Journal of KIIT. Vol. 23, No. 3, pp. 135-144, Mar. 31, 2025. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2025.23.3.135

소규모 SAR 데이터셋에 효과적인 딥러닝 표적 식별 모델 탐구

양유경*[1], 최여름*[2], 채대영*[3], 송정민*[4]

Exploring Effective Deeplearning Target Recognition Models for Small-Scale SAR Datasets

Yu Kyung Yang*[1], Yeoreum Choi*[2], Dae Young Chae*[3], and Jung Min Song*[4]

이 논문은 2024년 정부의 재원으로 수행된 연구 결과임

요 약

SAR(Synthetic Aperture Radar) 영상은 기상에 의존하지 않고 원거리 획득이 가능한 장점으로 인해 군의 감시정찰 분야나 국가 재난 감시 등 다양한 분야에 활용되고 있다. 하지만 SAR 영상은 데이터 획득 비용 및 획득 소요 시간으로 인해 큰 규모의 데이터셋 구축이 어려워 딥러닝 기반 표적 식별 모델의 성능향상을 어렵게 한다. 이러한 문제를 해결하기 위해 시뮬레이션 SAR영상을 생성하여 활용하거나, 대규모 광학 영상 데이터셋에서 학습된 모델을 소규모 SAR 영상에 전이학습하여 사용한다. 본 논문에서는 대규모 광학 영상 기반으로 개발된 최신의 고성능 식별 모델들을 선정하고 시뮬레이션 SAR 영상 및 소규모 SAR 실측 데이터셋에 전이학습하여 성능을 비교 분석함으로써 소규모 SAR 데이터셋에 효과적인 표적 식별 모델을 탐구한다.

Abstract

SAR(Synthetic Aperture Radar) images are used in various fields such as military surveillance and reconnaissance and national disaster monitoring due to the advantage of being able to acquire images from a long distance without relying on the weather condition. However, it is difficult to build a large-scale datasets due to data acquisition costs and time, making it difficult to improve the performance of deep learning-based target recognition models. To solve this problem, simulated SAR images are often utilized, and there is a method of transferring a model learned from a large-scale optical images to a small-scale SAR images. In this paper, we select the latest state-of-the-art recognition models developed based on large-scale optical images and explore effective target recognition models for small-scale SAR datasets by comparing and analyzing their performance through transfer learning on simulated SAR images and small-scale SAR real images.

Keywords

SAR target recognition, small-scale dataset, classification model, transfer learning

* 국방과학연구소 연구원(*[1] 교신저자)
- ORCID[1]: https://orcid.org/0009-0004-6817-1188
- ORCID[2]: https://orcid.org/0000-0003-1541-7477
- ORCID[3]: https://orcid.org/0000-0001-6664-4034
- ORCID[4]: https://orcid.org/0009-0000-1776-5550

· Received: Dec. 03, 2024, Revised: Dec. 11, 2024, Accepted: Dec. 14, 2024

· Corresponding Author: Yu Kyung Yang
Defense AI Center, Agency for Defense Development Yuseong P.O.Box 35, Daejeon, 24186, Korea
Tel.: +82-42-281-0438, Email: ykyang@add.re.kr

Ⅰ. 서 론

SAR(Synthetic Aperture Radar) 영상은 마이크로파를 방사하여 물체로부터 반사/산란된 신호를 수신하여 획득되는데, 기상에 의존하지 않고 원거리 획득이 가능한 장점으로 인해 군의 감시정찰 분야나 국가 재난 감시 등 다양한 분야에 활용되고 있다. 하지만 고가의 SAR 센서, 복잡한 획득 과정, 광학 센서에 비해 상대적으로 긴 획득 소요 시간으로 인해 SAR 영상 획득에는 많은 비용과 시간이 요구된다. 또한, SAR 영상은 반사된 전자파를 영상화 하기 때문에 그림 2에서 보는 바와 같이, 광학 영상과는 전혀 다른 형상으로 획득되어 레이블링 작업도 상대적으로 더 어렵다. 이러한 문제들로 인해 큰 규모의 SAR 영상 데이터셋 구축이 어렵다.

딥러닝 모델들의 성능은 데이터셋의 규모에 크게 의존하며, 데이터셋의 규모에 비례하여 모델의 규모를 키워야 성능이 향상되는 것으로 알려져 있다[1]. 최근 광학영상 도메인에서는 억단위의 이미지 수량을 갖는 거대 데이터셋이 등장하면서 이 데이터셋으로 학습한 큰 규모의 모델들이 SOTA (State-of-the-Art) 성능을 보여주고 있다[1][2]. 반면 SAR 영상은 데이터셋의 부족 문제로 인해 SAR영상 기반 딥러닝 모델의 성능향상이 제한되는 문제를 겪고 있다.

SAR 데이터셋의 부족 문제를 해결하기 위해 시뮬레이션 SAR 영상을 생성하여 활용할 수 있다[3][4]. 또한 대규모의 광학영상 데이터셋에서 학습된 모델을 소규모 SAR 영상 데이터셋에 전이학습하는 방법들이 적용되고 있다[5]-[7]. 본 논문에서는 대규모 광학영상 기반으로 개발된 최신의 고성능 식별 모델들을 선정하고, 이 모델들을 시뮬레이션 SAR 영상 및 소규모 SAR 실측 데이터셋에 전이학습하여 성능을 비교 분석함으로써 소규모 SAR 데이터셋에 효과적인 표적 식별 모델을 탐구한다.

