향상된 객체 인식을 위한 합성 개구면 레이더 영상 내 표적 및 그림자 영역의 융합 방안에 관한 연구

Fusion of Target and Shadow Regions in Synthetic Aperture Radar Imagery for Improved Object Recognition

초록

합성 개구면 레이더 (Synthetic Aperture Radar: SAR) 센서는 날씨 및 기후 조건에 상관없이 일정한 관측 품질을 보증하지만, 광학 및 적외선 센서에 비해 낮은 영상 해상도를 지니며, 넓은 그림자 영역을 수반하는 단점으로 인해 정확한 객체 인식이 불가능하다. 상기 문제를 해결하기 위해, 기존의 SAR 기반 객체 인식 기법들은 영상 내 그림자 영역을 사전에 제거하고, 표적 영역만 추출함으로써 정제된 정보만을 취득하는데 초점을 맞추었다. 반면, 본 논문에서는 그림자 정보를 역으로 활용하여 SAR 영상 내 정보량을 증폭하는 방안에 대하여 모색한다. 나아가 인공지능 알고리즘을 활용하여 표적 영역과 그림자 영역을 결합함으로써 객체 인식 정확도를 더욱 향상시킬 수 있는 방안을 제시한다.

Key Words : Automatic Target Recognition, Information Fusion of Target and Shadow (IFTS), Deep Learning (DL), Feature-level Fusion, Attention Mechanism, Synthetic Aperture Radar (SAR).

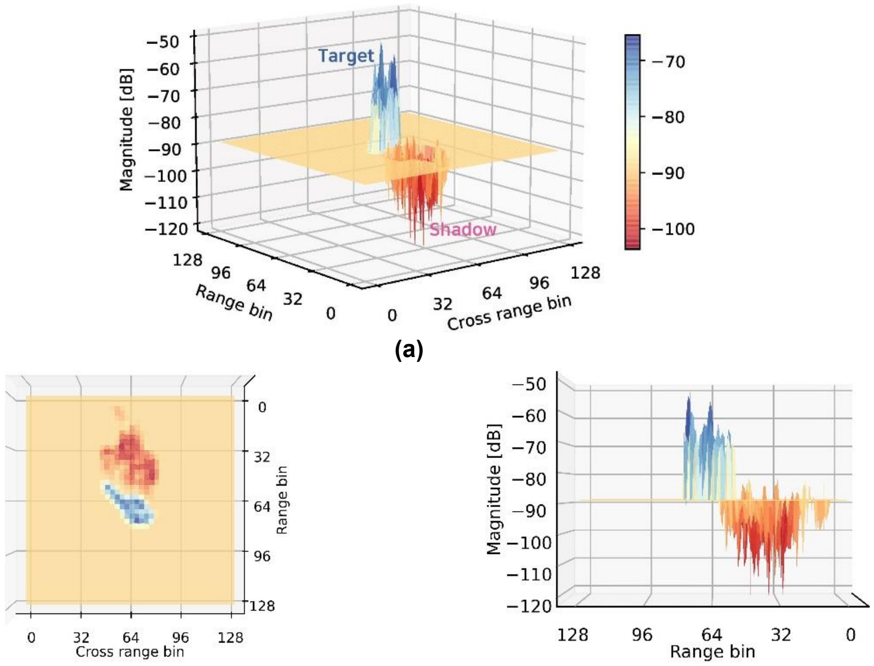
1. Introduction

Synthetic aperture radar (SAR) is an active imaging sensor characterized by its ability to generate high-resolution radar images in all-day, all-weather, and long-range conditions(1). Its unique electromagnetic properties compared with optical imagery allow it to be leveraged as a key information source in various surveillance and reconnaissance systems(2). However, on the other hand, its electromagnetic reflections also cause some complications such as differences in the scattering mechanism, anisotropic factors, contamination from clutter or jamming signals, and speckle noise, resulting in less intuitive visual interpretations(3~8). In this respect, the manual analysis of objects of interest from massive SAR image streams requires considerable human and material resources; this has consequently necessitated the development of SAR automatic target recognition (ATR) technologies.

Traditional SAR-ATR algorithms are primarily performed in three stages: detection, feature extraction, and classification(9~19). Detection is a process of identifying only the regions of interest (ROI) from a wide SAR scene and can generally be achieved using constant false alarm rate-based adaptive thresholding techniques(20). Subsequently, in the feature extraction stage, all the ROI templates are projected onto a lower-dimensional latent space such that each of them can be well aligned with respect to the corresponding category. Finally, some data-driven classifiers, such as AdaBoost, random forest, and support vector machine, can be exploited to automatically specify the detailed class of the target.

Among them, the feature extraction phase is typically regarded as one of the most essential and intractable factors in determining the ATR performance(21,22). Accordingly, most traditional SAR-ATR studies have focused on the manual construction of salient features suitable for SAR imagery, based on various signal processing algorithms such as scattering center extraction(12,13), advanced filtering(14), graphical modeling(15), image transformation(16), and compressed sensing(17~19). These approaches have realized significant achievements, but still retain clear drawbacks, i.e., their encoding pipelines remain computationally inefficient; in particular, the improvement rate of recognition accuracy saturates gradually because of the intrinsic limitations of heuristics in handcrafted feature (HF) engineering.

The advent of deep learning (DL) frameworks in the pattern analysis field has enabled the automatic formation of optimal feature descriptors from arbitrary raw input data(23~26). Inspired by this attractive property, a large body of studies have attempted for successful application of the DL framework to SAR-ATR(3,27~39), yielding remarkable performance enhancements compared with previous HF-based approaches(9~19). Despite a few methodological differences among these approaches, they share a common principle, i.e., to develop SAR-friendly DL-based algorithms by introducing an inductive bias associated with the unique domain characteristics of SAR imagery into the DL model. Chen *et al.*(3) proposed a parameter-efficient network architecture fully composed of convolutional modules to compensate for insufficient labeled SAR data. Moreover, in terms of input data for the network, SAR-specific data augmentation methods were presented using several image transformation techniques(27,28) or generative adversarial networks(29,30); while in terms of network training, domain adaptation-based learning strategies were suggested for transferring semantic representations from large-size optical data to SAR(31~33). To fully utilize the spatiotemporal properties of SAR sequences, convolutional neural network (CNN)-recurrent neural network (RNN) or three-dimensional (3D) CNN topologies were adopted(34~36). Meanwhile, to leverage the phase characteristics of SAR more effectively, Zhang *et al.*(37) introduced a complex CNN architecture. In recent years, to solve reliability issues arising from the mixed learning of clutter information within SAR training templates(38,40,41), novel SAR preprocessing schemes that enable the pre-removal of clutter components have also been proposed(38,39).



(b) (c)

Fig. 1. Electromagnetic amplitude levels in SAR imagery. Target and shadow regions are marked in blue and red, respectively, and clutter pixels are excluded for visual clarity. (a) 3D view. (b) Top view. (c) Side view.

The studies above demonstrate the importance of inducing a network to better reflect SAR domain properties such that ATR tasks can be performed successfully, instead of employing the typical DL algorithm developed for optical images as it is. Meanwhile, there also exists an additional domain characteristic of SAR that represents a distinct difference from other sensing modalities, i.e., the shadow information. Specifically, whereas electro–optical or infrared sensors form target images with a direct downward view, SAR signals are obtained from a slant transceiving path(42,43), which inevitably yields relatively wide shadow areas in the resultant images(27,44). Such a shadow region in the SAR image contains backprojected profiles of the object configuration and hence can be utilized for ATR, as with a target region(45~50). However, despite the useful information from the shadow domain, attempts to simultaneously use the target and shadow reflections for achieving further improvements in ATR performance are highly limited. To the best of our knowledge, with regard to the recent DL-based SAR-ATR approaches, the integration of the deep neural network (DNN)'s capability in automatic feature extraction with the information fusion of target and shadow (IFTS) has not been investigated.

The difficulty in incorporating the DL framework into the IFTS is primarily attributable to the inherent nature of shadow areas, which is clearly different from that of the target images. First, as shown in Fig. 1, shadow regions in SAR typically represent significantly lower intensity levels compared with the target or clutter regions. In addition, unlike the target area, whose internal pixels are composed of positive scattering peaks, most of the shadow pixels are based on negative scattering peaks. Nevertheless, conventional DNN architectures are not suitable for extracting complementary feature representations from each region, considering the contradictory characteristics of the target and shadow, but rather are configurated to focus only on target region. Consequently, semantic signatures from shadows are highly likely to be suppressed during the max-pool-based feature encoding in a general DNN model(51). Second, the shadow area in a SAR image is highly sensitive to the variation in depression angles, compared with the target area(27,52). Correspondingly, under practical conditions with significantly different depression angles between the training and test SAR data, the shadow information may not be beneficial at all; in fact, it may deteriorate the ATR performance.

Considering the aforementioned problems, unlike typical DL-based SAR-ATR approaches, which focus only on extracting salient features from target areas, we propose a novel framework that can achieve a successful IFTS through a parallelized processing scheme and adaptive multistage feature fusion, thereby affording further improved ATR performance. The main contributions of this study are threefold, as follows:

1) A major problem in current DL-based SAR-ATR methods is their incapability in performing independent processing specialized for each domain modality of targets and shadows, since their structural limitations cause each region to undergo uniform preprocessing and feature encoding. To tackle this problem, we propose a novel parallelized SAR processing pipeline, in which target and shadow areas are first segmented from an input SAR image, followed by parallel preprocessing and DNN-based feature encoding specifically customized for each region. Subsequently, the deep representations extracted from each region are fused to produce a single final decision. Consequently, domain-centric processing of target and shadow from a single image can be realized; in particular, the utility of shadow information can be improved significantly, even under the situations involving variable depression angles.
2) The target and shadow regions in a SAR template contain overlapping information and contribute differently from the perspective of ATR. In this respect, general feature fusion strategies(53,54) such as lateral concatenation and element-wise sum are not feasible for accommodating such information imbalances. Hence, we implement a novel domain fusion module (DFM), which induces the network to assign an adaptive weight in accordance with the influence of each region, thereby yielding complementary fusion. Moreover, a multistage fusion scheme is introduced to further improve the representation power of the fused feature.
3) It is noteworthy that the proposed IFTS framework is generic, so can easily be combined with other DL-based SAR-ATR techniques to strengthen their perception ability in the shadow region. Based on this flexibility, we investigate the performance gain by applying the proposed IFTS framework to various baseline DNN backbones, developed for image classification as well as SAR-ATR; the results confirm that our framework can activate each network to incorporate shadow information successfully and provide more precise recognition under both standard operation conditions (SOCs) and extended operation conditions (EOCs). To the best of our knowledge, this is the first study to report the efficacy of the IFTS for both SOCs and EOCs based on the DL approach.

