MD-Glow: Multi-task Despeckling Glow for SAR Image Enhancement

Shunsuke Takao
University of Tsukuba
Tsukuba, Japan
takao.s.work@gmail.com

Abstract

Synthetic Aperture Radar (SAR) is a widespread technology to measure data in earth observation or environmental monitoring, however severe speckle noise degrades SAR image quality which hinders in-depth image analysis. Speckle noise in a SAR image is mitigated by deep learning methods, although they suffer from a domain gap between training and testing data to pose problems such as oversmoothing in the recovered images. It is also challenging for practical applications to enhance satellite imagery by a single model, often having multiple scales and resolutions measured by different bands or sensing devices. In this paper, we propose a novel scheme of Multi-task Despeckling Glow, dubbed as MD-Glow*, for SAR image enhancement. Our deep model suppresses speckle noise (despeckling task) while simultaneously enhancing image resolution (super-resolution task) in a unified scheme, to effectively address the above problems for despeckling. We conduct numerical experiments on the despeckling task utilizing datasets composed of synthetic and real SAR images to demonstrate that the proposed MD-Glow produces competitive results in comparison to other despeckling methods.*

1. Introduction

Synthetic Aperture Radar (SAR) is a remote sensing technology to measure data for many applications in earth observation or environmental monitoring, such as change detection, terrain monitoring, and urban planning [3]. SAR images are beneficial to enable cost-effective sensing in all weather conditions in contrast to optical imagery, they, however, suffer from severe speckle noise that significantly degrades SAR image quality which impairs in-depth image analysis [8].

The speckle noise in a SAR image, which is caused by coherent imaging, is mitigated by despeckling methods along with rapid development of deep learning. Many supervised deep models are trained by synthetic SAR images, which are generated from optical images on a noise model, to cope with the difficulty to collect clear real SAR images [5, 13, 19, 20, 26]. While promoting performance progress of supervised despeckling methods, the trained deep models often suffer from domain gap among real and synthetic SAR images, which leads problems such as oversmoothing or hallucination in recovered images [8].

The other issue is that the scale and resolution of satellite imagery are diverse across the target bands due to being captured from different sensor devices, which may require retraining of deep models per input image resolution [21]. To better generalize over domains in practical scenarios, which often contain multiple scale satellite imagery, representation learning is recently incorporated to reconstruct multiscale optical images by Masked Autoencoder (MAE) for classification or segmentation tasks [17]. To the best of our knowledge, multiscale information is currently less utilized in despeckling for SAR images.

In this paper, we propose a novel scheme of *Multi-task* despeckling to effectively deal with the above issues for SAR images; our proposed scheme, dubbed as *MD-Glow*, consists of despeckling and super-resolution (SR) tasks tailored for SAR images that often accompany severe speckle noise. Our MD-Glow is equipped with the common encoder and decoder with respect to despeckling and SR tasks to augment image characteristics, such as edges or contours in the recovered image. As for encoder/decoder, we utilize *invertible* Glow network [12] to encode image characteristics into effectual feature representation by extracting bijective image features that is shared by the both tasks. The encoded feature representation is then transformed by task-specific feature translation modules of MD-Glow to produce despeckled and higher-resolution images for both tasks. We also apply random downsampling to the speckled image in training to produce multi-resolution images, similar to real satellite data observed from diverse sensors [17, 21]. Thereby, we can suppress speckle noise while effectively enhancing high frequency signals for SAR images at multiple scale. Overview of our MD-Glow is shown in Fig. 1. We conduct numerical experiments for the despeckling task to demonstrate that the proposed MD-Glow produces com-

