합성곱 신경망의 Channel Attention 모듈 및 제한적인 각도 다양성 조건에서의 SAR 표적영상 식별로의 적용

박지훈^{*,1)} · 서승모¹⁾ · 유지희¹⁾

¹⁾ 국방과학연구소 제3기술연구본부

Channel Attention Module in Convolutional Neural Network and Its Application to SAR Target Recognition Under Limited Angular Diversity Condition

Ji-Hoon Park^{*,1)} · Seung-Mo Seo¹⁾ · Ji Hee Yoo¹⁾

¹⁾ The 3rd Research and Development Institute, Agency for Defense Development, Korea

(Received 4 September 2020 / Revised 5 January 2021 / Accepted 14 January 2021)

Abstract

In the field of automatic target recognition(ATR) with synthetic aperture radar(SAR) imagery, it is usually impractical to obtain SAR target images covering a full range of aspect views. When the database consists of SAR target images with limited angular diversity, it can lead to performance degradation of the SAR-ATR system. To address this problem, this paper proposes a deep learning-based method where channel attention modules(CAMs) are inserted to a convolutional neural network(CNN). Motivated by the idea of the squeeze-and-excitation(SE) network, the CAM is considered to help improve recognition performance by selectively emphasizing discriminative features and suppressing ones with less information. After testing various CAM types included in the ResNet18-type base network, the SE CAM and its modified forms are applied to SAR target recognition using MSTAR dataset with different reduction ratios in order to validate recognition performance improvement under the limited angular diversity condition.

Key Words : Synthetic Aperture Radar(합성 개구면 레이더), Automatic Target Recognition(자동 표적 식별), Convolutional Neural Network(합성곱 신경망), Channel Attention Module(채널 특징 집중 모듈)

1. 서론

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

전파 송수신에 의해 형성되는 합성 개구면 레이더(Synthetic Aperture Radar, SAR)에서 획득된 영상은 광학영상과는 달리 전천후의 특성을 가지므로 군사나 민간 분야에서 감시 및 경찰의 목적으로 널리 활용 중

에 있으며, 이러한 SAR 영상으로부터 표적을 자동으로 식별(SAR-Automatic Target Recognition, SAR-ATR)하는 기법 또한 지난 30여 년 동안 중요한 연구주제로 다루어졌다. SAR-ATR을 위한 표적식별 기법으로는 크게 특징기반(feature-based)기법, 모델기반(model-based)기법, 반 모델기반(semi model-based)기법으로 분류되며^[1], 이 중 특징기반 기법의 경우 전통적인 템플릿 정합(template matching)과 함께 연구자에 의해 제작 및 선정된(hand-crafted) 특징벡터로 패턴인식 기반의 분류기를 훈련시킨 후 표적영상을 식별하는 방식이 대부분이었다. 그러나 최근 5년간에는 영상의 특징을 자동으로 학습하여 광학영상 분류에서 큰 성공을 거둔 합성곱 신경망(Convolutional Neural Network, CNN) 등의 딥러닝(deep learning)기반 기법이 SAR 표적 식별에 적용되어 성능을 크게 향상시켰으며, 현재 이 분야의 대세를 이루고 있다. M. David의 초기연구^[2]를 필두로, J. Ding 등은 SAR 표적영상의 합성곱 신경망 학습을 위한 데이터 증대기법을 제안하였으며^[3], S. Chen 등은 합성곱 신경망 종단의 완전 연결층(Fully Connected layer, FC)을 제거한 네트워크 구조를 제안하여 MSTAR 표적 식별성능 99 % 이상을 합성곱 신경망을 통해 최초로 달성하였다^[4]. 이후에는 합성곱 신경망에서 추출한 특징을 기존 분류기와 융합하거나 합성곱 신경망 내의 특징맵(feature map) 간 융합을 수행하여 표적 식별성능을 높이는 연구^[5-8], 광학영상 식별분야에서 도출된 최신 구조의 네트워크를 차용한 연구^[9-12], 여러 각도의 SAR 표적영상으로부터 멀티뷰(multi-view) 특징을 추출하여 식별성능을 개선한 연구^[13-15] 등이 수행되었다. 이들 연구에서는 모두 미 공군 연구소에서 수집하여 공개한 MSTAR SAR 표적영상 데이터 셋을 사용하였으며, S. Chen 등의 연구 이후에는 대부분의 연구사례에서 MSTAR 표준 운용조건(Standard Operating Condition, SOC)의 10개 표적에 대해 99 % 이상의 식별성능을 달성하였다.

전자파의 산란에 의해 형성되는 SAR 영상의 특성상 표적의 측면 각도에 따라 나타나는 영상의 특성이 매우 달라진다. 따라서 높은 표적 식별성능을 갖는 SAR-ATR 시스템을 구축하기 위해서는 가급적 0°부터 360°까지의 다양한 측면 각도를 포괄적으로 갖는 표적영상이 필요하다. 그러나 충분한 SAR 표적영상을 획득하기 위해서는 시간과 비용이 많이 드는 항공 또는 위성 SAR 촬영이 다수 필요하고, 군사용으로 많이 사용되는 SAR 특성 상 공개된 데이터 셋도 부족한

실정이다. 따라서 SAR-ATR용 네트워크를 학습시키기 위한 SAR 표적영상의 측면각도가 불충분한, 즉 제한적인 각도 다양성(limited angular diversity) 조건하에 놓일 가능성이 매우 높다. 앞서 언급한 선행연구와 그 외 딥러닝 기반 SAR-ATR 연구에서는 대부분 각도 다양성이 비교적 높은 (표적영상의 평균 각도간격 1° 내외) MSTAR 10개 표적에 대하여 100 %의 식별성능에 최대한 근접(99.86^[6], 99.90 %^[13] 등)하기 위한 네트워크 성능 향상 연구에 집중하였다. 그러나 2017년부터 일부 연구에서는 네트워크 학습을 위한 MSTAR 표적영상의 수량을 줄이면서 제한적인 각도 다양성 조건하의 식별성능을 연구하였다. Z. Lin 등은 네트워크 분기가 가변적으로 설정되는 ‘highway’ 네트워크 모듈을 합성곱 신경망에 도입하였으나 전체 수량 30 % 미만의 데이터로 학습 시에는 식별성능이 크게 저하되었다^[9]. F. Zhang 등은 여러 표적영상으로 만든 멀티뷰 시퀀스를 순환 신경망(Recurrent Neural Network, RNN)에 입력하여 전체 수량의 25 %인 학습용 표적영상을 이용해서도 95 % 이상의 식별성능을 얻었고^[13], J. H. Cho 등과 Q. Yu 등은 다중 스케일의 특징을 각자 고안한 방식으로 융합하여 20 %의 학습영상으로도 92 % 내외의 성능을 얻었다^[7,8]. 즉 이들은 네트워크 구조나, 합성곱 신경망 도출특징의 융합, SAR 표적영상의 특징을 멀티뷰 형태로 증대하는 방식 등으로 성능을 개선하였다.

