Deep Learning for Joint Image Reconstruction and Segmentation for SAR

Samia Kazemi
Department of Electrical, Computer and Systems Engineering,
Rensselaer Polytechnic Institute
110 8th Street, Troy, NY 12180 USA
Email: kazems@rpi.edu

Birsen Yazici
Department of Electrical, Computer and Systems Engineering
Rensselaer Polytechnic Institute
110 8th Street, Troy, NY 12180 USA
Email: yazici@ecse.rpi.edu

***Abstract*—We present an approach for joint image reconstruction and foreground-background separation for synthetic aperture radar (SAR) using deep learning (DL). Network structure of the deep model is derived by unwrapping the stages of an iterative algorithm that solves an underlying optimization problem. This leads to physical model based deep network with learned network parameters having meaningful interpretation. Combined image reconstruction and segmentation approach allows joint optimization of both tasks that enhances performance and prevent inadvertent loss of useful information. Numerical results are included to show feasibility of the proposed approach.**

I. Introduction

In recent year, there have been significant interest in applying DL based approach to a wide variety of SAR related problems including image reconstruction [1], [2], transmission signal recovery for passive SAR [3], automatic target recognition (ATR) [4]–[11], direct ATR from SAR received signal without intermediate imaging step [12], [13] etc.

Both image reconstruction and segmentation are important tasks in remote sensing, and image and video signal processing. They are often used as preliminary steps for subsequent processing tasks such as ATR, navigation etc. Segmentation refers to the process of isolating different parts of an image based on their content. In this paper, we consider segmentation as a foreground-background separation problem, where the foreground is composed of elements from a specific set of objects of interest while the background is refers to the remaining parts of the scene. Accurate foreground-background separation is highly desirable for ATR since locating and identifying various objects from a separated foreground is relatively easier than classification from the complete scene with complex background.

Joint reconstruction and segmentation instead of segmenting from a reconstructed image can be beneficial in terms of both reconstruction and segmentation quality. For the two step process, quality of the reconstructed image affects the performance of the segmentation algorithm. This is because the isolated reconstruction algorithm does not take subsequent processing steps into account and finer details that are important for the segmentation task may be lost while enhancing overall reconstruction quality.

Various joint reconstruction and segmentation methods have been previously studied in the literature including [14]–[16]. There are some recent works on DL based joint image reconstruction and segmentation problem in medical imaging field including [17], [18]. Both of these approaches are based on convolutional neural network (CNN). In [17], a joint multi-stage CNN model is proposed for parallel reconstruction and segmentation of 7T magnetic resonance (MR) images from 3T MR images in order to provide higher resolution and contrast quality. Reconstruction and segmentation tasks require separate cascaded CNN networks with mutual feedbacks during forward propagation step which allows for joint optimization. In [18], a joint reconstruction and segmentation network for magnetic resonance imaging (MRI) for compressed sensing (CS) measurement condition is presented. Proposed deep network, referred to as SegNetMRI, is composed of repetitive CNN based encoder-decoder pairs where the cascaded networks for reconstruction and segmentation share the same encoders. Outputs of the multiple decoders for the segmentation network are then combined to give the final output.

In this paper, we propose an approach that combines model based image reconstruction (MBIR) with the learning aspect of deep networks for simultaneous image reconstruction and segmentation. The network structure is derived based on an iteration algorithm for the optimization problem for this joint task. The resulting deep network takes a repetitive structure with various stages sharing common set of parameters. Such network architectures are commonly referred to as recurrent neural network (RNN), and the parameter sharing aspect makes them particularly suitable for SAR related task due to the scarcity in labeled dataset for training. Deep model based approach have the potential to utilize highly non-linear mapping capability to generate a better iteration sequence, optimized for a specific dataset, to improve performance while reducing computational cost by solving the optimization problem within a fixed number steps.

The rest of the paper is organized as follows: In Section II, we present the forward model and establish different notations. In Section III, we discuss the theory for the joint optimization problem and a corresponding iterative algorithm.

978-1-7281-6813-5/20/$31.00 ©2020 IEEE

In Section IV, we present the deep network. In Section VI, we present numerical results to verify feasibility of the model. Finally, Section VII concludes the paper.

