

차원축소 없는 채널집중 네트워크를 이용한 SAR 변형표적 식별

박지훈^{*.1)} · 최여름¹⁾ · 채대영¹⁾ · 임 호¹⁾

¹⁾ 국방과학연구소 국방인공지능기술센터

SAR Recognition of Target Variants Using Channel Attention Network without Dimensionality Reduction

Ji-Hoon Park^{*.1)} · Yeo-Reum Choi¹⁾ · Dae-Young Chae¹⁾ · Ho Lim¹⁾

¹⁾ Defense AI Technology Center, Agency for Defense Development, Korea

(Received 29 September 2021 / Revised 22 February 2022 / Accepted 6 May 2022)

Abstract

In implementing a robust automatic target recognition(ATR) system with synthetic aperture radar(SAR) imagery, one of the most important issues is accurate classification of target variants, which are the same targets with different serial numbers, configurations and versions, etc. In this paper, a deep learning network with channel attention modules is proposed to cope with the recognition problem for target variants based on the previous research findings that the channel attention mechanism selectively emphasizes the useful features for target recognition. Different from other existing attention methods, this paper employs the channel attention modules without dimensionality reduction along the channel direction from which direct correspondence between feature map channels can be preserved and the features valuable for recognizing SAR target variants can be effectively derived. Experiments with the public benchmark dataset demonstrate that the proposed scheme is superior to the network with other existing channel attention modules.

Key Words : Synthetic Aperture Radar(합성개구면레이더), Automatic Target Recognition(표적자동식별), Channel Attention(채널집중), Dimensionality Reduction(차원 축소), Deep Learning(딥러닝)

1. 서론

날씨와 주야에 관계없이 영상을 촬영할 수 있는 합성 개구면 레이더(Synthetic Aperture Radar, SAR) 센서

가 증가함에 따라, SAR 영상을 관독하는 전문 인력의 업무 부담을 경감시켜줄 표적자동식별(Automatic Target Recognition, ATR) 기술의 중요성도 지속적으로 부각되고 있다. SAR 영상에서의 표적자동식별 기술 또한 근 30여 년간 연구가 수행되어 왔으며, 초기에는 주로 표준영상 정합(template matching)^[1]이나 모델기반 컴퓨터 비전(Model-Based Vision, MBV)^[2]기반 기법 등이 차

^{*} Corresponding author, E-mail: dydynoel@add.re.kr
Copyright © The Korea Institute of Military Science and Technology

용되었다. 이후에는 서포트 벡터 머신(Support Vector Machine, SVM), 최 근접 이웃(Nearest Neighbor, NN) 분류기와 같은 패턴인식(pattern recognition) 기반의 기법이 SAR-ATR 분야에 적용되어 의미 있는 결과를 도출하였다^[3,4]. 최근 5년 동안에는 컴퓨터 비전 분야에서 큰 성공을 거둔 딥러닝(deep learning) 기반의 기법이 종래의 기법보다 우수한 SAR-ATR 성능을 도출함으로써 관련 연구의 대세가 되어가고 있다. 딥러닝의 대표적 기법인 합성곱 신경망(Convolutional Neural Network, CNN)을 SAR-ATR 분야에 적용하는 것 외에도^[5], 패턴인식 기법과의 연계^[6], 새로운 네트워크 구조의 제안^[7-9], 영상 시퀀스 식별^[10], 생성망(generative network)과 결합한 데이터 증대^[11], CNN 도출 특징 간 융합^[12] 등 실로 다양한 딥러닝 기반의 응용 기법이 SAR-ATR에 적용되었다.

강인한 SAR-ATR 시스템의 요건 중 가장 중요한 것이 식별을 위한 DB나 네트워크를 학습시키는데 사용되는 표적영상과 실제 입력되는 표적영상 간의 차이점이 적은 표준운용조건(Standard Operating Condition, SOC)에서 뿐만 아니라 차이가 커지는 확장운용조건(Extended Operating Condition, EOC)에서의 성능이 우수해야 한다는 것이다. EOC의 대표적인 사례로 ‘변형 표적(target variant)을 들 수 있는데, 이는 동종의 표적이라도 차량의 형태나 부가물의 종류 및 위치, 시리얼 번호, 생산 버전 등과 같은 다양한 요인이 발생하는 경우를 지칭한다. 전자파 산란에 의해 형성되는 SAR 영상의 특성 상, 표적의 종류가 동일해도 변형 내역에 따라 실제 형성되는 SAR 표적영상이 민감하게 달라질 수 있다. 이렇게 발생하는 표적영상 내의 작은 변화를 극복하고 일관적이고 정확한 식별을 수행하는 SAR-ATR 시스템이 강인하다고 할 수 있다.

본 논문에서는 채널집중(channel attention) 모듈을 CNN에 삽입한 채널집중 네트워크를 위에서 언급한 SAR-ATR에서의 변형표적 식별문제에 적용한다. 딥러닝 네트워크에서의 채널집중 기법의 경우, 2018년 ‘Squeeze-Excitation(SE)’ 네트워크^[13] 이래 활발한 연구가 수행되었으며, 식별에 유용한 특징을 선택적으로 강조하는 작용으로 인해 사진영상 뿐만 아니라 SAR 표적영상 식별에도 효과적으로 적용된 바 있다^[14,16]. 이러한 작용을 SAR-ATR에서의 변형표적 식별 성능 향상에 활용하고자 하는 것이 본 논문의 주요한 연구 동기이다. 또한 다양한 딥러닝의 SAR-ATR 적용 사례 중에서 채널집중 기법을 변형표적 문제에 적용한 사

례를 찾아보기 힘들다는 것도 또 다른 연구동기가 된다. 한편 이전의 연구사례에서 활용된 채널집중 기법들은 대부분 네트워크의 특징맵(feature map)의 채널특성 강화에 사용되는 여기벡터(excitation vector)를 채널방향의 차원축소(dimensionality reduction)를 통해 획득하였다. 그러나 일부 연구^[17]에서 지적한 바와 같이 이러한 차원축소 과정은 특징맵을 구성하는 채널 간의 직접적인 연관성을 파괴하여 성능향상에 한계를 야기할 수 있으므로, 본 논문에서는 차원축소 과정이 없는 채널집중 모듈을 딥러닝 네트워크 CNN에 도입하여 변형표적 식별 문제에 적용하고자 한다.

본 논문의 2장에서는 기본 CNN 구조 및 여기에 적용될 채널집중 모듈의 구조, 3장에서는 변형표적에 관한 MSTAR 공개 데이터셋 및 식별시험 결과, 4장에서는 채널집중 작용에 대한 고찰을 채널 활성화 맵을 통해 다룬다. 5장에서는 본 논문의 결론을 내린다.

2. 네트워크 구조

2.1장에서는 본 논문의 SAR 변형표적 식별에 관한 베이스라인(baseline)이 되는 CNN 형태의 기본 네트워크의 구조를 다루며, 2.2장에서는 기본 네트워크에 삽입될 3종의 채널집중 모듈의 구조를 다룬다.

