Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.30, NO.6, Dec. 2020 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2020.30.6.967

딥러닝을 활용한 전략물자 판정 지원도구 개발에 대한 연구

조 재 영,[1†] 윤 지 원[2‡]
[1,2]고려대학교 (대학원생, 교수)

A Study on the Development of a Tool to Support Classification of Strategic Items Using Deep Learning

Jae-Young Cho,[1†] Ji-Won Yoon[2‡]
[1,2]Korea University (Graduate Student, Professor)

요 약

전략물자관리 제도의 이행 확산에 따라 전략물자 판정의 중요성이 높아지고 있으나 전략물자 제도를 처음 접하는 수출기업은 전략물자의 개념을 이해하기 쉽지 않고, 전략물자를 통제하는 기준이 다양하여 전략물자 판정에 어려움이 따른다. 본 논문에서는 전략물자 제도를 처음 접하는 기업이나 전략물자 판정시스템 이용자에게 진입장벽을 낮추어 판정이라는 과정을 쉽게 접근할 수 있는 방법을 제안한다. 이용자가 전략물자 판정이라는 절차를 매뉴얼이나 카탈로그의 제공만으로 판정결과를 확인할 수 있게 된다면, 전략물자 판정 방법과 절차에 보다 편리하고 쉽게 다가설 수 있을 것이다. 본 연구 목적을 달성하기 위해 이미지 인식 및 분류에서 연구되고 있는 딥러닝과 OCR(광학문자판독) 기술을 활용하고, 전략물자 판정 지원도구에 대한 개발과 연구를 통하여 우리 기업의 전략물자 판정에 도움이 되는 정보를 제공한다.

ABSTRACT

As the implementation of export controls is spreading, the importance of classifying strategic items is increasing, but Korean export companies that are new to export controls are not able to understand the concept of strategic items, and it is difficult to classifying strategic items due to various criteria for controlling strategic items. In this paper, we propose a method that can easily approach the process of classification by lowering the barrier to entry for users who are new to export controls or users who are using classification of strategic items. If the user can confirm the decision result by providing a manual or a catalog for the procedure of classifying strategic items, it will be more convenient and easy to approach the method and procedure for classfying strategic items. In order to achieve the purpose of this study, it utilizes deep learning, which are being studied in image recognition and classification, and OCR(optical character reader) technology. And through the research and development of the support tool, we provide information that is helpful for the classification of strategic items to our companies.

Keywords: Deep Learning, Classification, CNN, OCR, Dual-use Item

I. 서 론

전략물자란 대량 파괴 무기, 재래식 무기, 그 운반수단인 미사일 및 이들의 제조·개발·사용 또는 보관 등의 용도로 전용(轉用)될 수 있는 군용 및 산업용 물품과 기술을 의미한다.

본 연구는 전략물자 통제품목 중 정보보안 품목군의 전략물자 해당여부 판정에 도움이 되는 전략물자

Received(05. 07. 2020), Modified(1st: 09. 02. 2020, 2nd: 11. 17. 2020), Accepted(11. 17. 2020)

† 주저자, minbill38@korea.ac.kr
‡ 교신저자, jiwon_yoon@korea.ac.kr(Corresponding author)

판정 지원도구 개발에 대한 연구로 딥러닝과 광학문자판독(OCR) 기술을 활용하여 전략물자 해당여부 판정에 쉽게 접근하도록 하는 것에 목적이 있다.

II. 전략물자 판정

2.1 판정 개념

전략물자의 판정이란 대상 물품 등이 별표2(이중용도품목), 별표 2의2(상황허가 대상품목), 별표 3(군용물자품목)에 해당되는 것인지 여부를 판단하는 것으로 자가판정과 전문판정으로 구분된다. 자가판정이란 무역거래자가 자체적으로 전략물자 여부를 판단하는 것으로 전문판정기관에 의한 판정인 전문판정과 구별된다. 전략물자의 판정은 대상 품목이 물질 혹은 장비인지 먼저 확인하고 기술사양이 통제기준에 충족하는지를 확인하여 전략물자 해당되는지 여부를 판정하는 것이다. 기본적으로 판정이라는 절차는 수출품목과 통제기준의 기술적 사양을 비교하는 과정이다. 따라서 현재의 기준으로 판정을 하기 위해서는 수출품목의 기술적 사양을 알 수 있는 문서(카탈로그, 매뉴얼, 도면 등)와 판정을 하고자 하는 품목과 연관된 적절한 통제기준이 필요하다. 이 2가지가 준비되어 있다면 그다음은 기술 사양이 각 통제기준에 부합하는지 분석하는 과정을 거쳐 최종적으로 전략물자 해당여부를 판단할 수 있게 되는 것이다.