II. Problem statement

Let $x \in \mathbb{R}$, $\boldsymbol{x} = [x_1, x_2] \in \mathbb{R}^2$ and $\mathbf{x} = [\boldsymbol{x}, \psi(\boldsymbol{x})] \in \mathbb{R}^3$ denotes the one, two and three dimensional co-ordinate variables, respectively. $\psi(\boldsymbol{x})$ denotes the surface topography at coordinate $\boldsymbol{x}$. Let $s \in [s_1, s_2]$ and $\omega \in [\omega_1, \omega_2]$ denote the slow-time points and fast-time frequencies, respectively. The SAR is mounted on a moving platform that travels along a known trajectory. We assume that start-stop approximation is applicable, i.e. the SAR location remains effectively unchanged while the pulse is being transmitted and reflected back due to relatively small platform velocity compared to the speed of light in free space c. Location of the SAR transmitter and receiver are denoted by $\boldsymbol{\gamma}_T(s) : \mathbb{R} \to \mathbb{R}^3$ and $\boldsymbol{\gamma}_R(s) : \mathbb{R} \to \mathbb{R}^3$, respectively. Received signal for slow-time point s and fast-time frequency ω, assuming single scattering model, can be expressed as [19]

$$d(s,\omega) = \int A(s,\omega,\mathbf{x}) e^{-\frac{\mathrm{i}\omega}{c} R(s,\mathbf{x})} \rho(\mathbf{x}) d\mathbf{x} + n(s,w), \quad (1)$$

$$= \mathcal{F}[\rho](s,\omega) + n(s,w). \quad (2)$$

$\rho(\mathbf{x}) : \mathbb{R}^3 \to \mathbb{C}$ denotes the reflectivity of the surface at co-ordinate $\mathbf{x}$. $\mathcal{F}[.](s,\omega)$ is the forward mapping operator. $R(s,x) = |\boldsymbol{\gamma}_T(s) - \mathbf{x}| + |\mathbf{x} - \boldsymbol{\gamma}_R(s)|$ is the total distance travelled by the transmitted signal and $n(s,w)$ is the additive Gaussian noise. Suppose a finite dimensional representation of the forward mapping operator $\mathcal{F}$ is denoted by $\mathbf{F} \in \mathbb{C}^{M \times N}$. Corresponding discrete forward model can be expressed as:

$$\mathbf{d} = \mathbf{F}\boldsymbol{\rho} + \mathbf{n}. \quad (3)$$

Here, $\mathbf{d} \in \mathbb{C}^M$ and $\boldsymbol{\rho} \in \mathbb{C}^N$ are the discrete data and reflectivity vectors, respectively. $\mathbf{n} \in \mathbb{C}^N$ is a complex Gaussian vector distributed as $\mathcal{N}(\mathbf{0}, \sigma^2(1+1i)\mathbf{I})$. Suppose the foreground and background parts of the image $\boldsymbol{\rho}$ are denoted by $\boldsymbol{\rho}_f$ and $\boldsymbol{\rho}_b$, respectively. Our goal is to jointly reconstruct $\boldsymbol{\rho}_f$ and $\boldsymbol{\rho}_b$ given the received signal $\mathbf{d}$.

III. Joint Image Reconstruction and Segmentation

We perform joint reconstruction of the image segments from the sensing data by estimating the complete image $\boldsymbol{\rho}$ and a mapping vector $\mathbf{w}$ of length N with its n^{th} element equal to 1 if the corresponding component of $\boldsymbol{\rho}$ belong to the foreground, else it is 0. The two segments, $\boldsymbol{\rho}_f$ and $\boldsymbol{\rho}_b$, can be calculated from these estimated quantities by $\boldsymbol{\rho}_f = diag(\mathbf{w})\boldsymbol{\rho}$ and $\boldsymbol{\rho}_b = (\mathbf{I} - diag(\mathbf{w}))\boldsymbol{\rho}$, where $diag(\mathbf{w})$ refers to a diagonal matrix with $\mathbf{w}$ along its main diagonal.