2.1 기본 네트워크 구조

Fig. 1에는 기본 네트워크의 전체 구조를 나타내었다. 먼저 ‘stage 1’에서는 128×128 크기의 입력영상을 받아서 커널크기가 3×3이고, 커널 개수가 64개이며, 커널계산 보폭(stride) 2 및 영상 둘레에 0을 배치시키는 폭의 개수(zero-padding) 1인 합성곱(conv)을 수행한다(conv 3×3 @64, s2p1). Fig. 1의 ‘conv’로 표기된 네트워크 계층에는 배치 정규화(batch normalization)와 ReLU 활성화가 포함되어 있다. ‘stage 1’의 연산 결과, 보폭에 의해 가로/세로 크기가 1/2로 줄고 채널방향 차원이 64로 늘어난 64×64×64 크기의 특징맵이 도출된다. 이어서 ‘stage 2’부터 ‘stage 5’까지는 입력되는 특징맵이 합성곱 연산 등에 의해 출력된 특징맵과 스킵 연결(skip connection)을 통해 합쳐지는 잔차(residual) 구조를 가지며 이는 K. He 등에 의해 고안된 레즈넷(ResNet)^[18]의 구조와 흡사하다. ‘stage 2’에서는 ‘stage 1’과 같이 64×64×64 크기의 특징맵이 다루어지고, ‘stage 3~5’에서는 가로/세로 크기가 1/2로 줄고, 채널

크기는 2배 늘어남으로써 각각 $32 \times 32 \times 128$, $16 \times 16 \times 256$, $8 \times 8 \times 512$ 크기의 특징맵이 다루어진다. 마지막 'stage 5'에서 도출되는 $8 \times 8 \times 512$ 크기 특징맵은 전역평균풀링(Global Average Pooling, GAP)에 의해 $1 \times 1 \times 512$ 크기로 압축되고, 4개의 노드를 갖는 완전 연결 계층(fully connected layer, fc) 및 소프트맥스(softmax) 계층에 의해 노드별 확률 값으로 치환된다. 이 때 각 노드는 식별대상 표적의 종류 개수와 일치한다. 소프트맥스 값을 바탕으로, 네트워크의 학습 과정에서는 이진 벡터 형태(one-hot encoded)의 실제 목표 값과 교차 엔트로피(cross-entropy) 손실을 구함으로써 네트워크를 구성하는 가중치를 기울기(gradient) 기반 오류 역전파(back-propagation)로 업데이트 한다. Fig. 1의 구조에서 빨간색 화살표로 표기된 부분은 뒤에서 설명할 채널 집중 모듈이 삽입될 위치를 가리킨다.

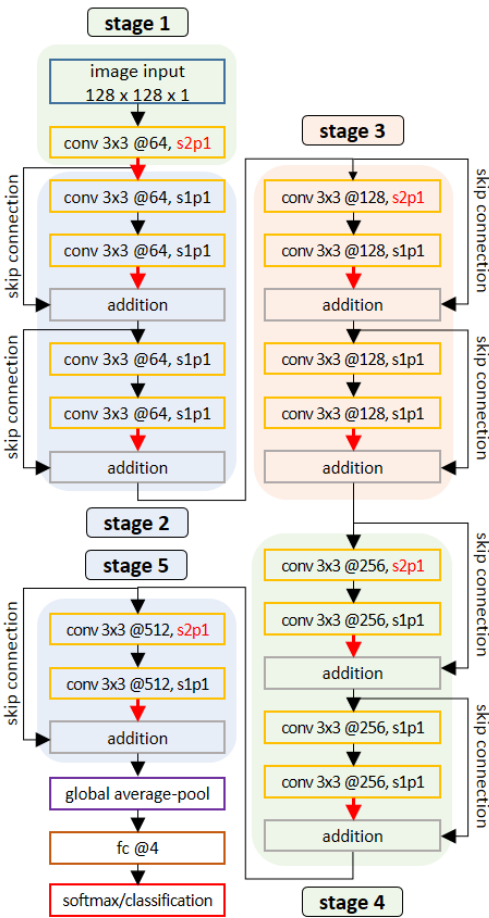


Fig. 1. Structure of basic network

2.2 채널집중 모듈 구조

이미 많은 연구사례에서 다양한 채널집중 모듈 구조가 제안된 바 있으며, 본 논문에서는 그중 대표적인 모듈 3종에 대해서만 다루도록 한다.

2.2.1 SE(Squeeze-Excitation) 채널집중 모듈

Fig. 2에는 대표적인 채널집중 모듈인 SE 모듈^[13]을 나타내었다. 모듈에 H (세로, 높이) $\times W$ (가로, 너비) $\times C$ 크기를 갖는 특징맵이 입력되었을 때, 전역평균풀링(GAP)에 의해 특징맵의 채널정보가 $1 \times 1 \times C$ 의 크기로 압축된다. 뒤이어 C/r 및 C 개의 노드를 갖는 2개의 완전 연결 계층 fc에 의해 채널방향 차원이 축소되었다가 복원되는 과정을 거침으로써 중요한 채널의 정보를 강조하고 덜 중요한 채널의 정보를 억제하는 방향으로 학습이 진행된다. 첫 번째 fc에는 ReLU 활성화가 사용되고 두 번째 fc에는 시그모이드(sigmoid) 활성화가 사용되어, 사실상 채널 별 중요도의 가중치로 치환된다. 이렇게 산출된 채널방향 여기(excitation)벡터를 입력 특징맵에 채널 방향으로 곱해주면, 채널방향으로 집중(attention) 효과가 반영된 결과를 도출할 수 있다. 이러한 SE 모듈 및 변형된 형태가 SAR 표적식별^[14,15] 뿐만 아니라, SAR 선박탐지에도 활용된 바 있다.

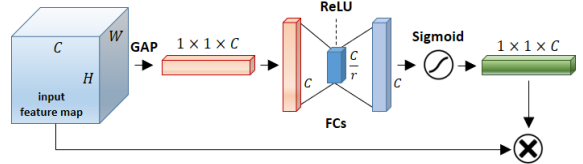


Fig. 2. Structure of SE module

2.2.2 CBAM(Convolutional Block Attention Module)

CBAM^[19]은 국내 연구진에 의해 개발된 집중 모듈로서, 채널방향 뿐만 아니라 공간(spatial)방향으로의 집중작용을 순차적으로 수행하여 SAR 표적식별^[16] 뿐만 아니라 SAR 선박탐지에도 활용되었으며, SAR 영상 외에도 다양한 광학영상에 적용되어 성능 향상효과를 입증하였다. Fig. 3에는 이러한 CBAM의 구조를 나타내었다. 먼저 입력된 특징맵으로부터 GAP과 전역 최대풀링(Global Maximum Pooling, GMP)에 의해 채널 방향으로 압축된 $1 \times 1 \times C$ 의 크기 벡터 2개를 산출한다. 그리고 각각의 벡터에 대해 SE 모듈과 같이 C/r 및 C 개의 노드를 갖는 2개의 완전 연결 계층 fc을 통과시킨다. 이렇게 통과된 2개의 $1 \times 1 \times C$ 의 크기 벡터를 상