Ⅱ. 관련 연구

2.1 SAR 데이터셋 부족 문제

SAR 데이터셋의 부족 문제를 해결하기 위해 그동안 많은 연구가 시도되었다. 먼저 SAR 데이터 도메인 내에서 해결하는 방법으로는 데이터 레이블이 일부만 있는 경우에 학습하는 준지도 학습(Semi-supervised learning)이 있는데, Wen은 영상을 그래프 구조로 모델링하고 어텐션 메커니즘을 적용하여 소수의 레이블 데이터만으로 높은 정확도를 보이는 SAR 표적식별 방법을 제안하였다[8]. 다른 방법으로는 클래스당 샘플수가 매우 적은 경우에 적용하는 Few-Shot 학습 방법이 있는데, Fu는 SAR 표적 식별 분야에 최초로 메타학습(Metalearning)을 적용하는 방법을 제안하였다[9].

SAR 데이터셋의 부족을 극복하기 위한 다른 방법으로는 데이터셋이 풍부한 도메인에서 모델을 학습한 후 습득된 지식을 SAR 데이터 도메인으로 전이하여 학습하는 방법이 있다[5]-[7]. 최근 영상개수를 3억개까지 키운 JFT-300M과 같은 거대 데이터셋에 규모가 매우 큰 모델들을 학습하여 SOTA성능을 보여주는 BiT[1], ViT[2]과 같은 모델들이 소개되고 있고, 이러한 모델들을 SAR 영상에 적용하려는 시도가 이어지고 있다. Singh는 큰 규모의 광학영상 데이터셋에서 사전 학습된 6개의 SOTA 모델들(DenseNet[10], MobileNet[11], XceptionNet[12], ConvNeXt[13], BiT[1], ViT[2])을 MSTAR 데이터셋[14]에 세부 튜닝(fine tuning)한 성능을 비교 분석하였는데, 별도의 도메인 적응 기법을 적용하지 않은 단순한 세부 튜닝 만으로도 95.8~99.7%의 높은 정확도를 보였으며 특히, 상대적으로 최신 모델인 ConvNeXt, BiT, ViT이 높은 성능을 보였다[7].

SAR 데이터셋의 부족을 극복하기 위한 또 다른 방법으로 시뮬레이션 데이터를 활용하는 방법들이 있다[3][4]. Malmgren-Hansen는 지형모델과 표적의 CAD 모델로부터 시뮬레이션 영상을 생성하는 프로그램인 SARSIM[15]를 이용해 SAR 시뮬레이션 영상을 생성하고, 시뮬레이션 영상에서 표적 식별 모델을 학습한 후 MSTAR 데이터셋에 전이학습하여, SAR 데이터셋만 이용하는 경우보다 성능향상됨을 보였다[3].

우리는 앞서 소개한 Singh와 Malmgren-Hansen의 연구로부터 시뮬레이션 데이터를 사용하고 측정 데이터셋이 매우 부족한 조건에 대한 표적 식별 연구로 범위를 확장고자 한다.

즉, 큰 규모의 광학 영상데이터셋에서 사전학습된 SOTA 모델들을 선정하고 SAR 시뮬레이션 영상과 매우 적은 SAR 실측 영상 데이터셋에 전이학습한 성능을 비교 분석하여 데이터가 매우 부족한 경우에 효과적인 식별 모델을 탐구 하였다.

2.2 SOTA 식별 모델

소규모 SAR 데이터셋에 대한 전이학습 비교를 위해 4개의 ConvNet 계열 모델과 1개의 Transformer 계열 모델을 선정하였다. ConvNet 계열 모델은 앞서 소개한 Singh[7]의 연구 결과에서 상대적으로 성능이 좋았던 BiT[1]과 ConvNeXt[13] 모델을 선정하였다. 또한 많은 비교실험 결과에서 꾸준히 좋은 성능이 보고되는 EfficientNet[16]과 ConvNeXt를 향상시킨 ConvNeXt V2[17] 모델을 추가적으로 선정하였다. Transformer 계열 모델로는 Singh[7] 연구에서 사용한 ViT 모델 대신 최근 더 높은 성능을 보이는 Swin Transformer[18] 모델을 선정하였다.

2.2.1 EfficientNet[16]

EfficientNet 모델은 2019년 Google Research Brain에서 제안한 모델로, 기존의 연구[19][20]이 더 높은 정확도를 달성하기 위해 단순히 모델의 크기를 키워가는 방향으로만 발전하였던 점을 보완하여, 효율적인 스케일 업(Scale up) 방법을 제안하였다. 네트워크의 깊이(Depth), 너비(Width), 해상도(Resolution)를 효율적으로 구성하도록 복합 계수(Compound Scaling) 기법을 적용하여 가장 기본이 되는 모델인 EffientNet-B0를 구성하였다. 기본 모델로부터 각 계수를 증가시켜가며 B1부터 B7까지 모델을 구성하였다. EfficientNet은 훨씬 적은 파라미터와 연산량을 가진 모델로 다양한 데이터셋에서 최신 모델의 성능과 동등하거나 높은 성능을 달성했다.