본 논문에서는 합성곱 신경망에 채널방향 집중 모듈(Channel Attention Module, CAM)을 삽입하여 제한적인 각도 다양성 조건 하의 SAR 표적 식별에 적용한다. 이는 ‘Squeeze-and-Excitation(SE)’이라는 대표적인 채널방향 집중 모듈을 제안하여 ILSVRC 2017의 ImageNET 영상 식별분야에서 우승한 SE Net^[16]의 아이디어에서 출발한다. 즉 채널방향 집중 모듈을 통해 합성곱 신경망에서 도출된 특징맵의 채널 간 상호 의존성(interdependencies)을 명시적으로 모델링한 결과를 바탕으로 유용한 특징은 강조하고 덜 유용한 특징은 억제함으로써, 영상 식별의 변별력을 높이게 된다. 기존에도 채널방향 집중 모듈 구조를 활용하여 SAR-ATR 연구에 적용한 연구사례는 있었으나^[11,12], 제한적인 각도 다양성 조건에서 적용하지는 않았다. 이중 G. Huang 등의 연구^[12]에서는 전역평균풀링(Global Average Pooling, GAP)에 의해 C×1 벡터로 줄어든 특징맵을 그룹 별로 묶어 완전 연결층에 연결함으로써 압축(Squeeze) 연산을 수행하였으며, 채널방향 집중 모듈 및 제안된 밀집

(dense) 네트워크 구조에 의해 99.79 %의 높은 식별 성능을 얻었다. 그러나 그룹 별 연산은 채널 간 정보흐름을 차단하는 결과를 낳기 때문에 그룹 연산으로 손실된 정보를 다른 연산으로 보강해야 하는 한계가 있다. L. Wang 등의 연구^[11]에서는 순수하게 SE 채널방향 집중 모듈의 세부 구조를 일부 변경한 형태의 Enhanced SE(ESE) 채널방향 집중 모듈을 제안하였다. 채널방향 집중 모듈의 개념을 본 논문의 문제에 접목하여 생각한다면, 채널방향 집중 모듈에 의해 합성곱 신경망의 표현력(representation power)을 강화하여 학습의 효율성을 높임으로써, 적은 학습영상으로도 높은 성능을 달성할 수 있으리라는 것이 본 논문의 핵심 추론이다. 또한 네트워크 구조 관점에서 제한적인 각도 다양성 조건하의 SAR 식별성능을 개선한다는 의미가 있다. 2장에서는 채널방향 집중 모듈이 삽입될 레즈넷18(ResNet18)^[17,18] 형태의 기본 네트워크 및 SE 채널방향 집중 모듈에 대해 다룬다. 3장에서는 다양한 형태로 구성된 채널방향 집중 모듈에 대하여 MSTAR 학습 데이터 셋의 25% 만으로 학습시킨 네트워크의 성능을 비교하고, 4장에서는 MSTAR 데이터 셋의 네트워크 학습 이용비율을 변경하여 3장에서 제안된 네트워크 구조의 식별성능을 비교한다. 5장에서는 결론을 맺는다.

2. 기본 네트워크 및 SE 채널방향 집중 모듈

2.1 레즈넷18 형태의 기본 네트워크

본 논문에서 채널방향 집중 모듈을 삽입할 합성곱 신경망은 K. He 등에 의해 제안된 18층 깊이의 ‘레즈넷18’이다. 레즈넷은 입출력을 바로 연결하는 ‘스킵(skip) 연결’이 주요 특징이며^[17,18], 본 논문에서는 가장 간단한 레즈넷의 형태인 레즈넷18을 차용한다. SAR 표적영상 식별에 레즈넷18을 적용하기 위하여 가로×세로 픽셀 개수 128×128과 채널 1개인 SAR 표적영상의 크기, 그리고 분류 개수(10개) 등의 조건을 고려하여 세부 구조에 변화를 주었다. Fig. 1에는 레즈넷18의 전체구조(위) 및 구성모듈의 세부 구조(아래)를 나타내었다.

128×128×1 크기의 입력 표적 SAR 영상은 ‘stage 1’에서 64×64 크기 및 채널 개수 64개의 특징맵으로 변환되고, ‘stage 2’에서도 동일한 크기의 특징맵에 대한 합성곱, 배치(batch) 정규화 등이 수행된다. 이후 ‘stage

3,4,5’에서는 특징맵의 가로×세로 크기가 1/2씩 줄어들면서 채널개수는 2배 증가한다. 이후 stage 5의 8×8×512 특징맵에 대해 전역평균풀링을 거쳐 채널수와 동일한 512×1 벡터가 형성되고, 10개의 뉴런(neuron, node)을 갖는 완전 연결층을 거쳐 10종의 소프트맥스(softmax) 분류기를 거친다. 각 ‘stage’는 2개의 ‘Residual 모듈’로 구성되고 각 모듈은 2층의 3×3 합성곱 및 배치 정규화와 ReLU 활성화 함수로 이루어지며, 여기서 연산된 결과와 입력을 스킵 연결로 더해준다(element-wise addition). 또한 ‘addition’ 경로에서 ‘pre-activation’ 구조^[18]를 갖는데, 동일 특징맵 매핑(identity mapping)의 목적보다는 추후 채널방향 집중 모듈을 용이하게 삽입하기 위함이다. Fig. 1 아래의 좌측은 ‘stage 2’의 모듈 및 ‘stage 3,4,5’의 두 번째 모듈의 구조를 나타내고, Fig. 1 아래의 우측은 ‘stage 3,4,5’의 첫 번째 모듈의 구조를 나타낸다. 우측 모듈의 경우, 특징맵의 가로×세로 크기를 절반으로 줄이고 입출력을 더하는 과정에서 차원을 일치시키기 위하여 스킵 연결에 stride 2인 1×1 합성곱을 위치시킨다. 여기서 stride는 합성곱 필터의 계산 보폭이며, ‘pad = same’은 합성곱 연산 결과 특징맵의 가로×세로 크기가 이전 크기의 1/stride 배가 되도록 입력 특징맵 둘레에 가변적으로 0을 위치(zero-padding)시킨다.