The remainder of this paper is organized as follows: In Section 2, the effects of shadows in the current DL-based SAR-ATR techniques are investigated in various perspectives. Subsequently, we discuss the necessity of parallelized processing pipelines customized for targets and shadows separately. In Section 3, the methodology of the proposed IFTS framework is described in detail. Section 4 presents the experimental results under various conditions using a public moving and stationary target acquisition and recognition (MSTAR) dataset(52) to validate the effectiveness of the proposed method. Finally, concluding remarks are provided in Section 5.

2. Motivation

In this section, the intrinsic incompatibility of current DL-based SAR-ATR approaches in addressing SAR shadow content is experimentally demonstrated. Subsequently, we explore the solutions for a successful IFTS.

2.1. Necessity for Considering Shadow-Centric Processing in SAR-ATR

Belloni *et al.*(51) designed an evaluation protocol to numerically confirm the global contribution of the target, shadow, and clutter regions for DL-based ATR. By training a baseline CNN with partially segmented SAR regions, classification scores were observed for all possible combinations. Their experimental results notably indicated that the effect of shadows on SAR-ATR performance is insignificant, implying that a CNN is incapable of appropriately leveraging semantic representations from the shadow.

This undesirable phenomenon is due to the general CNN architectures centered solely on the feature extraction of the target, not the shadow. In other words, even though the shadow areas retain unique domain characteristics, which are exactly opposite to the target as discussed before (i.e., maintaining significantly lower intensity levels compared with the target or clutter, and comprising pixels with negative local peaks), the structural nature of the CNN causes final features to be extracted based on pixels with higher intensity levels and positive local peaks. Eventually, as shown by the experimental results of Belloni *et al.*(51), a typical CNN is likely to concentrate only on a target region, thereby causing the loss of useful discriminatory information within a shadow.

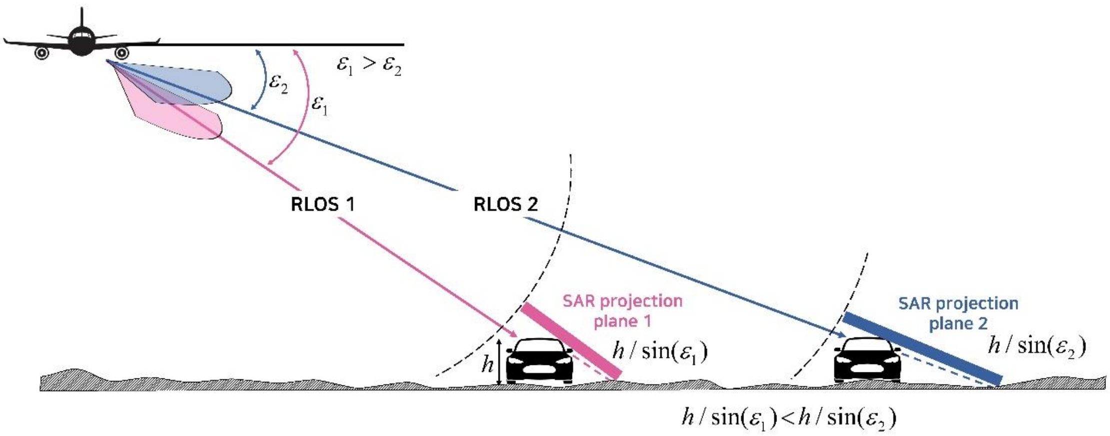


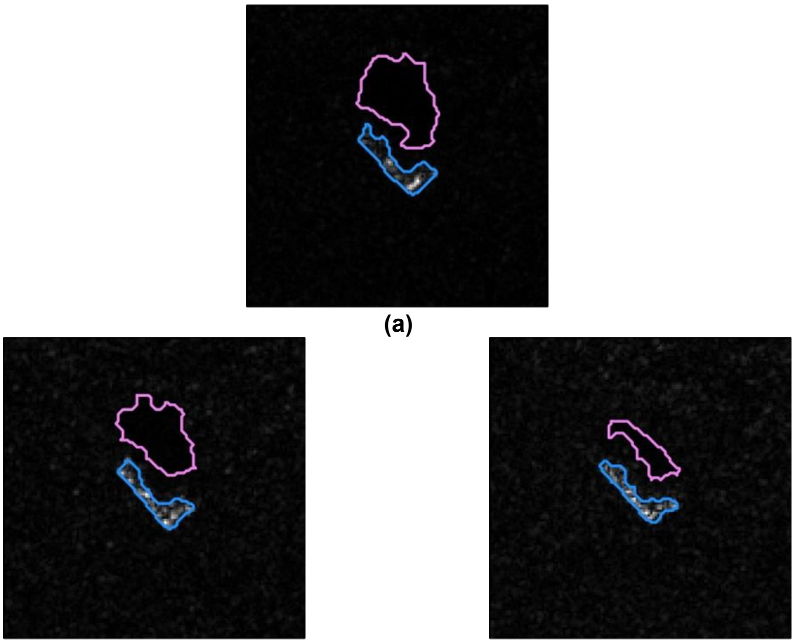
Fig. 2. Conceptual figure of SAR imaging system for different RLOS conditions (i.e., different depression angles).

It is noteworthy that the effect of shadows on ATR becomes further aggravated under EOCs, where the depression angle during the test session varies considerably compared with the training data. As illustrated in Fig. 2, a SAR system projects each object onto a slant image projection plane spanned by the radar line of sight (RLOS). During slant projection, a geometric distortion in the form of a scale transformation is inevitably incurred across the resulting SAR imagery. The problem is that the degree of such image distortion differs between the target and shadow regions, even within a single SAR image reflected from the same object(27). Formally, let the depression angle between a SAR platform and an object of interest be denoted as ε. Then, considering the spatial geometry, the target region in the SAR projection plane is scaled with a factor of $\cos(\varepsilon)$ into the range direction, whereas the shadow is scaled with a relatively large factor of $1/\sin(\varepsilon)$(27,52). For example, under the condition of EOC-1 in an MSTAR dataset, where the training and test data are constructed based on depression angles of $17°$ and $30°$ (a detailed description of the dataset is provided in Section 4.1), respectively, relative scaling ratio of the target region within the test SAR template is computed as follows:

$$\frac{\cos(\varepsilon_{test})}{\cos(\varepsilon_{train})} = \frac{\cos 30°}{\cos 17°} \approx 0.91 \tag{1}$$

where ε_{train} and ε_{test} represent the depression angles in the training and test data, respectively. By contrast, a much greater degree of compression is generated in the shadow region, as follows:

$$\frac{1/\sin(\varepsilon_{test})}{1/\sin(\varepsilon_{train})} = \frac{\sin 17°}{\sin 30°} \approx 0.58 \qquad (2)$$



(a) (b) (c)

Fig. 3. Example SAR templates for 2S1 vehicle at same azimuth angle but different depression angles. Target and shadow regions in each image are indicated with blue and red contours, respectively. Depression angles of (a) $15°$, (b) $17°$, and (c) $30°$.

Fig. 3 shows the contours of the target and shadow on the SAR images captured from the same object at different depression angles (i.e., $15°$, $17°$, and $30°$). As expected, unlike the target areas outlined with blue lines, which represent almost similar configurations irrespective of the depression angle, the shadow areas with pink lines undergo severe image variation as the depression angle of SAR platform changes.

Table 1. ATR Performance (Test Accuracy, %) Based on Several Regional Combinations of Partially Segmented SAR Images, under SOC and EOC-1 Setups

	SOC (Training: $17°$, Test: $15°$, 10 class)			EOC-1 (Training: $17°$, Test: $30°$, 4 class)		
	Target +Shadow	Target Only	Shadow Only	Target +Shadow	Target Only	Shadow Only
AConvNet[3]	96.15%	95.12%	79.58%	91.67%	91.83%	36.24%
LM-BN-CNN[38]	97.11%	96.30%	80.24%	93.54%	93.41%	39.22%
ESENet[39]	97.08%	96.69%	80.58%	93.03%	93.57%	38.67%
Average	96.78%	96.04%	80.13%	92.75%	92.94%	38.04%

To numerically investigate the discussions above in terms of ATR performance, we measured the recognition accuracies with respect to regional combinations of partially segmented MSTAR SAR images (i.e., Target+Shadow, Target Only, or Shadow Only), similar to the experiments of Belloni *et al.*[51]. However, in our case, an additional evaluation in EOC-1 was conducted as well, employing several backbone networks that were specifically designed for SAR, i.e., A-ConvNet[3], LM-BN-CNN[38], and ESENet[39], and the results are summarized in Table 1. Based on the results under the ideal SOC, it can be observed that utilizing the shadow region alone even enables the networks to attain a stable ATR performance of approximately 80%, indicating that the shadow modality clearly contains some backprojected signatures for the objects of interest. Despite the semantic potentiality of shadows, the comparison between the case involving both a segmented target and shadow and the case involving the target alone indicates only a slight performance improvement of 0.74% on average. Especially under a more practical condition (EOC-1), it is remarkable that all network configurations suffer from severe performance degradation when trained and inferenced with only shadows (i.e., exhibiting a level of accuracy with slight difference from that of simply randomized outputs in the EOC-1's four-class classification task). This implies that the networks cannot extract any informative indicators from the shadow, particularly when different depression angles are involved between training and testing. Accordingly, DNNs based on both target and shadow information cannot benefit from the shadow; in fact, they yield rather degraded ATR performance compared with those based on the target only.

The results shown in Table 1 explicitly demonstrate that the current approaches cannot combine the shadow information in a proper manner and that they depend only on the target information for SAR-ATR. For a successful IFTS, a network must be able to capture the unique domain characteristics of shadows as well as compensate for the region-wise scaling distortion with respect to the variation in the depression angle; this implies that shadow-centric processing must be accompanied together in the overall SAR-ATR mechanism, in addition to the target area.

2.2. Necessity for Parallelized Processing Pipeline

In addition to the necessity for designing separate processing for SAR shadow, a fundamental bottleneck exists when combining the shadow-centric processing with conventional DL-based ATR algorithms. Because the target and shadow regions are entangled within a single SAR template, independent processing optimized specifically for each modality is not feasible for implementation. In other words, when shadow-centric processing is applied to a specified SAR, the target area in the image will also be affected, and vice versa.