An optimization problem for this joint task can be derived from the formulation of the Maximum Apriori Probability (MAP) estimation of $\boldsymbol{\rho}$ and $\mathbf{w}$. Suppose $\boldsymbol{\rho}^*$ and $\mathbf{w}^*$ denote the MAP estimations of unknown signals $\boldsymbol{\rho}$ and $\mathbf{w}$, respectively. Then

$$\boldsymbol{\rho}^*, \mathbf{w}^* = \underset{\boldsymbol{\rho},\mathbf{w}}{\operatorname{argmax}} \log Pr(\boldsymbol{\rho}, \mathbf{w}|\mathbf{d}), \quad (4)$$

$$= \underset{\boldsymbol{\rho},\mathbf{w}}{\operatorname{argmax}} \log Pr(\mathbf{d}|\boldsymbol{\rho}, \mathbf{w}) + \log Pr(\boldsymbol{\rho}|\mathbf{w}) + \log Pr(\mathbf{w}), \quad (5)$$

$$= \underset{\boldsymbol{\rho},\mathbf{w}}{\operatorname{argmax}} -\|\mathbf{d} - \mathbf{F}\boldsymbol{\rho}\|^2 + \log Pr(\boldsymbol{\rho}|\mathbf{w}) + \log Pr(\mathbf{w}). \quad (6)$$

Data fitting term $\|\mathbf{d} - \mathbf{F}\boldsymbol{\rho}\|^2$ measures the consistency of the reconstructed image with the measured data and is a convex function of $\boldsymbol{\rho}$. Let $\Phi(\boldsymbol{\rho}, \mathbf{w}) := \log Pr(\boldsymbol{\rho}|\mathbf{w}) + \log Pr(\mathbf{w})$, and it imposes constraints on the unknown quantities $\boldsymbol{\rho}$ and $\mathbf{w}$. In case the joint prior distribution $Pr(\boldsymbol{\rho}, \mathbf{w})$ is known in advance, we can get an explicit expression of $\phi(.)$ as a function of the unknowns. For example, the separated foreground part is likely to be sparse with a higher expected value compared to the background part, while the background may or may not be sparse.

Similar to the formulation in [20], we assume generalized Gaussian distribution for the image $\boldsymbol{\rho}$ given $\mathbf{w}$. Let $\mathcal{N}_f$ and $\mathcal{N}_b$ denote the sets of indices of $\boldsymbol{\rho}$ that belong to the foreground and the background parts, respectively. Let the shape parameters for the generalized Gaussian distribution for the foreground and background segments are ζ_f and ζ_b, respectively. The scale parameters for these two parts are denoted by γ_f and γ_b, respectively. The probability distribution of $\boldsymbol{\rho}$ given $\mathbf{w}$ is as follows:

$$Pr(\boldsymbol{\rho}|\mathbf{w}) = \frac{1}{\left(2\gamma_f^{\frac{1}{\zeta_f}}\Gamma(1+\frac{1}{\zeta_f})\right)^{|\mathcal{N}_f|}} \exp\left(-\frac{\sum_{i\in\mathcal{N}_f}|\boldsymbol{\rho}[i]|^{\zeta_f}}{\gamma_f}\right) \times \frac{1}{\left(2\gamma_b^{\frac{1}{\zeta_b}}\Gamma(1+\frac{1}{\zeta_b})\right)^{|\mathcal{N}_b|}} \exp\left(-\frac{\sum_{j\in\mathcal{N}_b}|\boldsymbol{\rho}[j]|^{\zeta_b}}{\gamma_b}\right). \quad (7)$$

Mapping vector $\mathbf{w}$ is assumed to have the Gibbs distribution, i.e.,

$$Pr(\mathbf{w}) = \frac{1}{K}\exp\left(\sum_{n=1}^{N}\sum_{n_1\in\epsilon(n)}\delta(\mathbf{w}[n] - \mathbf{w}[n_1])\right), \quad (8)$$

$\epsilon(n)$ denotes the set of indices that are in the neighbourhood of the n^{th} element, $\delta(.)$ denotes the Kronecker delta function and K is the normalizing constant.

By replacing the expressions for $Pr(\boldsymbol{\rho}|\mathbf{w})$ and $Pr(\mathbf{w})$ in (6), we get

$$\boldsymbol{\rho}^*, \mathbf{w}^* = \underset{\boldsymbol{\rho},\mathbf{w}}{\operatorname{argmin}} J(\boldsymbol{\rho}, \mathbf{w}) + i_C(\mathbf{w}), \quad (9)$$

where

$$J(\boldsymbol{\rho}, \mathbf{w}) = \|\mathbf{d} - \mathbf{F}\boldsymbol{\rho}\|^2 + (N - \mathbf{1}^T\mathbf{w})\log\left(2\gamma_b^{\frac{1}{\zeta_b}}\Gamma\left(1+\frac{1}{\zeta_b}\right)\right) + \mathbf{1}^T\mathbf{w}\log\left(2\gamma_f^{\frac{1}{\zeta_f}}\Gamma\left(1+\frac{1}{\zeta_f}\right)\right) + \frac{1}{\gamma_b}\|(\mathbf{I} - \mathbf{W})\boldsymbol{\rho}\|^{\zeta_b}$$

$$+\frac{1}{\gamma_f}\|\mathbf{W}\boldsymbol{\rho}\|^{\zeta_f}+\|\mathbf{A}\mathbf{w}\|^2. \quad (10)$$

