

# Generating High Quality Visible Images from SAR Images Using CNNs

Puyang Wang                      Vishal M. Patel  
Department of Electrical and Computer Engineering  
Rutgers, The State University of New Jersey  
94 Brett Rd, Piscataway, NJ 08854  
{puyang.wang, vishal.m.patel}@rutgers.edu

**Abstract**—We propose a novel approach for generating high quality visible-like images from Synthetic Aperture Radar (SAR) images using Deep Convolutional Generative Adversarial Network (GAN) architectures. The proposed approach is based on a cascaded network of convolutional neural nets (CNNs) for despeckling and image colorization. The cascaded structure results in faster convergence during training and produces high quality visible images from the corresponding SAR images. Experimental results on both simulated and real SAR images show that the proposed method can produce visible-like images better compared to the recent state-of-the-art deep learning-based methods.

**Keywords**—Synthetic aperture radar image, despeckling, colorization.

## I. INTRODUCTION

Synthetic aperture radar (SAR) is a coherent radar imaging technology which is capable of producing high-resolution images of targets and landscapes. Due to its ability to capture images both at night and in bad weather conditions, SAR imaging has several advantages compared to optical and infrared systems. However, SAR images are often difficult to interpret mainly due to the following two reasons.

- 1) They are contaminated by multiplicative noise known as speckle. Speckle is caused by the constructive and destructive interference of the coherent returns scattered by small reflectors within each resolution cell [1].
- 2) Processed SAR images are often grayscale and they do not contain any color information.

These two issues often make the processing and interpretation of SAR images very difficult for both human interpreters and computer vision systems. Hence, despeckling and proper colorization are important for semantically interpreting the reflectivity field in SAR imaging.

Assuming that the SAR image is an average of  $L$  looks, the observed SAR image  $Y$  is related to the noise free image  $X$  by the following multiplicative model [2]

$$Y = F \odot X, \quad (1)$$

where  $F$  is the normalized fading speckle noise random variable and  $\odot$  denotes the element-wise multiplication. One common assumption on  $F$  is that it follows a Gamma distribution with unit mean and variance  $\frac{1}{L}$  and has the following



Fig. 1: A sample result of the proposed SAR-GAN method for SAR image to visible image translation. (a) Simulated input noisy SAR image. (b) Despeckled and colorized image.

probability density function [3]

$$p(F) = \frac{1}{\Gamma(L)} L^L F^{L-1} e^{-LF}, \quad (2)$$

where  $\Gamma(\cdot)$  denotes the Gamma function and  $F \geq 0$ ,  $L \geq 1$ .

Based on the above SAR observation model, various methods have been developed in the literature to suppress speckle. These include multi-look processing [4], [5], filtering methods [6], [7], [8], wavelet-based despeckling methods [9], [10], [11], [12], SAR block-matching 3D (SAR-BM3D) algorithm [13], Total Variation (TV) methods [14], and deep learning-based methods [15], [16]. Note that some of these methods apply homomorphic processing in which the multiplicative noise is transformed into an additive noise by taking the logarithm of the observed data [17].

Although state-of-the-art SAR image despeckling algorithms such as SAR-BM3D and wavelet-based methods are able to generate despeckled SAR images with sharp edges, the resulting despeckled images are often difficult to interpret due to their grayscale nature. For example, even after despeckling, it is difficult to distinguish between sandy land and grass field due to the grayscale nature of SAR images. Hence, generating a visible-like image from a SAR image is not only an interesting problem but also important for semantic segmentation and interpretation of SAR images. This problem shares some similarities with image colorization. However, there are some notable differences. First, in the image colorization domain

(grayscale image to RGB) the luminance is directly given by grayscale input, so only the chrominance needs to be estimated. Secondly, in the case of colorization techniques, in general noiseless grayscale images are given as input to obtain the RGB images. But in the case of SAR images, the input will have speckle and the expected output is the clean visible-like image with three RGB channels.

In this paper, we develop a deep learning-based method, called SAR-GAN, for the problem of SAR image to high quality visible image translation where we map a single channel noisy SAR image into a visible-like RGB image. Figure 1 shows a sample output from our SAR-GAN method. Given a simulated speckled SAR image shown in Figure 1 (a), SAR-GAN can generate not only the despeckled image but also the visible-like image as shown in Figure 1 (b). As can be seen by comparing Figure 1 (a) and Figure 1 (b), that our method is able to simultaneously denoise and colorize the simulated SAR image reasonably well.