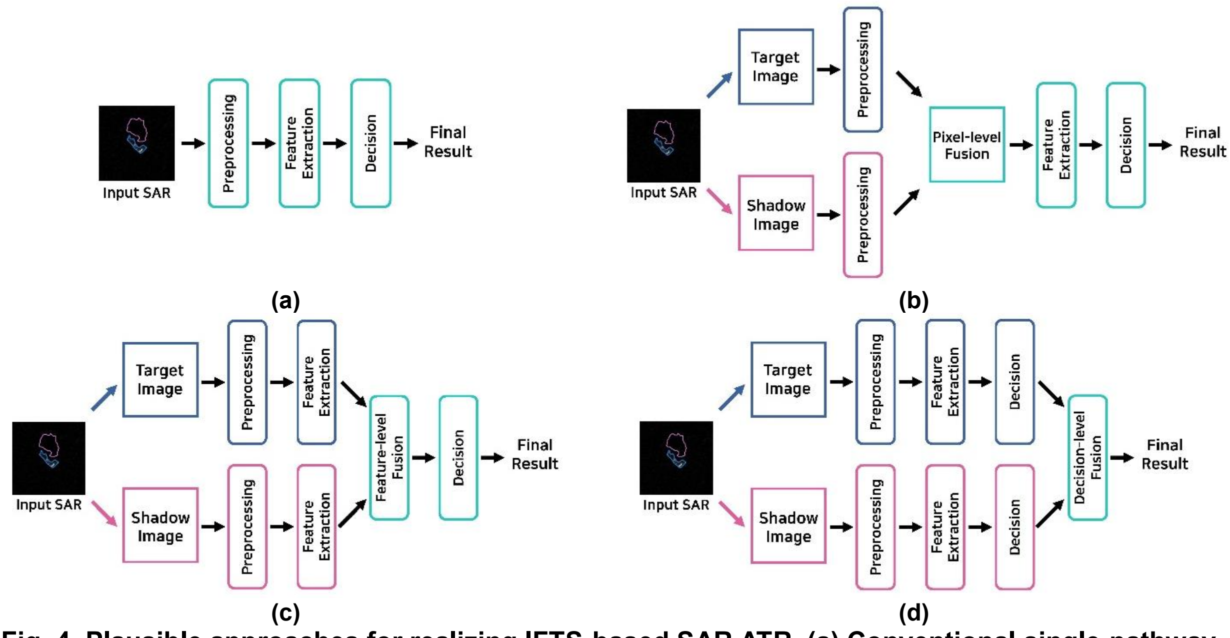


Fig. 4. Plausible approaches for realizing IFTS-based SAR ATR. (a) Conventional single-pathway-encoding pipeline. (b) Pixel-level fusion-based IFTS. (c) Feature-level fusion-based IFTS. (d) Decision-level fusion-based IFTS.

Essentially, this problem is attributable to the single-pathway-based pipeline of current ATR approaches, in which uniform preprocessing and deep feature encoding are applied based on a specified SAR input, as shown in Fig. 4(a). In this regard, parallelized processing pipelines for the target and shadow must be established to manage the problem and realize independent processing oriented toward each modality. To this end, we herein propose a novel SAR-IFTS framework that enables parallelized pipelines by regarding target and shadow regions within a single SAR as unique modalities. Specifically, the entangled target and shadow regions from the input SAR image are separated first using image segmentation techniques, followed by an independent processing suitable for each domain. Subsequently, the information pairs from the separated target and shadow are combined at a certain point to obtain a final single decision via a multimodal fusion scheme[55].

As shown in Figs. 4(b)-(d), the typical multimodal fusion algorithms can be categorized into three primary schemes based on the type of information to be combined: pixel-level, feature-level, and decision-level fusion. Among them, feature-level fusion [Fig. 4(c)] is expected to be the most suitable for performing the IFTS task, since pixel-level fusion [Fig. 4(b)] inevitably demands a sophisticated mechanism of integrating pixel-wise information pairs with high dimensionality and decision-level fusion [Fig. 4(d)] cannot readily consider hierarchical inter-relationships between the two modalities.

Hence, a parallel processing pipeline coupled with a feature-level fusion scheme is adopted in the proposed ATR framework to facilitate independent processing customized for the target and shadow, separately, while ensuring full benefits from multimodal fusion. In particular, the fusion is based on a newly developed DFM operation, which allows a network to extract complementary representations while considering the priority of each modality. In the next section, we present the methodology of the proposed framework in detail.

3. Methodology

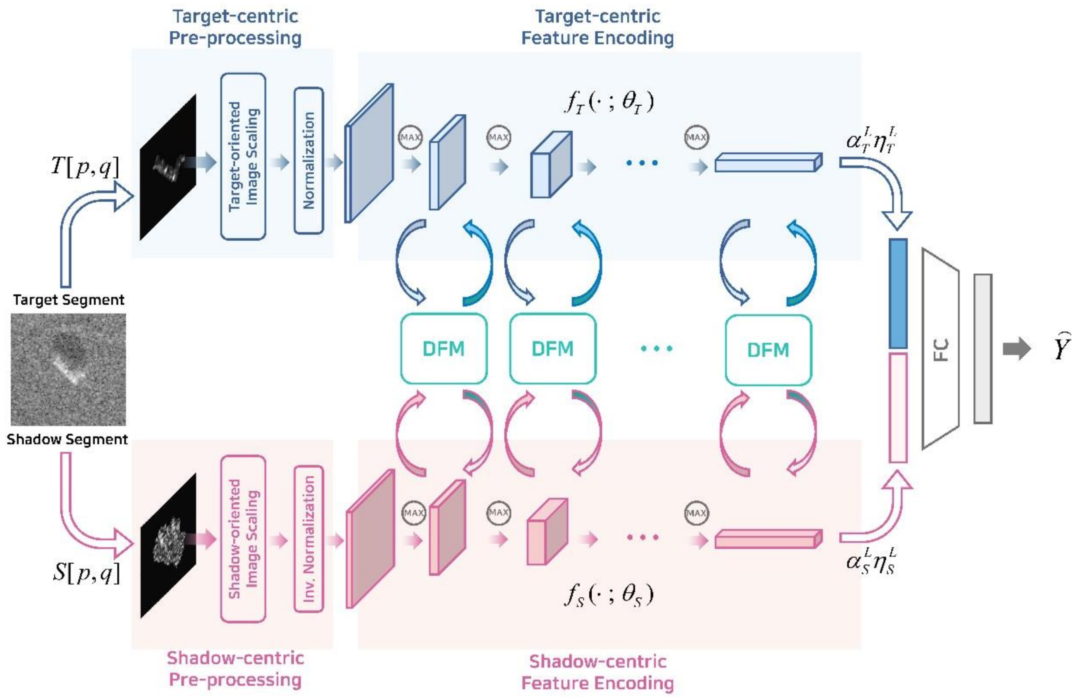


Fig. 5. Overall pipeline of the proposed SAR IFTS framework.

The overall concept of the proposed SAR-IFTS framework is presented in Fig. 5. As shown in the figure, the proposed framework first segments the target and shadow from a single SAR template, followed by image preprocessing and feature embedding specifically tailored for each region. Meanwhile, the representation pairs of the target and shadow extracted independently from the parallelized pipelines are combined based on novel feature fusion strategies to derive a single final decision. In this section, the stepwise procedures of the proposed framework are described in detail.

3.1. Segmentations of Target and Shadow Regions

Since a SAR image generally shows a mixture of backscattering reflections from the target, shadow, and clutter, separation of each component must be preceded to realize a parallel processing mechanism. Because the target is clustered in the high-intensity range, whereas the shadow is in the low-intensity range across the global SAR distribution, they can be separated using a series of image processing techniques, such as intensity-based binarization and morphological refinement. Inspired by the key idea of several segmentation techniques(16,38,56,57), we re-established a SAR segmentation algorithm aimed at separating target and shadow regions from a single SAR image.

Let a SAR image template be denoted as $I[m,n]$ with $1 \leq m \leq M$ and $1 \leq n \leq N$, where (m,n) are the coordinates on the down-range and cross-range dimensions, respectively. Because radar reflection represents different levels of electromagnetic intensity depending on its RLOS range between the mounting platform and targets of interest, the intensity variation in the SAR image must first be adjusted.

$$I_v[m,n] = \frac{I[m,n]}{\sum_{m=1}^{M}\sum_{n=1}^{N} I[m,n]}, \ \forall(m,n), \tag{3}$$

Based on the adjusted SAR template $I_v[m,n] \in \mathbb{R}^{M \times N}$, the target and shadow regions can be segmented via the following procedures:

Step 1: Select only pixels that correspond to the upper 3% intensity from the entire histogram of $I_v[m,n]$ to generate a binarized target mask (i.e., 1 for the target and 0 for the remainder). Likewise, select only the pixels with the lower 25% amplitude to create a corresponding binary shadow mask.

Step 2: Perform counting filtering for each mask such that spurious pixels can be suppressed to the maximum extent.

Step 3: Morphological closing is applied to bind the areas of interest and smooth the edge components.

Step 4: Obtain the final refined binary mask by extracting only the max-connected region.

Step 5: Multiply each binary mask with $I_v[m,n]$ to segment the target and shadow pixels. Subsequently, obtain the final target image $T[p,q] \in \mathbb{R}^{P \times Q}$ and shadow image $S[p,q] \in \mathbb{R}^{P \times Q}$ by cropping a rectangular area with a width of P and a height of Q around the center of mass from each region.

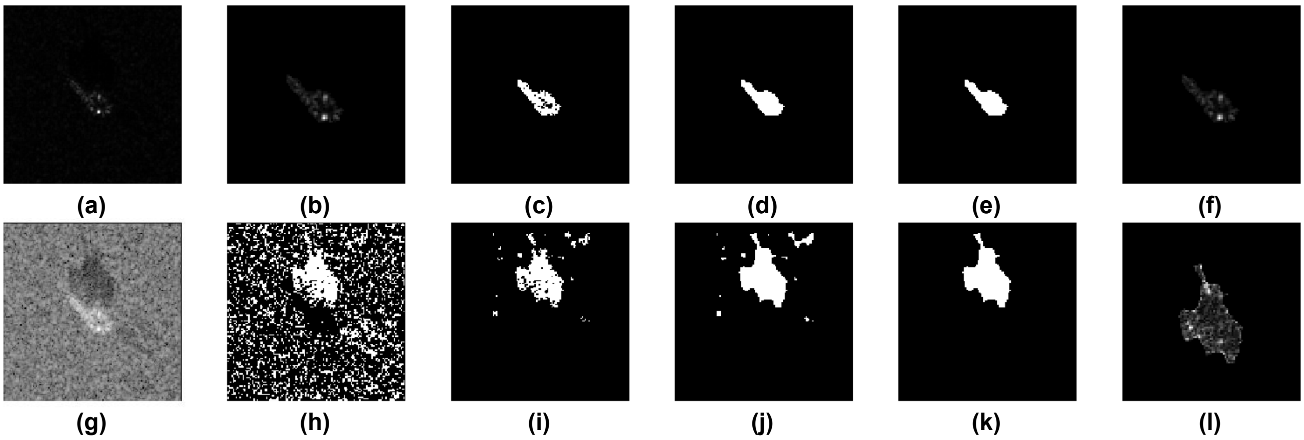


Fig. 6. Stepwise output examples of the proposed segmentation technique for SAR template. (a) Original SAR image for T-72 tank in linear scale. (b) Target mask after upper thresholding. (c) Target mask after counting filtering. (d) Target mask after morphology. (e) Refined target mask obtained by selecting only the max-connected region. (f) Final segmented target image. (g) Original SAR image for T-72 tank in log scale. (h) Shadow mask after lower thresholding. (i) Shadow mask after counting filtering. (j) Shadow mask after morphology. (k) Refined shadow mask obtained by selecting only the max-connected region. (l) Final segmented shadow image.