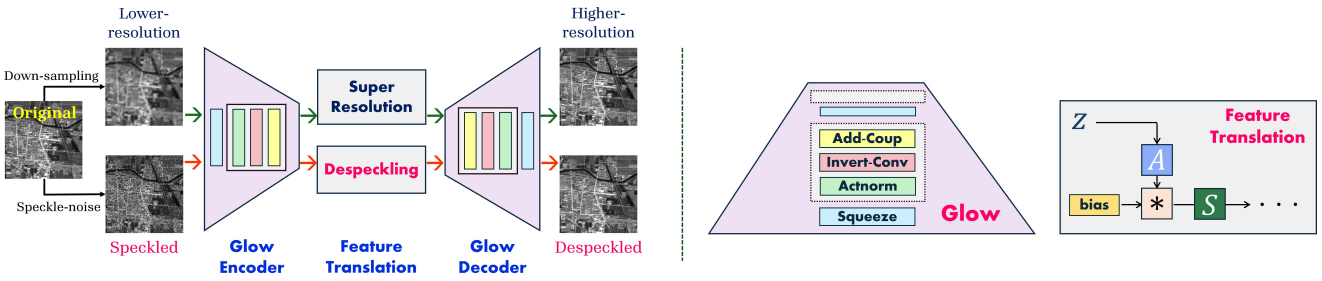


Figure 1. Overview of the MD-Glow. MD-Glow is trained by *multi-task* learning to simultaneously produce despeckled and higher-resolution images with the Glow modules [12]. Invertible convolution and additive coupling in Glow, which are defined in Section 3.2, are abbreviated by Add-Coup and Invert-Conv in the figure. Feature translation module is constructed on Π-Net [6] by the 3rd order NCP block with polynomial transform. The latent variable z is processed by affine transformation A and the last block S in feature translation.

petitive performance on both synthetic and real SAR images in comparison to the other despeckling techniques.

Our main contributions are summarized as following.

- We explore the potential of multi-task learning that is composed of despeckling and SR tasks to showcase the superiority over the single-task training for SAR image despeckling.
- We propose *MD-Glow* built upon invertible Glow to encode image characteristics into effective feature representation that is shared by despeckling and SR tasks.
- Experimental results on benchmark datasets for synthetic and real SAR images demonstrate the efficacy of the proposed MD-Glow in comparison to the other despeckling methods.

2. Related Works

2.1. Despeckling Methods for SAR Images

The speckle is a spatially correlated multiplicative noise which occurs in a coherent imaging system. The speckle noise in SAR images has been alleviated by classic despeckling methods such as filtering approaches or methods based on wavelet transform to smooth out the high frequency artifacts [3, 8]. Recent deep learning methods have addressed severe speckle noise by Convolutional Neural Network (CNN) [26], Generative Adversarial Network (GAN) [25], and Vision Transformer (ViT) [19], where often equipped with residual connections to preserve image characteristics. However, the despeckling ability may be limited in supervised deep models to sometimes produce outputs showing oversmoothing or hallucination [8], which is attributed to the domain gap problem between real and synthetic SAR images. To mitigate the problem, self-supervised learning is introduced for despeckling; these methods are trained without ground-truth SAR images under some assumptions such as uncorrelated speckle [8, 15], of which may not consistent with the real conditions.

In our MD-Glow, we apply multi-task learning for enhancing finer image characteristics to alleviate oversmoothing effectively in the despeckled image.

2.2. Multi-task Learning in Remote Sensing

Multi-task learning contributes to effective training and alleviates overfitting by simultaneously optimizing different tasks with similar characteristics [1, 4, 24]. Task-specific features are effectively learned through a soft-attention module for each task [9, 14]. The loss function weights for each task are dynamically set by means of task uncertainty [10], change ratio of the loss score [14] or gradient magnitude [9]. In the remote sensing literature, optical satellite imagery is converted by the single encoder to encode effective feature representation and then processed through task-specific decoders to produce two kinds of images corresponding to the segmentation task and the height estimation task [4].

The proposed MD-Glow is equipped with the task-specific feature translation modules in contrast to previous models having individual decoders for each task [1, 4]. Moreover, despeckling and SR tasks are trained by a unified, end-to-end manner in our MD-Glow compared to the single-task trained methods [13, 16, 19, 20, 23, 26].

3. Formulation and Method

3.1. Speckle Model of SAR Images

Let SAR intensity be denoted by Y which is degraded by multiplicative noise N as following equation [8, 19, 20].

$$Y = XN, \tag{1}$$

where X is the speckle-free image. The speckle noise N is assumed to subject to a Gamma distribution given by

$$p(N) = \frac{1}{\Gamma(N)} N^{L-1} L^L e^{-LN}, \tag{2}$$

where $\Gamma(\cdot)$ is the Gamma function and L is the number of looks of the SAR image. In this work, we address the single-look case $L = 1$.

3.2. MD-Glow Model

We describe the framework of proposed MD-Glow for SAR image despeckling, where overview is shown in Fig. 1.