2.2.2 BiT[1]

BiT 모델은 Google Research Brain에서 대규모 데이터셋에서 학습한 고성능의 범용모델을 제공하고자 개발한 모델이다. ResNet[19] 모델을 기반으로 모델의 깊이와 너비를 키운 다양한 규모의 모델들(ResNet-50x1, 50x3, 101x1, 152x2, 152x4 등)을 제안하였고, 큰 규모의 모델을 큰 규모의 데이터셋에 학습하는 것을 고려하여 배치 정규화(Batch Normalization)대신 그룹 정규화(Group Normalization) 및 가중치 표준화(Weight Standardization)을 적용하였다. 학습 데이터셋의 규모에 따라서 130만장의 ILSVRC-2012 데이터셋에서 학습된 BiT-S 모델, 1400만장의 ImageNet-22k 데이터셋에서 학습된 BiT-M 모델, 3억장의 JFT-300M 데이터셋에서 학습된 BiT-L 모델이 있으며 BiT-L모델로 ImageNet 벤치마크에서 2020년 SOTA 성능을 달성했다.

2.2.3 Swin Transformer[18]

Swin Transformer 모델은 Microsoft Research에서 개발한 모델로, 본래 자연어처리(National Language Processing, NLP)에서 사용되었던 Transformer 모델을 컴퓨터 비전 분야에 적용한 모델 중 하나이다. 기존의 Transformer 기반의 연구[2][21]에서 토큰(Token)으로 고정된 크기의 패치 영상을 사용하여 다양한 스케일의 정보를 활용하지 못하고, 고해상도 영상을 처리하기 어렵다는 단점을 보완하기 위해 계층적 특징맵과 Shifted Window Partitioning 기법을 적용하였다. ImageNet-1k 데이터셋과 ImageNet-22k 사전학습 모델에 대해서 모두 SOTA 모델과 비교하여 더 적은 파라미터와 연산량으로 높은 성능을 보였으며, 식별뿐만 아니라 객체탐지(Object detection)와 이미지 분할(Semantic segmentation) 분야에서도 SOTA 성능을 달성했다.

2.2.4 ConvNeXt[13]

ConvNeXt 모델은 Facebook AI Research와 UC Berkely에서 개발한 모델로, Transfomer의 등장으로 비전분야 주류에서 밀려나가고 있는 ConvNet의 현대화를 목적으로 개발하였다. ResNet[19] 모델을 기본으로 최신 모델들에서 적용하고 있는 다양한 학습 기술과 구조변화(패치화된 stem, depthwise convolution, inverted bottleneck, 커널 사이즈 증대 및 activation 함수 변경)에 대한 실험을 통해 각각의 성능 기여를 확인하여 적용하였다.

Residual 블록의 채널수와 블록의 반복 횟수로 구분되는 모델의 규모에 따라 ConvNeXt-T, -S, -B, -L, -XL의 모델이 있다. ConvNeXt는 ImageNet 벤치마크에서 2022년 Swin Transformer와 같은 최신 Transformer 계열의 모델의 성능을 뛰어넘는 식별 결과를 보였다.

2.2.5 ConvNeXt V2[17]

ConvNeXt V2는 ConvNeXt 모델에 Masked Auto Econder(MAE) 기반 자체 지도 사전 학습을 도입하여 개발된 모델이다. MAE는 일부 랜덤한 영역들을 제거한 영상을 입력으로 하여 특징을 추출하는 인코더와 추출된 특징으로부터 원래의 영상을 복원하는 디코더로 구성된다. ConvNeXt V2는 인코더와 디코더로 ConvNeXt 모델과 ConvNeXt 블록을 사용하는 fully convolutional MAE 구조를 갖는다. 또한, Multi-Layer Perceptron(MLP) 레이어의 특징 포화 문제를 해결하기 위해 ConvNeXt residual 블록에 추가적으로 Global Response Normalization(GRN)이라는 새로운 정규화 레이어를 적용하였다. ConvNeXt V2는 ImageNet 벤치마크에서 2023년 기존의 지도학습 기반인 ConvNeXt 결과를 뛰어넘는 결과를 보였다. ResNet과 ConvNeXt, ConvNeXt V2 모델의 각 residual 블록 구성의 차이점은 그림 1과 같다.