2.2 판정 방법

전략물자의 판정은 수출자가 직접 판정하는 자가판정과 판정 전문기관에 신청하여 그 결과를 받아보는 전문판정 2가지 방법이 있다. 자가판정은 전략물자관리시스템(www.yestrade.go.kr)에서 제공하는 온라인 자가판정시스템을 활용하거나 직접 전략물자 수출입고시 별표2의 통제기준 목록에 대해 판정하여 그 결과를 확인하는 방식이 있다. 자가판정의 경우 판정 결과를 즉시 확인하고 출력하여 효력을 얻을 수 있는 장점이 있으나 판정 결과에 대한 책임이 무역거래자에게 있다. 전문판정은 전략물자 해당 여부를 판정 전문기관에 의뢰하여 결과를 얻는 방법으로써, 해당 서비스는 무료로 진행되어 공인된 판정 결과를 얻을 수 있으나 법적 처리기간 15일이 소요된다.

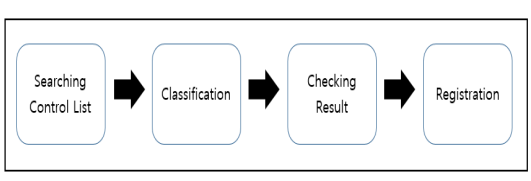


Fig. 1. Process of Online Self-Classification

III. 딥러닝(CNN) 관련 연구

3.1 AlexNet

AlexNet은 2012년 토론토대학의 Alex가 NIPS 학회에서 처음 발표하였는데, 2개의 GPU로 병렬연산을 수행한다는 점이 기존의 알고리즘과 다른 가장 큰 변화다. 총 8개의 레이어로 구성되어있는데 5개의 합성곱레이어(Convolutional Layer)와 3개의 완전히 연결된 레이어(Fully Connected Layer) 구조를 통하여 높은 성능을 구현하였다. AlexNet은 합성곱 신경망(CNN) 구조로 이미지 좌우 반전, 위치 변화, 평균값 연산 등의 Data Augmentation 함수를 사용하여 학습성능을 높여 다양한 분야에서 활용되고 있다. 또한, 학습 시 발생하는 과적합(over fitting)을 해결하기 위해 드롭아웃(Dropout) 기법을 적용하여 특정 뉴런에만 치우치지 않도록 하여 학습효율을 높였다.

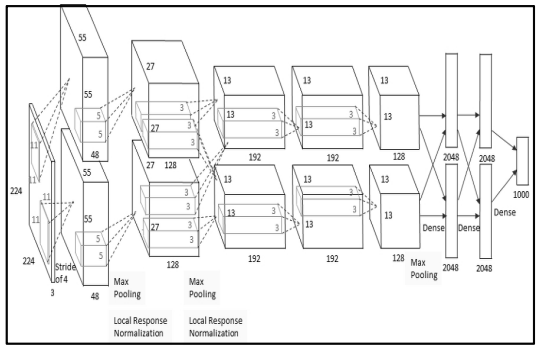


Fig. 2. Architecture of AlexNet[1]

3.2 GoogLeNET

GoogLeNet은 2014년 ILSVRC(ImageNet Large Scale Visual Recognition Challenge)에서 1등을 차지한 알고리즘으로 Inception이라는 이름으로 알려진 구글에서 발표한 네트워크이다. 일반적으로 딥러닝 네트워크의 성능은 구조가 깊어질수

록(deep), 그리고 레이어가 넓을수록(wide) 성능이 좋아진다고 알려져 있다. 하지만 실제로 학습 때는 파라미터가 많아지면서 과적합(over fitting)이나 기울기 소실(Gradient Vanishing)과 같은 문제가 생겨 학습이 어려워진다. 여기서 GoogLeNet의 핵심 아이디어는 제한된 계산 자원을 이용해 최적의 성능을 낼 수 있는 네트워크를 만들려고 한 것이다. GoogLeNet은 다른 알고리즘보다 작은(3×3, 1×1) 합성곱레이어 여러 개를 한 개의 모듈로 구성하고 결과를 합치는 형태로 네트워크를 구성하였다.