## II. PROPOSED METHOD

In this section, we provide details of the proposed SAR-GAN method in which we aim to learn a mapping from input speckled SAR images to visible images for both noise removal and colorization. The proposed method consists of three main components: despeckling sub-network  $G_D$ , colorization sub-network  $G_C$  and generative adversarial learning. The primary goal of the despeckling sub-network is to restore a clean image from a noisy observation. The colorization sub-network then transforms the despeckled image into a visible image. Inspired by recent works on using generative adversarial learning for image colorization, we add the adversarial loss by introducing a discriminator network  $D$ . The adversarial loss, empirically, can in principle become aware that gray looking outputs are unrealistic, and encourage a wider color distribution. The composition of the two sub-networks, despeckling and colorization, forms the generator  $G$  in a typical generative adversarial network (GAN) framework as follows:

$$G = G_C \circ G_D. \quad (3)$$

The overall structure of the proposed SAR-GAN method containing two sub-networks and the training procedure is shown in Figure 2, where black arrow lines indicate data flows and red arrow lines denote network parameter updating. The detailed architectures of both sub-networks and loss functions are discussed in the following subsections.

### A. Despeckling Network

The detailed architecture of the despeckling sub-network is shown in Figure 3, where Conv, BN and ReLU stand for Convolution and Batch Normalization and Rectified Linear Unit, respectively. The despeckling CNN is adopted from our previous work [15] on SAR image restoration. Using a specific CNN architecture, we learn a mapping from an input SAR image into a despeckled image. One possible solution to the despeckling problem would be to transform the image into a logarithm space and then learn the corresponding mapping via CNN [18]. However, this approach needs extra steps to transfer the image into a logarithm space and from a logarithm space back to an image space. As a result, the overall algorithm

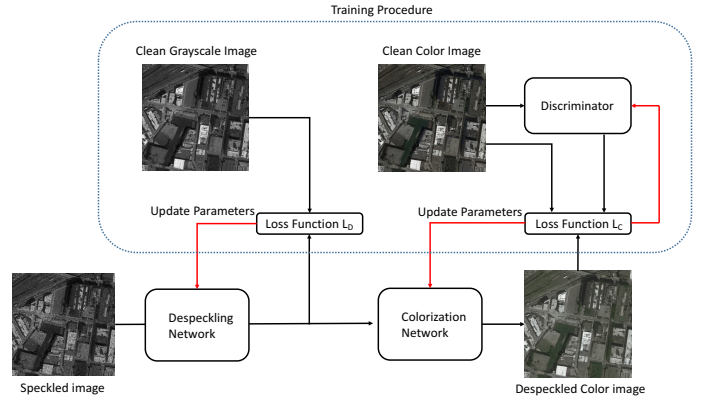


Fig. 2: Proposed SAR-GAN network architecture for SAR to visible image translation.

can not be learned in an end-to-end fashion. To address this issue, a division residual method is leveraged in our method where a noisy SAR image is viewed as a product of speckle with the underlying clean image (i.e. (1)). By incorporating the proposed component-wise division residual layer into the network, the convolutional layers are forced to learn the speckle component during the training process. In other words, the output before the division residual layer represents the estimated speckle. Then, the despeckled image is obtained by simply dividing the input image by the estimated speckle.

The noise-estimating part of despeckling sub-network consists of 8 convolutional layers (along with batch normalization and ReLU activation functions), with appropriate zero-padding to make sure that the output of each layer shares the same dimension with that of the input image. Batch normalization is added to alleviate the internal covariate shift by incorporating a normalization step and a scale and shift step before the nonlinearity in each layer. Each convolutional layer (except for the last convolutional layer) consists of 64 filters with stride of one. Then, the division residual layer with skip connection divides the input image by the estimated speckle. A hyperbolic tangent layer is stacked at the end of the network which serves as a non-linear function.