Fig. 1. Structures of ResNet18

2.2 Squeeze-and-Excitation(SE) 모듈

SE^[16]는 합성곱 신경망에서 도출된 특징맵에서 채널 방향 특징의 선택적인 강조 및 억제를 수행하는 대표적인 채널방향 집중 모듈로서, 기존의 네트워크에 삽입되는 형태를 갖는다. Fig. 2에는 이러한 SE 채널방향 집중 모듈의 구조를 나타내었다.

Fig. 2. Structure of SE-type module

먼저 $M \times N \times C$ 크기의 특징맵이 입력(CAM input)되면, 각 채널의 정보를 표현(일종의 encoding)하기 위한 전역평균풀링을 수행하여 $C \times 1$ 크기의 벡터로 변환시킨다. 이를 C/r 개의 뉴런을 갖는 완전 연결층에 연결하는데, r 은 통상 16으로 설정된다. 이렇게 채널정보가 C 에서 C/r 로 줄어드는 압축(Squeeze)이 수행된다. 이후 ReLU 활성화를 거쳐 원래 채널수인 C 개 뉴런을 갖는 완전 연결층에 연결하여 복원하고, 0에서 1의 값을 갖는 시그모이드(Sigmoid)함수에 의한 활성화를 통해 채널 간의 상대적인 중요도를 모델링한다. 이렇게 채널 간의 상호 의존성이 학습에 의해 도출된 결과를 채널방향 집중 모듈 입력 특징맵의 채널방향(channel-wise)으로 곱하는 자극(Excitation)을 통해, 영상 식별에 중요한 채널의 정보는 강조하고 덜 중요한 채널의 정보는 억제하는 효과를 얻을 수 있다. 즉 중요한 특징을 갖는 채널에 네트워크 학습의 주의를 집중시키는 작용이 이루어진다. 이는 주어진 네트워크 학습용 영상의 특징을 보다 효율적으로 사용하게 된다는 의미이므로, 본 논문에서 다루는 제한적인 각도 다양성 조건에서의 SAR 표적영상 식별문제에 효과적일 것임을 추론할 수 있다.

본 논문에서는 SE 모듈을 기본 네트워크인 레즈넷 18에 Fig. 1의 위쪽 그림과 같이 ‘CAM’으로 표시된 곳

에 삽입한다. 각 레지듀얼(Residual) 모듈에는 Fig. 1의 아래쪽 ‘CAM to be inserted’와 같이 스킵 연결 맞은편에 삽입된다. 또한 SE 채널방향 집중 모듈 연구결과^[16]에 따르면, 채널 개수가 많은 마지막 ‘stage 5’에는 채널방향 집중 모듈 내의 많은 완전 연결층의 뉴런에 비해 삽입에 의한 성능향상은 미미한 것으로 알려져 있으므로 채널방향 집중 모듈을 삽입하지 않는다.

3. 다양한 형태의 모듈에 대한 ablation 실험

3.1 MSTAR SAR 표적영상 데이터 셋

여러 형태의 채널방향 집중 모듈을 제시하기에 앞서 실험에 이용되는 MSTAR 표적영상 데이터 셋에 대해 소개한다.

Table 1. Number of target images in MSTAR dataset

표적명 (class)	일련번호 (serial #)	학습영상 수 (내림각 17°)	시험영상 수 (내림각 15°)
BMP-2	9563	233	196
BTR-70	c71	233	196
T-72	132	232	196
BTR-60	k10yt7532	256	195
2S1	b01	299	274
BRDM-2	E-71	299	274
D7	92v13015	299	274
T-62	A51	299	273
ZIL-131	E12	299	274
ZSU-234	D08	299	274

MSTAR 표적영상 데이터 셋은 미 공군 연구소(Air Force Research Laboratory, AFRL) 주도의 SAR-ATR 연구 프로젝트에서 획득된 공개용 표적 SAR 영상의 데이터 셋으로서, 본 논문에서는 표준 운용조건(SOC)의 MSTAR 데이터 셋만 사용한다. 네트워크를 학습시키기 위한 영상은 거의 모든 문헌에서 수행하는 바와 같이 내림각 17°인 표적영상을 사용하고, 네트워크의 식별성능을 측정하기 위한 시험영상은 내림각 15°인 표적영상을 사용한다. Table 1에는 MSTAR 데이터 셋

의 표적 종류 별 학습/시험영상의 개수를 나타내었고, Fig. 3에는 10종 표적의 사진 및 그에 해당하는 SAR 표적영상의 예시를 나타내었다. 주어진 MSTAR 데이터 셋의 네트워크 학습 영상은 0°부터 360°의 다양한 측면 각도의 영상을 포함하고 있다. 평균 각도 간격은 약 1.3°로서 충분한 SAR 표적영상을 위해 다수의 항공촬영이 필요하다는 사실을 감안할 때, 각도 다양성이 비교적 높다고 할 수 있다. 이전 딥러닝 기반 SAR-ATR 연구에서 달성한 99 % 이상의 식별성능은 Table 1의 학습영상을 모두 활용하여 획득한 결과이다 [4-6,7-9,11-14].

Fig. 3. Optical and SAR images MSTAR 10-target

3.2 다양한 형태의 채널방향 집중 모듈 구조

2장에서 설명한 SE 채널방향 집중 모듈 및 관련 선행연구를 기반으로, 본 논문에서는 보다 다양한 형태의 채널방향 집중 모듈 구조를 추가로 제시하여 MSTAR 데이터에 대한 SAR 표적영상 식별에 적용한다. 이렇게 제시된 채널방향 집중 모듈 각각을 기본 네트워크인 레즈넷18에 삽입하면서 그에 따른 식별 성능(식별확률)을 산출한다. 이와 같이 다양한 구조에 따라 성능을 산출하고 상호 비교하는 것을 애블레이션(ablation) 실험이라고 한다.