The MD-Glow model is composed of Glow modules [12] for encoder and decoder, and feature translation module. The Glow module encodes image characteristics of an original image into effective feature representation. For multi-task learning, encoded image features are processed by two feature translation modules to enhance degraded image features for despeckling and super-resolution (SR) tasks. The translated two kinds of feature representation are projected to the image space by the decoder to generate despeckled images and higher-resolution images. The network of MD-Glow is detailed as follows.

3.2.1 Glow Modules for Encoding and Decoding

We design the Glow modules by two Glow blocks which consist of a squeeze operator and eight flow blocks.

Input feature map $\boldsymbol{F} \in \mathbb{R}^{H \times W \times C}$ of height H, width W, and channel C is first processed by the $\texttt{squeeze}$ operation to expand feature map by factor of 2 [12]; $\boldsymbol{F}' = \texttt{squeeze}(\boldsymbol{F}) \in \mathbb{R}^{\frac{H}{2} \times \frac{W}{2} \times 4C}$. The squeezed feature map is translated by applying eight sets of flow blocks containing

$$\text{Actnorm: } \boldsymbol{F_1} = a\boldsymbol{F}_{in} + b, \tag{3}$$

$$\text{Invertible convolution: } \boldsymbol{F_2} = \boldsymbol{W} * \boldsymbol{F_1}, \tag{4}$$

$$\text{Additive coupling: } \boldsymbol{F}_{out} = \texttt{concat}(\boldsymbol{F_2^t}, \varphi_{\boldsymbol{\theta}}(\boldsymbol{F_2^t}) + \boldsymbol{F_2^b}), \tag{5}$$

where $\boldsymbol{F_2^t}, \boldsymbol{F_2^b}$ is the top and bottom halves of $\boldsymbol{F_2}$ along channel dimension, denoted by $\boldsymbol{F_2} = \texttt{concat}\left(\boldsymbol{F_2^t}, \boldsymbol{F_2^b}\right)$. φ is Convolutional Neural Network (CNN) parametrized by $\boldsymbol{\theta}$. Here, the actnorm is defined as a simple affine transformation having trainable parameters a, b. Invertible convolution $\boldsymbol{W}$ is implemented by using the 1×1 kernel to flexibly permute channel components by training [12]. The additive coupling layer [7] transforms feature maps by the simple invertible operation.

The Glow decoding module is constructed by the inverse processes with a reversed order of the encoder, as the above operations are all invertible. It is noteworthy that parameters of encoder and decoder are shared [2] in this study.

3.2.2 Feature Translation Module

We build the feature translation module upon the 3rd order NCP block of Π-Net [6], where feature maps are converted by polynomial transform.

The encoded feature maps $\boldsymbol{X} \in \mathbb{R}^{H' \times W' \times D}$ are first processed through the pre-mapping to produce $\boldsymbol{z} \in \mathbb{R}^{H' \times W' \times D'}$ with a series of eight layers of convolution and ReLU. Affine transformation A is then applied to the projected feature map z, which is subsequently fed into the last block S composed of two layers of convolution and ReLU. In our MD-Glow, we recursively repeat the above processing three times by the 3rd order polynomial [6] to correct degraded image features. The kernel size of all convolutional layers is set to 1.

3.2.3 Multi-Task Learning

We train the MD-Glow model on despeckling and SR tasks for SAR images. To train the MD-Glow in a multi-task way, the common Glow encoding module is applied to two kinds of images, a speckled image and a low-resolution image, for each task (Fig. 1). The transformed feature representations are processed by the two individual feature translation modules for the respective tasks, and then projected back to image space by the Glow decoder. Thereby, MD-Glow produces a despeckled image $\hat{I}_{desp}$ and a super-resolved image $\hat{I}_{SR}$ for both tasks $t \in \{desp, SR\}$. The MD-Glow is trained by the following loss function of

$$\mathcal{L} = \sum_{t \in \{desp, SR\}} \|\boldsymbol{I} - \hat{\boldsymbol{I}}_t\|_1, \tag{6}$$

where $\boldsymbol{I}$ is a ground-truth SAR image, which is common in the two tasks, and $\|\cdot\|_1$ represents L1 distance. We empirically set the loss coefficients for each task $t \in \{desp, SR\}$ in this work. Applying dynamic weighting to balance loss coefficient for each task [9, 10, 14] is our future work.