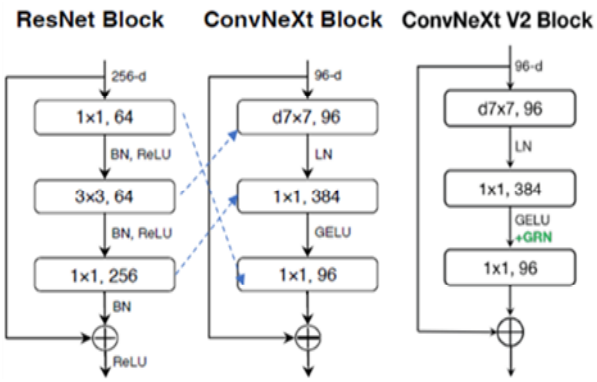


그림 1. ResNet, ConvNeXt, ConvNeXt V2 모델의 residual 블록 구성 비교

Fig. 1. Comparison of the residual block in ResNet, ConvNeXt and ConvNeXt V2

III. 소규모 SAR 데이터셋에 대한 식별 모델 성능 비교

3.1 소규모 SAR 표적 식별 데이터셋

본 연구에서 사용한 데이터셋은 총 10종의 지상차량에 대해 시뮬레이션 SAR 영상과 실측 SAR 영상으로 구성하였다. 실측영상은 운용조건을 고려한 실표적 측정 연구[22]에서 획득한 영상을 활용하였는데, 이는 X밴드 소형 SAR 장비인 PicoSAR를 항공기에 장착하여 지상에 배치된 표적 주변을 선회하면서 360도 전방위의 영상을 획득한 것으로, 획득된 영상에서 표적 중심의 180x180 화소 크기의 TOI(Target of Interest) 영상으로 잘라내어 식별용 데이터셋을 구성한 것이다. 본 연구에서는 0.3m의 해상도를 갖는 영상을 학습 데이터와 시험 데이터가 고른 방위각 분포를 가지도록 구성하였다. 시뮬레이션 영상은 복잡한 산란구조를 갖는 실표적의 전자기파 수치해석 시뮬레이션 연구[23]에서 구축한 데이터를 활용하였는데, 이는 10종의 지상차량에 대해 레이저 스캔으로 3D CAD 모델을 생성하고, 이것에 대한 전자파 수치 해석을 통해 SAR 영상을 생성하는 방식이다. 그림 2는 실측 및 시뮬레이션으로 생성한 표적 TOI 영상의 예제를 보여준다.

표 1은 본 연구를 위해 구성한 데이터셋의 각 클래스별 수량을 영상 수량을 보여준다. 소규모 측정 데이터만을 사용하기 위해 실측 SAR영상이 시뮬레이션 영상의 10%가 되는 2가지 학습 데이터를 구성하였다. 각 실측 학습 데이터는 676개와 338개로 매우적으며 시험 데이터는 공정한 평가를 위해 상대적으로 많은 1,981개의 영상으로 구성하였다.

3.2 소규모 SAR 데이터셋에 대한 식별 모델 성능 비교

본 연구에서는 5가지의 종류의 모델에 대해 총 14개의 세부모델을 선정하였으며 각 모델의 파라미터 수는 표 2와 같다. EfficientNet 모델은 ImageNet-1k 데이터셋, 그 외 모델은 ImageNet-22k 데이터셋에서 사전학습된 파라미터를 초기값으로하여 학습하였다. 시뮬레이션 데이터셋에 대해 먼저 150 에폭 동안 학습한 후에 실측 데이터셋에 추가 150에폭 동안 학습하였다.