Fig. 4에는 본 논문에서 고안되어 사용된 다양한 형태의 채널방향 집중 모듈을 나타내었다. ‘type A’의 경우 SE 채널방향 집중 모듈의 전역평균풀링을 ‘합성곱과 배치 정규화+ReLU 활성화’로 교체한 형태를 갖는다. 단순히 전역평균풀링의 평균 연산만으로 채널 정보를 표현하는 것보다는 특징맵의 3차원 크기와 동일한 크기의 필터를 채널의 개수만큼 적용하는 합성

Fig. 4. Various forms of CAMs used in this paper

곱 연산이 보다 정확한 채널정보를 표현할 것이라는 가정에서 고안되었다. 그러나 특징맵과 동일한 크기를 갖는 필터의 매개변수로 인해 네트워크 전체 용량이 크게 증가한다. ‘type B’의 경우 SE 채널방향 집중 모듈의 시그모이드($= e^x/(1+e^x)$) 대신 학습가능 매개변수 a, b 가 추가되는 ‘매개변수화(Parametric) 시그모이드($= e^{ax}/(b+e^{ax})$)’를 사용하였으며, 기존 시그모이드의 기울기와 평행 이동 정도에 변화를 줌으로써 채널 중요도 모델링의 적응성(adaptivity)을 높였다. ‘type C’의 경우 SE 모듈 이후에 공간정보 강화 모듈(Spatial Attention Module, SAM)을 추가하였다. 두 가지의 형태 중에서 ‘type C1’은 YOLO v4^[19]에서, ‘type C2’는 CBAM (Convolutional Block Attention Module)^[20]에 사용된 공간정보 강화 모듈이 적용된다. ‘type D’의 경우 SE 모듈과 달리 전역평균풀링 이후 2단의 완전 연결층을 추가 삽입하여 채널정보의 압축 및 복원이 점진적으로 이루어질 수 있도록 하였다. 예를 들어 SE 모듈에서는 64×1 벡터를 $r = 16$ 에 대해 4×1로 압축한 후 다시 복원하였다면, type D에서는 8×1 및 4×1로 압축하고, 역순으로 복원하여 자극(Excitation)에 이용한다. ‘type E’에서는 SE 모듈의 ReLU 대신 Parametric ReLU(PReLU)

를 사용하였다. 즉 활성화 입력이 음수인 경우에 대해서도 학습결과에 의해 0이 아닌 활성화 값을 부여한다. ‘type F’에서는 완전 연결층의 뉴런 개수가 상대적으로 많은 ‘stage 4’에 국한되어 사용되며, (‘stage 5’에서는 앞서 언급한 바와 같이 채널방향 집중 모듈 미사용) 완전 연결층 사이에 50 % 확률로 네트워크 학습 시 뉴런을 비 활성화하는 드롭아웃(drop-out)을 삽입한다. 드롭아웃은 다양한 네트워크의 형태를 학습에 이용하도록 함으로써 과적합을 방지하고 투표효과(voting effect)에 의한 일반화 성능을 향상시키는 것으로 알려져 있다. 다만 뉴런 개수가 적은 ‘stage 1~3’에서는 드롭아웃이 네트워크 학습의 불안정성을 초래하여 발산이 찾아지므로 삽입하지 않는다.

3.3 MSTAR 학습영상 25 % 이용 ablation 실험

앞서 제시된 다양한 형태의 채널방향 집중 모듈을 레즈넷18에 Fig. 1과 같이 삽입하여 Table 1과 Fig. 3에 나타난 10개 표적의 MSTAR 데이터 셋에 대해 식별성능을 산출하고 비교하는 실험을 수행한다. 이 때 MSTAR 학습영상에 대해서는 표적 별로 25 %의 표적 영상만 남김으로써 제한적인 각도 다양성 조건을 조성한다. 이 때 학습에 이용되는 표적영상은 임의로 (randomly) 선정된다. 예를 들어 BMP-2 표적영상은 233개 중 임의의 58개만 네트워크 학습에 사용된다. 본 실험에 사용된 25 % 수량 / 10표적 학습영상의 평균 측면각도 간격은 5.2° 이고, 표준편차는 5.12° 이다. 네트워크 입력영상의 크기는 $128 \times 128 \times 1$ 로 고정되고, NVidia Geforce GTX 1080 GPU 하드웨어 및 매트랩(MATLAB) 딥러닝 툴박스 소프트웨어를 이용하였으며, 학습 단위로 반복되는 미니배치 수는 32이고, 매 세대(epoch) 마다 학습영상을 섞이게(shuffle)하였다. 초기 학습률(learning rate)은 0.001이고 50세대마다 0.5배 감소되도록 설정하였으며, 최대 200세대까지 학습한다. 또한 각도 다양성이 부족해진 학습영상에 대해서 ± 4 픽셀만큼 좌우/상하로 학습영상을 평행이동(translation)하는 데이터 증대기법을 사용하였다. 이때 매트랩 딥러닝 툴박스 특성 상 실제로 증대된 영상을 생성시켜 GPU 메모리에 올리지 않고 매 세대 마다 임의로 평행이동 정도를 변경시키므로 가급적 많은 세대를 설정하여 데이터 증대가 학습에 반영되도록 하였다. 시험영상은 내림각 15° 의 MSTAR 영상 전체를 사용한다. Table 2에는 25 %의 MSTAR 학습영상으로 학습시킨 다양한 네트워크 구조에 대하여 MSTAR 시험영상

에 전체에 대한 식별성능(식별확률) 및 네트워크 학습 시간, 네트워크의 학습 가능한 매개변수 개수(네트워크 크기와 관련)를 나타내었다.