To provide a clear illustration, Fig. 6 shows the stepwise outputs of the proposed target/shadow segmentation process based on the SAR image chip of a T-72 tank. It is noticed that the target and shadow regions can be roughly be separated through hard thresholding [Figs. 6(b) and 6(h)] from $I_v[m,n]$, and then gradually refined by counting filtering [Figs. 6(c) and 6(i)] and morphological adjustments [Figs. 6(d) and 6(j)]. Finally, the segmented images for target $T[p,q]$ and shadow $S[p,q]$ can be obtained by multiplying each binary mask with $I_v[m,n]$ in an element-wise manner and readjusting the center points, as shown in [Figs. 6(f) and 6(l)], respectively.

3.2. Parallel Processing Customized for Target and Shadow Regions

The core of the proposed SAR-ATR framework lies in the novel concept of the parallelized processing mechanism (Fig. 5), which can structurally ensure independent processing for $T[p,q]$ and $S[p,q]$ by regarding each of them as a disparate input modality. This, in turn, enables a flexible implementation of the preprocessing and feature embedding pipelines customized for each modality, while accounting for their characteristic differences. It is noteworthy that the target and shadow show distinct characteristic differences in two major aspects, as mentioned in Section 2.1: 1) in cases where the depression angle between the training and test conditions fluctuates, each region undergoes different degrees of image distortion in the form of scale transformation; 2) when approaching the electromagnetic scattering centers of an object, the reflected amplitude levels in the target region tend to increase, whereas the amplitudes of the shadow region gradually decrease. The two domain discrepancies are first compensated using our parallelized preprocessing techniques.

3.2.1. Preprocessing for Target and Shadow Regions

Recall that the target and shadow images are scaled across the down range dimension with a factor of $\cos(\varepsilon)$ and $1/\sin(\varepsilon)$, respectively, with respect to the depression angle ε between a SAR platform and an object of interest. Hence, when geometrical adjustment is performed based on the target area, the shadow area would undergo another form of scaling distortion as well, and vice versa. Now that the independent processing of the target and shadow is guaranteed by the parallelized mechanism, we can mitigate the scaling distortion of both $T[p,q]$ and $S[p,q]$ using the region-wise scaling factor. To this end, we adopt a general affine transformation technique. Let the depression angle $\varepsilon_{\text{test}}$ in the test environment be changed from the training depression angle ε_{train} (i.e., $\varepsilon_{test} \neq \varepsilon_{train}$). Then, for the testing conditions, we transform $T[p,q]$ and $S[p,q]$ in the (p,q) Cartesian coordinates to $T_c[p',q']$ and $S_c[p',q']$ in the rescaled (p',q') Cartesian coordinates, respectively, where the (p,q) and (p',q') coordinates are correlated as follows[27]:

$$\begin{pmatrix} p' \\ q' \end{pmatrix} = \begin{pmatrix} \lambda & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix}, \tag{4}$$

where λ denotes the recalibrating parameter, which is obtained by the inverse of the scaling distortion within each test region.

$$\lambda = \begin{cases} \dfrac{\cos(\varepsilon_{train})}{\cos(\varepsilon_{test})}, & \text{for target} \\ \dfrac{1/\sin(\varepsilon_{train})}{1/\sin(\varepsilon_{test})}, & \text{for shadow} \end{cases} \tag{5}$$

In summary, in the proposed IFTS framework, the input pair of segmented images $T[p,q]$ and $S[p,q]$ are used without modification for the training phase, and the rescaling transformation in (4) is additionally applied for the test phase such that even the test input pairs under ε_{test} become consistent with the training depression angle ε_{train}.

Next, in terms of image intensity, we compensate for the conflicting domain characteristics of the target and shadow through region-wise normalization, ensuring consistent statistical distributions as well as appropriate dynamic range levels corresponding to each region. Similar to the region-wise rescaling, normalization is also performed in a parallelized manner to readily manage the unique distribution of each modality. For example, $T_c[p',q']$ and $S_c[p',q']$ in the test phase are normalized as follows:

$$T_n[p',q'] = \begin{cases} \dfrac{T_c[p',q'] - \mu_T}{\sigma_T}, & \forall (p',q') \in \mathbb{I}_T \\ \min_{p',q'} \left\{ \dfrac{T_c[p',q'] - \mu_T}{\sigma_T} \right\}, & \forall (p',q') \notin \mathbb{I}_T \end{cases} \tag{6}$$

$$S_n[p',q'] = \begin{cases} -\left(\dfrac{S_c[p',q'] - \mu_S}{\sigma_S} \right), & \forall (p',q') \in \mathbb{I}_S \\ \min_{p',q'} \left\{ -\left(\dfrac{S_c[p',q'] - \mu_S}{\sigma_S} \right) \right\}, & \forall (p',q') \notin \mathbb{I}_S \end{cases} \tag{7}$$

where

$$\mathbb{I}_T = \{ (p',q') \mid T_c[p',q'] \neq 0 \},$$
$$\mathbb{I}_S = \{ (p',q') \mid S_c[p',q'] \neq 0 \}.$$

Here, $T_n[p',q'] \in \mathbb{R}^{P\times Q}$ and $S_n[p',q'] \in \mathbb{R}^{P\times Q}$ represent the final preprocessed images for the target and shadow, respectively. In addition, μ_T and μ_S denote the sample means of the SAR images in the target and shadow regions, respectively; σ_T and σ_S denote sample variances for the corresponding regions, respectively. Note that the normalized output of the shadow is reversed considering its inverse characteristics, as shown in Fig. 1 (we refer to it as “inverse normalization” hereinafter); as such, consistent with the target image, the negative peaks of $S_c[p',q']$ can be converted into positive peaks.

3.2.2. DNN-based Feature Encoding for Target and Shadow Regions

Considering that the preprocessed images $T_n[p',q']$ and $S_n[p',q']$ separately involve the signatures of the object of interest, a deep CNN model can be utilized for an automatic extraction of the complementary representation from each modality. In this subsection, we formulate a detailed methodology to extract a domain-specific feature representation based on each preprocessed image, given an arbitrary CNN model.

A general CNN architecture is configured to iterate the nonlinear embedding and downsampling operations through the hierarchical combinations of convolutional mapping and pooling modules, thereby effectively deriving the global and local features from a high-dimensional input[58]. By unifying various combinations of internal network modules and their hierarchical connections, numerous CNN topologies can be constructed.

In this study, we do not focus on the topology of the deep network itself. Instead, we attempt to identify a method to convert a specified CNN-based feature extractor into an extended model aimed at the IFTS task. Let $f(\cdot\,;\theta)$ denote a CNN-based extractor with internal parameters θ, which is designed to project an arbitrary input image $A \in \mathbb{R}^{P\times Q}$ into the latent space $f(A\,;\theta) = \{\eta^1, \eta^2, \ldots, \eta^L\}$, where η^l represents the feature from the l-th encoding layer. Then, we duplicate it for extension to the parallelized encoding pair $f_T(\cdot\,;\theta_T)$ and $f_S(\cdot\,;\theta_S)$, which exhibit an identical network topology with $f(\cdot\,;\theta)$ but different internal parameters such that each pathway (or subnetwork) can be organized to address the multimodal input pair T_n and S_n, respectively (Fig. 5). Specifically, for a set of SAR images $\mathcal{I}$, a SAR template $I[m,n] \sim \mathcal{I}$ uniformly sampled from $\mathcal{I}$, and two preprocessed region-wise images $T_n \sim \mathcal{T}$ and $S_n \sim \mathcal{S}$, the parallelized encoding pipelines of $f_T(\cdot\,;\theta_T)$ and $f_S(\cdot\,;\theta_S)$ allow each subnetwork to be individually optimized from $\mathcal{T}$ and $\mathcal{S}$, thereby yielding the feedforward generation of region-wise feature pairs $f_T(T_n;\theta_T) = \{\eta_T^1, \eta_T^2, \ldots, \eta_T^L\}$ (for target) and $f_S(S_n;\theta_S) = \{\eta_S^1, \eta_S^2, \ldots, \eta_S^L\}$ (for shadow).

It is noteworthy that our pipeline enables a network to extract feature specifically oriented toward each conflicting region; but on the other hand, there remains an additional issue of how to properly combine the representation pairs from the target and shadow to derive a single final decision. In the next subsection, we introduce novel fusion strategies for the target and shadow, including the DFM and multistage fusion scheme.

3.2.3. Adaptive Fusion of Target and Shadow Features

In SAR imagery, the target and shadow regions have imbalanced significance for ATR. Namely, despite the usefulness of shadows, they definitely retain a lower level of content information than the target areas. In this respect, the region-wise features η_T and η_S must be adaptively incorporated with unequal weight ratios.

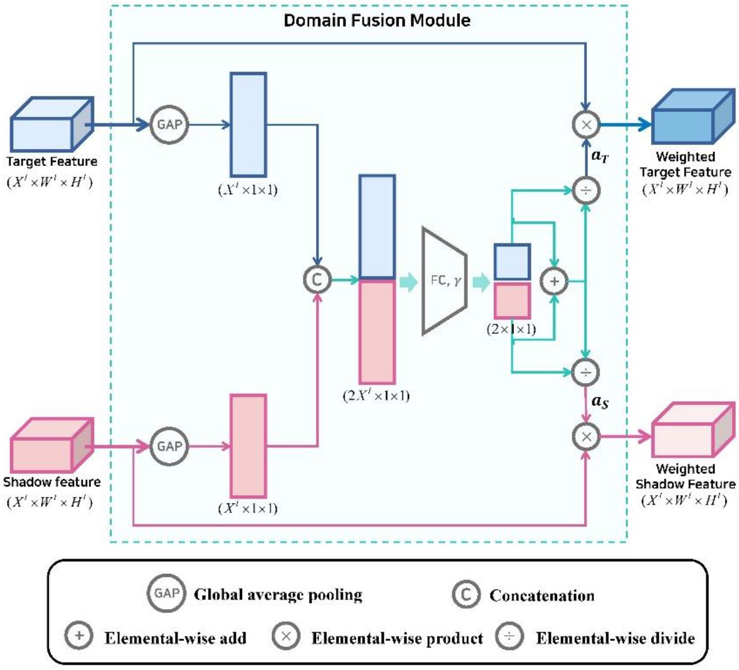


Fig. 7. Detailed workflow of DFM.