4. Experiments

4.1. Setting

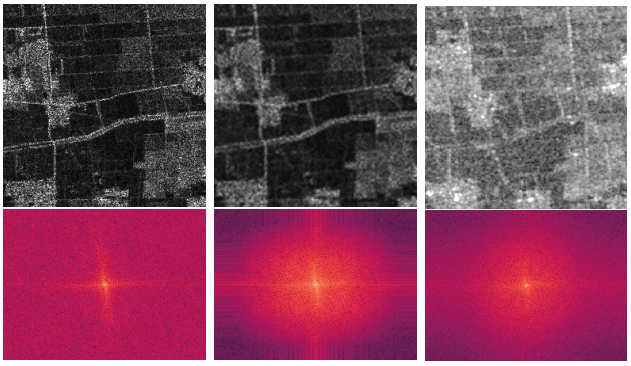
To train the MD-Glow, we synthesize speckled images from optical satellite imagery of *SEN1-2* [22] by following equation 1 and equation 2; we select 5025 images from the *SEN1-2* dataset of 256 × 256 resolution for fast training. We train MD-Glow by *multi-task* learning composed of despeckling and SR tasks. In the despeckling task, we apply *downsampling* to the synthetic speckled images at random to mimic real SAR images of multiple resolution that are often measured by different sensors [17, 21]. Meanwhile, in the SR task, bicubic downsampling is applied to the original images by factor of 2 to produce lower resolution input images. Random flipping is applied for both tasks in horizontal and vertical axes. In training our MD-Glow, we leverage Adam optimizer [11] with the learning rate 0.0001. We halve the learning rate every 40 epochs for 100-epoch training and set the batch size to 16.

Table 1. PSNR and SSIM of ablation study. The network (a) is trained by despeckling task (*single*-task) to produce speckle-free images, while the other (b), (c), and (d) are trained by *multi*-task learning (despeckling and super-resolution tasks) to simultaneously produce speckle-free image and higher-resolution image. The *invertible* (d) consists of the common encoder and decoder, having the same parameters, where the non-invertible (c) contains different parameters among encoder and decoder.

Network	(a)	(b)	(c)	(d)
Multi-task		✓	✓	✓
Downsampling			✓	✓
Invertibility				✓
SEN1-2 (LR) Dataset				
PSNR	21.996	22.579	25.361	**25.412**
SSIM	0.545	0.588	**0.625**	0.623
SEN1-2 Dataset [22]				
PSNR	25.755	**25.804**	25.746	25.726
SSIM	0.657	0.653	**0.663**	0.654

Input image (a) Single-Task (b) Multi-Task (c) Downsampling+(b) (d) Invertibility+(c)

Figure 2. Visual comparison in ablation study. Also refer to Table 1 which details the setting of the ablated results.



Speckled Downsampled Real

Figure 3. Samples of SAR images (top) and frequency features converted by Fourier Transform (bottom).

Input image (c) non-invertible (d) invertible

Figure 4. Real SAR image and recovered results by the non-invertible model (c) and the invertible model (d). High frequency artifacts are observed in (c) in comparison to (d). Setting is shown in Table 1.

To evaluate the MD-Glow, we leverage 525 test images, which are randomly selected in the *SEN1-2* dataset, and also utilize a low-resolution (LR) *SEN1-2* dataset to verify the robustness of MD-Glow; the LR dataset is produced by downsampling to halve its resolution (Fig. 3) for simulating diverse scaled real satellite imagery [17, 21]. The evaluations on the LR dataset are motivated by the similarity between the downsampled images and the low-resolution, real SAR images; In Fig. 3, the real SAR image (3rd column) contains less high-frequency features as observed in the downsampled image (2nd column) in comparison to the original image having more high-frequency signals (1st column).

We adopt Peak signal-to-noise ratio (PSNR) and Structural Similarity Index Measure (SSIM) to evaluate MD-

Figure 5. Comparison of despecked image and Super Resolution. The red box in the images is cropped for more details.