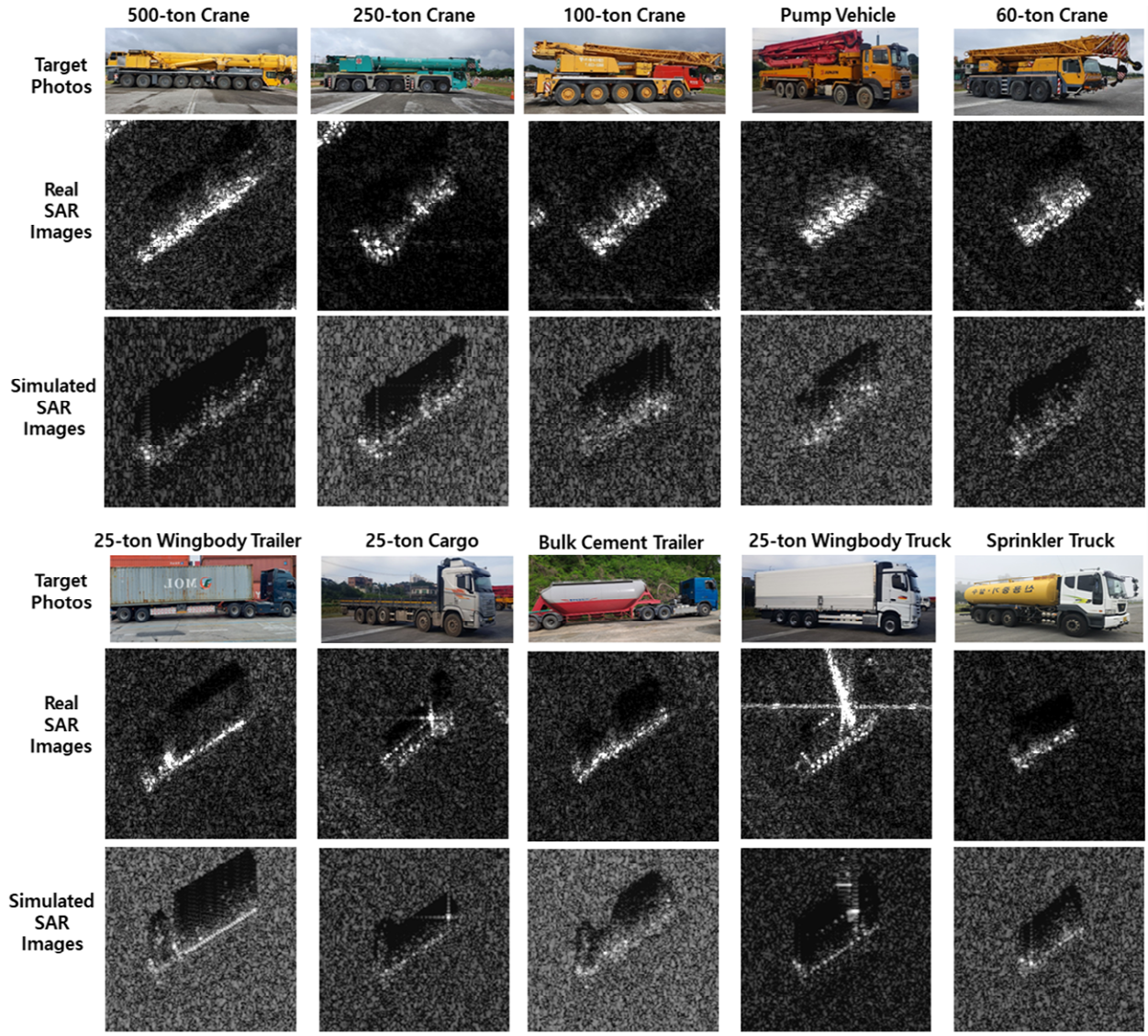


그림 2. 표적별 SAR 실측 영상 및 시뮬레이션 영상
Fig. 2. Acquired and simulated SAR target images

표 1. 소규모 SAR 표적 식별 데이터셋
Table 1. Small SAR target recognition datasets

Datasets \ Target categories		500 -ton crane	250 -ton crane	100 -ton crane	Pump vehicle	60 -ton crane	25 -ton Wing body trailer	25 -ton cargo	Bulk cement trailer	25 -ton wing body truck	Sprinkler truck	Total
Train Dataset A	Real	58	58	58	84	58	70	80	70	64	76	676
	Simulated	966	966	965	483	966	483	483	483	482	483	6,760
Tran Dataset B	Real	29	29	29	42	29	35	40	35	32	38	338
	Simulated	483	483	483	241	483	242	241	242	241	241	3,380
Test Dataset(real)		168	168	168	251	168	204	238	204	190	222	1,981

표 2. 식별 모델별 파라미터 수
Table 2. Number of parameters of models

Models	Specific models	Parameters
EfficientNet	EfficientNet-B0	5.3M
	EfficientNet-B1	7.8M
BiT	BiT-M-R50x1	24M
	BiT-M-R101x1	43M
	BiT-M-R50x3	211M
	BiT-M-R152x2	233M
Swin transformer	SwinT-Base	88M
	SwinT-Large	197M
ConvNeXt	ConvNeXt-S	50M
	ConvNeXt-B	89M
	ConvNeXt-L	198M
ConvNeXt V2	ConvNeXt V2-T	29M
	ConvNeXt-V2-B	89M
	ConvNeXt-V2-L	198M

각 모델의 학습 스케쥴링과 데이터 증강은 각 모델의 저자가 제안한 방법을 따랐다.

표 3은 학습데이터 A에 대해 학습한 모델들의 시험데이터에 대한 Top-1 식별 정확도를 비교한 것이다. 이 값은 실측 데이터에 학습하는 150 에폭 동안 모든 에폭에서의 시험데이터에 대한 Top-1 정확도중 가장 높은값을 기준으로 하였다.

EfficientNet 모델은 B0, B1 모델 모두 높은 정확도를 보였으며 EfficientNet-B1 모델이 1% 더 높은 성능을 보였다. 규모가 큰 모델이 더 높은 성능을 보였는데, 이는 EfficientNet이 가장 적은 파라미터를 보유한 모델인 만큼 학습데이터 A가 두 모델을 학습하기에 충분하였다는 것을 확인할 수 있다.

BiT 모델의 경우, 규모가 작은 50x1, 101x1 모델은 87.8 ~ 88.1%의 정확도를 보였으나 규모가 큰 152x2나 R50x3 모델은 성능이 매우 저조해 큰 규모의 모델은 작은 SAR 데이터셋으로 충분한 학습이 이루어지지 않는다는 것을 알 수 있다.

Swin Transformer 모델은 Base와 Large 모델 모두 97% 이상의 높은 성능을 보였다. 특히 비슷한 규모의 모델인 ConvNeXt-M/L 및 ConvNeXt-V2-B/L과 각각 비교하여도 더 높은 정확도를 보였다.

ConvNeXt 모델의 경우 규모가 작은 모델부터 큰 모델까지 93.4% ~ 97.1%의 성능을 보였다. ConvNeXt-L은 BiT-M-R50x3, 152x2와 비슷하게 큰 규모의 모델인데도 학습이 잘 이루어진 것을 알 수 있다.

표 3. Dataset A에 대한 모델별 Top-1 정확도 (%)
Table 3. Comparison of Top-1 accuracy for Dataset A (%)

Models	Specific models	500 -ton crane	250 -ton crane	100 -ton crane	Pump vehicle	60 -ton crane	25 -ton wing body trailer	25 -ton cargo	Bulk cement trailer	25 -ton wing body truck	Sprinkler truck	Total
Efficient Net	EfficientNet-B0	97.6	94.6	86.9	98.8	89.9	95.1	98.7	97.6	100	98.7	**96.2**
	EfficientNet-B1	99.4	96.4	89.9	98.4	91.7	97.6	99.6	97.1	100	99.1	**97.2**
BiT	BiT-M-R50x1	93.5	76.2	50.6	86.5	86.9	94.1	97.1	89.2	96.3	98.2	**87.8**
	BiT-M-R101x1	95.2	73.8	58.3	83.3	91.7	92.2	96.2	89.7	98.9	95.5	**88.1**
	BiT-M-R152x2	66.1	14.9	0.0	68.5	0.6	2.9	89.5	0.0	77.4	78.8	**42.9**
	BiT-M-R50x3	50.0	10.1	0.0	45.0	0.0	22.5	87.8	52.5	47.4	47.7	**39.0**
Swin transformer	SwinT-Base	100	98.2	86.9	98.8	95.2	98.5	100	99.5	100	99.6	**97.9**
	SwinT-Large	99.4	99.4	88.7	100	96.4	100	100	100	100	99.6	**98.6**
Conv NeXt	ConvNeXt-S	98.2	97.6	82.7	96.8	93.5	100.0	98.7	98.5	100.0	98.2	**96.7**
	ConvNeXt-B	98.2	95.8	82.1	96.4	92.3	99.5	99.2	98.0	100.0	98.2	**96.3**
	ConvNeXt-L	97.6	98.2	86.3	97.2	93.5	99.5	99.2	99.5	100.0	97.3	**97.1**
Conv NeXt V2	ConvNeXt V2-T	97.6	93.5	71.4	91.2	85.1	96.1	97.5	90.2	98.4	96.4	**92.2**
	ConvNeXt V2-B	95.8	92.3	78.0	93.6	89.9	98.5	99.6	97.1	100.0	96.4	**94.5**
	ConvNeXt V2-L	91.7	95.2	82.1	93.6	82.7	99.0	99.2	97.1	96.8	95.0	**93.7**