$i_C(.)$ is an indicator function that takes a large value if its argument vector lies outside of set C, and otherwise it is 0. Here, C is a set of all N dimensional vectors with components equal to 0 or 1. $\mathbf{1}$ denotes an N dimensional vector of all 1's. $\mathbf{A}$ is a deterministic $N\times N$ matrix whose component at the i^{th} diagonal element $A_{i,i}$ equals to $|\epsilon(i)|$, and $A_{i,j}$ for all $j\in\epsilon(i)$ equals to -1. $\mathbf{W}$ is a diagonal matrix with $\mathbf{w}$ along its main diagonal.

The optimization problem in (9) can be solved using proximal gradient descent algorithm with alternating minimization steps for $\boldsymbol{\rho}$ and $\mathbf{w}$. At each iteration, there is a gradient descent step and an projection step for each unknown. Current estimations of the scene $\boldsymbol{\rho}$ and the mapping function $\mathbf{w}$ at the l^{th} iteration, $\boldsymbol{\rho}^{(l)}$ and $\mathbf{w}^{(l)}$, respectively, are updated by the following optimization problems:

$$\boldsymbol{\rho}^{(l+1)}=\underset{\boldsymbol{\rho}}{\operatorname{argmin}}\, J_1(\boldsymbol{\rho};\mathbf{w}^{(l)}), \quad (11)$$

$$J_1(\boldsymbol{\rho};\mathbf{w}^{(l)})=\|\mathbf{d}-\mathbf{F}\boldsymbol{\rho}\|^2+\frac{1}{\gamma_b}\|(\mathbf{I}-\mathbf{W}^{(l)})\boldsymbol{\rho}\|^{\zeta_b}+\frac{1}{\gamma_f}\|\mathbf{W}^{(l)}\boldsymbol{\rho}\|^{\zeta_f}, \quad (12)$$

and

$$\mathbf{w}^{(l+1)}=\underset{\mathbf{w}}{\operatorname{argmin}}\, J_2(\mathbf{w};\boldsymbol{\rho}^{(l+1)})+i_C(\mathbf{w}), \quad (13)$$

$$\begin{aligned}&J_2(\mathbf{w};\boldsymbol{\rho}^{(l+1)})\\&=\|\mathbf{A}\mathbf{w}\|^2+(N-\mathbf{1}^T\mathbf{w})\log\left(2\gamma_b^{\frac{1}{\zeta_b}}\Gamma\left(1+\frac{1}{\zeta_b}\right)\right)\\&+\mathbf{1}^T\mathbf{w}\log\left(2\gamma_f^{\frac{1}{\zeta_f}}\Gamma\left(1+\frac{1}{\zeta_f}\right)\right)+\frac{1}{\gamma_b}\|(\mathbf{I}-\mathbf{W})\boldsymbol{\rho}^{(l+1)}\|^{\zeta_b}\\&+\frac{1}{\gamma_f}\|\mathbf{W}\boldsymbol{\rho}^{(l+1)}\|^{\zeta_f}. \quad (14)\end{aligned}$$

We set ζ_f and ζ_b as tuning parameters whose values are adjusted based on the dataset. As an example case, we assume $\zeta_f=1$ and $\zeta_b=2$, and formulate an algorithm and corresponding deep network for this selection. This case is equivalent to imposing sparsity constraint on the foreground and normal prior distribution on the background.

Let $K_1(\boldsymbol{\rho};\mathbf{w}^{(l)}):=\|\mathbf{d}-\mathbf{F}\boldsymbol{\rho}\|^2+\frac{1}{\gamma_b}\|(\mathbf{I}-\mathbf{W}^{(l)})\boldsymbol{\rho}\|^2$. Value of reconstructed image can be updated at the l^{th} iteration step by performing a forward step along the steepest descent direction of K_1 and then applying a proximity operator on the updated image to apply the sparsity constraint imposed by the term $\frac{1}{\gamma_f}\|\mathbf{W}^{(l)}\boldsymbol{\rho}\|^1$. The update steps for $\boldsymbol{\rho}$ are stated below:

$$\boldsymbol{\rho}_1^{(l+1)}=\boldsymbol{\rho}^{(l)}-\lambda_\rho\nabla_{\boldsymbol{\rho}}K_1(\boldsymbol{\rho},\mathbf{w})|_{(\boldsymbol{\rho},\mathbf{w})=(\boldsymbol{\rho}^{(l)},\mathbf{w}^{(l)})}, \quad (15)$$

$$\boldsymbol{\rho}^{(l+1)}=\mathcal{P}_\tau(\boldsymbol{\rho}_1^{(l+1)}), \quad (16)$$

where λ_ρ is the learning rate and

$$\nabla_{\boldsymbol{\rho}}K_1(\boldsymbol{\rho},\mathbf{w})=\overline{\left(\frac{\partial K_1}{\partial\boldsymbol{\rho}}\right)} \quad (17)$$

$$=-\mathbf{F}^H(\mathbf{d}-\mathbf{F}\boldsymbol{\rho})+\frac{1}{\gamma_b}(\mathbf{I}-\mathbf{W})^2\boldsymbol{\rho}. \quad (18)$$

Proximity function $\mathcal{P}_\tau(.)$ applies componentwise on its argument vector and has the following expression:

$$\boldsymbol{\rho}^{(l+1)}[i]=\mathcal{P}_\tau(\boldsymbol{\rho}_1^{(l+1)}[i]), \quad (19)$$

$$=\begin{cases}\boldsymbol{\rho}_1^{(l+1)}[i], & |\mathbf{w}^{(l)}[i]\boldsymbol{\rho}_1^{(l+1)}[i]|\geq\tau,\mathbf{w}^{(l)}[i]=1\\0, & |\mathbf{w}^{(l)}[i]\boldsymbol{\rho}_1^{(l+1)}[i]|<\tau,\mathbf{w}^{(l)}[i]=1\\\boldsymbol{\rho}_1^{(l+1)}[i], & otherwise\end{cases}, \quad (20)$$

for all $i\in[1,N]$ and $\tau\in\mathbb{R}^+$.

Similarly, we define $K_2(\mathbf{w};\boldsymbol{\rho}^{(l+1)}):=\|\mathbf{A}\mathbf{w}\|^2+(N-\mathbf{1}^T\mathbf{w})\log\left(\gamma_b^{0.5}\sqrt{\pi}\right)+\mathbf{1}^T\mathbf{w}\log(2\gamma_f)+\frac{1}{\gamma_b}\|(\mathbf{I}-\mathbf{W})\boldsymbol{\rho}^{(l+1)}\|^2$. Gradient descent update of $\mathbf{w}$ is based on the derivative of this term and the constraints imposed by $i_C(\mathbf{w})$ and $\frac{1}{\gamma_f}\|\mathbf{W}\boldsymbol{\rho}^{(l+1)}\|^1$ are applied via proximity operators $\mathcal{Q}$ and $\mathcal{R}_\beta$, respectively. At the l^{th} stage, estimation of $\mathbf{w}$ is updated as follows:

$$\mathbf{w}_1^{(l+1)}=\mathbf{w}^{(l)}-\lambda_w\nabla_{\mathbf{w}}K_2(\boldsymbol{\rho},\mathbf{w})|_{(\boldsymbol{\rho},\mathbf{w})=(\boldsymbol{\rho}^{(l+1)},\mathbf{w}^{(l)})}, \quad (21)$$

$$\mathbf{w}_2^{(l+1)}=\mathcal{Q}(\mathbf{w}_1^{(l+1)}), \quad (22)$$

$$\mathbf{w}^{(l+1)}=\mathcal{R}_\beta(\mathbf{w}_2^{(l+1)}). \quad (23)$$

where λ_w is the learning rate and

$$\nabla_{\mathbf{w}}K_2(\boldsymbol{\rho},\mathbf{w})=2\mathbf{A}^T\mathbf{A}\mathbf{w}+\left(\log(2\gamma_f)-\frac{1}{2}\log(\gamma_b\pi)\right)\mathbf{1}+\frac{2}{\gamma_b}diag(\boldsymbol{\rho}\odot\overline{\boldsymbol{\rho}})(\mathbf{w}-\mathbf{1}). \quad (24)$$