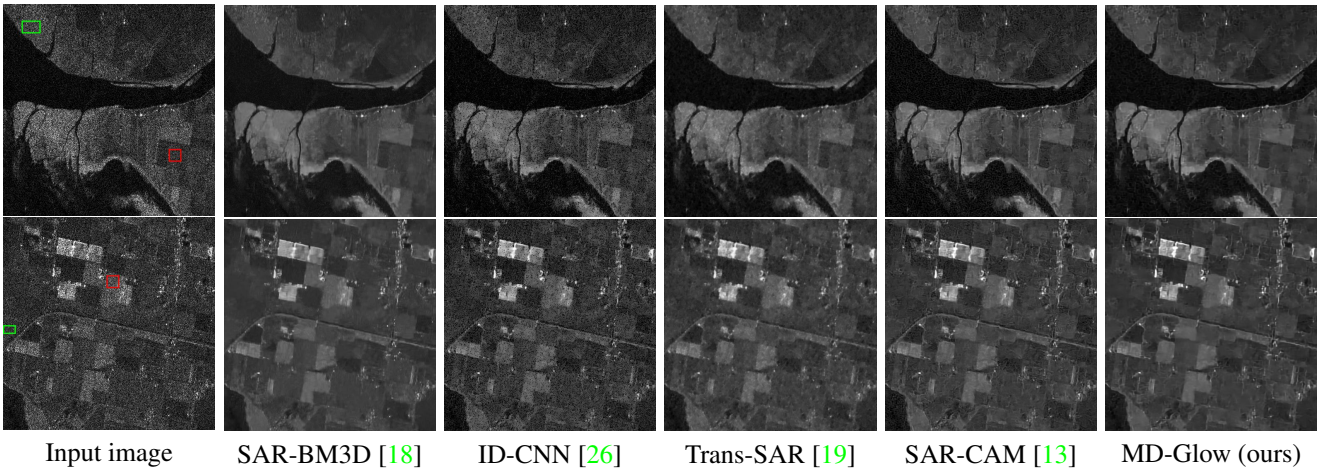


Figure 6. Despeckled results on SAR images of higher resolution. ENL is computed for Image-1 (top) and Image-2 (bottom).

Glow on the synthetic datasets, while Equivalent Number of Looks (ENL) is computed on real SAR images that usually contain no ground-truth images [3].

4.2. Ablation Study

4.2.1 Effect of MD-Glow Components

We conduct ablation studies of the proposed MD-Glow to verify the efficacy of Multi-task learning, random downsampling, and invertible structure for SAR image despeckling. We evaluate the despeckling performance of the ablated methods on the synthetic dataset from *SEN1-2* [22].

Table 1 shows performance comparison of the ablated methods. As shown in (a) and (b) of Table 1, multi-task learning facilitates the performance improvement in PSNR and SSIM scores of our MD-Glow. By comparing (c) with (b), it can be observed that random downsampling, which is applied in the despeckling task, contributes to improving PSNR and SSIM scores, especially on LR images. It may be due to the similarity of the task to enhance the LR dataset, where speckle reduction and image resolution improvement are required. Fig. 2 visually supports the despeckling ability of MD-Glow on a synthetic SAR image (1st column).

Table 2. ENL on real SAR images. ENL is computed for the red box and the green box within the images in Fig. 6.

Method	Image-1 Red box	Image-1 Green box	Image-2 Red box	Image-2 Green box
SAR-BM3D [18]	230.190	90.267	198.844	**1210.432**
ID-CNN [26]	4.565	6.924	10.238	6.027
Trans-SAR [19]	103.678	93.454	139.832	158.468
SAR-CAM [13]	111.704	80.620	157.028	303.969
MD-Glow (ours)	**311.466**	**106.018**	**263.328**	1170.122

The single-task trained model partially reduces the speckle noise, yet the results are oversmoothed and blurred (a). Multi-task training augments high-frequency, finer image structures within the recovered image (b), however, fail to remove speckle noises a little. By applying random downsampling, speckle noises are more removed while remaining clear contours in (c) compared to (b).

It should be noted that the invertible structure slightly drops PSNR and SSIM scores (d) in comparison to the non-invertible model (c), that may be caused by less parameters to share the same encoder and decoder in the MD-Glow. In