ConvNeXt V2는 작은 모델부터 큰 모델까지 92.2 ~ 94.6%의 성능을 보였다. ConvNeXt와 비교하면 ConvNeXt V2-B 모델이 ConvNeXt-B모델 보다 우세했지만 ConvNeXt V2-L모델이 ConvNeXt-L모델보다 낮은 결과를 보였다.

표 4는 각 학습데이터 B에 대해 학습한 모델들의 시험데이터에 대한 Top-1 식별 정확도를 비교한 것이다. 학습데이터 B는 학습데이터 A에 비해 수량이 반으로 줄어든 것으로 모든 모델들의 성능이 저하되었다.

EfficientNet 모델은 95.1%, 94.8%로 학습데이터 A보다 약 1~3%의 성능 감소가 발생하였다. 특히 EfficientNet-B1은 B0보다도 낮은 정확도를 보였는데, 이는 학습데이터 B가 EfficientNet-B1 모델을 충분히 학습하기에는 수량이 부족하였기 때문이라고 볼 수 있다.

BiT 모델은 50x1, 101x1 모델이 83.9%의 성능으로 학습데이터 A에서 학습한 경우보다 약 4%정도 성능이 감소하였고, 큰 모델인 153x2, 50x3은 9 ~ 17%의 큰 폭의 성능 감소가 있었다.

Swin Transformer 모델 역시 학습데이터 B에 대해서 두 모델 모두 약 2%의 성능 저하가 발생하였다. 다만 그럼에도 95% 이상의 높은 정확도를 보여, 소규모 SAR 데이터셋 전이학습에 효과적임을 확인할 수 있다.

ConvNeXt는 92.3 ~ 93.4%의 정확도로 학습데이터 A에 비해 약 1 ~ 4%의 성능 감소가 있었다. 학습데이터 A에서는 ConvNeXt-L 모델이 가장 성능이 좋았는데 학습데이터 B에서는 ConvNeXt-S모델의 성능이 더 좋았다. 즉, ConvNeXt모델도 학습 데이터의 규모에 따라 적정 규모의 모델을 선택하는 것이 중요하다는 것을 알 수 있다.

ConvNeXt V2는 84.0 ~ 89.5%의 정확도로 학습데이터 A에서 보다 약 4 ~ 8%의 성능 감소를 보여, ConvNeXt보다 더 크게 성능이 감소하였다. 이것으로 MAE 기반 사전 학습이 소규모 SAR 데이터셋으로의 전이학습에는 효과적이지 않다는 것을 알 수 있다.

그림 3은 학습데이터 A와 학습데이터 B에 대해 학습한 각 모델의 Top-1 정확도를 그래프로 비교한 것이다. 두 학습 데이터셋 모두 Swin Transformer 모델이 가장 성능이 좋았다. BiT 모델처럼 단지 ResNet 모델의 넓이와 깊이만 키운 모델 보다 적극적으로 최신 학습기술과 구조를 반영한 Swin Transformer, EfficientNet, ConvNeXt 모델이 소규모 SAR 데이터셋에서 보다 효과적인 모델임을 알 수 있다.

표 4. Dataset B에 대한 모델별 Top-1 정확도 (%)
Table 4. Comparison of Top-1 accuracy for Dataset B (%)

Models	Specific models	500-ton crane	250-ton crane	100-ton crane	Pump vehide	60-ton crane	25-ton wing body trailer	25-ton cargo	Bulk cement trailer	25-ton wing body truck	Sprinkler truck	Total
Efficient Net	EfficientNet-B0	97.0	92.3	88.7	97.2	92.3	93.1	99.6	93.1	100	94.6	**95.1**
	EfficientNet-B1	98.8	89.3	77.4	97.6	97.0	95.1	99.2	93.1	99..5	96.4	**94.8**
BiT	BiT-M-R50x1	89.3	77.4	57.7	70.9	76.8	90.7	95.8	81.4	98.4	95.0	**83.8**
	BiT-M-R101x1	88.1	71.4	53.6	74.9	83.3	85.3	97.9	82.4	97.9	96.4	**83.8**
	BiT-M-R152x2	16.1	0.0	0.0	33.1	0.0	0.0	57.6	50.0	46.8	32.0	**25.7**
	BiT-M-R50x3	0.0	0.0	0.0	91.6	0.0	48.5	82.4	0.0	37.9	0.0	**30.1**
Swin transformer	SwinT-Base	99.4	98.8	75.6	94.8	88.1	96.6	100	99.0	100	99.6	**95.6**
	SwinT-Large	99.4	97.6	83.9	96.8	91.1	97.1	100	100	100	98.7	**96.8**
Conv NeXt	ConvNeXt-S	96.4	96.4	76.8	90.4	89.3	95.1	96.6	96.1	99.5	95.0	**93.4**
	ConvNeXt-B	92.3	89.9	75.0	85.3	87.5	95.6	99.6	96.1	100.0	97.7	**92.3**
	ConvNeXt-L	93.5	93.5	75.0	87.6	88.1	96.1	98.7	96.1	100.0	96.4	**92.8**
Conv NeXt V2	ConvNeXt V2-T	91.7	70.8	58.9	82.9	66.7	89.7	99.2	86.8	95.3	87.8	**84.0**
	ConvNeXt V2-B	84.5	89.9	63.7	87.6	79.2	97.1	97.9	91.7	97.4	93.7	**89.0**
	ConvNeXt V2-L	92.9	88.7	69.0	88.4	78.6	92.2	96.6	91.2	100.0	91.9	**89.5**