Instead of determining the weight for each modality empirically, we allow the network to assign adaptive weights on its own, through a novel DFM. As illustrated in Fig. 7 the DFM is configured to take the feature pair of a certain layer $\{\eta_T^l, \eta_S^l\}$ as the input and then compute the corresponding weight ratio $\{\alpha_T^l, \alpha_S^l\}$ using the attention mechanism. Specifically, the 3D input features $\eta_T^l \in \mathbb{R}^{X^l \times W^l \times H^l}$ and $\eta_S^l \in \mathbb{R}^{X^l \times W^l \times H^l}$ are compressed first via a global average pooling operation to form one-dimensional vectors $\mathbf{z}_T^l \in \mathbb{R}^{X^l}$ and $\mathbf{z}_S^l \in \mathbb{R}^{X^l}$:

$$\{\mathbf{z}_T^l\}_x = \frac{1}{W^l \times H^l} \sum_{i=1}^{H^l} \sum_{j=1}^{W^l} \eta_T^l[x, i, j], \quad (8)$$

$$\{\mathbf{z}_S^l\}_x = \frac{1}{W^l \times H^l} \sum_{i=1}^{H^l} \sum_{j=1}^{W^l} \eta_S^l[x, i, j], \quad (9)$$

where $\{\cdot\}_x$ denotes the x-th element of the compressed feature vector. By concatenating $\mathbf{z}_T^l$ and $\mathbf{z}_S^l$ to form $\mathbf{z}_F^l \in \mathbb{R}^{2X^l}$, the weight value corresponding to each modality can be inferred through fully connected (FC) operation and sigmoid mapping, as follows:

$$[\omega_T^l, \omega_S^l] = \gamma\left(\mathbf{W}^l \mathbf{z}_F^l + \mathbf{b}^l\right), \quad (10)$$

Here, $\mathbf{W}^l$ and $\mathbf{b}^l$ represent the trainable weight and bias of the FC operation, respectively, and $\gamma(\cdot)$ refers to the sigmoid activation function[58]. Finally, ω_T^l and ω_S^l are normalized to obtain the fusion ratio for the target (i.e., $\alpha_T^l = \omega_T^l/(\omega_T^l + \omega_S^l)$) and shadow (i.e., $\alpha_S^l = \omega_S^l/(\omega_T^l + \omega_S^l)$).

In particular, we do not confine the application of the DFM to a specific layer; instead, we allow it to be leveraged across multiple encoding layers such that the network can further benefit from the effect of adaptive fusion. During the parallel encoding of the target and shadow, we allocate the DFM-driven adaptive weights for each layer where the region-wise features become resampled via the pooling operation, expressed as follows:

$$\eta_T^l \longleftarrow \alpha_T^l \eta_T^l \quad \forall l \in \mathbb{L}, \tag{11}$$

$$\eta_S^l \longleftarrow \alpha_S^l \eta_S^l \quad \forall l \in \mathbb{L}, \tag{12}$$

where $\mathbb{L}$ represents the set of the layers right after pooling operations in f. Now that the adaptive weights for the target and shadow regions are guaranteed across multiple layers, ATR can be performed by concatenating the final latent $\alpha_T^L \eta_T^L$ and $\alpha_S^L \eta_S^L$ from each subnetwork, followed by the application of the FC layers and softmax classifier to the fused feature vector.

Algorithm 1 Main learning algorithm of the proposed SAR IFTS framework.

Input:
training SAR samples $\mathcal{I} = \{I^{(1)}, I^{(2)}, \ldots, I^{(K)}\}$,
corresponding label $\mathcal{Y} = \{Y^{(1)}, Y^{(2)}, \ldots, Y^{(K)}\}$,
batch size B, network structure f.
Step one: Image segmentation & pre-processing
for $I^{(k)}[m,n] \in \mathcal{I}$ **do**
  for all $k \in \{1, \ldots, K\}$ **do**
    Compute $I_v^{(k)}[m,n]$ by (3).
    Segment target $T^{(k)}[p,q]$ and shadow $S^{(k)}[p,q]$ from $I_v^{(k)}[m,n]$ as in Section III-*A*.
    Compute $T_n^{(k)}[p,q]$ by (6). ▹ *Normalization*
    Compute $S_n^{(k)}[p,q]$ by (7). ▹ *Inv. normalization*
  end for
end for
return target samples $\mathcal{T} = \left\{T_n^{(1)}, T_n^{(2)}, \ldots, T_n^{(K)}\right\}$ and shadow samples $\mathcal{S} = \left\{S_n^{(1)}, S_n^{(2)}, \ldots, S_n^{(K)}\right\}$.
Step two: Network training
Construct f_F, composed of $f_T(\cdot;\theta_T)$, $f_S(\cdot;\theta_S)$, and DFMs, referring to the given structure f.
Initialize θ_T, θ_S, θ_F.
for sampled minibatch $\mathcal{B} \subset \{1, \ldots, K\}$, $\left\{T_n^{(b)}\right\} \subset \mathcal{T}$, $\left\{S_n^{(b)}\right\} \subset \mathcal{S}$, $\{Y^{(b)}\} \subset \mathcal{Y}$ **do**
  for all $b \in \mathcal{B}$ **do**
    $P(\tilde{y} \mid T_n^{(b)}, S_n^{(b)}) = f_F(T_n^{(b)}, S_n^{(b)}; \theta_T, \theta_S, \theta_F)$. ▹ *Parallel encoding*
  end for
  $\mathcal{L} = -\frac{1}{B} \sum_{b \in \mathcal{B}} \log P(\tilde{y} = Y^{(b)} \mid T_n^{(b)}, S_n^{(b)})$.
  $(\theta_T^*, \theta_S^*, \theta_F^*) = \underset{\theta_T, \theta_S, \theta_F}{\text{argmin}}\ \mathcal{L}$.
end for
return trained network $f_F(\cdot; \theta_T^*, \theta_S^*, \theta_F^*)$.

Algorithm 2 Inference algorithm of the proposed SAR IFTS framework.

Input:

test SAR sample $I[m,n]$,

depression angle for training ε_{train} and test ε_{test},

trained network $f_F(\cdot\,; \theta_T^*, \theta_S^*, \theta_F^*)$.

Step one: Image segmentation & pre-processing

Compute $I_v[m,n]$ by (3).

Segment the target $T[p,q]$ and shadow $S[p,q]$ from $I_v[m,n]$ as in Section III-*A*.

Compute $T_c[p',q']$ and $S_c[p',q']$ by (4). ▷ *Region-wise rescaling*

Compute $T_n[p',q']$ by (6). ▷ *Normalization*

Compute $S_n[p',q']$ by (7). ▷ *Inv. normalization*

return $T_n[p',q']$ and $S_n[p',q']$.

Step two: Network-based inference

$P(\tilde{y} \mid T_n, S_n) = f_F(T_n, S_n\,; \theta_T^*, \theta_S^*, \theta_F^*)$. ▷ *Parallel encoding*

$\tilde{Y} = \underset{Y}{\operatorname{argmax}}\, P(\tilde{y} = Y \mid T_n, S_n)$.

return recognized output $\tilde{Y}$.

To train the overall IFTS network $f_F(\cdot\,;\theta_T,\theta_S,\theta_F)$, θ_T in the target-centric subnetwork $f_T(\cdot\,;\theta_T)$, θ_S in the shadow-centric subnetwork $f_S(\cdot\,;\theta_S)$, and the parameters θ_F corresponding to multistage DFMs must be considered comprehensively; all of them can be optimized in an end-to-end manner. The detailed learning and inference algorithms of the proposed IFTS framework are shown in Algorithms 1 and 2, respectively.

4. Experimental Results

4.1. Dataset Description

Fig. 8. Optical (left) and SAR (right) images for 10 different ground vehicles.

To evaluate the proposed SAR-IFTS framework, we used the public MSTAR dataset[52] as a benchmark, which was established under the joint support of the Defense Advanced Research Projects Agency (DARPA) and the Air Force Research Laboratory (AFRL). The collection was based on the Sandia National Laboratory SAR sensor platform for 10 different categories of ground military vehicles (armored personnel carrier: BMP-2, BRDM-2, BTR-60, and BTR-70; tank: T-62, T-72; air defense unit: ZSU-234; bulldozer: D-7; rocket launcher:

2S1; truck: ZIL-131), which are shown in Fig. 8. For each category, the resulting SAR images were acquired from a full aspect coverage (i.e., from $0°$ to $360°$ azimuth angle varying at an interval of $5°$ to $6°$), with a spatial resolution of 0.3 m × 0.3 m and a size of 128 × 128 pixels.

Table 2. Training and Test SAR Samples under SOC Experimental Setup

Class	Serial No.	Training		Test	
		Depression	Number	Depression	Number
BRDM-2	E-71	17°	298	15°	274
BTR-60	7532	17°	256	15°	195
BTR-70	C71	17°	233	15°	196
T-62	A51	17°	299	15°	273
ZSU-234	d08	17°	299	15°	274
D-7	13015	17°	299	15°	274
2S1	B01	17°	299	15°	274
ZIL-131	E12	17°	299	15°	274
BMP-2	9563	17°	233	15°	195
	9566	17°	232	15°	196
	C21	17°	233	15°	196
T-72	132	17°	232	15°	196
	812	17°	231	15°	195
	S7	17°	228	15°	191

Table 3. Training and Test SAR Samples under EOC-1 Experimental Setup

Class	Serial No.	Training		Test	
		Depression	Number	Depression	Number
BRDM-2	E-71	17°	298	30°	287
ZSU-234	d08	17°	299	30°	288
2S1	B01	17°	299	30°	288
T-72	132	17°	232	-	-
	812	17°	231	-	-
	S7	17°	228	-	-
	A64	-	-	30°	288

Table 4. Training and Test SAR Samples under EOC-2 Experimental Setup

Class	Serial No.	Training		Test	
		Depression	Number	Depression	Number
BRDM-2	E-71	17°	298		-
BTR-70	C71	17°	233		-
BMP-2	9563	17°	233		-
T-72	132	17°	232		-
	S7	-	-	15°, 17°	419
	A32	-	-	15°, 17°	572
	A62	-	-	15°, 17°	573
	A63	-	-	15°, 17°	573
	A64	-	-	15°, 17°	573

For a comprehensive evaluation of various scattering scenarios, the MSTAR dataset can mainly be divided into two setups depending on the operating conditions: the SOC and EOC. The SOC is defined as a 10-class SAR classification problem for ground objects measured from exactly the same target configurations and serial numbers, as well as at similar depression angles, reflecting almost ideal scenarios. The EOC setups are designed to assess the performance under more practical scattering conditions than the SOC and can further be categorized into three different variants, i.e., EOC-1, EOC-2, and EOC-3, each of which reflects the scenario of significant depression angle change, target configuration variance, and version variance, respectively. Detailed target information and the number of available SAR templates in each dataset setup are listed in Tables 2, 3, 4, and 5.

Table 5. Training and Test SAR Samples under EOC-3 Experimental Setup

Class	Serial No.	Training		Test	
		Depression	Number	Depression	Number
BRDM-2	E-71	17°	298	-	-
BTR-70	C71	17°	233	-	-
BMP-2	9563	17°	233	-	-
	9566	-	-	15°, 17°	428
	C21	-	-	15°, 17°	429
T-72	132	17°	232	-	-
	812	-	-	15°, 17°	426
	A04	-	-	15°, 17°	573
	A05	-	-	15°, 17°	573
	A07	-	-	15°, 17°	573
	A10	-	-	15°, 17°	567

To segment the target and shadow regions from a specified MSTAR SAR chip (Section 3.1), we utilized a 5 × 5 counting filter with a threshold of 15, and a 5 × 5 morphological image mask with its vertex pixels set to 0, as follows:

$$\begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

The final segmented size of each region, i.e., (P, Q), was set to $(96{,}96)$.