Table 2. Ablation experiments for various networks

네트워크 구조	식별확률	학습시간	매개변수
ResNet18	85.74 %	35분	11.18M
ResNet22	85.99 %	46분	12.65M
ResNet26	90.17 %	56분	17.30M
ResNet34	84.01 %	76분	21.29M
ResNet50	48.48 %	154분	23.54M
ResNet18 + SE	94.30 %	146분	11.20M
ResNet18 + type A	89.13 %	257분	128.64M
ResNet18 + type B	94.55 %	160분	11.20M
ResNet18 + type C1	86.57 %	175분	12.79M
ResNet18 + type C2	92.02 %	184분	11.20M
ResNet18 + type D	94.38 %	171분	11.22M
ResNet18 + type E	90.74 %	172분	11.20M
ResNet18 + type BD	95.41 %	189분	11.23M
ResNet18 + type F	95.74 %	164분	11.20M

- ① 먼저 채널방향 집중 모듈을 적용하지 않았을 경우, 기본 네트워크인 레즈넷18로부터 레지듀얼 모듈을 추가하는 방식으로 네트워크의 깊이를 증가시키면, 26층인 레즈넷26에서 최고의 성능인 90.17 %이 산출되어 기본 네트워크의 성능인 85.74 % 보다 약 4.5 % 향상된다. 그러나 더 깊은 네트워크에 대해서는 성능이 오히려 떨어지는데, 이는 네트워크에서 학습할 수 있는 매개변수의 수가 증가해도 학습영상의 수량은 제한적이므로 과적합이 발생하여 일반화(generalization) 성능이 크게 떨어졌기 때문이다. 이는 네트워크의 깊이를 늘려도 제한적인 각도 다양성 조건과 같이 학습영상의 개수가 부족한 상황에서는 식별성능을 향상시키는데 한계가 있음을 의미한다.
- ② 다음으로 레즈넷18 기본 네트워크에 앞서 언급한 채널방향 집중 모듈을 추가한다. 먼저 SE 모듈을 삽입한 경우, 기본 네트워크의 식별성능보다 8.5 %

향상된 94.30 %의 식별성능을 얻었다. 이는 네트워크의 깊이를 늘린 결과보다 우수한 결과로서, SE 모듈로 인해 제한적으로 주어진 학습영상의 특징을 보다 효율적으로 학습하여 일반화 성능이 개선된 것을 알 수 있다. 뿐만 아니라, SE 모듈 또한 이미 광학영상 연구에서 다양한 구조에 대한 실험 수행의 결과로 도출되었으므로, 성능향상의 효과가 다방면으로 입증되었다고 볼 수 있다.

③ 그 외 Fig. 4에서 제시된 다양한 채널방향 집중 모듈 구조에 대해서는 모두 기본 네트워크의 식별성능을 상회하지만, 향상의 정도가 각기 다르게 나타난다. 먼저 type A의 경우 네트워크 크기 및 학습시간이 크게 증가한 것뿐만 아니라 식별성능 또한 SE 모듈 사용 대비 약 5 % 감소한다. 전역평균풀링을 합성곱으로 대체함에 따라 채널정보와 특징맵의 공간(spatial)정보가 뒤섞이게 되어 채널정보의 표현 정확도가 떨어졌음을 알 수 있다. type B는 채널 중요도를 모델링하는 시그모이드에 적응성을 부여함으로써 성능이 소폭(0.25 %) 증가하였다. type C의 경우 이전연구 결과^[19,20]와는 달리 SAR 표적식별 문제에서는 공간정보 강화 모듈이 채널방향 집중 모듈에 의해 향상된 식별성능을 열화시킴을 확인할 수 있다. 따라서 SAR 표적영상 식별에 적합한 공간정보 강화 모듈 구조를 고안하거나, 채널방향 집중 모듈과 공간정보 강화 모듈의 작용이 보다 독립적으로 수행되도록 조치할 필요가 있다. type D의 경우 매개변수의 개수 및 네트워크 학습시간이 증가한 것에 비해서는 소폭(약 0.2 %)이지만 세분화된 채널정보 압축과정에 의해 성능이 향상된 것을 알 수 있으며 type E의 경우, 매개변수화 시그모이드와는 달리 PReLU가 ReLU보다 3.5 % 가량 더 성능이 낮다. 즉 압축된 채널정보의 음수 값은 ReLU에 의해 0으로 만드는 것이, PReLU에 의해 작은 음수 값으로 만드는 것보다 더 성능이 좋다는 것을 확인할 수 있다.

④ type A-E들을 서로 조합하여 나타나는 성능 변화의 공통점은, SE 모듈에 비해 성능이 저하되는 공간정보 강화 모듈을 조합하면 성능이 향상되지 않고 오히려 더 열화하거나 변화가 없게 된다. 그러나 앞서 SE 공간정보 강화 모듈 대비 성능이 향상된 type B 및 D를 조합한 결과(type BD), 성능향상의 정도가 강화되어 SE 모듈 대비 1.1 % 향상된 95.41 %의 식별성능이 산출된다.

⑤ 추가적으로 type F의 경우, 네트워크 구조의 큰 변화 없이 stage 4에만 드롭아웃을 추가하였음에도 type BD보다 0.3 % 가량 더 높고 SE 모듈 대비 약 1.4 % 상승한 95.74 %의 식별성능이 산출되었다.

4. 여러 각도 다양성 제한 조건하의 표적식별 적용

4.1 여러 각도 다양성 제한 조건에 대한 네트워크 식별성능 비교

다양한 채널방향 집중 모듈 구조에 대한 실험을 바탕으로 MSTAR 학습영상에 대해 여러 각도 다양성 제한조건을 부여하여 네트워크 간의 식별성능을 비교한다. 이를 위해 10 %, 25 %, 50 %, 100 % 비율의 MSTAR 학습용 표적영상을 네트워크 학습에 이용되도록 한다. 각 비율에 대한 표적 별 학습영상의 개수는 다음과 같다. (개수 기재의 순서는 Table 1에서 10종의 표적이 기재된 순서와 동일)

- 100 % : 233/233/232/256/299/299/299/299/299/299
 - 50 % : 117/117/116/128/150/150/150/150/150/150
 - 25 % : 58/58/58/64/75/75/75/75/75/75
 - 10 % : 23/23/23/26/30/30/30/30/30/30

비교대상 네트워크는 총 4종으로서, 기본 네트워크인 레즈넷18과 함께, SE 모듈, type BD 모듈, type F 모듈을 기본 네트워크에 삽입한 세 종류의 네트워크이다. 학습틀이나 세대와 같은 네트워크 학습을 위한 매개변수는 앞선 실험과 동일하며, 이용비율 별 학습영상 데이터 셋에 대한 ± 4 픽셀 평행이동 기반 데이터 증대 또한 동일하게 적용하였다.