호 더한 다음, 시그모이드에 의해 도출된 채널방향 여기(excitation)벡터를 입력 특징맵에 채널 방향으로 곱해준다. 이렇게 채널방향 집중 효과를 반영한 특징맵에 대하여 채널방향 평균 풀링(Channel-wise Average Pooling, CAP) 및 채널방향 최대값 풀링(Channel-wise Maximum Pooling, CMP)으로 도출된 $H \times W \times 1$ 크기 특징맵 각각을 채널 방향으로 쌓은 후(concatenation), 커널 개수 1개인 1×1 합성곱 연산 및 시그모이드에 의해 $H \times W \times 1$ 크기의 공간방향 여기 맵(excitation map)을 구한다. 이러한 맵을 채널방향 집중 후의 특징맵에 공간방향으로 곱해주면 공간방향 집중 효과도 반영된 특징맵이 생성된다. 요약하면 CBAM은 SE 모듈과 유사한 채널차원 축소 형태의 채널집중을 거친 후, 공간방향 집중이 순차적으로 이루어진다.

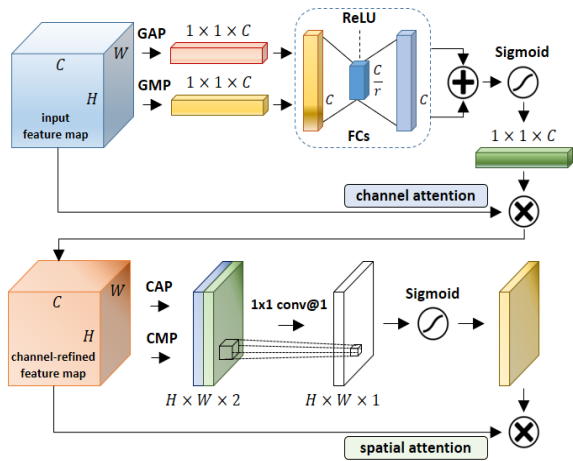


Fig. 3. Structure of CBAM

2.2.3 ECA(Efficient Channel Attention) 모듈

Fig. 4에는 채널집중 모듈의 하나인 ‘ECA(Efficient Channel Attention)’ 모듈을 나타내었다. 특징맵으로부터 GAP에 의해 압축된 채널벡터를 산출하고, 일련의 과정을 통해 도출된 벡터에 시그모이드로 여기벡터를 산출하여 특징맵에 곱한다는 점에서 SE 모듈과 흡사하다. 그러나 상기 ‘일련의 과정’에서 큰 차이가 있다. SE 모듈이나 CBAM의 경우 채널집중을 위해 압축된 채널벡터에 대해 fc 에 의한 차원축소를 수행하는 반면, ECA 모듈에서는 이러한 차원축소를 거치지 않는다. 대신 GAP으로 산출된 $1 \times 1 \times C$ 의 크기 채널벡터에 커널 크기 k 를 갖는 1차원 합성곱(1D conv)을 수행한다. 이때 채널벡터 양 끝단에 0을 추가하여 1차원 합성곱에

의해 채널벡터의 크기(C)가 변경되지 않도록 한다. ECA 모듈을 창안한 연구^[17]에서는 종래의 SE 모듈 및 그 변형 채널집중 모듈들에서 채널벡터의 차원이 C/r 개의 노드를 갖는 fc 에 의해 축소되는 것에 대한 문제를 제기하였다. 여기서 r 은 실험에 의해 결정되는 매개변수(hyper parameter)이며 통상 16으로 설정된다. 예를 들어 $C = 64$ 이면 SE 모듈에서는 채널방향 차원이 일거에 4로 축소되는데, 이러한 경우 전역적(global)인 채널 방향으로는 정보의 요약효과가 나타날 수 있으나, 국부적(local)으로는 인접 채널 간의 상호 연관성(direct correspondence)이 파괴되어 이후 C 개 노드를 갖는 fc 에 의해 원래 크기로 복원되더라도 손실된 채널 정보를 복원할 수 없다는 문제점이 발생한다. 이러한 상황에서는 여기벡터가 특징맵에 채널방향으로 곱해질 경우 중요한 정보를 담고 있는 채널에 인접한 모든 채널의 정보까지 강조되거나, 중요하지 않은 정보를 담고 있는 채널에 인접한 중요한 채널의 정보까지 억제되는 역효과가 발생할 수 있다. 따라서 Fig. 4에 나타낸 ECA 모듈과 같이 국부적인 관점에서 채널벡터에 대해 1차원 합성곱을 수행한다면 인접 채널 간의 상호 연관성을 충실히 반영할 수 있는 채널집중 효과가 나타날 것으로 기대된다. 더불어 ECA 모듈의 경우 채널집중을 위해 학습되는 가중치의 개수가 SE 모듈이나 CBAM에 비해 현저히 적기 때문에 모듈의 명칭에도 드러나 있듯이 효율성 측면에서도 우수하다. 본 논문에서는 이러한 ECA 모듈의 아이디어를 바탕으로 인접 채널 간의 상호 연관성이 충실히 반영된 채널집중 모듈을 2.1장의 기본 네트워크에 적용하여 SAR 영상에서의 변형표적 식별 문제에 적용하게 된다.

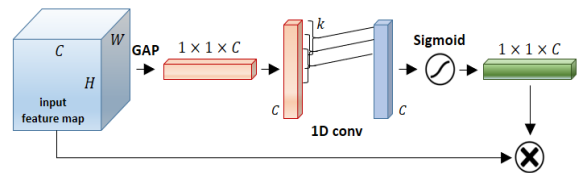


Fig. 4. Structure of ECA module

본 논문에서는 기본 네트워크 각 부분에 삽입되는 ECA 모듈의 1차원 합성곱 커널 크기 k 를 다음과 같이 실험적으로 결정하여 사용한다.

$$k_0 = \left\lfloor \frac{1}{2} (\log_2 C + 1) \right\rfloor \quad (1)$$

위 식 (1)에서 C 는 채널집중 모듈에 입력되는 특징맵의 채널수이고, $\lfloor \cdot \rfloor$ 는 소수점을 버리는 바닥(floor) 연산이다. 위의 k_0 가 짝수이면 커널의 크기는 $k = k_0 - 1$ 이고, 홀수이면 $k = k_0$ 이다. 즉 1차원 합성곱 커널의 크기가 입력되는 특징맵의 채널수에 따라 가변적으로 설정된다.