4.2. Experimental Setup

Note that the proposed IFTS framework is generic, so it can be easily combined with other classification models regardless of the network topology. Hence, we comprehensively investigated the ATR performance based on various network structures, including backbone models specialized for SAR imagery, such as AConvNet(3), LM-BN-CNN(38), and ESENet(39), as well as baseline models developed in the image classification field, such as AlexNet(23), VGGNet(59), and ResNet(60). By adding a global average pooling layer(60) to the last encoding layer (i.e., the layer right before the FC layers) and modifying the channel size of the input convolutional kernel to 1, each network was adjusted appropriately to manage the SAR images of an arbitrary input size.

Each model was trained for 300 epochs using the adaptive moment estimation (Adam) optimizer(61) at a learning rate of 0.001 and a batch size of 128. All the experiments were implemented based on the framework of Pytorch 1.6, which was executed on an Intel i7-9800K CPU with an Nvidia Titan RTX GPU (24 GB memory) and 64 GB of RAM.

4.3. Results Under SOC

As listed in Table 2, the training and test datasets in the SOC setup comprise 10-class SAR images of the same target configurations and serial numbers captured from similar depression angles of $17°$ and $15°$, respectively. Under this ideal setup, we measured the ATR accuracy of each baseline network by applying/not applying the proposed IFTS framework. Namely, one is based only on the target regions, similar to current SAR-ATR approaches(17,38,39,62), whereas the other utilizes the target and shadow simultaneously through our IFTS-based encoding. Results are reported as the averages of five independent trials for reliability.

Table 6. Experimental Results under SOC Setup

Speciality	Backbone	# Params	Accuracy [%]
Target Only (SAR)	AConvNet[3]	117K	95.12
	LM-BN-CNN[38]	141K	96.30
	ESENet[39]	555K	96.69
Target Only (Optical Image)	AlexNet[23]	57M	94.83
	VGGNet16[59]	134M	94.67
	VGGNet19[59]	140M	94.23
	ResNet18[60]	11M	96.87
	ResNet34[60]	21M	96.72
	ResNet50[60]	24M	96.61
Target + Shadow	AConvNet + IFTS	235K	97.59 **(+2.47)**
	ESENet + IFTS	1.1M	*98.45* **(+1.76)**
	ResNet18 + IFTS	22M	**98.90** **(+2.03)**

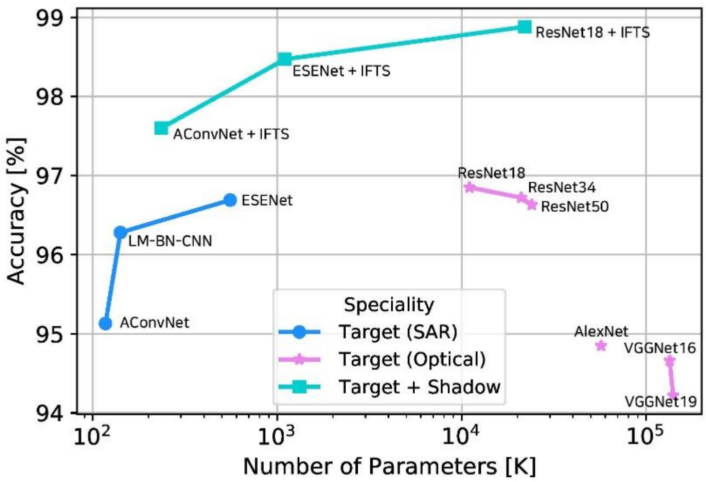


Fig. 9. SAR-ATR accuracy of various backbone models when trained with or without our IFTS framework, under SOC setup.

The ATR results for each model are summarized in Table 6. In the table, we emphasize the best performance by the bold-face font and the second-best performance by the italic font. To provide a clearer comparison of the results, we also present the ATR performance with respect to the parameter size of each backbone model, as shown in Fig. 9. By comparing the outcomes of conventional target-only encodings (top nine rows of Table 6), it is noticeable that the ATR techniques customized for SAR imagery[3,38,39] exhibit satisfactory performance even with small numbers of training parameters. By contrast, the backbone models specialized in optical data pursue deeper layers and larger parameter sizes to further improve the representation capability of the extracted features, thereby resulting in ResNet18 achieving the best ATR accuracy among the existing techniques. Meanwhile, the oversized network inevitably requires more training SAR data to prevent overfitting, and consequently indicates rather deteriorated performance over a certain number of training parameters. This demonstrates the fundamental limitations of current SAR-ATR approaches that attempt to improve the performance through changes in the model backbone, since data acquisition is substantially laborious and expensive particularly in SAR-based tasks.

Notably, applying the proposed IFTS framework allows all models to successfully exploit shadow modality and hence outperform each backbone counterpart to a large extent (the accuracy gain compared with its corresponding counterpart is also presented in the table). When coupled with the IFTS, even AConvNet can achieve an ATR performance superior to that of the previous best model based on target-only encoding, i.e., ResNet18+target-only encoding, using significantly fewer training parameters (approximately 47 times lower than that of ResNet18). ResNet18 coupled with the IFTS demonstrates state-of-the-art performance under the SOC setup, 2.03% higher than its counterpart. These results support our motivation to exploit shadow information together for improved ATR. Detailed ATR results for ResNet18+IFTS are shown in Table 7 in a confusion matrix format.

Table 7. ATR Results of ResNet18+IFTS in Confusion Matrix Form, under SOC Setup

Class	BRDM-2	BTR-60	BTR-70	T-62	ZSU-234	D-7	2S1	ZIL-131	BMP-2	T-72	Accuracy [%]
BRDM-2	268	3	1	0	0	0	0	1	1	0	97.81
BTR-60	5	189	0	0	1	0	0	0	0	0	96.92
BTR-70	0	0	194	0	0	0	2	0	0	0	98.98
T-62	0	0	0	269	1	0	0	1	0	2	98.53
ZSU-234	0	0	0	0	271	2	0	0	0	1	98.91
D-7	0	0	0	0	2	272	0	0	0	0	99.27
2S1	0	0	1	0	0	0	270	3	0	0	98.54
ZIL-131	0	0	0	0	3	0	0	270	0	1	98.54
BMP-2	0	1	0	0	0	0	0	0	583	3	99.32
T-72	0	0	0	1	0	0	0	0	0	581	99.83
Total											**98.88**

It should be noted that this study focused on the practicality and realization of the IFTS in the pattern analysis of radar imagery, i.e., we did not address detailed orthogonal factors that can further improve ATR performance, such as advanced segmentation algorithms of SAR imagery(63,64), domain adaptation(31~33), and modern network architectures(34~36,65). We believe that combining such factors with the proposed IFTS framework will further improve robustness and generality.

4.4. Results Under EOCs

Table 8. Experimental Results under EOC Setups

Speciality	Backbone	# Params	Accuracy [%] under EOC-1	under EOC-2	under EOC-3
Target Only (SAR)	AConvNet(3)	110K	91.83	88.23	87.25
	LM-BN-CNN(38)	141K	93.41	88.96	88.12
	ESENet(39)	555K	93.57	89.81	89.33
Target Only (Optical Image)	AlexNet(23)	57M	91.05	87.68	86.55
	VGGNet16(59)	134M	89.94	87.27	86.63
	VGGNet19(59)	140M	90.10	87.04	86.47
	ResNet18(60)	11M	93.75	90.21	90.31
	ResNet34(60)	21M	93.92	90.63	89.97
	ResNet50(60)	24M	93.66	90.44	89.83
Target + Shadow	AConvNet + IFTS	222K	95.04 **(+3.21)**	97.23 **(+9.00)**	94.48 **(+7.23)**
	ESENet + IFTS	1.1M	*95.83* **(+2.26)**	*97.90* **(+8.09)**	*95.09* **(+5.76)**
	ResNet18 + IFTS	22M	**96.86** **(+3.11)**	**98.28** **(+8.07)**	**95.46** **(+5.15)**

In this subsection, the numerical performance of each ATR model is investigated under different EOC setups, which reflect more practical circumstances than the SOC setup. The overall results for the EOC-1, EOC-2, and EOC-3 datasets are organized in Table 8.

4.4.1. Results Under EOC-1

The EOC-1 setup corresponds to the scenario of significant depression angle changes and comprises four-class training and test data obtained from depression angles of $17°$ and $30°$, respectively (Table 3). Based on the first column of Table 8, which shows the results for EOC-1, it can be observed that the overall ATR performance deteriorates compared with the SOC setup, despite the decrease in the number of classes to be categorized (i.e., from 10 to 4 classes). This indicates the difficulty in capturing an appropriate high-level representation under significant depression angle differences between the training and test SAR data. Fortunately, it is remarkable that applying the proposed IFTS framework can increase ATR accuracies with larger margins compared with SOC setup across all counterpart backbones: improvements by 3.21% for AConvNet, 2.26% for ESENet, and 3.11% for ResNet18. Detailed EOC-1 results of the ResNet18 backbone model combined with the IFTS are presented in Table 9 in a confusion matrix format.

Table 9. ATR Results of ResNet18+IFTS in Confusion Matrix Form, under EOC-1 Setup

Class	**BRDM-2**	**ZSU-234**	**2S1**	**T-72**	**Accuracy** [%]
BRDM-2	284	1	2	0	98.95
ZSU-234	0	282	2	4	98.26
2S1	7	0	277	4	96.18
T-72	6	7	3	272	94.44
Total					**96.87**

4.4.2. Results Under EOC-2

EOC-2 is a four-class classification setup for evaluating ATR performance in a scenario where the target configuration varies between training and testing (Table 4). The results under the EOC-2 setup (the second column of Table 8) reveal that the conventional models with target-only encoding yield further degraded accuracies when the target configurations are varied, compared with the EOC-1 case where the depression angles are varied. Nevertheless, when the proposed IFTS framework is applied, all networks indicate huge performance improvements (i.e., improvements by 9.00% for AConvNet, 8.09% for ESENet, and 8.07% for ResNet18), resulting in even higher accuracies compared to those under EOC-1. Therefore, we can infer that the shadow region can provide further complementary indicators in the scenario of configuration variant compared with the other setups. Table 10 shows the confusion matrix of the results of the ResNet18 backbone model with IFTS under the EOC-2 setup.