Table 3에는 네트워크 학습에 이용된 MSTAR 학습영상의 비율을 위에서 언급한 바와 같이 변화시키면서, 4종의 네트워크가 MSTAR 시험영상을 식별하여 성능을 산출한 결과를 나타내었다. 학습영상이 모두 이용된(100 %) 경우에는 채널방향 집중 모듈이 삽입된 네트워크의 성능이 더 높지만 기본 네트워크인 레즈넷18도 99.01 %의 준수한 성능을 나타내고 있다. 학습에 이용된 영상의 비율이 절반인 50 %로 줄어들면 레즈넷18의 식별성능은 5 % 감소하나, 채널방향 집중 모듈이 삽입된 경우 97 % 이상의 성능을 유지한다. 특히 type BD 모듈이 삽입된 네트워크는 98 % 이상의 성능을 나타낸다. 비율이 25 % 및 10 %인 경우,

표적 별 영상 간 측면각도 간격이 약 5°, 12°로 비교적 크기 때문에 네트워크 학습에 이용되는 영상을 임의로 선정하는 과정에서 매회 학습영상 데이터 셋 세부 구성의 편차가 커질 수 있다. 따라서 채널방향 집중 모듈 삽입에 따른 성능개선 효과 검증의 신뢰도를 높이기 위해 학습영상을 총 3회 임의 선정하여 각각에 대해 네트워크를 학습하고 시험영상에 대한 식별결과를 산출하였다. 각 임의 선정 결과의 10표적 영상에 대한 평균 측면각도 간격 및 표준편차는 다음과 같다.

- 25 %, 1~3회 : 5.21°/5.12°, 5.23°/4.56°, 5.18°/4.81°
- 10 %, 1~3회 : 12.83°/12.65°, 12.83°/13.05°, 12.84°/11.82°

Table 3. Recognition performances of networks for various limited angular diversity conditions

학습이용 영상비율		네트워크 구조			
		ResNet18	ResNet18 + SE	ResNet18 +type BD	ResNet18 + type F
100 %		99.01 %	99.13 %	99.30 %	99.38 %
50 %		94.01 %	97.27 %	98.06 %	97.98 %
25 %	1 st	85.74 %	94.30 %	95.41 %	95.74 %
	2 nd	87.60 %	92.48 %	93.51 %	94.26 %
	3 rd	83.14 %	94.26 %	94.83 %	94.67 %
	Avg.	85.49 %	93.68 %	94.58 %	94.89 %
10 %	1 st	67.93 %	75.43 %	78.93 %	72.73 %
	2 nd	61.78 %	71.16 %	79.50 %	72.27 %
	3 rd	70.37 %	80.12 %	82.19 %	82.11 %
	Avg.	66.69 %	75.57 %	80.21 %	75.70 %

학습영상 비율이 25 %인 경우, 기본 네트워크는 평균 85.49 %의 식별성능을 보이지만, 채널방향 집중 모듈이 삽입되면 94 % 전후의 식별성능을 보인다. 10 %인 경우에는 기본 네트워크의 성능은 66.69 %이지만, 채널방향 집중 모듈이 삽입된 네트워크는 75 % 이상의 성능을 보이고, type BD 모듈은 평균 80.21 %로써 가장 안정적인 성능을 보인다. 특히 모든 학습영상 이용비율에 대해, 본 논문에서 고안한 두 채널방향 집중 모듈을 삽입한 결과가 SE 모듈을 삽입한 결과에 비해 개선된 성능을 보임을 알 수 있다.

4.2 채널방향 집중 모듈에 의한 식별성능 개선확인

4.2.1 공간(spatial)방향의 확인

채널방향 집중 모듈을 삽입한 네트워크가 제한적인 각도 다양성 조건 하에서 어떻게 더 높은 성능을 낼 수 있는지를 영상의 공간 방향(가로×세로)으로 확인하기 위해, Fig. 5와 같이 부류 활성화 맵(class activation map)을 MSTAR 학습용 표적영상에 덧씌웠다. 부류 활성화 맵이란 네트워크가 표적영상의 어떤 부분에 집중하여 특정 부류(class)로 학습하는데 주요한 특징으로 활용하였는지를 직관적으로 보여주는 방법이다. Fig. 5에서 1~3번째 행은 각각 T-72 탱크(측면각 328.79°), ZIL-131 트럭(측면각 359.06°), BTR-60 장갑차(측면각 191.48°)의 SAR 표적영상을 나타낸다. RGB로 덧씌워진 그림은 부류 활성화 맵으로서 1~4번째 열은 25 % 비율로 학습된 레즈넷18, 레즈넷18 + SE 모듈, 레즈넷 + type BD 모듈, 레즈넷 + type F 모듈에 의한 결과를 나타낸다. 레즈넷18(첫 번째 열)의 경우 학습영상이 부족한 상태에서 중요한 특징으로의 집중(attention) 작용이 없으므로, 학습에 주로 이용된 부분이 표적에서 벗어나 있는 경우가 많다. 그러나 채널방향 집중 모듈이 삽입된 네트워크의 경우 보다 개선된 양상을 보인다. SE 모듈의 경우에는 학습에 주로 이용된 부분이 표적 영역에 집중된 것을 확인할 수 있으며, type BD 및 type F 모듈은 주어진 학습용 표적영상의 특징을 더 효과적 나타내는 부분에 선별적으로 집중된 것을 확인할 수 있다. 이러한 부류 활성화 맵의 분포양상은 본 논문의 서론에서 세운 추론을 간접적으로 입증한다고 할 수 있다.

Fig. 5. Class activation maps for training images

4.2.2 채널(channel)방향의 확인

채널방향 집중 모듈을 삽입한 네트워크의 성능 개선을 보다 직접적으로 확인할 수 있는 방법은 Fig. 6

과 같이 채널 별 특징맵으로부터 활성화가 어떻게 이루어지는 지를 관측하는 것이다. Fig. 6에 나타난 특징맵들은 모두 Fig. 5의 T72 탱크 표적영상을 입력받

Fig. 6. Illustration of channel-wise feature maps of stage 2 for the T72 target (a) ResNet18 (b) ResNet18 + SE (c) ResNet18 + type BD (d) ResNet18 + type F

아 Fig. 1 네트워크 구조의 ‘stage 2’를 거친 결과이며, 총 64개 채널의 특징맵 중 활성화 정도가 큰 순으로 36개만 나타난 것이다. ‘stage 2’ 특징맵은 상대적으로 덜 추상화되어, 보다 직관적인 관측이 가능하다. (a)부터 (d)는 25 % 비율로 학습된 레즈넷18, 레즈넷18 + SE 모듈, 레즈넷 + type BD 모듈, 레즈넷 + type F 모듈에 의한 채널 별 특징맵을 나타낸다.