3. SAR 변형표적 식별 실험

3.1 MSTAR SAR 변형표적 영상 데이터 셋

실험결과 제시에 앞서 SAR 변형표적(target variants) 식별 시험에 사용되는 MSTAR 표적영상 데이터 셋에 대해 소개한다. MSTAR는 미 공군 연구소(Air Force Research Laboratory, AFRL) 주도로 수행된 SAR-ATR 연구 프로젝트로서, 미국 군용 장비 및 적성 장비의 SAR 표적영상을 다양한 측면 각도에 대하여 약 10만 여장 획득하였다. 이 중 일부가 공개된 MSTAR 데이터 셋을 이용하여 많은 연구자가 이를 SAR-ATR 연구에 활용해왔다. 본 논문에서는 이중 변형표적 식별 시험에 해당하는 SAR 표적영상만 제한적으로 활용한다. MSTAR 데이터 셋의 변형표적 식별 문제를 다루는 모든 문헌^[7,9,10]에서는 Fig. 5와 같은 학습용(training) 표적 및 두 가지의 시험용(test) 표적을 나눈다. 표적 종류 및 시리얼 번호 별 세부 내역은 Table 1에 나타내었다. 시험 표적 중 ‘EOC-C’는 형태(configuration)에 의한 변형표적 식별문제로서, 학습 및 시험 데이터 셋 내의 표적영상이 반응장갑(reactive armor) 및 보조 연료통(fuel barrel)과 같은 요소 등에서 다른 구성 측면에서의 변형에 해당된다. ‘EOC-V’는 일부 구조물이 탱크 포탑과

Table 1. Information on dataset for target variants

사용목적	표적	시리얼번호	내림각	영상개수
학습 (training)	BMP2	9563	17°	233
	BRDM2	E71	17°	298
	BTR70	C71	17°	233
	T72	132	17°	232
시험 (EOC-C)	T72	S7	15°/17°	191/228
	T72	A32	15°/17°	274/298
	T72	A62	15°/17°	274/299
	T72	A63	15°/17°	274/299
	T72	A64	15°/17°	274/299
시험 (EOC-V)	BMP2	9566	15°/17°	196/232
	BMP2	C21	15°/17°	196/233
	T72	812	15°/17°	195/231
	T72	A04	15°/17°	274/299
	T72	A05	15°/17°	274/299
	T72	A07	15°/17°	274/299
	T72	A10	15°/17°	271/296

주포의 회전, 해치(hatch)의 개폐와 같이 학습 및 시험 데이터 셋 간의 상대적 움직임 변화가 나타나는 표적 버전 측면에서의 변형에 해당된다^[10]. EOC-C 및 EOC-V 모두 네트워크 학습을 위해 동일한 영상을 사용한다. 내림각(depression angle) 측면에서는 학습용 표적영상의 경우 17도의 영상만 사용되는 반면, 시험용 표적영상은 17도 뿐만 아니라 15도의 영상도 사용된다.

3.2 실험결과

3.1장에서 소개된 학습용 데이터셋을 이용하여 4개의 네트워크 구조, ‘Net 1’, ‘Net 2’, ‘Net 3’, ‘Net 4’를 각각 학습시킨다. 여기서 ‘Net 1-4’는 다음과 같으며, 각 구조의 매개변수 개수도 함께 나타내었다.

- Net 1 : 기본 네트워크 / 6,451,076
- Net 2 : 기본 네트워크 + SE 모듈 / 6,507,424
- Net 3 : 기본 네트워크 + CBAM / 6,563,793
- Net 4 : 기본 네트워크 + ECA 모듈 / 6,451,108

네트워크 입력영상의 크기는 Fig. 1에 나타낸 바와 같이 128×128로 고정되고, NVidia Geforce GTX 1080 GPU 하드웨어 및 MATLAB Deep Learning Toolbox 소프트웨어를 이용하였으며, 학습 단위로 반복되는 미니



Fig. 5. Target variants for training and tests

배치(mini-batch) 수는 32이고, 매 세대(epoch) 마다 학습영상을 섞이게(shuffle)하였다. 여기서 세대란, 모든 학습 대상영상이 1번 씩 네트워크 학습에 이용되는데 소요되는 반복의 주기를 의미한다. 초기 학습률(learning rate)은 0.001이고 50세대마다 0.5배 감쇄되도록 설정하였으며, 최대 200세대까지 학습한다. 또한 각도 다양성이 부족해진 학습영상에 대해서 ±4픽셀만큼 좌우/상하로 학습영상을 평행이동(translation)하는 데이터 증대기법을 사용하였다. 이 때 증대된 영상을 물리적으로 생성하기 보다는 매 세대마다 임의로 평행이동의 양을 ±4픽셀 범위 내에서 가변시킴으로써 네트워크 학습에 소요되는 GPU 메모리를 줄인다.

3.2.1 EOC-C 실험

EOC-C 실험에서는 5개의 다른 시리얼 번호를 갖는 T72 표적을 식별한다. Table 2~5에는 각각 Net 1~4에 의한 EOC-C 실험 혼동행렬(confusion matrix)을 나타내었다. 행렬의 행(row)에는 식별 대상 표적영상의 표적 종류 및 시리얼 번호, 내림각을 나타내었으며, 행렬의 열(column)에는 네트워크가 예측한 표적의 종류를 나타낸다. 예를 들어 시리얼 번호가 ‘S7’인 내림각 15도의 T72 표적영상 191장 중 7장이 BMP2 표적으로 잘못 식별되었으며, 184장이 T72 표적으로 올바르게 식별되었다. 여기서 예측된 ‘T72’의 경우 실제로는 네트워크 학습에 이용된 시리얼 번호 ‘132’ 및 내림각 17도에 해당하는 영상이나, 본 논문이나 다른 연구사례에서와 같이 시리얼 번호나 내림각 관계없이 표적의 종류가 일치하면 올바르게 분류한 것으로 간주한다.

Table 2. Confusion matrix for EOC-C of Net 1

	BMP2	BRDM2	BTR70	T72	P _{cc} (%)
T72_S7_15	7	0	0	184	96.34
T72_S7_17	28	0	2	198	86.84
T72_A32_15	4	0	0	270	98.54
T72_A32_17	17	0	0	281	94.30
T72_A62_15	0	0	0	274	100.00
T72_A62_17	8	0	0	291	97.32
T72_A63_15	4	0	1	269	98.18
T72_A63_17	20	0	2	277	92.64
T72_A64_15	1	0	4	269	98.18
T72_A64_17	33	0	3	263	87.96
합계	122	0	12	2,576	95.06

Table 3. Confusion matrix for EOC-C of Net 2

	BMP2	BRDM2	BTR70	T72	P _{cc} (%)
T72_S7_15	0	0	0	191	100.00
T72_S7_17	14	0	0	214	93.86
T72_A32_15	0	0	0	274	100.00
T72_A32_17	0	0	0	298	100.00
T72_A62_15	0	0	0	274	100.00
T72_A62_17	2	0	0	297	99.33
T72_A63_15	0	0	0	274	100.00
T72_A63_17	5	0	0	294	98.33
T72_A64_15	0	0	0	274	100.00
T72_A64_17	9	10	0	280	93.65
합계	30	10	0	2,670	98.52