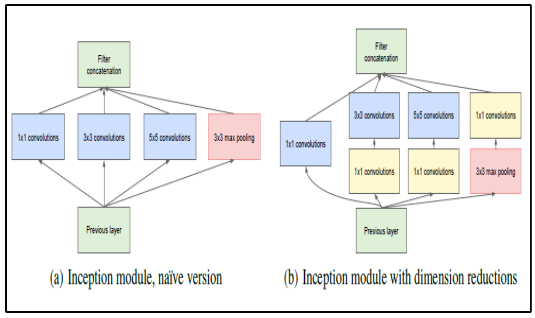


Fig. 3. Inception Module[2]

3.3 VGGNet

VGGNet은 2014년 영국의 K. Simonyan이 제안한 알고리즘으로 8레이어인 AlexNet 보다 깊은 구조의 19레이어를 갖는다. 합성곱 필터(Convolutional Filter)를 3×3을 사용한 것이 특징이며 8개의 합성곱 레이어(Convolutional Layer)와 3개의 완전히 연결된 레이어(Fully Connected Layer) 구조로 구성되어 있다. 깊이가 주는 영향을 밝히기 위해 receptive field의 크기는

ConvNet Configuration					
A 11 weight layers	A-LRN 11 weight layers	B 13 weight layers	C 16 weight layers	D 16 weight layers	E 19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Fig. 4. ConvNet configurations[3]

Table 1. Number of Parameters(in millions)[3]

Network	A,A-LRN	B	C	D	E
Number of parameter	133	133	134	138	144

가장 간단한 3×3으로 정하고 위 그림과 같이 6개의 구조에 대해 연구하였는데, 3×3 kernel을 사용하면 5×5, 7×7 혹은 그 이상의 합성곱을 인수분해하여 깊이는 깊어지지만 파라미터의 수를 줄여서 표현할 수 있다는 논리이다.

하지만 파라미터의 개수가 너무 많다는 단점이 있다. GoogLeNet의 파라미터의 개수가 5 million이었던 것에 비해 VGGNet은 가장 단순한 구조에서도 이미 133 million으로 상당히 많은 수준임을 확인할 수 있다.

3.4 ResNet

ResNet은 2015년 K. He 외 3명에 의해 제안된 네트워크로 GoogLeNet이 22레이어로 구성된 것에 비해 약 7배 이상 더 깊은 152 레이어를 갖는다. ImageNet 태스크에서 오류율 3.6%를 달성한 알고리즘으로 앞서 알아본 바와 같이 딥러닝의 네트워크는 구조가 깊을수록, 그리고 넓을수록 성능이 좋다고 하지만 레이어가 깊어질수록 파라미터가 많아지고 학습이 어려워지는 문제가 있다. K. He는 여기서 skip connection이라는 개념의 깊은 구조의 네트워크 모델을 제안하였다. 기존의 AlexNet, VGGNet, GoogLeNet은 합성곱레이어를 통해 특성맵(Feature maps)을 변환시켜왔지만, ResNet은 입력에 추가한 값을 합성곱으로 학습시켰다.

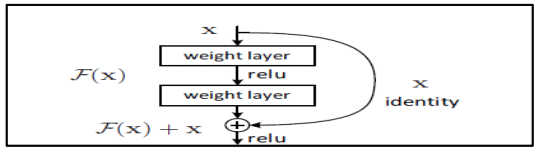


Fig. 5. Residual learning: a building block[4]

IV. 딥러닝 모델학습 및 OCR연구

4.1 자료수집을 위한 대상 선정

전략물자 이중용도 품목은 산업 전반에 걸쳐 있어 방대한 품목군으로 구성되어있는데 실제 1,800여 종

의 품목군이 전략물자에 해당될 수 있다. 본 연구에서는 정보보안 품목군에 딥러닝 기술을 적용하기 위해 방화벽과 집적회로(IC) 2가지 품목군과 비교 연구를 위해 화학 분야의 품목군인 밸브(valve)와 펌프(pump)를 선정하였다.