Table 10. ATR Results of ResNet18+IFTS in Confusion Matrix Form, under EOC-2 Setup

Class	**Serial No.**	**BRDM-2**	**BTR-70**	**BMP-2**	**T-72**	**Accuracy** [%]
T-72	S7	0	0	10	409	97.61
	A32	1	1	2	568	99.30
	A62	0	2	5	566	98.78
	A63	6	0	10	557	97.21
	A64	0	4	7	562	98.08
Total						**98.23**

4.4.3. Results Under EOC-3

Table 11. ATR Results of ResNet18+IFTS in Confusion Matrix Form, under EOC-3 Setup

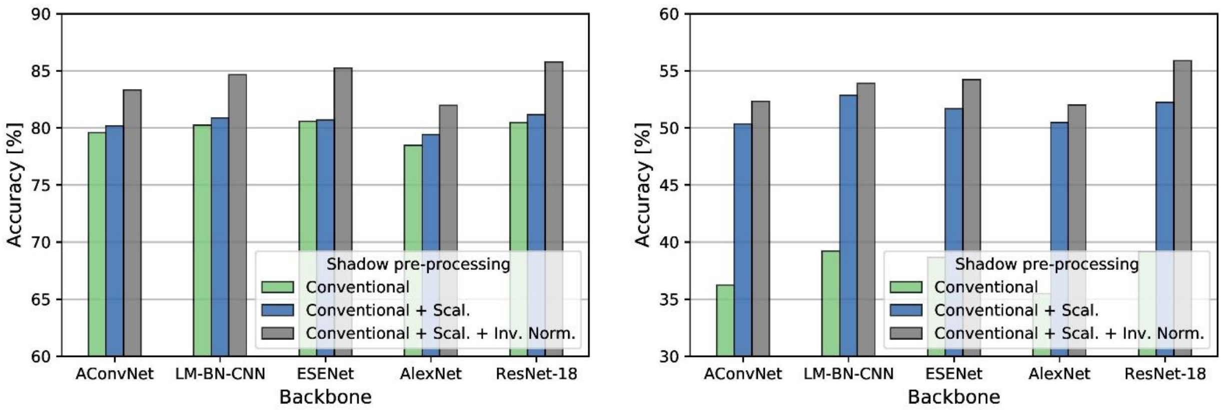
Class	**Serial No.**	**BRDM-2**	**BTR-70**	**BMP-2**	**T-72**	**Accuracy** [%]
BMP-2	9566	10	0	418	0	97.66
	C21	7	1	420	1	97.90
T-72	812	2	11	17	396	92.96
	A04	0	3	29	541	94.42
	A05	0	0	20	553	96.51
	A07	8	3	27	535	93.37
	A10	0	3	18	546	96.30
Total						**95.52**

This setup reflects the scenario where the target versions are varied, and it comprises four-class training SAR data and two-class test data with versions different from those in training (Table 5). The results of the EOC-3 setup (the third column of Table 8) show that the models based on IFTS encoding outperform those based on conventional target-only encoding by a large margin (i.e., improvements by 7.23% for AConvNet, 5.76% for ESENet, and 5.15% for ResNet18), demonstrating the effectiveness of the proposed IFTS framework even in the scenario involving different target versions between training and testing. The confusion matrix for the ResNet18+IFTS model is presented in Table 11.

4.4.4. Overall Discussion of EOC Results

Overall, the ATR models in the EOC setups highly benefit from the proposed IFTS framework regardless of the backbone topology; they exhibit even greater accuracy improvements compared with the SOC setup. Considering the aforementioned concern that the ATR performance can rather be degraded when the shadow information is incautiously incorporated without compensation for its unique domain characteristics (as discussed in Section 2.1), these results not only support our motivation to enable the cooperative implementation of shadow-centric processing based on a parallelized mechanism, but also validate the practicality of the proposed IFTS framework.

4.5. Ablation Study: Effectiveness of Shadow-Centric Preprocessing



(a) **(b)**

Fig. 10. SAR-ATR accuracy of various backbone models trained only on shadow modalities with different preprocessing. (a) Results under SOC setup. (b) Results under EOC-1 setup.

To systematically explore the effects of the proposed shadow-centric processing techniques (i.e., image rescaling for shadow and inverse normalization), we herein report the performance of several ATR models trained with only shadow information. We compared three different combinations of shadow preprocessing pipelines under the SOC and EOC-1 setups while fixing the other conditions. The results for each setup are illustrated in Fig. 10.

As presented in Fig. 10(a), which corresponds to the results of the SOC, adding a shadow-oriented rescaling algorithm increases the ATR accuracy across all backbone models (a 0.59% increase on average) compared with conventional processing, and further improves the performance when combined with the inverse normalization technique (an additional increase of 3.74% on average). In addition, it can be observed that appropriate preprocessing is much more influential than network architectures in managing shadow regions. The relative effectiveness of the shadow-oriented rescaling compared with inverse normalization is due to the marginal difference in the depression angle between training and test under the ideal SOC condition. As shown by the results under significant depression angle differences [Fig. 10(b)], it is clear that the shadow-oriented rescaling technique is extremely beneficial; it affords a significant improvement of 13.77% on average, compared with the additional increase of 2.14% when inverse normalization is added. In general, the results above explicitly demonstrate the validity of the proposed preprocessing techniques for shadows, particularly under practical sensing scenarios.

4.6. Ablation Study: Effectiveness of Fusion Strategies

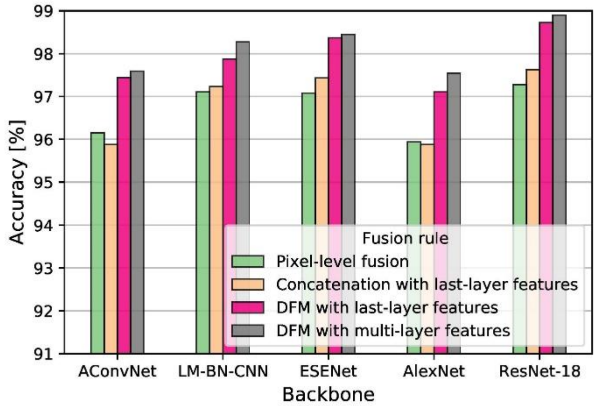


Fig. 11. SAR-ATR accuracy of various backbone models based on different fusion rules, under SOC setup.

To validate the effect of the proposed fusion strategies (i.e., adaptive fusion using multistage DFMs) for disparate target and shadow information, we evaluate the ATR performance of the SOC dataset under several plausible fusion rules: 1) pixel-level fusion of the preprocessed target and shadow images, 2) concatenation-based fusion of last-layer features, 3) DFM-based fusion of last-layer features, and 4) DFM-based fusion of multistage features.

As shown in Fig. 11, the DFM-based fusion rules outperform the pixel-level and concatenation-based fusion rules significantly, which indicates the importance of adaptive integration in leveraging shadows for SAR-ATR. In addition, our multilayer fusion strategy can further strengthen the representation capability of each network, leading to a 0.25% increment in ATR accuracy on average in comparison with last-layer fusion.

5. Conclusion

Shadow reflection in SAR imagery includes backscattered components of an object's configuration, as with direct target reflections. However, its unique domain properties, which are distinct from the target region, result in substantial incompatibility with the target, rendering shadows inoperative in current SAR-ATR tasks. To induce a network such that the benefit from the joint utilization of the target and shadow can be reaped effectively, we proposed an IFTS framework comprising novel solutions in three aspects. First, we designed a new series of preprocessing techniques specifically customized for the shadow region to compensate for its contradictory nature, as compared with the target. Second, we presented a parallelized SAR encoding pipeline such that independent processing for the target and shadow can be guaranteed structurally, thereby resulting in a representation pair oriented toward each modality. Third, we proposed a multistage fusion strategy based on DFMs, which enabled an adaptive fusion of the target and shadow while accounting for their relative significance. Based on extensive experiments using a public SAR benchmark dataset, we observed that our IFTS successfully enabled a network to improve its understanding of the shadow region, thereby achieving state-of-the-art performances under ideal SOC as well as practical EOC setups.

References

[1] M. Soumekh, Synthetic aperture radar signal processing, Wiley, New York, 1999.

[2] M. Kirscht, J. Mietzner, B. Bickert, A. Dallinger, J. Hippler, J. Meyer-Hilberg, R. Zahn, and J. Boukamp, "An airborne radar sensor for maritime and ground surveillance and reconnaissance—Algorithmic issues and exemplary results," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 9, no. 3, pp. 971–979, Mar. 2016.

[3] S. Chen, H. Wang, F. Xu, and Y. Jin, "Target classification using the deep convolutional networks for SAR images," IEEE Trans. Geosci. Remote Sens., vol. 54, no. 8, pp. 4806–4817, Aug. 2016.

[4] C. Wang, M. Zhang, Z.-W. Xu, C. Chen, and D.-S. Sheng, "Effects of anisotropic ionospheric irregularities on space-borne SAR imaging," IEEE Trans. Antennas Propag., vol. 62, no. 9, pp. 4664–4673, Sep. 2014.

[5] A. Fiche, S. Angelliaume, L. Rosenberg, and A. Khenchaf, “Analysis of X-band SAR sea-clutter distributions at different grazing angles,” IEEE Trans. Geosci. Remote Sens., vol. 53, no. 8, pp. 4650–4660, Aug. 2015.
[6] B. Zhao, L. Huang, J. Li, and P. Zhang, “Target reconstruction from deceptively jammed single-channel SAR,” IEEE Trans. Geosci. Remote Sens., vol. 56, no. 1, pp. 152–167, Jan. 2018.
[7] R. C. P. Marques, F. N. Medeiros, and J. Santos Nobre, “SAR image segmentation based on level set approach and G0 a model,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 10, pp. 2046–2057, Oct. 2012.
[8] V. S. Frost, J. A. Stiles, K. S. Shanmugan, and J. C. Holtzman, “A model for radar images and its application to adaptive digital filtering of multiplicative noise,” IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-4, no. 2, pp. 157–166, Mar. 1982.
[9] D. E. Dudgeon and R. T. Lacoss, “An overview of automatic target recognition,” Lincoln Lab. J., vol. 6, no. 1, 1993.
[10] L. M. Novak, G. J. Owirka, W. S. Brower, and A. L. Weaver, “The automatic target-recognition system in SAIP,” Lincoln Lab. J., vol. 10, no. 2, 1997.
[11] Q. Zhao and J. Principe, “Support vector machines for SAR automatic target recognition,” IEEE Trans. Aerosp. Electron. Syst., vol. 37, no. 2, pp. 643–654, Apr. 2001.
[12] G. Jones and B. Bhanu, “Recognition of articulated and occluded objects,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 21, no. 7, pp. 603–613, Jul. 1999.
[13] Z. Jianxiong, S. Zhiguang, C. Xiao, and F. Qiang, “Automatic target recognition of SAR images based on global scattering center model,” IEEE Trans. Geosci. Remote Sens., vol. 49, no. 10, pp. 3713– 3729, Oct. 2011.
[14] C. Enderli, L. Savy, and P. Refregier, “Application of the deflection criterion to classification of radar SAR images,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 9, pp. 1668–1672, Sep. 2007.
[15] U. Srinivas, V. Monga, and R. G. Raj, “SAR automatic target recognition using discriminative graphical models,” IEEE Trans. Aerosp. Electron. Syst., vol. 50, no. 1, pp. 591–606, Jan. 2014.
[16] J.-I. Park and K.-T. Kim, “Modified polar mapping classifier for SAR automatic target recognition,” IEEE Trans. Aerosp. Electron. Syst., vol. 50, no. 2, pp. 1092–1107, Apr. 2014.
[17] B. Ding, G. Wen, C. Ma, and X. Yang, “An efficient and robust framework for SAR target recognition by hierarchically fusing global and local features,” IEEE Trans. Image Process., vol. 27, no. 12, pp. 5983–5995, Dec. 2018.
[18] G. Dong, N. Wang, and G. Kuang, “Sparse representation of monogenic signal: With application to target recognition in SAR images,” IEEE Signal Process. Lett., vol. 21, no. 8, pp. 952–956, Aug. 2014.
[19] O. Kechagias-Stamatis and N. Aouf, “Fusing deep learning and sparse coding for SAR ATR,” IEEE Trans. Aerosp. Electron. Syst., vol. 55, no. 2, pp. 785–797, Apr. 2019.
[20] P. Gandhi and S. Kassam, “Analysis of CFAR processors in nonhomogeneous background,” IEEE Trans. Aerosp. Electron. Syst., vol. 24, no. 4, pp. 427–445, Jul. 1988.
[21] Y. Chen, E. Blasch, H. Chen, T. Qian, and G. Chen, “Experimental feature-based SAR ATR performance evaluation under different operational conditions,” in Proc. SPIE Signal Process., Sens. Fusion, and Target Recognit. XVII, vol. 6968, Orlando, Florida, USA, Apr. 2008, p. 69680F.
[22] A. Jain and D. Zongker, “Feature selection: evaluation, application, and small sample performance,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 2, pp. 153–158, Feb. 1997.
[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in Proc. Adv. Neural Inf. Process. Syst. (NIPS), vol. 25, Lake Tahoe, Nevada, USA, Dec. 2012, pp. 1097–1105.
[24] M. Tan, R. Pang, and Q. V. Le, “EfficientDet: Scalable and efficient object detection,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Seattle, WA, USA, Jun. 2020, pp. 10 778–10 787.
[25] H. Purwins, B. Li, T. Virtanen, J. Schl ¨ uter, S.-Y. Chang, and T. Sainath, “Deep learning for audio signal