Table 4. Confusion matrix for EOC-C of Net 3

	BMP2	BRDM2	BTR70	T72	P _{cc} (%)
T72_S7_15	2	0	0	189	98.95
T72_S7_17	5	0	0	223	97.81
T72_A32_15	0	0	0	274	100.00
T72_A32_17	0	0	0	298	100.00
T72_A62_15	0	0	0	274	100.00
T72_A62_17	0	0	0	299	100.00
T72_A63_15	0	0	0	274	100.00
T72_A63_17	1	0	0	298	99.67
T72_A64_15	0	0	0	274	100.00
T72_A64_17	0	0	0	299	100.00
합계	8	0	0	2,702	99.70

Table 5. Confusion matrix for EOC-C of Net 4

	BMP2	BRDM2	BTR70	T72	P _{cc} (%)
T72_S7_15	0	0	0	191	100.00
T72_S7_17	6	0	0	222	97.37
T72_A32_15	0	0	0	274	100.00
T72_A32_17	0	0	0	298	100.00
T72_A62_15	0	0	0	274	100.00
T72_A62_17	0	0	0	299	100.00
T72_A63_15	0	0	0	274	100.00
T72_A63_17	1	0	0	298	99.67
T72_A64_15	0	0	0	274	100.00
T72_A64_17	0	0	0	299	100.00
합계	7	0	0	2,703	99.74

위 결과들로부터, 집중모듈이 적용되지 않은 기본 네트워크의 경우, 총 2,710장의 T72 시험 표적영상 중에서 2,576장을 ‘T72’로 올바르게 분류함으로써 식별 확률(Probability of correct classification, P_{cc}) 95.06 %의 성능을 보인다. 시험영상 중 122장이 BMP2 표적으로 잘못 분류되는데, 이는 기존의 딥러닝 기반 변형표적 식별 연구사례에서도 확인되는 현상이다. 이는 보병수송을 위한 장갑차인 BTR70이나, BRDM2와는 달리 BMP2는 T72 탱크와 같은 궤도 형태의 하부구조 및 해치에 매달린 포로 인하여 SAR 표적영상의 형태가 비교적 유사함에 따른 결과로 보인다⁷⁾. Net 1에 SE 모듈을 추가한 Net 2의 경우 3.5 % 가량의 성능 향상을 보이며, 식별 확률은 98.52 %이다. SE 모듈의 채널 집중 효과에 의해 네트워크 특장점 채널 정보의 선택적 강조 및 억제가 반영되어, 네트워크의 변형표적 분별력을 강화시킨 것으로 보인다. 이로 인해 BMP2 및 BTR70으로 잘못 식별된 시험영상의 개수가 크게 줄어들었다. Net 1에 CBAM을 추가한 Net 3의 경우에는 SE 모듈을 추가한 Net 2에 비해 약 1.2 % 성능이 더 향상된 99.70 %의 성능을 보인다. Fig. 3과 같이 CBAM은 기존의 SE 모듈에 GMP에 의한 채널벡터 추가형성 및 공간방향 집중모듈을 추가하여, 보다 높은 성능을 보인다. 끝으로 Table 5에는 기본 네트워크에 ECA 모듈이 적용된 Net 4에 의한 EOC-C 결과를 나타내었다. 결과 성능은 99.74 %로서, 4종의 네트워크 구조 중 가장 우수한 성능을 보인다. Net 2와 Net 4의 공통점은 기본 네트워크에 채널집중 모듈이 적용된 점인데, Net 2의 SE 모듈은 앞서 언급한 바와 같이 채널방향 차원의 축소를 수반하고, Net 4의 ECA 모듈은 이러한 축소과정이 없다는 차이점이 있다. 두 네트워크 구조의 성능을 비교해보면 본 논문의 서론에서 가정한 바와 같이 채널 간의 직접 연관성 파괴 유무가 EOC-C 시험의 식별 성능에 영향을 미친 것으로 판단할 수 있다. CBAM이 적용된 Net 3과 Net 4는 큰 성능차이는 없다. 그러나 Net 3이 두 번의 채널차원 축소를 수반하는 채널집중 과정과 공간방향 집중과정을 모두 포함하고 있는 것에 비해서, Net 4는 채널차원 축소가 없는 채널집중만으로도 보다 우수한 식별 성능을 낸다는 것이 주목할 만하다. 뿐만 아니라, SE 모듈과 CBAM 모두 채널집중 과정에 사용되는 fc에 의해 기본 네트워크 대비 56,348개(+0.87 %) 및 112,717개(+1.72 %)가 늘어난 매개변수를 필요로 하는 반면, ECA 모듈의 경우 1차원 합성곱 연산에 의해 불과 32

개가 늘어난 매개변수만으로도 높은 EOC-C 식별 성능을 보임으로써, 차원축소가 없는 채널집중 작용의 효율성을 확인할 수 있다.

3.2.2 EOC-V 실험

EOC-V 실험에서는 2개 및 5개의 다른 시리얼 번호를 갖는 BMP2 및 T72 표적을 식별한다. Table 6-9에는 각각 Net 1~4에 의한 EOC-V 실험 혼동행렬을 나타내었다. 기본 표기는 동일하되, 식별 확률 P_{cc} 의 경우 BMP2 시험영상은 ‘BMP2’로 예측된 건수, T72 시험영상은 ‘T72’로 예측된 건수에 대해 산출된다. 또한 혼동행렬의 맨 아래 행에는 전체 시험영상 대비 맞게 식별된 영상 수로부터 산출된 식별 확률(볼드체)과 표적 별(BMP2, T72) 식별 확률을 나타내었다.

Table 6~9의 EOC-V 실험 결과로부터, EOC-C와 EOC-V 실험이 모두 동일한 학습용 영상으로 네트워크 구조들이 학습된다는 것을 감안할 때, T72 표적의 식별 성능보다는 BMP2 표적의 식별 성능에 전체 성능이 좌우된다는 것을 알 수 있다. 예를 들어 T72 표적만의 성능을 살펴보면 기본 네트워크인 Net 1도 99.56 %의 높은 성능을 보이지만, BMP2 표적에 대한

Table 6. Confusion matrix for EOC-V of Net 1

	BMP2	BRDM2	BTR70	T72	P_{cc} (%)
BMP2_9566_15	168	1	7	20	85.71
BMP2_9566_17	217	5	0	10	93.53
BMP2_C21_15	172	0	0	24	87.76
BMP2_C21_17	220	0	0	13	94.42
T72_812_15	1	0	0	194	99.49
T72_812_17	3	0	0	228	98.70
T72_A04_15	2	0	0	272	99.27
T72_A04_17	0	0	0	299	100.00
T72_A05_15	0	0	0	274	100.00
T72_A05_17	2	0	0	297	99.33
T72_A07_15	0	0	0	274	100.00
T72_A07_17	1	0	0	298	99.67
T72_A10_15	2	0	0	269	99.26
T72_A10_17	1	0	0	295	99.66
	BMP2	90.67	T72	99.56	97.42