먼저 채널방향 집중 모듈을 삽입하지 않은 Fig. 6(a)은 모듈이 삽입된 Fig. 6(b)~(d)에 비해 활성화 정도가 미약함을 알 수 있다. 활성화 정도의 크기순으로 선정한 36개의 특징맵 중에서도 10여개의 특징맵에서 활성화 특성의 관측이 어렵다. 일부 표적의 형태를 보여주는 특징맵들 또한 표적이 표적영상에서 실제로 차지하고 있는 비중에 비해 적고, 표적의 형태 또한 분명하지 못하다. 반면 Fig. 6(b)~(d)에서는 강한 활성화 특성이 나타나는 특징맵이 Fig. 6(a)보다 상대적으로 더 많고, 표적이 위치하는 부분뿐만 아니라 그림자 부분도 활성화가 됨을 알 수 있다. 이중 SE 모듈이 삽입된 Fig. 6(b)와 type F 모듈이 삽입된 Fig. 6(d)는 각각의 특징맵들이 표적, 표적과 그림자, 그림자 등의 형태로 나타나는 등 유사한 활성화 양상을 보인다. 이는 type F 모듈의 드롭아웃이 학습과정과는 달리 활성화 특성 관측 과정에서는 작용하지 않아 실질적으로 SE 모듈과 동일한 형태의 모듈을 갖기 때문으로 분석된다. type BD 모듈이 삽입된 Fig. 6(c)는 일부 그림자가 나타나는 특징맵도 있으나, 대부분의 특징맵이 표적의 전체 형태 및 세부적인 부분을 나타내므로 상대적으로 표적 부분의 활성화에 치중하는 양상을 보인다. 이상 채널방향 특징맵의 관측결과로부터, 채널방향 집중 모듈의 유형 별로 조금씩 다른 활성화 과정을 거치지만, 모듈이 삽입되지 않은 경우에 비해 명확한 활성화 양상을 보이며, 이를 통해 네트워크의 식별 성능을 개선시킨다는 것을 확인할 수 있다.

4.3 다른 연구사례와의 성능 비교

Table 4에서는 본 논문에서 제안된 ‘채널방향 집중 모듈이 삽입된 합성곱 신경망’을 이용하여, 제한적인 각도 다양성 조건에서의 SAR 표적영상을 식별한 결과를 타 연구사례^[7-9,13]에서 발표된 결과와 비교하였다. (소수 첫째자리 반올림) 그 외 Table 4의 부가정보는 다음과 같다.

① 일부 연구사례(- 표시)에서는 학습영상 이용비율이

- 10 %인 경우에 대해 식별성능을 산출하지 않았다.
- ② 25 % 산출결과에서 *표시가 된 수치는 해당 문헌에서 학습영상 이용비율 20 % 및 30 %에 대해 산출한 결과를 평균한 값이다. 20 % 및 30 %에 대해 J. H. Cho 등^[7]은 각각 91.08 %, 92.95 %를 산출하였고, Q. Yu 등^[8]은 각각 92.2 %, 95.4 %를 산출하였으며, Z. Lin 등^[9]은 각각 73.5 %, 94.97 %를 산출하였다.
- ③ Q. Yu 등^[8]과 Z. Lin 등^[9]의 연구에서는 학습영상 이용비율에 따른 식별성능을 수치(%)로 명시하지 않아 문헌의 그래프로부터 추정된 값을 기재하였다.
- ④ 본 논문에서 제시된 방법을 학습영상 이용비율 10 % 및 25 %인 경우에 적용한 경우, 최고성능 대신 Table 3에 나타냈던 평균값을 기재하였다.
- ⑤ 인용된 타 연구사례에서는 모두 본 논문과 같이 표준 운용조건(SOC)에서의 MSTAR SAR 표적영상 데이터 셋을 사용하였다. 다만 학습영상 이용비율이 50 % 이하인 경우, 수량이 동일하더라도 실제로 학습에 활용된 영상의 세부내역은 달라질 수 있다. 앞선 실험에서 25 % 및 10 %의 경우에 대해 학습영상을 세 가지로 임의 선정한 것과 연관된다.

Table 4. Performance comparison with other methods

연구사례	학습영상 이용비율			
	10 %	25 %	50 %	100 %
J. H. Cho et al. ^[7]	69.8 %	92.0 %*	95.0 %	95.5 %
Q. Yu et al. ^[8]	-	93.8 %*	97.7 %	99.8 %
Z. Lin et al. ^[9]	35.5 %	84.2 %*	97.9 %	99.1 %
F. Zhang et al. ^[13]	-	95.8 %	97.7 %	99.9 %
ResNet18 + SE	75.6 %	93.7 %	97.3 %	99.1 %
ResNet18 + typeBD	80.2 %	94.6 %	98.1 %	99.3 %
ResNet18 + typeF	75.7 %	94.9 %	98.0 %	99.4 %

Table 4로부터, 학습영상 이용비율이 50 % 이하인 제한적인 각도 다양성 조건에서 본 논문에서 제시된 채널방향 집중 모듈 삽입에 의한 합성곱 신경망의 식별성능이 다른 연구사례에서 도출된 식별성능과 필적하거나 보다 우수하다는 것을 확인할 수 있다. 10 %인 경우에는 다른 연구에서 공개된 성능보다 크게 개

선되었을 뿐만 아니라 성능의 열화 정도가 상대적으로 낮음을 알 수 있으며, 25 %인 경우에는 F. Zhang 등의 연구^[13]보다는 다소 낮지만 95 %에 가까운 성능을 보인다. 또한 50 %인 경우에는 98 % 이상의 식별 성능을 보이는데, 이는 전통적인 패턴인식 기반의 분류기에서 100 %의 MSTAR 학습영상을 모두 활용하여 성능을 산출한 결과를 상회하는 값이다. Table 4의 1~2번째 연구사례^[7,8]는 합성곱 신경망에서 도출된 특징을 융합하였고, 3번째 연구사례^[9]는 네트워크 구조적 관점에서 접근하였으며, 4번째 연구사례^[13]는 SAR 표적영상에서 멀티뷰 시퀀스를 도출하여 특징을 증대하였다. 본 논문에서는 서론에서 언급한 바와 같이 네트워크의 구조적인 관점에서 제한적인 각도 다양성 조건에서의 SAR 표적영상 식별성능 열화문제를 해결하고자 하였으며, 동일한 접근방식을 채택한 3번째 연구에 비해 부족한 학습영상에 대해 보다 강인한 식별 성능을 도출하였다.