4.2 딥러닝 기술 적용

4.2.1 품목 이미지 데이터셋 수집 및 정제

본 연구에서는 딥러닝 학습 절차를 총 4단계로 구성하였다. 학습을 수행하기 위한 선행절차로 학습대상 이미지(데이터셋)를 수집하는 것이 첫 번째다. 두 번째로 학습성능을 높이기 위해 이미지를 선별하고 정제하는 과정이다. 세 번째 과정은 충분한 데이터셋을 확보하기 위해 이미지를 가공하는 단계다. 마지막으로 데이터셋이 확보되면 딥러닝 모델학습을 수행하여 결과를 출력한다.

학습대상 데이터셋 구성을 위해 4가지 품목에 대하여 구글과 네이버에서 이미지를 수집하였다. 수집한 이미지에는 실물 사진이 아닌 도면이나 설명이 포함된 학습에 부적절한 이미지가 섞여 있기에 데이터 선별작업이 필요하다. 선별한 데이터는 삭제하고 정제된 데이터셋을 대상으로 학습을 수행한다.

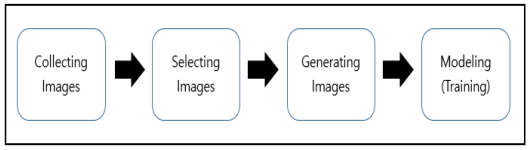


Fig. 6. Training Process of Deep Learning

Table 2. Collected Images

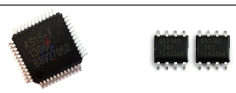
Items	Number of Images	Sample Images
Firewall	176	
Integrated Circuit	143	
Valve	312	
Pump	233	

Table 3. Selected Images

Category	number of images	
	Original	Selected
Firewall	176	147
Integrated Circuit	143	91
Valve	312	264
Pump	233	197

4.2.2 품목 이미지 데이터셋 가공

이미지 데이터셋을 수집한 후, 딥러닝 학습모델의 성능을 높이기 위해 이미지를 모두 동일한 크기로(100×100 픽셀) 맞추어주는 과정이 필요하여 파이썬으로 이미지의 크기 변경 작업을 수행하였다. 이미지 크기 변경 후에는 학습의 성능을 높일 수 있도록 Keras에서 제공하는 ImageDataGenerator 클래스를 활용하여 이미지의 수량을 증가시키는 과정을 수행한다. 데이터셋이 충분하지 않으면 과적합(over fitting) 문제가 발생하기 때문에 이미지를 효과적으로 분류하기 위해 대상의 핵심 특징을 학습하는 것에 중점을 두어야 한다. ImageDataGenerator 클래스를 이용해서 이미지를 증가시키는 목적이 바로 이 과적합(over fitting) 현상을 막는데 있다. 품목별 이미지의 데이터셋을 증가시킬 때 품목 이미지의 각도와 크기에 변화를 주고 반전 및 회전의 효과를 주어 품목별로 충분한 데이터셋을 확보한다. Keras ImageDataGenerator 클래스를 활용하여 파이썬 코드를 수행하면 이미지에 회전, 반전, 이동 등의 효과를 주어 학습대상 데이터셋을 생성하게 된다. 정제한 이미지 데이터셋을 품목별로 약 9,000개로 증가

Table 3. Generated Image Sample

original	rotation	flip	zoom
			

Table 4. Number of Generated Images

Type	number of images	
	Original	Generated
Firewall	147	8,581
Integrated circuit	91	8,399
Valve	264	9,625
Pump	197	8,959

시켜 충분한 학습대상 이미지를 확보한다.

4.2.3 학습(Training) 및 결과 도출

수집된 데이터셋을 대상으로 딥러닝 학습을 수행하는데 CNN 알고리즘을 적용하여 MATLAB 프로그램을 활용하여 수행하였다. 이미지의 특징(feature)을 추출하기 위해 Convolution 단계를 거쳐 특정 영역을 형성한 후에 가장 큰 값을 도출하는 Max Pooling 과정을 수행한다. 이후 완전히 연결된 레이어(Fully Connected Layer) 구조를 거쳐 최종적으로 이미지를 분류한다.

MATLAB 프로그램에서는 CNN 네트워크 수행 시 학습 옵션(TrainingOptions)을 설정할 수 있는데, 학습율(InitilalLearnRate)은 0.03과 0.05, 0.07로 설정하여 비교할 수 있도록 수행하였다. 학습율을 변경하여 수행한 결과 CNN 학습결과는 모두 상이하게 도출되었는데 학습율 0.05일 때 수행시간은 25분 26초였고, 예측 정확도는 95.1%가 도출되었다.