Functions $\mathcal{Q}$ and $\mathcal{R}_\beta$ operate componentwise on the input vectors and are defined as follows:

$$\mathbf{w}_2^{(l+1)}[i]=\mathcal{Q}(\mathbf{w}_1^{(l+1)}[i]), \quad (25)$$

$$=\begin{cases}1, & \mathbf{w}_1^{(l+1)}[i]\geq 0.5\\0, & otherwise\end{cases}, \quad (26)$$

and

$$\mathbf{w}^{(l+1)}[i]=\mathcal{R}_\beta(\mathbf{w}_2^{(l+1)}[i]), \quad (27)$$

$$=\begin{cases}\mathbf{w}_2^{(l+1)}[i], & \mathbf{w}_2^{(l+1)}[i]\boldsymbol{\rho}^{(l+1)}[i]\geq\beta\\0, & otherwise\end{cases}. \quad (28)$$

IV. Deep Network

Similar to the deep networks presented in [1], [3], [12], L stages of the proximal gradient descent updates are unwrapped to form two L stage RNNs whose parameters are learned from a training set. At the l^{th} stage of the network, the weight matrix, bias vector and activation function for the RNN corresponding to $\boldsymbol{\rho}$ are $\mathbf{W}_\rho^l=f(\mathbf{w}^{(l)}):=\mathbf{I}-\lambda_\rho\mathbf{F}^H\mathbf{F}-\frac{\lambda_\rho}{\gamma_b}(\mathbf{I}-\mathbf{W}^{(l)})^2$, $\mathbf{b}_\rho=\lambda_\rho\mathbf{F}^H\mathbf{d}$ and $\mathcal{P}_\tau(.)$, respectively. Similarly, for the mapping vector $\mathbf{w}$, weight matrix, bias vector and activation functions at the l^{th} stage are $\mathbf{W}_w^l:=g(\boldsymbol{\rho}^{(l+1)})=\mathbf{I}-2\lambda_w\mathbf{A}^T\mathbf{A}-\frac{2\lambda_w}{\gamma_b}diag(\boldsymbol{\rho}^{(l+1)}\odot\overline{\boldsymbol{\rho}}^{(l+1)})$, $\mathbf{b}_w^l:=h(\boldsymbol{\rho}^{(l+1)})=\frac{2\lambda_w}{\gamma_b}diag(\boldsymbol{\rho}^{(l+1)}\odot\overline{\boldsymbol{\rho}}^{(l+1)})\mathbf{1}-\lambda_w\left(\log(2\gamma_f)-\frac{1}{2}\log(\gamma_b\pi)\right)\mathbf{1}$ and $\mathcal{Q}\circ\mathcal{R}_\beta(.)$, respectively. Network diagram for joint image reconstruction and segmentation is shown in Figure 1.

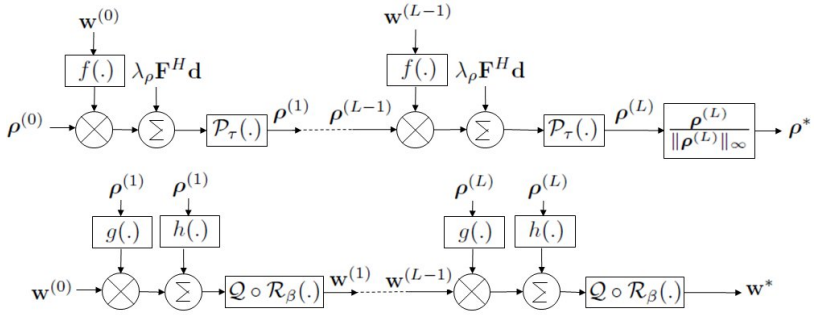

Fig. 1: Deep network for joint image reconstruction and foreground-background separation.

V. Training

Optimum values of the network parameters are learned by the training process from a set of examples. Suppose $\mathcal{T}$ denotes the training set and it has T samples. For a supervised training scenario, $\mathcal{T}$ comprises of a set of sensor data $\mathbf{d}_t$, for all t, collected by the SAR receiver from different scenes, and the corresponding true foreground and background image segments, $\boldsymbol{\rho}_{f,t}$ and $\boldsymbol{\rho}_{b,t}$, respectively. During training, a loss function $\mathcal{L}_{\boldsymbol{\Theta}}$ that quantifies the deviation of the estimated foreground segment $\boldsymbol{\rho}^*_{f,t} = \mathbf{w}^*_t \odot \boldsymbol{\rho}^*_t$ and the background segment $\boldsymbol{\rho}^*_{b,t} = (\mathbf{1} - \mathbf{w}^*_t) \odot \boldsymbol{\rho}^*_t$ from the ground truths, is minimized with respect to $\boldsymbol{\Theta}$. $\boldsymbol{\Theta}$ is the set containing all the network parameters to be learned and it includes λ_ρ, λ_w, γ_f, γ_b, τ and β. Additionally, from empirical evidence, learning process converges faster if $\mathbf{F}$ is included in $\boldsymbol{\Theta}$ with its known value used for initialization. Training loss function $\mathcal{L}_{\boldsymbol{\Theta}}$ is defined as follows:

$$\mathcal{L}_{\boldsymbol{\Theta}} = \frac{1}{T} \sum_{t=1}^{T} \|\boldsymbol{\rho}_{f,l} - \boldsymbol{\rho}^*_{f,l}(\boldsymbol{\Theta})\|^2 + \|\boldsymbol{\rho}_{b,l} - \boldsymbol{\rho}^*_{b,l}(\boldsymbol{\Theta})\|^2. \quad (29)$$

Learning process is based on stochastic gradient descent updates of the parameter set $\boldsymbol{\Theta}$, and at the k^{th} training its value is updated as

$$\boldsymbol{\Theta}^{(k+1)} = \boldsymbol{\Theta}^{(k)} - \lambda_t \nabla_{\boldsymbol{\Theta}} \mathcal{L}_{\boldsymbol{\Theta}}. \quad (30)$$

The RNN network requires initial values for the quantities to be estimated i.e. $\boldsymbol{\rho}^{(0)}$ and $\mathbf{w}^{(0)}$. We set $\boldsymbol{\rho}^{(0)}$ equal to the back-projected image $\mathbf{F}^H \mathbf{d}$, and $\mathbf{w}^{(0)}$ equal to all 1's.

VI. Numerical Results

To verify the feasibility of the proposed approach, we evaluate model performance on simulated SAR received signal. For simulation purpose, we have considered both ζ_f and ζ_b equal to 1. In the full version of the paper, we will include numerical results for more datasets, and comparative performances for different choices of ζ_f and ζ_b.

The sample scenes in the training and test set are simulated using MATLAB. Each scene is 32×32 pixels in dimension and contains one of the two kinds of objects, triangular and circular, located at a random position with different orientations. The background has non-zero reflectivity and the pixel values in the background varies from sample to sample.

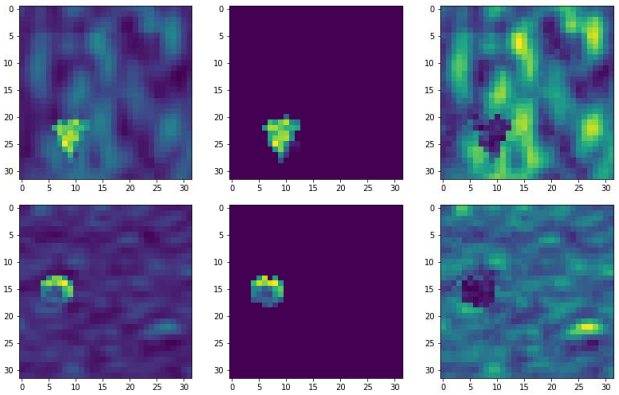

Fig. 2: Reconstruction and segmentation output for sample scenes with triangular and circular objects. On each row, the left-most images are the original scenes, the images on the center are the reconstructed foreground images and the images on the right-most column are the reconstructed backgrounds.

The SAR received signal is simulated assuming that the data collection geometry remains unchanged for different samples. We assume a mono-static SAR on a moving platform traveling in circular trajectory at a fixed height of 10km and radius 10km with the center of the scene as its origin. The sampled received signal has 28 slow-time and 28 fast-time points. Number of samples in the training and test set are 29500 and 500, respectively. Reconstructed foreground and background separated images for two samples in the test set are shown in Figure 2. For both types of object, reconstruction and segmentation performance appears to be accurate.

VII. Conclusion

We have proposed a framework for joint MBIR and segmentation approach for SAR received signal using DL. This approach allows for the incorporation of different prior models for the foreground and background parts of the image. Based on the prior models, derived network structures takes different forms. Preliminary numerical results on simulated SAR dataset shows promising reconstruction and segmentation performance.

VIII. Acknowledgements

This work was supported by the Air Force Office of Scientific Research (AFOSR) under the agreement FA9550-16-1-0234, Office of Naval Research (ONR) under the agreement N00014-18-1-2068 and by the National Science Foundation (NSF) under Grant No ECCS-1809234.