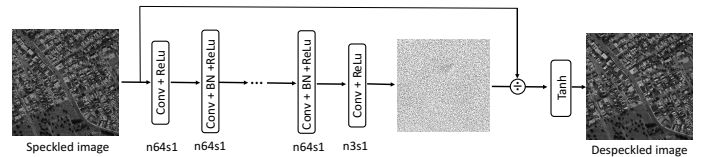


Fig. 3: Proposed network architecture for image despeckling.

### B. Colorization Network

Deep learning-based image colorization has been studied over the last couple of years [19], [20]. The key part of an image colorization neural network is to fully leverage the contextual information of an image for color translation. To extract and utilize the contextual information, one common way in deep learning is to use an encoder-decoder architecture in which an input image is encoded into a set of feature maps in the middle of the network. However, such a network

requires that all information flow passes through all the layers. For the image colorization problem, the sharing of low-level information between the input and output is important since the input and output should share the location of prominent edges. For the above reason, we add skip connections, following the general shape of an encoder-decoder CNN [21] as shown in Figure 4.

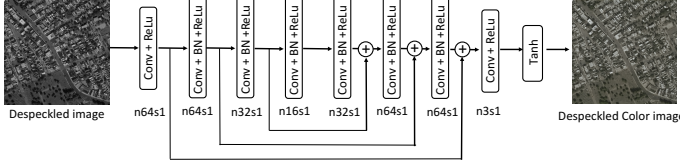


Fig. 4: Proposed network architecture for image colorization.

The colorization sub-network forms a symmetric encoder-decoder with 8 convolution layers and 3 skip connections. For each convolution layer, the kernel size is  $3 \times 3$ . Note that the  $n64s1$  in Figure 3 and 4 stands for 64 feature maps with one stride.

### C. Loss Functions

In a SAR image translation problem, it is important that the output image is noise free and realistic. One common loss function used in many image translation problems is the  $L_1$  loss. Given an image pair  $(X, Y)$ , where  $Y$  is the noisy input image and  $X$  is the corresponding ground truth, the per-pixel  $L_1$  loss is defined as

$$L_{L_1}(Y, X; G) = \frac{1}{CWH} \sum_{w=1}^W \sum_{h=1}^H \|G(Y^{c,w,h}) - X^{c,w,h}\|_1,$$

where  $G$  is the learned network and  $G(Y^{c,w,h})$  is the filtered image. Note that we have assumed that  $X$  and  $Y$  are of the same size  $C \times W \times H$  where  $C$  stands for the number of color channels. In this case, the network is trained to minimize the  $L_1$  distance between the output and the ground truth on the training set. Although the  $L_1$  loss has been shown to be very effective for image de-noising problem, it will incentivize an average, grayish color when it is uncertain which of several plausible color values a pixel should take on. In particular,  $L_1$  will be minimized by choosing the median of the conditional probability density function over possible colors. Hence, the  $L_1$  loss alone is not suitable for image colorization. Recent studies have shown adversarial loss, on the other hand, can in principle become aware that gray looking outputs are unrealistic, and encourage matching the true color distribution.

Given a set of  $N$  despeckled and colorized images,  $\{\hat{X}_i^{c,w,h}\}_{i=1}^N$ , generated from the generator  $G$ , the adversarial loss to guide the generator is defined as

$$L_A(\hat{X}; D, G) = -\frac{1}{N} \sum_{i=1}^N \log(D(\hat{X}_i^{c,w,h})), \quad (4)$$

where  $\hat{X}_i^{c,w,h} = G(Y_i^{c,w,h})$ . One of the issues with the adversarial loss is that it does not rely on the ground truth  $X$ . Hence, the results often contain artifacts that are not present

in the clean ground truth image. It tries to make the 'style' of the output closer to the training images.