processing," IEEE J. Sel. Topics Signal Process., vol. 13, no. 2, pp. 206–219, May 2019.
[26] B. Yang, R. Guo, M. Liang, S. Casas, and R. Urtasun, "RadarNet: Exploiting radar for robust perception of dynamic objects," in Proc. Eur. Conf. Comput. Vis. (ECCV), Glasgow, UK, Aug. 2020, pp. 496–512.
[27] S. A. Wagner, "SAR ATR by a combination of convolutional neural network and support vector machines," IEEE Trans. Aerosp. Electron. Syst., vol. 52, no. 6, pp. 2861–2872, Dec. 2016.
[28] J. Ding, B. Chen, H. Liu, and M. Huang, "Convolutional neural network with data augmentation for SAR target recognition," IEEE Geosci. Remote Sens. Lett., vol. 13, no. 3, pp. 364–368, Mar. 2016.
[29] Z. Cui, M. Zhang, Z. Cao, and C. Cao, "Image data augmentation for SAR sensor via generative adversarial nets," IEEE Access, vol. 7, pp. 42 255–42 268, Mar. 2019.
[30] C. Zheng, X. Jiang, and X. Liu, "Semi-supervised SAR ATR via multi-discriminator generative adversarial network," IEEE Sensors J., vol. 19, no. 17, pp. 7525–7533, Sep. 2019.
[31] K. Wang, G. Zhang, and H. Leung, "SAR target recognition based on cross-domain and cross-task transfer learning," IEEE Access, vol. 7, pp. 153 391–153 399, Oct. 2019.
[32] Z. Huang, Z. Pan, and B. Lei, "What, where, and how to transfer in SAR target recognition based on deep CNNs," IEEE Trans. Geosci. Remote Sens., vol. 58, no. 4, pp. 2324–2336, Apr. 2020.
[33] F. Ye, W. Luo, M. Dong, H. He, and W. Min, "SAR image retrieval based on unsupervised domain adaptation and clustering," IEEE Geosci. Remote Sens. Lett., vol. 16, no. 9, pp. 1482–1486, Sep. 2019.
[34] J. Pei, Y. Huang, Z. Sun, Y. Zhang, J. Yang, and T.-S. Yeo, "Multiview synthetic aperture radar automatic target recognition optimization: Modeling and implementation," IEEE Trans. Geosci. Remote Sens., vol. 56, no. 11, pp. 6425–6439, Nov. 2018.
[35] X. Bai, R. Xue, L. Wang, and F. Zhou, "Sequence SAR image classification based on bidirectional convolution-recurrent network," IEEE Trans. Geosci. Remote Sens., vol. 57, no. 11, pp. 9223–9235, Nov. 2019.
[36] R. Xue, X. Bai, and F. Zhou, "Spatial–temporal ensemble convolution for sequence SAR target classification," IEEE Trans. Geosci. Remote Sens., vol. 59, no. 2, pp. 1250–1262, Feb. 2021.
[37] Z. Zhang, H. Wang, F. Xu, and Y.-Q. Jin, "Complex-valued convolutional neural network and its application in polarimetric SAR image classification," IEEE Trans. Geosci. Remote Sens., vol. 55, no. 12, pp. 7177–7188, Dec. 2017.
[38] F. Zhou, L.Wang, X. Bai, and Y. Hui, "SAR ATR of ground vehicles based on LM-BN-CNN," IEEE Trans. Geosci. Remote Sens., vol. 56, no. 12, pp. 7282–7293, Dec. 2018.
[39] L. Wang, X. Bai, and F. Zhou, "SAR ATR of ground vehicles based on ESENet," Remote Sens., vol. 11, no. 11, pp. 1–16, Jun. 2019.
[40] A. Filippidis, L. Jain, and N. Martin, "Fusion of intelligent agents for the detection of aircraft in SAR images," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 4, pp. 378–384, Apr. 2000.
[41] R. Schumacher and J. Schiller, "Non-cooperative target identification of battlefield targets - classification results based on SAR images," in Proc. IEEE Int. Radar Conf., Arlington, VA, USA, May 2005, pp. 167–172.
[42] S.-H. Kim, W.-J. Song, and S.-H. Kim, "Robust ground target detection by SAR and IR sensor fusion using Adaboost-based feature selection," Sens., vol. 16, no. 7, pp. 1–27, Jul. 2016.
[43] A. W. Doerry, "Comments on airborne ISR radar utilization," in Proc. SPIE Radar Sens. Technol. XX, vol. 9829, Baltimore, Maryland, USA, May 2016, pp. 504–515.
[44] A. M. Raynal, D. L. Bickel, and A. W. Doerry, "Stationary and moving target shadow characteristics in synthetic aperture radar," in Proc. SPIE Radar Sens. Technol. XVIII, vol. 9077, Baltimore, Maryland, USA, May 2014, pp. 413–427.
[45] S. Papson and R. M. Narayanan, "Classification via the shadow region in SAR imagery," IEEE Trans. Aerosp. Electron. Syst., vol. 48, no. 2, pp. 969–980, Apr. 2012.
[46] M. Jahangir, D. Blacknell, C. P. Moate, and R. D. Hill, "Extracting information from shadows in SAR imagery," in Proc. Int. Conf. on Mach. Vis. (IMCV), Isalambad, Pakistan, Dec. 2007, pp. 107–112.

[47] K. Tang, X. Sun, H. Sun, and H. Wang, “A geometrical-based simulator for target recognition in high-resolution SAR images,” IEEE Geosci. Remote Sens. Lett., vol. 9, no. 5, pp. 958–962, Sep. 2012.

[48] P. Lombardo, M. Sciotti, and L. Kaplan, “SAR prescreening using both target and shadow information,” in Proc. IEEE Radar Conf. (RadarConf), Atlanta, GA, USA, May 2001, pp. 147–152.

[49] J. Cui, J. Gudnason, and M. Brookes, “Radar shadow and super-resolution features for automatic recognition of MSTAR targets,” in Proc. IEEE Int. Radar Conf., Arlington, VA, USA, May 2005, pp. 534–539.

[50] F. Gao, J. You, J. Wang, J. Sun, E. Yang, and H. Zhou, “A novel target detection method for SAR images based on shadow proposal and saliency analysis,” Neurocomputing, vol. 267, pp. 220–231, Dec. 2017.

[51] C. Belloni, A. Balleri, N. Aouf, J.-M. Le Caillec, and T. Merlet, “Explainability of deep SAR ATR through feature analysis,” IEEE Trans. Aerosp. Electron. Syst., vol. 57, no. 1, pp. 659–673, Feb. 2021.

[52] E. R. Keydel, S. W. Lee, and J. T. Moore, “MSTAR extended operating conditions: A tutorial,” in Proc. SPIE Algo. Synthetic Aperture Radar Imagery III, vol. 2757, Orlando, FL, USA, Jun. 1996, pp. 228–242.

[53] C. Feichtenhofer, H. Fan, J. Malik, and K. He, “SlowFast networks for video recognition,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Seoul, Korea, Nov. 2019, pp. 6201–6210.

[54] Y.-G. Jiang, Z. Wu, J. Wang, X. Xue, and S.-F. Chang, “Exploiting feature and class relationships in video categorization with regularized deep neural networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 2, pp. 352–364, Feb. 2018.

[55] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, “Multimodal deep learning,” in Proc. Int. Conf. Mach. Learn. (ICML), Bellevue, Washington, USA, Jul. 2011, pp. 689–696.

[56] R. Meth, “Target/shadow segmentation and aspect estimation in synthetic aperture radar imagery,” in Proc. SPIE Algo. Synthetic Aperture Radar Imagery V, vol. 3370, Orlando, FL, USA, Sep. 1998, pp. 188–196.

[57] M. Chang and X. You, “Target recognition in SAR images based on information-decoupled representation,” Remote Sens., vol. 10, no. 1, pp. 1–19, Jan. 2018.

[58] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, Cambridge, 2016.

[59] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in Proc. Int. Conf. Learn. Representations (ICLR), San Diego, CA, USA, May 2015.

[60] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[61] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in Proc. Int. Conf. Learn. Representations (ICLR), San Diego, CA, USA, May 2015.

[62] H. Zhu, T. Wong, N. Lin, W. Wang, and S. Thedoridis, “Synthetic aperture radar target classification based on 3-D convolutional neural network,” in Proc. IEEE Int. Conf. on Signal and Image Process. (ICSIP), Nanjing, China, Oct. 2020, pp. 440–445.

[63] D. Malmgren-Hansen and M. Nobel-Jogensen, “Convolutional neural networks for SAR image segmentation,” in 2015 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Abu Dhabi, UAE, Dec. 2015, pp. 231–236.

[64] M. Heiligers and A. Huizing, “On the importance of visual explanation and segmentation for SAR ATR using deep learning,” in Proc. IEEE Radar Conf. (RadarConf), Oklahoma City, OK, USA, Apr. 2018, pp. 0394–0399.

[65] M. Tan and Q. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in Proc. Int. Conf. Mach. Learn. (ICML), vol. 97, Long Beach, California, USA, Jun. 2019, pp. 6105–6114.