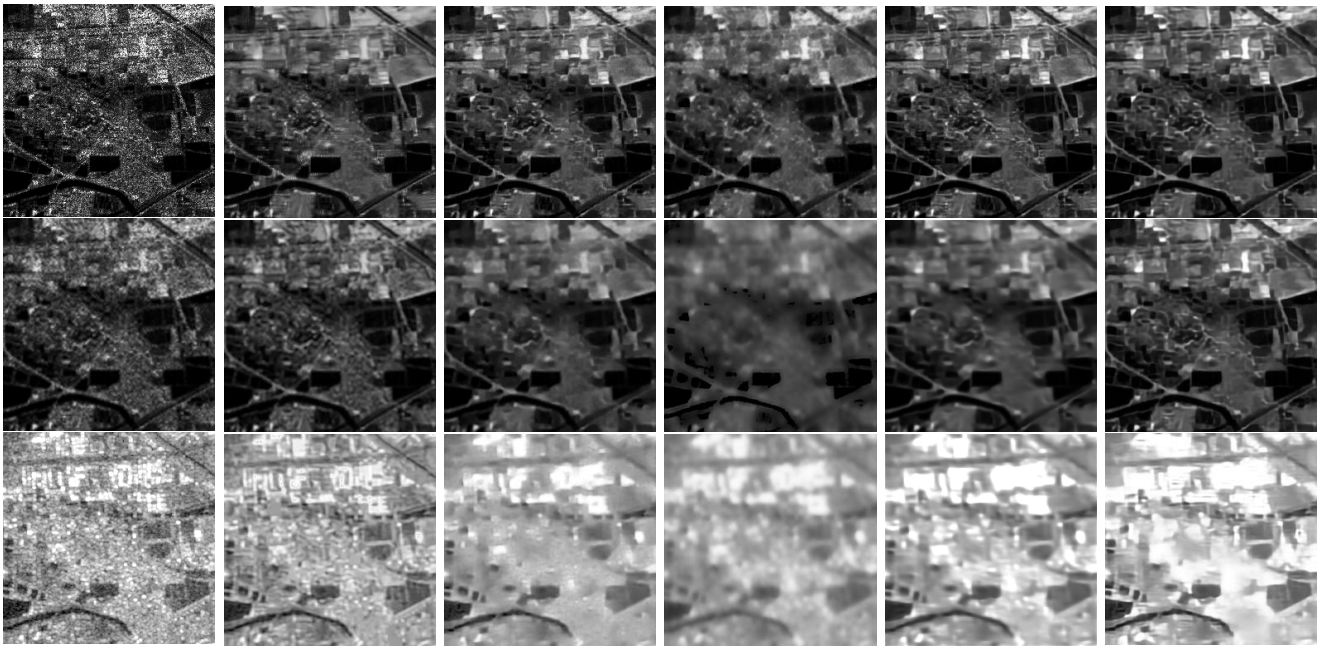


Figure 7. Despeckled results on SAR images of lower resolution. The synthetic speckled image (1st row) and the lower resolution image (2nd row), as well as real SAR image (3rd row) are shown.

the recovered image by the invertible model (d), we observe that the high-frequency artifacts are suppressed in a real SAR image (Fig. 4) compared to the non-invertible model, therefore, we adopt the invertible structure for MD-Glow.

4.2.2 Super Resolution to further Enhance Despeckled Images

We explore the possibility to enhance image resolution of a despeckled image by applying SR to the recovered image that is simultaneously trained in the MD-Glow. Fig. 5 shows visual sample images of a synthetic SAR image. The synthesized speckle noise is removed by the MD-Glow (2nd column), while its finer contours are slightly blurred in the recovered image (3rd column). The despeckled image is then fed to the SR branch of MD-Glow to produce more clear image (4th column), having sharper image characteristics (5th column). The above attempt may be beneficial for widespread remote sensing applications to effectively utilize lower resolution images [16].

4.3. Comparison to other Despeckling Methods

We compare our MD-Glow with the other despeckling methods; ID-CNN [26], Trans-SAR [19], SAR-CAM [13], and SAR-BM3D [18]. We leverage codes released by the authors and retrained the methods on the *SEN1-2* dataset except for the unsupervised SAR-BM3D.

Table 3. PSNR and SSIM on synthetic SAR images

Method	*SEN1-2* (LR) Dataset PSNR / SSIM	*SEN1-2* Dataset PSNR / SSIM
SAR-BM3D [18]	22.672 / 0.593	25.165 / 0.637
ID-CNN [26]	23.600 / 0.565	25.486 / 0.655
Trans-SAR [19]	21.434 / 0.446	23.371 / 0.545
SAR-CAM [13]	23.802 / 0.548	**26.012 / 0.683**
MD-Glow (ours)	**25.412 / 0.623**	25.726 / 0.654

We examine the despeckling ability on SAR images for higher resolution in Section 4.3.1 and lower resolution in Section 4.3.2, to explore the applicability for real remote sensing imagery that often measured at multiple scale or targeting bands [17, 21].