Considering the pros and cons of both losses, we combine the per-pixel  $L_1$  loss and the adversarial loss together with appropriate weights to form our new refined loss function. The proposed loss function is defined as follows

$$L_D = L_{L_1}(\text{gray}(Y), \text{gray}(X); G_D), \quad (5)$$

$$L_C = L_{L_1}(Y, X; G_C) + \lambda_a L_A(\hat{X}; D, G_C), \quad (6)$$

$$L = L_D + L_C, \quad (7)$$

where  $\text{gray}(X)$  and  $\text{gray}(Y)$  are the corresponding grayscale versions of ground truth  $X$  and noisy  $Y$  with single channel, respectively. Here,  $L_D$  and  $L_C$  are loss functions for despeckling and colorization sub-network, respectively. The overall function  $L$  is the sum of  $L_D$  and  $L_C$ . The  $L_1$  loss in (5) makes the despeckling network  $G_D$  learn a mapping between the speckled input and clean ground truth. Loss function  $L_C$  for the colorization sub-network is a weighted sum of  $L_1$  and adversarial loss. Note that, for SAR images, the number of color channels is equal to 1. Hence, the dimension of input  $Y$  and  $G_D(Y)$  should be  $1 \times W \times H$  and  $3 \times W \times H$  for  $X$  and  $G(Y)$ .  $\lambda_a$  is a pre-defined weight for adversarial loss to balance the scale difference between losses. Because of the single combined loss function  $L$  we are able to train the network  $G$  which contains two sub-networks in an end-to-end fashion.

## III. EXPERIMENTAL RESULTS

To evaluate the effectiveness and performance of our proposed method, we present and compare results of our SAR-GAN with others methods. Since no similar work on despeckling and colorization of SAR images simultaneously has been done, we compare the performance of our method with that of the two CNN methods (CNN [22] and pix2pix [23]) and their combinations with the state-of-the-art despeckling algorithm SAR-BM3D [13]. For all the compared methods, parameters are set as suggested in their corresponding papers. For the basic CNN method, we adopt the network structure proposed in [22] and train the network using the same training dataset as used to train our network.

To train the proposed SAR-GAN network, we generate a dataset that contains 3292 image pairs. Training images are collected from the scraped Google Maps images [23] and the corresponding speckled images are generated using (1). All images are resized to  $512 \times 512$ . The entire network is trained using the ADAM optimization method [24], with mini-batches of size 12 and learning rate of 0.0002. During training, we set  $\lambda_a = 0.1$ . The architecture of the discriminator  $D$  is adapted from that in [25].

### A. Despeckling Performance

One key part of generating high quality visible images from SAR images is about removing as much speckle as possible meanwhile retaining the fine details. Therefore, we perform an experiment comparing despeckling performance of the proposed SAR-GAN and other SAR image despeckling algorithms including the state-of-the-art SAR-BM3D on synthetic SAR images. The outputs of the despeckling network  $G_D$  are used for comparison.





Fig. 5: (a) Ground truth. (b) Synthetic SAR image. (c) SAR-GAN despeckled. (d) SAR-GAN. (e) CNN. (f) CNN w/ despeckling. (g) pix2pix. (h) pix2pix w/ despeckling.

We randomly selected 85 speckled images out of all images in the dataset. The remaining images are used for training the network. Experiments are carried out on three different noise levels. In particular, the number of looks  $L$  is set equal to be 1, 4 and 10, respectively. The Peak Signal to Noise Ratio (PSNR), Structural Similarity Index (SSIM) [26], Universal Quality Index (UQI) [27], and Despeckling Gain (DG) [28] are used to measure the denoising performance of different methods. Average results of 85 test images corresponding this experiment are shown in Table I. As can be seen from this table, in all three noise levels, SAR-GAN provides the best performance compared to the other despeckling methods.

TABLE I: Quantitative results for various experiments on synthetic images.

	Metric	Noisy	Lee	Kuan	PPB	SAR-BM3D	ID-CNN
$L = 1$	PSNR	14.53	21.48	21.95	21.74	22.99	<b>24.74</b>
	SSIM	0.369	0.511	0.592	0.619	0.692	<b>0.727</b>
	UQI	0.374	0.450	0.543	0.488	0.591	<b>0.621</b>
	DG	-	16.01	17.08	14.30	17.17	<b>23.51</b>
$L = 4$	PSNR	18.49	22.12	22.84	23.72	24.96	<b>26.89</b>
	SSIM	0.525	0.555	0.650	0.725	0.782	<b>0.818</b>
	UQI	0.527	0.485	0.594	0.605	0.679	<b>0.723</b>
	DG	-	8.35	10.00	10.52	14.89	<b>19.33</b>
$L = 10$	PSNR	20.54	22.30	23.11	24.92	26.45	<b>28.07</b>
	SSIM	0.602	0.571	0.671	0.779	0.834	<b>0.853</b>
	UQI	0.599	0.498	0.613	0.678	0.745	<b>0.765</b>
	DG	-	4.06	5.93	7.75	13.61	<b>17.35</b>

