.org/10.5302/J.ICROS.2022.21.0178>
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). *Grad-cam: Visual explanations from deep networks via gradient-based localization*. In Proceedings of the IEEE international conference on computer vision (pp. 618-626). Retrieved from https://openaccess.thecvf.com/content_iccv_2017/html/Selvaraju_Grad-CAM_Visual_Explanations_ICCV_2017_paper.html
- Yang, D. H. (2020). A Vehicle Classification Method in Thermal Video Sequences using both Shape and Local Features. *Journal of IKEEE*, 24(1), 97-105. <https://doi.org/10.7471/ikeee.2020.24.1.97>
- Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). *Learning deep features for discriminative localization*. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2921-2929). https://openaccess.thecvf.com/content_cvpr_2016/html/Zhou_Learning_Deep_Features_CVPR_2016_paper.html

원 고 접 수 일 2023년 07월 15일

원 고 수 정 일 2023년 11월 01일

계 재 확 정 일 2023년 12월 28일

지상 무기체계 분류 신뢰성 향상을 위한 설명 가능한 인공지능 모델

배기민* · 이장형**

국문초록

본 연구는 감시 및 정찰 애플리케이션에서 지상 무기 시스템의 분류 신뢰성을 향상시키기 위한 신뢰할 수 있는 인공지능(AI) 모델 개발에 중점을 두었다. 전차, 자주포, 다연장 로켓과 같은 공군 물체에 대한 제한된 데이터 가용성으로 인해 제안된 AI 모델은 전이학습 및 미세 조정 기술을 활용하여 이러한 문제를 극복한다. Kaggle에서 공개적으로 사용 가능한 Military-Vehicles 데이터 세트를 사용하여 35개의 딥 러닝 모델을 종합적으로 평가한 결과 MobileNet이 지상 무기 시스템 분류에 가장 적합한 모델임을 확인하였다. 선택한 MobileNet 모델은 5가지 유형의 지상 무기 시스템으로 구성된 데이터 세트에서 테스트했을 때 평균 F1 점수 92%를 달성하였다. 또한 설명 가능한 AI 기술인 Grad-CAM을 적용하여 제안 모델의 의사 결정 프로세스에 대한 통찰력을 제공하고 그 신뢰성을 검증하였다. 훈련 비디오에서 추출한 프레임을 사용한 실제 평가는 전차, 자주포 및 다연장 로켓에 대해 유망한 정확도를 보여주었다. 전반적으로 이 연구는 지상 감시 및 정찰 시스템의 성능을 향상시키기 위한 설명 가능하고 신뢰할 수 있는 AI 모델 개발에 기여하였다.

주제어 : 지상무기체계분류, 설명가능한 인공지능, 전이학습, 모바일넷, 기율기 가중 범주 활성화 영역

* (제1저자) 육군미래혁신연구센터, 감시정찰영상AI연구담당, rlals3501@gmail.com, <https://orcid.org/0000-0001-6587-4209>.

** (교신저자) 육군미래혁신연구센터, 첨단감시정찰기술팀장, leejangh@gmail.com, <https://orcid.org/0009-0003-2751-7560>.