4.3.1 Despeckling on high resolution (HR) SAR images

Fig. 6 shows higher resolution, real SAR images degraded by speckle noise (1st column), where ENL is computed on the red and green boxes. The proposed MD-Glow despeckles the high-frequency noise while maintaining edges or contours effectively in the recovered image. SAR-CAM [13] and Trans-SAR [19] fail to remove finer speckle noise in the recovered images. SAR-BM3D [18] slightly blurs contours in the recovered image. Severe high frequency artifacts are left in ID-CNN [26] results.

In Table 2 for quantitative comparison, the proposed MD-Glow produces higher or competitive ENL scores compared to the other techniques for despeckling. Note that SAR-BM3D is subject to its hyper parameters and needs parameter tuning in practical setting.

4.3.2 Despeckling on low resolution (LR) SAR images

Table 3 show PSNR and SSIM results computed on the *SEN1-2* dataset. In Table 3, our MD-Glow produces competitive scores in comparison to the other despeckling techniques, getting the highest score in the LR dataset and the second result in the original *SEN1-2* dataset.

Meanwhile, in the visual sample images shown in Fig. 7, the proposed MD-Glow despeckles high-frequency noises on the synthetic SAR image (1st, 2nd row) and a real SAR image (3rd row). SAR-CAM [13] slightly blurs contours of the real SAR image (3rd row). SAR-BM3D [18] cannot remove artifacts in the low-resolution SAR image (2nd and 3rd row). Recovered results are blurred in Trans-SAR [19] or ID-CNN [26]. Note that the detailed image structure is degraded by the high-frequency speckle noise to be lost in the recovered images (3rd row), due to the low resolution of the original image.

5. Conclusion

We proposed Multi-task Despeckling Glow, dubbed as *MD-Glow*, for SAR image despeckling. The proposed multi-task learning framework was composed of despeckling and super-resolution (SR) tasks to exploit the feature representation, that is common to each task, for effectively enhancing high frequency signals with less artifacts in the recovered images. We demonstrate the efficacy of our MD-Glow on synthetic and real SAR images to produce competitive performance compared to the other despeckling techniques.

Acknowledgment

In conducting this research, Yukihiro Yano and Takahiro Kumura in Visual Intelligence Research Labs, NEC Corporation, provided valuable suggestions and kind support regarding satellite data. The author would like to express the sincere gratitude.

References

[1] Mahmoud Afifi and Michael S Brown. Deep white-balance editing. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pages 1397–1406, 2020. 2

[2] Jie An, Siyu Huang, Yibing Song, Dejing Dou, Wei Liu, and Jiebo Luo. Artflow: Unbiased image style transfer via reversible neural flows. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 862–871, 2021. 3

[3] Fabrizio Argenti, Alessandro Lapini, Tiziano Bianchi, and Luciano Alparone. A tutorial on speckle reduction in synthetic aperture radar images. *IEEE Geoscience and remote sensing magazine*, 1(3):6–35, 2013. 1, 2, 5

[4] Marcela Carvalho, Bertrand Le Saux, Pauline Trouvé-Peloux, Frédéric Champagnat, and Andrés Almansa. Multitask learning of height and semantics from aerial images. *IEEE Geoscience and Remote Sensing Letters*, 17(8):1391–1395, 2019. 2

[5] Giovanni Chierchia, Davide Cozzolino, Giovanni Poggi, and Luisa Verdoliva. Sar image despeckling through convolutional neural networks. In *2017 IEEE international geoscience and remote sensing symposium (IGARSS)*, pages 5438–5441. IEEE, 2017. 1

[6] Grigorios G Chrysos, Stylianos Moschoglou, Giorgos Bouritsas, Jiankang Deng, Yannis Panagakis, and Stefanos Zafeiriou. Deep polynomial neural networks. *IEEE transactions on pattern analysis and machine intelligence*, 44(8):4021–4034, 2021. 2, 3

[7] Laurent Dinh, David Krueger, and Yoshua Bengio. Nice: Non-linear independent components estimation. *arXiv preprint arXiv:1410.8516*, 2014. 3