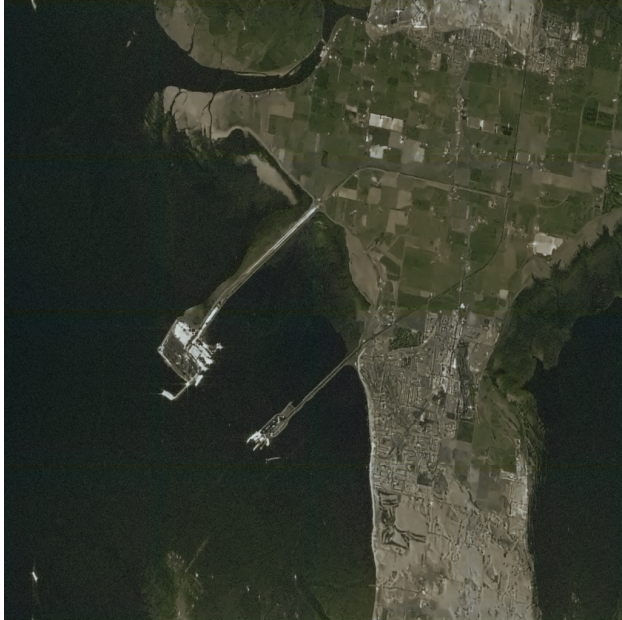




(a) SAR image.



(b) despeckled image by SAR-GAN.



(c) Visible image by SAR-GAN.



(d) Satellite image.

Fig. 6: Results of SAR-GAN on a real SAR image.

### B. Results on Synthetic Images

The despecking as well as colorization results on a synthetic image corresponding to different methods are shown in Figure 5. The details of the four compared methods are as follows:

- **CNN Network** is adopted from [22] and trained with only the  $L_1$  loss. The input and output are speckled image and generated visible image, respectively.
- **CNN w/ SAR-BM3D** The input images are first despeckled by SAR-BM3D and then fed into the network which is trained on image pairs of grayscale

image and the corresponding color image.

- **pix2pix** The  $L_1+cGAN$  network in [23] trained with the  $L_1$  and the adversarial losses. The input and output are speckled image and generated visible image, respectively.
- **pix2pix w/ SAR-BM3D** Similar to CNN w/ SAR-BM3D, the colorization network is replaced by the pix2pix network.

From Figure 5, we can clearly see that our proposed SAR-GAN performs the best overall. Compared with Figure 5 (e) and (g), our result in Figure 5 (d) suffers from less artifacts

because of better despeckling performance of the despeckling network. Furthermore, from (f) and (h) we see that SAR-BM3D helps to suppress speckle but at the cost of losing some detail information. Note that (e) and (f) both have some gray color in the final output. We believe that this is mainly due to the use of only the  $L_1$  loss in their networks.

### C. Results on Real SAR Images

Finally, we evaluate the performance of the proposed SAR-GAN on a real SAR image. Results are shown in Figure 6. The real SAR image shown in Figure 6 (a) is from the Vancouver scene of RADARSAT-1 operating on the C band [29]. Parameters of RADARSAT-1 for the Vancouver scene are as follows:

Sampling Rate  $F_r$  is 32.317 MHz, pulse duration  $T_r$  is 41.7  $\mu$ s and radar frequency  $f_0$  is 5.3 GHz. Figure 6 (d) is the satellite image captured on the same date as in (a). By comparing Figure 6 (c) and (d), we can clearly see that the proposed SAR-GAN is capable of generating high quality visible-like image from a real SAR image.

## IV. CONCLUSION

We proposed a novel approach for generating high quality visible-like images from SAR images using GAN architectures. The proposed approach is based on the usage of a cascaded model for despeckling and colorization in a progressive way. The cascaded structure allows a fast convergence during the training and obtains a greater similarity between the given SAR image and the corresponding visible image. The proposed approach has been evaluated on both simulated and real SAR images and it is shown that the proposed approach can provide better colorization compared to some of the recent deep learning-based methods.