Table 7. Confusion matrix for EOC-V of Net 2

	BMP2	BRDM2	BTR70	T72	P _{cc} (%)
BMP2_9566_15	182	0	3	11	92.86
BMP2_9566_17	230	0	1	1	99.14
BMP2_C21_15	176	0	0	20	89.80
BMP2_C21_17	233	0	0	0	100.00
T72_812_15	0	0	0	195	100.00
T72_812_17	1	0	0	230	99.57
T72_A04_15	0	0	0	274	100.00
T72_A04_17	1	0	0	298	99.67
T72_A05_15	0	0	0	274	100.00
T72_A05_17	8	0	0	291	97.32
T72_A07_15	0	0	0	274	100.00
T72_A07_17	2	0	0	297	99.33
T72_A10_15	0	0	0	271	100.00
T72_A10_17	0	0	0	296	100.00
	BMP2	95.80	T72	99.56	98.66

Table 9. Confusion matrix for EOC-V of Net 4

	BMP2	BRDM2	BTR70	T72	P _{cc} (%)
BMP2_9566_15	185	0	9	2	94.39
BMP2_9566_17	227	0	5	0	97.84
BMP2_C21_15	183	1	0	12	93.37
BMP2_C21_17	233	0	0	0	100.00
T72_812_15	0	0	0	195	100.00
T72_812_17	0	0	0	231	100.00
T72_A04_15	0	0	0	274	100.00
T72_A04_17	0	0	0	299	100.00
T72_A05_15	0	0	0	274	100.00
T72_A05_17	0	0	0	299	100.00
T72_A07_15	0	0	0	274	100.00
T72_A07_17	0	0	0	299	100.00
T72_A10_15	0	0	0	271	100.00
T72_A10_17	0	0	0	296	100.00
	BMP2	96.62	T72	100.00	99.19

Table 8. Confusion matrix for EOC-V of Net 3

	BMP2	BRDM2	BTR70	T72	P _{cc} (%)
BMP2_9566_15	182	0	3	11	92.86
BMP2_9566_17	229	0	0	3	98.71
BMP2_C21_15	183	0	0	13	93.37
BMP2_C21_17	231	0	0	2	99.14
T72_812_15	0	0	0	195	100.00
T72_812_17	3	0	0	228	98.70
T72_A04_15	0	0	0	274	100.00
T72_A04_17	2	0	0	297	99.33
T72_A05_15	1	0	0	273	99.64
T72_A05_17	0	0	0	299	100.00
T72_A07_15	0	0	0	274	100.00
T72_A07_17	0	0	0	299	100.00
T72_A10_15	0	0	0	271	100.00
T72_A10_17	0	0	0	296	100.00
	BMP2	96.27	T72	99.78	98.94

식별 확률이 90.67 %에 불과하여 전체적으로 97.42 %의 식별 성능이 나타난다. SE 채널집중 모듈이 추가된 Net 2는 BMP2 표적의 식별 성능이 개선되어 98.66 %의 전체 식별 확률을 나타냄을 알 수 있다. 그에 비해 Table 9에 나타난 채널방향 차원축소가 없는 채널집중 모듈에 의한 Net 4는 BMP2 표적에 대한 식별 확률은 96.62 %로 보다 개선되었고, 모든 T72 표적을 올바르게 식별함으로써 99.19 %의 성능을 낸다. 이로부터 채널차원 축소가 없는 채널집중 모듈의 효과가 EOC-V에 해당하는 변형표적 식별능력에 긍정적으로 작용하였음을 다시 한 번 확인할 수 있다. 이는 보다 복잡한 구조를 갖는 CBAM이 적용된 Net 3보다도 0.25 % 개선된 성능이다. 표적 별 결과를 살펴보면, T72 표적의 경우 잘못 식별된 시험영상은 주로 BMP2 표적으로 예측된 것을 알 수 있으며, 그 수량이 EOC-C 실험에 비해서는 적기 때문에 EOC-V에 쓰인 시리얼 번호의 T72 표적영상들이 네트워크 학습에 쓰인 시리얼 번호 132인 T72 표적영상과 상대적으로 유사함을 알 수 있다. BMP2 표적의 경우에는 반대로 T72 표적과 혼동하는 비율이 높으며, 보병 수송차량인 BTR70이나 BRDM2에 대해서는 모든 네트워크가 비교적 강인한

변별력을 가짐을 알 수 있다. 이렇게 3장에 제시된 MSTAR 표적영상의 EOC-C 및 EOC-V 실험 결과는 본 논문에서 제안한 차원축소 없는 채널집중 모듈이 사용된 네트워크의 변형표적 식별 능력을 입증한다.

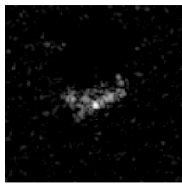
4. 결과에 대한 고찰

4장에서는 네트워크에서 도출된 특징맵으로부터의 채널 활성화 맵을 이용하여 3장에 기술된 실험결과에 대해 고찰한다. 지면 관계상 모든 표적영상에 대해 채널 활성화 맵을 제시할 수는 없으므로, EOC-C 및 EOC-V에서 가장 낮은 식별 확률을 나타낸 표적에 대해 대표적인 영상 각 1장에 대해 결과를 분석한다. 채널 활성화 맵은 주로 특징의 추상화가 덜 된 낮은 계층(본 논문에서는 ‘stage 2’)에서 도출된 특징맵으로부터 관측하기 때문에, 네트워크가 최종적으로 예측한 부류(class)와 연관성이 다소 적을 수 있으나, 네트워크가 결과를 도출하는 데 있어서 특징맵의 채널을 어떻게 활용하였는지를 직관적으로 보여주는 도구이다.

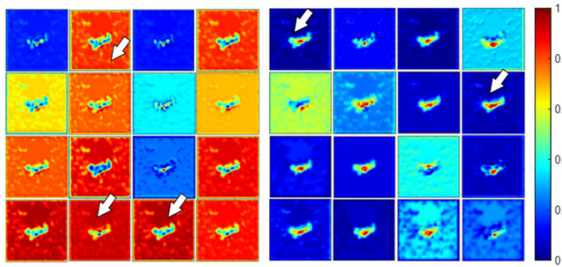
먼저 Fig. 6에는 EOC-C에서 4종의 네트워크 구조 공통적으로 가장 낮은 성능을 나타낸 시리얼 번호 S7인 T72 표적영상 중 방위각이 61도인 영상에 대해 네트워크의 ‘stage 2’ 출력 특징맵의 채널 별 활성화 맵을 나타내었다. 이때 출력 특징맵의 크기는 $64 \times 64 \times 64$ 이며, 64개의 채널 활성화 맵 중 크기가 큰 16개만 선별하여 Fig. 6에 나타내었다. Fig. 6(a)은 표적영상을, Figs. 6(b)-(e)는 각각 Net 1-4의 ‘stage 2’ 출력 특징맵에서 선별한 채널 별 활성화 맵이다. 표적영상을 살펴보면 대체적으로 표적 부분 외의 배경 클러터(background clutter) 레벨이 낮고, 표적의 좌측 면에 강한 산란점이 나타난다. Fig. 6(b)에 나타난 Net 1에 의한 활성화 맵을 관측하면 주로 배경 클러터 부분에 강한 활성화 신호가 형성되는 것으로 미루어 볼 때, 네트워크가 표적 부분 보다는 배경 부분에 더 집중하여 학습했다는 것을 알 수 있다(화살표 참조). 이는 다른 MSTAR 표적영상 식별 관련 연구사례^{19,20)}에서 지적된 바와 같이 MSTAR 표적영상의 높은 배경 클러터 상관도(background correlation)를 보여주는 것으로서, Net 1이 EOC-C 실험에서 95 %의 성능을 도출한 것에 배경 클러터의 기여가 상당할 수 있음을 보여주는 것이다. SE 채널집중 모듈이 적용된 Net 2의 결과인 Fig. 6(c)에는 이러한 문제가 상당히 완화되어, 네