모델학습에서 학습율이 너무 낮게 설정되면 훈련시간이 오래 걸리고, 학습률이 너무 높으면 훈련이 최적의 결과보다 못한 값에 도달하거나 발산될 수 있으므로 적절한 값을 설정할 필요가 있다.

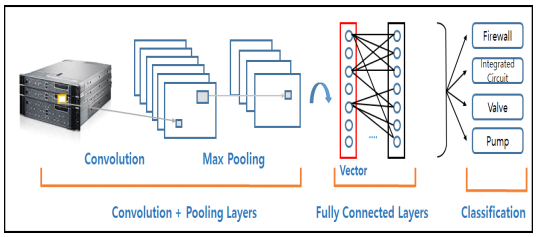


Fig. 7. CNN Architecture

Table 5. Result of Training

Training rate	Time	Accuracy(%)				
		F/W	IC	Valve	Pump	Avg.
0.03	25'18"	91.0	95.0	99.0	91.5	94.1
0.05	25'26"	91.5	97.5	99.5	92.0	95.1
0.07	23'29"	86.0	93.5	97.5	88.0	91.2

4.3 OCR을 활용한 문자 분석

앞에서는 이미지 인식과 분류를 위해 딥러닝 CNN 모델을 활용하였다. 이미지 분류와 함께 이미지에 포함된 문자열을 인식하면 복합적이고 고도화된 전략물자 판정정보를 제공할 수 있다. 광학 문자의 판독을 위해 오픈소스 OCR 프로그램인 Tesseract를 통해 이미지에서 텍스트파일을 추출한다.

정보보안 품목은 암호화 알고리즘이나 프로토콜이 포함된 경우 전략물자에 해당될 가능성이 높아지는데 주요 키워드는 아래 표와 같다. 따라서 매뉴얼이나 카탈로그에 포함된 주요 암호화 알고리즘이나 프로토콜 키워드를 판독하여 해당 정보를 제공함으로써 전략물자에 해당될 가능성이 높음을 알려줄 수 있다.

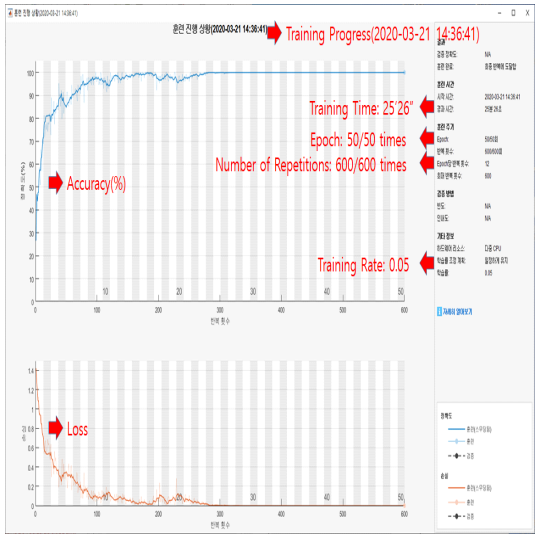


Fig. 8. Training Progress In MATLAB

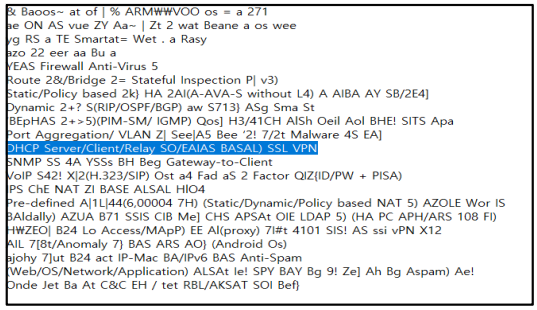


Fig. 9. Result of Optical Character Reader

Table 6. Keywords for Classification of Information Security Items

Type	Contents
Algorithm	AES, TEA3, 3-DES, SEED, RSA, ARIA, ECC, ECDH, ECDSA, RC5, Diffie-Hellman over Z/pZ, Diffie-Hellman over an elliptic curve
Protocol	SSL, TLS, SSH, SSHv2, HTTPS, IPsec, MACsec, VPN, SCP, SFTP, SNMP v3, LTO, IBM IMM, FDE, PGP, WEP, WPA-PSK, WPA-TKIP, WPA-EAP, AES-CCMP, WPA, WPA2, WPA-2PSK, TKIP, HP Insight control, TrustSec, Supervisor Engine, HP iLO4, GETVPN, LDAP, H.235, EMC OneFSOS, SED, wireless HART, SRTP, ZRTP, S/MIME