## ACKNOWLEDGMENT

This work was supported by an ARO grant W911NF-16-1-0126.

## REFERENCES

- [1] J. W. Goodman, "Some fundamental properties of speckle," *Journal of the Optical Society of America*, vol. 66, no. 11, pp. 1145–1150, Nov 1976.
- [2] F. Ulaby and M. C. Dobson, *Handbook of Radar Scattering Statistics for Terrain*. Norwood, MA: Artech House, 1989.
- [3] F. T. Ulaby and M. C. Dobson, *Handbook of radar scattering statistics for terrain*. Artech House, 1989.
- [4] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*. Norwood, MA: Artech House, 1998.
- [5] P. Thompson, D. E. Wahl, P. H. Eichel, D. C. Ghiglia, and C. V. Jakowatz, *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach*. Norwell, MA, USA: Kluwer Academic Publishers, 1996.
- [6] J.-S. Lee, "Speckle analysis and smoothing of synthetic aperture radar images," *Computer graphics and image processing*, vol. 17, no. 1, pp. 24–32, 1981.
- [7] 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 Transactions on pattern analysis and machine intelligence*, no. 2, pp. 157–166, 1982.
- [8] A. Baraldi and F. Parmiggiani, "A refined gamma map sar speckle filter with improved geometrical adaptivity," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, no. 5, pp. 1245–1257, 1995.
- [9] H. Xie, L. E. Pierce, and F. T. Ulaby, "Sar speckle reduction using wavelet denoising and markov random field modeling," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2196–2212, Oct 2002.
- [10] F. Argenti and L. Alparone, "Speckle removal from sar images in the undecimated wavelet domain," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 11, pp. 2363–2374, Nov 2002.
- [11] A. Achim, P. Tsakalides, and A. Bezerianos, "Sar image denoising via bayesian wavelet shrinkage based on heavy-tailed modeling," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 8, pp. 1773–1784, Aug 2003.
- [12] V. M. Patel, G. R. Easley, R. Chellappa, and N. M. Nasrabadi, "Separated component-based restoration of speckled sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 2, pp. 1019–1029, Feb 2014.
- [13] S. Parrilli, M. Poderico, C. V. Angelino, and L. Verdoliva, "A nonlocal sar image denoising algorithm based on lmmse wavelet shrinkage," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 2, pp. 606–616, 2012.
- [14] J. M. Bioucas-Dias and M. A. T. Figueiredo, "Multiplicative noise removal using variable splitting and constrained optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 7, pp. 1720–1730, July 2010.
- [15] P. Wang, H. Zhang, and V. M. Patel, "Sar image despeckling using a convolutional neural network," *IEEE Signal Processing Letters*, vol. 24, no. 12, pp. 1763–1767, Dec 2017.
- [16] —, "Generative adversarial network-based restoration of speckled sar images," in *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*. IEEE, 2017.
- [17] F. Argenti, A. Lapini, T. Bianchi, and L. Alparone, "A tutorial on speckle reduction in synthetic aperture radar images," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 3, pp. 6–35, Sept 2013.
- [18] G. Chierchia, D. Cozzolino, G. Poggi, and L. Verdoliva, "Sar image despeckling through convolutional neural networks," *arXiv preprint arXiv:1704.00275*, 2017.
- [19] R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," in *European Conference on Computer Vision*. Springer, 2016, pp. 649–666.
- [20] S. Izuka, E. Simo-Serra, and H. Ishikawa, "Let there be color!: Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 110:1–110:11, Jul. 2016.
- [21] X. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Advances in Neural Information Processing Systems*, 2016, pp. 2802–2810.
- [22] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, July 2017.
- [23] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *arXiv preprint arXiv:1611.07004*, 2016.
- [24] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [25] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.
- [26] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [27] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE signal processing letters*, vol. 9, no. 3, pp. 81–84, 2002.
- [28] G. Di Martino, M. Poderico, G. Poggi, D. Riccio, and L. Verdoliva, "Benchmarking framework for sar despeckling," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 3, pp. 1596–1615, 2014.
- [29] I. Cumming and F. Wong, *Digital Processing of Synthetic Aperture Radar Data: Algorithms and Implementation*. Norwood, MA: Artech House, 2005.