트워크가 보다 표적부분에 집중하여 학습을 수행한 것을 알 수 있다(화살표 참조). 일부 활성화 맵에서는 클러터의 비중이 높은 편이나, 16개의 맵 모두 표적 부분의 강한 산란점에 집중하여 활성화 양상이 형성되었음을 알 수 있다. 이는 네트워크가 채널집중 작용에 의해 보다 표적 부분의 산란특성을 충실히 학습하게 됨으로써, 98.52 %의 개선된 성능으로 이어졌음을 예상할 수 있다. 그러나 마치 표적 부분에 평균 필터를 적용한 것과 같이 평탄화(smoothing)되어, 표적 부분 내의 강한 부분 외에 산란점 구성의 세밀함이 다소 떨어진다. 반면에 SE 모듈에 공간방향 집중 효과가 더해진 CBAM은 Fig. 6(d)와 같이 표적영상을 구성하는 산란점의 세밀한 분포를 보다 잘 나타냄을 알 수 있다(화살표 참조). 그러나 추가적인 공간방향 집중 작용은 배경 클러터로의 네트워크 학습이 집중되는 부작용도 일부 야기하여 Fig. 6(b)에 나타난 문제점이 관측되는 채널 활성화 맵도 일부 존재함을 알 수 있다. EOC-C 실험에서 Net 3의 높은 성능 99.70 %는 표적 부분에 대한 세밀함의 개선과 함께 배경 클러터가 일부 기여하였다고 추정할 수 있다. Fig. 6(e)에 나타난 채널방향 차원축소를 거치지 않는 채널집중 네트워크 Net 4에 의한 결과는 Net 2와 Net 3의 장점이 모두 나타나고(화살표 참조), 배경 클러터 강조나 표적 부분의 평탄화와 같은 단점은 완화되었음을 알 수 있다. Fig. 7에는 Fig. 6에 대한 그라디언트 부류 활성화 맵(Gradient Class Activation Map, Grad-CAM)을 네트워크 별로 나타내었다. Net 1, Net 2, Net 3에 비해서 Net 4의 결과가 표적의 좌측 산란점에 부류 활성화의 중심이 위치한 것을 알 수 있으며, 표적 중심영역에 대한 집중도도 높은 것을 알 수 있다.

Fig. 8에는 EOC-V에서 시리얼 번호 9566인 BMP2 표적영상 중 방위각이 202도인 영상에 대해 네트워크의 ‘stage 2’ 출력 특징맵의 채널 별 활성화 맵을 나타내었다. Fig. 8(a)의 표적영상을 살펴보면, 상대적으로 배경 클러터의 레벨이 높고, 표적의 그림자가 12시 방향으로 뚜렷하게 형성되었음을 알 수 있다. Fig. 8(b)의 Net 1에 의한 활성화 맵을 관측하면, 주로 중심의 표적 영역에 활성화 양상이 집중되어 있음을 알 수 있으나, 일부 맵의 경우에는 배경 클러터 및 그림자 위쪽의 표적 영역과 무관한 부분에 학습이 집중된 것을 알 수 있다(화살표 참조). Fig. 8(c) 및 8(d)의 Net 2, Net 3에 의한 활성화 맵에서는 일부 클러터 활성화 집중도가 높은 맵을 제외하면 중심의 영역에 학습의

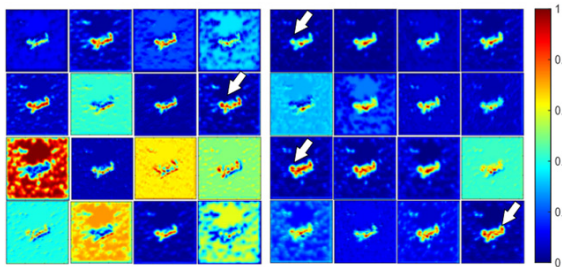


(a)



(b)

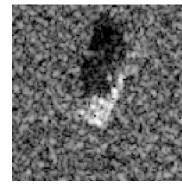
(c)



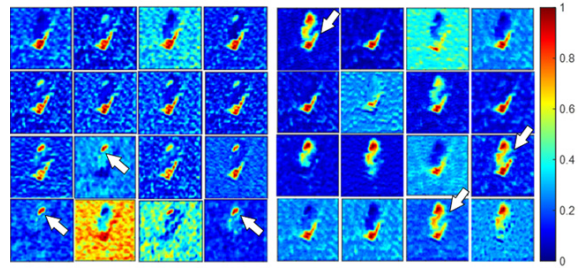
(d)

(e)

Fig. 6. (a) SAR target image of T72(S/N:S7) and channel activation maps from 'stage 2' of (b) Net 1, (c) Net 2, (d) Net 3, (e) Net 4

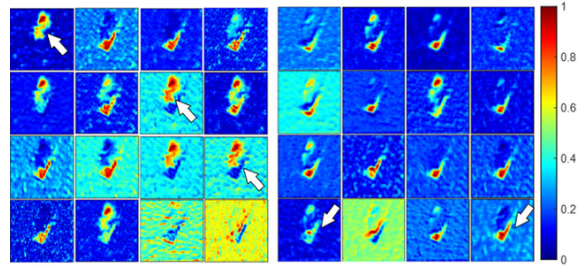


(a)



(b)

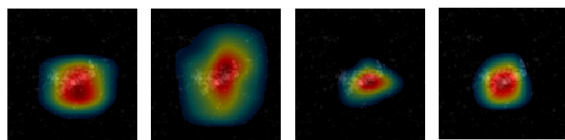
(c)



(d)

(e)

Fig. 8. (a) SAR target image of BMP2(S/N:9566) and channel activation maps from 'stage 2' of (b) Net 1, (c) Net 2, (d) Net 3, (e) Net 4



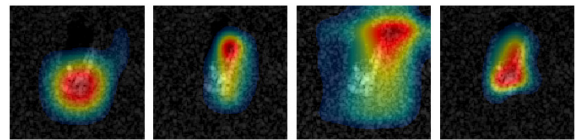
(a)

(b)

(c)

(d)

Fig. 7. Gradient Class Activation Maps for Fig. 6 in case of (a) Net 1, (b) Net 2, (c) Net 3, (d) Net 4



(a)

(b)

(c)

(d)

Fig. 9. Gradient Class Activation Maps for Fig. 8 in case of (a) Net 1, (b) Net 2, (c) Net 3, (d) Net 4