V. 결 론

본 연구에서는 딥러닝과 OCR 기술을 활용한 전략물자 판정 지원도구 개발에 대해 연구하였다. 이미지 분류 분야에서 이미 딥러닝 기술은 인간의 오류율(5%)보다 낮은 성과를 도출하였다. 이러한 성과를 전략물자의 판정에 적용하여 검증절차를 마련한다면 많은 시행착오와 잘못된 판정결과를 비교함으로써 판정의 오차를 줄일 수 있을 것으로 기대한다. 방대한 산업 분야에 분포되어 있는 전략물자 DB에 대해 충분한 데이터셋을 확보하고 우리나라의 우수한 전략물자 판정기술과 노하우를 접목하여 딥러닝 기반의 판정 시스템을 정비하게 된다면 전략물자 제도를 처음 접하는 사용자라 할지라도 전략물자 판정결과 혹은 유용한 정보를 쉽게 확인할 수 있게 될 것이다.

또한, 이미지 분류에서 발생하는 오류율을 보완하기 위해 OCR 기술을 활용하여 텍스트를 추출하여 정보보안품목의 주요 판정키워드와 비교하였다. OCR 기술은 이미 1920년대 후반부터 시작되었고 물류와 금융 산업을 필두로 우편물 관리, 자동차 번호판 인식, 명함 인식 등 다양한 산업 분야에서 활용되고 있어 전략물자 판정 분야에도 충분히 활용할 수 있는 기술이다. 연구에서처럼 판정 대상 품목의 주요 키워드를 추출하여 비교 분석함으로써 전략물자 판정에 결정적인 기준정보를 제공할 수 있다.

References

[1] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", Advances in Neural Information Processing Systems, pp. 1097-1105, 2012

[2] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, "Going Deeper With Convolutions," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-9, 2015

[3] Karen Simonyan, Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-14, 2014

[4] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, 2016

[5] Sunggyun Im, Youngbae Jeon, Junghwan Hwang, Jiwon Yoon, "Ship Detection using CNN based on Contrast Fusion Technique in Satellite Images : Accuracy Enhancement," Journal of The Korean Institute of Communications and Information Sciences, 46(08), pp. 823-833, Aug. 2019

[6] Myung ho Kim, "Application of Deep Learning Technique for Detecting Construction worker wearing Safety Helmet Based on Computer Vision," Journal of the Korean Society of Safety, 34(6), pp. 29-37, Dev. 2019

[7] Bongmo Kim, "Deep learning based image classification technology trend,"

Journal of The Korean Institute of Communications and Information Sciences, 35(12), pp. 8-14, Nov. 2018

〔8〕 Oh Wonsik, Lee Ugwiyeon, Oh Jeongseok, "Deep Learning(CNN) based Worker Detection on Infrared Radiation Image Analysis," Journal of The Korean Institute of Gas, 22(6), pp. 8-15, 2018

〔9〕 Chae-won Shin, Chulyun Kim, "Image Category Classification Based on Deep Learning," Jourrnal of The Korean Institute of Communications and Information Sciences, pp. 95-97, Jun. 2018

〈저자소개〉

조 재 영 (Jae-Young Cho) 정회원
2007년 2월: 성균관대학교 컴퓨터공학 졸업
2007년~2011년: 현대그린푸드 IT사업부
2011년 10월~현재: 전략물자관리원 선임연구원
2020년 9월: 고려대학교 정보보호대학원 사이버보안학과 석사
〈관심분야〉 정보보호, 딥러닝, 인공지능

윤 지 원 (Ji-Won Yoon) 종신회원
2008년: University of Cambridge 통계신호처리 박사
2011년~2012년: IBM Research Lab
2012년~현재: 고려대학교 정보보호 대학원 교수
2016년~현재: 정보보호학회 이사
2016년 6월~현재: 서울경찰청 사이버 보안 자문위원
2019년 10월~현재: 국립전파연구원 전파보안 자문위원
〈관심분야〉 통계신호처리, 베이지안 기법, 인공지능