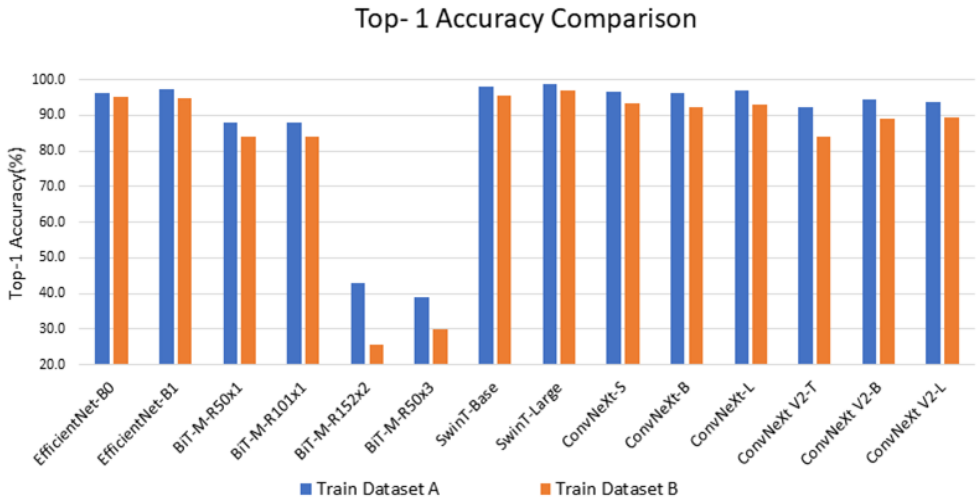


그림 3. 모델별 Top-1 정확도 비교
Fig. 3. Comparison of Top-1 Accuracy among models

Ⅳ. 결론 및 향후 과제

본 논문에서는 대규모 광학영상에서 학습된 SOTA 식별 모델들을 SAR 시뮬레이션 영상 및 매우 적은 수의 SAR 실측 영상으로 구성된 데이터셋에 전이학습 하였을 때 어떠한 성능을 보이는지 비교 분석하였다. ImageNet 벤치마크에서는 모두 뛰어난 성능을 보였던 모델들도 데이터가 매우 적은 SAR 영상 데이터셋에 전이학습한 결과는 성능 차이가 매우 컸다. EfficientNet, SwinTransformer, ConvNeXt의 성능이 우수하였으며 BiT 모델은 성능이 매우 저조하였다. 가장 성능이 좋았던 모델은 Swin Transformer-Large이며, 매우 작은 규모의 SAR 영상 데이터셋에도 효과적으로 전이학습되는 것을 확인 할 수 있었다.

본 연구를 통해 소규모 SAR 데이터셋에 대한 효과적인 식별 모델을 선정하였고, 향후에는 해당 모델에 대한 다양한 학습 방법 및 구조적 변화 등을 시도하여 세부튜닝의 성능에서 보다 향상된 식별 성능을 얻고자 한다.

References

[1] A. Kolesnikov, L. Beyer, and X. Zhai, "Big transfer (bit): General visual representation learning", European Conference on Computer Vision, Online, pp. 491-507, Aug. 2020. https://doi.org/10.48550/arXiv.1912.11370.

[2] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is 11 worth 16x16 words: Transformers for image recognition at scale", International Conference on Learning Representations, Online, May 2021. https://doi.org/10.48550/arXiv.2010.11929.

[3] D. Malmgren-Hansen, A. Kusk, and J. Dall, "Improving sar automatic target recognition models with transfer learning from simulated data", IEEE Geoscience and Remote Sensing Lett.ers, Vol. 14, No. 9, pp. 1484-1488, Sep. 2017. https://doi.org/10.1109/LGRS.2017.2717486.

[4] Y. Choi, "Simulated SAR target recognition using Image-to-Image translation based on complex-valued CycleGAN", Journal of KIIT, Vol. 20, No. 7, pp. 19-28, Jul. 2022. http//dx.doi.org/10.14801/jkiit.2022.20.7.19.

[5] Z. Huang, Z. Pan, and B. Lei, "What, where and how to transfer in SAR target recognition based on deep CNNs", IEEE Geoscience and Remote Sensing, Vol. 58, No. 4, pp. 2324-2336, Apr. 2020. https://doi.org/10.1109/TGRS.2019.2947634.

[6] Z. Huang, Z. Pang, and B. Lei, "Transfer learning with deep convolutional neural network for sar target classification with limited labeled data", Remote Sensing, MDPI, Vol. 9, No. 9, pp. 907, Aug. 2017. https://doi.org/10.3390/rs9090907.