기여도가 높다는 것을 알 수 있다. 그러나 화살표로 표시한 것과 같이, 중심 영역 중에서 표적 영역뿐만 아니라, 그림자 영역에 의한 기여가 크다. 이는 SE 모듈이나 CBAM에 의해 네트워크가 그림자 영역을 학

습에 더 많이 이용하였다고 추정할 수 있다. SAR 표적영상에 형성된 그림자의 경우도 표적식별의 주요 특징으로 삼을 수 있으나, 표적 영역에 비해서는 선명도가 떨어지며 일부 SAR 영상에서는 그림자 영역에

대한 식별이 불가능하다는 한계가 있다. Fig. 8(e)의 나타난 채널방향 차원축소를 거치지 않는 Net 4에 의한 결과는 다른 활성화 맵에 비하여 표적 영역에 대한 학습의 집중도가 높으며, 이것이 높은 성능으로 이어지는 요인 중 하나라고 생각할 수 있다. Fig. 9에는 Fig. 8에 대한 그라디언트 부류 활성화 맵을 나타내었다. 집중작용이 없는 Fig. 9(a)의 경우 활성화 중심이 표적 후미에 집중되나, SE 모듈과 CBAM에 의한 결과는 Fig. 8에서 관측한 바와 같이 표적의 그림자 쪽 부분에 활성화 중심이 이동한 것으로 알 수 있다. 그러나 ECA에 의한 Fig. 8(d)은 상대적으로 표적 부분에서 부류 활성화 집중도가 더 높다.

5. 결론

본 논문에서는 채널방향의 차원 축소 과정이 없는 채널집중 모듈을 딥러닝 네트워크에 도입하여 SAR 변형표적 식별 문제에 적용하였다. 차용한 채널집중 모듈의 성능 개선에 대한 기본 가정을 바탕으로 MSTAR SAR 변형표적 영상에 적용한 결과, 기본 네트워크는 물론 기존의 채널집중 모듈을 적용한 네트워크보다도 효율적이면서 우수한 성능을 보였다. 또한 채널 활성화 맵을 통하여 제안된 채널 집중 모듈의 성능 향상에 미치는 영향을 직관적으로 확인하였다. 본 논문에서는 SAR-ATR 연구에서 널리 활용되는 MSTAR 데이터 셋으로부터 식별성능을 산출/분석함으로써 본 논문에서 제안된 기법의 유효성을 입증하였다. 그러나 MSTAR 데이터 셋은 표적영상의 중심에 항상 표적이 단독으로 위치하고 있는 형태이고, 데이터 셋 자체가 통제된 조건 하에 획득되었으므로 클러스터 형태 또한 일정하다. 따라서 MSTAR에서 접하지 못한 다양한 운용조건을 갖는 데이터 셋에 본 논문의 기법을 추가 적용함으로써 차원축소가 없는 채널집중 작용의 효과를 보다 면밀히 검증할 수 있을 것이라 판단된다.

References

- [1] L. M. Novak. et al., "The Automatic Target-Recognition System in SAIP," *Linc. Lab. J.*, Vol. 10, No. 2, pp. 187-202, 1997.
- [2] R. Hummel, "Model-based ATR Using Synthetic Aperture Radar," in *Proc. IEEE Radar Conf.*, Alexandria, VA, USA, 2000.
- [3] Q. Zhao., et al., "Support Vector Machines for SAR Automatic Target Recognition," *IEEE Trans. Aerosp. Electron. Syst.*, Vol. 37, No. 2, pp. 643-654, Apr. 2001.
- [4] Y. J. Sun, et al., "Adaptive Boosting for SAR Automatic Target Recognition," *IEEE Trans. Aerosp. Electron. Syst.*, Vol. 43, No. 1, pp. 112-125, Jan. 2007.
- [5] M. David, "Deep Convolutional Neural Networks for ATR from SAR Imagery," *Proc. SPIE, Algorithms for Synthetic Aperture Radar Imagery XXII*, 2015.
- [6] O. Kechagias-Stamatis, et al., "Fusing Deep Learning and Sparse Coding for SAR ATR," *IEEE Trans. Aerosp. Electron. Syst.*, Vol. 55, No. 2, pp. 785-797, Apr. 2019.
- [7] S. Chen, et al., "Target Classification Using the Deep Convolutional Networks for SAR Images," *IEEE Trans. Geoscience and Remote Sensing*, Vol. 54, No. 8, pp. 4806-4817, 2016.
- [8] Z. Lin, et al., "Deep Convolutional Highway Unit Network for SAR Target Classification with Limited Labeled Training Data," *IEEE Geosci. Remote Sens. Lett.*, Vol. 14, No. 7, pp. 1091-1095, Jul. 2017.
- [9] F. Zhou, et al., "SAR ATR of Ground Vehicles based on LM-BM-CNN," *IEEE Trans. Geosci. Remote Sens.*, Vol. 56, No. 12, pp. 7282-7293, Dec. 2018.
- [10] R. Xue, et al., "Spatial-Temporal Ensemble Convolution for Sequence SAR Target Classification," *IEEE Trans. Geosci. Remote Sens.*, Vol. 59, No. 2, pp. 1250-1262, Feb. 2021.
- [11] Z. Cui, et al., "Image Data Augmentation for SAR Sensor via Generative Adversarial Nets," *IEEE Access*, Vol. 7, pp. 42255-42268, 2019.
- [12] Q. Yu, et al., "High-Performance SAR Automatic Target Recognition Under Limited Data Condition based on a Deep Feature Fusion Network," *IEEE Access*, Vol. 7, pp. 165646-165658, 2019.
- [13] J. Hu, et al., "Squeeze-and-Excitation Networks," in

- Proc. IEEE Comput. Soc. Conf. Comp. Vis. Pattern Recognit., Salt Lake City, UT, USA, 2018.
- [14] L. Wang, et al., "SAR ATR of Ground Vehicles based on ESEnet," *Remote Sens.*, Vol. 11, No. 1316, pp. 1-16, 2019.
- [15] J. H. Park, et al., "SAR ATR for Limited Training Data Using DS-AE Network," *Sensors*, Vol. 21, No. 4358, pp. 1-19, 2021.
- [16] M. Zhang, et al., "Convolutional Nerual Network with Attention Mechanism for SAR Automatic Target Recognition," *IEEE Geosci. Remote Sens. Lett.*, Early Access. 2020.
- [17] Q. Wang, et al., "ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks," in *Proc. IEEE Comput. Soc. Conf. Comp. Vis. Pattern Recognit.*, Virtual Conference, 2020.
- [18] K. He, et al., "Deep Residual Learning for Image Recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.(CVPR)*, pp. 770-778, Jun. 2016.
- [19] S. Woo, et al., "CBAM: Convolutional Block Attention Module," in *Proc. Eur. Conf. Comput. Vis.*, pp. 3-19, 2018.
- [20] O. Kechagias-Stamatis, et al., "SAR ATR by a Combination of Convolutional Neural Network and Support Vector Machines," *IEEE Aerosp. Electron. Syst. Mag.*, pp. 56-81, Mar. 2021.