[8] Giulia Fracastoro, Enrico Magli, Giovanni Poggi, Giuseppe Scarpa, Diego Valsesia, and Luisa Verdoliva. Deep learning methods for synthetic aperture radar image despeckling: An overview of trends and perspectives. *IEEE Geoscience and Remote Sensing Magazine*, 9(2):29–51, 2021. 1, 2

[9] Ankit Jha, Awanish Kumar, Biplab Banerjee, and Subhasis Chaudhuri. Adamt-net: An adaptive weight learning based multi-task learning model for scene understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 706–707, 2020. 2, 3

[10] Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7482–7491, 2018. 2, 3

[11] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 3

[12] Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. *Advances in neural information processing systems*, 31, 2018. 1, 2, 3

[13] Jaekyun Ko and Sanghwan Lee. Sar image despeckling using continuous attention module. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:3–19, 2021. 1, 2, 5, 6, 7

[14] Shikun Liu, Edward Johns, and Andrew J Davison. End-to-end multi-task learning with attention. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1871–1880, 2019. 2, 3

[15] Andrea Bordone Molini, Diego Valsesia, Giulia Fracastoro, and Enrico Magli. Speckle2void: Deep self-supervised sar despeckling with blind-spot convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–17, 2021. 2

[16] Max Muzeau, Chengfang Ren, Jeremy Fix, Frederic Brigui, and Jean Philippe Ovarlez. On the impact of despeckling for supervised sar super-resolution. In *EUSAR 2024; 15th European Conference on Synthetic Aperture Radar*, pages 469–474. VDE, 2024. 2, 6

[17] Mubashir Noman, Muzammal Naseer, Hisham Cholakkal, Rao Muhammad Anwer, Salman Khan, and Fahad Shahbaz Khan. Rethinking transformers pre-training for multi-spectral satellite imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27811–27819, 2024. 1, 3, 4, 6

[18] Sara Parrilli, Mariana Poderico, Cesario Vincenzo Angelino, and Luisa Verdoliva. A nonlocal sar image denoising algorithm based on llmmse wavelet shrinkage. *IEEE Transactions on Geoscience and Remote Sensing*, 50(2):606–616, 2011. 5, 6, 7

[19] Malsha V Perera, Wele Gedara Chaminda Bandara, Jeya Maria Jose Valanarasu, and Vishal M Patel. Transformer-based sar image despeckling. In *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, pages 751–754. IEEE, 2022. 1, 2, 5, 6, 7

[20] Malsha V Perera, Nithin Gopalakrishnan Nair, Wele Gedara Chaminda Bandara, and Vishal M Patel. Sar despeckling using a denoising diffusion probabilistic model. *IEEE Geoscience and Remote Sensing Letters*, 20:1–5, 2023. 1, 2

[21] Colorado J Reed, Ritwik Gupta, Shufan Li, Sarah Brockman, Christopher Funk, Brian Clipp, Kurt Keutzer, Salvatore Candido, Matt Uyttendaele, and Trevor Darrell. Scale-mae: A scale-aware masked autoencoder for multiscale geospatial representation learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4088–4099, 2023. 1, 3, 4, 6

[22] Michael Schmitt, Lloyd Hughes, and Xiao Xiang Zhu. The sen1-2 dataset for deep learning in sar-optical data fusion. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 141–146, 2018. 3, 4, 5

[23] Huanfeng Shen, Liupeng Lin, Jie Li, Qiangqiang Yuan, and Lingli Zhao. A residual convolutional neural network for polarimetric sar image super-resolution. *ISPRS Journal of Photogrammetry and Remote Sensing*, 161:90–108, 2020. 2

[24] Di Wang, Jing Zhang, Minqiang Xu, Lin Liu, Dongsheng Wang, Erzhong Gao, Chengxi Han, Haonan Guo, Bo Du, Dacheng Tao, et al. Mtp: Advancing remote sensing foundation model via multi-task pretraining. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024. 2

[25] Puyang Wang, He Zhang, and Vishal M Patel. Generative adversarial network-based restoration of speckled sar images. In *2017 IEEE 7th international workshop on computational advances in multi-sensor adaptive processing (CAMSAP)*, pages 1–5. IEEE, 2017. 2

[26] Puyang Wang, He Zhang, and Vishal M Patel. Sar image despeckling using a convolutional neural network. *IEEE Signal Processing Letters*, 24(12):1763–1767, 2017. 1, 2, 5, 6, 7