5. 결 론

본 논문에서는 SAR-ATR을 위한 네트워크를 학습시킬 표적영상의 각도 다양성이 제한적인 조건 하에서 채널방향 집중 모듈을 합성곱 신경망에 삽입하여 표적 식별성능을 개선하는 연구를 수행하였다. 레즈넷18 형태의 기본 네트워크에 SE 모듈 및 이를 기반으로 고안된 2종의 채널방향 집중 모듈을 삽입하여 여러 각도 다양성 제한 조건에 대한 SAR 표적식별에 적용하였고, 기본 네트워크에 의한 결과와 비교하거나 유사한 조건에서의 다른 연구사례와 비교하여 성능을 입증하였다. 추후의 연구 방향은 다음과 같다.

- ① MSTAR SAR 표적영상 데이터 셋 외에도 다른 해상도를 갖거나 다른 표적군에 대한 표적영상 데이터 셋에 적용하여, 본 논문에서 제시된 기법의 일반성을 추가 입증할 필요가 있다.
- ② 채널방향 집중 모듈의 장점을 유지하면서 보다 적은 학습시간을 요하는 효율적인 모듈의 구조연구 및 높은 식별성능이 입증된 멀티뷰 특징 도입연구 등으로 보다 각도 다양성이 제한된 조건에서 SAR 표적영상에 대한 식별성능을 개선해나갈 계획이다.
- ③ 또한 앞서 채널방향 집중 모듈의 네트워크 성능 개선 양상을 직관적으로 확인한 것처럼 본 논문

서 제시된 모듈이 확장 운용조건(Extended Operating Condition, EOC)에서는 어떻게 식별성능에 변화를 주는지를 확인하는 것도 가치 있는 연구의 방향이 될 것이다.

References

- [1] K. El-Darymli, et al., "Automatic Target Recognition in Synthetic Aperture Radar Imagery: A State-of-the-Art Review," IEEE Access, Vol. 6, pp. 6014-6058, 2016.
- [2] M. David, "Deep Convolutional Neural Networks for ATR from SAR Imagery," Proc. SPIE, Algorithms for Synthetic Aperture Radar Imagery XXII, Vol. 9475, 2015.
- [3] J. Ding, et al., "Convolutional Neural Network with Data Augmentation for SAR Target Recognition," IEEE Geoscience and Remote Sensing Letters, Vol. 13, No. 3, pp. 364-368, 2016.
- [4] 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.
- [5] S. Wagner, "SAR ATR by Combination of Convolutional Neural Network and Support Vector Machine," IEEE Trans. Aerospace and Electronic Systems, Vol. 52, No. 6, pp. 2861-2872, 2016.
- [6] O. Kechagias-Stamatis et al., "Fusing Deep Learning and Sparse Coding for SAR ATR," IEEE Trans. Aerospace and Electronic Systems, Vol. 55, No. 2, pp. 785-797, 2019.
- [7] J. H. Cho, et al., "Multiple Feature Aggregation Using Convolutional Neural Networks for SAR Image-Based Automatic Target Recognition," IEEE Geoscience and Remote Sensing Letters, Vol. 15, No. 12, pp. 1882-1886, 2018.
- [8] Q. Yu, et al., "High-Performance SAR ATR Under Limited Data Condition Based on a Deep Feature Fusion Network," IEEE Access, Vol. 7, pp. 165646-165658, 2019.
- [9] Z. Lin, et al., "Deep Convolutional Highway Unit Network for SAR Target Classification with Limited

- Labeled Training Data,” IEEE Geoscience and Remote Sensing Letters, Vol. 14, No. 7, pp. 1091-1095, 2017.
- [10] F. Zhou, et al., “SAR ATR of Ground Vehicles Based on LM-BN-CNN,” IEEE Trans. Geoscience and Remote Sensing, Vol. 56, No. 12, pp. 7282-7293, 2018.
- [11] L. Wang, et al., “SAR ATR of Ground Vehicles Based on ESENet,” Remote Sensing, Vol. 11, No. 11, pp. 1316-1331, 2019.
- [12] G. Huang, et al., “A Novel Group Squeeze Excitation Sparsely Connected Convolutional Networks for SAR Target Classification,” International Journal of Remote Sensing, Vol. 40, No. 11, pp. 4346-4360, 2019.
- [13] F. Zhang, et al., “Multi-Aspect-Aware Bidirectional LSTM Networks for Synthetic Aperture Radar Target Recognition,” IEEE Access, Vol. 5, pp. 26880-26891, 2017.
- [14] J. Pei, et al., “SAR Automatic Target Recognition Based on Multiview Deep Learning Framework,” IEEE Trans. Geoscience and Remote Sensing, Vol. 56, No. 4, pp. 2196-2210, 2018.
- [15] P. Zhao, et al., “Multi-Stream Convolutional Neural Network for SAR Automatic Target Recognition,” Remote Sensing Vol. 10, No. 9, pp. 1473-1494, 2018.
- [16] J. Hu, et al., “Squeeze-and-Excitation Networks,” Proc. IEEE Conference on Computer Vision and Pattern Recognition(CVPR 2018), pp. 7132-7141, 2018.
- [17] K. He, et al., “Deep Residual Learning for Image Recognition,” Proc. IEEE Conference on Computer Vision and Pattern Recognition(CVPR 2016), pp. 770-778, 2016.
- [18] K. He, et al., “Identity Mappings in Deep Residual Networks,” Proc. European Conference in Computer Vision(ECCV 2016), pp. 630-645, 2016.
- [19] A. Bochkovskiy, et al., “YOLOv4 : Optimal Speed and Accuracy of Object Detection,” arXiv 2020, Available Online:Arxiv.org/abs/2004.10934.
- [20] S. Woo, et al., “CBAM : Convolutional Block Attention Module,” Proc. European Conference in Computer Vision(ECCV 2018), pp. 3-9, 2018.