[7] A. Singh and V. K. Singh, "Exploring deep learning methods for classification of SAR Images: Towards nextgen convolutions via transformers", arXiv preprint, arXive:2303.15852, Mar. 2023. https://doi.org/10.48550/arXiv.2303.15852.

[8] L. Wen, X. Huang, and S. Qin, "Semi-Supervised SAR target recognition with graph attention network", 13th European Conference on Synthetic Aperture Radar, online, Mar. 2021. https://ieeexplore.ieee.org/document/9472569.

[9] K. Fu, T. Zhang, and Y. Zhang, "Few-Shot SAR target classification via metalearning", IEEE Geoscience and Remote Sensing, Vol. 60, Feb. 2021. https://doi.org/10.1109/TGRS.2021.3058249.

[10] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolution Networks", IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, pp. 4700-4708, Jul. 2017. https://doi.org/10.1109/CVPR.2017.243.

[11] A. G. Howard, M. Zhu, and B. Chen, "MobileNets: Efficient convolution neural networks for mobile vision applications", arXiv preprint, arXive:1704.04861, Ap. 2017. https://doi.org/10.48550/arXiv.1704.04861.

[12] F. Chollet, "Xception: Deep Learning with depthwise separable convolutions", IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, pp. 1251-1258, Jul. 2017. https://doi.org/10.1109/CVPR.2017.195.

[13] Z. Liu, H. Mao, and C. Wu, "A ConvNet for the 2020s", IEEE Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, pp. 11976-11986, Jun. 2022. https://doi.org/10.1109/CVPR52688.2022.01167.

[14] U.S.AirForce Research Laboratory, "The air force moving and stationary target recognition database", https://www.sdms.afrl.af.mil/index.php?collection=mstar [accessed: Oct. 10, 2024]

[15] A. Kusk, A. Abulaitijiang, and J. Dall, "Synthetic SAR image generation using sensor, terrain and target models", 11th European Conference on Synthetic Aperture Radar, Hamburg, Germany, Jun. 2016. https://ieeexplore.ieee.org/document/7559326.

[16] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks", International Conference on Machine Learning, California, USA, pp. 6105-6114, Jun. 2019. https://doi.org/10.48550/arXiv.1905.11946.

[17] S. Woo, S. Debnath, and R. Hu, "ConvNeXt V2: Co-designing and scailing convNets with masked autoencoders", arXiv preprint, arXiv:2301.00808, Jan. 2023. https://doi.org/10.48550/arXiv.2301.00808.

[18] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," Proc. of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, pp. 10012-10022, Oct. 2021. https://doi.org/10.1109/ICCV48922.2021.00986.

[19] K. He, X. Zhang, and S. Ren, "Deep Residual Learning for Image Recognition", IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, pp. 770-778, Jun. 2016. https://doi.org/10.1109/CVPR.2016.90.

[20] Y. Huang, Y. Cheng, D. Chen, H. Lee, J. Ngiam, Q. V. Le, and Z. Chen, "Gpipe: Efficient training of giant neural networks using pipeline parallelism", arXiv preprint, arXiv:1811.06965, Nov. 2018. https://doi.org/10.48550/arXiv.1811.06965.

[21] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. J´egou, "Training data-efficient image transformers & distillation through attention", International Conference on Machine Learning, Online, pp. 10347-10357, Jul. 2021. https://doi.org/10.48550/arXiv.2012.12877.

[22] D. Y. Chae, H. Lim, J. Park, and J. H. Yoo, "SAR Real Target Measurement Considering Operating Conditions", KIEES Fall Conference, Seoul, Korea,, Vol. 30, No. 1, Nov. 2020.

[23] J. Ko, W. Kim, J. Park, and H. Lim,"Simulated SAR image generation of real large complex target", KIMST Conference, Jeju, Korea, pp. 257-258, Jun. 2021.

저자소개

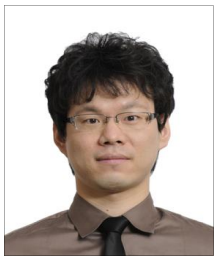
양 유 경 (Yu Kyung Yang)

2002년 2월 : 전북대학교 정보통신공학과(공학사)
2004년 2월 : 한국과학기술원 전기 및 전자공학과(공학석사)
2004년 3월 ~ 2007년 1월 : KTF Technologies SW 엔지니어
2007년 2월 ~ 현재 : 국방과학연구소 연구원
관심분야 : 영상처리, 딥러닝 표적 탐지/식별

최 여 름 (Yeoreum Choi)

2015년 2월 : 한국과학기술원 전기 및 전자공학과(공학사)
2017년 2월 : 한국과학기술원 전기 및 전자공학과(공학석사)
2017년 2월 ~ 현재 : 국방과학연구소 연구원
관심분야 : SAR 표적 식별, 딥러닝 표적 식별, 딥러닝 이미지 변환

채 대 영 (Dae Young Chae)



2006년 2월 : 성균관대학교 전자공학과(공학사)
2008년 2월 : POSTECH 전자전기공학과(공학석사)
2008년 2월 ~ 현재 : 국방과학연구소 연구원
관심분야 : 영상처리, 레이더 표적 식별

송 정 민 (Jung Min Song)

2018년 3월 : 고베대학교 전기전자공학과(공학사)
2020년 3월 : 도쿄대학교 전기계 공학과(공학석사)
2020년 4월 ~ 2022년 11월 : 라온피플 AI SW 엔지니어
2022년 12월 ~ 현재 : 국방과학연구소 연구원
관심분야 : 영상처리, 딥러닝 표적/식별 추적