References

[1] B. Yonel, E. Mason, and B. Yazici, "Deep learning for passive synthetic aperture radar," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 90–103, 2017.

[2] Y. Chun and J. A. Fessler, "Deep bcd-net using identical encoding-decoding cnn structures for iterative image recovery," in *2018 IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*. IEEE, 2018, pp. 1–5.

[3] B. Yonel, E. Mason, and B. Yazici, "Deep learning for waveform estimation and imaging in passive radar," *IET Radar, Sonar & Navigation*, vol. 13, no. 6, pp. 915–926, 2019.

[4] Z. Sun, L. Xue, and Y. Xu, "Recognition of SAR target based on multilayer auto-encoder and SNN," *International Journal of Innovative Computing, Information and Control*, vol. 9, no. 11, pp. 4331–4341, 2013.

[5] D. A. E. Morgan, "Deep convolutional neural networks for ATR from SAR imagery," in *Algorithms for Synthetic Aperture Radar Imagery XXII*, vol. 9475, 2015, p. 94750F.

[6] H. Wang, S. Chen, F. Xu, and Y. Jin, "Application of deep-learning algorithms to MSTAR data," in *2015 IEEE International Geoscience and Remote Sensing Symposium*, 2015, pp. 3743–3745.

[7] S. Chen, H. Wang, F. Xu, and Y. Jin, "Target classification using the deep convolutional networks for SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4806–4817, 2016.

[8] J. L. Michael Wilmanski, Chris Kreucher, "Modern approaches in deep learning for SAR ATR," in *Algorithms for Synthetic Aperture Radar Imagery XXIII*, vol. 9843, 2016, p. 98430N.

[9] K. Du, Y. Deng, R. Wang, T. Zhao, and N. Li, "SAR ATR based on displacement-and rotation-insensitive cnn," *Remote Sensing Letters*, vol. 7, no. 9, pp. 895–904, 2016.

[10] Z. Huang, Z. Pan, and B. Lei, "Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data," *Remote Sensing*, vol. 9, no. 9, p. 907, 2017.

[11] H. Furukawa, "Deep learning for end-to-end automatic target recognition from synthetic aperture radar imagery," *arXiv preprint arXiv:1801.08558*, 2018.

[12] S. Kazemi, B. Yonel, and B. Yazici, "Deep learning for direct automatic target recognition from SAR data," in *2019 IEEE Radar Conference (RadarConf)*, 2019, pp. 1–6.

[13] M. B. Alver, S. Atito, and M. Çetin, "SAR ATR in the phase history domain using deep convolutional neural networks," in *Image and Signal Processing for Remote Sensing XXIV*, vol. 10789, 2018, p. 1078913.

[14] Q. Zhang, R. Plemmons, D. Kittle, D. Brady, and S. Prasad, "Joint segmentation and reconstruction of hyperspectral data with compressed measurements," *Applied optics*, vol. 50, no. 22, pp. 4417–4435, 2011.

[15] C. Hane, C. Zach, A. Cohen, R. Angst, and M. Pollefeys, "Joint 3d scene reconstruction and class segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 97–104.

[16] M. Storath, A. Weinmann, J. Frikel, and M. Unser, "Joint image reconstruction and segmentation using the potts model," *Inverse Problems*, vol. 31, no. 2, p. 025003, 2015.

[17] K. Bahrami, I. Rekik, F. Shi, and D. Shen, "Joint reconstruction and segmentation of 7t-like MR images from 3t MRI based on cascaded convolutional neural networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2017, pp. 764–772.

[18] L. Sun, Z. Fan, X. Ding, Y. Huang, and J. Paisley, "Joint CS-MRI reconstruction and segmentation with a unified deep network," in *International Conference on Information Processing in Medical Imaging*, 2019, pp. 492–504.

[19] K. Duman and B. Yazici, "Moving target artifacts in bistatic synthetic aperture radar images," *IEEE Transactions on Computational Imaging*, vol. 1, no. 1, pp. 30–43, 2015.

[20] N. Zhao, A. Basarab, D. Kouamé, and J.-Y. Tourneret, "Joint segmentation and deconvolution of ultrasound images using a hierarchical bayesian model based on generalized gaussian priors," *IEEE transactions on Image Processing*, vol. 25, no. 8, pp. 3736–3750, 2016.